Programme du cours de Statistique mathématique

- Partie I : Estimation ponctuelle
- Partie II : Tests statistiques

Bibliographie

- Bernard Prum, La démarche statistique (disponible à la BU).
- Milhaud, X., Statistique, Belin.
- Saporta, G., Probabilités, analyse des données et statistiques. Technip.
- Jean-François Delmas : Une introduction aux Probabilités et aux Statistiques - cours et exercices corrigés (disponible en ligne sur sa page web).

Chapitre I

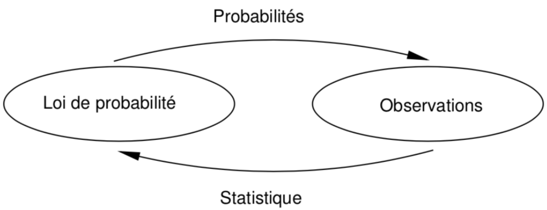
Estimation ponctuelle

Plan

La statistique

- A pour but de fournir un ensemble de méthodes permettant, à partir d'observations,
  - d'analyser de décrire les données observées,
  - mais également d'en déduire des traitements modèles interprétations : aide à la décision, mise en évidence de facteurs explicatifs, prédiction du comportement futur, ...
- Statistique descriptive : Structurer, synthétiser, représenter les données.
- Statistique inférentielle/mathématique :
  - Etendre à toute une population les propriétés constatées sur un échantillon.
  - Confirmer ou infirmer certaines suppositions
  - Prendre des décisions
  - Repose sur la modélisation probabiliste des observations

Les observations sont vues comme des réalisations de variables aléatoires (tirages indépendants d'une variable aléatoire) définie sur un espace probabilisable $(\Omega, \mathcal{F})$



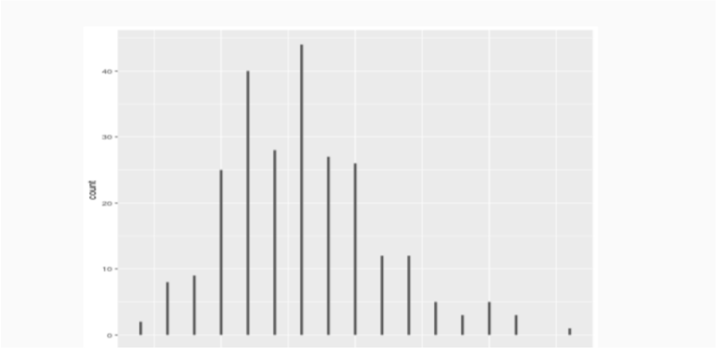
- ☛ La théorie des probabilités vise à évaluer le comportement des observations (espérance, moments, probabilités de dépassement, comportement de sommes,...) étant donné la loi de probabilité $\mathbb{P}$.
- ☛ La statistique fournit des méthodes pour résoudre le problème inverse dit d'inférence statistique : caractériser $\mathbb{P}$ au vu des observations

En général, l'objectif de déterminer complètement $\mathbb{P}$ à partir d'observations est trop ambitieux et il est nécessaire

- de faire des hypothèses plus restrictives sur la loi $\mathbb{P}$; ces hypothèses reflètent nos connaissances à priori sur le processus qui génère les données.
- de considérer des observations dont la structure probabiliste est raisonnablement simple

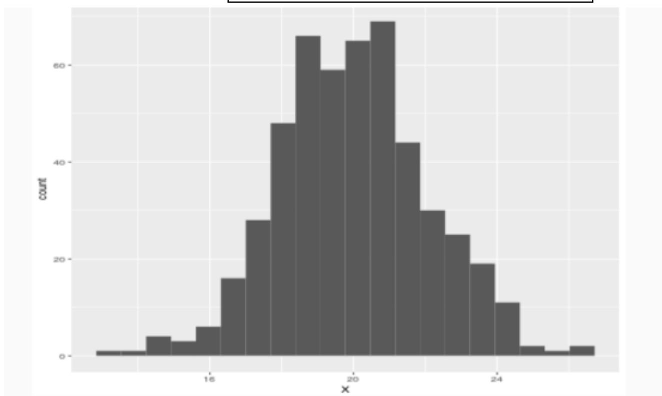
Exemple 1 : Nombre de voitures à un feu rouge

- Afin de mieux gérer la circulation, on s'intéresse au nombre de voitures à un feu rouge sur un créneau donné.
- Expérience : On compte le nombre de voiture dans la file d'attente à chaque fois que le feu passe au vert.
- On récolte $n = 250$ observations : $\boxed{5, 9, 9, 9, 11, 9, \ldots}$



Exemple 2 : Durée d'un trajet

- J'ai une réunion à mon travail à 8h, à quelle heure dois-je partir pour " avoir de grandes chances" d'être à l'heure ?
- Expérience : je mesure la durée de trajet domicile/travail pendant plusieurs jours.
- je récolte $n = 100$ observations : $\boxed{20.87,\ 22.12,\ 20.90,\ 21.33, \ldots}$



Question

Comment utiliser au mieux ces données pour gérer le nombre de voitures au feu/mon heure de départ ?

Quantité d'intérêt

- Il serait intéressant d'avoir de l'information sur la loi de probabilité du nombre de voitures arrêtées au feu à ce créneau/de la durée de trajet domicile-travail.
- On dispose juste de mesures, cette loi est donc inconnue.
- Le travail statistique va donc consister à essayer de reconstruire au mieux cette loi (discrète ou continue) à partir des mesures effectuées.

On introduit un formalisme (mathématique) précis pour représenter (modéliser) ces problèmes.

Plan

Cadre général

- On suppose que les données ont été collectées.
- Ces données sont le résultat d'une expérience répétée n fois.
- On va les noter $x_1, \dots, x_n$ des réalisations de variables aléatoires $X_1, \dots, X_n$.
- $x_1 = X_1(\omega), \dots, x_n = X_n(\omega)$

Hypothèse

On va supposer que les variables sont indépendantes et de même loi de probabilité inconnue $\mathbb{P}$.

Le problème de l'estimation

Il consiste à trouver (estimer) la loi $\mathbb{P}$ à partir de $X_1, \dots, X_n$.

Modèle statistique

Poser un modèle revient à supposer que la loi de probabilité inconnue $\mathbb{P}$ appartient à une famille de lois $(\mathbb{P}_\theta)_{\theta\in\Theta}$ où $\Theta \subseteq \mathbb{R}^d$ avec $d \in \mathbb{N}^*$.

Définition 1

On appelle modèle statistique la donnée du triplet $(\Omega, \mathcal{F}, (\mathbb{P}_\theta))_{\theta\in\Theta}$ *où*

- Ω *est l'espace des observations (l'ensemble dans lequel les observations prennent valeurs) ;*
- $\mathcal{F}$ *est une tribu sur* Ω *;*
- $(\mathbb{P}_\theta)_{\theta\in\Theta}$ *est une famille de probabilités définies sur* $(\Omega, \mathcal{F})$*,* Θ *est appelé espace des paramètres.*

- Ω et $\mathcal{F}$ ne sont généralement pas difficile à caractériser.
- Le statisticien ou le praticien doit par contre choisir une famille de loi de probabilité susceptible de contenir la loi inconnue $\mathbb{P}$.

Exemple

- On souhaite tester l'efficacité d'un nouveau traitement à l'aide d'un essai clinique.
- On traite $n = 100$ patients atteints de la pathologie.
- A l'issue de l'étude, 72 patients sont guéris.

Modélisation

- On note $x_i = 1$ si le i$^{\text{ème}}$ patient a guéri, 0 sinon.
- On suppose que x_i est la réalisation d'une variable aléatoire X_i de loi de Bernoulli de paramètre inconnu $p \in]0,1[$
- Si les individus sont choisis de manière indépendante et ont tous la même probabilité de guérir (ce qui peut revenir à dire qu'ils en sont au même stade de la pathologie), il est alors raisonnable de supposer que les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes.

Spécification du triplet

- $\Omega = \{0, 1\}$.
- $\mathcal{F}$ ensemble des parties de $\{0, 1\}$
- $(\mathbb{P}_\theta)_{\theta \in \Theta} = (\mathcal{B}(p))_{p \in]0,1[}$.

- A travers ce modèle, on suppose que la variable aléatoire X_i qui représente la réaction du i$^{\text{ème}}$ patient au traitement suit une loi de Bernoulli de paramètre inconnu $p \in [0, 1]$.
- Le problème statistique : reconstruire ou estimer ce paramètre à l'aide de l'échantillon $X_1, \ldots, X_n$.

Autres exemples

- Exemple 1 : Nombre de voitures au feu rouge (modèle de Poisson).
- Exemple 2 : Durée de trajet domicile/travail (modèle gaussien)

	Ω	$\mathcal{F}$	$(\mathbb{P}_\theta)_{\theta\in\Theta}$
Exemple 1	$\mathbb{N}$	$\mathcal{P}(\mathbb{N})$	$(\mathcal{P}(\lambda))_{\lambda>0}$
Exemple 2	$\mathbb{R}$	$\mathcal{B}(\mathbb{R})$	$\{\mathcal{N}(m,\sigma^2), m\in\mathbb{R}, \sigma\in\mathbb{R}_+\}$

Le problème statistique sera d'estimer λ ou (m,σ^2) à partir de $X_1,\ldots,X_n$.

Démarche statistique

1. On récolte n observations (n valeurs) $x_1, \ldots, x_n$ qui sont les résultats de n expériences aléatoires indépendantes.
2. Modélisation : on suppose que les n valeurs sont des réalisations de n variables aléatoires indépendantes $X_1, \ldots, X_n$ et de même loi $\mathbb{P}_\theta$. Ce qui nous amène à définir le modèle $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$.
3. Estimation : chercher dans le modèle une loi $\mathbb{P}_{\widehat{\theta}}$ qui soit la plus proche possible de $\mathbb{P}_\theta \Rightarrow$ chercher un estimateur $\widehat{\theta}$ de θ.

Plan

Définition 2

On appelle n−échantillon associé au modèle statistique $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ *une suite* $X_1, \dots, X_n$ *de n variables aléatoires indépendantes et de même loi sous* $\mathbb{P}_\theta$*, pour* $\theta \in \Theta$.

Le statisticien dispose d'une observation de l'échantillon :

$$x_1 = X_1(\omega), \dots, x_n = X_n(\omega),$$

Définition 3

Soit $(X_1, \dots, X_n)$ *un n-échantillon,*

- *On appelle* ***moyenne empirique*** *la v.a.* $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.
- *On appelle* ***variance empirique*** *la v.a.* $V_n^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}_n^2$.

Soit $X_1, \ldots, X_n$ i.i.d. de loi commune $\mathbb{P}_\theta$ avec $\theta \in \Theta$ inconnue. On désigne par $\mathbb{E}_\theta$ et Var_θ l'espérance et la variance respectives sous la loi $\mathbb{P}_\theta$.

Définition 4

- *On appelle statistique toute application de Ω^n dans un espace $\mathbb{R}^p$.*
- *On appelle estimateur du paramètre θ toute statistique à valeurs dans Θ.*

Remarque :

Une statistique ne dépend pas de θ mais seulement de l'échantillon. C'est une variable aléatoire qui a sa propre distribution (qui peut dépendre de θ)

Chercher le "meilleur " estimateur $\widehat{\theta}(X_1, \ldots, X_n)$.
A la fin, calculer l'estimation $\widehat{\theta}(x_1, \ldots, x_n)$ (renvoyé par le logiciel).

Exemple 1 (Traitement)

Les v.a. $\widehat{p}_1 = X_1$ *et* $\widehat{p}_2 = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$ *sont des estimateurs de* p.

Exemple 2

Soit $X_1, \ldots, X_n$ *i.i.d. de loi* $\mathbb{P}_\theta$ *avec* $\theta \in \Theta$ *inconnue avec* $m_\theta = \mathbb{E}_\theta(X)$ *et* $\sigma_\theta^2 = \mathrm{Var}(X_1)$.

- *La moyenne empirique* $\overline{X}_n = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$ *est un estimateur de* $m_\theta = \mathbb{E}_\theta(X)$. $\widehat{\theta}_n = X_1$ *et* $\widetilde{\theta}_n = \sum_{i=1}^{[n/2]} X_{2i}/[n/2]$ *sont aussi des estimateurs de* $m_\theta = \mathbb{E}_\theta(X)$.
- *Le moment empirique simple d'ordre* k, $\widehat{m}_k = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^k$ *est un estimateur de* $m_k(\theta) = \mathbb{E}_\theta(X^k)$, $k \in \mathbb{N}^*$.

Exemple 3

- *Le moment empirique centré d'ordre* k, $\widehat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^k$ *est un estimateur de* $\mu_k(\theta) = \mathbb{E}_\theta\left[(X - \mathbb{E}(X))^k\right]$, $k \in \mathbb{N}^*$.
- *La variance empirique* $V_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ *est un estimateur de* $\sigma_\theta^2 = \mathrm{Var}_\theta(X_1)$.
- *Les valeurs extrêmes de l'échantillon :*

$$\max_{1 \leq i \leq n} X_i \text{ et } \min_{1 \leq i \leq n} X_i$$

.

Biais d'un estimateur

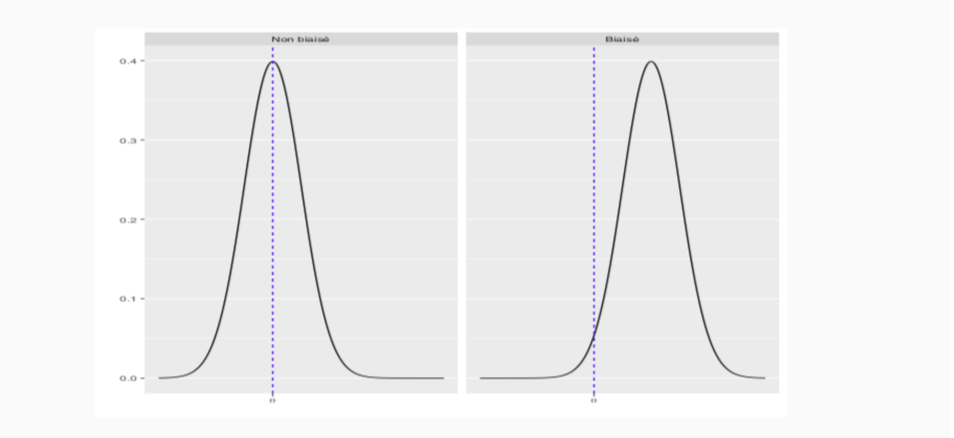
Soit T_n un estimateur de θ, défini à partir d'un échantillon $X_1, \ldots, X_n$ de loi $\mathbb{P}_\theta$. On désigne par $\mathbb{E}_\theta$ l'espérance sous la loi $\mathbb{P}_\theta$.

Définition 5

- *Un estimateur T_n de θ est dit sans biais si $\mathbb{E}_\theta(T_n) = \theta$.*
- *Un estimateur T_n de θ est dit biaisé si $\mathbb{E}_\theta(T_n) \neq \theta$.*
- *La quantité $B_\theta(T_n) = \mathbb{E}_\theta(T_n) - \theta$ est appelé biais de l'estimateur T_n en θ.*
- *Un estimateur T_n de θ est dit asymptotiquement sans biais si $\lim_{n\to+\infty} \mathbb{E}_\theta(T_n) = \theta$.*

- biais= erreur systématique due au fait que T_n fluctue en moyenne autour de $\mathbb{E}_\theta(T_n)$ au lieu de θ.
- Il est souhaitable d'utiliser des estimateurs sans biais.

- On représente ci-dessous les lois de probabilité de 2 estimateurs de θ.



- L'estimateur de gauche semble être préférable à celui de droite.
- Sa loi de probabilité est en effet centrée sur le paramètre inconnu $\Rightarrow \mathbb{E}_\theta[T_n] \approx \theta$.

Biais d'un estimateur

Exemple 4

Soit $X_1, \ldots, X_n$ *i.i.d. de loi commune* $\mathbb{P}_\theta$ *avec* $\theta \in \Theta$ *inconnue où* $\mathbb{E}_\theta(X_1) = m_\theta$ *et* $\mathrm{Var}_\theta(X_1) = \sigma_\theta^2$.

- *La moyenne empirique* $\overline{X}_n$ *est un estimateur sans biais de* m_θ.
- *La variance empirique* $V_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ *est un estimateur biaisé de* σ_θ^2.
- *La variance empirique corrigée* $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ *est un estimateur sans biais de* σ_θ^2.
- V_n^2 *est un estimateur asymptotiquement sans biais de* σ_θ^2.

Risque quadratique

On cherche souvent des estimateurs ayant un biais et une variance faible. Le risque quadratique permet de prendre en compte simultanément ces deux critères.

Définition 6

- *On appelle risque (ou erreur) quadratique moyenne de l'estimateur T_n de θ la quantité*

$$\mathcal{R}(\theta, T_n) = \mathbb{E}_\theta\left[(T_n - \theta)^2\right]$$

- *Soient T_n et T_n' deux estimateurs de θ. On dit que T_n est préférable à T_n' au sens du risque quadratique si*

$$\mathcal{R}(\theta, T_n) < \mathcal{R}(\theta, T_n') \quad \forall \theta \in \Theta.$$

Remarque 1

On a la décomposition suivante

$$\mathcal{R}(\theta, T_n) = Var_\theta(T_n) + B_\theta^2(T_n).$$

Exemple 5

Soit $X_1, \ldots, X_n$ i.i.d. de loi commune $\mathcal{B}(\theta)$, $\theta \in]0,1[$ inconnue. L'estimateur $\overline{X}_n$ de θ est sans biais. Par suite,

$$\mathcal{R}(\theta, \overline{X}_n) = \frac{\theta(1-\theta)}{n}$$

et

$$\mathcal{R}(\theta, X_1) = \theta(1-\theta)$$

Convergence en moyenne quadratique

Soit T_n un estimateur de θ, défini à partir d'une observation de loi $\mathbb{P}_\theta$.

Définition 7

On dit que T_n est un estimateur convergent en moyenne quadratique de θ lorsque

$$\lim_{n\to\infty} \mathcal{R}(\theta, T_n) = 0$$

Exemple 6

On considère $X_1, \ldots, X_n$ i.i.d. de loi commune $\mathbb{P}_\theta$ avec $m_\theta = \mathbb{E}_\theta(X_1)$ et de variance $\mathrm{Var}_\theta(X_1)$ inconnus. Alors, la moyenne empirique $\overline{X}_n$ converge en moyenne quadratique vers θ.

Estimateur convergent

Soit T_n un estimateur de θ, défini à partir d'une observation de loi $\mathbb{P}_\theta$.

Définition 8

- *T_n est consistent si T_n converge en probabilité vers θ quand $n \to \infty$, i.e.*

$$\forall \theta \in \Theta, \forall \varepsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}_\theta \left[|T_n - \theta| > \varepsilon \right] = 0$$

- *T_n est fortement consistent si T_n converge presque sûrement vers θ quand $n \to \infty$:*

$$\forall \theta \in \Theta, \ \ \mathbb{P}_\theta \left(\lim_{n\to\infty} T_n = \theta \right) = 1$$

Propriété 1

T_n converge en moyenne quadratique vers θ implique que T_n converge en probabilité vers θ.

En effet, l'inégalité de Bienaymé Tchebychev implique

$$\forall \varepsilon > 0, \ \ \mathbb{P}\left[|T_n - \theta| > \varepsilon \right] \leq \frac{\mathcal{R}(\theta, T_n)}{\varepsilon^2} \to 0.$$

Exemple 7

On considère $X_1, \ldots, X_n$ i.i.d. de loi commune $\mathbb{P}_\theta$ avec $m_\theta = \mathbb{E}_\theta(X_1)$ et de variance $\sigma_\theta^2 = \mathrm{Var}_\theta(X_1)$ inconnus. Alors,

- *La moyenne empirique $\overline{X}_n$ est un estimateur fortement consistent de m_θ.*
- *La variance empirique $V_n^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ est un estimateur fortement de la variance σ_θ^2.*

Illustration 1

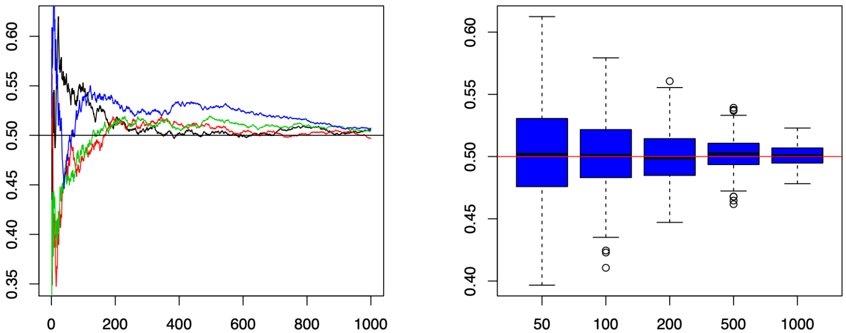


FIGURE – Evolution de $\overline{x}_n$ par rapport à n pour 4 échantillons (à gauche) et boxplots pour les $\overline{x}_n$ obtenus pour $n = 50,\ 100,\ 200,\ 500, 1000$ à l'aide de 4000 échantillons

Illustration 2

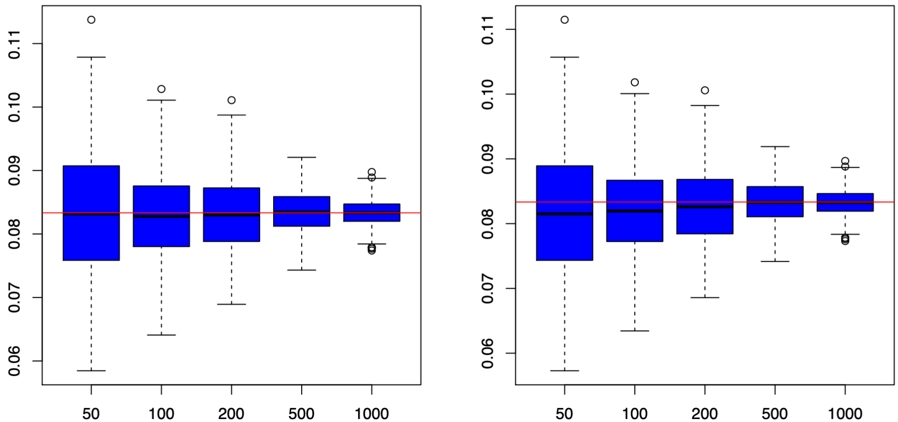


FIGURE – Boxplots pour l'estimateur non biaisé S_n^2 (à gauche) et biaisé V_n^2 (à droite)

Exemple 8

Modèle gaussien : On considère $\mathbf{X} = (X_1, \ldots, X_n)$ *où les* X_i *sont i.i.d et* $X_1 \rightsquigarrow \mathcal{N}(m, \sigma^2)$ *où* $(m, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$ *avec* m *et* σ^2 *inconnus.*
On a S_n^2 *et* V_n^2 *sont des estimateurs consistants de* σ^2 *et*

$$\begin{aligned} \mathcal{R}(\sigma^2, S_n^2) &= \frac{2\sigma^4}{n-1} \\ \mathcal{R}(\sigma^2, V_n^2) &= \frac{2n-1}{n^2}\sigma^4. \end{aligned}$$

V_n^2 *est préférable à* S_n^2 *au sens du risque quadratique.*

Exercice

On suppose que la durée de vie moyenne d'un modèle de téléphones portables peut être modélisée par une v.a. $X \rightsquigarrow \mathcal{E}(\frac{1}{\theta})$ où $\theta > 0$ est inconnu. On cherche à estimer la durée de vie moyenne θ. Pour cela, on considère un échantillon $X_1, \ldots, X_n$ de n portables choisis au hasard.

1. Soit $T_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Montrer que T_n est un estimateur sans biais de θ et calculer son risque quadratique.
2. On pose $Y_n = \min_{1 \leq i \leq n}(X_i)$. Déterminer la f.d.r. de Y_n et en déduire sa loi.
3. Soit $\widehat{T}_n = nY_n$. Montrer que $\widehat{T}_n$ est un estimateur sans biais de θ. Calculer son risque quadratique.
4. Comparer les deux estimateurs.

Plan

Estimation par maximum de vraisemblance

Exemple 9

On cherche à évaluer la fréquence des jours où l'indice ATMO (mesurant la qualité de l'air) à Lille dépasse le niveau 8. On choisit au hasard 10 *jours et on regarde si le niveau 8 est dépassé. On considère les v.a* X_i *avec*

$$X_i = \begin{cases} 1 \text{ si la qualité de l'air dépasse le niveau 8 au jour } i \\ 0 \text{ sinon.} \end{cases}$$

On obtient les réponses suivantes : Non, Oui, Oui, Non, Oui, Oui, Oui, Non, Non, Oui (4 fois Non et 6 fois Oui). le paramètre $\theta = \mathbb{P}(X = 1)$ *et on observe*

$$(x_1, \ldots, x_{10}) = (0, 1, 1, 0, 1, 1, 1, 0, 0, 1).$$

$$\mathbb{P}(X_1 = 0, X_2 = 1, \ldots, X_{10} = 1) = \theta^6(1-\theta)^4$$

θ	0.2	0.3	0.4	0.5	0.6	0.7	0.8
$\theta^6(1-\theta)^4 \times 10^4$	0.26	1.75	5.31	9.77	11.94	9.53	4.19

Il est naturel de choisir comme estimation de θ*, celle pour laquelle la probabilité d'observer l'échantillon est la plus forte i.e.* $\widehat{\theta}_{10} = 0.6$.

- On se place dans le cadre d'un modèle paramétrique $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\Theta})$: On note $\mathbf{X} = (X_1, \ldots, X_n)$ où les X_i sont i.i.d. on note $f(x, \theta)$ la densité de X_1.
- Dans le cas discret, on a pour tout $x \in \mathbb{R}$, $f(x, \theta) = \mathbb{P}_\theta(X = x)$.
- Le but est d'estimer le paramètre θ de la variable aléatoire générique du modèle à partir de la réalisation $(x_1, \ldots, x_n)$ de l'échantillon $\mathbf{X}$.

Idée du maximum de vraisemblance

Estimer le paramètre θ par la valeur qui rend la réalisation de l'échantillon tiré la plus probable.

- Difficulté : Si on a des lois continues, comme des phénomènes modélisées par une loi normale, la réalisation d'un échantillon $x_1, \ldots, x_n$ de $\mathbf{X}$ aura une probabilité nulle. Par exemple :

$$\text{Si} \quad X \rightsquigarrow \mathcal{N}(0,1), \quad \text{alors} \quad \mathbb{P}(X = 2) = 0$$

- On contourne le problème en considérant la densité jointe.

Définition 9

On appelle vraisemblance *de l'échantillon* $\mathbf{X}$ *la fonction* L *définie pour tout* $\theta \in \Theta$ *par :*

- *Dans le cas discret,*

$$\begin{aligned} L(x_1, \ldots, x_n; \theta) &= \mathbb{P}_\theta(X_1 = x_1, \ldots, X_n = x_n) \\ &= \prod_{i=1}^{n} \mathbb{P}_\theta(X_i = x_i). \end{aligned}$$

- *Dans le cas continu,*

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

Commentaires

- La vraisemblance est aléatoire. Une fois que l'échantillon est réalisé, $x_1, \ldots, x_n$ sont des valeurs numériques connues et la vraisemblance ne dépend plus que de la valeur inconnue du paramètre (ou vecteur des paramètres).
  - **Cas discret :** La réalisation de la vraisemblance est la probabilité d'obtenir cette réalisation de l'échantillon.
  - **Cas continue :** La réalisation de la vraisemblance est la densité de $\mathbf{X}$ en la réalisation de l'échantillon.

- Supposons que pour 2 valeurs θ_1 et θ_2 du paramètre, $L(x_1, \ldots, x_n, \theta_1) > L(x_1, \ldots, x_n, \theta_2)$, la valeur du paramètre θ_1 est "plus vraisemblable" que θ_2 étant donné la réalisation $x_1, \ldots, x_n$.

Cas discret

Exemple 10

- *Modèle de Bernouilli : On considère* $\mathbf{X} = (X_1, \ldots, X_n)$ *où les* X_i *sont i.i.d et* $X_1 \rightsquigarrow \mathcal{B}(\theta)$ *avec* $\theta \in]0, 1[$.
- *Retour à l'exemple (ATMO) : la vraisemblance s'écrit* $L\,(0, 1, 1, 0, 1, 1, 1, 0, 0, 1; \theta) = \theta^6(1-\theta)^4$.
- *Modèle de Poisson : On considère* $\mathbf{X} = (X_1, \ldots, X_n)$ *où les* X_i *sont i.i.d et* X_1 *suit une loi de Poisson de paramètre* $\theta > 0$ *.*

Cas continu

Exemple 11

- *Modèle gaussien : On considère* $\mathbf{X} = (X_1, \ldots, X_n)$ *où les* X_i *sont i.i.d et* $X_1 \rightsquigarrow \mathcal{N}(\theta, 1)$, $\theta \in \mathbb{R}$.

- *Modèle exponentiel : On considère* $\mathbf{X} = (X_1, \ldots, X_n)$ *où les* X_i *sont i.i.d et* $X_1 \rightsquigarrow \mathcal{E}(\theta)$ *avec* $\theta > 0$.

- *On considère* $\mathbf{X} = (X_1, \ldots, X_n)$ *où les* X_i *sont i.i.d et* X_1 *admet pour densité* $f(\cdot, \theta)$ *définie pour tout* $x \in \mathbb{R}$ *par*

$$f(x, \theta) = \frac{\theta}{x^2} \mathbb{1}_{\{x \geq \theta\}}, \text{ avec } \theta > 0.$$

Définition 10 (Estimateur par maximum de vraisemblance)

On appelle estimateur par maximum de vraisemblance (EMV) du paramètre θ un estimateur $\theta_n \in \Theta$ défini par

$$L(\mathbf{X}, \theta_n) = \sup_{\theta \in \Theta} L(\mathbf{X}, \theta).$$

Remarque 2

Comme la fonction $\log$ *est strictement croissante de* $\mathbb{R}_+^*$ *dans* $\mathbb{R}$, $\sup_{\theta \in \Theta} L(\mathbf{x}, \theta) = \sup_{\theta \in \Theta} \log L(\mathbf{x}, \theta)$.

Définition 11

La fonction $\ell(\mathbf{x}, \theta) = \log L(\mathbf{x}, \theta)$ *est appelée la log-vraisemblance :*

$$\ell(\mathbf{x}, \theta) = \sum_{i=1}^{n} \log f(x_i, \theta)$$

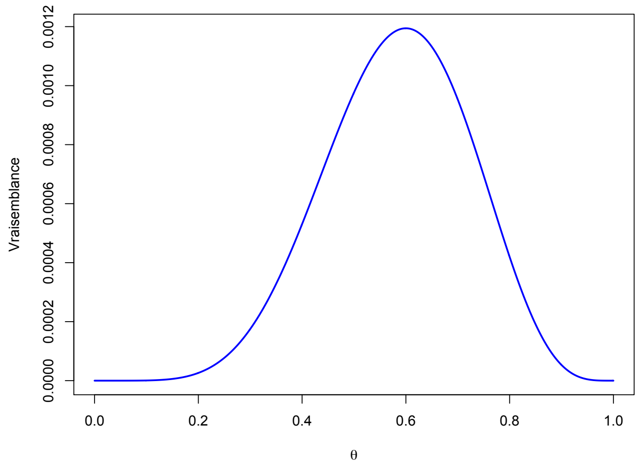
Cas discret

Exemple 12

- *Modèle de Bernouilli : On considère $\mathbf{X} = (X_1, \ldots, X_n)$ où les X_i sont i.i.d et $X_1 \rightsquigarrow \mathcal{B}(\theta)$ avec $\theta \in]0, 1[$.*

- *Modèle de Poisson : On considère $\mathbf{X} = (X_1, \ldots, X_n)$ où les X_i sont i.i.d et X_1 suit une loi de Poisson de paramètre $\theta > 0$.*

Reprenons l'exemple de la qualité de l'air ATMO



Cas continu

Exemple 13

- *Modèle gaussien : On considère $\mathbf{X} = (X_1, \ldots, X_n)$ où les X_i sont i.i.d et $X_1 \rightsquigarrow \mathcal{N}(\theta, 1)$, $\theta \in \mathbb{R}$.*

- *Modèle exponentiel : On considère $\mathbf{X} = (X_1, \ldots, X_n)$ où les X_i sont i.i.d et $X_1 \rightsquigarrow \mathcal{E}(\theta)$ avec $\theta > 0$.*

- *On considère $\mathbf{X} = (X_1, \ldots, X_n)$ où les X_i sont i.i.d et X_1 admet pour densité $f(\cdot, \theta)$ définie pour tout $x \in \mathbb{R}$ par*

$$f(x, \theta) = \frac{\theta}{x^2} \mathbb{1}_{\{x \geq \theta\}},$$

avec $\theta > 0$.

Exemple 14

On considère $\mathbf{X} = (X_1, \ldots, X_n)$ *où les* X_i *sont i.i.d et* $X_1 \rightsquigarrow \mathcal{N}(m, \sigma^2)$, $\theta = (m, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$ *(*m *et* σ^2 *sont inconnus). La log-vraisemblance est*

$$\ell(\mathbf{x}, m, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - m)^2.$$

Si on pose :

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \quad \text{et} \quad \widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2,$$

alors $\widehat{\theta}_n = (\overline{X}_n, \widehat{\sigma}_n^2)$ *est l'EMV du paramètre* $\theta = (m, \sigma^2)$.

Exercice

On rappelle qu'une variable aléatoire X suit une loi géométrique de paramètre θ, $\theta > 0$ si

$$\mathbb{P}_\theta(X = k) = \theta(1-\theta)^{k-1}.$$

Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi géométrique de paramètre $\theta > 0$.

1. Déterminer la fonction de Log-vraisemblance associée à cet échantillon.
2. Calculer l'estimateur du maximum de vraisemblance de θ.

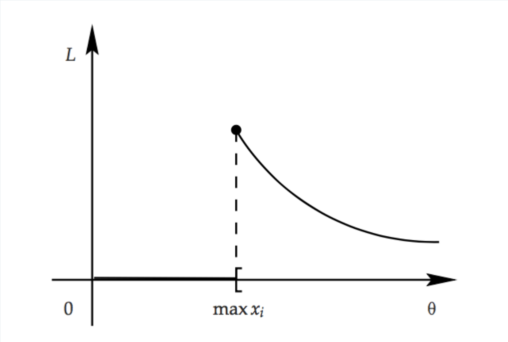
Remarque 3

- ☛ *L'EMV n'existe pas toujours car la maximisation se fait sur un ensemble ouvert* Θ.
- ☛ *La vraisemblance n'est pas a priori dérivable en tout point* $\theta \in \Theta$.
- ☛ *Il n'y a aucune raison pour que l'EMV soit sans biais.*
- ☛ *L'EMV n'a aucune raison d'être unique.*

Exemple 15

On considère $\mathbf{X} = (X_1, \ldots, X_n)$ *où les* X_i *sont i.i.d et* $X_1 \rightsquigarrow \mathcal{U}_{[0,\theta]}$ *avec* $\theta > 0$. *La vraisemblance s'écrit*

$$L(\theta) = \frac{1}{\theta^n} \mathbb{1}_{\{0 \leq \max_{1 \leq i \leq n} x_i \leq \theta\}}$$



$\widehat{\theta}_n = \max_{1 \leq i \leq n} X_i$ *est l'EMV de* θ.