

# Statistique Empirique Bayésienne

## Estimation Optimale et Parcimonie

Albert BIJAOU

Université de Nice Sophia Antipolis  
UMR CNRS 6202 Cassiopée  
Observatoire de la Côte d'Azur, BP 4229, 06304 Nice Cedex 04, France  
Albert.Bijaoui@oca.eu

**Résumé** – Le débruitage d'un signal est considéré dans le cadre de la statistique empirique bayésienne. Avec un critère d'estimation basé sur l'espérance a posteriori, on est conduit, dans le cadre d'un bruit gaussien, à l'estimateur optimal de Miyasawa. La mesure est corrigée d'un terme proportionnel à la dérivée logarithmique de la densité de probabilité de cette mesure. La variance de l'erreur d'estimation en dérive. Le pourcentage moyen de réduction de bruit  $R$  peut être estimé à partir de cette loi, ce qui constitue un critère sur la parcimonie du signal étudié. L'estimation de  $R$  doit tenir compte du nombre limité de mesures. Plusieurs méthodes d'évaluation ont été mises en œuvre. Dans le cas d'un bruit blanc, toute transformation unitaire  $U$  conserve cette propriété, mais modifie  $R_U$ . L'optimisation de  $R_U$  conduit au meilleur débruitage. Différentes méthodes d'évaluation de  $R$  sont présentées. Une application sur des images tests illustre cette analyse.

**Abstract** – The signal denoising is considered in the framework of the empirical Bayesian statistics, taking into account a Gaussian noise. The posterior expectation leads to the Miyasawa optimal estimate. The measurement  $y$  is corrected with a term proportional to the logarithmic derivative of the posterior probability density function  $q(y)$ . The variance of the estimate is then deduced. The noise variance is reduced by a factor  $R$  which can be obtained only from  $q(y)$ . The  $R$  estimation has to take into account the limited number of measurements. Different methods are proposed. In the case of a white noise, any unitary transform  $U$  keeps this property. The best signal denoising can be derived from the  $R_U$  maximization.

## 1 La relation de Miyasawa.

L'approche statistique dite bayésienne empirique a été introduite par Robbins [1]. Les estimateurs associés sont basés sur la formule de Bayes, mais en partant, comme en statistique classique, des données observées, via la loi a posteriori. De nombreux travaux ont été publiés sur cette approche [2]. Dans ce contexte l'un des problèmes principaux est celui du filtrage.

Considérons une variable aléatoire  $x$ , de loi a priori  $p(x)$  dont la mesure est  $y$ .  $P(y/x)$  désigne la loi de dispersion des mesures. La loi a posteriori  $Q(x/y)$  est obtenue par la règle de Bayes :

$$Q(x/y) = \frac{p(x)P(y/x)}{q(y)}, \quad (1)$$

où

$$q(y) = \int_D p(x)P(y/x)dx \quad (2)$$

est la loi observée de la mesure  $y$  et  $D$  le domaine de définition de la variable  $x$ . Par la suite cet interval sera l'ensemble des réels.

Dans l'estimation bayésienne classique, l'estimation de  $x$  après mesure est  $\hat{x}_{MAP}$ , le maximum de  $Q(x/y)$ . Nous considérons ici l'espérance bayésienne :

$$\hat{x} = \frac{\int_D xp(x)P(y/x)dx}{q(y)}. \quad (3)$$

Il est facile de montrer que c'est l'estimateur qui minimise le carré de l'erreur moyenne :

$$C_2 = E[(\hat{x} - x)^2]. \quad (4)$$

Miyasawa [3] a montré que, dans le cas d'un bruit gaussien,  $\hat{x}$  dérive très simplement de la distribution de mesures  $y$ . En effet nous avons :

$$P(y/x) = \frac{1}{\sqrt{2\pi N}} e^{-\frac{(y-x)^2}{2N}}. \quad (5)$$

$\hat{x}$  est donc :

$$\hat{x} = \frac{\int xp(x)e^{-\frac{(y-x)^2}{2N}} dx}{\int p(x)e^{-\frac{(y-x)^2}{2N}} dx}. \quad (6)$$

En tenant compte que  $x = y + (x - y)$  et :

$$ue^{-\frac{u^2}{2N}} = -N \frac{de^{-\frac{u^2}{2N}}}{du}, \quad (7)$$

la relation (6) peut s'écrire comme :

$$\hat{x} = y + N \frac{q'(y)}{q(y)}, \quad (8)$$

ou

$$\hat{x} = y + N(\log q(y))'. \quad (9)$$

Dans le cas d'une loi a priori gaussienne, la loi a posteriori est aussi gaussienne :

$$q(y) = \frac{1}{\sqrt{2\pi(S+N)}} e^{-\frac{y^2}{2(S+N)}}. \quad (10)$$

On en déduit donc :

$$(\log q(y))' = -\frac{y}{S+N}, \quad (11)$$

ce qui conduit à :

$$\hat{x} = y - \frac{N}{S+N}y = \frac{S}{S+N}y. \quad (12)$$

Ceci correspond exactement au filtre de Wiener. La relation de Miyasawa peut être considérée comme la généralisation du filtre de Wiener au cas d'une variable de loi *a priori* quelconque, mais toujours dans le cas d'un bruit gaussien.

## 2 Un critère de parcimonie d'une distribution.

### 2.1 Réduction moyenne de la variance.

Par construction cette approche permet de réduire en moyenne la variance de l'écart entre  $x$  et  $\hat{x}$ . Considérons les moments :

$$m_k = \int_D x^k p(x) G(x-y) dx. \quad (13)$$

Par dérivation on obtient :

$$m'_k = \frac{m_{k+1} - y m_k}{N}. \quad (14)$$

Ce qui permet de déduire :

$$m_2 = N m'_1 + y m_1, \quad (15)$$

avec :

$$m_1 = N q' + y q \quad (16)$$

et :

$$m'_1 = N q'' + y q' + q. \quad (17)$$

D'où :

$$m_2 = N^2 q'' + 2N y q' + (N + y^2) q. \quad (18)$$

La variance résiduelle est donc :

$$M_2 = \frac{m_2}{q} - \hat{x}^2. \quad (19)$$

Ce qui conduit à :

$$M_2 = N^2 \log'' q(y) + N. \quad (20)$$

$\log'' q(y)$  n'est pas nécessairement toujours négatif. Pour certaines valeurs, l'estimation peut donc augmenter l'erreur. On peut examiner la variance moyenne :

$$D = N - N^2 \int_D \frac{(q')^2(y)}{q(y)} dy. \quad (21)$$

La variance moyenne de la mesure  $D$  est donc toujours réduite après filtrage. La réduction en pourcentage est :

$$R = N \int_D \frac{(q')^2(y)}{q(y)} dy. \quad (22)$$

### 2.2 Quelques propriétés.

Dans le cas d'une distribution *a priori* gaussienne, on obtient aisément :

$$R = \frac{N}{S + N}. \quad (23)$$

La réduction du bruit est d'autant plus forte que le rapport signal sur bruit (RSB) est faible.

Considérons l'inégalité de Schwarz :

$$\left[ \int_D y \log' q(y) dF \right]^2 \leq \int_D y^2 dF \int_D (\log' q(y))^2 dF, \quad (24)$$

$F(y)$  étant la fonction de répartition. On en déduit :

$$R \geq \frac{N}{V}, \quad (25)$$

$V$  désignant la variance de la mesure. L'égalité n'est possible que dans le cas d'une distribution gaussienne, où  $\log' q(y) = -y/V$ . À variance donnée  $V$ , c'est avec une distribution *a priori* gaussienne qu'on minimise la réduction du bruit.

Lorsque la variance du bruit  $N$  tend vers 0,  $q(y)$  tend vers  $p(y)$ .  $R$  tend donc vers  $NI$  où :

$$I = \int_D \frac{(p')^2(y)}{p(y)} dy. \quad (26)$$

En appliquant la relation (25) on en déduit  $I \geq 1/V_p$ , où  $V_p$  désigne la variance de la loi *a priori*. Si  $N$  est très faible, toutes les valeurs sont statistiquement significatives,  $R$  est proche de 1 et le signal n'est pas du tout parcimonieux.

Si  $N$  devient très grand,  $q(y)$  tend vers la loi gaussienne associée au bruit (au décalage de la moyenne près).  $R$  tend donc vers  $N/(S + N)$  où  $S$  est négligeable par rapport à  $N$ .  $R$  tend donc vers 1. Si le bruit est très grand, il n'y a plus que très peu de valeurs significatives et  $R$  devient proche de 1.

Compte tenu de son comportement par rapport aux valeurs significatives,  $R$  correspond bien à un critère de parcimonie.

### 2.3 Effet d'une transformation unitaire.

Du point de vue statistique, c'est avec cette approche qu'on obtient le meilleur débruitage des mesures. Ceci peut être exploité pour débruiter un signal ou une image [4] [5]. Dans ce cas, il est bien évident que les signaux doivent être préalablement transformés de manière à ce que la distribution des coefficients soit la plus parcimonieuse possible, la parcimonie étant alors comprise comme un nombre faible de coefficients. En concentrant l'information dans le minimum de coefficients, on conçoit aisément qu'on peut mieux estimer le signal.

L'optimisation du débruitage du signal a conduit à la mise en œuvre de méthodes parcimonieuses, basées sur le critère naturel de la minimisation du nombre de coefficients significatifs. Le développement théorique précédent montre que l'effet du filtrage optimal, sous l'angle de l'estimation bayésienne empirique, peut être parfaitement quantifié par la quantité  $R$ . Considérons un signal discret  $s(k)$  et transformons le avec une matrice unitaire  $U$ . Ceci conduit à une transformée de coefficients  $y(l)$ . La transformation unitaire conserve la nature du bruit blanc gaussien, par contre la distribution des coefficients varie avec  $U$ . La statistique des coefficients conduit à la quantité  $R_U$ .

Pour le signal  $s$ , la transformée la plus parcimonieuse est celle qui minimise  $R_U$ , si on ne tient pas compte de la différence de répartition des coefficients transformés. Considérons en effet un signal formé de deux composantes de proportion  $a_1$  et  $a_2$  ( $a_1 + a_2 = 1$ ) et de lois  $q_1(y)$  et  $q_2(y)$ . Si on débruite avec la loi moyenne  $a_1 q_1(y) + a_2 q_2(y)$  on obtient une réduction  $R$ , alors que séparément on obtient  $R_1$  et  $R_2$ . On montre aisément que  $R \leq a_1 R_1 + a_2 R_2$ . En général, les coefficients transformés n'ont pas la même distribution et il est donc naturel de débruiter séparément les différents coefficients. Ceci n'est possible que si on applique un estimateur de  $R$  robuste au nombre de valeurs considérées.

Toute la difficulté réside dans la construction d'un tel estimateur.

### 3 Estimation du critère de parcimonie.

$R$  peut s'écrire :

$$R = N \int_D \left( \frac{q'(y)}{q(y)} \right)^2 dF, \quad (27)$$

où  $F$  désigne la fonction de répartition.

L'estimation nécessite celle de  $q(y)$  et  $q'(y)$  à partir des coefficients. L'utilisation de l'histogramme ne fournit en général que des estimations très bruitées et inexploitables. Plusieurs voies ont été explorées :

**Somme de gaussiennes** La loi de distribution est décomposée en gaussiennes [6]. De ce fait, il suffit d'estimer les paramètres associés. Une partie de l'information peut disparaître, par un choix trop limité du nombre de composantes. La décomposition en somme de gaussiennes est sans doute l'une des voies les plus élégantes pour résoudre ce problème.

**Pyramide de lissages.** L'histogramme est une version bruitée de la loi de distribution. Pour obtenir une estimation suffisamment précise de la densité, il faut que le nombre d'événements dépasse un seuil  $n$ , fixé initialement. Une série de lissage est effectuée, par un algorithme à trous [7], de manière à avoir pour chaque valeur au moins  $n$  valeurs utilisées. Ceci permet de déduire une bonne estimation de la distribution et de sa dérivée [4].

**Distribution de Voigt** L'application de l'algorithme précédent a montré sur plusieurs images astronomiques que la distribution des coefficients était plutôt du type Voigt (convolution d'une loi de Gauss avec une loi de Cauchy) [4]. Les coefficients de la loi peuvent être estimés avec le principe de maximum de vraisemblance. Sur les images considérées la partie "Cauchy" est marginale par rapport à la partie gaussienne. Les paramètres peuvent être estimés grossièrement en calculant d'abord les paramètres de la gaussienne, avec un rejet itératif à  $k - \sigma$  [8]. Le pourcentage de valeurs rejetées permet ensuite d'estimer le paramètre de Cauchy.

**Utilisation des quantiles** Les coefficients  $y$  sont ordonnés en valeur croissante. On peut alors exprimer  $R$  en fonction des dérivées des quantiles. Les quantiles peuvent être obtenus par méthodes variationnelles, sans rangement préalable [9].

**Ajustement local** Raphan & Simoncelli modélisent la loi de distribution autour d'une valeur  $y$  par une loi exponentielle tronquée [5]. Le calcul de sa dérivée logarithmique en devient immédiat. On peut aussi faire un ajustement local avec une loi gaussienne tronquée, ce qui conduit à un filtre de Wiener local. Cet algorithme est très sensible au nombre de valeurs.

**Méthode de Parzen** Il s'agit d'estimer la loi de la densité en introduisant une fenêtre  $\phi(y)$  avec l'expression [10] :

$$q(y) = \sum_l \phi(y - y_l). \quad (28)$$

Le choix de la fenêtre est critique. La transformée en ondelettes permet d'obtenir une meilleure approximation [11].

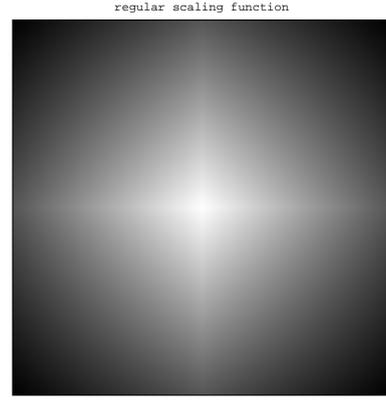


FIG. 1 – Fonction d'échelle régulière.

Ces différentes méthodes ont été exploitées sur plusieurs images et sur leurs transformées.  $R$  est dominé par les termes qui maximisent la dérivée seconde du logarithme de la distribution, c'est-à-dire la partie centrale. Avec l'algorithme du rejet itératif à  $k - \sigma$ , on détermine les paramètres d'une distribution gaussienne qui colle bien à la région centrale.  $R$  est bien estimé par  $N/V_k$ , où  $V_k$  est la variance obtenue. Par contre, seule une fraction  $\alpha_k$  des valeurs a été utilisée. Les valeurs rejetées, plus lointaines, contribuent de manière négligeable à l'intégrale. Nous proposons donc comme estimateur  $R_k = \alpha_k N/V_k$ .

Cette quantité dépend du seuil  $k$  de rejet. Avec un seuil inférieur à 3, le biais dans la variance est important, alors qu'au-dessus de 4, le nombre de valeurs rejetées, dans le cas d'une loi gaussienne, est faible. Dans nos expériences nous avons utilisé  $k = 3, 5$ .

## 4 Exemple d'application sur des images et leurs transformées.

### 4.1 Images tests.

Pour clarifier l'exposé, deux images issues de simulation ont été choisies pour ces tests. Elles correspondent à des fonctions d'échelle satisfaisant l'équation :

$$\begin{aligned} \frac{1}{4} \phi\left(\frac{x}{2}, \frac{y}{2}\right) &= 0,25 \phi(x, y) + \\ a[\phi(x-1, y) + \phi(x+1, y) + \phi(x, y-1) + \\ \phi(x, y+1)] &+ b[\phi(x-1, y-1) + \phi(x+1, y-1) + \\ \phi(x-1, y+1) + \phi(x+1, y+1)]. \end{aligned} \quad (29)$$

Pour converger, on doit avoir  $a + b = 0,1875$ . Avec la valeur  $a = 0,125$  et  $b = 0,0625$ , on obtient l'image régulière de la Figure (1), correspondant au produit de la fonction triangle en  $x$  et en  $y$ .

Avec  $a = 0,1275$  et  $b = 0,06$ , on obtient l'image à la structure fractale de la Figure (2). On peut noter la très grande sensibilité du résultat avec le paramètre  $a$ .

Ces images sont ensuite bruitées avec un bruit gaussien d'écart-type 0,001, 0,01, 0,1 et 0,5. La fractalité sur la seconde image disparaît avec l'accroissement du bruit. Le paramètre  $R$  est ensuite calculé sur chaque image et sur leurs transformées en ondelettes. La transformation de Haar et celle de Daubechies

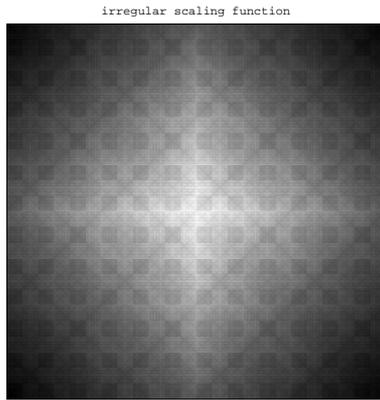


FIG. 2 – Fonction d'échelle irrégulière.

TAB. 1 – Valeurs de  $R$  sur l'image régulière.

Image	0,001	0,01	0,1	0,5
Brute	0,00004	0,004	0,295	0,918
gHaar	0,399	0,930	0,996	0,996
mHaar	0,665	0,959	0,994	0,996
gD8	0,989	0,996	0,996	0,996
mD8	0,987	0,994	0,995	0,995

à 8 coefficients sont appliquées [12].  $R$  a été calculé sur les transformations globalement (g) et avec une moyenne pondérée sur chaque type de coefficients (m). Quatre échelles ont été utilisées.

Sur les tableaux 1 et 2, on a résumé les résultats obtenus. On constate :

- L'effet de la transformation sur la parcimonie en fonction du RSB. À faible bruit, la réduction avec une transformation plus riche est beaucoup plus efficace. À fort bruit, les résultats sont similaires, limités par la précision de l'algorithme d'estimation.
- $R$  est beaucoup plus élevé sur l'image régulière, dont les transformées sont parcimonieuses, que sur l'image irrégulière. À fort bruit l'effet des irrégularités disparaît.

D'autres transformations unitaires, associées à l'algorithme de construction des images, devraient conduire à des valeurs de  $R$  plus grandes. La recherche de la meilleure transformation peut s'examiner sous cet angle. Rappelons qu'à l'ordre 2, on retrouve le filtre de Wiener. Si on considère un environnement local, la meilleure transformation est l'expansion de Karhunen-Loève, associée aux vecteurs propres de la matrice d'autocorrélation. Il s'agira donc d'étendre cet approche à des ordres élevés.

TAB. 2 – Valeurs de  $R$  sur l'image irrégulière.

Image	0,001	0,01	0,1	0,5
Brute	0,00004	0,004	0,281	0,912
gHaar	0,0057	0,351	0,977	0,996
mHaar	0,0087	0,445	0,981	0,996
gD8	0,008	0,435	0,985	0,996
mD8	0,012	0,528	0,988	0,995

## 5 Conclusion.

Dans cette communication, nous avons tracé une nouvelle voie pour quantifier la parcimonie d'un signal. Elle dérive de l'approche statistique bayésienne empirique. La relation de Miyasawa permet d'obtenir le meilleur filtrage d'un signal donné, à condition que celui-ci soit transformé de manière optimale. Le critère de parcimonie proposé tient donc compte d'une part de la transformation effectuée et, d'autre part, de la variance du bruit, supposé blanc gaussien.

Toute la difficulté de la mise en œuvre de ce critère réside dans l'estimation de la loi de distribution de la mesure et dans celle de sa dérivée. Dans les expériences effectuées, il est apparu qu'une simple modélisation pouvait suffire. Les résultats obtenus sur des images tests mettent en relief le rôle de la transformation en fonction du rapport signal sur bruit. La recherche de la meilleure transformation associée à une image donnée peut s'examiner dans ce cadre.

## Références

- [1] H. Robbins, *An empirical bayes approach to statistics*, Proc. Third Berkley Symposium on Mathematical Statistics, 1, 157–163, 1956.
- [2] J. S. Maritz and T. Lwin, *Empirical Bayes Methods*, Chapman & Hall, 2nd ed., 1989.
- [3] K. Miyasawa, *An empirical bayes estimator of the mean of a normal population*, Bull. Inst. Internat. Statist., 38, 181–188, 1961.
- [4] A. Bijaoui, *A Universal Bayesian Estimate. Application to denoising*, Astronomical Data Analysis IV, ed. C. Surace, [http://www.oamp.fr/conf/ada4/pub/19/DSD-ADAIV\\_Bijaoui.pdf](http://www.oamp.fr/conf/ada4/pub/19/DSD-ADAIV_Bijaoui.pdf), 2006.
- [5] M. Raphan, E.P. Simoncelli, *Empirical Bayes Least Squares Estimation without an Explicit Prior*, Computer Science Technical Report TR2007-900, Courant Institute of Mathematical Sciences, New York University, May 2007.
- [6] A. Bijaoui, *Wavelets, Gaussian Mixtures and the Wiener Filtering*, *Signal Processing*, 82, 709-712, 2002
- [7] J.L. Starck, F. Murtagh, A. Bijaoui, *Image processing and data analysis. The multiscale approach*, p.21, C.U.P., 1998.
- [8] J.L. Starck, F. Murtagh, A. Bijaoui, *Image processing and data analysis. The multiscale approach*, p.48, C.U.P., 1998.
- [9] R. Koenker, G. Bassett, *Quantile regression*, *Econometrica*, 46 :1, 33-50, 1978.
- [10] E. Parzen, *On estimation of a probability density function and mode*, *Ann. Math. Stat.* 33, 1065-1076, 1962.
- [11] I. M. Johnstone, G. Kerkyacharian, D. Picard, *Estimation d'une densité de probabilité par méthode d'ondelettes*, *C.R.Ac.Sc. Paris*, A315, 211-216, 1992.
- [12] I. Daubechies, *Orthonormal bases of compactly supported wavelets*, *Comm. Pure & Appl. Math.*, 41 (7), 909-996, 1988.