

Estimation du coefficient de régularité locale d'une trajectoire de processus

Delphine Blanke ^{a,b}

^a L.M.A.H., Université du Havre, 25, rue Philippe Lebon, BP 540, 76058 Le Havre cedex, France

^b L.S.T.A., Université Paris 6, France

Reçu le 7 octobre 2001 ; accepté après révision le 1^{er} décembre 2001

Note présentée par Paul Deheuvels.

Résumé

Nous considérons dans cette Note, les processus satisfaisant localement une condition de type Hölder avec un coefficient inconnu γ_0 . Nous étudions deux familles d'estimateurs pour γ_0 : l'une basée sur la connaissance globale de la trajectoire sur $[0, T_N]$, $T_N \nearrow \infty$, et la deuxième fondée sur n observations échantillonnées aux temps $i\delta_n$ avec $i = 0, \dots, n-1$, $\delta_n \rightarrow 0$ et $n\delta_n \rightarrow \infty$ pour $n \rightarrow \infty$. Nous donnons les vitesses de convergence presque sûre pour ces deux familles distinctes. Pour citer cet article : D. Blanke, C. R. Acad. Sci. Paris, Ser. I 334 (2002) 145–148. © 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

Estimation of the local regularity index of a sample path

Abstract

In this Note, we consider processes that satisfied a local Hölder condition with unknown coefficient γ_0 . We study two families of estimators for γ_0 : the first one based upon the whole sample path over $[0, T_N]$, $T_N \nearrow \infty$, and the second one constructed with n observations at sampling rate δ_n , $\delta_n \rightarrow 0$ and $n\delta_n \rightarrow \infty$ as $n \rightarrow \infty$. For the almost sure convergence, we give the rates for these two families. To cite this article: D. Blanke, C. R. Acad. Sci. Paris, Ser. I 334 (2002) 145–148. © 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

1. Introduction et notations

Soit $\{X_t, t \in \mathbb{R}\}$ un processus réel, séparable et mesurable dont les trajectoires satisfont localement une condition de type Hölder :

$$\limsup_{\delta \searrow 0} \sup_{t_1 \leq t \leq t_2 - \delta} \sup_{0 \leq s \leq \delta} |X_{t+s} - X_t|/s^\gamma L(s) \leq 1, \quad 0 \leq t_1 < t_2, \quad (1.1)$$

où $0 < \gamma \leq 1$ et $L(\cdot)$ est une fonction à variation lente. Notre but est alors d'estimer γ_0 avec $\gamma_0 = \sup\{\gamma, \gamma \in \Gamma\}$ où Γ est l'ensemble des γ vérifiant (1.1).

Dans le cas Gaussien, les estimateurs fondés sur les accroissements et les variations quadratiques généralisées ont fait l'objet de nombreuses études, voir par exemple Constantine et Hall [3], Istas et Lang [5], Peltier et Levy-Vehel [6], Taylor et Taylor [8]. D'autres types d'estimateurs ont également été utilisés notamment par Philippe et Thilly [7].

Adresse e-mail : Delphine.Blanke@univ-lehavre.fr (D. Blanke).

Dans cette Note, nous nous plaçons dans un cadre beaucoup plus général (qui inclut les processus Gaussiens) et nous étudions le comportement presque sûr de nos estimateurs. Deux grandes familles sont abordées : soit la trajectoire est observée globalement sur $[0, T_N]$ où (T_N) est une suite de réels tels que $T_N \nearrow \infty$, $T_{N+1} - T_N \geq a > 0$, $N \geq 1$, avec a constante; soit on dispose de n observations discrétisées $X_0, \dots, X_{(n-1)\delta_n}$ où δ_n est le pas d'échantillonnage tel que $\delta_n \rightarrow 0$, $n\delta_n \rightarrow \infty$ quand $n \rightarrow \infty$ (c'est-à-dire que les observations sont fréquentes et disponibles sur une longue durée). Dans la section suivante, nous donnons les comportements asymptotiques obtenus pour la première famille d'estimateurs; le cas discrétisé est présenté dans la section 3. La section 4 compare et discute les résultats obtenus.

2. Le cas continu

Supposons tout d'abord que la trajectoire est observée globalement sur $[0, T_N]$, et considérons la statistique :

$$S_{T_N, \ell}^{(k)}(\gamma_0) = \frac{1}{T_N - \ell\delta_{T_N}} \int_0^{T_N - \ell\delta_{T_N}} \left| \frac{X_{t+\ell\delta_{T_N}} - X_t}{(\ell\delta_{T_N})^{\gamma_0}} \right|^k dt,$$

où $k \in [1, \infty[$, $\ell \in \mathbb{R}$, sont des nombres donnés et $\delta_{T_N} \rightarrow 0$ quand $T_N \nearrow \infty$.

Pour obtenir des résultats asymptotiques, nous utilisons le coefficient de forte mélangeance classique défini par : $\alpha_X(u) = \sup_{t \in \mathbb{R}} \sup_{\substack{A \in \sigma(X_s, s \leq t) \\ B \in \sigma(X_s, s \geq t+u)}} |P(A \cap B) - P(A)P(B)|$, $u \geq 0$. Rappelons que (X_t) est dit géométriquement fortement mélangeant (GFM) s'il existe des constantes $c > 0$ et $0 \leq \rho < 1$ telles que $\alpha_X(j) \leq c\rho^j$, $j > 0$.

THÉORÈME 1. – Si $\{X_t, t \in \mathbb{R}\}$ est un processus GFM vérifiant

$$\exists \delta_0 > 0, \exists a > 0, \forall \delta \leq \delta_0, \sup_{t \in \mathbb{R}^+} E(\exp(a\delta^{-\gamma_0}|X_{t+\delta} - X_t|)) \leq M < \infty, \tag{2.1}$$

alors, nous obtenons pour une constante positive C_1

$$\limsup_{T_N \nearrow \infty} \sqrt{T_N / \ln T_N} |S_{T_N, \ell}^{(k)}(\gamma_0) - ES_{T_N, \ell}^{(k)}(\gamma_0)| \leq C_1 \text{ p.s.} \tag{2.2}$$

La constante C_1 est compliquée mais explicite, elle dépend des valeurs de M, ρ, α_X, k . La condition locale (1.1) est vérifiée sous l'hypothèse (2.1) (Csáki and Csörgő [4]). De plus, si $T \mapsto Z_T := (1/(\sqrt{T \ln T})) \int_0^{T-\ell\delta_T} \left| \frac{X_{t+\ell\delta_T} - X_t}{(\ell\delta_T)^{\gamma_0}} \right|^k dt$ est p.s. à trajectoires uniformément continues sur \mathbb{R}^+ , alors d'après une version du lemme de Borel–Cantelli en temps continu (voir Bosq [2]) le résultat (2.2) reste vrai avec T_N remplacé par T . Pour l'étude du terme de biais, nous avons :

LEMME 1. – Supposons : $\exists \delta_0 > 0, \forall \delta \leq \delta_0, \forall v \geq 0, E|X_{v+\delta} - X_v|^k = \delta^{k\gamma_0} m_{v, \delta}^{(k)}$ avec $\frac{1}{T_N - \ell\delta_{T_N}} \int_0^{T_N - \ell\delta_{T_N}} m_{v, \ell\delta_{T_N}}^{(k)} dv = m_k + R_{T_N}^{(k)}(\ell\delta_{T_N})$,

- (1) si de plus $R_{T_N}^{(k)}(\delta_{T_N}) \rightarrow 0$ quand $T_N \nearrow \infty$, alors $\lim_{T_N \nearrow \infty} (ES_{T_N, \ell}^{(k)}(\gamma_0) - m_k) = 0$,
- (2) si $\lim_{T_N \nearrow \infty} \delta_{T_N}^{-\beta_0} R_{T_N}^{(k)}(\delta_{T_N}) = b_k$ ($\beta_0 > 0, b_k > 0$), alors $\lim_{T_N \nearrow \infty} \delta_{T_N}^{-\beta_0} (ES_{T_N, \ell}^{(k)}(\gamma_0) - m_k) = b_k \ell^{\beta_0}$.

Les hypothèses du lemme 1(1) sont notamment satisfaites pour des processus à accroissements stationnaires dès que $E|X_u - X_0|^k / u^{k\gamma_0} \rightarrow m_k, u \rightarrow 0$. L'ensemble de ces résultats nous permettent de considérer un premier type d'estimateurs pour γ_0 :

$$\tilde{\gamma}_{0, T_N}^{(\ell, k)} = \ln \left(\frac{1}{T_N - \ell\delta_{T_N}} \int_0^{T_N - \ell\delta_{T_N}} |X_{v+\ell\delta_{T_N}} - X_v|^k dv \right) / k \ln(\ell\delta_{T_N}). \tag{2.3}$$

La décomposition $\tilde{\gamma}_{0, T_N}^{(\ell, k)} = \gamma_0 + (\ln S_{T_N, \ell}^{(k)}(\gamma_0) / k \ln(\ell\delta_{T_N}))$ et les résultats précédents conduisent au

THÉORÈME 2. – *Sous les hypothèses du théorème 1 et du lemme 1(1), nous obtenons*

$$\lim_{T_N \nearrow \infty} \ln(\delta_{T_N}^{-k}) (\gamma_0 - \tilde{\gamma}_{0,T_N}^{(\ell,k)}) = \ln(m_k) \quad p.s.$$

La vitesse de convergence obtenue indique que cet estimateur ne pourra être performant que pour de très petites valeurs de δ_{T_N} . Néanmoins, il apparaît comme relativement robuste car sa construction ne dépend pas de paramètres inconnus. De plus, la vitesse de convergence de la statistique $S_{T_N,\ell}^{(k)}(\gamma_0)$ n'intervenant pas, la condition (2.1) peut être assouplie en conséquence. Des estimateurs plus élaborés peuvent également être recherchés sous la forme :

$$\hat{\gamma}_{0,T_N}^{(\ell_1,\ell_2,k)} = (\ln(\ell_2 \delta_{T_N}) \tilde{\gamma}_{0,T_N}^{(\ell_2,k)} - \ln(\ell_1 \delta_{T_N}) \tilde{\gamma}_{0,T_N}^{(\ell_1,k)}) / \ln(\ell_2/\ell_1). \quad (2.4)$$

THÉORÈME 3. – *Sous les hypothèses du théorème 1 et du lemme 1(2), et pour tout choix $\delta_{T_N} \rightarrow 0$ tel que $\delta_{T_N} = o((\ln T_N/T_N)^{1/(2\beta_0)})$, $\limsup_{T_N \nearrow \infty} \sqrt{T_N/\ln T_N} (\hat{\gamma}_{0,T_N}^{(\ell_1,\ell_2,k)} - \gamma_0) \leq 2C_1/(km_k \ln(\ell_2/\ell_1))$ p.s.*

3. Le cas discrétisé

Considérons maintenant le cas où l'on dispose de n observations discrétisées $X_{i\delta_n}$, pour $i = 0, \dots, n-1$, avec les conditions $\delta_n \rightarrow 0, n\delta_n \rightarrow \infty$. La statistique d'intérêt se présente maintenant sous la forme :

$$S_{n,\ell}^{(k)}(\gamma_0) = \frac{1}{n(\ell\delta_n)^{k\gamma_0}} \sum_{i=0}^{n-\ell} |X_{(i+\ell)\delta_n} - X_{i\delta_n}|^k,$$

où $k \in [1, \infty[$ et $\ell \in \mathbb{N}$ ($1 \leq \ell < n$) sont des nombres donnés. De manière équivalente à la section précédente, nous obtenons successivement :

THÉORÈME 4. – *Soit $\{X_t, t \in \mathbb{R}\}$ un processus GFM vérifiant la condition (2.1) et si $n\delta_n/(\ln n)^{2k+3} \rightarrow \infty$ quand $n \rightarrow \infty$, on obtient pour une constante positive C_2 ,*

$$\limsup_{n \rightarrow \infty} \sqrt{n\delta_n/\ln n} |S_{n,\ell}^{(k)}(\gamma_0) - \mathbb{E}S_{n,\ell}^{(k)}(\gamma_0)| \leq C_2 \quad p.s.$$

LEMME 2. – *Supposons : $\exists \delta_0 > 0, \forall \delta \leq \delta_0, \forall v \geq 0, \mathbb{E}|X_{v+\delta} - X_v|^k = \delta^{k\gamma_0} m_{v,\delta}^{(k)}$ avec $\frac{1}{n} \sum_{i=0}^{n-\ell} m_{i\delta_n, \ell\delta_n}^{(k)} = m_k + R_n^{(k)}(\ell\delta_n)$,*

(1) *si de plus $R_n^{(k)}(\delta_n) \rightarrow 0$ quand $n \rightarrow \infty$, alors $\lim_{n \rightarrow \infty} (\mathbb{E}S_{n,\ell}^{(k)}(\gamma_0) - m_k) = 0$,*

(2) *si $\lim_{n \rightarrow \infty} \delta_n^{-\beta_0} R_n^{(k)}(\delta_n) = b_k$ ($\beta_0 > 0, b_k > 0$), alors $\lim_{n \rightarrow \infty} \delta_n^{-\beta_0} (\mathbb{E}S_{n,\ell}^{(k)}(\gamma_0) - m_k) = b_k \ell^{\beta_0}$.*

Considérons les estimateurs définis par (2.3) et (2.4) sous une forme discrétisée :

$$\tilde{\gamma}_{0,n}^{(\ell,k)} = \ln \left(\frac{1}{n} \sum_{i=0}^{n-\ell} |X_{(i+\ell)\delta_n} - X_{i\delta_n}|^k \right) / k \ln(\ell\delta_n) \quad (3.1)$$

$$\hat{\gamma}_{0,n}^{(\ell_1,\ell_2,k)} = (\ln(\ell_2 \delta_n) \tilde{\gamma}_{0,n}^{(\ell_2,k)} - \ln(\ell_1 \delta_n) \tilde{\gamma}_{0,n}^{(\ell_1,k)}) / \ln(\ell_2/\ell_1). \quad (3.2)$$

THÉORÈME 5. –

(1) *Sous les conditions du théorème 4 et du lemme 2(1),*

$$\lim_{n \rightarrow \infty} \ln(\delta_n^{-k}) (\gamma_0 - \tilde{\gamma}_{0,n}^{(\ell,k)}) = \ln(m_k) \quad p.s.$$

(2) Sous les conditions du théorème 4 et du lemme 2(2), avec le choix $\delta_n = (\ln n/n)^{1/(2\beta_0+1)}$,

$$\limsup_{n \rightarrow \infty} (n/\ln n)^{\beta_0/(2\beta_0+1)} (\widehat{\gamma}_{0,n}^{(\ell_1, \ell_2, k)} - \gamma_0) \leq (2C_1 + |b_k|(\ell_1^{\beta_0} + \ell_2^{\beta_0}))/km_k \ln(\ell_2/\ell_1) \quad p.s.$$

4. Discussion

Les résultats de la section 3 sont établis dans Blanke [1]. De plus, les estimateurs définis par la relation (3.2) étant particulièrement sensibles à la valeur de β_0 , un estimateur de ce paramètre y est proposé et étudié. De même, des simulations permettent de comparer le comportement de l'ensemble de ces estimateurs. Les résultats de la section 2 sont nouveaux mais découlent des mêmes techniques de démonstration. Enfin, des exemples de processus satisfaisant les hypothèses de cette Note se trouvent détaillés dans le même article.

Les résultats précédents nous montre que la connaissance globale de la trajectoire sur $[0, T_N]$ apporte beaucoup plus d'informations que dans le cas discrétisé, ce qui améliore le comportement général des estimateurs. Par exemple, selon le théorème 5(1), la condition $n\delta_n \rightarrow +\infty$ implique que la vitesse de convergence $\ln(\delta_n^{-1})$ est simplement de l'ordre de $(\ln n)^{-1}$! Parallèlement, en temps continu (théorème 2), le choix de δ_{T_N} étant libre, la vitesse de convergence peut être optimisée en conséquence. Pour les mêmes raisons et avec δ_{T_N} suffisamment petit, les estimateurs du type (2.4) seront moins sensibles au choix de β_0 que ceux donnés par la relation (3.2). Néanmoins, le cas discrétisé conserve l'avantage de s'appliquer à un plus grand nombre de situations pratiques.

Remerciements. Je tiens à remercier le rapporteur pour l'ensemble de ses remarques.

Références bibliographiques

- [1] Blanke D., Estimation of local smoothness coefficients for continuous time processes, *Statist. Inf. Stochastic Process*, 2002, to appear.
- [2] Bosq D., *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction*, 2nd ed., Lecture Notes in Statist., Vol. 110, Springer-Verlag, New York, 1998.
- [3] Constantine A.G., Hall P., Characterizing surface smoothness via estimation of effective fractal dimension, *J. Roy. Statist. Soc. Ser. B* 56 (1) (1994) 97–113.
- [4] Csáki E., Csörgő M., Inequalities for increments of stochastic processes and moduli of continuity, *Ann. Probab.* 20 (2) (1992) 1031–1052.
- [5] Istas J., Lang G., Quadratic variations and estimation of the local Hölder index of a Gaussian process, *Ann. Inst. H. Poincaré Probab. Statist.* 33 (4) (1997) 407–436.
- [6] Peltier R.F., Lévy-Véhel J., A new method for estimating the parameter of fractional Brownian motion, Technical Report 2396, I.N.R.I.A., Novembre 1994.
- [7] Philippe A., Thilly E., Identification of locally self-similar Gaussian process by using convex rearrangements, *Publications I.R.M.A.* 53 (4) (2000), Université des Sciences et Technologies de Lille.
- [8] Taylor C.C., Taylor S.J., Estimating the dimension of a fractal, *J. Roy. Statist. Soc. Ser. B* 53 (1991) 353–364.