Statistics

Nonparametric estimation of a trend based upon sampled continuous processes

David Degras

Laboratoire de statistique théorique et appliquée, Université Paris 6, boîte 158, 175, rue du Chevaleret, 75013 Paris, France

Abstract

Let $X = \{X(t), \ t \in [0, T]\}$ be a second order random process of which n independent realizations are observed on a fixed grid of p time points. Under mild regularity assumptions on the sample paths of X, we show the asymptotic normality of suitable nonparametric estimators of the trend function $\mu = \mathbb{E}X$ in the space $C([0, T])$ as $n, p \to \infty$ and, using Gaussian process theory, we derive approximate simultaneous confidence bands for μ. ***To cite this article: D. Degras, C. R. Acad. Sci. Paris, Ser. I 347 (2009).***

Résumé

Estimation non paramétrique d'une tendance à partir de réalisations d'un processus à temps continu. Soit $X=\{X(t),$ $t \in [0, T]\}$ un processus aléatoire du second ordre dont on observe n réalisations indépendantes sur une grille de p points déterministes. Sous de faibles conditions de régularité sur les trajectoires de X, nous prouvons la normalité asymptotique d'estimateurs non paramétriques de la tendance $\mu = \mathbb{E}X$ dans l'espace $C([0, T])$ lorsque $n, p \to \infty$, puis nous obtenons des bandes de confiance simultanées approchées pour μ à l'aide de la théorie des processus Gaussiens. ***Pour citer cet article : D. Degras, C. R. Acad. Sci. Paris, Ser. I 347 (2009).***

1. Introduction

In various applied fields such as internet traffic monitoring, medical imagery, or signal processing, modern technology has allowed one to collect data routinely from population samples with a high temporal and/or spatial resolution. Indeed, such datasets should be viewed as (collections of) curves or functions rather than as high-dimensional vectors; they are thus commonly termed *functional data*. (See [8,13] for a comprehensive introduction to functional data analysis.) Typical functional data may be modeled as observations of independent realizations $X_1, \ldots, X_n$ of a second order random process $X = \{X(t), t \in \mathcal{D}\}$ at fixed design points $t_1, \ldots, t_p$, where $\mathcal{D}$ denotes a continuous temporal and/or spatial domain. In this framework, the observed data are

$$Y_{ij} = X_i(t_j) + \varepsilon_{ij}, \quad 1 \leqslant i \leqslant n, \ 1 \leqslant j \leqslant p, \tag{1}$$

E-mail address: david.degras@upmc.fr.

where the ε_{ij} are mean zero random variables (r.v.) representing potential measurement errors.

The trend function $\mu = \mathbb{E}X$ often appears as a population mean response function, which motivates its inference. The nonparametric regression literature contains several results on the asymptotic properties of estimators of μ as the sample sizes n and p go to infinity. For instance when $\mathcal{D} = [0, 1]$, mean-square convergence rates of kernel and spline estimators can be found in [2,3,6,9]. When $\mathcal{D}$ is a compact metric space, [5] gives a universal consistency result as well as the asymptotic normality of all usual regression estimators in the sense of finite dimensional distributions and of the space $L_2(\mathcal{D})$, with an application to simultaneous confidence intervals. The task of building (nonparametric and simultaneous) confidence bands for μ, which proves useful in various problems of prediction, model diagnostic, or calibration (e.g. [1,11]), has received considerable attention in the classical regression setting (e.g. [7,14]) but not, to our knowledge, for functional data.

In this Note, we study the model (1) in the case where the random process X is indexed by a compact interval $\mathcal{D} = [0, T]$ and has mildly regular sample paths. In Section 2, we state the asymptotic normality of suitable nonparametric estimators of μ in the space $C([0, T])$ of all continuous functions on $[0, T]$ as $n, p \to \infty$. In Section 3, we use previous results to build approximate simultaneous confidence bands for μ. Finally in Section 4, some potential applications and extensions of our results are discussed.

2. Asymptotic normality of nonparametric estimators

We state here the assumptions made on the random process X and on the model (1) of Section 1.

(A.1) X is mean-square continuous on $\mathcal{D} = [0, T]$.
(A.2) The sample paths $X(\omega, \cdot)$ are almost surely (a.s.) variation-bounded, with their total variation bounded by $B(\omega)$, where B is a r.v. with finite variance; or
(A.2′) $|X(\omega, s) - X(\omega, t)| \leqslant C|s - t|^{\beta}$ a.s. for some positive constants C and β.
(A.3) μ has two bounded derivatives on $[0, T]$.
(B.1) The data form a triangular array: $Y_{ij} = Y_{ij}(n)$, $t_j = t_j(n)$, and $p = p(n)$, with $p(n) \to \infty$ as $n \to \infty$.
(B.2) The random errors ε_{ij} are mutually independent and independent of the X_i; they have mean zero and common variance $\sigma^2 \geqslant 0$.
(B.3) The t_j are ordered ($0 \leqslant t_1 < \cdots < t_p \leqslant T$) and they have a quasi-uniform repartition, i.e., writing $t_0 = 0$ and $t_{p+1} = T$, it holds that $\frac{\max_{0 \leqslant j \leqslant p}(t_{j+1} - t_j)}{\min_{1 \leqslant j \leqslant p-1}(t_{j+1} - t_j)} = \mathcal{O}(1)$ as $n, p \to \infty$.

Note that (A.2′) implies (A.2). Also, (B.3) ensures that the t_j are regularly spaced in $[0, T]$; it is fulfilled e.g. when the t_j are equally spaced or are generated by a regular probability density function (p.d.f.).

For each assumption (A.2) and (A.2′), we now introduce a suitable nonparametric estimator of μ and give its asymptotic distribution. Under (A.2), we use the interpolation-type estimator of [4], denoted by $\hat{\mu}_C$, with a boundary correction. We recall here its definition. Let $\overline{Y}_j = (\sum_{i=1}^{n} Y_{ij})/n$ for $1 \leqslant j \leqslant p$, and let $\overline{Y}(t)$ be the process obtained by linear interpolation of the $(t_j, \overline{Y}_j)$ such that $\overline{Y}(t) = \overline{Y}_1$ if $t \leqslant t_1$ and $\overline{Y}(t) = \overline{Y}_p$ if $t \geqslant t_p$. The estimator $\hat{\mu}_C$ is the convolution of a kernel function K with $\overline{Y}$:

$$\hat{\mu}_C(t) = \frac{1}{\int_0^T K_h(t - u)\,\mathrm{d}u} \int_0^T K_h(t - u)\overline{Y}(u)\,\mathrm{d}u, \quad K_h(\cdot) = K(\cdot/h)/h. \tag{2}$$

For convenience we take K as a symmetric, compactly supported, Lipschitz-continuous p.d.f. The real $h > 0$ is a fixed bandwidth. Following [12], we say that a sequence (Z_n) of random elements of $C([0, T])$ converges weakly to a limit Z in $C([0, T])$ if $\mathbb{E}\varphi(Z_n) \to \mathbb{E}\varphi(Z)$ as $n \to \infty$ for all uniformly continuous functional φ on $C([0, T])$ equipped with the sup-norm. We denote by R the covariance function of the process X and by $\mathcal{G}(0, C)$ any Gaussian process indexed by $[0, T]$ with mean zero and covariance C. We are now in position to state the weak convergence of $\hat{\mu}_C$ in $C([0, T])$ as $n, p \to \infty$ (recall that $p = p(n)$).

Theorem 2.1. *Assume that* (A.1), (A.2), (B.1)–(B.3) *hold and that* $h = h(n, p) \to 0$ *and* $ph^2 \to \infty$ *as* $n, p \to \infty$. *Then* $n^{1/2}(\hat{\mu}_C - \mathbb{E}\hat{\mu}_C)$ *converges weakly to* $\mathcal{G}(0, R)$ *in* $C([0, T])$. *If in addition* (A.3) *holds and* $n = o(p)$, $nh^2 \to 0$ *as* $n, p \to \infty$, *then* $n^{1/2}(\hat{\mu}_C - \mu)$ *also converges to* $\mathcal{G}(0, R)$ *in* $C([0, T])$.

Next, we address the case where X satisfies (A.2′) (Hölder continuity). We consider the local linear estimator, denoted here by $\hat{\mu}_L$ and defined by

$$\hat{\mu}_L(t) = \hat{\theta}_0, \qquad (\hat{\theta}_0, \hat{\theta}_1) = \operatorname*{argmin}_{(\theta_0, \theta_1)} \sum_{j=1}^{p} \big(\overline{Y_j} - \theta_0 - (t_j - t)\theta_1\big)^2 K_h(t_j - t). \tag{3}$$

Theorem 2.2. *Assume that* (A.1), (A.2′), (B.1)–(B.3) *hold and that* $h = h(n, p) \to 0$ *and* $ph^2 \to \infty$ *as* $n, p \to \infty$. *Then* $n^{1/2}(\hat{\mu}_L - \mathbb{E}\hat{\mu}_L)$ *converges weakly to* $\mathcal{G}(0, R)$ *in* $C([0, T])$. *If in addition* (A.3) *holds,* $n = \mathrm{o}(p^2)$, *and* $nh^4 \to 0$ *as* $n, p \to \infty$, *then* $n^{1/2}(\hat{\mu}_L - \mu)$ *also converges to* $\mathcal{G}(0, R)$ *in* $C([0, T])$.

Remarks.

(1) The proofs of the these theorems are similar and rely on the following steps: (i) note that the estimator is linear in the data; (ii) use the functional central limit Theorem 10.6 of [12] for the estimator applied to the data without noise $X_i(t_j)$; (iii) show that under condition $ph^2 \to \infty$, the estimator applied to the errors ε_{ij} becomes negligible in probability before $n^{-1/2}$ as $n, p \to \infty$, uniformly over $[0, T]$; (iv) impose additional conditions on μ and on the joint rates of n, p and h to make the asymptotic bias of the estimator as $\mathrm{o}(n^{-1/2})$ uniformly over $[0, T]$ (in particular the rate $n = \mathrm{o}(p)$ used in Theorem 2.1 is only needed to control the boundary effects in the bias of $\hat{\mu}_C$).
(2) Under (A.2), we can prove asymptotic normality in $C([0, T])$ only for $\hat{\mu}_C$. This is because $\hat{\mu}_C$, as opposed to more classical estimators, has the remarkable feature of preserving monotonicity and thus satisfies (A.2) like X, which makes the step (ii) of the theorem proof straightforward. On the other hand, under the stronger assumption (A.2′) the asymptotic normality in $C([0, T])$ can be obtained for various classical kernel or projection estimators, as well as for $\hat{\mu}_C$. The choice of $\hat{\mu}_L$ here was motivated by the popularity of this estimator and by its good bias properties.
(3) The condition $ph^2 \to \infty$ can be dropped in both theorems if there is no noise in the data ($\sigma = 0$). Besides, the results carry over to the case of correlated errors, e.g. of autoregressive or mixing type.
(4) Theorem 2.2 corrects a mistake in the Section 4 of [15] which gives $(nph)^{1/2}$ as the normalizing rate for the weak convergence of $\hat{\mu}_L(t) - \mathbb{E}\hat{\mu}_L(t)$. (The condition (C3*) ($ph \to 0$ as $n \to \infty$) of this paper does not produce a well-defined estimator for large n.) On the other hand, our normalizing rate $n^{1/2}$ is consistent with the variance rate n^{-1} found in the literature.

3. Simultaneous confidence bands

We build here approximate simultaneous confidence bands for μ at the level $1 - \gamma \in (0, 1)$. First assume that either Theorem 2.1 or 2.2 applies, i.e. that $n^{1/2}(\hat{\mu} - \mu)$ converges weakly in $C([0, T])$, where $\hat{\mu}$ denotes the corresponding estimator (2) or (3). Assume also that the covariance function R is nondegenerate and let $\widehat{R}(t, t)$ be any uniformly consistent estimator of $R(t, t)$ with respect to $t \in [0, T]$. With Slutsky, one sees that $n^{1/2}(\hat{\mu}_{np} - \mu)/\widehat{R}(t, t)^{1/2}$ converges to $Z = \mathcal{G}(0, \rho)$ in $C([0, T])$, where ρ is the correlation function of X. It suffices then to apply a classical result of [10] to get that

$$\lim_{\lambda \to \infty} \lambda^{-2} \log \mathbb{P}\Big\{ \sup_{t \in [0, T]} Z(t) > \lambda \Big\} = -\Big(2 \sup_{t \in [0, T]} \rho(t, t)\Big)^{-1} = -\frac{1}{2}. \tag{4}$$

Finally, use (4) along with the inequality $\mathbb{P}\{\sup_{t \in [0, T]} |Z(t)| > \lambda\} \leqslant 2\mathbb{P}\{\sup_{t \in [0, T]} Z(t) > \lambda\}$ to derive the following approximate simultaneous confidence bands for μ:

$$\hat{\mu}(t) \pm \big(-2 \log(\gamma/2)\widehat{R}(t, t)/n\big)^{1/2} \quad (0 \leqslant t \leqslant T). \tag{5}$$

4. Discussion

The asymptotic normality results presented in this paper provide a new tool for making simultaneous inference on a trend function μ in the context of functional data. They can plausibly be extended to the framework of multivariate and/or vector-valued random processes and to the inference of derivatives of μ. To the best of our knowledge, the

simultaneous confidence band procedure of Section 3 is the only one appearing in the literature for functional data. Its implementation only requires the estimation of the mean and of the variance of X (not the whole covariance R) and some simulations have indicated its good performances in terms of empirical coverage probability. It would benefit from additional features such as bias correction or data-driven bandwidth selection. The asymptotic normality results may also be applied to constructing tests for μ. A linearity test based on these results is currently under study.

References

[1] A. Azzalini, A.W. Bowman, W. Härdle, On the use of nonparametric regression for model checking, Biometrika 76 (1989) 1–11.

[2] K. Benhenni, M. Rachdi, Nonparametric estimation of the average growth curve with a general nonstationary error process, Comm. Statist. Theory Methods 36 (2007) 1173–1186.

[3] H. Cardot, Nonparametric estimation of smoothed principal components analysis of sampled noisy functions, J. Nonparametr. Statist. 12 (2000) 503–538.

[4] R.M. Clark, Non-parametric estimation of a smooth regression function, J. Roy. Statist. Soc. B 39 (1977) 107–113.

[5] D. Degras, Asymptotics for the nonparametric estimation of the mean function of a random process, Statist. Probab. Lett. 78 (2008) 2976–2980.

[6] D. Degras, R. Jallet, Convergence de l'estimateur spline cubique de lissage dans un modèle de régression longitudinale avec erreur de type processus, C. R. Math. Acad. Sci. Paris, Ser. I 340 (2005) 851–854.

[7] R.L. Eubank, P.L. Speckman, Confidence bands in nonparametric regression, J. Amer. Statist. Assoc. 88 (1993) 1287–1301.

[8] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis, Springer, New York, 2006.

[9] J.D. Hart, T.E. Wehrly, Kernel regression estimation using repeated measurements data, J. Amer. Statist. Assoc. 81 (1986) 1080–1088.

[10] H. Landau, L.A. Shepp, On the supremum of a Gaussian process, Sankhyā 32 (1970) 369–378.

[11] G. Knafl, C. Spiegelman, J. Sacks, D. Ylvisaker, Nonparametric calibration, Technometrics 26 (1984) 233–241.

[12] D. Pollard, Empirical Processes: Theory and Applications, Regional Conference Series in Probability and Statistics, vol. 2, Institute of Mathematical Statistics, Hayward, CA, 1990.

[13] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, second ed., Springer, New York, 2005.

[14] Y. Xia, Bias-corrected confidence bands in nonparametric regression, J. Roy. Statist. Soc. B 60 (1998) 797–811.

[15] F. Yao, Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data, J. Multivariate Anal. 98 (2007) 40–56.