

Comparaison de deux estimateurs de la variance relative à des fluctuations d'échelle de transitoires, en présence de bruit coloré

O. Meste, E. Bataillou, H. Rix

Laboratoire I3S-URA 1376 CNRS, 250 Av. Albert Einstein, Sophia Antipolis, 06560 Valbonne,
FRANCE

RÉSUMÉ

ABSTRACT

Lors de l'acquisition de certains signaux physiologiques répétitifs, le signal se répète d'une réalisation à l'autre avec une légère fluctuation aléatoire d'échelle. La présence de bruit peut alors empêcher l'estimation de chaque changement d'échelle. On montre dans cet article que l'on peut estimer directement la variance de ces fluctuations à partir de l'ensemble des réalisations en utilisant les résultats déjà obtenus lorsque le changement d'échelle est remplacé par un retard aléatoire. Les conditions d'acquisition font apparaître un bruit coloré ce qui nous conduit à utiliser un estimateur linéaire à biais nul et à variance minimale. Nous proposons alors deux estimateurs non biaisés de la matrice de covariance du bruit (qui dans notre approche devient quadratique) dont nous calculons les variances respectives.

I. INTRODUCTION

Les signaux dits répétitifs le sont, soit de façon spontanée comme les électrocardiogrammes, soit en réponse à une stimulation périodique comme les potentiels évoqués cérébraux ou musculaires. Nous nous affranchirons de l'incertitude due au processus d'alignement des signaux en considérant uniquement le deuxième cas. Cependant, en présence de signaux physiologiques, la répétitivité n'est jamais parfaite et il peut apparaître de légères fluctuations soit d'échelle soit de position. L'estimation de ces fluctuations aléatoires peut permettre de caractériser plus finement le signal ou bien de les éliminer en vue d'utiliser la technique du moyennage synchrone dans le cas bruité. Nous nous intéresserons non pas à l'estimation de cette fluctuation pour chaque répétition, ce qui dans certains cas est rendu difficile à cause du bruit, mais plutôt à l'estimation de sa variance ce qui dans le cas gaussien centré caractérise sa statistique. Cette quantité peut être évaluée sans passer par l'estimation de chaque fluctuation. On montre que la transposition du problème des fluctuations de position au problème des fluctuations d'échelle se fait simplement, ce qui permet un formalisme commun. L'approche proposée fait appel à des fonctions quadratiques qui ont pour effet de modifier non linéairement le bruit d'observation qui sera considéré coloré. La solution du problème passera par

When studying some particular physiological signals, like stimulated and repetitive signals, the presence of a random scaling factor from a realization to another is observed. The physiological noise may make it impossible to properly estimate each individual factor. In this paper we show that the estimation of the scaling factors variance can be made directly, using the result obtained when we are interested in time delay estimation instead of scaling factor. The relevant signal is embedded in a colored noise leading to apply the Best Linear Unbiased Estimator (BLUE) formulation. In our approach the initial noise becomes quadratic so we propose two unbiased estimators of the noise correlation matrix. To compare these estimators, we calculate their variance.

l'utilisation d'un estimateur sans biais à variance minimale (BLUE en anglais) ce qui nécessitera la connaissance de la matrice de covariance de ce bruit.

II. APPROCHE CLASSIQUE

Le traitement classique des signaux répétitifs bruités est le moyennage synchrone. Outre les hypothèses sur le bruit, une contrainte importante qu'il faut imposer au signal est la répétitivité d'une réalisation à l'autre. L'écart à cette hypothèse couramment rencontré est soit un décalage aléatoire, soit un changement d'échelle aléatoire. Si le premier phénomène est rencontré, la moyenne des N réalisations est [4] (en négligeant le bruit résiduel et pour N suffisamment grand) :

$$\bar{r}_d(t) = (s * p_d)(t) \quad (1)$$

c'est à dire le signal $s(t)$ convolué par la densité de probabilité des décalages.

Lorsque ce décalage est remplacé par un facteur d'échelle a , transformant $s(t)$ en $s(at)$, cette moyenne devient :

$$\bar{r}_a(t) = \int_0^{+\infty} \frac{1}{t} s(\xi) p_a\left(\frac{\xi}{t}\right) d\xi \quad t > 0 \quad (2)$$

où p_a est la densité de probabilité (d.p) de a .

On retrouve alors l'équation du filtrage linéaire invariant par



changement de forme [1] (mais non invariant par translation).

Un des moyens de limiter ces effets de filtrage (de type basse-fréquence) est d'estimer soit les décalages, soit les facteurs d'échelle, afin de reconstituer le signal répétitif. Le niveau de bruit et ses caractéristiques spectrales peuvent induire non pas un rehaussement du signal moyen mais au contraire une dégradation. Si l'opération de déconvolution est maîtrisée, on peut alors se pencher sur l'estimation de la densité de probabilité de ces variables aléatoires, ce qui dans certains cas peut se réduire au calcul de la variance. Cette dernière quantité sera l'objet de cette communication.

III. FORMULATION DU PROBLÈME

Lorsque chaque réalisation $x_i(t)$ est constituée par la somme d'un signal de référence $s(t)$ retardé aléatoirement d'une quantité d_i supposée indépendante des autres, l'ensemble des réalisations s'écrit :

$$x_i(t) = s(t - d_i) + n_i(t) \quad (i = 1 \dots N; 0 \leq t \leq T) \quad (3)$$

Les séquences de bruit $n_i(t)$ sont supposées indépendantes dans leur ensemble mais chaque séquence est un processus gaussien stationnaire, centré, coloré. On montre dans [3] que sous certaines hypothèses sur le contenu fréquentiel de $s(t)$, la variance des x_i en fonction du temps peut s'exprimer par :

$$\sigma_x^2(t) \approx s^{(1)2}(t)\sigma_d^2 + \hat{\sigma}_n^2(t) \quad (4)$$

avec $s^{(1)2}(t)$ la dérivée de $s(t)$ au carré, σ_d^2 la variance à estimer des retards d_i et $\hat{\sigma}_n^2(t)$ la variance des séquences de bruit. Cette formule pouvant être approchée à des ordres supérieurs.

On voit alors apparaître une fonction linéaire de la variance des retards pour chaque temps. La variance du bruit peut être remplacée par une constante plus une erreur, fonction du temps, qui sera l'erreur d'estimation de la variance pour chaque temps. Cette erreur n'est pas blanche si les séquences de bruit ne le sont pas, d'où la nécessité d'utiliser BLUE. Le calcul de la matrice de covariance sera traité par la suite.

Si la présence d'un retard aléatoire est remplacée par celle d'une fluctuation aléatoire d'échelle, le modèle (3) devient :

$$x_i(t) = s(a_i t) + n_i(t) \quad (i = 1 \dots N; 0 \leq t \leq T) \quad (5)$$

Les variables aléatoires a_i seront supposées indépendantes et identiquement distribuées, de moyennes égales à 1. L'équation (5) peut se réécrire :

$$x_i(t) = s(t - (1 - a_i)t) + n_i(t) \quad (i = 1 \dots N; 0 \leq t \leq T) \quad (6)$$

La comparaison de (6) et (3) nous conduit à considérer le terme $(1 - a_i)t$ comme un retard aléatoire fonction du temps dont on peut calculer la variance également fonction du temps :

$$\sigma_d^2(t) = \sigma_a^2 t^2 \quad (7)$$

Il suffit de remplacer la variance des retards de l'équation (4) par l'équation (7) pour obtenir la variance des x_i lorsque le modèle (5) est considéré :

$$\sigma_x^2(t) \approx s^{(1)2}(t)\sigma_a^2 t^2 + \hat{\sigma}_n^2(t) \quad (8)$$

Ces différentes équations étant obtenues dans le cas continu, nous allons les exprimer dans le cadre discret.

IV. DISCRETISATION

L'opération la plus contraignante est la dérivation du signal de référence discrétisé $s(kT)$, où T représente la période d'échantillonnage. Nous simplifierons les équations en omettant le terme T . Cette dérivation sera approximée par un filtre RIF dont les coefficients sont : $-0.5, 0, 0.5$. L'intérêt d'un tel filtre est de déphaser le signal d'un nombre entier d'échantillons.

Le signal de référence échantillonné, dérivé et élevé au carré puis multiplié par les instants d'échantillonnage au carré, c'est à dire $s^{(1)2}(t)t^2$ discrétisé, sera noté \mathbf{h} . L'équation continue (8) s'exprimera sous forme échantillonnée :

$$\begin{aligned} \sigma_x^2 &\approx (\mathbf{h} | \mathbf{I}) \begin{pmatrix} \sigma_a^2 \\ \sigma_n^2 \end{pmatrix} + \mathbf{b} \\ &\approx \mathbf{A} \sigma^2 + \mathbf{b} \end{aligned} \quad (9)$$

avec \mathbf{I} un vecteur unité et $\hat{\sigma}_n^2(k) = \sigma_n^2 + b(k)$. Le bruit \mathbf{b} correspond à l'erreur d'estimation de la variance des bruits n_i . Nous verrons que ce bruit a une longueur de corrélation inférieure à celle des n_i , que l'on pourra déterminer en estimant la matrice de covariance.

La solution de l'équation (9) est alors donnée par la formule des moindres carrés pondérés correspondant à l'estimateur linéaire à biais nul et à variance minimale :

$$\hat{\sigma}^2 = (\mathbf{A}^T \mathbf{R}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{R}^{-1} \sigma_x^2 \quad (10)$$

où \mathbf{R} est la matrice de covariance du bruit, c'est à dire $\mathbf{R} = E[\mathbf{b}\mathbf{b}^T]$.

V. DÉTERMINATION DE LA MATRICE DE COVARIANCE

La détermination de cette matrice est nécessaire si l'on veut prendre en considération la coloration du bruit. Nous verrons que la connaissance de la fonction d'autocorrélation du bruit initial ne suffit pas pour déterminer simplement cette matrice. En effet, l'approche proposée pour le calcul de la variance des fluctuations d'échelle ou de position implique de modifier non linéairement le bruit initial ce qui va rendre délicat le calcul de la fonction d'autocorrélation du bruit ainsi transformé.

Le calcul de la matrice de covariance (ou autocovariance) \mathbf{R} passe par le calcul de la fonction de covariance $C(k, \tau)$ définie par :

$$C(k, \tau) = E [(\hat{\sigma}_n^2(k) - E(\hat{\sigma}_n^2(k)))(\hat{\sigma}_n^2(k + \tau) - E(\hat{\sigma}_n^2(k + \tau)))] \quad (11)$$

ce qui correspond également à la fonction de covariance de $b(k)$. Le bruit étant supposé stationnaire, on pourra remplacer $C(k, \tau)$ par $C(\tau)$.

Il reste à déterminer la quantité $\hat{\sigma}_n^2(k)$, de moyenne σ_n^2 , correspondant au calcul de la variance des séquences de bruit $n_i(k)$:

$$\hat{\sigma}_n^2(k) = \frac{1}{N} \sum_{i=1}^N n_i^2(k) - \left[\frac{1}{N} \sum_{j=1}^N n_j(k) \right]^2 \quad (12)$$

On peut montrer que lorsque l'estimateur (12) est employé, $C(\tau)$ s'écrit :

$$C(\tau) = 2 \frac{N-1}{N^3} E^2 [n(0)n(\tau)] - \frac{(N-1)^2}{N^3} \sigma_n^4 + \frac{(N-1)^2}{N^3} E [n^2(0)n^2(\tau)] \quad (13)$$

Si le bruit $n(k)$ est gaussien et centré, on a la relation [2] :

$$E [n^2(0)n^2(\tau)] = 2E^2 [n(0)n(\tau)] + \sigma_n^4 \quad (14)$$

Ce qui permet d'exprimer (13) de deux façons différentes :

$$C(\tau) = 2 \frac{N-1}{N^2} E^2 [n(0)n(\tau)] \quad (15)$$

ou bien

$$C(\tau) = \frac{N-1}{N^2} (E [n^2(0)n^2(\tau)] - \sigma_n^4) \quad (16)$$

Ces deux équations sont équivalentes lorsque l'espérance mathématique est employée. Nous proposons deux estimateurs non biaisés des expressions (15) et (16) divisées par $\frac{N-1}{N^2}$.

On montre facilement la relation suivante:

$$E \left[\left(\frac{1}{N} \sum_{i=1}^N n_i(0)n_i(\tau) \right)^2 \right] = \frac{N+1}{N} E^2 [n(0)n(\tau)] + \frac{1}{N} \sigma_n^4 \quad (17)$$

On peut en déduire un premier estimateur non biaisé $e_1(\tau, N)$, provenant de (15) :

$$e_1(\tau, N) = 2 \frac{N}{N+1} \left(\frac{1}{N} \sum_{i=1}^N n_i(0)n_i(\tau) \right)^2 - \frac{2}{N+1} \sigma_n^4 \quad (18)$$

De même, on peut proposer un estimateur non biaisé $e_2(\tau, N)$ issu de (16) :

$$e_2(\tau, N) = \frac{1}{N} \sum_{i=1}^N n_i^2(0)n_i^2(\tau) - \sigma_n^4 \quad (19)$$

Nous obtenons donc deux estimateurs non biaisés, dont les performances sont évaluées en calculant leur variance respectives $v_1(\tau, N)$ et $v_2(\tau, N)$:

$$v_1(\tau, N) = E \left[\left\{ \frac{2N}{N+1} \left[\frac{1}{N} \sum_{i=1}^N n_i(0)n_i(\tau) \right]^2 \right\}^2 \right] - E^2 \left[\frac{2N}{N+1} \left[\frac{1}{N} \sum_{i=1}^N n_i(0)n_i(\tau) \right]^2 \right] \quad (20)$$

et

$$v_2(\tau, N) = E \left[\left\{ \frac{1}{N} \sum_{i=1}^N n_i^2(0)n_i^2(\tau) \right\}^2 \right] - E^2 \left[\frac{1}{N} \sum_{i=1}^N n_i^2(0)n_i^2(\tau) \right] \quad (21)$$

Ces calculs font appel aux résultats issus de [2] permettant de calculer les expressions $E[n^p(k)]$ et $E[n^p(k)n^p(k+\tau)]$. On obtient alors les expressions des deux variances en fonctions du nombre de réalisations N et de la fonction d'intercorrélation du bruit initial $n(k)$.

$$v_1(\tau, N) = \frac{8(2N+3)}{(N+1)N} E^4 [n(0)n(\tau)] + \frac{16(N^2+7N+9)}{(N+1)^2 N} \sigma_n^4 E^2 [n(0)n(\tau)] + \frac{8(N+3)}{(N+1)^2 N} \sigma_n^8 \quad (22)$$

et

$$v_2(\tau, N) = \frac{20}{N} E^4 [n(0)n(\tau)] + \frac{68}{N} \sigma_n^4 E^2 [n(0)n(\tau)] + \frac{8}{N} \sigma_n^8 \quad (23)$$

Ce qui nous permet d'établir la différence de ces deux variances $V(\tau, N) = v_2(\tau, N) - v_1(\tau, N)$:

$$V(\tau, N) = \frac{4(N-1)}{(N+1)N} E^4 [n(0)n(\tau)] + \sigma_n^4 \frac{4(13N+19)(N-1)}{(N+1)^2 N} E^2 [n(0)n(\tau)] + \frac{8(N+2)(N-1)}{(N+1)^2 N} \sigma_n^8 \quad (24)$$

On peut vérifier que $V(\tau, 1) = 0$, ce qui correspond au résultat $e_1(\tau, 1) = e_2(\tau, 1)$. On retrouve également $V(\tau, \infty) = 0$ car les estimateurs tendent vers l'espérance mathématique.

Le constat important est que les polynômes en N de l'équation (24) sont tous strictement positifs pour $N \in]1, \infty[$. Ceci nous permet de conclure que l'estimateur $e_1(\tau, N)$ présente une variance inférieure ce qui justifiera son choix.

Pour illustrer ces résultats, nous avons généré 500 ensembles de 500 réalisations de bruits colorés indépendants d'une réalisation à l'autre dont la longueur de corrélation est de 30 échantillons. La fonction d'autocorrélation du bruit étant parfaitement connue, nous comparons le résultat théorique (24) avec la moyenne de 500 estimations de $V(\tau, N)$ à partir de $e_1(\tau, N)$ et $e_2(\tau, N)$. Nous constatons dans la figure 1 l'adéquation de la pratique avec les résultats théoriques.

Le passage de la fonction $e_1(\tau, N)$, estimateur de l'équation (15), à une estimation de la matrice de covariance $\mathbf{R} = E[\mathbf{b}\mathbf{b}^T]$ se fera en créant une matrice de Toeplitz à partir du vecteur constitué des valeurs de $e_1(\tau, N)$ pour $\tau \in [0, L-1]$. La matrice ainsi constituée sera de dimension $L \times L$.

VI. SIMULATION

Le but de cette simulation est de vérifier que dans les conditions expérimentales proposées, le modèle introduit au

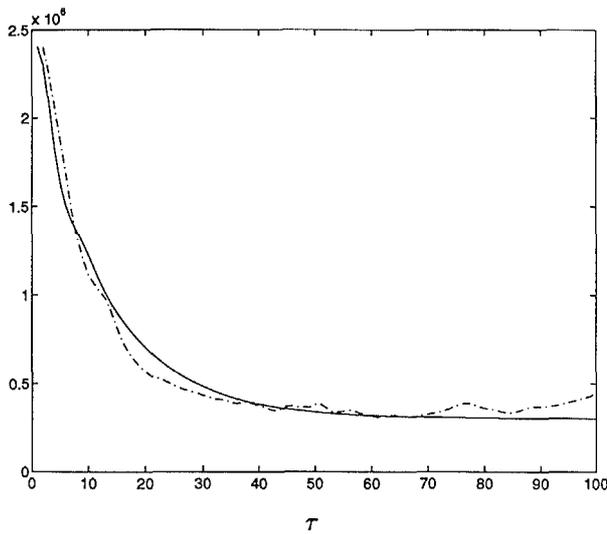


Fig. 1. $V(\tau)$ pour $n_i(t)$ supposé stationnaire et avec $N = 500$. Courbe théorique : —, courbe estimée à partir de 500 ensembles de 500 réalisations de bruits $n_i(t)$: - - - -.

début de cette communication permet de retrouver la variance des fluctuations d'échelle. Nous utilisons dans l'algorithme d'estimation, l'estimation de la fonction d'autocorrélation (20) qui théoriquement présente la variance la plus faible. Cette simulation a également pour intérêt de vérifier que l'ajout d'une matrice de pondération estimée \mathbf{R}^{-1} n'induit pas d'incertitude supplémentaire quant à la détermination de σ . Cette incertitude serait alors due à la variance présente dans l'estimation de \mathbf{R}^{-1} . L'acquisition de bruit seul permettra d'estimer l'écart-type du bruit σ_n et par conséquent de calculer (20). La quantité σ_n nous permettra également de simplifier (11) en soustrayant directement à σ_x^2 la valeur σ_n^2 .

Nous simulons 500 réalisations d'un même signal soumis à un facteur d'échelle aléatoire gaussien, de moyenne 1 et de variance $\sigma_a^2 = 0.39 \cdot 10^{-3}$. La figure (2) représente les 500 réalisations non bruitées. Le bruit additif coloré a une longueur de corrélation moyenne de 30 échantillons. Le RSB des signaux ainsi créés est de 0dB. La variance de $x_i(t)$, c'est à dire σ_x^2 , est représentée dans la figure (3), de même que la variance estimée après calcul de σ_a^2 à l'aide du BLUE.

Nous avons généré 100 ensembles de 500 réalisations afin de caractériser la diminution de la variance de l'erreur d'estimation de σ_a^2 lorsque le BLUE est utilisé. La variance σ_a^2 étant de $0.39 \cdot 10^{-3}$, on obtient pour écart-type empirique de l'erreur d'estimation la valeur $7 \cdot 10^{-5}$ avec BLUE et $8 \cdot 10^{-5}$ sans BLUE. On a donc une amélioration d'environ 12% sur l'incertitude relative.

VII. CONCLUSION

Nous avons montré que l'on pouvait estimer directement la variance de la fluctuation d'échelle à laquelle est soumis un signal répétitif. Pour cela, nous avons transposé des résultats obtenus précédemment, concernant les décalages aléatoires.

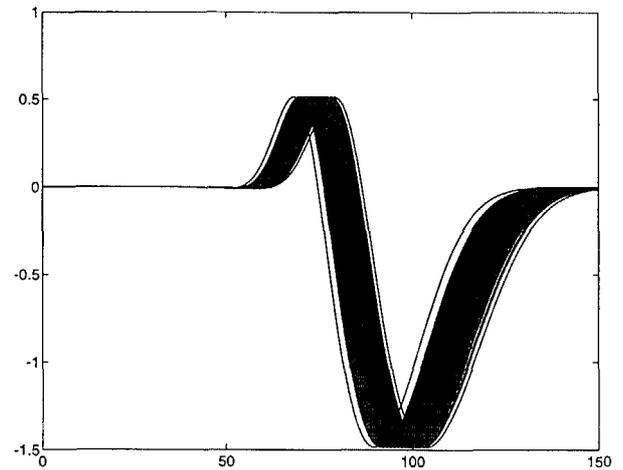


Fig. 2. 500 réalisations non bruitées avec $\sigma_a^2 = 0.39 \cdot 10^{-3}$

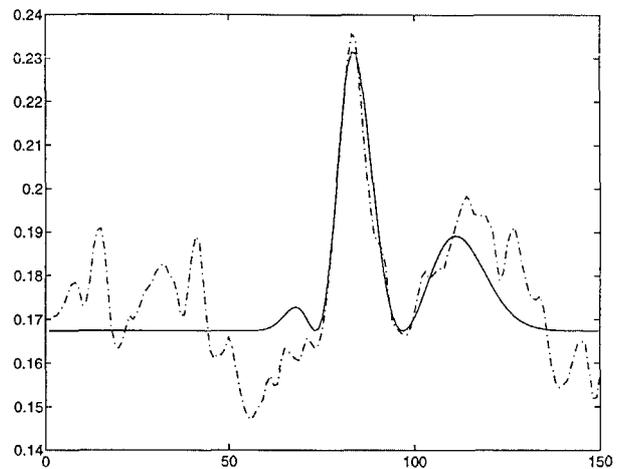


Fig. 3. Variance des $x_i(k)$ $\sigma_x^2(k)$. Mesuré avec la forme (12): - - - -, estimée après résolution de (10): —

Ces signaux étant acquis dans un contexte bruité, nous avons montré comment choisir théoriquement entre deux estimateurs asymptotiquement identiques. Les résultats de simulations d'une part confirme la qualité de notre estimateur des fluctuations d'échelle, d'autre part permettent de mesurer l'amélioration apportée par BLUE sur la précision de l'estimation.

Nous disposons donc d'un outil adapté à l'étude de la variabilité des signaux biomédicaux répétitifs.

VIII. REFERENCES

- [1] C. Braccini, G. Gambardella, "Form-invariant linear filtering: theory and applications", *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-34, pp. 1612-1627, Dec. 1986
- [2] S. Kikkawa, M. Ishida, "Number of degree of freedom, correlation times, and equivalent bandwidths of a random process", *IEEE Trans. Information Theory*, Vol. 34, No. 1, pp. 151-155, 1988.
- [3] O. Meste, H. Rix, "Jitter statistics estimation in alignment processes", à paraître dans *Signal Processing*.
- [4] O. Rompelman, H. H. Ros, "Coherent averaging technique: a tutorial review; Part 2: trigger jitter, overlapping responses and non-periodic stimulation", *Journal of Biomed. Eng.*, No. 8, pp. 30-35, 1986