PLD and SDE: quasi-likelihood analysis for stochastic differential equations

Nakahiro YOSHIDA

Graduate School of Mathematical Sciences, University of Tokyo,
3-8-1 Komaba, Meguro-ku, Tokyo 153, Japan.

March 21, 2007, Universite du Maine, Le Mans

Summary In SAPS V in 2005, a polynomial type large deviation (PLD) inequality for a statistical random field was presented and an application to inference for a sampled diffusion process was discussed. This talk is a continuation to the previous one. Based on the PLD, an asymptotic theory of the quasi-likelihood analysis is presented. It enables us to obtain convergence of the statistical random field, asymptotic distributions and convergence of the moments of M (i.e. quasi-MLE type) and Bayesian type estimators over an abstract sequence of statistical experiments with graded parameter spaces. We apply those results to quasi-likelihood analysis for a sampled stochastic differential equation. In particular, quasi-MLE and Bayesian type estimators for a mixing diffusion process are asymptotically normal and admit convergence of moments. We need precise estimates of the tail probability for the estimator to develop theories of higher-order asymptotic statistics, prediction and information criteria, resampling methods, etc, and there PLD is inevitable.

The role of PLD

Ibragimov-Has'minskii-Kutoyants' program

$\downarrow$ PLD

Likelihood analysis for SDE

$$\left\{\begin{array}{l}\circ \text{ Ibragimov-Has'minskii-Kutoyants' program}\\ \circ \text{ Inference for sampled stochastic processes}\end{array}\right.$$

$\downarrow$ PLD (quasi-likelihood random field)

Quasi-likelihood analysis for SDE under sampling

Menu

1. PLD (polynomial type large deviation inequality)
2. Graded parameter spaces
3. M-estimator
4. Bayesian type estimator
5. Applications to diffusion processes
6. Comments on applications to SDE with jumps
7. Comments on PLD for a random field associated with a Z-estimator

Introduction

- $\mathcal{E}^\epsilon = \{\mathcal{X}^\epsilon, \mathcal{A}^\epsilon, P_\theta^\epsilon, \theta \in \Theta\}$: a sequence of statistical experiments with $\epsilon \in (0, 1]$.
- Θ: a parameter space in $\mathbb{R}^m$.
- $\varphi(\epsilon)$: a positive normalizing factor $\to 0$ as $\epsilon \downarrow 0$.
- $\theta_0 \in \Theta$
- a random field

$$Z_\epsilon(u) = \frac{dP^\epsilon_{\theta_0+\varphi(\epsilon)u}}{dP^\epsilon_{\theta_0}}(X^\epsilon) \quad (u \in \mathbb{R}^m)$$

Theorem 0. *(Ibragimov-Has'minskii [4, 5, 6]) Suppose that Z_ϵ satisfies the following conditions:*

(i) *There exist $\alpha > m$ and $k \geq \alpha$ such that for some constant C,*

$$P^\epsilon \left[|Z_\epsilon(u_2)^{1/k} - Z_\epsilon(u_1)^{1/k}|^k \right] \leq C|u_2 - u_1|^\alpha \quad (\forall u_1, u_2, \epsilon). \tag{1}$$

(ii) *For some $\gamma > 0$ and $c > 0$,*

$$P^\epsilon \left[Z_\epsilon^{1/2}(u) \geq e^{-c|u|^\gamma} \right] \leq e^{-c|u|^\gamma}. \tag{2}$$

(iii) *Finite-dimensional convergence: $Z_\epsilon \to^{d_f} Z$, where Z is a $\hat{C}(\mathbb{R}^m)$-valued random variable.*[1]

Then $(P^\epsilon)^{Z_\epsilon} \to \mathcal{L}\{Z\}$. Moreover,

$$P^\epsilon \left[\sup_{u:|u|\geq r} Z_\epsilon(u) > e^{-c_1 r^\gamma} \right] \leq e^{-c_1 r^\gamma}.$$

(↓)

[1] $\hat{C}(\mathbb{R}^m)$ is the sapce of continuous functions that tends to zero at the infinity.

(↓)

If $\hat{u}$ *uniquely attains the maximum of* $Z(u)$*, then for any sequence of the maximum likelihood estimator* $\hat{\theta}_\epsilon$ *for* θ*,* $\hat{u}_\epsilon := \varphi(\epsilon)^{-1}(\hat{\theta}_\epsilon - \theta_0) \to^d \hat{u}$ *and moreover*

$$P^\epsilon\left[f(\hat{u}_\epsilon)\right] \to P\left[f(\hat{u})\right]$$

for every $f \in C_\uparrow(\mathbb{R}^m)$*.*[2]

[2]The set of continuous functions on $\mathbb{R}^m$ of at most polynomial growth.

Why PLD?

Estimation of the probaility in theoretical statistics:

- The expected value of the plug-in functional

$$P^n \left[\log\ \text{likelihood}\ (\hat{\theta}_n) \right]$$

  in prediction theory and in construction of information criteria.

-

$$\mathbf{P} \left[|\hat{\theta}_n - \theta_0| > n^{-\beta} \right] \leq n^{-N}$$

  for $\beta < 1/2$ in the higer-order asymptotic theory.

- Bayes estimator

History:

- Kutoyants' program: Ibragimov-Has'minskii's scheme works for stochastic processes including diffusion type processes and point processes.
  - A complete theory for processes with small diffusions (Kutoyants [9], Kutoyants [10].)
  - Note: the Ibragimov-Has'minskii-Kutoyants' theory (as well as the Malliavin calculus) was effectively used in derivation of asymptotic expansions ([26, 28]).
- The large deviation inequality (2) is not easy to obtain for most of stochastic processes, even for nonlinear ergodic diffusions
- Without such a strong assumption, the convergence of statistical random fields was proved in [25], however without moment convergence.

- An important observation: the exponential type large deviation inequality like (2) is much stronger than our use. The polynomial type large deviation inequality is sufficient to develop a theory:

$$P^{\epsilon}\left[\sup_{u:|u|\geq r} Z_{\epsilon}(u) \geq r^{-N}\right] \leq \frac{C_N}{r^N} \quad (r > 0).$$

  Some exponential in place of r^{-N} on the left-hand side is very often possible.

- Kutoyants [12] presented a polynomial type large deviation inequality for one-dimensional diffusions by means of the local time.

Polynomial type large deviation inequality

- Θ: a bounded open set in $\mathbb{R}^m$
- The unknown parameter space $\Xi = \Theta \times \mathcal{T}$, where $\mathcal{T}$ is an arbitrary set.
- The statistical inference for the unknown parameter θ will be done based on the random fields $\mathbb{H}_T : \Omega \times \Xi \to \mathbb{R}$.
- ξ_0: an index of the probability measures, in many case $\xi_0 = (\theta_0, \tau_0)$, the true value of the unknown parameter.
- Replace θ by another local parameter u to focus on the vicinity of $\theta_0 \in \Theta$.
- Let $\mathbb{U}_T(\xi_0) = \{u \in \mathbb{R}^m;\ \theta_0 + a_T(\xi_0)u \in \Theta\}$
- $a_T(\xi_0) \in \mathrm{GL}(m)$: a deterministic sequence in $\mathrm{GL}(m)$ satisfying $|a_T(\xi_0)| \to 0$ as $T \to \infty$.
- We mimic the local asymptotic theory to define the random field $\mathbb{Z}_T : \mathbb{U}_T \times \mathcal{T} \to (0, \infty)$ by

$$\mathbb{Z}_T(u, \tau; \theta_0) = \exp\left\{\mathbb{H}_T(\theta_0 + a_T(\xi_0)u, \tau) - \mathbb{H}_T(\theta_0, \tau)\right\}.$$

We assume that the random field $\mathbb{Z}_T$ is

partially locally asymptotically quadratic at $\theta_0 \in \Theta$ under P_{ξ_0}:

$$\mathbb{Z}_T(u,\tau;\theta_0) = \exp\left(\Delta_T(\tau;\xi_0)[u] - \frac{1}{2}\Gamma(\tau;\xi_0)[u,u] + r_T(u,\tau;\xi_0)\right) \quad (3)$$

for $(u,\tau) \in \mathbb{U}_T(\theta_0) \times \mathcal{T}$.

- $\Delta_T(\tau;\xi_0)$ is a random linear form on $\mathbb{R}^m$.
- $\Gamma(\tau;\xi_0)$ a random bilinear form on $\mathbb{R}^m \times \mathbb{R}^m$.
- The residual term $r_T(u,\tau;\xi_0)$ is expected, under P_{ξ_0}, to be small when $T \to \infty$ but we shall specify its property more precisely later.

In what follows, we fix a set $K \subset \Theta \times \mathcal{T}$ and also fix a positive number L.

Let $V_T(r,\xi_0) = \{u \in \mathbb{U}_T(\xi_0);\ r \le |u|\}$. The following theorem gives a polynomial type large deviation inequality for the statistical random field.

Theorem 1. *Given $L > 0$, assume Conditions* [A1]-[A6]. *Then there exists a constant C_L such that*

$$\sup_{\xi_0 \in K} P_{\xi_0}\left[\sup_{(u,\tau)\in V_T(r,\xi_0)\times\mathcal{T}} \mathbb{Z}_T(u,\tau;\theta_0) \ge e^{-r^{2-(\rho_1\vee\rho_2)/2}}\right] \le \frac{C_L}{r^L} \tag{4}$$

for all $T > 0$ and $r > 0$.

Remark 1. See Y. ('05, preprint) for details of this result and applications to SDE discussed below. For the consequence of Theorem 1, essentially, we only assume the L^p-boundedness of certain random variables.

Graded parameter spaces

- For each $k \in \{1, ..., K\}$, let Θ_k be a bounded open set [3] in $\mathbb{R}^{d_k}$. $\theta = (\theta_1, ..., \theta_K)$ is a blocked vector in $\Theta_1 \times \cdots \times \Theta_K$. The true value of θ is denoted by $\theta^* = (\theta_1^*, ..., \theta_K^*)$. We assume that $\mathcal{T}$ is a one-point set and omit τ from notation (or more precisely, $\mathcal{T}$ is the spaces with indices bigger than k at k-th step).

- $a_T^k(\theta_k^*)$ is deterministic invertible $d_k \times d_k$-matrix and set $a_T(\theta^*) = \mathrm{diag}(a_T^1(\theta_1^*), ..., a_T^K(\theta_K^*))$. We write $\theta_k^\dagger$ for $\theta_k^* + a_T^k(\theta_k^*)u_k$, $u_k \in \mathbb{R}^{d_k}$.

- Also write $\underline{\theta}_k = (\theta_1, ..., \theta_k)$ and $\overline{\theta}_k = (\theta_k, ..., \theta_K)$. Parameters $\underline{\theta}_k^*$ and $\overline{\theta}_k^*$ are defined in a similar manner.

- Write $u = (u_1, ..., u_K)$ and $\theta^\dagger = (\theta_1^\dagger, ..., \theta_K^\dagger)$ and $d = \sum_{k=1}^K d_k$.

[3] A bounded closed domain is also available if the reader would like insurance for the existence of an exact M-estimator though an asymptotic M-estimator is sufficient to consider in asymptotic theory.

Let $\Theta = \prod_{k=1}^{K} \Theta_k$. We consider a sequence of statistical random fields $\mathbb{H}_T : \Omega \times \Theta \to \mathbb{R}$. The exponential statistical random field is now given by

$$\mathbb{Z}_T(u; \theta^*) = \exp\left(\mathbb{H}_T(\theta^\dagger) - \mathbb{H}_T(\theta^*)\right).$$

The k-th random field $\mathbb{Z}_T^k(u_k; \underline{\theta}_{k-1}, \theta_k^*, \overline{\theta}_{k+1})$ is defined by

$$\mathbb{Z}_T^k(u_k; \underline{\theta}_{k-1}, \theta_k^*, \overline{\theta}_{k+1}) = \exp\left(\mathbb{H}_T(\underline{\theta}_{k-1}, \theta_k^\dagger, \overline{\theta}_{k+1}) - \mathbb{H}_T(\underline{\theta}_{k-1}, \theta_k^*, \overline{\theta}_{k+1})\right). \tag{5}$$

By convention, we neglect symbols with index $K+1$ like $\overline{\theta}_{K+1}$ and ones with index 0 like $\underline{\theta}_0$.

Let $V_T^k(r, \theta_k^*) = \{u_k \in \mathbb{U}_T^k(\theta_k^*);\ r \leq |u_k|\}$, where $\mathbb{U}_T^k(\theta_k^*) = \{u_k \in \mathbb{R}^{d_k};\ \theta_k^* + a_T^k(\theta_k^*)u_k \in \Theta_k\}$.

Let L be a given positive number.

We obtain a polynomial type estimate for the M-estimator if we apply Theorem 1 stepwise. More precisely, we have

Proposition 1. *Let $\hat{\theta} \equiv \hat{\theta}_T = (\hat{\theta}_1, ..., \hat{\theta}_K)$ be any sequence of M-estimators maximizing $\mathbb{H}_T$.*[4] *Given $L > 0$ and $\mathcal{K} \subset \Theta$, suppose that certain polynomial type estimates are valid. Then the normalized M-estimator $\hat{u}_T = a_T(\theta^*)^{-1}(\hat{\theta}_T - \theta^*)$ satisfies the polynomial type large deviation inequality*

$$\sup_{\theta^* \in \mathcal{K}} P_{\theta^*} \left[|\hat{u}_T| > r \right] \leq \frac{C'_L}{r^L} \tag{6}$$

for all $T > 0$ and $r > 0$, for some constant C'_L.

In particular, this proposition ensures the tightness of $\{\hat{u}_T\}_{T>T_0}$.

[4] When there is no maximum, one can assign any number for the M-esitimator.

We considered a marginal (elementwise) convergence, but here we are interested in a joint convergence. For this purpose, we need only weak convergence of the statistical random field on compact sets because we already have tihgtness (or more strongly estimate of tail probability) of the M-estimator.

Write $B(R) = \{u \in \mathbb{R}^d;\ |u| \leq R\}$.

Theorem 2. *Suppose that the following conditions are fulfilled:*

(i) *For every* $R > 0$, $\mathbb{Z}_T(\cdot;\theta^*) \to^d \mathbb{Z}(\cdot;\theta^*)$ *in* $C(B(R))$ *as* $T \to \infty$.

(ii) *There exists a measurable mapping* $\hat{u}$ *that is a unique maximum point of* $\mathbb{Z}(\cdot;\theta^*)$.

Then

(a) $\hat{u}_T \to^d \hat{u}$ *as* $T \to \infty$ *if* $\{\hat{u}_T\}_{T>T_0}$ *is tight for some* $T_0 \geq 0$.

(b) $P_{\theta^*}[f(\hat{u}_T)] \to \mathbb{P}[f(\hat{u})]$ *as* $T \to \infty$ *for all continuous functions* f *satisfying* $\lim_{|u|\to\infty} |f(u)||u|^{-p} = 0$ *if* $\limsup_{T\to\infty} \|\hat{u}_T\|_p < \infty$, *in particular, if (6) holds for some* $L > p$.

Convergence of the integration type functionals and Bayes estimators

Two Bayesian estimators:

$$\begin{cases} \circ \text{ simultaneous Bayes type estimator} \\ \\ \circ \text{ adaptive Bayes type estimator} \end{cases}$$

Theorem 11. *Let $\mathcal{K} \subset \Theta$. Suppose that the following conditions are satisfied:*

(i) *For every $k = 1, ..., K$ and $R > 0$, $\mathbb{Z}_T^k(\cdot; \widetilde{\underline{\theta}}_{k-1}, \theta_k^*, \overline{\theta}^{\star}_{k+1}) \to^{d(P_{\theta^*})} \mathbb{Z}^k(\cdot; \underline{\theta}^*_{k-1}, \theta_k^*, \overline{\theta}^*_{k+1})$ in $C(\{u_k; |u_k| \le R\})$ uniformly in $\theta^* \in \mathcal{K}$ as $T \to \infty$.*

(ii) *Certain PLD inequalities are valid.*

Then the following assertions hold:

(a) *For any $f \in C(\mathbb{R}^{d_k}; \mathbb{R}^s)$ satisfying $\sup_{u_k \in \mathbb{R}^d} (1 + |u_k|^p)^{-1} |f(u_k)| < \infty$, the convergence of the integration type functionals hold.*

(↓)

(↓)

(b) *When $p \geq 1$, if each family $\mathcal{L}\{(\int_{\mathbb{R}^{d_k}} \mathbb{Z}^k(u_k;\theta^*)du_k)^{-1}|P_{\theta^*}\}$ $(\theta^* \in \mathcal{K})$ is uniformly tight, then for the adaptive Bayes type estimators,*

$$\tilde{u}_{k,T} \rightarrow^{d(P_{\theta^*})} \tilde{u}_k$$

uniformly in $\theta^* \in \mathcal{K}$ as $T \rightarrow \infty$ for each $k = 1, ..., K$.

Moreover, if the uniform joint convergence hold for "Z_T^k" random fields, then the adaptive Bayes type estimator

$$\tilde{u}_T \rightarrow^{d(P_{\theta^*})} \tilde{u}$$

uniformly in $\theta^* \in \mathcal{K}$ as $T \rightarrow \infty$.

(↓)

(↓)

(c) *Let $p \geq 1$ and $L > p+1$. Then under certain non-degeneracy conditions,*

$$P_{\theta^*}\left[f(\tilde{u}_{k,T})\right] \to \mathbb{P}\left[f(\tilde{u}_k)\right]$$

uniformly in $\theta^* \in \mathcal{K}$ as $T \to \infty$, for any continuous function f satisfying $\limsup_{|u_k|\to\infty} |u_k|^{-p}|f(u_k)| < \infty$.

Moreover,

$$P_{\theta^*}\left[g(\tilde{u}_T)\right] \to \mathbb{P}\left[g(\tilde{u})\right]$$

for any continuous function g satisfying $\limsup_{|u|\to\infty} |u|^{-p}|g(u)| < \infty$ under the joint convergence assumption in (b).

- There are theorems that ensure the convergences of "$\mathbb{Z}_T$" random fields on compacts in a general setting. We do not go into details here.
- In a similar way as $\mathbb{Z}_T$, a PLD inequality for the random field ψ_T for a "Z-estimator" (an estimator that is a root of a random equation $\psi_T = 0$) has been proved.

An example: estimators for a stochastic differential equation based on sampled data

A d-dimensional stationary diffusion process satisfying the stochastic differential equation

$$dX_t = a(X_t, \theta_2)dt + b(X_t, \theta_1)dw_t, \quad X_0 = x_0.$$

- w_t is an r-dimensional standard Wiener process independent of the initial value x_0.
- θ_1 and θ_2 are unknown parameters with $\theta_i \in \Theta_i$, a bounded open set in $\mathbb{R}^{m_i}$ admitting Sobolev's inequalities for embedding $W^{1,p}(\Theta_i) \hookrightarrow C(\Theta_i)$ or the GRR inequality.[5]
- The distribution of x_0 depends on those parameters.
- The true value of the unknown parameter is denoted by $\theta^* = (\theta_1^*, \theta_2^*)$.

[5] Those inequalities are valid for *daily-life* domains. The corn condition is sufficient for it; see Adams [1]. On the other hand, it is possible to replace estimates by Sobolev's inequality appearing below by Kolmogorov's type criteria for tightness of random fields.

We assume the following condition.

[D1] (i) The mappings $a : \mathbb{R}^d \times \Theta_2 \to \mathbb{R}^d$ and $b : \mathbb{R}^d \times \Theta_1 \to \mathbb{R}^d$ have continuous derivatives satisfying

$$\sup_{\theta_2 \in \Theta_2} \left|\partial_{\theta_2}^i a(x,\theta_2)\right| \leq C(1+|x|)^C \quad (0 \leq i \leq 4)$$

and

$$\sup_{\theta_1 \in \Theta_1} \left|\partial_x^j \partial_{\theta_1}^i b(x,\theta_1)\right| \leq C(1+|x|)^C \quad (0 \leq i \leq 4\,; 0 \leq j \leq 2)$$

for some constant C.

(ii) $B(x,\theta_1) = bb'(x,\theta_1)$ is elliptic uniformly in (x,θ_1).

(iii) For some constant C,

$$\begin{aligned} &\sup_{\theta_2 \in \Theta_2} |a(x_1,\theta_2) - a(x_2,\theta_2)| + \sup_{\theta_1 \in \Theta_1} |b(x_1,\theta_1) - b(x_2,\theta_1)| \\ \leq\ & C|x_1 - x_2| \quad (x_1, x_2 \in \mathbb{R}^d). \end{aligned}$$

(iv) $X_0 \in \bigcap_{p>0} L^p(P_{\theta^*})$.

Now we want to estimate the unknown parameters with the discrete-time observations

$$\mathbf{x}_n = (X_{t_i})_{i=0}^n,$$

where $t_i = ih$ with $h = h_n$ depending on $n \in \mathbb{N}$.

For this purpose, we consider a quasi-likelihood function

$$\begin{aligned} & p_n(\mathbf{x}_n, \theta) \\ = & \prod_{i=1}^n \frac{1}{(2\pi h)^{d/2} |B(X_{t_{i-1}}, \theta_1)|^{1/2}} \\ & \times \exp\left(-\frac{1}{2h} B(X_{t_{i-1}}, \theta_1)^{-1} \left[(\Delta_i X - ha(X_{t_{i-1}}, \theta_2))^{\otimes 2} \right] \right) \end{aligned}$$

with $\Delta_i X = X_{t_i} - X_{t_{i-1}}$.

The maximum likelihood type estimator $\hat{\theta}_n = (\hat{\theta}_{1,n}, \hat{\theta}_{2,n})$ is an estimator that maximizes $p_n(\mathbf{x}_n, \theta)$ in $\theta = (\theta_1, \theta_2) \in \Theta = \Theta_1 \times \Theta_2$.

We assume a mixing property for X:

[D2] There exists a positive constant a such that

$$\alpha_X(h) \le a^{-1} e^{-ah} \quad (h > 0),$$

where $\alpha_X(h) = \sup_{t\in\mathbb{R}_+} \sup_{\substack{A\in\sigma[X_r; r\le t],\\ B\in\sigma[X_r; r\ge t+h]}} \left| P_{\theta^*}[A\cap B] - P_{\theta^*}[A]\, P_{\theta^*}[B] \right|.$

We will assume that $h \to 0$ and $nh^2 \to 0$ as $n \to \infty$.[6]
We also assume a set of identifiability conditions as [D3] and [D4].

[6] The last one is the so-called condition for rapidly increasing experimantal design. Relaxing this condition is possible, but we do not do it here for simplicity. We will comment on this point later.

We obtain the following theorem from Theorem 2.

Theorem 12. *Suppose that Conditions* [D1]-[D4] *are fulfilled. Then for any sequence of M-estimators for* $\theta = (\theta_1, \theta_2)$*, it holds that*

$$\left(\sqrt{n}(\hat{\theta}_1 - \theta_1^*), \sqrt{nh}(\hat{\theta}_2 - \theta_2^*)\right)$$
$$\to^d \ (\zeta_1, \zeta_2) \sim N_{m_1+m_2}\left(0, \mathrm{diag}\left[\Gamma_1(\theta^*)^{-1}, \Gamma_2(\theta^*)^{-1}\right]\right)$$

as $n \to \infty$*. Moreover,* $P_{\theta^*}\left[f(\sqrt{n}(\hat{\theta}_1 - \theta_1^*), \sqrt{nh}(\hat{\theta}_2 - \theta_2^*))\right] \to \mathbb{P}\left[f(\zeta_1, \zeta_2)\right]$ *as* $n \to \infty$ *for all continuous functions* f *of at most polynomial growth.*

Remark 2. Prakasa Rao [16, 17] presented asymptotic results for an ergodic diffusion process under a sampling scheme. The joint weak convergence in Theorem 12 was in Yoshida [27]. Kessler [7] treated a local Gaussian approximation with higher order correction terms to relax the rate of convergence of h to zero. Shimizu and Yoshida [21] obtained asymptotic normality for a maximul likelihood type estimator for a diffusion process with jumps. An asymptotic expansion formula, as well as the first order results, is known for an M-estimator for a sampled diffusion process with jumps (possibly infinitely many on compacts).

Theorem 13. *Suppose that Conditions* [D1]-[D4] *are fulfilled. Then for adaptive Bayes type estimator* $(\tilde{\theta}_{1,n}, \tilde{\theta}_{2,n})$ *for* $\theta = (\theta_1, \theta_2)$*, it holds that*

$$\begin{aligned}&\left(\sqrt{n}(\tilde{\theta}_{1,n} - \theta_1^*), \sqrt{nh}(\tilde{\theta}_{2,n} - \theta_2^*)\right) \\ &\to^d \ (\zeta_1, \zeta_2) \sim N_{m_1+m_2}\left(0, \mathrm{diag}\left[\Gamma_1(\theta^*)^{-1}, \Gamma_2(\theta^*)^{-1}\right]\right)\end{aligned}$$

as $n \to \infty$*. Moreover,* $P_{\theta^*}\left[f(\sqrt{n}(\tilde{\theta}_1 - \theta_1^*), \sqrt{nh}(\tilde{\theta}_2 - \theta_2^*))\right] \to \mathbb{P}\left[f(\zeta_1, \zeta_2)\right]$ *as* $n \to \infty$ *for all continuous functions* f *of at most polynomial growth.*

Remark 3. From our discussions, we have already had the same asymptotic normality of the simultaneous Bayes estimator.

Remark 4. The obtained asymptotic variance implies that our Bayesian type estimators are first-order efficient.

Remark 5. We confined our attention to estimation by sampled data from a diffusion process. However, the weak convergence of Bayes type estimators, to say nothing of the moment convergence, and the weak plus moment convergence of M-estimators are new at least for multi-dimensional diffusion processes even under continuous observations. Futhermore, it is clear that our method can apply to semimartingales.

The log Likelihood function for a semimartingale has a representation

$$\begin{aligned} \mathbb{H}_T(\theta) - \mathbb{H}_T(\theta_0) &= \delta(\theta) \cdot M^c_T - \frac{1}{2}\delta(\theta)^{\otimes 2} \cdot < M^{\otimes 2} >_T \\ &\quad + (Y(\theta) - 1) * \tilde{\mu}_T + (\log Y(\theta) - Y(\theta) + 1) * \mu_T \end{aligned}$$

under the true model. Obviously, a set of conditions (ergodicity, differentiability, integrability and global identifiability) implies a polynomial type large deviation inequality for the statistical random field $\mathbb{Z}_T(u;\theta_0)$, the weak convergence of $\mathbb{Z}_T(u;\theta_0)$ in $\hat{C}(\mathbb{R}^m)$, the asymptotic normality of any sequence of the maximum likelihood estimators $\hat{\theta}_T$ for θ, and the convergence of moments of $\hat{u}_T = \sqrt{T}(\hat{\theta}_T - \theta_0)$. It is easy to present a set of sufficient conditions for each result in this article applied to this semimartingale model.

Comments on applications to SDE with jumps

Our general asymptotic results apply to the stochastic differential equation with jumps. Hoevever, a variety of results is possible and it is beyond a simple description, so we will do it elsewhere and, instead, give some comments here. Z-estimation scheme seems most feasible in a general case, but here we only consider a quasi-likelihood approach.

Consider a parametric model of stochastic integral equations:

$$\begin{aligned} X_t &= X_0 + \int_0^t a(X_s, \theta_2) ds + \int_0^t b(X_s, \theta_1) dw_t \\ &\quad + \int_0^t \int_E c(X_{s-}, v, \theta_2)(\mu(ds, dv) - \nu_{\theta_2}(ds, dv)). \end{aligned}$$

w is an r-dimensional Wiener process, and μ is a Poisson random measure on $\mathbf{R}_+ \times E$, $E = \mathbf{R}^k - \{0\}$, with compensator given by $\nu_{\theta_2}(dt, dv) = f_{\theta_2}(v) dt dv$. Assume that $X = (X_t)_{t \in \mathbf{R}_+}$ is stationary with distribution π.

We want to estimate θ from the sampled data $(X_{t_i})_{i=0}^{n}$. We assume that $h \to 0$, $nh \to \infty$ and $nh^2 \to 0$.

In this talk, we present a random field $\mathbb{H}_n(\theta)$ of partially quasi-log ikelihood type.

Comments on PLD for a random field associated with a Z-estimator

- In a similar way as $\mathbb{Z}_T$, a PLD inequality for the random field ψ_T for a "Z-estimator" (an estimator that is a root of a random equation $\psi_T = 0$) has been proved.
- One can construct a Z-estimator fairly freely even when a maximum likelihood type estimator is difficult to construct or analyze. It is caused e.g. by complexity of the jump structure and by degeneracy of coefficients of the SDE. Even in such a case, the Z-estimation method works and the asymptotic normality of the estimator can be proved.

References

[1] Adams, R. A: Sobolev spaces. Pure and Applied Mathematics, Vol. 65. Academic Press, New York-London, 1975.

[2] Doukhan, P.: Mixing: properties and examples. Lect. Notes in Statisit. **85**, Springer 1995

[3] Doukhan, P., Oppenheim, G., Taqqu, M. S. (eds.): Theory and applications of long-range dependence. Birkhauser Boston, Inc., Boston, MA, 2003

[4] Ibragimov, I.A., Has'minskii, R.Z.: The asymptotic behavior of certain statistical estimates in the smooth case. I. Investigation of the likelihood ratio. (Russian) Teor. Verojatnost. i Primenen. 17 (1972), 469–486.

[5] Ibragimov, I.A., Has'minskii, R.Z.: Asymptotic behavior of certain statistical estimates. II. Limit theorems for a posteriori density and for Bayesian estimates. (Russian) Teor. Verojatnost. i Primenen. 18 (1973), 78–93.

[6] Ibragimov, I.A., Has'minskii, R.Z.: Statistical estimation: asymptotic theory. Springer, New York 1981

[7] Kessler, M.: Estimation of diffusion processes from discrete observations. Scand. J. Statist. **24**, 211-229 (1997)

[8] Kusuoka, S., Yoshida, N.: Malliavin calculus, geometric mixing, and expansion of diffusion functionals. Prob. Theory Related Fields **116**, 457-484 (2000)

[9] Kutoyants, Yu. A.: Parameter estimation for stochastic processes. Translated and edited by B.L.S.Prakasa Rao, Berlin: Herdermann 1984

[10] Kutoyants, Yu.: Identification of dynamical systems with small noise. Dordrecht Boston London: Kluwer 1994

[11] Kutoyants, Yu. A.: Statistical inference for spatial Poisson processes. Lect. Notes in Statistics **134**, Berlin Heiderberg New York London Paris Tokyo Hong Kong: Springer 1998

[12] Kutoyants, Yu. A.: Statistical inference for ergodic diffusion processes. Springer Series in Statistics. Springer-Verlag London, Ltd., London, 2004

[13] Meyn, S. P., Tweedie, R. L.: Stability of Markovian processes. I. Criteria for discrete-time chains. Adv. in Appl. Probab. **24**, 542–574 (1992)

[14] Meyn, S. P., Tweedie, R. L.: Stability of Markovian processes. II. Continuous-time processes and sampled chains. Adv. in Appl. Probab. **25**, 487–517 (1993)

[15] Meyn, S. P., Tweedie, R. L.: Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. Adv. in Appl. Probab. **25**, 518–548 (1993)

[16] Prakasa-Rao, B. L. S.: Asymptotic theory for non-linear least square estimator for diffusion processes, Math. Operationsforsch. Statist. Ser. Stat. **14**, 195-209 (1983)

[17] Prakasa-Rao, B. L. S.: Statistical inference from sampled data for stochastic processes, Contemp. Math. **80**, 249-284 (1988)

[18] Rio, E.: Inegalites de moments pour les suites stationnaires et fortement melangeantes. (French) [Moment inequalities for stationary strongly mixing sequences] C. R. Acad. Sci. Paris Ser. I Math. 318, 355–360.(1994)

[19] Sakamoto, Y., Yoshida, N.: Higher order asymptotic expansions for a functional of a mixing process and applications to diffusion functionals. unpublished manuscript (1999)

[20] Sakamoto, Y., Yoshida, N.: Asymptotic expansion formulas for functionals of epsilon-Markov processes with a mixing property. Annals of the Institute of Statistical Mathematics, 56, 545-597 (2004)

[21] Shimizu, Y.; Yoshida, N.: Estimation of diffusion processes with jumps from discrete observations. (2002) preprint, to appear in Statistical Inference for Stochastic Processes

[22] Uchida, M., Yoshida, N.: Information criteria in model selection for mixing processes Statistical Inference for Stochastic Processes, 4, 73-98 (2001)

[23] Veretennikov, A. Yu. : Bounds for the mixing rate in the theory of stochastic equations. Theory Probab. Appl. **32**, 273-281 (1987)

[24] Veretennikov, A. Yu. : On polynomial mixing bounds for stochastic differential equations. Stoch. Processes Appl. **70**, 115-127 (1997)

[25] Yoshida, N.: Asymptotic behavior of M-estimator and related random field for diffusion process. Annals of the Institute of Statistical Mathematics, **42**, No. 2, 221-251 (1990)

[26] Yoshida, N: Asymptotic expansion for small diffusions via the theory of Malliavin-Watanabe. Prob. Theory Related Fields. **92**, 275-311 (1992)

[27] Yoshida, N: Estimation for diffusion processes from discrete observation. J. Multivar. Anal. **41**, 220-242 (1992)

[28] Yoshida,N.: Asymptotic expansion of Bayes estimators for small diffusions. Probab. Theory Relat. Fields **95**, 429-450 (1993)

[29] Yoshida, N.: Partial mixing and conditional Edgeworth expansion. Probab. Theory Related Fields 129, 559-624 (2004)

[30] Yoshida, N.: Asymptotic expansion for general M-estimation for stochastic differential equation with jumps. *Zenkin Tenkai 2003*, Hiroshima International University (2003)

[31] Yoshida, N.: Polynomial type large deviation inequality and its applications. preprint (2005)