Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 344 (2007) 727–731
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Theory of Signals/Statistics

On ergodic filters with wrong initial data

Marina L. Kleptsyna [a,b], Alexander Yu. Veretennikov [b,c]

[a] *Université du Maine, avenue Olivier-Messiaen, 72085 Le Mans cedex 09, France*
[b] *Institute of Information Transmission Problems, Moscow, Russia*
[c] *University of Leeds, Leeds LS2 9JT, UK*

Received 26 September 2006; accepted 30 March 2007

Available online 23 May 2007

Presented by Paul Deheuvels

Abstract

For a class of non-uniformly ergodic partly observable Markov processes, under observations subject to a Wiener process or i.i.d. noise, it is shown that a wrong initial data is forgotten with a certain rate. ***To cite this article: M.L. Kleptsyna, A.Yu. Veretennikov, C. R. Acad. Sci. Paris, Ser. I 344 (2007).***

Résumé

Sur des filtres ergodiques avec des données initiales erronées. On démontre que pour une classe de processus de Markov non-uniformément ergodiques avec des observations perturbées par un mouvement Brownien, le fait d'avoir des données initiales erronées est asymptotiquement oublié. La vitesse de convergence est explicitée. ***Pour citer cet article : M.L. Kleptsyna, A.Yu. Veretennikov, C. R. Acad. Sci. Paris, Ser. I 344 (2007).***

Version française abrégée

On considère le problème de la stabilité du filtre optimal par rapport à sa condition initiale pour un système, aussi bien en temps continu (voir (2), (3)) qu'en temps discret (4), (5). Soit $b:\mathbb{R}^d\to\mathbb{R}^d$; $h:\mathbb{R}^d\to\mathbb{R}^\ell$, $(W,B)_{t\geqslant 0}$ un mouvement Brownien de dimension $(d+\ell)$-et en temps discret (W_n,B_n) – une suite i.i.d. centrée, X_0 et (W,B) indépendantes. On suppose que la loi exacte μ_0 du signal X_0 est inconnue et on utilise son approximation ν_0. On étudie le comportement asymptotique (lorsque $t\to\infty$) de la différence $\|\mathbf{P}_t^{\mu_0,Y}(\cdot)-\mathbf{P}_t^{\nu_0,Y}(\cdot)\|_{TV}$, où $\mathbf{P}_t^{\mu_0,Y}(\cdot)=P_{\mu_0}(X_t\in\cdot\mid F_t^Y)$ est la vraie mesure conditionnelle et $\mathbf{P}_t^{\nu_0,Y}(\cdot)=\mathbf{P}_t^{\mu_0,Y}(\cdot)|_{\mu_0=\nu_0}$ son approximation obtenue en remplaçant μ_0 par ν_0.

Les cas du signal ergodique uniformément ou non-uniformément ergodique sous des hypothèses supplémentaires assez fortes, ont été étudié dans plusieurs articles ([2,4,8,1,10], et al.). Dans ce travail on propose des conditions générales et en même temps faciles à verifier, pour avoir,

E-mail addresses: Marina.Kleptsyna@univ-lemans.fr (M.L. Kleptsyna), A.Veretennikov@leeds.ac.uk (A.Yu. Veretennikov).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2007.04.015

$$\lim_{t\to\infty} E_{\mu_0}\big\|\mathbf{P}_t^{\mu_0,Y}(\cdot)-\mathbf{P}_t^{\nu_0,Y}(\cdot)\big\|_{TV}=0.$$

De plus la vitesse de convergence est précisée, voir Éq. (1).

Hypothèses

($A1_p$) La fonction b est bornée et il existe $p\in\{0,1\}$, $M>0$ et $r\in(0,+\infty]$ telles que :

$$\left\langle b(x),\frac{x}{|x|^{1-p}}\right\rangle \leqslant -r,\quad |x|\geqslant M\ ;$$

(dans le cas où $p=1$ on suppose de plus que r est *suffisamment grand*).

(A2) La mesure μ_0 est absolument continue par rapport à ν_0 et $\|\frac{d\mu_0}{d\nu_0}\|_{L_\infty(\nu_0)}\leqslant C<\infty$. De plus elles ont chacune un moment exponentiel fini, $\int e^{c|x|}(\mu_0(dx)+\nu_0(dx))<\infty$.

(A3) *En temps continu*, la fonction $h\in C^2$ et $\|\nabla h\|_{C^1}<\infty$. *En temps discret,* h est bornée localement et la densité de (W_1,B_1) est localement bornée et est localement séparée de zéro.

Théorème 0.1. *Sous les Hypothèses* ($A1_p$)–(A3)*, la majoration suivante pour les systèmes* (2), (3) *et* (4), (5) *est vérifiée* : $\forall m>0\ \exists C_m\ (p=1)$ *et* $\exists C\ \&\ c\ (p=0)$ *telles que* :

$$E_{\mu_0}\big\|\mathbf{P}_t^{\mu_0,Y}(\cdot)-\mathbf{P}_t^{\nu_0,Y}(\cdot)\big\|_{TV}\leqslant\begin{cases}C_m t^{-m}, & p=1,\\ C\exp(-ct), & p=0\end{cases}\tag{1}$$

en temps continu ($t\geqslant 0$) *et en temps discret* ($t=0,1,\ldots$).

La démonstration est basée sur le principe de contraction de Birkhoff et l'estimation des moments du temps d'arrêt [11]. En temps continu le principe de Birkhoff est applicable grâce à l'inégalité de Harnack [7].

1. Introduction

1.1. Continuous time

For $0\leqslant t<\infty$, we consider a Markov diffusion $(X_t,\ t\geqslant 0)$ with values in the Euclidean space $\mathbb{R}^d$ and observations $(Y_t,\ t\geqslant 0)$ from $\mathbb{R}^\ell$, satisfying the following system of non-linear Itô's equations with a $(d+\ell)$-dimensional Wiener process (W_t,B_t),

$$X_t=X_0+\int_0^t b(X_s)\,ds+W_t,\tag{2}$$

$$Y_t=\int_0^t h(X_s)\,ds+B_t,\quad t\geqslant 0,\tag{3}$$

with $b:\mathbb{R}^d\to\mathbb{R}^d$; $h:\mathbb{R}^d\to\mathbb{R}^\ell$ and X_0 is the initial data independent from W and B.

1.2. Discrete time

For $n=0,1,\ldots,$ we consider a homogeneous Markov process X_n, satisfying the following non-linear autoregression equation with an i.i.d. centered sequence (W_n,B_n), not necessarily Gaussian,

$$X_{n+1}=X_n+b(X_n)+W_n,\quad n\geqslant 0,\tag{4}$$

$$Y_n=h(X_n)+B_n,\quad n\geqslant 1,\tag{5}$$

with the initial data X_0 independent from (W,B).

The exact initial distribution of X_0 in both cases is denoted by μ_0. The main question addressed in this Note is about an asymptotical behaviour of the filtering measure when μ_0 is *not known*. Under *uniform ergodicity* assumptions, the limiting independence of the optimal filter algorithm on a wrong initial data was established in [2] and in [8] along with certain bounds. For non-uniformly non-linear ergodic case, certain partial results under assumption of small noise in observations have been established, [1]; in [10] another special class of divergence form diffusions is tackled under assumptions close to large deviations. Let us also mention the paper [3] with a very important counterexample. We are studying a generic non-uniformly ergodic case. For more detailed presentation and references see [5,6].

2. Assumptions

($A1_p$) We assume that the function b is bounded and there exist $p \in \{0,1\}$, $M > 0$ and $r \in (0, +\infty]$ such that

$$\left\langle b(x), \frac{x}{|x|^{1-p}} \right\rangle \leqslant -r, \quad |x| \geqslant M;$$

if $p = 1$ then r is assumed to be *large enough*.

(A2) The measure μ_0 is absolutely continuous with respect to ν_0, and $\|\frac{d\mu_0}{d\nu_0}\|_{L_\infty(\nu_0)} \leqslant C < \infty$, and both initial measures μ_0 and ν_0 possess some exponential moment, $\int e^{c|x|}(\mu_0(dx) + \nu_0(dx)) < \infty$.

(A3) *In continuous time case*, the function h is C^2-smooth and satisfies the condition $\|\nabla h\|_{C^1} < \infty$; *in discrete time case, h* is locally bounded and the densities of (W_1, B_1) are locally bounded and locally separated from zero.

3. Setting and main results

Let us define the following random operators (S_t^Y and $\bar{S}_t^{Y,\mu}$) on the space of measures on the Euclidean space:

$$\mu S_t^Y(A) := \int E_{x_0}^\gamma\big(1(X_t \in A)\rho_{0,t}(X,Y)|Y\big)\mu(dx_0), \qquad \mu\bar{S}_t^{Y,\mu}(A) := d_t^\mu(Y)\mu S_t^Y(A)$$

where $d_t^\mu(Y)$ is a normalization constant and $\gamma = \rho_{0,1}^{-1}$,

in continuous time case $\rho_{s,t}(X,Y) = \exp(\int_s^t h(X_s)\,dY_s - (1/2)\int_s^t h^2(X_s)\,ds),\ s < t,\ h^2 := |h|^2$,

in discrete time case $\rho_{0,t}(X,Y) = \prod_{i=1}^t q_B(Y_i - h(X_i))$, ($q_B$ is the density of B_1).

Then for every $t > 0$, P_{μ_0}-almost surely,

$$\mathbf{P}_t^{\mu_0,Y}(\cdot) = P_{\mu_0}(X_t \in \cdot \mid Y) = \mu_0\bar{S}_t^{Y,\mu_0}(\cdot).$$

Regular conditional measures are to be considered in continuous time case at $t = 0, 1, 2, \ldots$, see [9]. Similarly, we can define the filtering measure with a wrong initial data as $\mathbf{P}_t^{\nu_0,Y}(\cdot) := \nu_0\bar{S}_t^{Y,\nu_0}(\cdot)|_{\tilde{Y}=Y}$, where $(\tilde{X}, \tilde{Y})$ is a solution of the same SDE system (2), (3) with a new initial distribution ν_0 of X_0 on some independent probability space.

Theorem 3.1. *Under the assumptions* ($A1_p$)–(A3)*, the following bounds hold true for the systems* (1–2) *and* (3–4): $\forall m > 0\ \exists C_m\ (p = 1)$ *and* $\exists C\ \&\ c\ (p = 0)$ *such that*:

$$E_{\mu_0}\big\|\mu_0\bar{S}_t^{Y,\mu_0} - \nu_0\bar{S}_t^{Y,\nu_0}\big\|_{TV} \leqslant \begin{cases} C_m t^{-m}, & p = 1, \\ C\exp(-ct), & p = 0. \end{cases}$$

Here $t \geqslant 0$ in continuous time case, and $t = 0, 1, \ldots$ in discrete time one.

For discrete time, Borel–Cantelli Lemmae also imply a point-wise convergence with suitable bounds.

4. The sketch of the proof (for continuous time)

The filtering measure at time t is a multiple integral, and it can be split into a sum of integrals with all variables of integration restricted to some fixed bounded domains (to be taken large) or to the complement of this domain. The terms that have a relatively large number of restrictions of the first kind do have good contraction properties in a *Birkhoff metric*, see [2]. In continuous time this is due to the Harnack inequality, [7]. The remaining terms can be controlled via the mixing bounds for the state process. Let us show some formulae. Denote a (non-random) vector of dimension n with coordinates 1 or 0 at every place by δ, $\#1(\delta)$ the (non-random) number of ones in the vector δ. For fixed R, which can be chosen large enough, we introduce the following events and indicators (here $0^0 = 1$):

$$D_i := \Big\{\max\big(|X_i|, |\tilde X_i|\big) \leqslant R;\ \max\Big(\sup_{i\leqslant s\leqslant i+1} |X_s|, \sup_{i\leqslant s\leqslant i+1} |\tilde X_s|\Big) < R+1\Big\},$$

$$1_\delta(X, \tilde X) := \prod_{i=0}^{n-1} \big(1(D_i)\big)^{\delta_i} \times \big(1 - 1(D_i)\big)^{1-\delta_i}.$$

It is convenient and useful to switch to processes of a *doubled dimension* here, which corresponds to the use of the *coupling method*; however, the latter is not applied directly here: we have,

$$\big\|\mu_0 \bar S_t^Y - \nu_0 \bar S_t^{\tilde Y}|_{\tilde Y=Y}\big\|_{TV} \leqslant 2 \sum_{\delta\in\Delta} e_t^{\delta;\mu_0,\nu_0;Y} \sup_{D\in B(R^{2d})} \big((\mu_0,\nu_0)\hat S_t^{R;\delta;\mu_0,\nu_0;Y}(D) - (\nu_0,\mu_0)\hat S_t^{R;\delta;\mu_0,\nu_0;Y}(D)\big),$$

where Δ is the whole set of possible values of δ,

$$e_t^{\delta;\mu_0,\nu_0;Y} = E_{\mu_0,\nu_0}\big(1_\delta(Z) \mid Y, \tilde Y\big)|_{\tilde Y=Y},$$

$$(\mu,\nu)\hat S_t^{R;\delta;\mu_0,\nu_0;Y}(A\times B) := \big(e_t^{\delta;\mu_0,\nu_0;Y}\big)^{-1} d_t^{\mu_0} d_t^{\nu_0} \times \int E_{x_0,\tilde x_0}^{\gamma}\big(1(X_t\in A, \tilde X_t \in B) 1_\delta(X,\tilde X)\rho_{0,t}(X,Y)\rho_{0,t}(\tilde X, Y) \mid Y\big)\mu(\mathrm{d}x_0)\nu(\mathrm{d}\tilde x_0).$$

To estimate the terms $\sup_D((\mu_0,\nu_0)\hat S_t^{R;\delta;\mu_0,\nu_0;Y}(D) - (\nu_0,\mu_0)\hat S_t^{R;\delta;\mu_0,\nu_0;Y}(D))$, we use the Birkhoff metric for positive measures, cf. [2]. We get,

$$E_{\mu_0}\big\|\mu_0 \bar S_n^Y - \nu_0 \bar S_n^Y\big\|_{TV} \leqslant C\pi_R^{\epsilon n-1} + C E_{\mu_0,\nu_0} \sum_{\delta:\ \#1(\delta)<\epsilon n} e_n^{\delta;\mu_0,\nu_0;Y},$$

where π_R is a contraction constant for the Birkhoff metric (see [8]) on the ball $B_R \subset R^{2d}$. Denote

$$\#1(X)_R := \sum_{k=0}^{n-1} 1\big(|X_k| \leqslant R, \sup_{k\leqslant s\leqslant k+1} |X_s| < R+1\big).$$

The rest of the proof is based on the inequality

$$E_{\mu_0}\left(E_{\mu_0,\nu_0}\left(\sum_{\delta:\ \#1(\delta)<\epsilon n} 1_\delta(X,\tilde X) \mid Y, \tilde Y\right)\Bigg|_{\tilde Y=Y}\right) \leqslant E_{\mu_0} 1\left(\#1(X)_R < \frac{1+\epsilon}{2} n\right) + C E_{\nu_0} 1\left(\#1(\tilde X)_R < \frac{1+\epsilon}{2} n\right),$$

and on the hitting time bound with R large enough from [11],

$$E_{\mu_0} 1\big(\#1(X)_R < \epsilon n\big) \leqslant \begin{cases} C_m n^{-m}, & p=1,\ \forall m>0,\\ C\exp(-cn), & p=0. \end{cases}$$

References

[1] R. Atar, Exponential stability for nonlinear filtering of diffusion processes in a noncompact domain, Ann. Probab. 26 (4) (1998) 1552–1574.

[2] R. Atar, O. Zeitouni, Exponential stability for nonlinear filtering, Ann. Inst. H. Poincaré Probab. Statist. 33 (6) (1997) 697–725.

[3] P. Baxendale, P. Chigansky, R. Liptser, Asymptotic stability of the Wonham filter: ergodic and nonergodic signals, SIAM J. Control Optim. 43 (2) (2004) 643–669 (electronic).

[4] A.S. Budhiraja, D.L. Ocone, Exponential stability in discrete-time filtering for bounded observation noise, Systems Control Lett. 30 (1997) 185–193.

[5] M. Kleptsyna, A. Veretennikov, On discrete time filtres with wrong initial data (revised version), Prépublication no 06-3, de LSP, Univ. du Maine, 2006; http://www.univ-lemans.fr/sciences/statist/download/Kleptsyna/filt19fg.pdf.

[6] M. Kleptsyna, A. Veretennikov, On continuous time filtres with wrong initial data, Prépublication no 06-2 de LSP, Univ. du Maine, 2006; http://www.univ-lemans.fr/sciences/statist/download/Kleptsyna/filt36c.pdf.

[7] N.V. Krylov, M.V. Safonov, A property of the solutions of parabolic equations with measurable coefficients, Izv. Akad. Nauk SSSR Ser. Mat. 44 (1) (1980) 161–175, 239 (in Russian).

[8] F. Le Gland, N. Oudjane, A robustification approach to stability and to uniform particle approximation of nonlinear filters: the example of pseudo-mixing signals, Stochastic Process. Appl. 106 (2003) 279–316.
[9] K.R. Parthasarathy, Probability Measures on Metric Spaces, Academic Press, New York and London, 1967.
[10] W. Stannat, Stability of the pathwise filter equations for a time dependent signal on R^d, Appl. Math. Optim. 52 (2005) 39–71.
[11] A.Yu. Veretennikov, On polynomial mixing and the rate of convergence for stochastic differential and difference equations, Teor. Veroyatn. i Primenen. 44 (2) (1999) 312–327; Engl. transl. in Theory Probab. Appl. 44 (2) (2000) 361–374.