DETERMINATION DU NOMBRE DE CLASSES D'UN HISTOGRAMME PAR LE CRITERE D'AKAIKE

par Denis de BRUCQ

LACIS-ITEPEA U.E.R. des sciences B.P. 67
Université de ROUEN 76130 Mont-Saint-Aignan

Résumé: La modélisation d'un phénomène physique nécessite le choix des variables X(1),...,X(m) d'état du système et dans de nombreuses situations pratiques,le nombre m de variables réelles à estimer est grand, voire infini.Par ailleurs,l'expérience se répète indépendamment un nombre n fini de fois ou se prolonge pendant une durée n ΔT limitée.

Connaissant l'observation à n dimensions, le critère d'Akaike fournit la dimension p du vecteur X à estimer.Cette dimension p est inférieure ou égale à la dimension m ,théorique,nécessaire pour la modélisation.Ce critère s'appuie sur une fonction de coût,utilisant l'information de Kullbach.

Au lieu d'appliquer classiquement le critère d'Akaike au choix de l'ordre p d'un modèle auto-régressif,la détermination du nombre p de classes d'un histogramme en fonction du nombre n d'observations indépendantes,est effectuée.

Summary: In order to modelize experimental phenomena,some unknowns X(1),...,X(m) are requested to characterize the state of the system.In various situations,the number m is great not to say infinite.However either datas contain only n values obtained independently,or experiment extends over a n ΔT duration.

From a sample with n values,we have to assure statistical stability and Akaike's criterium gives obtimal p to estimate m.This dimension p is less than m and increases with n.The criterium minimize a cost function coming from Kullbach information theory.

Instead of an auto-regressive process,we consider an histogram modelizing an empirical density of probability and we obtain a practical formula to determine the optimal number p of classes.So we give a solution of the agregation problem.

Nous considérons une structure statistique $(\Omega,A,P_\theta)_{\theta\in I}$ dans laquelle θ est absolument continue par rapport à une probabilité μ a priori.Nous possédons n observations indépendantes $Y_1,\ldots,Y_n$ notée $Y(1\mapsto n)$de la variable aléatoire X de densité $f(\lambda)=\frac{d\lambda}{d\mu}$;le paramètre λ,vraie valeur de θ est identifiée avec la probabilité elle-même.Nous introduisons une tribu $\mathcal{B}(M)$ discrète provenant de m ensembles mesurables $(A_1,\ldots,A_m)$ constituant une partition M de Ω.Introduisons les densités de probabilité f discrètisées sur $\mathcal{B}(M)$.

Il s'agit de construire l'estimateur θ du maximum de vraisemblance ainsi que le critère d'Akaike permettant de choisir au mieux un éventuel regroupement desensembles $A_1,\ldots,A_m$ en $B_1,\ldots,B_k$ avec $k\leqslant m$.

1:HISTOGRAMME APPROCHE

DEFINITION 1-1:
Soit $M=\{A_1,\ldots,A_m\}$,une partition de Ω en éléments de $\mathcal{A}$ telle que pour tout $1\in\{1,\ldots,m\}$,nous ayons $\mu(A_1)>0$.

Nous définissons la densité discrètisée de θ par rapport à μ par $\forall x\in\Omega$

$$(1.1)\qquad f(x;\theta)\overset{\Delta}{=}\sum_{A\in M}\frac{\theta(A)}{\mu(A)}\,1_A(x)$$

Comme θ est une probabilité,nous avons:

$$\sum_{l=1}^{m}\theta(A_l)=1 \text{ et } \theta(A_l)\geqslant 0$$

pour tout $1\in\{1,\ldots,m\}$.

Ainsi l'ensemble $\mathcal{I}(M)$ est le convexe:

$$(1.2)\qquad \mathcal{I}(M)\overset{\Delta}{=}\{\theta\in\mathbf{R}^m;\forall A\in M\ \theta(A)\geqslant 0 \text{ et } \sum_{A\in M}\theta(A)=1\}$$

DEFINITION 1-2:
Soit $K=\{B_1,\ldots,B_k\}$,une sous-partition mesurable de M;plus précisémment tout B de K appartient à $\mathcal{A}$ et pour tout A de M,il existe B de K tel que $A\subset B$. Nous définissons la densité $f(\theta(K))$ de $\theta(K)$ par rapport à μ pour tout x de Ω par:

$$(1.3)\qquad f(x;\theta(K))=\sum_{B\in K}\frac{\theta(B)}{\mu(B)}\,1_B(x)$$

ainsi le nombre k de paramètres $\theta(B_1),\ldots,\theta(B_k)$ pour définir $f(\theta(K))$ relativement à K est une variable entière libre entre 1 et m.

2-ESTIMATEURS DU MAXIMUM DE VRAISEMBLANCE

Lorsque la partition K mesurable de $(\Omega,\mathcal{A},\mu)$ est donnée,nous introduisons l'estimateur $\theta(K,Y(1\mapsto n))$ du maximum de vraisemblance pour lequel les calculs sont aisés.De plus,cet estimateur jouit de nombreuses propriétés asymptotiques lorsque le nombre n d'observations indépendantes tend vers l'infini.

PROPOSITION 2-1:
Soit $Y(1\mapsto n)$,un échantillon de taille n de la variable aléatoire X de densité f alors

la Log vraisemblance relativement à la partition mesurable K de (Ω,A), vaut:

$$(2.1)\quad L(K,Y(1\mapsto n))=\sum_{B\in K} n\,\nu(B)\,\mathrm{Log}\frac{\theta(B)}{\mu(B)} \qquad \text{avec}$$

$$n\nu(B)\overset{\Delta}{=}\{i\in\{1,\dots,n\};1_B(y_i)=1\}$$

qui est le nombre d'observations $Y_i=y_i$ appartenant à B.
Ainsi $\nu(B;Y(1\mapsto n))$ est la fréquence statistique de l'ensemble B.

PROPOSITION 2-2:
Soit $Y(1\mapsto n)$, un échantillon de taille n de la variable aléatoire X de densité f_λ alors pour toute partition K mesurable, l'estimateur $\theta(Y(1\mapsto n);K)$ du maximum de vraisemblance de λ relativement à K est tel que pour tout $B\in K$

$$\theta(K,Y(1\mapsto n))=\nu(B)$$

fréquence statistique de B.

La modélisation choisie conduit donc à un résultat très simple: si B est un ensemble d'une partition K mesurable de (Ω,A), l'estimateur $\theta(B)$ de $\lambda(B)$ du maximum de vraisemblance n'est autre que la fréquence statistique $\nu(B)$. Cet estimateur est indépendant de la partition K dont fait partie B, la notation $\theta(B)=\theta(B;Y(1\mapsto n))$ qui n'introduit pas la partition K est cohérente avec cette propriété.

PROPOSITION 2-3:
Soit K, une partition mesurable de A. Si $\theta(Y(1\mapsto n))$ est l'estimateur du maximum de vraisemblance de la loi λ alors le vecteur aléatoire

$$(\sqrt{n}\,(\theta(B;Y(1\mapsto n))-\lambda(B));B\in K)$$

converge en loi vers la loi gaussienne centrée dégénérée de covariance:

$$\Gamma=\begin{bmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_K\end{bmatrix}-\begin{bmatrix}\lambda_1\\ \vdots\\ \lambda_K\end{bmatrix}\begin{bmatrix}\lambda_1,\dots,\lambda_K\end{bmatrix}$$

La matrice C d'information de Fisher peut-être calculée aisément:

$$(2.3)\quad C(i,j)=E\left[\left(\frac{\partial \mathrm{Log}f(\theta)}{\partial\theta_i}\,\frac{\partial \mathrm{Log}f}{\partial\theta_j}\right)\right]_{\theta=\lambda}$$

Cependant pour notre structure statistique θ n'est pas libre dans $\mathbb{R}^k$ mais est contraint de vérifier:

$$\Sigma\,\theta(l)=1$$

Nous prenons $\theta(1),\dots,\theta(k-1)$ comme variables libres et

$$\theta(k)=1-\theta(1)-\dots-\theta(k-1).$$

PROPOSITION 2-4:
Soit $(\Omega,B(K),(f(\theta))_{\theta\in I(K)})$, la structure statistique où:

$$f(x;\theta(K))=\sum_{l=1}^{k}\frac{\theta(l)}{\mu(l)}\,1_{B(l)}(x)$$

alors en fonction des paramètres libres $\theta(1),\dots,\theta(k-1)$, la matrice de Fisher vaut:

$$C=\begin{bmatrix}\lambda(1)^{-1}+\lambda(k)^{-1} & \dots & \lambda(k)^{-1}\\ \dots & \dots & \dots\\ \lambda(k)^{-1} & \dots & \lambda(k-1)^{-1}+\lambda(k)^{-1}\end{bmatrix}$$

PROPOSITION 2-5:
Les matrices C et $\Gamma/k-1$ sont inverses l'une de l'autre.

Comme $n\theta(B;Y(1\mapsto n))$ suit une loi binomiale, la convergence de $\theta(B;Y(1\mapsto n))$ vers sa limite $\lambda(B)$ s'obtient aisément par un calcul direct.

PROPOSITION 2-6:
Pour tout ensemble B mesurable de A, la fréquence expérimentale $\theta(B;Y(1\mapsto n))$ converge en moyenne quadratique vers $\lambda(B)$.

De cette proposition il résulte que toute combinaison linéaire de variables $\theta(B;Y(1\mapsto n))$ avec $B\in A$ converge en moyenne quadratique vers sa limite.

3-PERTE ET APPROXIMATIONS QUADRATIQUES

Comme $\theta(Y(1\mapsto n))$ estime la vraie valeur λ du paramètre et remplace donc celui-ci dans les applications, une erreur s'introduit qu'il faut évaluer. La perte W à définir doit représenter le coût pour l'expérimentateur de la différence entre $\theta(Y(1\mapsto n))$ et λ.
Cependant le choix du type d'estimateur, ici l'estimateur du maximum de vraisemblance s'effectue par une comparaison sur le risque R définit comme la perte moyenne R=E(W) lorsque l'échantillon $Y(1\mapsto n)$ est aléatoire.
Il s'agit d'utiliser les définitions classiques sur la fonction de perte et la fonction de risque R pour indiquer les limites de cette théorie et de la changer éventuellement afin d'avoir des expressions utilisables par la suite.

PROPOSITION 3-1:
Si $f(x;\theta)=\sum_{A\in M}\frac{\theta(A)}{\mu(A)}\,1_A(x)$ avec $\theta\in I$, alors la fonction de perte:

$$(3.1)\quad W(\lambda,\theta)=E_\lambda[\Phi(\frac{f(X;\theta)}{f(X;\lambda)})]$$

$$=\sum_{A\in M}\lambda(A)\,\Phi(\frac{\theta(A)}{\lambda(A)}) \qquad \text{où}$$

$$\Phi(x)=-2\,\mathrm{Log}(x)$$

est positive ou nulle et s'annule si et seulement si $\theta=\lambda$.
Nous utilisons la fonction

$$(3.2)\qquad \Phi(x)=4\,(\sqrt{x}-1)^2$$

convexe et positive. La proposition 3-1 reste valable

La forme générale de la fonction de perte pour comparer λ et $\theta(K)$, s'écrit:

$$(3.3)\quad W(\lambda,\theta(K))=E_\lambda[\Phi(\frac{d\theta(K)}{d\mu(K)}/\frac{d\lambda}{d\mu})]$$

$$=\sum_{A\in M}\lambda(A)\,\Phi(\frac{\theta(B(A))}{\mu(B(A))}/\frac{\lambda(A)}{\mu(A)})$$

Dans cette expression, chaque ensemble A de la partition M détermine de façon unique un ensemble B, noté B(A), de la sous-partition π

mesurable, c'est l'ensemble B∈K telle que A⊆B.

PROPOSITION 3-2:
Pour toute sous-partition mesurable K de M et pour tout paramètre θ(K), la perte W(λ,θ(K)) est majorée par 8.

Il est naturel de prendre comme probabilité μ, la probabilité représentant la connaissance a priori sur la structure statistique. Dans le cas d'un ensemble Ω=[0,1], la mesure de Lebesgue représente l'absence de connaissance a priori.
Nous introduisons dans R^m, un produit scalaire $\langle , \rangle$ et une norme $\| \|$ qui transforment cet espace en un espace de Hilbert et nous remplaçons la perte W(λ,θ(K)) par une expression approchée $\|\theta(K)-\lambda\|^2$.

DEFINITION 3-3:
Soient K et H, deux sous-partitions mesurables de M et soit T(K), l'opérateur de $\mathcal{I}(K)$ dans $\mathcal{I}$ défini par

$$(3.4)\quad [T(K)\theta](A)=T(\theta;K)(A)=\frac{\theta(B(A))}{\mu(B(A))}\,\mu(A)$$

Le produit scalaire de θ(K) avec θ(H)', est défini par:

$$(3.5)\quad \langle\theta(K),\theta(H)'\rangle=\sum_{A\in M}\frac{1}{\lambda(A)}\,T(\theta;K)(A)\;T(\theta';H)(A)$$

Nous observons que θ(K) défini uniquement sur les ensembles B de K, doit être prolongé aux ensembles A de M afin de pouvoir effectuer une comparaison entre θ(K) et λ; c'est le rôle de l'opérateur T qui utilise la probabilité a priori μ pour étendre θ(K) par règle de trois.
Par ailleurs, l'erreur entre W(λ,θ(K)) et $\|\lambda-\theta(K)\|^2$ fait intervenir la grandeur:

$$(3.6)\quad |\theta(K)-\theta(H)|_\infty=$$
$$=\sup\left(\frac{1}{\lambda(A)}\,|T(\theta;K)(A)-T(\theta';H)(A)|\;;A\in M\right)$$
$$=\sup\left\{\frac{1}{\lambda(A)}\left|\frac{\theta(B(A))}{\mu(B(A))}-\frac{\theta'(C(A))}{\mu(C(A))}\right|\mu(A)\;;\;A\in M,B(A)\in K,C(A)\in H\right\}$$

En particulier, prenons θ(K)=λ(K), H=M et θ'=λ d'où

$$(3.7)\quad |\lambda(K)-\lambda|_\infty=\sup\left\{\left|\frac{\lambda(B(A))}{\mu(B(A))}\,\frac{\mu(A)}{\lambda(A)}-1\right|\;;A\in M\right\}$$

Avec les notations introduites, le risque s'écrit:

$$(3.8)\quad W(\lambda,\theta(K))=\|\theta(K)-\lambda\|^2\;(1+O(|\theta(K)-\lambda|_\infty))$$

PROPOSITION 3-4:
Si pour ϵ donné positif, $|\lambda(K)-\lambda|_\infty\leqslant\epsilon$ et si $\theta_n(K)$ est un estimateur de λ(K), convergent alors les variables aléatoires $W(\lambda,\theta_n(K))$ et $\|\theta_n(K)-\lambda\|^2$ ont asymptotiquement les mêmes lois de probabilités à ϵ près.

Il faut noter que la limite de $W(\lambda,\theta_n(K))$ vaut W(λ,λ(K)).
Le risque $W(\lambda,\theta_n(K))$ compare les structuresassociée à deux partitions différentes K et M.
Ce risque n'a aucune raison d'être nulle sauf dans certains cas particuliers; par contre, comme les deux normes $|\ |_\infty$ et $\|\ \|$ sont équivalentes, dès que $|\lambda(K)-\lambda|_\infty$ est petit, il en est de même pour $\|\lambda(K)-\lambda\|$ et donc pour W(λ,λ(K)).

PROPOSITION 3-5:
Pour tout ϵ donné positif, soit K une sous-partition mesurable de M telle que $|\lambda(K)-\lambda|_\infty\leqslant\epsilon$, alors à ϵ près, la variable aléatoire $n\,\|\theta(Y(1\mapsto n);K)-\lambda(K)\|^2$ suit asymptotiquement une loi du χ^2 à k-1 degrés de liberté.

4-RISQUE:

Par hypothèse, le risque R est la moyenne de la fonction de perte W: pour chaque échantillon Y(1↦n) de la variable X, un estimateur $\theta_n(K)$ du paramètre λ est défini. Dans ce texte, θ(Y(1↦n)) est l'estimateur du maximum de vraisemblance. Lorsque λ(K) est remplacé par $\theta_n(K)$, le coût de l'erreur est évalué par $W(\lambda,\theta_n(K))$. Si de nombreuses expériences ont lieu, c'est le coût moyen R appellé risque qui intervient soit:

$$(4.1)\quad R(n;\lambda,K)=E(W(\lambda,\theta_n(K)) \quad \text{où}$$

$W(\lambda,\theta_n(K))$ est défini en (3.3).
D'après la Prop 3-2, la perte W est majorée par 8 et par conséquent, l'espérance mathématique de celle-ci est définie.

PROPOSITION 4-1:
Soit K, une sous-partition mesurable de M, si l'estimateur $\theta_n(k)$ converge presque sûrement vers λ(K) alors le risque R(n;λ,K) est défini et converge vers:

$$(4.2)\quad R(\infty;\lambda,K)=$$
$$=4\sum_{A\in M}\lambda(A)\left\{\left(\frac{\lambda(B(A))}{\mu(B(A))}\,\frac{\mu(A)}{\lambda(A)}\right)^{\frac{1}{2}}-1\right\}^2$$

PROPOSITION 4-2:
Soit K, une sous-partition mesurable de M, si l'estimateur $\theta_n(K)$ converge en moyenne quadratique vers λ(K) alors $\|\theta_n(K)-\lambda\|^2$ converge en moyenne vers la variable certaine $\|\lambda(K)-\lambda\|^2$.

Le risque asymptotique R(∞;λ,K) peut être approché par $\|\lambda(K)-\lambda\|^2$.

5-CRITERE D'AKAIKE:

Le choix de la sous-partition K de M dépend du nombre n, taille de l'échantillon Y(1↦n) à la disposition de l'expérimentateur; cette sous-partition K devrait rendre minimum le risque R(n;λ,K) (4.1).
Comment à l'aide d'un seul échantillon approcher ce risque R(n;λ,K) où, de plus, la probabilité λ est inconnue afin de savoir, s'il y a lieu ou non, de procéder à un nouveau découpage?
Tout d'abord

$$R(n;\lambda,K)\overset{\Delta}{=}E(W(\lambda,\theta(Y(1\mapsto n;K)))$$

doit être approché par

$$\|\theta(Y(1\mapsto n;K)-\lambda\|^2$$

L'estimateur du maximum de vraisemblance $\theta(B;Y(1\mapsto n))$ de $\lambda(B)$ vaut zéro avec la probabilité $(\lambda(B))^n$ et par conséquent la dérivée troisième $\Phi(x)'''=-3\,x^{\frac{5}{2}}$ de $\Phi(x)$ doit être considérée sur $[0,+\infty[$ intervalle sur lequel elle n'est pas bornée.

PROPOSITION 5-1:

Soit $c \overset{\Delta}{=} \{|\Phi'''(x)|;|x-1|\leqslant|\lambda(K)-\lambda|_\infty\}$, une majoration de la dérivée troisième de la fonction $\Phi(x)=4\,(\sqrt{x}-1)^2$ alors il existe une constante C telle que:

$$|R(n;\lambda,K)-E(\|\theta(K)-\lambda\|^2)|\leqslant c\,|\lambda(K)-\lambda|_\infty^3 + E(\|\theta(K)-\lambda(K)\|)\;(C+E(\|\theta(K)-\lambda(K)\|))$$

PROPOSITION 5-2:

Pour toute sous-partition mesurable K de M, la structure Hilbertienne sur $\mathbf{R}^m$ vérifie pour tout $\theta(K)$ de $\mathcal{I}(K)$ et tout λ de $\mathcal{I}(M)$:

$$(5.1)\quad \|\theta(K)-\lambda\|^2=\|\theta(K)-\lambda(K)\|^2+\|\theta-\theta(K)\|^2 -\|\theta-\lambda-(\theta(K)-\lambda(K)\|^2-2\langle\theta-\lambda,\lambda-\lambda(K)\rangle$$

Pour $\theta(Y(1\mapsto n))$, estimateur du maximum de vraisemblance de λ, le produit scalaire $\langle\theta(Y(1\mapsto n))-\lambda,\lambda-\lambda(K)\rangle$ est d'espérance mathématique nulle.

Prenons l'espérance mathématique des deux membres de (5.1):

$$(5.2)\quad E(\|\theta(K)-\lambda\|^2)=E(\|\theta(K)-\lambda(K)\|^2) +E(\|\theta-\theta(K)\|^2)-E(\|\theta-\lambda-(\theta(K)-\lambda(K))\|^2)$$

Pour simplifier, l'estimateur du maximum de vraisemblance est noté $\theta=\theta(Y(1\mapsto n))$.

D'après la Proposition 3-5, la somme

$$n\,\|\theta(K)-\lambda(K)\|^2$$

suit asymptotiquement, à $|\lambda(K)-\lambda|_\infty$ près, une loi du χ^2 à k-1 degrés de liberté; comme la convergence des $\theta(B)$ vers $\lambda(B)$ a lieu en moyenne quadratique, le premier terme du second membre a pour partie principale $\frac{k-1}{n}$ lorsque n tend vers l'infini.

Nous montrons également que la somme $n\|\theta-\lambda-(\theta(K)-\lambda(K))\|^2$ suit asymptotiquement, à $|\lambda(K)-\lambda|_\infty$ près, une loi du χ^2 à m-k degrés de liberté.

De plus, la loi des n $\theta(A)$ pour A inclu dans B conditionnellement à la tribu

$$\mathcal{B}(K)=\sigma(\theta(B);B\in K)$$

est une loi multinomiale de paramètres n $\theta(B)$ et de fréquences $\frac{\lambda(A)}{\lambda(B)}$.

PROPOSITION 5-3:

Soit K, une sous-partition mesurable de M, soit $\theta(Y(1\mapsto n))$, l'estimateur du maximum de vraisemblance de λ alors:

$$E(\|\theta(Y(1\mapsto n))-\lambda-(\theta(Y(1\mapsto n);K)-\lambda(K))\|^2)$$

a pour partie principale $\frac{m-k}{n}$ lorsque n tend vers l'infini à $\frac{1}{n}\,|\lambda(K)-\lambda|_\infty$.

Le critère pour choisir la meilleure partition K à partir de l'observation $Y(1\mapsto n)$, utilise l'approximation $E(\|\theta(K)-\lambda\|^2)$ de $R(n;\lambda,K)$ puis l'expression (5.2) approchée par:

$$E(\|\theta(K)-\lambda\|^2)\approx \frac{k-1}{n} + E(\|\theta(K)-\theta\|^2)-\frac{m-k}{n}$$

L'espérance mathématique

$$-2\sum_{B\in K}\lambda(B)\;E(\mathrm{Log}\frac{\theta(B)}{\mu(B)})$$

est remplacée par la variable aléatoire:

$$-2\sum_{B\in K}\theta(B)\;\mathrm{Log}\frac{\theta(B)}{\mu(B)}$$

et seuls les termes dépendant explicitement de la partition K, sont conservés dans l'approximation de $E(\|\theta(k)-\theta\|^2)$, soit en définitive:

$$c\overset{\Delta}{=}\frac{2k-1}{n}-2\sum_{B\in K}\theta(B;Y(1\mapsto n))\;\mathrm{Log}\frac{\theta(B;Y(1\mapsto n))}{\mu(B)}$$

II-6-CONCLUSION

Le critère introduit dans ce texte est original et est fonction du nombre k d'ensembles retenus pour calculer l'histogramme, et du nombre n d'observations.

Le critère présentée fournit une approximation statistique de:

$$\frac{2k-1}{n}-\sum_{B\in K}\lambda(B)\;\mathrm{Log}\;\frac{\lambda(B)}{\mu(B)}$$

Par l'inégalité de Jensen, le second terme décroit avec tout nouveau découpage:

$$-\sum_{B}\lambda(B)\;\mathrm{Log}\frac{\lambda(B)}{\mu(B)}\;\geqslant\;-\sum_{A\in M}\lambda(A)\;\mathrm{Log}\;\frac{\lambda(A)}{\mu(A)}\;\geqslant 0$$

Lorsque la probabilité a priori μ, restreinte à un ensemble B, est égale à la probabilité λ a posteriori, restreinte au même ensemble B, tout nouveau découpage de B est inutile: k augmente d'une unité et la partie en Log ne change pas. Le résultat est même plus précis: si pour tout ensemble A inclu dans B, nous pouvons prédire linéairement la mesure $\lambda(A)$ par la formule:

$$\lambda(A)=\frac{\lambda(B)}{\mu(B)}\;\mu(A)$$

alors la partie en Log n'est pas modifiée pour tout nouveau découpage de B.

La mise en oeuvre du critère s'effectue par découpage successif des ensembles retenus. Le découpage doit être arrêté lorsque la valeur du critère augmente en raison de l'augmentation de l'erreur statistique d'estimation entre $\theta(B)$ et sa limite $\lambda(B)$.

BIBLIOGRAPHIE:

AKAIKE H., Information theory and an Extension of the Maximum Likelihood Principle, Budapest, Akadémia Kiado, Symposium on Information Theory Eds. B.N. Petrov and F. Csaki I973 pp267-281.

de BRUCQ D., Théorie du Signal, modélisation statistique, automatique et traitement, Masson (à paraître).

DELECROIX M., Histogrammes et estimation de la densité, Que sais-je, Presses Universitaires de France, 1983.

KULLBACH S., Information Theory and Statistics, John Wiley and Sons, New York 1959.