MA201 - Séance 1
Estimation paramétrique ponctuelle - Approche fréquentiste
Rappels et compléments

H. Piet-Lahanier - L. Meyer

Content

Estimation statistique

Contexte

- Données collectées associées à un même phénomène
- Peuvent présenter une certaine variabilité
- $\Rightarrow$ représentées comme des variables aléatoires avec une loi de probabilité associée

Objectif

À partir d'un échantillon de mesures du phénomène physique, quelles informations peut-on déduire sur cette loi ?

C'est l'objectif inverse de la théorie des probabilités où la loi est connue et on cherche à donner les caractéristiques des variables qui suivent cette loi.

Estimation statistique

Problématique

- On dispose d'un ensemble de n réalisations $x_1, ..., x_n$ de n variables aléatoires $X_1, ..., X_n$ indépendantes et de même loi.
- On associe à cette loi une densité $f(x, \theta)$ ou une probabilité $\mathbb{P}_\theta(x)$ qui dépend d'un paramètre θ inconnu.
- $\Rightarrow$ Comment calculer ou approcher θ à partir de cet ensemble de réalisations ?

Exemple : Détection de défauts (1/2)

Problématique

Recherche de la proportion d'objets défectueux au sein d'un lot de N objets.

Démarche retenue :

- n tirages indépendants d'un objet parmi les N objets du lot,
- Pour chaque tirage : analyse qualité de cet objet $\Rightarrow$ caractère défectueux ou non de l'objet ?

Question : quelle proportion d'objets défectueux dans le lot total (de N objets) ?

Exemple : Détection de défauts (2/2)

Démarche

- Soit X_i, $i \in \{1, ..., n\}$, la variable aléatoire valant 1 si l'objet issu du i-ème tirage est défectueux, 0 sinon.
- On suppose que les X_i suivent toutes la même loi, une loi de Bernouilli de paramètre θ : $P(X = x) = \theta^x(1 - \theta)^{1-x}$, avec $x \in \{0, 1\}$.
- L'espérance de cette loi est $\mathrm{E}[X] = \theta$
- $\Rightarrow$ Le calcul de la moyenne statistique des réalisations des n v.a. X_i peut fournir une estimation de θ.

Content

Définitions

Definition

1. Population : ensemble d'*individus*.
   Exemple : lot de pièces.
2. Individu : membre d'une population.
   Exemple : une pièce au sein d'un lot.
3. Echantillon : sous-ensemble extrait d'une population
   Exemple : ensemble de n pièces extraites d'un lot.
4. n-échantillon : ensemble $(X_1, ..., X_n)$ de n variables (ou vecteurs) aléatoires réelles (v.a.r.) indépendantes et identiquement distribuées (i.i.d.). Tous les X_i suivent la même loi. Chaque v.a. X_i représente une caractéristique du i-ème individu de l'échantillon.
   Exemple : $X_i \sim \mathcal{B}(p)$, loi de Bernouilli, et vaut 1 si la pièce i est défecteuse, 0 sinon.
5. Observation : réalisation d'une variable aléatoire. On notera x_i l'observation issue de la variable aléatoire X_i.
   Exemple : après avoir effectué une *mesure*, on détermine que la i-ème pièce est défectueuse. On a alors : $x_i = 1$.

Cadre de l'estimation paramétrique

Hypothèses de l'estimation paramétrique :

- Nous supposons disposer d'un n-échantillon de vecteurs aléatoires réels $X_1, ..., X_n$ (en particulier les vecteurs aléatoires sont i.i.d.).
- Le modèle du processus aléatoire est partiellement connu : la structure de la loi est supposée connue, mais un (ou plusieurs) paramètre(s) sont supposés inconnus.

On dispose d'un *modèle paramétrique* donc d'un ensemble $\{\mathbb{P}_\theta; \theta \in \Theta \subset \mathbb{R}^p\}$, où $\mathbb{P}_\theta$ est une loi de probabilité *connue*, de paramètre θ *inconnu*, Θ étant l'ensemble *connu* 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,
- $\{\mathbb{P}_\theta; \theta \in \Theta\} = \{\mathcal{U}_{[-\theta;\theta]}; \theta \in \mathbb{R}_+^*\}$, 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.

Définition d'un estimateur

Définition

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

- Un estimateur du paramètre θ est un vecteur aléatoire réel $\hat{\theta}_n$, fonction du n-échantillon. On définit donc une fonction $F : \mathbb{R}^n \to \mathbb{R}^p$ telle que $\hat{\theta}_n = F(X_1, ..., X_n)$.
  Exemple : $\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ (estimateur de la moyenne empirique du n-échantillon).
- La valeur estimée de θ est fonction des réalisations $x_1, ..., x_n$ des v.a. $X_1, ..., X_n$. C'est donc une réalisation de $\hat{\theta}_n$.
  Exemple : $\mathit{realisation}(\hat{\theta}_n) = \frac{1}{n} \sum_{i=1}^{n} x_i$ (estimation de la moyenne empirique du n-échantillon).

Remarque

La définition précédente est celle d'un estimateur ponctuel. On peut aussi s'intéresser à la recherche d'un intervalle de confiance tel que le paramètre s'y trouve avec une probabilité donnée -> Estimation par intervalle de confiance (hors cadre de ce cours).

Définition d'un estimateur (cont.)

Exemples

Exemples d'estimateurs pour le paramètre θ d'un n-échantillon $(X_1, ..., X_n)$ qui suit une loi normale $\mathcal{N}(\theta, 1)$:

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

Définition d'un estimateur (cont.)

Objectif de l'estimation

Une statistique d'un *n*-échantillon est une variable aléatoire $F(X_1, ..., X_n)$ où F est une fonction.
Un estimateur $\hat{\theta} = F(X_1, ..., X_n)$ du paramètre θ est une statistique à valeurs dans le domaine de définition de θ
L'estimation consiste à déterminer l'estimateur qui associe à un *n*-échantillon, une variable aléatoire *proche* (en un sens à définir) du paramètre θ caractérisant la loi de probabilité à l'origine de la réalisation de ce *n*-échantillon.

Questions liées à l'estimation paramétrique ponctuelle

- Comment construire la fonction F ? Sur quels critères ?
- Comment évaluer la qualité d'un estimateur ?
- Comment étudier les propriétés de l'erreur d'estimation $\hat{\theta} - \theta$, où $\hat{\theta} = F(X_1, ..., X_n)$ est l'estimé de θ le paramètre réel ?
- Quelles sont les caractéristiques principales de l'erreur d'estimation qui qualifient la qualité du choix de F ?

Consistance d'un estimateur

Définition

Un estimateur $\hat{\theta}_n$ est dit *consistant* pour la valeur $\theta \in \Theta \subset \mathbb{R}^p$ (ou *convergent* vers θ) si $\hat{\theta}_n$ tend vers θ presque sûrement (par rapport à la probabilité $\mathbb{P}_\theta$) 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 : L'estimateur $\hat{\theta}$ est une variable aléatoire.

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

Notations

Soit X un vecteur aléatoire. On note $\mathrm{E}[X]$ son espérance mathématique, $Cov(X) = \mathrm{E}[(\mathrm{E}[X] - X)(\mathrm{E}[X] - X)^T]$ sa matrice de covariance, et $Var(X) = tr(Cov(X))$.

Définition

Le *biais* d'un estimateur $\hat{\theta}_n$ de $\theta \in \Theta$ 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 $\mathcal{N}(\theta, 1)$:

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

Exemples (cont.)

- $\hat{\theta}^{(4)} = \frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{n^2}$,
  $\mathrm{E}[\hat{\theta}^{(4)}] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[X_i] + \frac{1}{n^2} = \frac{1}{n}\sum_{i=1}^{n}\theta + \frac{1}{n^2} = \theta + \frac{1}{n^2} => B(\hat{\theta}^3, \theta) = \frac{1}{n^2} =>$ 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 :

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

$$Cov(\tilde{\theta}_n) \geq Cov(\hat{\theta}_n), \tag{1}$$

où, pour deux matrices symétriques M et N, on note $M \geq N$ si la matrice $M - N$ est positive.

Propriété

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

Définition

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

Définition de la vraisemblance

Définition

Dans un modèle paramétrique, 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 $f_\theta(x_1, ..., x_n)$:

$$\theta \mapsto L_n(\theta; x_1, ..., x_n) = f_\theta(x_1, ..., x_n) = \prod_{i=1}^{n} f_\theta(x_i). \tag{2}$$

Dans le cas discret, on a :

$$\theta \mapsto L_n(\theta; x_1, ..., x_n) = \mathbb{P}_\theta(x_1, ..., x_n) = \prod_{i=1}^{n} \mathbb{P}_\theta(X = x_i). \tag{3}$$

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

$$l_n(\theta; x) = \ln(L_n(\theta; x_1, ..., x_n)) = \ln(f_\theta(x_1, ..., x_n)) = \sum_{i=1}^{n} \log f_\theta(x_i). \tag{4}$$

Définition de la vraisemblance (cont.)

Les dernières égalités de (2) et de (3) 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.

Théorème de Cramer-Rao

Théorème

Soit un n-échantillon $(X_1, ..., X_n)$ de loi $\mathbb{P}_\theta$ avec $\theta \in \Theta$ paramètre inconnu.
Notons $L_n(\theta; x_1, ..., x_n)$ sa vraisemblance.
Supposons vérifiée la condition de régularité suivante : $\mathrm{E}[\frac{\partial \ln L_n(\theta; X_1, \ldots, X_n)}{\partial \theta}] = 0$.
Alors, quel que soit l'estimateur non biaisé $\hat{\theta}_n$ de θ, sa matrice de covariance vérifie :

$$Cov(\hat{\theta}_n) \geq M_f(\theta)^{-1}, \tag{5}$$

où M_f est la matrice d'information de Fisher définie par :

$$M_f(\theta) = Cov\left(\frac{\partial \ln L_n(\theta; X_1, \ldots, X_n)}{\partial \theta}\right) = \mathrm{E}\left[\frac{\partial \ln L_n(\theta; X_1, \ldots, X_n)}{\partial \theta} \frac{\partial \ln L_n(\theta; X_1, \ldots, X_n)}{\partial \theta}^T\right]. \tag{6}$$

Théorème de Cramer-Rao (cont.)

D

e plus, s'il existe une fonction F (indépendante de θ) telle que :

$$\frac{\partial \ln L_n(\theta; X_1, ..., X_n)}{\partial \theta} = M_f(\theta)(F(X_1, ..., X_n) - \theta), \tag{7}$$

alors on peut trouver un estimateur non biaisé qui atteint la borne (i.e. tel que $Cov(\hat{\theta}_n) = M_f^{-1}(\theta)$).
Celui-ci est donné par $\hat{\theta}_n = F(X_1, ..., X_n)$, qui est donc un estimateur *efficace*.

Exemple de recherche d'estimateur sans biais à minimum de variance

Soit un n-échantillon $X_1, ..., X_n$ suivant une loi normale $\mathcal{N}(\theta, \sigma^2)$, avec σ supposé connu, et $\theta \in \mathbb{R}$ le paramètre inconnu.
La vraisemblance s'écrit :

$$L(\theta; x_1, ..., x_n) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{x_i - \theta}{2\sigma^2}). \tag{8}$$

Ainsi :

$$\frac{\partial \ln f_\theta(X_1, ..., X_n)}{\partial \theta} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \theta). \tag{9}$$

On trouve alors $E[\frac{\partial \ln f_\theta(X_1,...,X_n)}{\partial \theta}] = 0$, et $M_f(\theta) = Cov(\frac{\partial \ln f_\theta(X_1,...,X_n)}{\partial \theta}) = n/\sigma^2$.

Ainsi, en appliquant le théorème précédent, la borne de Cramer-Rao vaut σ^2/n.

De plus, on a : $\frac{\partial \ln f_\theta(X_1,...,X_n)}{\partial \theta} = \frac{n}{\sigma^2} \left[\frac{1}{n} \sum_{i=1}^{n} X_i - \theta \right]$.

Finalement, l'estimateur sans biais de variance minimale, est donné par $\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$, (où l'on reconnait la moyenne empirique).

Risque Quadratique Moyen

Définition

Soit $\hat{\theta}_n$ un estimateur de $\theta \in \Theta \subset \mathbb{R}^p$. On appelle *Risque Quadratique Moyen (RQM)* (ou *Erreur Quadratique Moyenne (EQM)*) la valeur : $RQM(\hat{\theta}_n, \theta) = \mathrm{E}[(\hat{\theta}_n - \theta)^T(\hat{\theta}_n - \theta)]$.

Propriété

$$RQM(\hat{\theta}_n, \theta) = B(\hat{\theta}_n, \theta)^T B(\hat{\theta}_n, \theta) + tr(Cov(\hat{\theta}_n)). \tag{10}$$

Ainsi, pour un estimateur sans biais : $RQM(\hat{\theta}_n, \theta) = tr(Cov(\hat{\theta}_n)$.

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

Estimateur à RQM minimal

Définition

- Un estimateur $\hat{\theta}_n$ est dit *(quadratiquement) préférable* à un estimateur $\tilde{\theta}_n$ pour la valeur $\theta \in \Theta$ si $RQM(\hat{\theta}_n, \theta) \leq RQM(\tilde{\theta}_n, \theta)$.
- 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.

Exemple : Recherche d'un estimateur à RQM minimal

Soit un n-échantillon $X_1, ..., X_n$ suivant une loi normale $\mathcal{N}(\theta, \sigma^2)$, avec σ supposé connu, et $\theta \in \mathbb{R}$ le paramètre inconnu.

On se propose de chercher un estimateur sous la forme

$$\hat{\theta}_n = \frac{a}{n} \sum_{i=1}^{n} X_i, \tag{11}$$

avec un RQM minimal.

Il s'agit donc de déterminer la valeur de a qui permet d'atteindre ce RQM minimal.

Exemple : Recherche d'un estimateur à RQM minimal

On a : $E[\hat{\theta}_n] = a\theta$, et $Var(\hat{\theta}_n) = a^2\sigma^2/n$. D'où :

$$RQM(\hat{\theta}_n, \theta) = (a-1)^2\theta^2 + \frac{a^2\sigma^2}{n}. \tag{12}$$

Ainsi :

$$a_{opt} = argmin_a RQM(\hat{\theta}_n) = \frac{\theta^2}{\theta^2 + \sigma^2/n}, \tag{13}$$

qui dépend de θ inconnu !

Avec ce choix d'estimateur, on a :

$$B(\hat{\theta}_n, \theta) = \frac{\theta\sigma^2}{n\theta^2 + \sigma^2}, \quad Var(\hat{\theta}_n) = \frac{\theta^4 n\sigma^2}{(N\theta^2 + \sigma^2)^2}, \quad RQM(\hat{\theta}_n) = \frac{\theta^2\sigma^2}{n\theta^2 + \sigma^2}. \tag{14}$$

Meilleurs estimateurs ?

Quel(s) critère(s) pour le choix d'un estimateur ?

- (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.

Content

Estimateurs empiriques

Principe

- Estimateurs basés sur les caractéristiques empiriques de l'échantillon.
- Grande importance pratique : de nombreux problèmes résolus de façon plus efficace par un estimateur empirique que par aucun autre des estimateurs présentés dans la suite
- Dans ce cours : nous nous limitons aux deux estimateurs empiriques de base : moyenne et variance empiriques.

Moyenne empirique

Définition

Soit un n-échantillon $(X_1, ..., X_n)$ suivant une loi $\mathbb{P}$ de moyenne et de variance inconnues. L'estimateur de la moyenne empirique s'écrit :

$$\widehat{moy}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{15}$$

Il correspond à la moyenne des variables aléatoires composant l'échantillon.

Propriété

L'espérance de $\widehat{moy}_n$ vaut :

$$\mathrm{E}[\widehat{moy}_n] = \mathrm{E}[X_i], \qquad \forall i \in \{1, ..., n\}. \tag{16}$$

C'est donc un estimateur sans biais de l'espérance de la loi $\mathbb{P}$ (ou de manière équivalente de l'espérance de la variable X_i, pour tout $i \in \{1, ..., n\}$.

Variance empirique

Définition

Soit un n-échantillon $(X_1, ..., X_n)$ suivant une loi $\mathbb{P}$ de moyenne et de variance inconnues. L'estimateur de la moyenne empirique s'écrit :

$$\widehat{var}_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \widehat{moy}_n)^2 \tag{17}$$

Il correspond à la moyenne des variables aléatoires composant l'échantillon.

Propriété

- L'espérance de $\widehat{var}_n$ vaut :

  $$\mathrm{E}[\widehat{var}_n] = \frac{n-1}{n} Var(X_i), \qquad \forall i \in \{1, ..., n\}. \tag{18}$$

  C'est donc un estimateur biaisé de la variance de la loi $\mathbb{P}$ (ou de manière équivalente de la variance de la variable X_i, pour tout $i \in \{1, ..., n\}$.
- L'estimateur $\frac{n}{n-1}\widehat{var}_n$ est un estimateur sans biais de la variance de la loi $\mathbb{P}$.

Content

Estimateur des moments

Principe

Soit un n-échantillon d'observations $(X_1, ..., X_n)$ 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 :

$$m_k(\hat{\theta}_n^{(k)}) = \hat{m}_k, \tag{19}$$

avec pour tout $k \geq 0$:

- $m_k(\theta) = \mathrm{E}_\theta[X_i^k]$, $\forall i \in \{1, ..., n\}$ (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 :

$$\mathrm{E}_\theta[X] \approx \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{20}$$

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

- Mais, à l'ordre 2 :

$$Var_\theta(X) = \mathrm{E}[X^2] - E[X]^2 \approx \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\frac{1}{n}\sum_{i=1}^{n} X_i)^2. \tag{21}$$

  Or : $Var_\theta(X) = \theta$
  Ainsi : $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\frac{1}{n}\sum_{i=1}^{n} X_i)^2$ est aussi un estimateur des moments de θ.

Conclusion : non unicité de l'estimateur des moments d'un paramètre donné.

Propriété

Questions :

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

Propriété

Si $\mathrm{E}[|X|^k] < \infty$, 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 $\hat{\theta}_n$ de l'équation :

$$\mu(\hat{\theta}_n) = \hat{\mu}, \tag{22}$$

avec :

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

Propriété

Si $E[|g(X)|] < \infty$, 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

Estimateur du Maximum de Vraisemblance

Intuition : Choisir comme estimateur la valeur du paramètre qui rend les réalisations du n-échantillon les plus probables.

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 :

$$\hat{\theta}(x_1, ..., x_n) = argmax_{\theta \in \Theta} L_n(\theta; x_1, ..., x_n) = argmax_{\theta \in \Theta} l_n(\theta; x_1, ..., x_n). \tag{23}$$

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 $i \in \{1, ..., n\}$:

$$f_p(x_i) = \mathbb{P}_p(X = x_i) = p^{x_i}(1-p)^{(1-x_i)} \tag{24}$$

La log-vraisemblance s'écrit :

$$\begin{aligned} l_n(p; x_1, ..., x_n) &= \textstyle\sum_{i=1}^n \log f_p(x_i) \\ &= \textstyle\sum_{i=1}^n \left(x_i \log p + (1-x_i)\log(1-p)\right) \\ &= (\textstyle\sum_{i=1}^n x_i)\log p + (n - (\textstyle\sum_{i=1}^n x_i))\log(1-p) \end{aligned} \tag{25}$$

On recherche ensuite p tel que :

$$\frac{\partial l_n(p; x_1, ..., x_n)}{\partial p} = 0, \tag{26}$$

et on trouve finalement :

$$\hat{p} = \frac{1}{n}\sum_{i=1}^n X_i. \tag{27}$$

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 $i \in \{1, ..., n\}$:

$$f_\lambda(x_i) = \lambda \text{exp}^{-\lambda x_i} \tag{28}$$

La log-vraisemblance s'écrit :

$$\begin{array}{l} l_n(p; x_1, ..., x_n) = \sum_{i=1}^n \log f_p(x_i) \\ = \sum_{i=1}^n \left(\log \lambda - \lambda x_i \right) \\ = n \log \lambda - \lambda \sum_{i=1}^n x_i \end{array} \tag{29}$$

On recherche ensuite λ tel que :

$$\frac{\partial l_n(\lambda; x_1, ..., x_n)}{\partial \lambda} = 0, \tag{30}$$

et on trouve finalement :

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n X_i}. \tag{31}$$

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 à l'intérieur de Θ, alors l'estimateur du Maximum de Vraisemblance est consistant.

Discussion

- Avantages
  - Méthode universelle
  - L'estimateur prend ses valeurs dans l'espace des paramètres
  - Bonnes propriétés asymptotiques
- Limites (liées aux difficultés de rechercher un maximum)
  - Existence non nécessairement assurée
  - Existence de plusieurs maximum locaux
  - Instabilité numérique
  - Paramètres cachées... Cf. modèles de mélange

Content

Résumé

Durant ces séances les principaux points abordés ont été les suivants :

- Cadre de l'estimation fréquentiste ponctuelle
- Définition d'un estimateur
- Principales propriétés d'un estimateur : biais, variance, consistance, Risque Quadratique Moyen (RQM)
- Classes d'estimateur : Sans Biais à minimum de Variance / à RQM minimal
- Estimateurs empiriques
- Méthode des moments pour la construction d'estimateurs
- Estimateur du Maximum de Vraisemeblance (EMV)