Chapitre 3: Exhaustivité, complétion, information de Fischer et estimation sans biais

Rabah Messaci

Département de Probabilités-Statistique
USTHB

Octobre 2011

Définition : Fonction de vraisemblance

Soit le modèle statistique $(\mathfrak{X}, \mathfrak{B}, (f_\theta)_{\theta\in\Theta})$ On appelle fonction de vraisemblance du modèle (ou de l'observation), la fonction :

$L : \Theta \rightarrow R_+$

$\quad \theta \rightarrow L(\theta, x) = f_\theta(x)$

Si le modèle est un modèle d'échantillon $(\mathfrak{X}, \mathfrak{B}, (f_\theta)_{\theta\in\Theta})^{(n)}$

$L(\theta, x_1, x_2, ..., x_n) = L(\theta, \widetilde{x}) = \prod_{i=1}^{n} f_\theta(x_i)$

Le logarithme de la fonction de vraisemblance est dit fonction de log-vraisemblance notée :

$l(\theta, \widetilde{x}) = \log L(\theta, \widetilde{x})$

Définition : Fonction de vraisemblance

1) n-échantillon loi de Poisson $P(\lambda)$. On a

$$L(\lambda, \widetilde{x}) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \frac{\lambda^{\sum\limits_{i=1}^{n} x_i}}{\prod\limits_{i=1}^{n} x_i!}$$

$$\mathrm{l}(\lambda, \widetilde{x}) = \log L(\lambda, \widetilde{x}) = -n\lambda + (\sum_{i=1}^{n} x_i) \log \lambda - \log(\prod_{i=1}^{n} x_i!).$$

2) n-échantillon loi normale, σ^2 connu.

$$L(m, \widetilde{x}) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x_i - m)^2}{\sigma^2}} \right) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{1}{2} \sum\limits_{i=1}^{n} \frac{(x_i - m)^2}{\sigma^2}} \; l(m, \widetilde{x}) =$$

$\log L(m, \widetilde{x})$

Exhaustivité

En statistique les observations (données) peuvent être très nombteuses (des fois de l'ordre de plusieurs milliers, par exemple pluviométrie quotidienne ou taux d'inflation mensuel relevés sur plusieurs années. Les statistiques associés à un échantillon d'observations servent à résumer ces observations. Ainsi si on a n observations réelles $\widetilde{x} = (x_1, x_2, ..., x_n)$ $T(\widetilde{x}) = \sum_{i=1}^{n} x_i$ les résume en une seule donnée.

Question : Le résumé est-il un bon résumé ?

En résumant ainsi, i.e en considérant que $T(x)$ au lieu de l'échantillon initiale, perd-on de l'information sur le paramètre θ ?

Quand on lit un livre ou un résumé de ce livre, est-ce la même chose du point de vue de l'information sur un point donné ?

Si oui, on dira que le résumé est exhaustif.

Problème : définir mathématiquement cette notion d'exhaustivité.

Exhaustivité

Définition : Statistique exhaustive

Soit un modèle statistique $(\mathfrak{X}, \mathfrak{B}, (\mathcal{P}_\theta)_{\theta \in \Theta})$ et une statistique T définie sur ce modèle. T est dite exhaustive si la loi conditionnelle $P_{X/T(X)=t}$ ne dépend pas de θ .

Théorème de factorisation

T est exhaustive si et seulement si :
il existe des fonctions g et h telles que

$$L(\theta, x) = g(\theta, T(x))h(x)$$

Ce théorème est à admettre.

Exaustivité

Exemples

- Echantillon d'une loi de Poisson

$$L(\lambda,\widetilde{x}) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}\frac{1}{\prod_{i=1}^{n} x_i!} = g(\lambda,\sum_{i=1}^{n} x_i)h(\widetilde{x})$$

donc $T(\widetilde{x})) = \sum_{i=1}^{n} x_i$ est une statistique exhaustive pour λ.

- Echantillon d'une loi normale à variance connue :

$$L(m,\widetilde{x}) = \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\frac{(x_i-m)^2}{\sigma^2}}\right) =$$

$$\left(\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\frac{m^2}{\sigma^2}}\right)^n e^{\frac{m}{\sigma^2}\sum_{i=1}^{n} x_i} e^{-\frac{1}{2}\frac{\sum_{i=1}^{n} x_i^2}{\sigma^2}} = g(m,\sum_{i=1}^{n} x_i)h(\widetilde{x})$$

donc $T(\widetilde{x})) = \sum_{i=1}^{n} x_i$ est une statistique exhaustive pour m.

Exaustivité

Exemples

3) n-échantillon loi normale

$$L(m,\sigma^2,\widetilde{x}) = \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\frac{(x_i-m)^2}{\sigma^2}}\right) =$$

$$\left(\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\frac{m^2}{\sigma^2}}\right)^n e^{\frac{m}{\sigma^2}\sum_{i=1}^{n}x_i} e^{-\frac{1}{2}\frac{\sum_{i=1}^{n}x_i^2}{\sigma^2}} = g(m,\sigma^2,\sum_{i=1}^{n}x_i,\sum_{i=1}^{n}x_i^2)h(\widetilde{x})$$

avec $h(\widetilde{x}) = 1$.

$T(\widetilde{x})) = (\sum_{i=1}^{n}x_i, \sum_{i=1}^{n}x_i^2)$ est une statistique exhaustive pour m et σ^2. C'est une statistique de dimension 2 (vectorielle).

Conclusion : Quand on veut faire de l'inférence statistique sur λ, si on a un échantillon poissonien, travailler avec l'échantillon tout entier $\widetilde{x} = (x_1, x_2, ..., x_n)$ ou avec $T(\widetilde{x}) = \sum_{i=1}^{n}x_i$

ne change rien. Toute l'information contenue sur λ dans $\widetilde{x}$ se retouvre dans

Exaustivité

Exemples

4)$\widetilde{x} = (x_1, x_2, ..., x_n)$ n-échantillon de $X \backsim \mathcal{U}_{[0,\theta]}$ de densité $f_\theta(x) = \frac{1}{\theta} 1_{[0,\theta]}(x)$. On a

$$L(\theta, \widetilde{x}) = \prod_{i=1}^{n} f_\theta(x_i) = \frac{1}{\theta^n} 1_{[0,\theta]}(\max_{1 \leq i \leq n}(x_i)) = g(\theta, \max_{1 \leq i \leq n}(x_i)) h(\widetilde{x}) \text{ avec } h(\widetilde{x}) = 1$$

$\Longrightarrow\ T(\widetilde{x}) = \max_{1 \leq i \leq n}(x_i)$: statistique exhaustive pour θ.

Exhaustivité

Théorème

Soit f_θ appartenant à la famille des lois exponentielles $\mathcal{E}_1$ d'ordre 1, i.e

$$f_\theta(x) = c(\theta)h(x)\exp(a(x)\alpha(\theta))$$

Alors $T(x) = a(x)$ est exhaustive pour θ.

Si on a un n-échantillon alors $T(\widetilde{x}) = \sum_{i=1}^{n} a(x_i)$ est exhaustive pour θ.

Démonstration : Théorème de factorisation.

Exhaustivité

Théorème

Soit f_θ appartenant à la famille des lois exponentielles $\mathcal{E}_k$ d'ordre k, i.e

$$f_\theta(x) = c(\theta)h(x)\exp(\sum_{i=1}^{k} a_i(x)\alpha_i(\theta))$$

Alors $T(x) = (a_1(x), a_2(x)..., a_k(x))$ est exhaustive pour θ.

Si on a un n-échantillon alors $T(\widetilde{x}) = (\sum_{i=1}^{n} a_1(x_i), \sum_{i=1}^{n} a_2(x_i)...., \sum_{i=1}^{n} a_k(x_i))$ est exhaustive pour θ.

Démonstration : Théorème de factorisation.

Complétion

Problème :
Il peut exister pour un paramètre donné plusieurs statistiques ehaustives (résumés exhaustifs). On a vu que n données peuvent résumées de manière exhaustive dans 3, 4 données ou même 1 donnée. Il y a donc réduction plus ou moins importante.
Question : Quelle est la plus grande réduction que l'on peut avoir ?
C'est la notion de statistique exhaustive minimale liée à la notion de statistique exhaustive complète.

Définition : statistique complète

Une statistique T est dite complète si :
$E(h_1(T)) = E(h_2(T)) \implies h_1(T) = h_2(T)$ $P_\theta-$presque sûrement $\forall\theta \in \Theta$.

Complétion

Théorème

Soit f_θ appartenant à la famille des lois exponentielles $\mathcal{E}_k$ d'ordre k.
Si $dim(\Theta) = k$, alors la statistique $T(\widetilde{x}) = (\sum_{i=1}^{n} a_1(x_i), \sum_{i=1}^{n} a_2(x_i) ...., \sum_{i=1}^{n} a_k(x_i))$ exhaustive est complète.

Information de Fischer

L'expression d'information sur un paramètre contenue dans des observations est souvent employée. Cette notion est jusqu'à présent intuitive.
Question : peut-on définir mathématiquement cette notion

Il existe plusieurs définitions possibles (Fisher, Shannon, Kullblack....). Nous introduisons dans la suite la notion d'information au sens de Fischer. Soit un modèle statistique
L'information de Fischer est définie sous certaines hypothèses dites : conditions de régularité de Fischer :
1) Le support de f_θ : $Supp(f_\theta) = \{x \diagup \ f_\theta(x) \neq 0\}$ ne dépend pas de θ
2) On peut dériver sous le signe somme dans $\int f_\theta(x)dx$
i.e $\dfrac{\partial}{\partial\theta} \int f_\theta(x)dx = \int \dfrac{\partial}{\partial\theta} f_\theta(x)dx.$

Information de Fischer

Définition : information de Fischer

Soit un modèle statistique associé à l'observation d'une v.a de loi P_θ de densité f_θ.

On appelle quantité d'information (de Fisher) ramenée par le modèle(ou par l'observation) sur θ le nombre

$$I(\theta) = Var(\frac{\partial}{\partial\theta} \log L(\theta, x))$$

Exemple

Information ramenée par une observation d'une variable de loi $N(m, \sigma^2)$ sur m (σ^2 connu)

$$L(m,x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}} \implies \log L(m,x) = \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{1}{2}\frac{(x-m)^2}{\sigma^2}$$

$$\implies \frac{\partial}{\partial m}\log L(m,x) = \frac{(X-m)}{\sigma^2} \implies I(m) = Var(\frac{\partial}{\partial m}\log L(m,x)) = \frac{1}{\sigma^2} Var\left(\frac{X-m}{\sigma}\right)$$

$$\implies I(m) = \frac{1}{\sigma^2}$$

Information de Fischer

Propriétés :

1) $I(\theta) = E((\frac{\partial}{\partial\theta} \log L(\theta, x))^2)$

2)Si on suppose en plus $\frac{\partial^2}{\partial\theta^2} \int f_\theta(x)dx = \int \frac{\partial^2}{\partial\theta^2} f_\theta(x)dx$. on a

$I(\theta) = -E(\frac{\partial^2}{\partial\theta^2} \log L(\theta, x))$

3) Si on a un n-échantillon l'information ramenée par le n-échantillon $I_n(\theta)$ est égale à $nI(\theta)$

4) Si T est exhaustive $I_T(\theta) = I(\theta)$

Information de Fischer

Démonstration 1)On a $E(\frac{\partial}{\partial\theta}\log L(\theta,x)) = 0$ $E(\frac{\partial}{\partial\theta}\log L(\theta,x)) = \int \frac{\partial}{\partial\theta}$

$\log f_\theta(x)dx = \int \frac{\frac{\partial}{\partial\theta}f_\theta(x)}{f_\theta(x)} f_\theta(x)dx = \int \frac{\partial}{\partial\theta}\ f_\theta(x)dx = \frac{\partial}{\partial\theta}\int\ f_\theta(x)dx = 0$

2) $E(\frac{\partial^2}{\partial\theta^2}\log L(\theta,x)) = \int \frac{\partial}{\partial\theta}\left(\frac{\frac{\partial}{\partial\theta}f_\theta(x)}{f_\theta(x)}\right) f_\theta(x)dx =$

$\int \frac{f_\theta(x)\frac{\partial^2}{\partial\theta^2}f_\theta(x) - \left(\frac{\partial}{\partial\theta}f_\theta(x)\right)^2}{f_\theta(x)^2} f_\theta(x)dx =$

$\int \frac{\partial^2}{\partial\theta^2}\ f_\theta(x)dx - \int\left(\frac{\frac{\partial}{\partial\theta}f_\theta(x)}{f_\theta(x)}\right)^2 f_\theta(x)dx = -E(\left(\frac{\partial}{\partial\theta}\log L(\theta,x)\right)^2)$

Estimation sans biais

amélioration d'un estimateur

Soit T un estimateur sans biais de $g(\theta)$ et S une statistique exhaustive pour θ.
Alors
1) $E_\theta(T/S)$ est un estimateur sans biais de $g(\theta)$
2) $Var_\theta(E_\theta(T/S)) \leq Var_\theta(T)$

Conclusion : $E_\theta(T/S)$ est un meilleur estimateur sans biais de $g(\theta)$

Estimation sans biais

Démonstration :
$E_\theta(T/S)$ est un estimateur de $g(\theta)$ car ne dépend pas de θ.
1) $E_\theta(E_\theta(T/S)) = E_\theta(T) = g(\theta)$
(Théorème de l'espérance conditionnelle)
2) $Var_\theta(T) = Var_\theta(E_\theta(T/S)) + E_\theta(Var_\theta(T/S))$
(Théorème de la variance conditionnelle)
$\Longrightarrow Var_\theta(E_\theta(T/S)) = Var_\theta(T) - E_\theta(Var_\theta(T/S)) \leq Var_\theta(T)$
Remarque : Si $T = h(S)$ alors $E_\theta(h(S)/S) = h(S) = T$
Dans ce cas il n'y a pas d'amélioration.

Estimation sans biais

Théorème de Lehmann-Scheffé

Soit T un estimateur sans biais de $g(\theta)$ et S une statistique exhaustive et complète pour θ. Alors
$E_{\theta}(T/S)$ est l'unique estimateur sans biais de $g(\theta)$ dit estimateur sans biais de variance uniformément minimale (ESBVUM)

Démonstration

Estimation sans biais

Théorème de Rao-Cramer

Soit T un estimateur sans biais de $g(\theta)$. On suppose les conditions de régularité de Fischer vérifiées, alors

$$Var_{\theta}(T) \geq \frac{(g'(\theta))^2}{I(\theta)}$$

La quantité $\dfrac{(g'(\theta))^2}{I(\theta)}$ est dite borne de Rao-Cramer (BRC)

$$Var_{\theta}(T) = \frac{(g'(\theta))^2}{I(\theta)}$$

Estimation sans biais

Démonstration : Rappel :

$$|\rho_{X,Y}| = \left|\frac{Cov(X,Y)}{\sigma_X \sigma_Y}\right| \leq 1 \implies Cov(X,Y)^2 \leq \sigma_X^2 \sigma_Y^2$$

$$E_\theta(T(X)) = \int_{\mathcal{X}} T(x) f_\theta(x) dx = \int_{\mathcal{X}} T(x) L(\theta, x) dx = g(\theta) \implies$$

$$\frac{\partial}{\partial\theta}(E_\theta(T(X)) = \int_{\mathcal{X}} T(x) \frac{\partial}{\partial\theta}(L(\theta,x)) dx = g'(\theta)$$

$$= \int_{\mathcal{X}} T(x) \frac{\partial}{\partial\theta}(\log L(\theta,x)) L(\theta,x) dx = Cov(T(X), \frac{\partial}{\partial\theta}(\log L(\theta,X))$$

$$\implies Cov^2(T(X), \frac{\partial}{\partial\theta}(\log L(\theta,X)) \leq Var_\theta(T(X)) Var(\frac{\partial}{\partial\theta}(\log L(\theta,X))$$

$$\implies Var_\theta(T(X)) \geq \frac{Cov^2(T(X), \frac{\partial}{\partial\theta}(\log L(\theta,X))}{Var(\frac{\partial}{\partial\theta}(\log L(\theta,X))} = \frac{g'(\theta)^2}{I(\theta)}$$

Estimation sans biais

Cas particulier : $g(\theta) = \theta$

$BCR = \dfrac{1}{I(\theta)}$

Remarque :

Pour que $Var_\theta(T(X)) \geq \dfrac{Cov^2(T(X), \frac{\partial}{\partial\theta}(\log L(\theta, X))}{Var(\frac{\partial}{\partial\theta}(\log L(\theta, X))} = \dfrac{g'(\theta)^2}{I(\theta)}$

Il est nécessaire et suffisant que $\dfrac{\partial}{\partial\theta}(\log L(\theta, X) = k(\theta)T(X) + l(\theta)$

$$\implies E_\theta(\frac{\partial}{\partial\theta}(\log L(\theta, X)) = k(\theta)E_\theta(T(X)) + l(\theta)$$

$$\implies E_\theta(T(X)) = g(\theta) = -\frac{l(\theta)}{k(\theta)}$$

$$\iff \frac{\partial}{\partial\theta}(\log L(\theta, X) = k(\theta)(T(X) - g(\theta))$$

Estimation sans biais

Définition : estimateur efficace

Un estimateur sans biais T de $g(\theta)$ est dit efficace si

$$Var_\theta(T) = \frac{(g'(\theta))^2}{I(\theta)}$$

Information de Fischer : cas vectoriel

Cas d'un paramètre vectoriel $\theta = (\theta_1, \theta_2, ..., \theta_p)$
L'information de Fischer est alors donnée par la matrice des variances-covariances

$$\mathcal{I}(\theta) = \mathcal{V}(Grad \log L(\theta, X))$$

du vecteur $Grad \log L(\theta, X) = \begin{cases} \dfrac{\partial}{\partial \theta_1} \log L(\theta, X) \\ \dfrac{\partial}{\partial \theta_2} \log L(\theta, X) \\ . \\ . \\ \dfrac{\partial}{\partial \theta_p} \log L(\theta, X) \end{cases}$

Comme pour le cas réel on peut montrer que

$$\mathrm{I}(\theta) = -(E(\frac{\partial}{\partial \theta_i} \frac{\partial}{\partial \theta_j} \log L(\theta, X))_{1 \leq i \leq p, 1 \leq j \leq p},$$

Information de Fischer : cas vectoriel

Exemple

Soit x une observation de de $X \backsim N(m, \sigma^2)$

$$\mathcal{I}(m,\sigma^2) = -\begin{pmatrix} E(\frac{\partial^2}{\partial m^2}\log L(\theta,X) & E(\frac{\partial}{\partial m}\frac{\partial}{\partial \sigma^2}\log L(\theta,X) \\ E(\frac{\partial}{\partial m}\frac{\partial}{\partial \sigma^2}\log L(\theta,X) & E(\frac{\partial^2}{\partial (\sigma^2)^2}\log L(\theta,X) \end{pmatrix}$$

$$\mathcal{I}(m,\sigma^2) = -\begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

Théorème de Rao-Cramer : cas vectoriel

Soit deux matrices définies positives A et B, on dit que $A \geq B$ si
$Q_A(x) = x^t Ax \geq Q_B(x) = x^t Bx$
On rappelle que les matrices de variances-covariances sont définies positives.

Théorème de Rao-Cramer : cas vectoriel

Soit T un estimateur sans biais de $g(\theta)$ (à valeurs dans R^p). On suppose les conditions de régularité de Fischer vérifiées, alors

$$\mathcal{V}_\theta(T) \geq Grad(g(\theta)^t \mathcal{I}(\theta)^{-1} Grad(g(\theta)$$

$\mathcal{V}_\theta(T)$ est la matrice de variances-covariances de T.