

Estimation de la moyenne et statistiques d'ordre

Steeve Zozor, Eric Moisan et Pierre-Olivier Amblard

CEPHAG, ENSIEG, Domaine Universitaire, B.P. 46
38402 Saint Martin d'Hères, France

RÉSUMÉ

Nous étudions le problème de l'estimation de la moyenne d'un échantillon de n variables aléatoires, en utilisant leurs statistiques d'ordre. Nous avons en cela complété le travail de Bovik qui concernait des variables indépendantes, au cas de variables corrélées, notamment en étendant les estimateurs basés uniquement sur les statistiques d'ordre à de nouveaux estimateurs, basés à la fois sur les variables et leurs statistiques d'ordre. Nous montrons ainsi qu'on peut améliorer l'estimation de la moyenne en tenant compte de la densité de probabilité et de la corrélation des variables, sans connaissances *a priori*. Enfin un algorithme adaptatif est donné, et une application pratique est présentée.

ABSTRACT

We study in this paper the estimation of the mean using the order statistics of a sample of n random variables. This kind of estimating have been done by Bovik for independent identically distributed variables. In this paper we extend this work to correlated variables. In particular we extend this kind of estimator to a new estimator using simultaneously the variables and their order statistics. We show that this new estimator performs better than the previous one by "learning" the correlation and the probability density function of the variables, without *a priori* knowledge. At last an adaptive algorithm is given and a practical application is presented.

1 Introduction.

En traitement du signal on a souvent besoin d'estimer la moyenne d'une variable aléatoire à partir d'une séquence de longueur finie $\{X_i\}$. Cette estimation est usuellement effectuée à partir de la moyenne empirique. Il est en effet bien connu que cet estimateur est le meilleur (au sens de l'erreur quadratique moyenne (EQM) minimum) lorsque les variables sont gaussiennes, indépendantes, identiquement distribuées (iid). Mais cette propriété n'est plus valable si l'une au moins des hypothèses n'est plus vérifiée. Considérons par exemple des variables iid, uniformes; l'estimateur du maximum de vraisemblance fait alors apparaître deux statistiques d'ordre [5] : $\hat{m}_{mle} = \frac{Min_i X_i + Max_i X_i}{2}$. Afin d'estimer la moyenne sans connaissances a priori (contrairement à la méthode du maximum de vraisemblance), Bovik [1] a travaillé sur des estimateurs non linéaires, mais linéaires suivant les statistiques d'ordre (OS). Il a étudié ces estimateurs uniquement pour des séquences iid, et a montré que le meilleur estimateur, non biaisé, linéaire suivant les statistiques d'ordre (BOSUE) est au moins aussi bon que la moyenne empirique.

2 Résultats préliminaires.

Considérons une séquence de n variables aléatoires X_1, \dots, X_n . Les statistiques d'ordre sont définies en ordonnant ces variables $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. $X_{(i)}$ est alors appelée i ème statistique d'ordre de l'échantillon. Ce procédé non linéaire complique considérablement l'étude, mais de nombreux résultats peuvent être trouvés dans l'ouvrage de David [2]. On notera m l'espérance mathématique des X_i . On peut donc écrire $X_i = m + B_i$, où B_i est une variable aléatoire

centrée. On cherche alors un estimateur de m sous la forme $\hat{m} = \underline{a}^t \underline{\tilde{x}}$ en utilisant les notations suivantes :

\underline{x}	$= [X_1, \dots, X_n]^t$	vecteur des échantillons
$\underline{\tilde{x}}$	$= [X_{(1)}, \dots, X_{(n)}]^t$	vecteur des OS
\underline{b}	$= [B_1, \dots, B_n]^t$	vecteur des données centrées
$\underline{\tilde{b}}$	$= [B_{(1)}, \dots, B_{(n)}]^t$	vecteur des OS des B_i
$\underline{1}$	$= [1, \dots, 1]^t$	vecteur constant
\underline{a}	$= [a_1, \dots, a_n]^t$	coefficients de l'estimateur
$\underline{\Gamma}$	$= E[\underline{b} \underline{b}^t]$	matrice de covariance de \underline{b}
\underline{R}	$= E[\underline{\tilde{b}} \underline{\tilde{b}}^t]$	matrice de corrélation de $\underline{\tilde{b}}$

Considérons dans un premier temps que les variables sont indépendantes. En supposant que la densité de probabilité f_b des variables B_i est paire, David [2] et Bovik [1] ont construit le BOSUE de la moyenne m comme suit :

- estimateur non biaisé : $E[\underline{a}^t \underline{\tilde{x}}] = m$
- minimisation de l'erreur quadratique $E[(\underline{a}^t \underline{\tilde{x}} - m)^2]$

La contrainte de non biais s'écrit $\underline{a}^t E[\underline{\tilde{b}}] + m \underline{a}^t \underline{1} = m$. Sous l'hypothèse de parité de f_b on montre que $E[B_{(i)}] = -E[B_{(n+1-i)}]$. Si on suppose que compte tenu de la symétrie du problème les composantes de \underline{a} vérifient $a_i = a_{n+1-i}$, la contrainte se réduit à

$$\underline{a}^t \underline{1} = 1 \quad (1)$$

Il faudra néanmoins vérifier la "symétrie" de \underline{a} a posteriori. En minimisant alors l'erreur quadratique sous la contrainte (1), c'est-à-dire en minimisant le Lagrangien $E[(\underline{a}^t \underline{\tilde{b}})^2] + \lambda(\underline{a}^t \underline{1} - 1)$, et en déterminant le coefficient de Lagrange λ pour vérifier la contrainte, Bovik a abouti à l'estimateur suivant :

$$a_{\text{bosue}} = \frac{\underline{R}^{-1} \underline{1}}{\underline{1}^t \underline{R}^{-1} \underline{1}} \quad (2)$$

On vérifie facilement [2, 3], pour f_b paire, que les composantes de \underline{R} satisfont $R_{i,j} = R_{n+1-i,n+1-j}$ ce qui implique la symétrie de $\underline{a}_{\text{bosue}} \cdot \underline{m}_{\text{bosue}} = \underline{a}_{\text{bosue}}^t \underline{\tilde{x}}$ est donc l'estimateur de la moyenne des variables X_i , non biaisé, linéaire suivant les statistiques d'ordre, et d'erreur quadratique minimum (BOSUE). Par la suite, par abus de langage, nous appellerons \underline{a} "estimateur". On peut remarquer que l'erreur quadratique moyenne est donnée par $EQM_{\text{bosue}} = \frac{1}{\underline{1}^t \underline{R}^{-1} \underline{1}}$. Bovik a montré que pour des variables gaussiennes $\underline{a}_{\text{bosue}}^t \underline{\tilde{x}}$ est la moyenne empirique, et que pour des variables uniformes cet estimateur est $\frac{\text{Min}_i X_i + \text{Max}_i X_i}{2}$. Il s'agit dans les deux cas de l'estimateur du maximum de vraisemblance [5]. De la même manière, étant données n variables aléatoires X_i , le meilleur estimateur, linéaire et non biaisé (BLUE) est défini par

$$\underline{a}_{\text{blue}} = \frac{\underline{\Gamma}^{-1} \underline{1}}{\underline{1}^t \underline{\Gamma}^{-1} \underline{1}} \quad (3)$$

L'erreur quadratique est donnée par $EQM_{\text{blue}} = \frac{1}{\underline{1}^t \underline{\Gamma}^{-1} \underline{1}}$. Si les X_i sont iid, $\underline{\Gamma}$ est proportionnelle à la matrice identité, et donc quelle que soit la densité de probabilité des variables, le BLUE est la moyenne empirique. Il en découle que pour des variables indépendantes, le BOSUE est au moins aussi performant que le BLUE. En effet ce dernier est, dans ce cas, un estimateur particulier linéaire suivant les statistiques d'ordre (il est évident que $\frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n X_{(i)}$).

La question à laquelle nous voulons répondre dans ce papier est donc la suivante : que se passe-t-il si les variables ne sont pas indépendantes ?

3 Estimation de la moyenne d'un bruit coloré.

La figure 1 présente les estimateurs BOSUE et BLUE obtenus dans deux cas particuliers de bruit : gaussien corrélé et uniforme corrélé (cf. [4]).

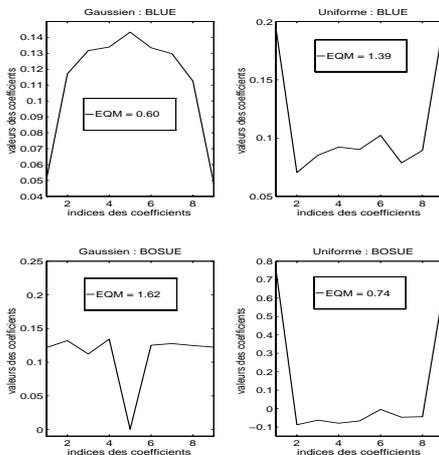


FIG. 1 — BOSUE and BLUE pour des bruits gaussien et uniforme colorés.

On s'aperçoit que cette fois le BOSUE n'est pas toujours meilleur que le BLUE, ceci à cause de la corrélation. Nous

proposons alors de combiner les deux approches (linéaire et OS simultanément), espérant que la partie linéaire de l'estimateur fera l'apprentissage de la corrélation du signal, tandis que la partie non linéaire fera l'apprentissage de la densité de probabilité des variables. Il est entendu que dans ces conditions la taille des vecteurs est doublée. Nous considérons donc le vecteur $\underline{y} = [\underline{x}^t \underline{\tilde{x}}^t]^t$. Nous supposons que la densité de probabilité conjointe des variables centrées est paire : $f_{1,\dots,n}(b_1, \dots, b_n) = f_{1,\dots,n}(-b_1, \dots, -b_n)$. La contrainte de non biais s'écrit alors $\underline{a}^t E[\underline{y}] = m$. On peut montrer que pour des variables corrélées, sous l'hypothèse de symétrie de $f_{1,\dots,n}$, $E[\underline{b}]$ est encore "antisymétrique", ainsi en supposant que la partie de \underline{a} qui s'applique sur les statistiques d'ordre est symétrique, la contrainte de non biais peut toujours être écrite $\underline{a}^t \underline{1} = 1$. Minimiser la variance de l'estimateur sous la précédente contrainte conduit ainsi à l'équation

$$E[[\underline{b}^t \underline{\tilde{b}}^t]^t \cdot [\underline{b}^t \underline{\tilde{b}}^t]] \cdot \underline{a} = -\frac{\lambda}{2} \underline{1} \quad (4)$$

où λ est le coefficient de Lagrange. Malheureusement la matrice de corrélation qui apparaît dans le premier membre de (4) est singulière. En effet on peut voir que $\sum_{i=1}^n B_i = \sum_{i=1}^n B_{(i)}$, et par conséquent

$$E[[\underline{b}^t \underline{\tilde{b}}^t]^t \cdot [\underline{b}^t \underline{\tilde{b}}^t]] \cdot \underline{k} = 0$$

où $\underline{k}^t = [\underbrace{\underline{1}^t}_n \quad \underbrace{-\underline{1}^t}_n]$. On peut montrer que cette matrice est

de rang $2n - 1$ et que, par conséquent, le vecteur $\underline{1}$ étant orthogonal à \underline{k} il appartient à son image : il existe donc des vecteurs \underline{a} satisfaisant à l'équation (4). Compte tenu du rang de $E[[\underline{b}^t \underline{\tilde{b}}^t]^t \cdot [\underline{b}^t \underline{\tilde{b}}^t]]$ nous proposons de réduire l'équation (4) en supprimant la première composante X_1 du vecteur \underline{y} afin d'obtenir une solution unique de taille $2n - 1$. Le meilleur estimateur combinant les variables et les statistiques d'ordre, non biaisé (BLOSUE) est alors

$$\underline{a}_{\text{blosue}} = \frac{\underline{\phi}^{-1} \underline{1}}{\underline{1}^t \underline{\phi}^{-1} \underline{1}} \quad (5)$$

où $\underline{\phi} = E[[B_2 \dots B_n \ B_{(1)} \dots B_{(n)}]^t \cdot [B_2 \dots B_n \ B_{(1)} \dots B_{(n)}]]$ et où le coefficient de Lagrange a été déterminé de telle sorte que la contrainte soit vérifiée. Nous avons choisi de supprimer X_1 mais on pourrait supprimer n'importe quel X_i . Par contre, nous conservons toutes les statistiques d'ordre, afin de ne pas briser la symétrie de la "partie statistique d'ordre" de l'estimateur. Ainsi, de même que précédemment, nous vérifions *a posteriori* que la partie de \underline{a} qui s'applique sur les statistiques d'ordre est bien symétrique. L'erreur quadratique correspondant à cet estimateur (BLOSUE) est alors $EQM_{\text{blosue}} = \frac{1}{\underline{1}^t \underline{\phi}^{-1} \underline{1}}$. La figure 2 nous montre les résultats obtenus par cet estimateur sur les exemples présentés en figure 1.

La figure 3 représente les résultats obtenus pour un cas plus "atypique" : nous avons filtré non linéairement le bruit gaussien corrélé avec le filtre nl

$$nl(x) = \begin{cases} x & \text{si } |x| \leq 8 \\ -0.8x & \text{si } 8 < |x| \leq 10 \\ 0.4x & \text{si } |x| > 10 \end{cases} \quad (6)$$

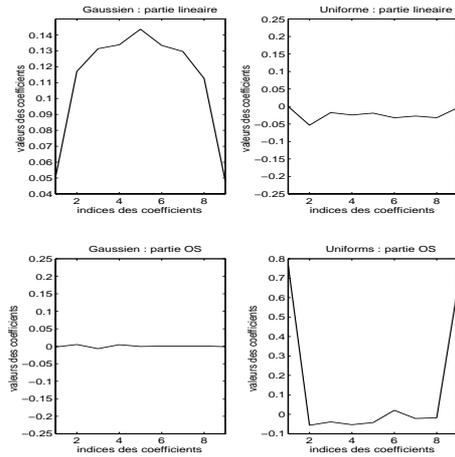


FIG. 2 — Parties linéaires et OS du BLOSUE pour les bruits gaussien et uniforme colorés. Cas gaussien : $EQM_{\text{blosue}} = 0.60 = EQM_{\text{blue}}$; cas uniforme : $EQM_{\text{blosue}} = 0.74 = EQM_{\text{bosue}}$.

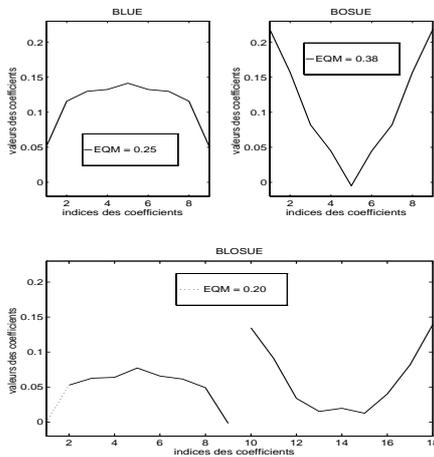


FIG. 3 — BLUE, BOSUE et BLOSUE pour le bruit gaussien coloré filtré non linéaire. $EQM_{\text{blosue}} < EQM_{\text{blue}}$ et $EQM_{\text{blosue}} < EQM_{\text{bosue}}$.

Comme nous l’espérons, le BLOSUE est toujours au moins aussi performant que le BOSUE et le BLUE. En particulier, il “choisit” l’estimation via les OS dans le cas uniforme blanc et via les variables dans le cas gaussien corrélé (supprimer X_1 a pour effet de rendre la partie OS non pas nulle, mais constante). Dans le cas “atypique”, il utilise les deux types de variables simultanément et améliore donc l’estimation.

Dans l’ensemble des exemples précédents, les matrices \underline{R} , $\underline{\Gamma}$ et $\underline{\phi}$ ont été estimées par moyennage empirique sur 9×10^4 observations (soit 10^4 réalisations, de b). En pratique il sera intéressant de disposer alors d’un algorithme adaptatif.

4 Algorithme adaptatif.

Dans la plupart des cas pratiques, la densité de probabilité et la corrélation des données ne sont pas connues. Les matrices de corrélation utilisées par les estimateurs ne pouvant être

calculées, elles seront estimées selon un algorithme adaptatif, similaire à un RLS par exemple. Soit alors \hat{m}_k une estimation de m à l’instant k . On désigne alors par \underline{M} la matrice \underline{R}^{-1} (resp. $\underline{\Gamma}^{-1}$, resp. $\underline{\phi}^{-1}$), et par \underline{v}_k le vecteur \tilde{x}_k (resp. \underline{x}_k , resp. $[x_{k-1} \dots x_{k-n+1} \tilde{x}_k^t]^t$), \underline{x}_k représentant le vecteur contenant les n derniers échantillons à l’instant k , et \tilde{x}_k étant issu de \underline{x}_k après avoir réordonné. \underline{M}_k est alors l’estimation de \underline{M} (via un moyennage empirique et une décomposition de Woodbury) à l’instant k , et \underline{a}_k est le vecteur des coefficients estimateurs évalués à partir de \underline{M}_k . On écrit facilement l’algorithme suivant :

- Etape d’initialisation : \hat{m}_1 et \underline{M}_1 .
- De l’étape k à l’étape $k + 1$:

$$\underline{d}_{k+1} = \underline{v}_{k+1} - \hat{m}_k \underline{1}$$

$$\underline{M}_{k+1} = \frac{k+1}{k} (\underline{M}_k - \frac{\underline{M}_k \underline{d}_{k+1} \underline{d}_{k+1}^t \underline{M}_k}{k + \underline{d}_{k+1}^t \underline{M}_k \underline{d}_{k+1}})$$

$$\underline{a}_{k+1} = \frac{\underline{M}_{k+1} \underline{1}}{\underline{1}^t \underline{M}_{k+1} \underline{1}}$$

$$\hat{m}_{k+1} = \underline{a}_{k+1}^t \underline{v}_{k+1}$$

5 Application pratique.

Supposons à présent que nous voulions estimer une moyenne variable dans le temps, constante par morceaux, bruitée additivement par un bruit centré. Considérons par exemple le cas d’un signal binaire, constant uniquement sur des durées finies. Nous avons ainsi appliqué les trois estimateurs (BLUE, BOSUE, BLOSUE) de manière adaptative sur de tels signaux dans le but d’estimer leur partie signal. Les résultats sont présentés en figure 4 pour le cas où le bruit est uniforme blanc, et en figure 5 pour le cas où le bruit est gaussien corrélé.

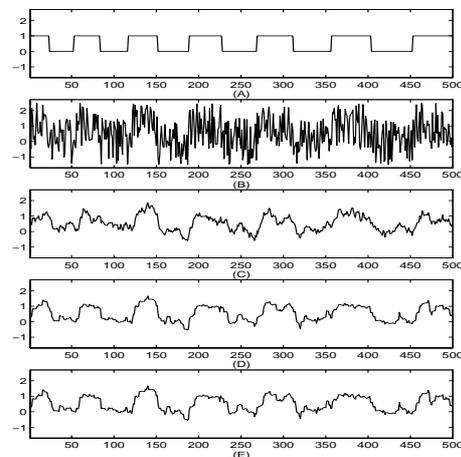


FIG. 4 — Cas iid uniforme : (A), signal ; (B), signal bruité ; (C), estimation BLUE ; (D) estimation BOSUE ; (E) estimation BLOSUE.

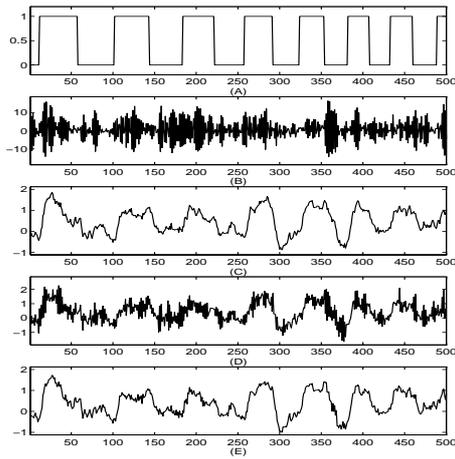


FIG. 5 — Cas gaussien corrélé : (A), signal ; (B), signal bruité ; (C), estimation BLUE ; (D) estimation BOSUE ; (E) estimation BLOSUE.

Nous avons évalué la durée de la phase transitoire en regardant l'évolution de deux coefficients particuliers, ainsi nous ne présentons ici que la phase stationnaire de ces algorithmes. Enfin, les figures 6 et 7 nous montrent les coefficients finaux de ces estimateurs ainsi que l'évolution des deux coefficients observés pour déterminer la phase stationnaire.

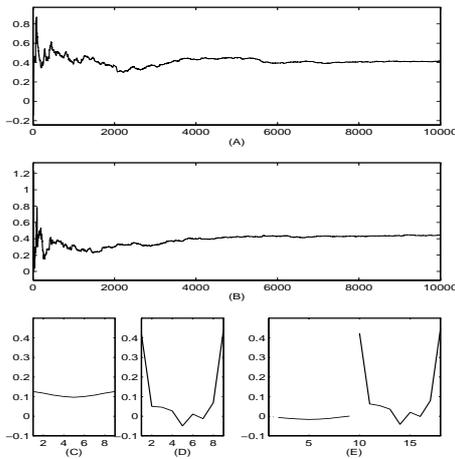


FIG. 6 — Cas iid uniforme : (A) et (B), évolutions de a_{bosue_1} et a_{bosue_n} ; (C) et (D), BLUE et BOSUE finaux ; (E), BLOSUE final (partie linéaire (1ère) et partie OS (2nd) séparées).

Dans le cas du bruit blanc uniforme, on s'aperçoit que le BLOSUE choisit clairement l'estimation via les statistiques d'ordre, qui est le meilleur estimateur dans le cas du bruit blanc. Par contre dans le cas du bruit gaussien corrélé, le BLOSUE choisit l'estimation via les variables non ordonnées, qui est dans ce cas le meilleur estimateur.

6 Conclusion.

Le BLUE et le BOSUE peuvent être facilement "mixés" pour donner le BLOSUE. De plus, ce dernier est toujours le meilleur des trois estimateurs car il est capable d'apprendre la

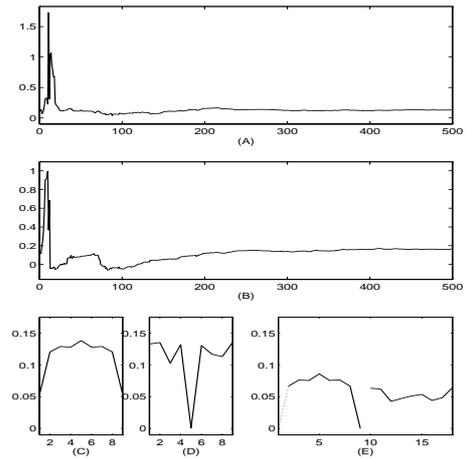


FIG. 7 — Cas gaussien corrélé : (A) et (B), évolutions de a_{bosue_1} et a_{bosue_n} ; (C) et (D), BLUE et BOSUE finaux ; (E), BLOSUE final (partie linéaire (1ère) et partie OS (2nd) séparées).

corrélation des variables (à travers sa partie linéaire) et leur densité de probabilité (à travers sa partie non linéaire). Une solution simple nous a permis de résoudre le problème de singularité de la matrice mise en jeu. Un algorithme adaptatif est alors proposé, il permet en définitive l'apprentissage des caractéristiques des données. Finalement, nous avons illustré l'intérêt de cette méthode d'estimation sur un problème courant.

Références

- [1] A.C. Bovik, T.S. Huang and D.C. Munson *A Generalization of Median Filtering Using Linear Combination of Order Statistics*, IEEE Trans. on ASSP, vol.31, no.6, 1983.
- [2] H.A. David, *Order statistics*, New York, J. Wiley, 1981.
- [3] S. Zozor, *Statistiques d'ordre pour l'estimation de la moyenne*, rapport DEA cepha 19 bis/95, 1995.
- [4] G.E. Johnson, *Constructions of Particular Random Processes*, IEEE proc. of the IEEE, vol.82, no.2, 1994.
- [5] T.T. Pham and R.J.P. De Figueiredo, *Maximum Likelihood of a Class of Non-Gaussian Densities with Application to L_p Deconvolution*, IEEE Trans. on ASSP, vol.37, no.1, 1989.