Démarche statistique

M2R santé publique

Université Paris-Sud

6 octobre 2017

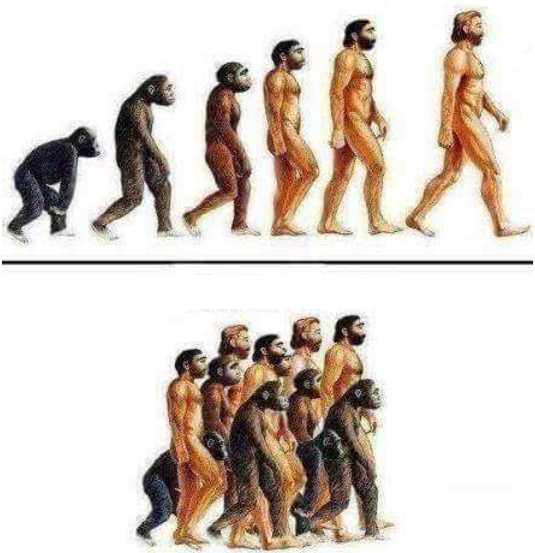
Contenu

Estimateurs
- Statistique, échantillon
- Erreur quadratique moyenne
- Estimateur sans biais
- Estimateur de variance minimale

Méthodes d'estimation
- Méthode des moments
- Estimation par maximum de vraisemblance
- Information et efficacité
- Étude asymptotique du MLE

Une loi de probabilité est un modèle

"all models are wrong, but some are useful!" Georges Box.



La statistique

Inférence statistique

- Objectif et principes
  - Se faire une idée précise sur un ou plusieurs paramètres (caractéristiques de la population).
  - Nécessité de collecter des données sur la population de l'étude (l'échantillon).
  - Des conclusions peuvent être fondées sur des diverses quantités calculées à partir de l'échantillon.
    - Par exemple, soit μ (un paramètre) représentant la durée d'une anesthésie (de courte durée). Nous avons $n = 10$ observations $x_1, \ldots, x_{10}$. La durée moyenne dans l'échantillon est $\bar{x}$ peut être utilisée pour se faire une idée sur la valeur de μ. De manière similaire, si σ^2 est la variance de la distribution des durées (variance de la population, un autre paramètre), la valeur de de s^2 peut être utilisée pour inférer σ^2.
- Notations
  - On utilise les lettres grecques comme notations conventionnelles et génériques pour désigner les paramètres.
  - Nous adoptons θ comme notation pour désigner le paramètre d'un modèle générique.

Définition d'une statistique

Une ***statistique*** est une quantité calculée à partir des données. C'est donc une **variable aléatoire**. On notera donc en minuscule, la réalisation ou l'observation d'une statistique.

Définition d'un échantillon

Les variables aléatoires $X_1, \ldots, X_n$ forment un ***échantillon aléatoire*** de taille n si

- Les X_i sont indépendantes.
- Les X_i ont la même loi de probabilité.

On parlera ainsi, de variables aléatoires iid (indépendantes identiquement distribuées).

Définition d'un échantillon

Les variables aléatoires $X_1, \ldots, X_n$ forment un ***échantillon aléatoire*** de taille n si

- Les X_i sont indépendantes.
- Les X_i ont la même loi de probabilité.

On parlera ainsi, de variables aléatoires iid (indépendantes identiquement distribuées).

Exemple : Le temps de passage d'un client à la caisse d'un magasin suit une loi exponentielle de paramètre λ. On s'intéresse au temps de passage de deux clients qu'on notera $T_2 = X_1 + X_2$. Les deux clients sont indépendants. Peut-on dire que T_2 est une statistique ? Calculer la loi de T_2, son espérance, sa variance.

Notations

Dans la suite, on notera f_θ ou $f(x;\theta)$ la densité (ou la fonction de masse) d'une variables aléatoire. θ désignera le paramètre de la loi.

Notations

Dans la suite, on notera f_θ ou $f(x;\theta)$ la densité (ou la fonction de masse) d'une variables aléatoire. θ désignera le paramètre de la loi.

Exemple : préciser le paramètre θ pour chacune des lois suivantes : Loi de Bernoulli, binomiale, géométrique, Poisson, exponentielle, gamma et normale.

Définition d'un estimateur

L'***estimateur*** d'un paramètre θ est une valeur qui peut être considérée comme *raisonnable* de θ. Un estimateur est le résultat d'une *statistique* évaluée sur un jeu de données (*échantillon*).

Définition d'un estimateur

L'***estimateur*** d'un paramètre θ est une valeur qui peut être considérée comme *raisonnable* de θ. Un estimateur est le résultat d'une *statistique* évaluée sur un jeu de données (*échantillon*).

Exemple : 20 observations de la tension de rupture diélectrique de morceaux de résine époxy. On suppose que la tension de rupture est gaussienne de moyenne μ inconnue.

24.46	25.61	26.25	26.42	26.66
27.15	27.31	27.54	27.74	27.94
27.98	28.04	28.28	28.49	28.50
28.87	29.11	29.13	29.50	30.88

Définition d'un estimateur

L'***estimateur*** d'un paramètre θ est une valeur qui peut être considérée comme *raisonnable* de θ. Un estimateur est le résultat d'une *statistique* évaluée sur un jeu de données (*échantillon*).

Exemple : 20 observations de la tension de rupture diélectrique de morceaux de résine époxy. On suppose que la tension de rupture est gaussienne de moyenne μ inconnue.

24.46	25.61	26.25	26.42	26.66
27.15	27.31	27.54	27.74	27.94
27.98	28.04	28.28	28.49	28.50
28.87	29.11	29.13	29.50	30.88

a. L'estimateur $\bar{X}$ dont la valeur est $\bar{x} = \sum x_i/n = 555.86/20 = 27.793$

Définition d'un estimateur

L'***estimateur*** d'un paramètre θ est une valeur qui peut être considérée comme *raisonnable* de θ. Un estimateur est le résultat d'une *statistique* évaluée sur un jeu de données (*échantillon*).

Exemple : 20 observations de la tension de rupture diélectrique de morceaux de résine époxy. On suppose que la tension de rupture est gaussienne de moyenne μ inconnue.

24.46	25.61	26.25	26.42	26.66
27.15	27.31	27.54	27.74	27.94
27.98	28.04	28.28	28.49	28.50
28.87	29.11	29.13	29.50	30.88

a. L'estimateur $\bar{X}$ dont la valeur est $\bar{x} = \sum x_i/n = 555.86/20 = 27.793$

b. L'estimateur $\tilde{X}$ dont la valeur est $\tilde{x} = (27.94 + 27.98)/2 = 27.960$

Définition d'un estimateur

L'***estimateur*** d'un paramètre θ est une valeur qui peut être considérée comme *raisonnable* de θ. Un estimateur est le résultat d'une *statistique* évaluée sur un jeu de données (*échantillon*).

Exemple : 20 observations de la tension de rupture diélectrique de morceaux de résine époxy. On suppose que la tension de rupture est gaussienne de moyenne μ inconnue.

24.46	25.61	26.25	26.42	26.66
27.15	27.31	27.54	27.74	27.94
27.98	28.04	28.28	28.49	28.50
28.87	29.11	29.13	29.50	30.88

a. L'estimateur $\bar{X}$ dont la valeur est $\bar{x} = \sum x_i/n = 555.86/20 = 27.793$

b. L'estimateur $\tilde{X}$ dont la valeur est $\tilde{x} = (27.94 + 27.98)/2 = 27.960$

c. L'estimateur $\bar{X}_e = [\min(X_i) + \max(X_i)]/2 =$ *the midrange*, dont la valeur $\bar{x}_e = (24.46 + 30.88)/2 = (24.46 + 30.88)/2$

Définition d'un estimateur

L'***estimateur*** d'un paramètre θ est une valeur qui peut être considérée comme *raisonnable* de θ. Un estimateur est le résultat d'une *statistique* évaluée sur un jeu de données (*échantillon*).

Exemple : 20 observations de la tension de rupture diélectrique de morceaux de résine époxy. On suppose que la tension de rupture est gaussienne de moyenne μ inconnue.

24.46	25.61	26.25	26.42	26.66
27.15	27.31	27.54	27.74	27.94
27.98	28.04	28.28	28.49	28.50
28.87	29.11	29.13	29.50	30.88

a. L'estimateur $\bar{X}$ dont la valeur est $\bar{x} = \sum x_i/n = 555.86/20 = 27.793$

b. L'estimateur $\tilde{X}$ dont la valeur est $\tilde{x} = (27.94 + 27.98)/2 = 27.960$

c. L'estimateur $\bar{X}_e = [\min(X_i) + \max(X_i)]/2 =$ *the midrange*, dont la valeur $\bar{x}_e = (24.46 + 30.88)/2 = (24.46 + 30.88)/2$

d. L'estimateur $\bar{X}_{\mathrm{tr}(10)}$, (*the* 10% *trimmed mean*) dont la valeur est

$$\bar{x}_{\mathrm{tr}(10)} = \frac{555.86 - 24.46 - 25.61 - 29.50 - 30.88}{16} = 27.838$$

Erreur quadratique moyenne

Idéalement, on veut que $\widehat{\theta} = \theta$. On sait que $\widehat{\theta}$ est une fonction des X_i (l'échantillon) donc pour certains échantillons $\widehat{\theta} \geq \theta$ et pour d'autres $\widehat{\theta} \leq \theta$.

Si on écrit

$$\widehat{\theta} = \theta + \text{ erreur d'estimation}$$

Naturellement, quantifier la proximité entre $\widehat{\theta}$ et θ revient à mesurer $|\widehat{\theta} - \theta|$ ou $(\widehat{\theta} - \theta)^2$

Erreur quadratique moyenne

Idéalement, on veut que $\widehat{\theta} = \theta$. On sait que $\widehat{\theta}$ est une fonction des X_i (l'échantillon) donc pour certains échantillons $\widehat{\theta} \geq \theta$ et pour d'autres $\widehat{\theta} \leq \theta$.

Si on écrit

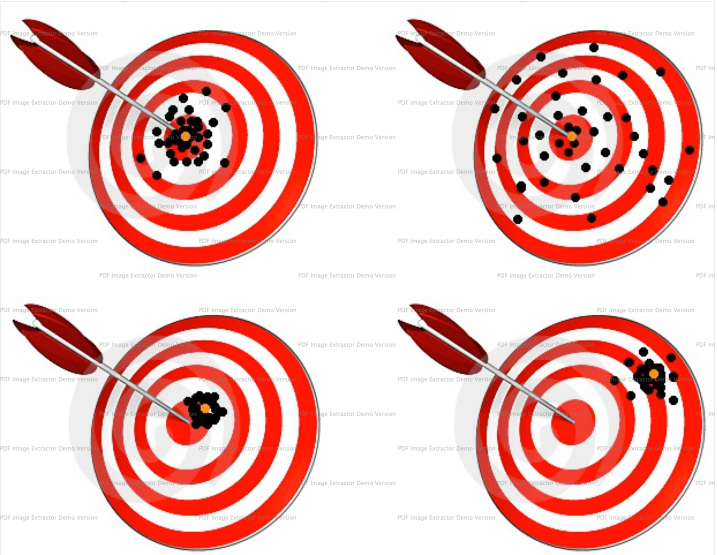
$$\widehat{\theta} = \theta + \text{ erreur d'estimation}$$

Naturellement, quantifier la proximité entre $\widehat{\theta}$ et θ revient à mesurer $|\widehat{\theta} - \theta|$ ou $(\widehat{\theta} - \theta)^2$

Erreur quadratique moyenne notée **MSE**
Le **MSE** d'un estimateur $\widehat{\theta}$ est $\mathbb{E}\big[(\widehat{\theta} - \theta)^2\big]$, où

$$\begin{aligned}
\textbf{MSE} &= \mathbb{E}\Big[\big(\widehat{\theta} - \theta\big)^2\Big] \\
&= \mathbb{E}\Big[\big(\widehat{\theta} - \mathbb{E}(\widehat{\theta}) + \mathbb{E}(\widehat{\theta}) - \theta\big)^2\Big] \\
&= \mathbb{E}\left[\big(\widehat{\theta} - \mathbb{E}(\widehat{\theta})\big)^2 + \big(\mathbb{E}(\widehat{\theta}) - \theta\big)^2 + 2\big(\widehat{\theta} - \mathbb{E}(\widehat{\theta})\big)\big(\mathbb{E}(\widehat{\theta}) - \theta\big)\right] \\
&= \text{ la variance de l'estimateur } + \text{biais}^2
\end{aligned}$$

Erreur quadratique moyenne



Estimateurs sans biais

Un estimateur $\widehat{\theta}$ est dit ***sans biais*** pour θ si $\mathbb{E}(\widehat{\theta}) = \theta$. Si $\mathbb{E}(\widehat{\theta}) \neq \theta$ alors la différence $\mathbb{E}(\widehat{\theta}) - \theta$ est appelée ***biais*** de $\widehat{\theta}$.

Estimateurs sans biais

Un estimateur $\widehat{\theta}$ est dit ***sans biais*** pour θ si $\mathbb{E}(\widehat{\theta}) = \theta$. Si $\mathbb{E}(\widehat{\theta}) \neq \theta$ alors la différence $\mathbb{E}(\widehat{\theta}) - \theta$ est appelée ***biais*** de $\widehat{\theta}$.

Lorsque $X \sim \mathcal{B}(n, p)$, la proportion $\hat{p} = X/n$ est un estimateur sans biais de p.

Estimateurs sans biais

Un estimateur $\widehat{\theta}$ est dit ***sans biais*** pour θ si $\mathbb{E}(\widehat{\theta}) = \theta$. Si $\mathbb{E}(\widehat{\theta}) \neq \theta$ alors la différence $\mathbb{E}(\widehat{\theta}) - \theta$ est appelée ***biais*** de $\widehat{\theta}$.

Lorsque $X \sim \mathcal{B}(n, p)$, la proportion $\hat{p} = X/n$ est un estimateur sans biais de p.

Si $X_1, X_2, \ldots, X_n$ est un échantillon provenant d'une distribution de moyenne μ, alors $\bar{X}$ est un estimateur sans biais de μ. Si de plus la distribution est continue et symétrique, alors, tout estimateur $\widetilde{X}$ (basé sur une troncature symétrique de l'échantillon) est aussi sans biais pour μ.

Estimateur de variance minimale

Supposons que $\widehat{\theta}_1$ et $\widehat{\theta}_2$ sont deux estimateurs sans biais de θ. Bien que la distribution de chaque estimateur est centrée sur la vraie valeur de θ, les dispersions des ces distributions autour de la vraie valeur peuvent être différentes.

Estimateur de variance minimale

Supposons que $\widehat{\theta}_1$ et $\widehat{\theta}_2$ sont deux estimateurs sans biais de θ. Bien que la distribution de chaque estimateur est centrée sur la vraie valeur de θ, les dispersions des ces distributions autour de la vraie valeur peuvent être différentes.

Parmi tous les estimateurs sans biais de θ, un estimateur $\widehat{\theta}$ minimisant la variance est appelé **MVUE** (*minimum variance unbiased estimator*) de θ. Un tel estimateur minimise le **MSE**.

Estimateur de variance minimale

Supposons que $\widehat{\theta}_1$ et $\widehat{\theta}_2$ sont deux estimateurs sans biais de θ. Bien que la distribution de chaque estimateur est centrée sur la vraie valeur de θ, les dispersions des ces distributions autour de la vraie valeur peuvent être différentes.

Parmi tous les estimateurs sans biais de θ, un estimateur $\widehat{\theta}$ minimisant la variance est appelé **MVUE** (*minimum variance unbiased estimator*) de θ. Un tel estimateur minimise le **MSE**.

Soit $X_1, X_2, \ldots, X_n$ un échantillon d'une loi normale de paramètres μ et σ. $\hat{\mu} = \bar{X}$ est un estimateur **MVUE sans biais de variance minimale** de μ.

Introduction

- Jusqu'à présent, les estimateurs ponctuels que nous avons introduit ont été obtenus par intuition et / ou via des connaissances *a priori*.
- Nous verrons maintenant deux méthodes **constructives** pour obtenir les estimateurs ponctuels : ***la méthode des moments*** et ***la méthode du maximum de vraisemblance***.
- Bien que les estimateurs par la méthode du maximum de vraisemblance sont généralement préférables à des estimateurs par la méthode des moments en raison de certaines propriétés d'efficacité, ils ont souvent besoin de beaucoup plus de calculs que les estimateurs par la méthode des moments.
- Parfois les deux méthodes donnent des estimateurs sans biais.

Fondement : la loi des grands nombres

Soit $X_1, X_2, \ldots, X_n$ un échantillon d'une variables aléatoire X. La loi des grands nombres assure que lorsque n est grand,

$$\frac{X_1 + \cdots + X_n}{n} \quad \text{est une bonne approximation de} \quad \mathbb{E}(X).$$

De plus, si g est une fonction continue, alors

$$\frac{g(X_1) + \cdots + g(X_n)}{n} \quad \text{est une bonne approximation de} \quad \mathbb{E}(g(X)).$$

Méthode des moments

L'idée de base de cette méthode est d'assimiler certaines caractéristiques de l'échantillon comme la moyenne aux valeurs attendues dans la population correspondante. Ensuite, les estimateurs sont les solutions de ces équations en considérant les paramètres comme inconnues.

Méthode des moments

L'idée de base de cette méthode est d'assimiler certaines caractéristiques de l'échantillon comme la moyenne aux valeurs attendues dans la population correspondante. Ensuite, les estimateurs sont les solutions de ces équations en considérant les paramètres comme inconnues.

Soit $X_1, X_2, \ldots, X_n$ un échantillon issu d'une fonction de masse ou d'une densité $f(x)$. Pour $k = 1, 2, 3, \ldots$, le $k^{\text{ème}}$ ***moment théorique*** (dans la population) des X_i est $\mathbb{E}\big(X^k\big)$. Le $k^{\text{ème}}$ ***moment empirique*** de l'échantillon est donné par $(1/n) \sum_{i=1}^{n} X_i^k$.

Méthode des moments

L'idée de base de cette méthode est d'assimiler certaines caractéristiques de l'échantillon comme la moyenne aux valeurs attendues dans la population correspondante. Ensuite, les estimateurs sont les solutions de ces équations en considérant les paramètres comme inconnues.

Soit $X_1, X_2, \ldots, X_n$ un échantillon issu d'une fonction de masse ou d'une densité $f(x)$. Pour $k = 1, 2, 3, \ldots$, le $k^{\text{ème}}$ ***moment théorique*** (dans la population) des X_i est $\mathbb{E}\big(X^k\big)$. Le $k^{\text{ème}}$ ***moment empirique*** de l'échantillon est donné par $(1/n)\sum_{i=1}^{n} X_i^k$.

Le premier moment théorique est $\mathbb{E}(X) = \mu$ et le premier moment empirique est $\bar{X} = \sum X_i/n$. Le deuxième moment théorique est $\mathbb{E}(X^2)$ et le deuxième moment empirique est $\sum X_i^2/n$.

Méthode des moments : définition

Soit $X_1, X_2, \ldots, X_n$ un échantillon issu d'une loi dont la fonction de masse ou la densité est $f(x; \theta_1, \ldots, \theta_m)$, où $\theta_1, \ldots, \theta_m$ sont les paramètres inconnus. Les estimateurs par la méthode des moments $\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_m$ obtenus en égalisant les m premier moments théoriques aux m premiers moments empiriques. $\widehat{\theta}_1, \ldots, \widehat{\theta}_m$ sont les solutions de ces équations.

Méthode des moments : définition

Soit $X_1, X_2, \ldots, X_n$ un échantillon issu d'une loi dont la fonction de masse ou la densité est $f(x; \theta_1, \ldots, \theta_m)$, où $\theta_1, \ldots, \theta_m$ sont les paramètres inconnus. Les estimateurs par la méthode des moments $\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_m$ obtenus en égalisant les m premier moments théoriques aux m premiers moments empiriques. $\widehat{\theta}_1, \ldots, \widehat{\theta}_m$ sont les solutions de ces équations.

Soit $X_1, X_2, \ldots, X_n$ un échantillon d'une loi exponentielle de paramètre λ. Ici nous n'avons qu'un seul paramètre à estimer, l'estimateur de λ est la solution de l'équation $\mathbb{E}(X) = \bar{X}$. Or on sait que $\mathbb{E}(X) = 1/\lambda$ dans le cas d'une loi exponentielle. Ainsi l'estimateur par la méthode des moments de λ est donné par $\hat{\lambda} = 1/\bar{X}$.

Exemple : paramètres d'une loi gamma

Soit $X_1, \ldots, X_n$ un échantillon issu d'une loi gamma de paramètres α et β, c'est-à-dire

$$f(x; \alpha, \beta) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases} \quad \text{où } \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \mathrm{d}x.$$

On sait que $\alpha > 0$, $\beta > 0$, $\mathbb{E}(X) = \alpha\beta$ et $\mathrm{Var}(X) = \alpha\beta^2$.

Exemple : paramètres d'une loi gamma

Soit $X_1, \ldots, X_n$ un échantillon issu d'une loi gamma de paramètres α et β, c'est-à-dire

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases} \quad \text{où } \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \mathrm{d}x.$$

On sait que $\alpha > 0$, $\beta > 0$, $\mathbb{E}(X) = \alpha\beta$ et $\mathrm{Var}(X) = \alpha\beta^2$.

Proposer des estimateurs de α et β par la méthode des moments.

Exemple : loi binomiale négative

Soit $X_1, \ldots, X_n$ un échantillon issu d'une loi binomiale négative de paramètres r et p. La fonction de masse est donnée par

$$\mathbb{P}\big(X = x\big) = \binom{x+r-1}{x} p^r (1-p)^x \text{ où } x = 0, 1, 2, \ldots$$

On sait que $\mathbb{E}(X) = \frac{r(1-p)}{p}$ et $\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$.
Proposer des estimateurs de r et p par la méthode des moments.

Corriger le biais ! !

Soit $X_1, \ldots, X_n$ un échantillon issu d'une loi uniforme $\mathcal{U}(0, \theta)$. Proposer un estimateur de θ. Calculer son biais. Proposer un estimateur sans biais à partir du précédent.

La vraisemblance

Soit $X_1, X_2, \ldots, X_n$ un échantillon issu d'une loi dont la fonction de masse ou la densité est $f(x; \theta_1, \ldots, \theta_m)$ où $\theta_1, \ldots, \theta_m$ sont les paramètres inconnus. La fonction de masse ou la densité jointe du vecteur $\left(X_1, \ldots, X_n\right)$ est donnée par

$$f(x_1, \ldots, x_n; \theta_1, \ldots, \theta_m) = \prod_{i=1}^{n} f(x_i; \theta_1, \ldots, \theta_m).$$

C'est une fonction des $\theta_1, \ldots, \theta_m$ quand les $X_1, \ldots, X_n$ sont connues. On appelle cette fonction ***la vraisemblance***.

Estimateur par maximum de vraisemblance

Soit $X_1, X_2, \ldots, X_n$ un échantillon issu d'une loi dont la fonction de masse ou la densité est $f(x; \theta_1, \ldots, \theta_m)$, où $\theta_1, \ldots, \theta_m$ sont les paramètres inconnus. Les estimateurs par maximum de vraisemblance **MLE** (*maximum likelihood estimator*) $\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_m$ sont les valeurs de $\theta_1, \ldots, \theta_m$ qui maximisent la vraisemblance, c'est-à-dire

$$f(x_1, \ldots, x_n; \widehat{\theta}_1, \ldots, \widehat{\theta}_m) \geq f(x_1, \ldots, x_n; \theta_1, \ldots, \theta_m) \text{ pour tout } \theta_1, \ldots, \theta_m.$$

Le **MLE** est obtenu en remplaçant les x_i par les X_i.

Remarque fondamentale

La fonction ln est continue et croissante donc maximiser $f(x_1, \ldots, x_n; \theta_1, \ldots, \theta_m)$ en $(\theta_1, \ldots, \theta_m)$ revient à maximiser la fonction

$$\begin{aligned} \ell(x_1, \ldots, x_n; \theta_1, \ldots, \theta_m) &= \ln f(x_1, \ldots, x_n; \theta_1, \ldots, \theta_m) \\ &= \sum_{i=1}^{n} \ln f(x_i; \theta_1, \ldots, \theta_m), \end{aligned}$$

appelée la ***log-vraisemblance***.

Principe d'invariance

Soit $\widehat{\theta}$ l'estimateur par maximum de vraisemblance de θ. Pour toute fonction bijective g, $g\left(\widehat{\theta}\right)$ est l'estimateur par maximum de vraisemblance de $g(\theta)$.

Comportement asymptotique du **MLE**

Sous des conditions générales sur la loi théorique des observations, lorsque la taille de l'échantillon tend vers l'infini, le **MLE** de tout paramètre θ est proche de θ, asymptotiquement sans biais et de variance minimale. Autrement dit, asymptotiquement, le **MLE** est approximativement **MVUE** de θ.

Exemple : processus de Poisson spatial

On modélise le comptage du nombre d'événements dans une région R de surface $a(R)$ par une loi de Poisson de paramètre $\lambda a(R)$ (λ est le nombre moyen d'événements sur une région de surface 1) et l'hypothèse de non chevauchement des régions implique l'indépendance des comptages dans les différentes régions. Des écologistes s'intéressent au comptage du nombre de plantes d'une certaine espèce sur des régions non chevauchantes $R_1, \ldots, R_n$, c'est-à-dire

$$\mathbb{P}\left(X_i = x_i\right) = \frac{\left[\lambda a\left(R_i\right)\right]^{x_i} e^{-\lambda a\left(R_i\right)}}{x_i!}.$$

- Écrire la vraisemblance et la log-vraisemblance.
- Déduire le **MLE** de λ.
- Une modélisation alternative consiste à fixer n points et observer Y_i : la distance entre le $i^{\text{ème}}$ point et la plante la plus proche.
  - Calculer la densité de Y ?
  - Déduire un autre **MLE** de λ basé sur $Y_1, Y_2, \ldots, Y_n$.

Objectif

- Introduire l'idée d'***information de Fisher*** et ses applications.
- Déterminer le minimum de variance dans la classe des estimateurs sans biais.
- Montrer que le **MLE** est asymptotiquement sans biais et normal (quand n tend vers l'infini) de moyenne la vraie valeur du paramètre θ et une variance minimale.

Information apportée par une observation

- On conserve la notation $f(\theta; x)$ pour désigner la fonction de masse ou la densité de probabilité.
- L'information de Fisher mesure l'information (parfois on dit la précision) d'une observation.
- On considère la variable aléatoire U obtenue en prenant la dérivée de $\ln\big[f(x;\theta)\big]$ par rapport à θ en remplaçant x par X :

$$U = \frac{\partial}{\partial\theta} \ln\big[f(X;\theta)\big].$$

Information apportée par une observation

- On conserve la notation $f(\theta; x)$ pour désigner la fonction de masse ou la densité de probabilité.
- L'information de Fisher mesure l'information (parfois on dit la précision) d'une observation.
- On considère la variable aléatoire U obtenue en prenant la dérivée de $\ln\left[f(x;\theta)\right]$ par rapport à θ en remplaçant x par X :

$$U = \frac{\partial}{\partial\theta} \ln\left[f(X;\theta)\right].$$

Par exemple, si $f(x;\theta) = \theta x^{\theta-1}$ où $0 < x < 1(\theta > 0)$, alors

$$\frac{\partial}{\partial\theta} \ln\left[\theta x^{\theta-1}\right] = \frac{\partial}{\partial\theta} \ln\left[\,\ln(\theta) + (\theta-1)\ln(x)\right] = \frac{1}{\theta} + \ln(x),$$

on obtient

$$U = \ln(X) + \frac{1}{\theta}$$

Le score

La variable aléatoire U s'appelle le ***score*** de l'observation X.

Le score

La variable aléatoire U s'appelle le ***score*** de l'observation X.

On sait que

$$\int f(x;\theta)\mathrm{d}x = 1,\ \frac{\partial}{\partial\theta}\int f(x;\theta)\mathrm{d}x = 0 \text{ et } [\ln f]' = \frac{f'}{f},$$

donc

$$\begin{aligned} 0 &= \frac{\partial}{\partial\theta}\int f(x;\theta)\mathrm{d}x = \int \frac{\partial}{\partial\theta} f(x;\theta)\frac{f(x;\theta)}{f(x;\theta)}\mathrm{d}x \\ &= \int \frac{\partial}{\partial\theta}\left[\ln f(x;\theta)\right] f(x;\theta)\mathrm{d}x = \mathbb{E}\left[\frac{\partial}{\partial\theta}\ln\left[f(X;\theta)\right]\right] = \mathbb{E}\big(U\big). \end{aligned}$$

Information de Fisher

L'***information de Fisher*** $\mathcal{I}(\theta)$ d'une observation issue de $f(x;\theta)$ est la variance de la variable aléatoire $U = \dfrac{\partial}{\partial\theta} \ln\left[f(X;\theta)\right]$:

$$\mathcal{I}(\theta) = \mathrm{Var}\left[\frac{\partial}{\partial\theta} \ln\left[f(X;\theta)\right]\right].$$

Information de Fisher

L'***information de Fisher*** $\mathcal{I}(\theta)$ d'une observation issue de $f(x;\theta)$ est la variance de la variable aléatoire $U = \dfrac{\partial}{\partial\theta} \ln\left[f(X;\theta)\right]$:

$$\mathcal{I}(\theta) = \mathrm{Var}\left[\frac{\partial}{\partial\theta} \ln\left[f(X;\theta)\right]\right].$$

On peut montrer qu'on peut faire appel à l'expression

$$\mathcal{I}(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} \ln\left[f(X;\theta)\right]\right]$$

Exemple

Soit X une v.a qui suit une loi de Bernoulli de paramètre θ, *i.e.* $f(x;\theta) = \theta^x(1-\theta)^{1-x}$ où $x = 0, 1$.
Montrons que

$$\mathcal{I}(\theta) = \mathrm{Var}\left[\frac{\partial}{\partial\theta}\ln\left[f(X;\theta)\right]\right] = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ln\left[f(X;\theta)\right]\right] = \frac{1}{\theta(1-\theta)}.$$

Information d'un échantillon (additivité)

On sait que

$$\frac{\partial}{\partial\theta}\ln\left[f(X_1, X_2, \ldots, X_n;\theta)\right] = \frac{\partial}{\partial\theta}\ln\left[f(X_1;\theta)\right] + \frac{\partial}{\partial\theta}\ln\left[f(X_2;\theta)\right] + \cdots + \frac{\partial}{\partial\theta}\ln\left[f(X_n;\theta)\right]$$

Information d'un échantillon (additivité)

On sait que

$$\frac{\partial}{\partial\theta}\ln\left[f(X_1,X_2,\ldots,X_n;\theta)\right] = \frac{\partial}{\partial\theta}\ln\left[f(X_1;\theta)\right] + \frac{\partial}{\partial\theta}\ln\left[f(X_2;\theta)\right] + \cdots + \frac{\partial}{\partial\theta}\ln\left[f(X_n;\theta)\right]$$

On déduit que

$$\mathbb{E}\left[\frac{\partial}{\partial\theta}\ln\left[f(X_1,X_2,\ldots,X_n;\theta)\right]\right] = 0$$

et

$$\begin{aligned}\mathcal{I}_n(\theta) = \mathrm{Var}\left[\frac{\partial}{\partial\theta}\ln\left[f(X_1,X_2,\ldots,X_n;\theta)\right]\right] &= n\mathrm{Var}\left[\frac{\partial}{\partial\theta}\ln\left[f(X_1;\theta)\right]\right] \\ &= n\mathcal{I}(\theta)\end{aligned}$$

Borne de Cramer-Rao

Soit $X_1, X_2, \ldots, X_n$ un échantillon de densité de probabilité $f(x;\theta)$ telle que l'ensemble des valeurs de x ne dépend pas de θ. Si la statistique $T = t(X_1, X_2, \ldots, X_n)$ est un estimateur sans biais du paramètre θ, alors

$$\mathrm{Var}\big(T\big) \geq \frac{1}{\mathcal{I}_n(\theta)}$$

Borne de Cramer-Rao

Soit $X_1, X_2, \ldots, X_n$ un échantillon de densité de probabilité $f(x;\theta)$ telle que l'ensemble des valeurs de x ne dépend pas de θ. Si la statistique $T = t(X_1, X_2, \ldots, X_n)$ est un estimateur sans biais du paramètre θ, alors

$$\mathrm{Var}\big(T\big) \geq \frac{1}{\mathcal{I}_n(\theta)}$$

Soit T un estimateur sans biais de θ. On dit que l'estimateur T est efficace lorsque sa variance atteint la borne de Cramer-Rao. C'est ce qu'on a noté précédemment un **MVUE**.

Loi asymptotique du MLE

Soit un échantillon $X_1, X_2, \ldots X_n$ d'une loi de densité $f(x;\theta)$. Supposons que l'ensemble des valeurs possibles de x ne dépend pas de θ. Lorsque n tend vers l'infini, l'estimateur par maximum de vraisemblance $\widehat{\theta}$ est asymptotiquement normal de moyenne θ et de variance $\mathcal{I}_n(\theta)^{-1} = n^{-1}\mathcal{I}(\theta)$. Autrement dit

$$\sqrt{n}\big(\widehat{\theta} - \theta\big) \longrightarrow \mathcal{N}\big(0, \mathcal{I}(\theta)^{-1}\big).$$

Exemple