Chapitre 3. Estimation Statistique. (1)

1 - Modèle Statistique.

Soit une population dont on veut étudier un caractère.

a - Calcul des probabilités

permet d'organiser le modèle

a) espace de probabilité $(\Omega, \mathcal{A}, P)$

b) une v.a $(\Omega, \mathcal{A}, P) \xrightarrow{X} \mathcal{X}$

c) Le type de distribution de X.

Exemple : Répartition des garçons et filles dans les familles de 7 enfants.

X va qui compte le nbre de filles

$X : (\Omega, \mathcal{P}(\Omega), P) \longrightarrow \{\{0,1,2,\ldots,7\}, \mathcal{P}(1\sim 3), \mathcal{B}(7,p)_{P_X}\}$ $\quad P_X = P(X=k)$.

$\Omega = \{G, F\}$

b - Soit le calcul des statistiques donne p un % de chances.

1 - les calculs statistiques ont pour objet d'informer sur la valeur du paramètre de 2 manières.

* En indiquant un nombre (probable) (Estimation ponctuelle)
* " " " intervalle de Confiance (Est. ensembliste)

c) Soit $X : (\Omega, \mathcal{A}, P) \to \mathcal{X}$ $\qquad \mathcal{X} = X(\Omega)$.

Définition 1 : Un modèle statistique est un système $(\mathcal{X}, \mathcal{B}, f(x,s), s \in S \subset \mathbb{R}^d)$ où $f(x,s)$ est la densité de la loi de probabilité de la va X.

$S \subset \mathbb{R}^d$ ensemble des paramètres.

$(\forall s \in S)$, à $f(x,s)$ on peut associer une probabilité P_s (la loi de proba. de X)

$$dP_s(x) = f(x,s)\,dx. \qquad (x \in \mathcal{X}).$$

$$P_s(C) = \int_C f(x,s)\,dx$$

$$= \sum_{x \in C} f(x,s).$$

Exemples : (1) Modèle statistique pour la va de bernoulli $\Omega \xrightarrow{X} \mathbb{R}$.

$$\Big(\underset{\mathcal{X}}{\underset{\shortparallel}{\{0,1\}}},\ \mathcal{P}(\{0,1\}),\ f(x,p),\ p \in [0,1] \Big)$$

$$\text{où} \quad f(x,p) = (1-p)\left(\frac{p}{1-p}\right)^{x}$$

(2) Modèle statistique pour la va binomiale

$$\Big(\mathcal{X} = \{0,1,2,\dots,n\},\ \mathcal{P}(\mathcal{X}),\ f(x,p),\ p \in [0,1] \Big).$$

$$\text{où} \quad f(x,p) = C_n^{x}\, p^{x} \cdot (1-p)^{n-x}.$$

(3) Modèle statistique pour la va de Poisson.

$$\Big(\mathcal{X} = \mathbb{N},\ \mathcal{P}(\mathbb{N}),\ f(x,\lambda),\ \lambda \in \mathbb{R}^{+} \Big)$$

$$f(x,\lambda) = e^{-\lambda} \cdot \frac{\lambda^{x}}{x!}.$$

(4) Modèle statistique de la loi normale.

$$\Big(\mathcal{X} = \mathbb{R},\ \mathcal{B}_{\mathbb{R}},\ f(x,m,\sigma),\ (m,\sigma) \in \mathbb{R} \times \mathbb{R}^{+} \Big)$$

$$\text{où} \quad f(x,m,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}(x-m)^2}.$$

Rappel :

Une va X est dite complètement centrée si $E_p(X) = 0 \quad \forall p \in S$.

" " " " " " de carré sommable si $E_p(X^2) < +\infty \; \forall p \in S$.

Définition 2 : Soit $X_1, X_2, \dots, X_n$ n va iid

On appelle fonction de vraisemblance d'un n-échantillon, la densité conjointe de n v.a.

$$L(x_1^n, p) = L_{(X_1, X_2, \dots, X_n)}(x_1, \dots, x_n, p) = \prod_{i=1}^{n} f(x_i, p).$$

Remarque : ① Si $(\mathcal{X}, \mathcal{B}, f(x,p), p \in S)$ est le modèle de la va X alors le modèle de l'échantillon de taille n est

$$(\mathcal{X}^n, \mathcal{B}^{(n)}, L(x_1^n, p), p \in S) = (\mathcal{X}, \mathcal{B}, f(x,p), p \in S)^n.$$

Nécessairement

On a $\displaystyle\int_{\mathcal{X}^n} L(x_1^n, p)\, dx_1 dx_2 \dots dx_n = 1$

② $L(x_1^n, p) = P_p(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$

$= P_p(X_1 = x_1)\, P_p(X_2 = x_2) \dots P_p(X_n = x_n)$.

Exemples : ① Si $X \sim \mathcal{P}(\lambda)$.

$(\mathbb{N}^n, \mathcal{P}(\mathbb{N}^n), L(x_1^n, \lambda), \lambda \in \mathbb{R}^+)$

où $L(x_1^n, \lambda) = \prod_{i=1}^{n} f(x_i, \lambda) = \prod_{i=1}^{n} e^{-\lambda} \cdot \frac{\lambda^{x_i}}{(x_i)!} = e^{-n\lambda} \cdot \frac{\lambda^{\sum x_i}}{\prod_{i=1}^{n} (x_i!)}$.

(2) Si $X \rightsquigarrow \mathcal{B}(k, p)$.

$$\left(\{0,1,\dots,k\}^n,\ \mathcal{P}(\{\dots\}^n),\ \mathcal{L}(x_1^n, p),\ p \in [0,1]\right)$$

$$\mathcal{L}(x_1^n, p) = \prod_{i=1}^{n} \left(C_k^{x_i}\, p^{x_i} (1-p)^{k-x_i} \right)$$

$$= p^{\sum_{i=1}^{n} x_i} \cdot (1-p)^{nk - \sum x_i} \cdot \prod_{i=1}^{n} C_k^{x_i}$$

(3) Si $X \rightsquigarrow N(m, \sigma)$.

$$\left(\mathbb{R}^n,\ \mathcal{B}_{\mathbb{R}^n},\ \mathcal{L}(x_1^n, m, \sigma),\ (m,\sigma) \in \mathbb{R} \times \mathbb{R}^+\right)$$

où $\mathcal{L}(x_1^n, m, \sigma) = (2\pi\sigma^2)^{-\frac{n}{2}} \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2}$.

Définition 3 : Soit $X : \Omega \to (\mathcal{X}, \mathcal{B}, P_\theta, \theta \in \Theta)$.

Soit un n-échantillon $(\mathcal{X}^n, \mathcal{B}^n, P_\theta^n, \theta \in \Theta)$.

On appelle statistique (d'échantillonnage), une application, notée T, mesurable de $(\mathcal{X}^n, \mathcal{B}^n) \to (E, \mathcal{E})$

$$(x_1, \dots, x_n) \longmapsto T(x_1, \dots, x_n).$$

$E = \mathbb{R}$ statistique Scalaire
$E = \mathbb{Z}$ " discrète.
$E = \mathbb{R}^d$ " Vectorielle.

Exemples : (1) $X \rightsquigarrow \Gamma(a, 1)$ dont le modèle est

$$\left(\mathbb{R}^+,\ \mathcal{B}_{\mathbb{R}^+},\ f(x,a),\ a \in \mathbb{R}^+\right) \text{ où } f(x,a) = \frac{1}{\Gamma(a)} e^{-x} \cdot x^{a-1}.$$

Son échantillon de taille n $\left(\mathbb{R}^{+n},\ \mathcal{B}_{\mathbb{R}^{+n}},\ \mathcal{L}(x_1^n, a),\ a \in \mathbb{R}^+\right)$

$$\mathcal{L}(x_1^n, a) = (\Gamma(a))^{-n} \cdot e^{-\sum_{i=1}^{n} x_i} \cdot \left(\prod_{i=1}^{n} x_i\right)^{a-1}.$$

Une statistique d'échantillonnage $T : \mathbb{R}_+^n \to \mathbb{R}$
$$(x_1, \dots, x_n) \longmapsto T(x_1^n) = \frac{1}{n}\sum_{i=1}^{n} \log x_i .$$

$$\boxed{T = \frac{1}{n}\sum_{i=1}^{n} \log X_i}$$

2. Soit $Z = (X, Y)$ de modèle $\left(\mathbb{R}_+^2, \mathcal{B}_{\mathbb{R}_+^2}, f(x, y, s), s \in \mathbb{R}^+\right)$

où $f_s(x, y) = \exp(-sx - y/s)$

(Exo : Vérifier que $f_s(x, y)$ est une densité de proba.)

1-échantillon
$$T : \mathbb{R}_+^2 \to \mathbb{R}$$
$$(x, y) \longmapsto T(x, y) = xy - 1 \quad \text{est une statistique.}$$

$$\boxed{T = XY - 1}$$

Remarque : Une statistique ne dépend pas du paramètre.

Définition : Soit X une va
Soit $X_1, \dots, X_n$ un n-échantillon et T une statistique.
$$T : (\mathcal{X}^n, \mathcal{C}^n) \longrightarrow (E, \mathcal{E}).$$
On appelle modèle statistique image par T le modèle statistique la va T $\left(E, \mathcal{E}, P_s^T, s \in S\right)$ où P_s^T est la loi de proba de la va T

(En général T admet une densité de proba. $f_T(t, s)$ où
$$dP_s^T(t) = f(t, s)\, dt.)$$

2. Estimateur Sans biais
Convergent
Exhaustif.

Définition 1 : Un estimateur T du paramètre s du modèle statistique $(\mathcal{X}, \mathcal{B}, f(x,s), s \in S)$ est dit sans biais si

$$E_s(T) = s \qquad \forall s \in S.$$

Exemples : ① Si $X \rightsquigarrow b(p)$ bernoulli

$E(X) = p$ $\qquad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad E_p(\bar{X}) = p$

$\bar{X}$ est un estimateur sans biais de p.

② Si $X \rightsquigarrow \mathcal{B}(n,p)$

$E(X) = np \qquad E_p(\bar{X}) = np$ est biaisé.

③ Si $X \rightsquigarrow \mathcal{P}(\lambda)$.

$T = \frac{n}{n-1} S_n'^2 = S_n^2$ est sans biais de λ car.

$$E_\lambda(T) = \frac{n}{n-1} E(S'^2) = \frac{n}{n-1} \cdot \frac{n-1}{n} \cdot Var_\lambda(X) = \lambda.$$

car
$E(S'^2) = \frac{1}{n}\sum \left(E(X_i^2) - E(\bar{X}^2)\right) = \frac{1}{n}\sum\left(\lambda^2+\lambda - \frac{\lambda}{n} - \lambda^2\right) = \frac{1}{n}\sum \frac{(n-1)}{n}\lambda = \frac{n-1}{n}\lambda.$

Souvent on veut estimer non pas $s \in S$ mais une fonction de s.

Exemple : Si $X \rightsquigarrow \Gamma(a,p) \quad (a,p) \in \mathbb{R}^+ \times \mathbb{R}$.

On sait que $E(X) = \frac{a}{p}$ et $Var X = \frac{a}{p^2}$.

On pourra chercher à estimer non pas a et p mais

$$h(a,p) = \frac{a}{p} \quad \text{et} \quad k(a,p) = \frac{a}{p^2}.$$

Définition 2 : Soit $(\mathcal{X}, \mathcal{B}, f_s, s \in S)$ est un modèle statistique,
on appelle paramètre du modèle une application $h : S \to F$.
En particulier, on retrouve le paramètre initial si $h = Id_E : S \to S$

Définition 3 : Soit $(\mathcal{X}, \mathcal{B}, f_s, s \in S)$ et $h : S \to F$ un paramètre
On appelle estimateur du paramètre h toute statistique $T : \mathcal{X}^n \to F$
Un tel estimateur est dit sans biais si

$$\boxed{\forall s \in S \quad E_s(T) = h(s).}$$

Exemples : (1) $X \rightsquigarrow \Gamma(a, p)$. $\quad (a, p) \in \mathbb{R}^+ \times \mathbb{R}$.

$h : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$
$(a, p) \mapsto \frac{a}{p} = h(a, p)$.

$g : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$
$(a, p) \mapsto g(a, p) = \frac{a}{p^2}$,

Soit $T_1 = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i : \mathcal{X}^n \to \mathbb{R}$ $\quad T_1 = \bar{X}$ est un esti. s.b pour h.
$E(\bar{X}) = \frac{a}{p}$.

$T_2 = \frac{n}{n-1} S_n'^2 = S_n^2 : \mathcal{X}^n \to \mathbb{R}$

On a $E(T^2) = E(S_n^2) = \frac{a}{p^2}$ $\quad T_2$ est. s.b de g.

(2) $X \rightsquigarrow \mathcal{B}(n, p)$ $\quad p \in [0, 1]$.

Soit $h : [0, 1] \to \mathbb{R}$
$p \mapsto h(p) = np$.

$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i : \mathcal{X}^n \to \mathbb{R}$ $\quad \bar{X}$ est un esti. sans biais de h car
$E(\bar{X}) = np$.

et $T = \frac{\bar{X}}{n}$ est un est. s.b de p.

Définition 3: Soit h un paramètre de $(\mathcal{X}, \mathcal{B}, f(x,\theta), \theta \in S)$ (modèle).
Soit T_n un estimateur de h sur l'échantillon de taille n.

(1) On dit que T_n est asymptotiquement sans biais si

$$\forall \theta \in S \qquad E_\theta(T) \xrightarrow[n\to\infty]{} h(\theta) .$$

(2) Si $T_n \xrightarrow{\text{Pr}} h(\theta)$, $\forall \theta$, On dit qu'il est CV en Proba.

$$\Longleftrightarrow \quad P_\theta(\cdot\,|T_n - h(\theta)| > \varepsilon) \xrightarrow[n\to\infty]{} 0$$

Exemple: $X \leadsto N(m, \sigma)$ σ connu.
alors $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ est un est. sans biais de m.

$$\overline{X} \xrightarrow[\text{Ps}]{\text{Pr}} E(X) = m$$

$$\overline{X}_n \xrightarrow[L_2]{} m \qquad \text{car } \underbrace{E_\theta(|\overline{X}_n - m|^2)}_{=\frac{\sigma^2}{n}} \xrightarrow[n\to\infty]{} 0 .$$

Définition 4: Soit $(\mathcal{X}, \mathcal{B}, f_\theta, \theta \in S)$ un modèle pour la va X.
Une statistique $T(X_1, \ldots, X_n)$ est dite exhaustive si la ft ~~est indépendante de θ.~~

~~(2) $E_\theta(X_1, \ldots, X_n \,|\, T = t) = k(t)$~~

~~càd si $Y = (X_1, \ldots, X_n)$~~ $E_\theta(Y|T) = \cancel{k(T, \theta)} = k(T)$

~~Exemple~~:

(1) $P(X | T = t)$. non ft de θ.

Exemples

1. $X \rightsquigarrow b(p)$. $p \in [0,1]$

$$f(x,p) = p^x (1-p)^{1-x}.$$

Soit $T = \sum_{i=1}^{n} X_i$ une statistique.

T est exhaustive car :

$$P\left((X_1 = x_1, \dots, X_n = x) \mid T = t\right) = \frac{P(X_1 = x_1, \dots, X_n = x_n, \sum X_i = t)}{P(\sum X_i = t)}$$

$$= \begin{cases} 0 & \text{si } \sum_{i=1}^{n} X_i \neq t \\ \dfrac{P(X_1 = x_1, \dots, X_n = x_n)}{P(\sum X_i = t)} & \end{cases}$$

$$= \begin{cases} 0 & \text{si } \sum X_i \neq t \\ \dfrac{\prod_{i=1}^{n} p^{x_i} \cdot (1-p)^{1-x_i}}{C_n^t \, p^t \cdot (1-p)^{n-t}} = \dfrac{\cancel{p^{\sum x_i}} \cancel{(1-p)^{n-\sum x_i}}}{C_n^t \cancel{p^t} \cancel{(1-p)^{n-t}}} = \dfrac{1}{C_n^t} & \end{cases}$$

Indépendante de p.

La statistique est exhaustive.

2. $X \rightsquigarrow \mathcal{P}(\lambda)$. $T = \sum_{i=1}^{n} X_i$ est exhaustive pour λ.

$(T \rightsquigarrow \mathcal{P}(\lambda n))$ $$P(X_1, \dots, X_n \mid T = t) = \frac{P(X_1, \dots, X_n, T = t)}{P(T = t)}$$

$$P(x_1,\dots,x_n \mid T=t) = P\left(x_1,\dots,x_{n-1},\, t-\sum_{i=1}^{n-1} x_i\right)$$

$$= \prod_{i=1}^{n-1}\left(\frac{e^{-\lambda}\cdot\lambda^{x_i}}{x_i!}\right)\cdot e^{-\lambda}\cdot\frac{\lambda^{t-\sum x_i}}{(t-\sum x_i)!}.$$

$$= \frac{e^{-n\lambda}\cdot\lambda^{t}}{x_1!\,x_2!\dots x_{n-1}!\,(t-\sum x_i)!}$$

D'où $\quad$ et $P(T=t) = e^{-n\lambda}\cdot\frac{(\lambda n)^t}{t!} = e^{-\lambda n}\cdot\frac{\lambda^t\cdot n^t}{t!}$

$$P(x_1,\dots,x_n \mid T=t) = \frac{t!}{n^t\cdot x_1!\dots x_{n-1}!\,(t-\sum x_i)!}$$ Indép de λ

donc T est exhaustive.

Théorème de factorisation de Fisher.

Soit $(\mathcal{X}, \mathcal{B}, f_s, s\in S)$ un modèle pour la va X dont un n-échantillon est $(\mathcal{X}^n, \mathcal{B}^n, \mathcal{L}(x_1^n, s), s\in S)$. Alors une statistique T définie sur $\mathcal{X}^n$ et à valeurs dans E, est exhaustive pour s si et seulement si ① $\exists$ une ft mesurable $g_s : E\to\mathbb{R}^+$

② $\exists$ une fonction mesurable $h : \mathcal{X}^n\to\mathbb{R}^+$ indépendante de s

tq $\mathcal{L}(x_1^n, s) = (g_s\circ T)(x_1,\dots,x_n)\cdot h(x_1,\dots,x_n) \quad \forall s.$

$$\begin{tikzcd}
\mathcal{X}^n \arrow[r, "T"] \arrow[dr, "\mathcal{L}(x_1^n, s)"] \arrow[dr, bend right, "h"'] & E \arrow[d, "g_s"] \\
& \mathbb{R}^+
\end{tikzcd}$$

Exemples : ① $X \rightsquigarrow b(p)$.

$(x_1, \dots, x_n)$ de densité $L(x_1^n, p) = p^{\sum x_i} \cdot (1-p)^{n-\sum x_i} = (1-p)^n \cdot \left(\frac{p}{1-p}\right)^{\sum x_i}$.

$T = \sum X_i$

$$L(x_1^n, p) = (1-p)^n \cdot \left(\frac{p}{1-p}\right)^{T(x_1, \dots, x_n)}$$

$$= (g_p \circ T)(x_1, \dots, x_n) \cdot 1$$

$T = \sum X_i$ est exhaustive.

$g_p(t) = (1-p)^n \cdot \left(\frac{p}{1-p}\right)^t$. et $h = 1$.

② $X \rightsquigarrow N(m, \sigma)$.

$$L(x_1^n, m, \sigma) = (2\pi\sigma^2)^{-\frac{n}{2}} \cdot \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2\right).$$

ⓐ <u>Si m est connue</u> $T = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$

T est une statistique exhaustive de σ

$$L(x_1^n, \sigma) = (2\pi\sigma^2)^{-\frac{n}{2}} \cdot \exp\left(-\frac{1}{2\sigma^2} T(x_1, \dots, x_n)\right).$$

$$= (g_\sigma \circ T)(x_1, \dots, x_n) \cdot 1$$

ⓑ <u>Si σ est connue</u>

$$L(x_1^n, m) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum x_i^2 - \frac{m^2}{2\sigma^2} + \frac{m}{\sigma^2} \sum x_i\right).$$

$$= \exp\left(-\frac{m^2}{2\sigma^2} + \frac{mn}{\sigma^2} \cdot \frac{1}{n} \sum_1^n x_i\right) \cdot (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum x_i^2\right)$$

$$= (g_m \circ \overline{X})(x_1^n) \cdot h(x_1^n)$$

c) Si <u>m et σ inconnues</u>

Alors $(\overline{X}, S^2)$ est exhaustif pour (m, σ) $S^2 = \frac{1}{n} \sum_1^n (X_i - \overline{X})^2$

Estimation à variance minimale
Information de Fisher.
Estimateur efficace.

Sur la classe des estimateurs sans biais $h : S \to E$.

$$\mathcal{E}_b = \{ T : \mathcal{X}^n \to E \mid E_s(T) = h(s) \ \forall s \}.$$

On peut mettre un pré-ordre

$T_1 \leq T_2$ ssi $Var_s(T_1) \geqslant Var_s(T_2) \quad \forall s \in S$

Définition : On dit que $T \in \mathcal{E}_b$ est à variance minimale si

$$\forall T' \in \mathcal{E}_b \qquad Var_s(T) \leq Var_s(T') \qquad (\forall s \in S)$$

Proposition : Un estimateur sans biais T est à variance minimale si $\forall s$, $E_s(TY) = 0$ pour toute statistique Y centrée et de carré sommable,

Preuve : Soit un autre estimateur sans biais S de h, alors $T-S$ est centrée donc $E_s(T(T-S)) = 0$ (car par hypothèse centrée).

$E(T^2 - TS) = 0$ donc $Var_s S - Var_s T \geqslant 0$

En effet $Var S - Var T = E(S^2) - E^2(S) - E(T^2) + E^2(T)$
$= E(S^2) - E(T^2) - \cancel{(h(s))^2} + \cancel{(h(s))^2} = E(S^2) - E(T^2)$.

$Var_s(S-T) = E_s((S-T)^2) = E_s(S^2 + T^2 - 2ST) = E_s(S^2) + E_s(T^2) - 2E(TS) = E(S^2) - E(T^2)$

donc $Var_s S - Var_s T = Var_s(S-T) \geqslant 0$

$\Rightarrow Var S \geqslant Var T$. $\forall S$ estimateur sans biais

donc T est à variance minimale.

Proposition : Si T est un estimateur sans biais du paramètre h et si Y est une statistique exhaustive de s alors $E(T/Y)$ est un estimateur sans biais de variance moindre que T.

Preuve : Puisque Y est exhaustive.

$$E_s(T/Y) = E(T/Y) \text{ ne dépend pas de } s.$$

et est sans biais $E_s(E(T/Y)) = E_s(T) = h(s)$.

et d'autre part $Var_s(E(T/Y)) \leqslant Var_s T$.

voir au dos

Proposition : Il ne peut exister qu'un seul estimateur sans biais de variance minimale d'un paramètre h.

Preuve : Supposons qu'il existe T_1 et T_2 de même variance et minimale. (T_1 et T_2 sans biais) alors $T = \frac{T_1+T_2}{2}$ est un estimateur sans biais

et $Var_s\left(\frac{T_1+T_2}{2}\right) \geqslant Var T_1 = Var T_2$.

$$Var\left(\frac{T_1+T_2}{2}\right) = \frac{1}{4} Var_s T_1 + \frac{1}{4} Var_s T_2 + \frac{1}{2} Cov(T_1,T_2) \geqslant Var_s T_i$$

$$\frac{1}{2} Var_s T_1 + \frac{1}{2} Cov(T_1,T_2) \geqslant Var_s T_i$$

$$\Rightarrow Cov(T_1,T_2) \geqslant Var_s T_i \quad (1)$$

Voir au dos

D'autre part on sait que

$$Cov_s(T_1,T_2) \leqslant (Var T_1 \,, Var T_2)^{1/2} \quad (2).$$

$$Cov_s(T_1,T_2) \leqslant Var_s T_i \quad (3).$$

(1) et (3) $\Rightarrow$ $Cov_s(T_1,T_2) = Var_s T_i \overset{(2)}{\Rightarrow} (T_1 - E_s(T_1))$ colinéaire à $(T_2 - E_s(T_2))$

donc $\exists\, k(s)$ tq $(T_1 - E_s(T_1)) = k(s)(T_2 - E(T_2))$ (4).

donc $Cov_s(T_1,T_2) = k(s)\, Var_s T_1$ (5) $\Longrightarrow k(s) = 1 \overset{(4)}{\Rightarrow} T_1 = T_2$

Définition : Soit $(\mathcal{X}, \mathcal{B}, f_\theta, \theta \in S)$ un modèle.

Soit un n-échantillon dont la fonction de vraisemblance est $\mathcal{L}(x_1^n, \theta)$, $\theta \in S$, $(x_1^n) \in \mathcal{X}^n$

Soit la statistique

$$\boxed{V_\theta(x_1^n) = \frac{d}{d\theta} \log \mathcal{L}(x_1^n, \theta)} \quad (\mathcal{X}^n \longrightarrow \mathbb{R}). \text{ (ft score)}$$

Si V_θ est définie

est complètement centrée

est de carré sommable

Alors on appelle <u>Information de Fisher</u> l'application $I_n(\theta) = E_\theta(V_\theta^2)$

$$I_n(\theta) = \begin{cases} \displaystyle\int_{\mathcal{X}^n} \left(\frac{d}{d\theta} \log \mathcal{L}(x_1^n, \theta)\right)^2 \cdot \mathcal{L}(x_1^n, \theta)\, dx_1 - dx_n & \text{(cas continue)} \\ \displaystyle\sum \left(\frac{d}{d\theta} \log \mathcal{L}(x_1^n, \theta)\right)^2 \mathcal{L}(x_1^n, \theta) & \text{(cas discret)} \end{cases}$$

Remarque : ① $I_n(\theta) = E_\theta(V_\theta^2) = \mathrm{Var}_\theta(V_\theta)$

② On demande que V_θ soit centrée pour s'assurer de la dérivabilité sous le signe $\int$ ou $\sum$ de $\displaystyle\int_{\mathcal{X}^n} \mathcal{L}(x_1^n, \theta)\, dx_1^n = 1$

Si on a pu dériver $\displaystyle\int_{\mathcal{X}^n} \frac{d}{d\theta} \mathcal{L}(x_1^n, \theta)\, dx_1^n = 0$

$$\int_{\mathcal{X}^n} \frac{d}{d\theta} \log \mathcal{L}(x_1^n, \theta) \cdot \mathcal{L}(x_1^n, \theta)\, dx_1^n = 0$$

$$E(V_\theta) = 0.$$

③ V_θ peut être regardée comme une v.a $V_\theta : \Omega \longrightarrow \mathbb{R}$

$$V_\theta = \frac{d}{d\theta} \log \mathcal{L}(X_1, \dots, X_n, \theta).$$

4°) $V_\theta(x_1^n) = \sum_{i=1}^{n} \frac{d}{d\theta} \log f_\theta(x_i)$

Exemples : Si $X \rightsquigarrow N(m,\sigma)$.

Supposons σ connu.

Recherche de V_θ.

$$f(x,m,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-m)^2}$$

$$\log f(x,m,\sigma) = -\log \sigma\sqrt{2\pi} - \frac{1}{2\sigma^2}(x-m)^2$$

$$\frac{d}{dm} \log f(x,m,\sigma) = \frac{2}{2\sigma^2}(x-m) = \frac{x-m}{\sigma^2}$$

$$V_\theta(x_1^n) = \sum_{i=1}^{n} \frac{d}{dm} \log f(x_i,m,\sigma) = \sum_{i=1}^{n} \left(\frac{x_i-m}{\sigma^2}\right) = \frac{1}{\sigma^2}\left(\sum_1^n x_i - mn\right)$$

$$= \frac{n}{\sigma^2}\left(\overline{X}-m\right)$$

$$E_m\left(V_m^2\right) = E_m\left(\frac{n^2}{\sigma^4}\left(\overline{X}-m\right)^2\right) = \frac{n}{\sigma^2} E_m\left(\left(\frac{\overline{X}-m}{\sigma/\sqrt{n}}\right)^2\right)$$

$$\frac{\overline{X}-m}{\sigma/\sqrt{n}} \rightsquigarrow N(0,1)$$

$$\left(\frac{\overline{X}-m}{\sigma/\sqrt{n}}\right)^2 \rightsquigarrow \Gamma\left(\frac{1}{2},\frac{1}{2}\right) \Rightarrow E\left(\left(\frac{\overline{X}-m}{\sigma/\sqrt{n}}\right)^2\right) = 1$$

$$\Rightarrow \boxed{I_n(m) = \frac{n}{\sigma^2}}$$

Inégalité de Cramer-Rao (Frechet-Darmois-Cramer-Rao FDCR).

Soit un n-échantillon $(\mathcal{X}^n, \mathcal{B}^n, L(x_1^n, \theta), \theta \in S \subset \mathbb{R})$

Soit T un estimateur de θ.

Alors $$\mathrm{Var}_\theta(T) \geqslant \frac{\left(\frac{d}{d\theta} E_\theta(T)\right)^2}{I_n(\theta)}$$ (Si $I_n(\theta)$ existe).

Démonstration :

$$E_\theta(T) = \int T(x_1^n) L(x_1^n, \theta)\, dx_1^n$$

$$\frac{d}{d\theta} E_\theta(T) = \int_{\mathcal{X}^n} T(x_1^n) \frac{d}{d\theta} L(x_1^n, \theta)\, dx_1^n$$

$$= \int_{\mathcal{X}^n} T(x_1^n) \frac{d}{d\theta} \operatorname{Log} L(x_1^n, \theta) . L(x_1^n, \theta)\, dx_1^n$$

$$= E_\theta(T . V_\theta)$$

Alors $E_\theta\big((T - E_\theta(T)) V_\theta\big) = E_\theta\big(T V_\theta - E_\theta(T) V_\theta\big) = E_\theta(T V_\theta) - E(T) \underset{\overset{\|}{0}}{E(V_\theta)} = E(T V_\theta)$

Donc $E_\theta\big((T - E_\theta(T)) V_\theta\big) = \frac{d}{d\theta} E_\theta(T)$.

Inégalité de Schwartz $\rightarrow$ $\left(\frac{d}{d\theta} E_\theta(T)\right)^2 \leqslant E_\theta\big((T - E_\theta(T))^2\big) E_\theta(V_\theta^2)$

D'où $\left(\frac{d}{d\theta} E_\theta(T)\right)^2 \leqslant \mathrm{Var}_\theta T . I_n(\theta)$

Corollaire : Si h est un paramètre du modèle, h dérivable et si T est un estimateur sans biais de h

Alors $$\boxed{\mathrm{Var}_\theta(T) \geqslant \frac{\left(\frac{d}{d\theta} h(\theta)\right)^2}{n I_1(\theta)}}$$

càd que la variance de tout estimateur sans biais est minorée par un nombre qui ne dépend pas de cet estimateur.

$\Rightarrow$ C'est donc un minorant universel pour la variance des estimateurs sans biais de h.

Donc si $\mathcal{E}_b(h) = \{$des estimateurs sans biais de $h\}$

$$\operatorname{Inf} \operatorname{Var}_s(T) \geqslant \frac{(h'(s))^2}{I_n(s)} \qquad \forall T \in \mathcal{E}_b(h).$$

Définition : Un estimateur sans biais T de h est dit <u>efficace</u> si $\operatorname{Var}_s T = \dfrac{(h'(s))^2}{I_n(s)}$

Donc un estimateur efficace est à variance minimale (Réciproque est fausse).

Proposition : 1°) $I_n(s) = -E_s\left[\frac{d}{ds} V_s\right]$

2°) Si $I_1(s)$ = est l'information donnée par le 1-échantillon alors $I_n(s) = n I_1(s)$.

Preuve : (1) On sait que $\displaystyle\int_{\mathcal{X}^n} L(x_1^n, s)\, dx_1^n = 1$.

$$\int_{\mathcal{X}^n} \frac{d}{ds} \log L(x_1^n, s) . L(x_1^n, s)\, dx_1^n = 0$$

$$\int_{\mathcal{X}^n} \frac{d^2}{ds^2} \log L(x_1^n, s) . L(x_1^n, s)\, dx_1^n + \int_{\mathcal{X}^n} \left(\frac{d}{ds} \log L(x_1^n, s)\right)^2 L(x_1^n, s)\, dx_1^n = 0.$$

$$\Rightarrow E(V_\theta^2) = -\int \frac{d}{d\theta} V_\theta(x) \cdot L(x_1^n, \theta)\, dx_1^n$$

$$\Rightarrow \quad E_\theta(V_\theta^2) = -E_\theta\left[\frac{d}{d\theta} V_\theta\right]$$

2°) $L(x_1^n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$

$$I_n(\theta) = -E_\theta\left[\frac{d^2}{d\theta^2} \operatorname{Log} L(x_1^n, \theta)\right]$$

$$= -E_\theta\left[\sum_{i=1}^{n} \frac{d^2}{d\theta^2} \operatorname{Log} f(x_i, \theta)\right] = -\sum_{i=1}^{n} E_\theta\left(\frac{d^2}{d\theta^2} \operatorname{Log} f(x_i, \theta)\right) =$$

$$= -\sum_{i=1}^{n} E_\theta\left(\frac{d^2}{d\theta^2} \operatorname{Log} f(x, \theta)\right) = -\, n E_\theta\left(\frac{d^2}{d\theta^2} \operatorname{Log} f(x, \theta)\right)$$

$$= n\left[-E_\theta\left(\frac{d}{d\theta} V_\theta\right)\right] = n I_1(\theta).$$

Exemple : $X \rightsquigarrow N(m, \sigma)$ $\quad$ m connue.

Information sur σ. $\quad f(x, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\cancel{x} - m)^2}$

$I_n(\sigma) = n I_1(\sigma)$

$$\operatorname{Log} f(x, \sigma) = -\frac{1}{2}\operatorname{Log}(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x-m)^2$$

$$\frac{d}{d\sigma} \operatorname{Log} f(x, \sigma) = -\frac{1}{\cancel{2}} \frac{\cancel{2\pi}\, \cancel{2}\sigma}{\cancel{2\pi}\, \sigma^{\cancel{2}}} - (x-m)^2\left(-\frac{\cancel{2}\sigma}{\cancel{2}\sigma^4}\right)$$

$$V_\sigma(x) = \frac{(x-m)^2}{\sigma^3} - \frac{1}{\sigma}$$

$$\frac{d}{d\sigma} V_\sigma(x) = -\frac{3(x-m)^2}{\sigma^4} + \frac{1}{\sigma^2}.$$

$$E_\sigma\left(\frac{d}{d\sigma} V_\sigma(x)\right) = E\left[\frac{1}{\sigma^2} - \frac{3}{\sigma^2}\left(\frac{x-m}{\sigma}\right)^2\right] = \frac{1}{\sigma^2} - \frac{3}{\sigma^2} E\left(\left(\frac{x-m}{\sigma}\right)^2\right) = \frac{1-3}{\sigma^2} = -\frac{2}{\sigma^2}$$

donc $I_1(\sigma) = -E_\sigma\left(\frac{d}{d\sigma} V_\sigma\right) = \frac{2}{\sigma^2} \Rightarrow \boxed{I_n(\sigma) = n I_1(\sigma) = \frac{2n}{\sigma^2}}$