

Estimation locale : compromis regression-variance

Olivier BARRIERE¹, Jacques LEVY-VEHEL²

¹IRCCyN, 1 rue de la Noe BP 92101, 44321 Nantes Cedex 03, France

²INRIA Rocquencourt, Domaine de Voluceau BP 105, 78153 Le Chesnay Cedex, France
 olivier.barriere@irccyn.ec-nantes.fr, jacques.levy-vehel@inria.fr

Resume – Nous proposons une methode permettant sous certaines hypotheses de diminuer la variance d’un estimateur optimal sans degrader son risque, en utilisant conjointement plusieurs estimateurs grâce a une technique tres simple pour arriver a un estimateur hybride pro tant du meilleur des deux. Cette methode peut s’appliquer dans de nombreux cas, typiquement a des estimateurs obtenus par regression lineaire.

Abstract – We present a method allowing under some hypothesis to decrease the variance of an optimal estimator without causing damage to its risk, using many estimators jointly, thanks to a very basic technique leading to a hybrid estimator that takes advantage of the best of them. This method can be applied in various fields, typically on estimators that were obtained using linear regressions.

1 Introduction

Soit $X(t)$ un processus qui depend d’un parametre fonctionnel (t) . L’estimation de (t) est un probleme usuel qui a de nombreuses applications en traitement du signal et des images. Classiquement, le probleme a resoudre s’exprime comme la recherche d’un estimateur minimisant le risque, c’est a dire realisant le meilleur « compromis biais-variance » : on parvient alors a la solution optimale en resolvant l’equation $biais^2 \approx variance$ [1] [2]. Cependant, en ce qui concerne l’aspect visuel, la variance apparaî t plus perturbante que le biais parce qu’elle introduit des « fausses irregularites » que l’on peut confondre avec les « vraies irregularites » propres a [3]. A risque constant, un estimateur ayant une variance plus faible est donc souvent preferable.

Nous proposons une methode generale permettant de reduire la variance d’un estimateur $\hat{\gamma}_2$ sans augmenter son risque, a condition que l’on dispose d’un autre estimateur $\hat{\gamma}_1$ et que certaines hypotheses reliant les biais et variances de ces deux estimateurs soient verifiees. Cette technique tres simple ne necessite presque aucun calcul supplementaire une fois qu’ $\hat{\gamma}_1$ et $\hat{\gamma}_2$ ont ete obtenus.

Une situation typique ou notre methode s’applique est la suivante : supposons que l’on connaisse un estimateur $\hat{\gamma}_1$ qui est performant dans le cas ou c’est bien le processus X qui est observe, mais fortement biaise quand c’est $G \cdot X$ qui est observe, ou G est une constante. Pour eliminer le biais dû a G , on definit generalement un deuxieme estimateur $\hat{\gamma}_2$ calcule via une regression lineaire sur plusieurs echelles [4]. Cette technique elimine bien le biais deterministe dû a G , mais souvent au prix d’une forte augmentation de la variance. Nous disposons alors de deux estimateurs dont le premier presente un fort biais mais une petite variance, et le second possede une grande variance.

La technique que nous proposons permet ici de garder le « meilleur des deux mondes », c’est a dire de diminuer la variance de $\hat{\gamma}_2$ tout en gardant son aspect optimal.

Nous presentons la methode dans un cadre general dans la section 2, et nous l’appliquons pour l’estimation de la regularite locale d’un processus fractal dans la section 3. Nous pensons que cette approche pourrait être mise a profit dans divers cadres autres que celui de la regression.

2 Principe

2.1 Hypotheses

On cherche ici a estimer une fonction (t) , parametre d’un processus $X(t) = X_{(t)}(t)$ indexe par $t \in I$ un intervalle de (t) , a partir d’un echantillonnage sur N points. On suppose que l’on dispose de deux estimateurs de cette fonction, $\hat{\gamma}_1$ et $\hat{\gamma}_2$, eux-mêmes parametres par des reels γ_1 et γ_2 . Notons alors :

$$\begin{aligned} B_i^{(N)}(\gamma_i, t) &= \mathbb{E}(\hat{\gamma}_i(\gamma_i, t) - (t)) \\ V_i^{(N)}(\gamma_i, t) &= \mathbb{E}(\hat{\gamma}_i(\gamma_i, t) - \mathbb{E}(\hat{\gamma}_i(\gamma_i, t)))^2 \\ R_i^{(N)}(\gamma_i, t) &= B_i^{(N)}(\gamma_i, t)^2 + V_i^{(N)}(\gamma_i, t) \\ \text{et } \left\| B_i^{(N)}(\gamma_i) \right\|_{\infty} &= \sup_{t \in I} B_i^{(N)}(\gamma_i, t), \text{ etc. } i = 1, 2 \end{aligned}$$

Posons les hypotheses suivantes : Il existe des parametres γ_1 et γ_2 tels que :

$$\begin{aligned}
 (H_1) \quad & \exists R^{(N)} \xrightarrow{N \rightarrow \infty} 0 \text{ tel que } \|R_i^{(N)}(i)\|_\infty \leq R^{(N)} \\
 (H_2) \quad & \forall t \ B_1^{(N)}(1, t) = B_{1C}^{(N)} + B_{1V}^{(N)}(1, t) \\
 (H_3) \quad & \|B_{1V}^{(N)}(1)\|_\infty = B_{1C}^{(N)} \\
 (H_4) \quad & \|B_2^{(N)}(2)\|_\infty = B_{1C}^{(N)} \\
 (H_5) \quad & \|V_1^{(N)}(1)\|_\infty = \|V_2^{(N)}(2)\|_\infty
 \end{aligned}$$

Nous disposons alors de deux estimateurs consistants (H_1) : le premier presente un biais en deux parties, dont une qui est independante de t et de α_1 (H_2) et preponderante (H_3). Le second possede un biais pas necessairement independant de t , mais dont la norme infinie est bien plus faible (H_4) alors que sa variance est beaucoup plus forte (H_5). Un cas typique ou ces conditions sont verifiees est celui ou l'on observe non pas le processus $X(t)$ mais $G \cdot X(t)$, ou G est une constante deterministe, mais inconnue. Le biais du premier estimateur est du a ce facteur d'echelle et il est necessaire d'effectuer une regression lineaire pour s'en aranchir et ainsi obtenir le second estimateur au prix d'une augmentation de la variance.

Supposons de plus $t \in [0, 1]$ et que pour des valeurs suffisamment eloignees en temps, les valeurs des estimateurs soient independantes (cette hypothese peut etre affaiblie et remplacee par une hypothese sur la vitesse de decroissance des correlations) :

$$(H_6) \quad |t - t'| \geq N^{\delta_i - 1} \Rightarrow \hat{\alpha}_i(i, t) \text{ et } \hat{\alpha}_i(i, t') \text{ independants}$$

2.2 Resultat principal

Theoreme 2.1. *oit l'estimateur $\hat{\alpha}$ de α pour tout t par :*

$$\hat{\alpha}(t) = \hat{\alpha}_1(t) - \langle \hat{\alpha}_1 \rangle + \langle \hat{\alpha}_2 \rangle$$

où $\langle . \rangle$ designe la moyenne temporelle :

$$\langle \hat{\alpha}_i \rangle = \frac{1}{N} \sum_{k=1}^N \hat{\alpha}_i(t_k)$$

oient B et V respectivement le biais et la variance de $\hat{\alpha}$. lors, sous les hypotheses H_1 à H_6 , on a :

$$\begin{aligned}
 |B^{(N)}(1, 2, t)| & \leq 2 \|B_{1V}^{(N)}(1)\|_\infty + \|B_2^{(N)}(2)\|_\infty \\
 & \stackrel{H_3, H_4}{=} \|B_1^{(N)}(1)\|_\infty \\
 V^{(N)}(1, 2, t) & \leq V_1(1, t) + \mathcal{O}(N^{\delta_2 - 1}) \|V_2^{(N)}(2)\|_\infty \\
 & \stackrel{H_5}{=} \|V_2^{(N)}(2)\|_\infty
 \end{aligned}$$

Ce nouvel estimateur cumule le meilleur des deux mondes : son biais ne comporte plus la partie preponderante (constante) du biais de $\hat{\alpha}_1$, et dans le même temps sa variance reste negligeeable par rapport a celle de $\hat{\alpha}_2$. La figure 1 met en evidence ce resultat. On constate en outre que le second estimateur n'est utilise que dans l'unique but de faire disparaître cette partie constante.

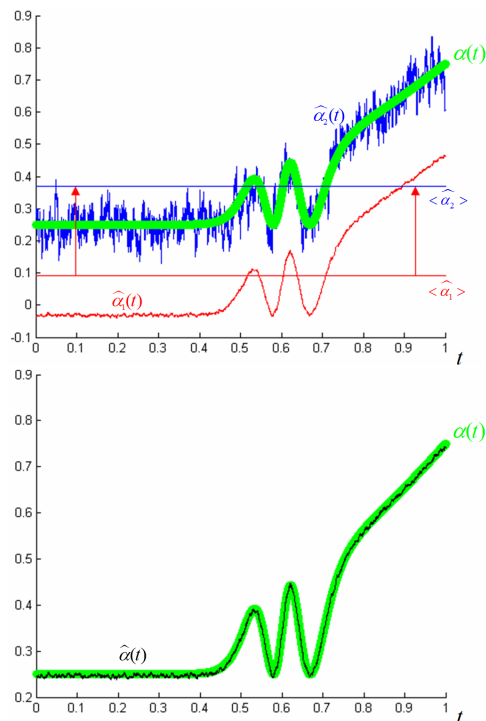


FIG. 1 – A partir de $\hat{\alpha}_1$ et $\hat{\alpha}_2$ (en haut), on construit $\hat{\alpha}$ (en bas)

Le parametre δ_1^* du premier estimateur sera classiquement choisi de sorte qu'il realise le meilleur compromis entre $B_{1V}^{(N)}(1)^2$ et $V_1^{(N)}(1)$. Il est important de noter qu'en revanche, δ_2^* ne sera pas la valeur qui minimise $R_2^{(N)}(2)$, mais celle minimisant $B_2^{(N)}(2)^2 + N^{(\delta_2 - 1)} V_2^{(N)}(2)$.

3 Exemple d'application

3.1 Definition d'un processus fractal a partir de ses coefficients d'ondelettes

Pour illustrer le theoreme 2.1, nous definissons un processus a regularite prescrite par ses coefficients d'ondelettes : ils sont independants et definis en loi par :

$$c_{j,k} \rightsquigarrow G \cdot \mathcal{N}(0, \frac{2}{j,k}) \text{ v.a gaussienne centree}$$

pour tout $j = 0 \dots n - 1$ $k = 0 \dots 2^j - 1$

ou $c_{j,k} = 2^{-j} (k^{2^{-j}})$ $(t) = H(t) + \frac{1}{2}$ suamment reguliere et $G > 0$ est un facteur d'echelle deterministe

Soit $(t_1^n \dots t_{2^n}^n)$ un echantillonnage de l'intervalle $I = [0, 1]$ sur $N = 2^n$ points. En tout point $t = t_k^n$, on considere l'ensemble $c_{j,k(j,t)}$ $j = 0 \dots n - 1$ des coefficients d'ondelettes situes « au dessus » de lui, avec $k(j, t) = \lfloor (t + 1)2^{j+1-n} \rfloor$ (Pour une raison de simplicite d'ecriture, il sera simplement note k par abus de notation).

Dans le but d'identifier la fonction nous definissons

$\hat{\alpha}_1(j, t)$		
<i>iaais</i>	$< \frac{2}{3}$	$\mathcal{O} \frac{2^{-j\delta}}{j} - \frac{\log_2(G)}{j}$
	$\geq \frac{2}{3}$	$\mathcal{O} j\beta^2 2^{j(2\delta-2)} - \frac{\log_2(G)}{j}$
	<i>optimal</i>	$B_1^* \approx \frac{2}{3} - \frac{\ln \frac{j^2 \beta^2 \ln(2)^2}{6}}{3j \ln(2)}$ $j = n - 1$
<i>ariance</i>		$\mathcal{O} \frac{2^{-j\delta}}{j^2}$
<i>isque optimal (G = 1)</i>		$R_1^* \approx \frac{4}{5} - \frac{\ln \frac{j^4 \beta^4 \ln(2)^4}{72}}{5j \ln(2)}$ $j = n - 1$
$\hat{\alpha}_2(m, t)$		
<i>iaais</i>	$< \frac{2}{3}$	$\mathcal{O} \frac{2^{-(n-m)\delta}}{m^2}$
	$\geq \frac{2}{3}$	$\mathcal{O} \frac{(n-m)^2 \beta^2 2^{(n-m)(2\delta-2)}}{m^2}$
	<i>optimal</i>	$B_2^* \approx \frac{2}{3} - \frac{\ln \frac{(n-m)^2 \beta^2 \ln(2)^2}{6}}{3(n-m) \ln(2)}$ $m \approx 2$
<i>ariance</i>		$\mathcal{O} \frac{2^{-(n-m)\delta}}{m^4}$
<i>isque optimal</i>		$R_2^* \approx \frac{4}{5} - \frac{\ln \frac{(n-m)^4 \beta^4 \ln(2)^4}{24}}{5(n-m) \ln(2)}$ $m \approx 6$

deux estimateurs :

$$ec\ premo\ ennage : \hat{\alpha}_1(j, t) - \frac{1}{2j} \log_2 \left(\frac{1}{2^{p+1}} \sum_{i=-p}^p c_{j,k+i}^2 \right)$$

$$remo\ ennage\ et\ regression : \hat{\alpha}_2(m, t) = \frac{12}{m(m-1)(m+1)} \sum_{j=n-m}^{n-1} j - \frac{2n-m-1}{2} \left(\frac{-\log_2 \frac{1}{2^{p+1}} \sum_{i=-p}^p c_{j,k+i}^2}{2} \right)$$

Le moyennage s'effectue sur $2p + 1 = \lfloor 2^{n\delta} \rfloor$ coefficients avec $0 < \delta < 1$ et la regression sur les m derniers niveaux.

Grâce à ce processus simple, à voir comme un outil permettant de mener des calculs explicites, nous pouvons alors obtenir des équivalents précis pour les biais, variances, et paramètres optimaux de ces estimateurs, en fonction du nombre de coefficients $N = 2^n$, de la largeur de l'intervalle de moyennage ajustée par δ , de l'échelle j , du nombre de niveaux de regression m et de la valeur de la dérivée de α en t , $\beta_{j,k}$. Ces résultats sont regroupés dans le tableau suivant :

Il reste maintenant à s'assurer que les hypothèses du théorème 2.1 sont vérifiées : les deux estimateurs sont consistants (H_1), le biais de $\hat{\alpha}_1$ est bien en deux parties (H_2) et la seconde partie en $-\frac{\log_2(G)}{j}$ est prépondérante devant $\frac{2^{-j\delta}}{j}$ dès que G est différent de 1 (H_3). En outre, étant donné qu'il n'y a pas de chevauchement pour des points suffisamment éloignés en temps, on obtient (H_6)

grâce à l'indépendance des coefficients d'ondelettes.

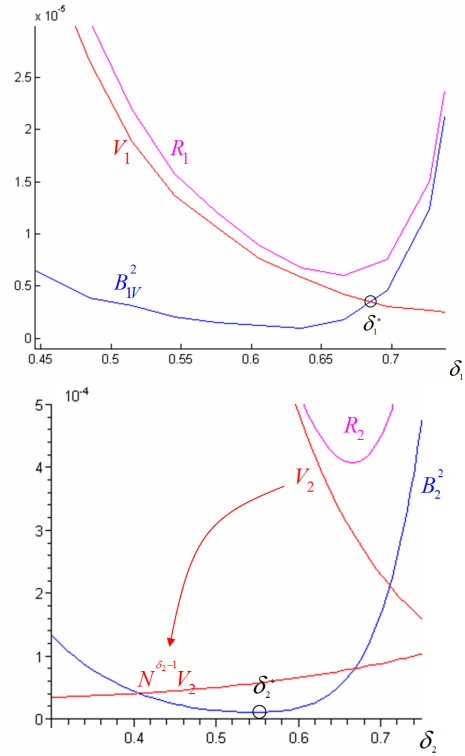


FIG. 2 - δ_1 choisi pour minimiser le risque standard de $\alpha_1(\delta, j, t)$ (en haut), et δ_2 pour minimiser le biais de $\alpha_2(\delta, m, t)$ (en bas)

En n , un réglage optimal des paramètres des deux estimateurs va permettre de valider les deux dernières hypothèses. Pour $\hat{\alpha}_1$, il faut choisir j le plus grand possible, soit $j = n - 1$ et δ_1 de sorte qu'il minimise le risque de l'estimateur sans la partie constante du biais. Nous prendrons donc $\delta_1 = \delta_{R_1}^*$ (figure 2 à gauche).

En ce qui concerne $\hat{\alpha}_2$, il faut chercher à minimiser $B_2^{(N)}(\delta_2)^2 + N^{(\delta_2-1)} V_2^{(N)}(\delta_2)$. Or $\forall \delta_2 N^{(\delta_2-1)} V_2^{(N)}(\delta_2) = \mathcal{O} N^{-1}$, ou δ_2 n'apparaît plus en puissance. Par conséquent, la minimisation revient essentiellement à minimiser le biais, soit une regression sur deux niveaux, et $\delta_2 = \delta_{B_2}^*$ (figure 2 à droite).

Comparons maintenant la variance de ce nouvel estimateur avec ses paramètres optimaux $\hat{\alpha}_1(\delta_{R_1}^*, n - 1)$ ($\delta_{B_2}^*, 2$) avec celle de $\hat{\alpha}_2(\delta_{R_2}^*, 6)$, qui réalise le meilleur compromis sans cette méthode, en se plaçant dans le cas asymptotique :

$$V_{R_1}(\delta_{R_1}^*, n - 1) - V_{B_2}(\delta_{B_2}^*, 2) = \mathcal{O} \frac{2^{-n\delta_{R_1}^*}}{n^2} + 2^{-n} = \mathcal{O} \frac{2^{-n\delta_{R_1}^*}}{n^2}$$

et $V_{R_2}(\delta_{R_2}^*, 6) = \mathcal{O} 2^{-n\delta_{R_2}^*}$

On remarque que même si les taux exponentiels de convergences asymptotiques optimaux sont les mêmes pour $\hat{\alpha}_1$ et $\hat{\alpha}_2$, la présence du terme $1/n^2$ assure de réduire la variance. De fait, $\hat{\alpha}_1$ donne de meilleurs résultats « à l'œil » que $\hat{\alpha}_2$. Ceci est illustré par la figure 1, obtenue en im-

plementant les trois estimateurs pour le processus decrit ci-dessus echantillonne sur 2^{16} points.

Notons par ailleurs que les parametres optimaux asymptotiques $4/5$ et $2/3$ ne sont atteints que pour des tailles de signaux inaccessibles, de l'ordre de 2^{200} . En pratique, ils sont proches de 0.65 et 0.5 pour des valeurs « raisonnables » des parametres, comme on le constate sur la figure 2 : celle-ci represente les choix optimaux de α_1 et α_2 pour un echantillonnage sur 2^{16} points et la fonction presentee sur la figure 1.

La même principe a été appliqué avec succès sur des mouvements Browniens multifractionnaires [4], où $H(t)$ est estimée soit par des oscillations, soit par des variations quadratiques généralisées, et il est implémenté dans la boîte à outils FracLab [5].

References

- [1] Jacques Istas et Gabriel Lang, Quadratic variations and estimation of the local Holder index of a gaussian process. *Ann. Inst. Henri Poincaré, Probab. Stat.*, n 4, p. 47–64, 1997.
- [2] X. Guyon et G. Leon, Convergence en loi des h -variations d'un processus Gaussien stationnaire, *Ann. Inst. Henri Poincaré* 5, 65–75, 1989.
- [3] David L. Donoho, De-noising by soft thresholding, *Transactions on Information Theory*, 1994.
- [4] Antoine Ayache et Jacques Levy-Vehel, Identification of the pointwise Holder exponent of Generalized Multifractal Brownian Motion, *Stochastic Processes and their Applications*, 95, 2004.
- [5] FracLab, *Fractal analysis toolbox for signal and image processing*, <http://www.irccyn.ec-nantes.fr/FracLab/>