REVISITING THE ESTIMATION OF THE MEAN USING ORDER STATISTICS

S. Zozor , E. Moisan and P.O. Amblard

CEPHAG-ENSIEG UPRESA CNRS 5086
BP 46, 38402 Saint-Martin d'Hères Cédex, FRANCE
Steeve.Zozor@cephag.inpg.fr

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 estimation has been done by Bovik for independent identically distributed variables. In this paper we extend this work to correlated variables. In particularly 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 an *a priori* knowledge. At last an adaptive algorithm is given and a practical application is presented.

1. INTRODUCTION

Signal processing often needs to estimate the expected value of a finite length sequence of random variables $\{X_i\}$. Generally, this estimation is performed with a sample mean. It is well known that such an estimator is the best one (in the sense of the mean squared error (MSE)) when the data are Gaussian and independent identically distributed (iid). But this property is no longer valid if the previous assumptions are not satisfied. As an example, for a uniform iid sample, the maximum likelihood estimator of the mean makes use of two order statistics [5]: $\widehat{m}_{mle} = \frac{\mathrm{Min}_i X_i + \mathrm{Max}_i X_i}{2}$.

Bovik developed a linear combination of order statistics (OS) to estimate expected values without an *a priori* knowledge, contrary to the maximum likelihood method [2]. He only worked with iid noise and showed that the best OS unbiased estimator (BOSUE) performed better than the sample mean (in the sense of the least MSE).

In this paper, Bovik's work is extended to the case of colored noise, by mixing OS and usual linear estimator, taking into account the noise correlation.

2. PRELIMINARY RESULTS

Given a sequence of n random variables $X_1, \ldots, X_n$, OS are defined by arranging these values $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$. $X_{(i)}$ is called the ith OS of the sample. This nonlinear process complicates considerably the analysis, but a lot of results can be found in David [1]. Let m be the expected value of the sample, it can be written $X_i = m + B_i$, where B_i is a zero-mean variable. Hence, it can also be written $X_{(i)} = m + B_{(i)}$. The following notations will be used

$\underline{x}$	$=$	$[X_1, \ldots, X_n]^t$	sample vector (resp. $\underline{b}$)
$\tilde{\underline{x}}$	$=$	$[X_{(1)}, \ldots, X_{(n)}]^t$	vector of the OS (resp. $\tilde{\underline{b}}$)
$\underline{\underline{R}}$	$=$	$E[\tilde{\underline{b}}\,\tilde{\underline{b}}^t]$	correlation matrix of $\tilde{\underline{b}}$
$\underline{\underline{\Gamma}}$	$=$	$E[\underline{b}\,\underline{b}^t]$	covariance matrix of $\underline{b}$
$\underline{1}$	$=$	$[1, \ldots, 1]^t$	constant vector
$\underline{a}$	$=$	$[a_1, \ldots, a_n]^t$	coefficients vector of the estimator

Assume now that the B_i are iid. Assuming that the probability density function f_b of the variables B_i is even, David [1], Bovik [2] and Pitas [3] have built the best OS unbiased estimator (BOSUE) of the constant m as follows

- the estimator $\widehat{m}$ is linear to the OS: $\widehat{m} = \underline{a}^t\,\tilde{\underline{x}}$
- the estimator is unbiased: $E[\underline{a}^t\,\tilde{\underline{x}}] = m$, i.e. $\underline{a}^t E[\tilde{\underline{b}}] + m\underline{a}^t\,\underline{1} = m$
- the estimator is of least MSE: $E[(\underline{a}^t\,\tilde{\underline{x}} - m)^2]$ is minimum

Under the assumption of symmetry of f_b, $E[\tilde{\underline{b}}]$, which is not null, verifies $E[B_{(n+1-i)}] = -E[B_{(i)}]$ [1, 4]. The estimator $\underline{a}$ is then considered to be symmetric, i.e. $a_{n+1-i} = a_i$, but this symmetry has to be verified *a posteriori*. As a consequence, the unbiasness condition holds

$$\underline{a}^t\,\underline{1} = 1 \tag{1}$$

Minimizing the MSE under condition (1), i.e. minimizing $E[(\underline{a}^t\,\tilde{\underline{b}})^2] + \lambda(\underline{a}^t\,\underline{1} - 1)$, and setting the Lagrange multiplier λ to verify the constraint, David [1] and Bovik [2] have evaluated the BOSUE as

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

David has shown that the entries of $\underline{\underline{R}}$ verify $R_{n+1-i,n+1-j} = R_{i,j}$. As a consequence, the symmetry of $\underline{a_{\text{bosue}}}$ can easily be verified. Notice that the mean squared error is given by $MSE_{\text{bosue}} = \frac{1}{\underline{1}^t\underline{\underline{R}}^{-1}\underline{1}}$. For a Gaussian sample, Bovik showed that $\underline{a_{\text{bosue}}}^t\,\underline{\tilde{x}}$ is the average of the variables X_i, while for a uniform sample the estimator is equal to $\frac{\text{Min}_i X_i + \text{Max}_i X_i}{2}$, which are the maximum likelihood estimator in these cases [5].

Similarly, given a finite number n of random values X_i, the best linear unbiased estimator (BLUE) is defined by

$$\underline{a_{\text{blue}}} = \frac{\underline{\underline{\Gamma}}^{-1}\underline{1}}{\underline{1}^t\underline{\underline{\Gamma}}^{-1}\underline{1}} \tag{3}$$

Then, the mean squared error is given by $MSE_{\text{blue}} = \frac{1}{\underline{1}^t\underline{\underline{\Gamma}}^{-1}\underline{1}}$. If the variables X_i are iid, $\underline{\underline{\Gamma}}$ is proportional to the identity matrix and then, whatever their probability density, the BLUE is the average of the sample, so that the minimum of the MSE is reached only if the X_i are Gaussian. Finally, when the data are independent, the BOSUE performs always at least as well as the BLUE because the BLUE is, in this case, a particular OS estimator. The question we want to answer in this paper is: what happens if data are not independent?

3. ESTIMATION OF THE MEAN OF CORRELATED VARIABLES

Figure 1 shows two examples for colored noise, Gaussian or uniform [6], to compare BOSUE and BLUE. Notice that the results have been estimated by averaging the matrix used (e.g. $\underline{\underline{R}}$) on a 9×10^4 points simulated signal (10^4 realizations of $\underline{b}$). It can be seen that BOSUE is no more better than BLUE when the data are correlated. So, we propose to combine both estimators, hoping that the linear part can learn the correlation of the data, while the nonlinear one learns the probability density!

We now consider vector $\underline{y} = [\underline{x}^t\ \ \underline{\tilde{x}}^t]^t$. Let $\underline{a_l}$ (respectively $\underline{a_{os}}$) be the linear part (respectively the OS part) of the estimator: $\underline{a} = [\underline{a_l}^t\ \ \underline{a_{os}}^t]^t$. The joint probability density function of the zero-mean sample is assumed to be even $f_{1,\dots,n}(b_1,\dots,b_n) = f_{1,\dots,n}(-b_1,\dots,-b_n)$. First we will insist that the estimator be unbiased, that is $E[\underline{a}^t\underline{y}] = m$. It leads to the condition $\underline{a_{os}}^t\,E[\underline{\tilde{b}}] + m\underline{a}^t\underline{1} = m$. It can be shown, under the assumption of symmetry of $f_{1,\dots,n}$ that $E[\underline{\tilde{b}}]$ is also antisymmetric ($E[B_{(n+1-i)}] = -E[B_{(i)}]$). Also considering that $\underline{a_{os}}$

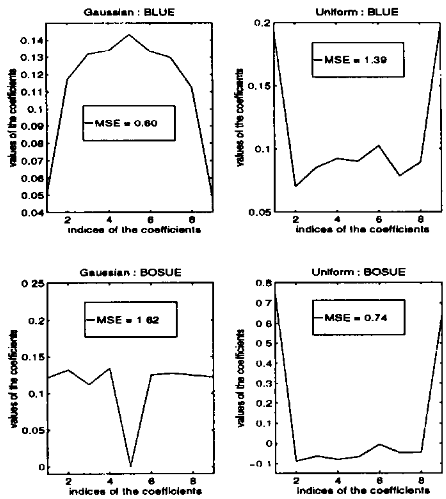

Figure 1: BOSUE and BLUE for colored Gaussian and uniform noises ($n = 9$).

is symmetric, the unbiasness constraint can always be written as $\underline{a}^t\,\underline{1} = 1$. Minimizing the MSE of the estimator under the previous constraint leads to

$$E\left[\begin{bmatrix}\underline{b}\\ \underline{\tilde{b}}\end{bmatrix}[\underline{b}^t\ \ \underline{\tilde{b}}^t]\right]\underline{a} = -\frac{\lambda}{2}\underline{1} \tag{4}$$

where λ is the Lagrange multiplier. Unfortunately, the autocorrelation matrix in the left hand side is singular. Indeed, $\sum_{i=1}^{n} B_i = \sum_{i=1}^{n} B_{(i)}$. As a consequence $\underline{k} = [\underbrace{\underline{1}^t}_{n}\ \ \underbrace{-\underline{1}^t}_{n}]^t$ is in the kernel of $E[[\underline{b}^t\ \ \underline{\tilde{b}}^t]^t.[\underline{b}^t\ \ \underline{\tilde{b}}^t]]$. It can be shown that its rank is $2n-1$, hence vector $\underline{1}$, which is orthogonal to $\underline{k}$, is in the range of this matrix, so that there are vectors $\underline{a}$ satisfying equation (4). The proposed solution is to remove the first component X_1 of $\underline{y}$ to obtain a unique vector, whose dimension is $2n-1$

$$\underline{a_{\text{blosue}}} = \frac{\underline{\underline{\phi}}^{-1}\underline{1}}{\underline{1}^t\underline{\underline{\phi}}^{-1}\underline{1}} \tag{5}$$

where

$$\underline{\underline{\phi}} = E[[B_2\dots B_n\,B_{(1)}\dots B_{(n)}]^t.[B_2\dots B_n\,B_{(1)}\dots B_{(n)}]]$$

and where the Lagrange multiplier has been adjusted to satisfy the constraint. As in the previous section, we verify *a posteriori* that $\underline{a_{os}}$ is symmetric. The associated mean squared error is then $MSE_{\text{blosue}} = \frac{1}{\underline{1}^t\underline{\underline{\phi}}^{-1}\underline{1}}$.

Notice that any X_i can be cancelled instead of X_1, but we have chosen to not suppress an OS to conserve the symmetry of this part.

Figure 2 shows the two same examples as the Figure 1 for colored noise, Gaussian or uniform, to compare BLOSUE , BOSUE and BLUE. For the colored

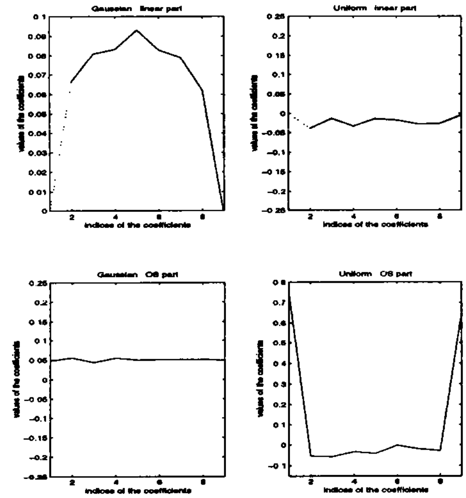

Figure 2: OS and linear parts of the BLOSUE for colored noises. Gaussian case, $MSE_{\text{blosue}} = 0.60 = MSE_{\text{blue}}$; uniform case, $MSE_{\text{blosue}} = 0.74 = MSE_{\text{bosue}}$. The dotted line recalls that the first variable has been cancelled.

Gaussian case, the BLOSUE chooses clearly the mean estimation through the variables, and for the colored uniform case, the estimation through the OS is chosen. Notice that the OS part of the BLOSUE for the colored Gaussian noise is not null but constant, due to the cancelation of X_1. Figure 3 depicts the BLUE, the BOSUE and the BLOSUE for colored Gaussian noise filtered nonlinearly as follows

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

As expected, BLOSUE is seen to be always the best one.

In all the previous simulations, matrices $\underline{\underline{R}}$, $\underline{\underline{\Gamma}}$ and $\underline{\underline{\phi}}$ have been estimated by averaging. Practically these matrices are not known and hence have been estimated by this method.

4. APPLICATION TO A PRACTICAL PROBLEM

In most cases, matrices $\underline{\underline{R}}$, $\underline{\underline{\Gamma}}$ or $\underline{\underline{\phi}}$ of the data are not known and must be approximated, using an adaptive scheme, like a RLS method, for example [3]. Consider $\{x_k\}$ to be a signal. Let $\underline{\underline{M}}$ denotes $\underline{\underline{R}}^{-1}$ (resp. $\underline{\underline{\Gamma}}^{-1}$,

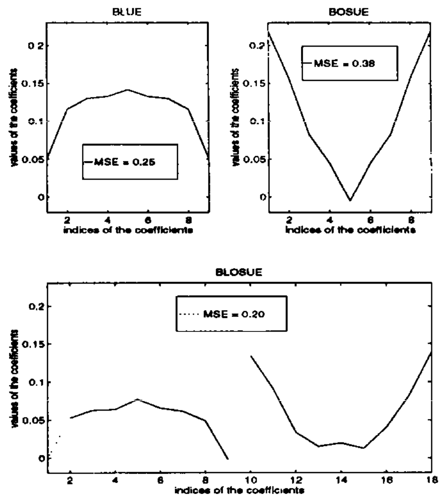

Figure 3: BLUE, BOSUE and BLOSUE for the non-linear filtered colored Gaussian noise. $MSE_{\text{blosue}} < MSE_{\text{blue}}$ and $MSE_{\text{blosue}} < MSE_{\text{bosue}}$.

resp. $\underline{\underline{\phi}}^{-1}$), and let $\underline{v}_k$ denotes $\underline{x}_k = [x_k \ldots x_{k+1-n}]^t$ the n last data (resp. $\tilde{\underline{x}}_k$, resp. $[x_{k-1} \ldots x_{k+1-n} \;\; \tilde{\underline{x}}_k]$). Let $\widehat{m}_k$ be an estimation of m at step k. $\underline{\underline{M}}$ is then estimated using an average scheme (through the woodbury decomposition), which leads to the following recursive algorithm

- Initialization step: $\widehat{m}_1$ and $\underline{\underline{M}}_1$
- From step k to step $k+1$:

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

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

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

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

Now let suppose that we have to estimate a time-varying mean, constant by step, corrupted by addition of zero-mean noise. As an example, we consider the case of a binary signal, which is constant only for finite durations. The three estimators BLUE, BOSUE and BLOSUE have been applied to such signals. Results are depicted on Figure 4 for iid uniform noise, and 6 for colored Gaussian noise. Finally, Figures 5 and 7 show the coefficients of those estimators and the evolutions of two of them. In these simulations we have "guessed" the transitive times of the algorithm seeing the evolutions of the two shown coefficients, hence only

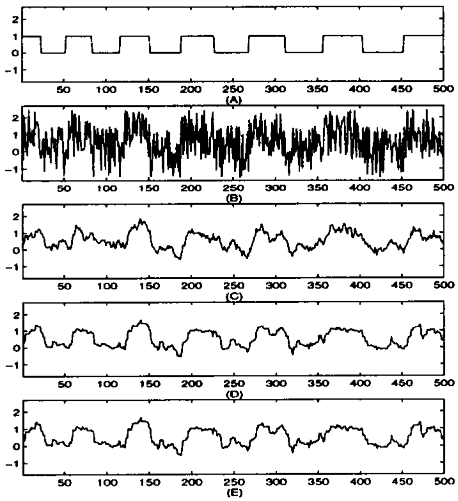

Figure 4: (A), binary signal; The signal is then corrupted by iid uniform noise (B) ; (C), estimation with the BLUE; (D) estimation with the BOSUE; (E) estimation with the BLOSUE.

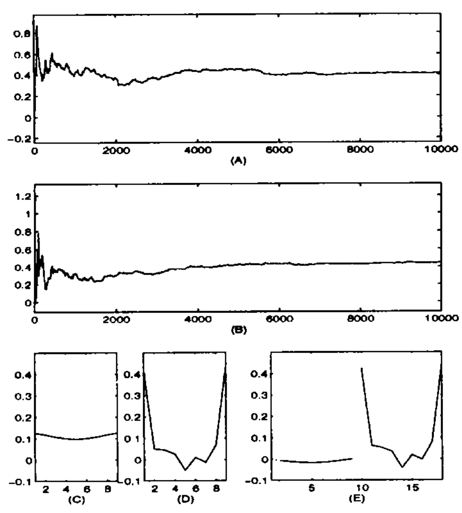

Figure 5: Uniform iid case: (A) and (B) represent the evolutions of a_{bosue_1} and a_{bosue_n}; (C) and (D) depict the BLUE and the BOSUE at the last step; (E) depicts the BLOSUE at the last step (linear part (left) and OS part (right) separated; a dotted line recalls that the first variable has been cancelled).

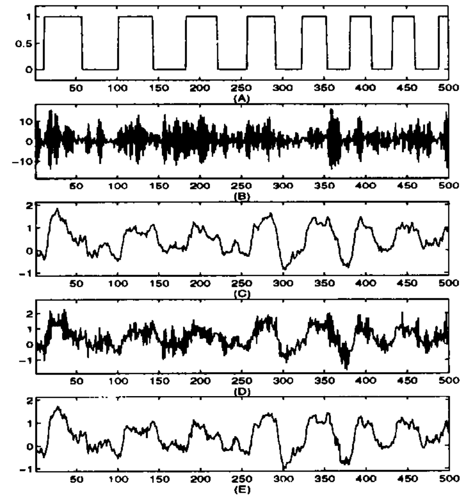

Figure 6: (A), binary signal; The signal is then corrupted by colored Gaussian noise (B) ; (C), estimation with the BLUE; (D) estimation with the BOSUE; (E) estimation with the BLOSUE.

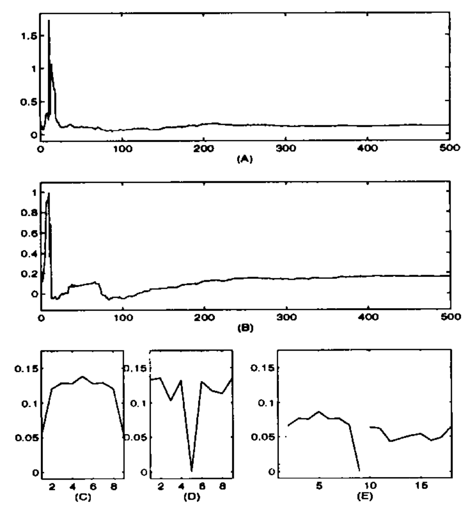

Figure 7: Colored Gaussian case: (A) and (B) represent the evolutions of a_{bosue_1} and a_{bosue_n}; (C) and (D) depict the BLUE and the BOSUE at the last step; (E) depicts the BLOSUE at the last step (linear part (left) and OS part (right) separated; a dotted line recalls that the first variable has been cancelled).

the steady state of our algorithm is presented. It can be seen that when the noise is iid uniform, BLOSUE clearly chooses the OS estimator, which is the best one for iid noise. In return, when the noise is colored Gaussian noise, BLOSUE chooses the non-ordered variables: in this case, the BLUE is the best estimator.

5. CONCLUSION

BLUE and BOSUE can easily be mixed to produce BLOSUE. This one is always the best, because it is able to learn the correlation (through its linear part) and the probability density (through its nonlinear part) of the data. A simple solution has been established to remove the singularity of the problem. An adaptive algorithm is then proposed which is, in fact, able to learn the characteristics of the data. Finally, the method is seen to be attractive on an usual example.

6. REFERENCES

[1] H.A. David, *Order Statistics*, New York, J. Wiley & Sens, 1981.

[2] A.C. Bovik, T.S. Huang and D.C. Munson, *A Generalization of Median Filtering Using Combination of Order Statistics*, IEEE transaction on Acoustics, Speech, and Signal Processing, vol. ASSP 31, no. 6, pp. 1342-1350, December 1983.

[3] I. Pitas and A.N. Venetsanopoulos, *Adaptive Filters based on Order Statistics*, IEEE transactions on Signal Processing, vol. 39, no. 2, pp. 518-522, February 1991.

[4] S. Zozor, *Statistiques d'ordre pour l'estimation de la moyenne* , rapport DEA cephag 19 bis/95, Septembre 1995.

[5] T.T. Pham and R.J.P. De Figueiredo, *Maximum Likelihood Estimation of a Class of Non-Gaussian Densities with Application to Lp Deconvolution*, IEEE transaction on Acoustics, Speech, and Signal Processing, vol. 37, no. 1, pp. 73-82, January 1989.

[6] G.E. Johnson, *Constructions of Particular Random Processes*, IEEE proceedings of the IEEE, vol. 82, no. 2, pp. 270-285, february 1994.