Identification d'un Système Non-Linéaire Partiellement Observé par la Méthode de la Distance Minimale

Yurii KUTOYANTS
Département de Mathématiques
Université du Maine, BP 535
72017 Le Mans Cedex, France.

François LE GLAND
IRISA / INRIA
Campus universitaire de Beaulieu
35042 Rennes Cedex, France.

Rivo RAKOTOZAFY
Département de Mathématiques
Université de Fianarantsoa, BP 1264
301 Fianarantsoa, Madagascar.

Résumé : Nous considérons le problème de l'estimation de paramètres par la méthode de la distance minimale, pour les processus de type de diffusion partiellement observés. Nous construisons l'estimateur de la distance minimale (EDM) à partir de n observations indépendantes, et nous montrons la consistance et la normalité asymptotique de l'EDM quand n tend vers l'infini.

Mots-clé : processus de type diffusion, observation partielle, estimateur paramétrique, estimateur de la distance minimale.

Identification of a Partially Observed Nonlinear System by the Minimum Distance Method

Abstract : We consider the problem of parameter estimation by the minimum distance method, for partially observed diffusion type processes. We consider the minimum distance estimate (MDE) based on n independent observations, and we prove consistency and asymptotic normality for the MDE when n goes to infinity.

Keywords : diffusion type process, partially observation, parameter estimation, minimum distance estimate.

1 Introduction

Les processus stochastiques sont bien adaptés pour formaliser l'évolution d'un système dynamique lorsque cette évolution subit une perturbation aléatoire. Formellement, un processus stochastique est une famille $X = \{X_t\,,\, t \geq 0\}$ de variables aléatoires définies sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, à valeurs dans un espace mesurable $(S, \mathcal{S})$ appelé *espace d'état*. On supposera dans la suite que la variable X_t décrit l'état d'un système dynamique ou d'un phénomène physique, et vérifie l'équation différentielle stochastique suivante :

$$\frac{d}{dt} X_t = b(X_t) + \sigma(X_t)\, \xi_t \ , \tag{1}$$

où ξ_t correspond à un bruit de dynamique et représente soit une perturbation aléatoire due à la physique du problème, soit les incertitudes de modélisation. En pratique, on ne dispose souvent que d'une observation partielle et bruitée z_t de l'état du système X_t, décrite par :

$$z_t = h(X_t) + \varepsilon_t \ , \tag{2}$$

où ε_t est le bruit de mesure.

Dans la suite, nous considérons $X = \{X_t\,,\, 0 \le t \le T\}$ et $Y = \{Y_t\,,\, 0 \le t \le T\}$ deux processus à valeurs dans $\mathbb{R}^m$ et $\mathbb{R}^d$ respectivement. Pour chaque valeur $\theta \in \Theta \subset \mathbb{R}^p$, nous considérons la probabilité $\mathbb{P}_\theta$ sur l'espace canonique $C([0,T]\,;\,\mathbb{R}^{m+d})$ telle que, sous $\mathbb{P}_\theta$:

$$\left\{ \begin{array}{lllll} dX_t &=& b_t(\theta,X)\,dt + \sigma_t(\theta,X)\,dW_t^\theta \ , & X_0 &=& x_0 \\[2ex] dY_t &=& h_t(\theta,X)\,dt + dV_t^\theta \ , & Y_0 &=& 0 \ , \end{array} \right. \tag{3}$$

où $\theta \in \Theta$ est le paramètre inconnu. Nous supposons aussi que sous $\mathbb{P}_\theta$, $\{W_t^\theta\,,\, 0 \le t \le T\}$ et $\{V_t^\theta\,,\, 0 \le t \le T\}$ sont des processus de Wiener standard indépendants, à valeurs dans $\mathbb{R}^m$ et $\mathbb{R}^d$ respectivement.

Nous faisons l'hypothèse suivante sur les coefficients du système (3) :

(I) *Pour tout $\theta \in \Theta$, les fonctions $(t,x) \mapsto b_t(\theta,x)$, $(t,x) \mapsto \sigma_t(\theta,x)$ et $(t,x) \mapsto h_t(\theta,x)$ sont nonanticipatives, à valeurs dans $\mathbb{R}^m$, $\mathbb{R}^{m\times m}$ et $\mathbb{R}^d$ respectivement. En outre, ces fonctions sont lipschitziennes et à croissance au plus linéaire : pour tout $0 \le t \le T$ et tout $x, x' \in C([0,T]\,;\,\mathbb{R}^m)$*

$$|b_t(\theta,x) - b_t(\theta,x')| + |\sigma_t(\theta,x) - \sigma_t(\theta,x')| + |h_t(\theta,x) - h_t(\theta,x')| \le L \sup_{0\le s\le t} |x_s - x'_s| \ ,$$

$$|b_t(\theta,x)| + |\sigma_t(\theta,x)| + |h_t(\theta,x)| \le L\ (1 + \sup_{0\le s\le t} |x_s|\,) \ ,$$

et la constante $L > 0$ est indépendante de $\theta \in \Theta$.

Sous l'hypothèse **(I)**, le système (3) admet une solution unique, voir Liptser–Shiryayev [4], et on peut vérifier que cette solution possède des moments de tout ordre.

Nous nous intéressons à l'estimation du paramètre θ dans le modèle décrit ci–dessus, par la méthode de la distance minimale. Pour l'étude asymptotique d'autres estimateurs (estimateur du maximum de vraisemblance, estimateur bayésien) dans les processus stochastiques, on consultera par exemple Basawa–Prakasa Rao [1], Ibragimov–Khasminskii [2] et Kutoyants [3]. Nous considérons ici une asymptotique de type *grand nombre d'échantillons indépendants*.

A partir de n observations indépendantes $\{Y^j\ ,\ 1 \leq j \leq n\}$, nous définissons le processus moyenne empirique

$$\widehat{Y}^n_t = \frac{1}{n} \sum_{j=1}^{n} Y^j_t\ ,$$

et nous définissons la fonction $\widehat{h}^n_t(\theta)$ par :

$$\widehat{h}^n_t(\theta) = \frac{1}{n} \sum_{j=1}^{n} h_t(\theta, X^j)\ , \tag{4}$$

où $\{X^j\ ,\ 1 \leq j \leq n\}$ sont n copies indépendantes de la solution de l'équation :

$$dX_t = b_t(\theta, X)\, dt + \sigma_t(\theta, X)\, dW^\theta_t\ , \quad X_0 = x_0\ .$$

Sous la probabilité $\mathbb{P}_\theta$, le processus moyenne empirique vérifie l'équation d'Itô suivante :

$$d\widehat{Y}^n_t = \widehat{h}^n_t(\theta)\, dt + \frac{1}{\sqrt{n}}\, d\widehat{V}^{n,\theta}_t\ , \quad \widehat{Y}^n_0 = 0\ , \tag{5}$$

où pour tout $n \geq 1$, $\widehat{V}^{n,\theta} = \{\widehat{V}^{n,\theta}_t\ ,\ 0 \leq t \leq T\}$ est un processus de Wiener standard, et la fonction $\widehat{h}^n_t(\theta)$ est définie par (4).

D'après la loi forte des grands nombres, nous avons pour tout $0 \leq t \leq T$ la convergence suivante :

$$\widehat{h}^n_t(\theta) \longrightarrow \bar{h}_t(\theta) = \mathbb{E}_\theta[h_t(\theta, X)]\ , \quad \mathbb{P}_\theta\text{–p.s.}$$

quand $n \to \infty$. Notons par $y(\theta)$ la solution de l'équation déterministe limite de l'équation (5) quand n tend vers l'infini. Nous avons :

$$y_t(\theta) = \int_0^t \bar{h}_s(\theta)\, ds\ . \tag{6}$$

L'estimateur de la distance minimale pour le paramètre inconnu θ dans le modèle statistique ci–dessus, est fondé sur la remarque suivante : quand la taille n de l'échantillon est grand, le processus moyenne empirique $\widehat{Y}^n$ est proche de $y(\theta_0)$, où θ_0 désigne la *vraie* valeur du paramètre. Nous définissons donc l'estimateur de la distance minimale comme l'ensemble des points $\theta \in \Theta$ tels que $\widehat{Y}^n$ soit proche de $y(\theta)$. Pour mesurer cette distance, nous considérons une mesure positive bornée μ sur $[0, T]$, et nous notons $\|\cdot\|$ et $\langle\cdot,\cdot\rangle$ la norme et le produit scalaire dans $L^2_\mu([0, T])$.

Définition 1.1 *L'estimateur de distance minimale (EDM) est défini par :*

$$\theta^*_n \in \operatorname*{argmin}_{\theta \in \Theta} \|\widehat{Y}^n - y(\theta)\|\ , \tag{7}$$

où $\widehat{Y}^n$ et $y(\theta)$ sont les solutions respectives des équations (5) et (6).

2 Consistance de l'EDM

Nous faisons l'hypothèse d'*identifiabilité* suivante :

(II) *Pour tout* $\nu > 0$

$$g(\nu) \stackrel{\triangle}{=} \inf_{|\theta-\theta_0|>\nu} \|y(\theta) - y(\theta_0)\| > 0 \ ,$$

où θ_0 *désigne la vraie valeur du paramètre inconnu.*

Nous obtenons une vitesse de convergence polynomiale de tout ordre, pour l'estimateur de distance minimale.

Théorème 2.1 *Sous les hypothèses* **(I)** *et* **(II)**, *pour tout* $r \geq 1$ *il existe une constante* $C_r > 0$ *telle que*

$$\mathbb{P}_{\theta_0}(|\theta_n^* - \theta_0| > \nu) \leq \frac{C_r}{n^r \, g^{2r}(\nu)} \ .$$

La démonstration de ce théorème repose sur les deux résultats suivants.

Lemme 2.2 *Pour tout* $\nu > 0$

$$\{|\theta_n^* - \theta_0| > \nu\} \subset \{ \, \|\widehat{Y}^n - y(\theta_0)\| > \tfrac{1}{2} \, g(\nu) \, \} \ .$$

Considérons le processus $\xi^n(\theta) = \{\xi_t^n(\theta) \, , \, 0 \leq t \leq T\}$, où $\xi_t^n(\theta) = \sqrt{n} \, [\widehat{Y}_t^n - y_t(\theta)]$, et où $\widehat{Y}^n$ et $y(\theta)$ sont les solutions respectives des équations (5) et (6). Ce processus vérifie l'équation d'Itô suivante :

$$d\xi_t^n(\theta) = \sqrt{n} \, [\widehat{h}_t^n(\theta) - \bar{h}_t(\theta)] \, dt + d\widehat{V}_t^n \ . \tag{8}$$

où $\{\widehat{V}_t^n \, , \, 0 \leq t \leq T\}$ est un processus de Wiener standard.

Lemme 2.3 *Sous l'hypothèse* **(I)**, *le processus* $\xi^n(\theta_0) = \{\xi_t^n(\theta_0) \, , \, 0 \leq t \leq T\}$, *solution de l'équation (8) possède sous* $\mathbb{P}_{\theta_0}$ *des moments de tout ordre, bornés indépendamment de* n.

Les preuves de ces deux Lemmes peuvent être consultées dans la thèse de Rivo Rakotozafy [6, pages I-10 et I-11].

Preuve du Théorème 2.1. Le Lemme 2.2 ci–dessus nous permet d'établir l'inégalité :

$$\mathbb{P}_{\theta_0}(|\theta_n^* - \theta_0| > \nu) \ \leq \ \frac{4^r}{n^r \, g^{2r}(\nu)} \mathbb{E}_{\theta_0} \|\xi^n(\theta_0\|^{2r} \ , \tag{9}$$

d'après l'inégalité de Markov. Ici, $\| \cdot \|$ désigne la norme dans $L^2_\mu([0,T])$, et $\xi^n(\theta_0)$ est la solution de l'équation (8). En utilisant l'inégalité de Jensen, nous avons

$$\|\xi^n(\theta_0)\|^{2r} \ \leq \ \mu_T^{r-1} \int_0^T |\xi_t^n(\theta_0)|^{2r} \, d\mu(t) \ .$$

Nous en déduisons que

$$\mathbb{E}_{\theta_0} \|\xi^n(\theta_0)\|^{2r} \ \leq \ \mu_T^{r-1} \int_0^T \mathbb{E}_{\theta_0} |\xi_t^n(\theta_0)|^{2r} \, d\mu(t) \ \leq \mu_T^r \sup_{0 \leq t \leq T} \mathbb{E}_{\theta_0} |\xi_t^n(\theta_0)|^{2r} \ .$$

D'après le Lemme 2.3, nous pouvons trouver une constante $C_r > 0$ qui ne dépend pas de n, telle que :

$$\mathbb{E}_{\theta_0} \|\xi^n(\theta_0)\|^{2r} \leq C_r \ . \qquad \square$$

3 Normalité asymptotique de la moyenne empirique

Considérons le processus $\xi^n(\theta) = \{\xi^n_t(\theta)\,,\, 0 \le t \le T\}$ solution de l'équation (8). Nous avons le résultat de convergence faible suivant, qui sera utile pour démontrer la normalité asymptotique de l'EDM.

Proposition 3.1 *Sous l'hypothèse* **(I)** *:*

$$\xi^n(\theta_0) \Rightarrow \xi(\theta_0) \ ,$$

dans $C([0,T]\,;\mathbb{R}^d)$, quand $n \to \infty$, où $\xi(\theta_0) = \{\xi_t(\theta_0)\,,\, 0 \le t \le T\}$ est un processus gaussien centré.

Le processus gaussien $\xi(\theta_0)$ a pour fonction de corrélation $\Xi(\theta_0) = \{\Xi_{s,t}(\theta_0)\,, 0 \le s,t \le T\}$, définie par

$$\Xi_{s,t}(\theta_0) = \mathbb{E}_{\theta_0}[\xi_s(\theta_0)\xi^*_t(\theta_0)] = \int_0^t \int_0^s \mathbb{E}_{\theta_0}[\zeta_u(\theta_0)\zeta^*_v(\theta_0)]\, du\, dv + (t \wedge s) \ , \qquad (10)$$

avec $\zeta_t(\theta_0) = h_t(\theta_0, X) - \bar{h}_t(\theta_0)$.

La preuve de cette Proposition peut être consultée dans la thèse de Rivo Rakotozafy [6, pages I-13 à I-16].

4 Normalité asymptotique de l'EDM

Nous faisons les hypothèses de *différentiabilité* et de *non–dégénérescence* suivantes :

(III) *La fonction $y(\theta)$ est différentiable par rapport à θ au point θ_0, c'est–à–dire qu'il existe une fonction $\dot{y}(\theta_0) \in L^2_\mu([0,T]\,;\mathbb{R}^{d\times p})$ telle que :*

$$\|y(\theta) - y(\theta_0) - \dot{y}(\theta_0)\,(\theta - \theta_0)\| = |\theta - \theta_0|\ \varepsilon(|\theta - \theta_0|) \ , \qquad (11)$$

où $\varepsilon(r) \to 0$ quand $r \to 0$.

(IV) *La matrice $p \times p$ définie par :*

$$J(\theta_0) = \int_0^T \dot{y}^*_t(\theta_0)\ \dot{y}_t(\theta_0)\, d\mu(t) \ ,$$

est non–singulière, c'est–à–dire qu'il existe une constante $c_0 > 0$ telle que : pour tout $u \in \mathbb{R}^p$

$$\|\dot{y}(\theta_0)\ u\| = [u^*\ J(\theta_0)\ u]^{1/2} \ge c_0\ |u| \ . \qquad (12)$$

Nous définissons le vecteur aléatoire

$$\widehat{u}(\theta_0) \triangleq J(\theta_0)^{-1}\ m(\theta_0) \ , \qquad \text{où :} \quad m(\theta_0) \triangleq \int_0^T \dot{y}^*_t(\theta_0)\ \xi_t(\theta_0)\, d\mu(t) \ .$$

Le processus gaussien $\xi(\theta_0)$ a pour fonction de corrélation $\Xi(\theta_0) = \{\Xi_{s,t}(\theta_0)\,,\, 0 \le s,t \le T\}$ définie par (10), donc $\widehat{u}(\theta_0)$ est un vecteur aléatoire gaussien centré et de matrice de covariance

$$I(\theta_0) = J(\theta_0)^{-1}\ \Big[\int_0^T \int_0^T \dot{y}^*_t(\theta_0)\ \Xi_{s,t}(\theta_0)\ \dot{y}_s(\theta_0)\, d\mu(t)\, d\mu(s) \ \Big]\ J(\theta_0)^{-1} \ . \qquad (13)$$

Théorème 4.1 *Sous les hypothèses* **(II)**, **(III)** *et* **(IV)**

$$\sqrt{n}\,(\theta_n^* - \theta_0) \Rightarrow \widehat{u}(\theta_0)\ ,$$

quand $n \to \infty$, où $\widehat{u}(\theta_0)$ est un vecteur aléatoire gaussien centré, de matrice de covariance $I(\theta_0)$ définie par (13).

Preuve. Nous commençons par vérifier les hypothèses (2.2), (2.4), (2.5) et (2.6) du Théorème 2.9 de Millar [5, Section X]. L'hypothèse d'identifiabilité (2.2) découle de notre hypothèse **(II)**. L'hypothèse de convergence (2.4) résulte de la Proposition 3.1 ci–dessus. Nos hypothèses **(III)** et **(IV)** entraînent respectivement l'hypothèse de différentiabilité (2.5) et l'hypothèse de non–singularité (2.6).

Nous introduisons alors les notations suivantes. L'hypothèse de différentiabilité **(III)** assure l'existence du vecteur gradient $\dot{y}(\theta_0) = (\dot{y}^1(\theta_0), \ldots, \dot{y}^p(\theta_0))$ avec

$$\dot{y}^i(\theta_0) = \frac{\partial}{\partial \theta^i}\, y(\theta_0)\ , \qquad i = 1, \ldots, p\ .$$

L'hypothèse de non–singularité **(IV)** signifie que les vecteurs $\{\dot{y}^1(\theta_0), \ldots, \dot{y}^p(\theta_0)\}$ sont linéairement indépendants dans $L^2_\mu([0,T];\mathbb{R}^d)$. Nous considérons le sous–espace $B \subset L^2_\mu([0,T];\mathbb{R}^d)$ engendré par ces vecteurs, et nous notons π la projection de $L^2_\mu([0,T];\mathbb{R}^d)$ sur B.

D'après le Théorème 2.9 de Millar [5, Section X], avec une probabilité qui approche 1 quand $n \to \infty$, l'EDM θ_n^* existe et est unique, et nous avons

$$\xi^n(\theta_n^*) - \xi^n(\theta_0) \Rightarrow \pi\,\xi(\theta_0)\ ,$$

dans $L^2_\mu([0,T];\mathbb{R}^d)$, quand $n \to \infty$, et

$$\sqrt{n}(\theta_n^* - \theta_0) \Rightarrow \Upsilon\ ,$$

dans $\mathbb{R}^p$, quand $n \to \infty$, où le vecteur aléatoire Υ est défini par :

$$\pi\,\xi(\theta_0) = \dot{y}(\theta_0)\Upsilon\ .$$

En prenant le produit scalaire dans $L^2_\mu([0,T])$ avec $\dot{y}(\theta_0)$, nous obtenons :

$$\begin{aligned} m(\theta_0) &= \int_0^T \dot{y}_t^*(\theta_0)\xi_t(\theta_0)\, d\mu(t) \\ &= \int_0^T \dot{y}_t^*(\theta_0)\pi\,\xi_t(\theta_0)\, d\mu(t) = \int_0^T \dot{y}_t^*(\theta_0)\dot{y}_t(\theta_0)\, d\mu(t)\ \Upsilon = J(\theta_0)\,\Upsilon\ , \end{aligned}$$

et nous en déduisons :

$$\Upsilon = J^{-1}(\theta_0)\, m(\theta_0) = \widehat{u}(\theta_0)\ . \qquad \square$$

5 Conclusion

L'identification des systèmes non-linéaires partiellement observés par l'estimateur de la distance minimale (EDM) est obtenue pour une asymptotique de type *grand nombre d'observations indépendantes.* L'EDM est facile à mettre en oeuvre, et peut être construit pour un grand nombre de problèmes, y compris dans des cas où la fonction de vraisemblance ne peut pas être définie.

References

[1] Ishwar V. Basawa and B. L. S. Prakasa Rao. *Statistical Inference for Stochastic Processes.* Probability and Mathematical Statistics. Academic Press, London, 1980.

[2] Ildar A. Ibragimov and Rafail Z. Khasminskii. *Statistical Estimation. Asymptotic Theory*, volume 16 of *Applications of Mathematics.* Springer–Verlag, New York, 1981.

[3] Yurii A. Kutoyants. *Parameter Estimation for Stochastic Processes*, volume 6 of *Research and Exposition in Mathematics.* Heldermann Verlag, Berlin, 1984.

[4] Robert Sh. Liptser and Albert N. Shiryayev. *Statistics of Random Processes I. General Theory*, volume 5 of *Applications of Mathematics.* Springer–Verlag, New York, 1977.

[5] P. Warwick Millar. The minimax principle in asymptotic statistical theory. In Paul-Louis Hennequin, editor, *Ecole d'Eté de Probabilités de Saint–Flour XI, 1981*, volume 976 of *Lecture Notes in Mathematics*, pages 75–265. Springer–Verlag, Berlin, 1983.

[6] Rivo Rakotozafy. *Estimateur de la Distance Minimale, Approximation Particulaire pour le Filtrage Non Linéaire.* Thèse de Doctorat, Université de Provence, October 1993.