Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
C. R. Acad. Sci. Paris, Ser. I 343 (2006) 551–554
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistique

Estimation de l'espérance structurelle d'une fonction aléatoire

Elie Maza

Laboratoire de statistique et probabilités, UMR C5583, université Paul-Sabatier, 118, route de Narbonne, 31062 Toulouse cedex 4, France

Reçu le 7 février 2005 ; accepté après révision le 26 septembre 2006

Disponible sur Internet le 18 octobre 2006

Présenté par Paul Deheuvels

Résumé

Nous présentons un modèle de déformation de courbe dans lequel la déformation est supposée être un processus aléatoire inconnu. Cette hypothèse nous permet de définir, de manière unique et intrinsèque, l'espérance structurelle de la fonction aléatoire sous-jacente. Nous proposons des estimateurs empiriques de cette espérance structurelle et des fonctions individuelles de déformation. Nous donnons des résultats asymptotiques, en nombre de courbes et de points de mesure. ***Pour citer cet article : E. Maza, C. R. Acad. Sci. Paris, Ser. I 343 (2006).***

Abstract

Estimation of the structural expectation of a stochastic function. We present a curve warping model in which the warping function is supposed to be an unknown stochastic process. This assumption allows us to define, in an unique and intrinsic way, the structural expectation of the underlying stochastic function. We propose empirical estimators of this structural expectation and of the warping functions. We give some asymptotic results, in number of curves and of measured points. ***To cite this article: E. Maza, C. R. Acad. Sci. Paris, Ser. I 343 (2006).***

1. Modèle de déformation et espérance structurelle

Lorsque la caractéristique mesurée sur un échantillon statistique d'individus est une fonction, par exemple une fonction du temps, deux sortes d'aléas peuvent apparaître : un aléa en amplitude (sur les ordonnées) et un aléa en phase (sur les abscisses). Dans ce cas, nous observons des courbes dont la moyenne est un mauvais représentant de la population sous-jacente, puisque les deux effets (en amplitude et en phase) ne sont pas dissociés. De nombreuses méthodes ont été proposées ces dernières années pour pallier à ce problème, par exemple dans [8,6,4,3] ou [2]. Dans ces travaux, les fonctions de déformation sont considérées comme des paramètres inconnus et à estimer, pour ensuite estimer (par déformation inverse) les fonctions originelles. Ainsi, l'asymptotique en nombre de courbes des estimateurs proposés n'est pas étudiée. Par contre, dans [7], les déformations sont des translations supposées aléatoires, et bien que le modèle soit restreint aux translations, l'asymptotique en nombre de courbes y est étudiée. Aussi, les modèles proposés dans ces travaux ne sont pas identifiables (voir, par exemple, [3]).

Adresse e-mail : Elie.Maza@math.ups-tlse.fr (E. Maza).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2006.09.022

Pour résoudre ces deux problématiques, nous allons supposer ici que les fonctions de déformation sont des réalisations indépendantes d'un processus stochastique H appelé « processus aléatoire de déformation ». Ce processus H est défini sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbf{P})$ inconnu, et à valeurs dans $\mathcal{C}([a,b])$, l'ensemble des fonctions réelles continues sur l'intervalle $[a,b]$, muni de la norme de la convergence uniforme $\|\cdot\|_\infty$ et de la tribu de ses boréliens $\mathcal{B}(\mathcal{C}([a,b]))$. De plus, nous supposons que H est presque sûrement strictement croissant, et que $H(a) \overset{ps}{=} a$ et $H(b) \overset{ps}{=} b$. Sous ces hypothèses, l'espérance $\phi = \mathbb{E}(H)$ et le moment d'ordre deux $\gamma = \mathbb{E}(H^2)$ du processus H existent. Nous notons la covariance de H : $r(t,s) = \mathrm{Cov}(H(t), H(s))$, pour tout $(t,s) \in [a,b]^2$.

Notre modèle de déformation est défini par

$$Y_{ij} = f \circ H_i^{-1}(t_j), \quad i = 1, \ldots, m, \ j = 0, \ldots, n, \ \text{où} \tag{1}$$

- la fonction $f : [a,b] \to \mathcal{I} \subset \mathbb{R}$ est une fonction réelle continue inconnue,
- pour tout $j \in \{0, \ldots, n\}$, $t_j = a + j(b-a)/n$,
- les processus $(H_i)_i$ sont i.i.d. et de même loi que le processus aléatoire de déformation inconnu H.

Afin d'alléger les notations, pour tout $i \in \{1, \ldots, m\}$, nous posons $F_i = f \circ H_i^{-1}$ la fonction aléatoire et $f_i = f \circ h_i^{-1}$ une réalisation de F_i. De plus, nous notons A l'ensemble des trajectoires $f \circ h^{-1} \in \mathcal{C}([a,b])$ où h est une fonction strictement croissante, avec $h(a) = a$ et $h(b) = b$.

Définition 1.1. Sous les hypothèses du modèle (1), la fonction $f \circ (\mathbb{E}(H))^{-1} = f \circ \phi^{-1}$ est appelée « espérance structurelle » de la fonction aléatoire $f \circ H^{-1}$.

Cette espérance structurelle s'interprète comme la déformation moyenne de la fonction f par le processus aléatoire H. Ainsi, en introduisant dans notre modèle (1) l'espérance ϕ du processus H :

$$Y_{ij} = \underbrace{f \circ \phi^{-1}}_{\text{Espérance structurelle}} \circ \underbrace{\phi \circ H_i^{-1}(t_j)}_{\text{Fonction de déformation}}, \quad i = 1, \ldots, m, \ j = 0, \ldots, n,$$

le couple $(f \circ \phi^{-1}, \phi \circ H_i^{-1})$ est unique, et notre problème d'indentifiabilité est résolu. En effet, l'on ne s'intéresse plus à l'estimation du couple (f, H_i^{-1}) qui lui n'est pas unique.

2. Estimations dans le cas où la fonction f est strictement croissante

Lorsque la fonction $f : [a,b] \to \mathcal{I}$ est strictement croissante, la fonction réciproque $f^{-1} : \mathcal{I} \to [a,b]$ existe et est également strictement croissante. De plus, une déformation en phase de la fonction f correspond à une déformation en amplitude de la fonction f^{-1}. Il est donc naturel, pour estimer l'inverse de l'espérance structurelle, de considérer la moyenne de l'ensemble des fonctions $(f_i^{-1})_i$. Ainsi, nous définissons l'estimateur empirique de l'inverse de l'espérance structurelle par

$$\forall y \in \mathcal{I}, \quad \widehat{\phi \circ f^{-1}}_{mn}(y) = \frac{1}{m} \sum_{i=1}^{m} T_i,$$

où

$$\forall i = 1, \ldots, m, \ T_i = \arg \min_{t_j \in \{t_0, \ldots, t_n\}} \left| F_i(t_j) - y \right|.$$

La notion de convergence d'une suite quelconque U_{mn} indicée par $(m,n) \in \mathbb{N}^2$ vers C (notion qui est utilisée dans les résultats asymptotiques donnés ci-après) est telle que, pour m et n assez grands, la quantité $|U_{mn} - C|$ peut être rendue aussi petite que souhaité.

Théorème 2.1 *(Consistance et normalité asymptotique). Sous les hypothèses du modèle* (1)*, et lorsque la fonction f est strictement croissante, l'estimateur $\widehat{\phi \circ f^{-1}}_{mn}$ converge presque sûrement vers $\phi \circ f^{-1}$:*

$$\left\| \widehat{\phi \circ f^{-1}}_{mn} - \phi \circ f^{-1} \right\|_\infty \xrightarrow[m,n \to +\infty]{ps} 0.$$

De plus, soit $y = (y_1, \ldots, y_k)' \in \mathcal{I}^k$ *fixé* ($k \in \mathbb{N}^*$)*, et supposons que* $\sqrt{m}/n \xrightarrow[m,n\to+\infty]{} 0$*, alors,*

$$\sqrt{m}\left(\begin{pmatrix} \widehat{\phi \circ f^{-1}}_{mn}(y_1) \\ \vdots \\ \widehat{\phi \circ f^{-1}}_{mn}(y_k) \end{pmatrix} - \begin{pmatrix} \phi \circ f^{-1}(y_1) \\ \vdots \\ \phi \circ f^{-1}(y_k) \end{pmatrix}\right) \xrightarrow[m,n\to+\infty]{\mathcal{L}} \mathcal{N}_k\big(0, \Gamma(y)\big),$$

où $\Gamma(y)$ *est la matrice de variance-covariance de taille* k *définie, pour tout couple* $(i,j) \in \{1, \ldots, k\}^2$*, par*

$$\Gamma(y)_{ij} = r\big(f^{-1}(y_i), f^{-1}(y_j)\big).$$

Par construction, la fonction $\widehat{\phi \circ f^{-1}}_{mn}$ est une fonction constante par morceaux. Soit $K(m,n)$ le nombre de sauts en les points $(v_k)_k$ tels que $f(a) = v_0 < v_1 < \cdots < v_{K(m,n)} < v_{K(m,n)+1} = f(b)$. Alors, nous pouvons écrire

$$\forall y \in \big]f(a), f(b)\big[\setminus (v_k)_k, \quad \widehat{\phi \circ f^{-1}}_{mn}(y) = \sum_{k=0}^{K(m,n)} u_k \mathbb{1}_{]v_k, v_{k+1}[}(y),$$

où les valeurs $(u_k)_k$ sont telles que $a = u_0 < u_1 < \cdots < u_{K(m,n)-1} < u_{K(m,n)} = b$. Ainsi, il est naturel de définir un estimateur de l'espérance structurelle $f \circ \phi^{-1}$ par

$$\forall t \in [a,b], \quad \widehat{f \circ \phi^{-1}}_{mn}(t) = \sum_{k=0}^{K(m,n)-1} \left(\frac{v_{k+1} - v_k}{u_{k+1} - u_k}(t - u_k) + v_k\right) \mathbb{1}_{[u_k, u_{k+1}[}(t) + v_{K(m,n)} \mathbb{1}_{\{b\}}(t).$$

Par construction, notre estimateur $\widehat{f \circ \phi^{-1}}_{mn}$ est continu et strictement croissant.

Théorème 2.2 *(Consistance). Sous les hypothèses du modèle* (1)*, et lorsque la fonction* f *est strictement croissante, l'estimateur* $\widehat{f \circ \phi^{-1}}_{mn}$ *converge presque sûrement vers l'espérance structurelle* $f \circ \phi^{-1}$ *:*

$$\big\|\widehat{f \circ \phi^{-1}}_{mn} - f \circ \phi^{-1}\big\|_\infty \xrightarrow[m,n\to+\infty]{ps} 0.$$

Remarque 1. De manière analogue à l'estimation de l'espérance structurelle, nous construisons des estimateurs des fonctions $(\phi \circ h_i^{-1})_i$, et en montrons la consistance et la normalité asymptotique (voir [5]).

3. Cas où la fonction f est strictement monotone par morceaux

Dans cette section, nous étendons nos estimateurs à l'ensemble $\mathcal{F}$ des fonctions strictement monotones par morceaux. L'idée est de transformer une fonction $f \in \mathcal{F}$ en une fonction strictement croissante. Soit $f \in \mathcal{F}$ et $(s_k)_k \in]a,b[^r$ l'ensemble des $r \in \mathbb{N}^*$ points de changement de sens de variation de f. Notons $s_0 = a$ et $s_{r+1} = b$. Soit $\pi :]a,b[\setminus \{s_1, \ldots, s_r\} \to \{-1, 1\}$ la fonction définie par $\pi(t) = 1$ si f est strictement croissante en t et $\pi(t) = -1$ si f est strictement décroissante en t.

Définition 3.1. Soit $f \in \mathcal{F}$. La transformation $\mathcal{G}(\cdot, f)$ est définie, pour tout $t \in]a,b[\setminus \{s_0, \ldots, s_{r+1}\}$, par

$$\mathcal{G}(t,f) = f(t)\pi(t) - \sum_{k=0}^{r} \pi(t) f(s_k) \mathbb{1}_{]s_k, s_{k+1}[}(t) + f(s_0) + \sum_{k=1}^{r} \big|f(s_{k-1}) - f(s_k)\big| \mathbb{1}_{]s_k, b[}(t),$$

et par continuité pour tout $s_k \in \{s_0, \ldots, s_{r+1}\}$.

Proposition 3.2. *Soit la fonction* $f \in \mathcal{F}$ *et* $g = \mathcal{G}(\cdot, f)$ *sa transformée. Soient les trajectoires* $(f_i)_i \in A^m$ *et* $(g_i = \mathcal{G}(\cdot, f_i))_i$ *leurs transformées. Alors, les fonctions* g *et* $(g_i)_i$ *sont strictement croissantes et telles que*

$$g_i = g \circ h_i^{-1}, \quad i = 1, \ldots, m.$$

La transformation $\mathcal{G}$ effectuée sur chacune des fonctions $(f_i)_i$ nous permet donc de remplacer le modèle (1) par le modèle

$$Z_{ij} = g \circ H_i^{-1}(t_j), \quad i = 1, \ldots, m, \ j = 0, \ldots, n.$$

Ainsi, les estimateurs construits dans la Section 2 peuvent être utilisés, en remplaçant les variables $(Y_{ij})_{ij}$ par les variables $(Z_{ij})_{ij}$.

Le problème qui se pose ici est que nous ne connaissons pas les points de changement de sens de variation des fonctions $(f_i)_i$. Nous ne pouvons donc pas utiliser directement la définition de $\mathcal{G}$ pour calculer les valeurs des variables $(Z_{ij})_{ij}$. Pour résoudre ce problème, d'après la structure de la transformation $\mathcal{G}$, il paraît naturel d'approcher les valeurs $(Z_{ij})_{ij}$ par les accroissements successifs des valeurs $(Y_{ij})_{ij}$. Nous définissons, pour tout $i \in \{1, \ldots, m\}$ fixé, la suite de variables aléatoires

$$\tilde{Z}_{i0} = Y_{i0} \quad \text{et pour tout } j \in \{1, \ldots, n\}, \quad \tilde{Z}_{ij} = \tilde{Z}_{ij-1} + |Y_{ij} - Y_{ij-1}|.$$

Proposition 3.3. *Soit $f_i \in \mathcal{F}$, $i \in \{1, \ldots, m\}$ fixé. Soit $t \in [a, b]$ fixé et $(j(n))_{n \in \mathbb{N}}$ une suite telle que $j(n)/n \xrightarrow[n \to +\infty]{} t$. Alors,*

$$\tilde{Z}_{ij(n)} - Z_{ij(n)} \xrightarrow[n \to +\infty]{ps} 0.$$

Cette dernière proposition nous permet d'utiliser les estimateurs de la Section 2 dans le cas d'une fonction f strictement monotone par morceaux.

Remarque 2. Pour une application à des données réelles, on peut se référé à [1].

Remerciements

Je remercie mon directeur de thèse, Fabrice Gamboa, pour l'intérêt qu'il a accordé aux développements mathématiques des idées présentées ici.

Références

[1] J.-F. Dupuy, J.-M. Loubes, E. Maza, Non parametric estimation of the structural expectation of a stochastic increasing function, soumis, 2006.
[2] F. Gamboa, J.-M. Loubes, E. Maza, Semi-parametric estimation of shifts, soumis, 2005.
[3] D. Gervini, T. Gasser, Self-modelling warping functions, J. R. Stat. Soc. Ser. B Stat. Methodol. 66 (4) (2004) 959–971.
[4] A. Kneip, X. Li, K.B. MacGibbon, J.O. Ramsay, Curve registration by local regression, Canad. J. Statist. 28 (1) (2000) 19–29.
[5] E. Maza, Prévision de trafic routier par des méthodes statistiques – Espérance structurelle d'une fonction aléatoire, Thèse du Laboratoire de Statistique et Probabilités, Université Toulouse III, 2004.
[6] J.O. Ramsay, B.W. Silvermann, Applied Functional Data Analysis, Methods and Case Studies, Springer Series in Statistics, Springer-Verlag, New York, 2002.
[7] B.B. Rønn, Nonparametric maximum likelihood estimation for shifted curves, J. R. Stat. Soc. Ser. B Stat. Methodol. 63 (2) (2001) 243–259.
[8] K. Wang, T. Gasser, Synchronizing sample curves nonparametrically, Ann. Statist. 27 (2) (1999) 439–460.