Quasi-likelihood analysis and its applications

Nakahiro Yoshida¹

Received: 4 October 2021 / Accepted: 27 December 2021 / Published online: 23 February 2022
© The Author(s) 2022

Abstract

The Ibragimov–Khasminskii theory established a scheme that gives asymptotic properties of the likelihood estimators through the convergence of the likelihood ratio random field. This scheme is extending to various nonlinear stochastic processes, combined with a polynomial type large deviation inequality proved for a general locally asymptotically quadratic quasi-likelihood random field. We give an overview of the quasi-likelihood analysis and its applications to ergodic/non-ergodic statistics for stochastic processes.

Keywords Ibragimov–Khasminskii theory · Quasi-likelihood · Locally asymptotically quadratic random field · Polynomial type large deviation · Quasi-maximum likelihood estimator · Quasi-Bayesian estimator · Diffusion process · Adaptive estimator · Jump-diffusion process · Lévy process · Volatility · Jump filter · Point process · Non-synchronous observations · Information criterion · Sparse estimation

1 Introduction

We consider an m -dimensional semimartingale $Y = (Y_t)_{t \in [0, T]}$ having a decomposition

$$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma(X_s, \theta) dw_s, \quad t \in [0, T] \quad (1.1)$$

on a stochastic basis $(\Omega, \mathcal{F}, \mathbf{F}, P)$ with a filtration $\mathbf{F} = (\mathcal{F}_t)_{t \in [0, T]}$. The time horizon T is fixed. The process $w = (w_t)_{t \in [0, T]}$ is an r -dimensional Wiener process with respect to $\mathbf{F}$, and $\sigma : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^m \otimes \mathbb{R}^r$ is a given function. The process $b = (b_t)_{t \in [0, T]}$ and $X = (X_t)_{t \in [0, T]}$ are respectively m -dimensional and d -dimensional progressively measurable processes. The process b is unobservable but we observe the data $(X_{t_j}, Y_{t_j})_{j=0,1,\dots,n}$ for $t_j = t_j^n = jT/n$. We aim at estimation of the unknown parameter $\theta \in \Theta$, a p -dimensional bounded open set. The model (1.1) includes diffusion processes on the time interval $[0, T]$. The model (1.1) is called a stochastic regression model.

Dedicated to Professor Rafail Z. Khasminskii on the occasion of his 90th anniversary

✉ Nakahiro Yoshida
nakahiro@ms.u-tokyo.ac.jp

¹ Graduate School of Mathematical Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan

When $Y = X$ and $b_t = b^o(X_t)$ for some function $b^o : \mathbb{R}^d \rightarrow \mathbb{R}^d$, the likelihood function is

$$L_n(\theta) = \prod_{j=1}^n p_h(X_{t_{j-1}}, X_{t_j}, \theta)$$

if the distribution of the initial value X_0 does not depend on θ , where $p_h(x, y, \theta)$ is the transition density of the diffusion process. However, we do not know about the function b^o by assumption, so we cannot use the function L_n for estimation. Even if we know the function b^o , it is not in general easy to compute $p_h(X_{t_{j-1}}, X_{t_j}, \theta)$ since it is a solution to a partial differential equation, besides, optimization is necessary after getting $L_n(\theta)$. The situation is more severe when $Y \neq X$ since no information about the structure of X is given. In any case, it is more realistic to replace the likelihood function L_n by some other easily handled utility function.

A candidate of the utility function is the quasi-log likelihood function $\mathbb{H}_n$ defined by

$$\begin{aligned} \mathbb{H}_n(\theta) = & -\frac{1}{2} \sum_{j=1}^n \left\{ \log \det S(X_{t_{j-1}}, \theta) \right. \\ & \left. + h^{-1} S(X_{t_{j-1}}, \theta)^{-1} [(\Delta_j Y)^{\otimes 2}] \right\}, \end{aligned} \quad (1.2)$$

where $S = \sigma \sigma^*$, $\star$ denoting the matrix transpose, and $\Delta_j Y = Y_{t_j} - Y_{t_{j-1}}$. The brackets $[\cdot \cdot \cdot]$ stand for the inner product. For example, $M[v^{\otimes}] = \sum_{i,j} M_{i,j} v^i v^j$ for a square matrix $M = (M_{i,j})$ and a vector $v = (v^i)$. A quasi-maximum likelihood estimator $\hat{\theta}_n$ is obtained by maximizing $\mathbb{H}_n$ with respect to θ . Then $\hat{\theta}_n$ is asymptotically mixed normal and asymptotically efficient in Hájek's sense. To establish this property, we need to show that the risk function, e.g., the L^p -risk $E_\theta[|\sqrt{n}(\hat{\theta}_n - \theta)|^p]$ (locally) asymptotically attains the lower bound of risks, in particular, that the L^p -norm of (the scaled error of) $\hat{\theta}_n$ is bounded.

By using differentiability in θ of $\mathbb{H}_n$, we can derive a stochastic expansion of $\hat{\theta}_n$, namely,

$$\hat{u}_n := \sqrt{n}(\hat{\theta}_n - \theta^*) = \Gamma^{-1}(M_n + n^{-1/2}N_n),$$

where θ^* is the true value of θ and Γ is the random Fisher information matrix define by (5.1). The variable M_n is given by

$$M_n = -\frac{1}{2\sqrt{n}} \sum_{j=1}^n (\sigma^*(\partial_\theta S^{-1})\sigma)(X_{t_{j-1}}, \theta^*) [h^{-1}(\Delta_j w)^{\otimes 2} - I_r],$$

where I_r is the r -dimensional identity matrix and $\Delta_j w = w_{t_j} - w_{t_{j-1}}$, and the variable N_n is of $O_p(1)$ having a complicated expression involving multiple stochastic integrals. Asymptotic expansion of the distribution of $\hat{u}_n$ can be obtained if we apply the martingale expansion by Yoshida (2013) (updated by arXiv:1210.3680v3 (2012)). More precisely, we can derive asymptotic expansion of the joint distribution of (Z_n, Γ) for $Z_n = M_n + n^{-1/2}N_n$ and next obtain the expansion for $\hat{u}_n$ by transforming (Z_n, Γ) into it. In this procedure, the Malliavin calculus is applied and then we need L^p -boundedness of $\hat{u}_n$. As a matter of fact, boundedness of a smooth deformation of $\hat{u}_n$ in the Sobolev space $\mathbb{D}^{s,p}(\mathbb{R}^p)$ can be shown with the aid of the L^p -boundedness of $\hat{u}_n$. Relatively high order of integrability (i.e., a large p) is necessary to carry out this plot because the integration-by-parts formula in the Malliavin calculus requires algebras of variables. Asymptotic expansion of Skorokhod integrals recently presented by

Nualart and Yoshida (2019), combined with Yoshida (2020), is also applicable to this problem in place of the martingale expansion.

The L^p -boundedness of an estimator is a key to develop the asymptotic theory. The celebrated Ibragimov–Khasminskii theory (Ibragimov and Khasminskii 1973a, b; Ibragimov and Hasminskii 1981) answered this important question. Epoch-making was their introduction of the notion of weak convergence of the likelihood ratio random field, from which the asymptotic properties of the likelihood estimators (i.e., the maximum likelihood estimator and the Bayesian estimator) are induced in a unified way. The likelihood ratio random field $\mathbb{Z}_n$ is defined by $\mathbb{Z}_n(u) = L_n(\theta^* + a_n u) / L_n(\theta^*)$ with the likelihood function L_n and a scaling matrix a_n . They proved the convergence

$$\mathbb{Z}_n \rightarrow^d \mathbb{Z} \quad (n \rightarrow \infty) \quad (1.3)$$

in a certain space of continuous functions on $\mathbb{R}^p$ for suitably extended $\mathbb{Z}_n$. When $\mathbb{Z}_n$ is locally asymptotically normal (Le Cam 1960), the limit becomes

$$\mathbb{Z}(u) = \exp(\Delta[u] - 2^{-1} I(\theta^*)[u^{\otimes 2}]) \quad (1.4)$$

with the Fisher information matrix $I(\theta^*)$ at θ^* and a random vector $\Delta \sim N_p(0, I(\theta^*))$. Thanks to the functional convergence (1.3), roughly speaking, we can apply the $\operatorname{argmax}_{\theta}$ operator to the both sides of (1.3) to obtain the convergence $\hat{u}_n = a_n^{-1}(\hat{\theta}_n - \theta^*) \rightarrow^d I(\theta^*)^{-1} \Delta$. For the Bayesian estimator

$$\tilde{\theta}_n = \left(\int L_n(\theta) \varpi(\theta) d\theta \right)^{-1} \int \theta L_n(\theta) \varpi(\theta) d\theta$$

with respect to the quadratic loss function and a prior density ϖ , the error of $\tilde{\theta}_n$ has the expression

$$\tilde{u}_n := a_n^{-1}(\tilde{\theta}_n - \theta^*) = \left(\int \mathbb{Z}_n(u) \varpi(\theta^* + a_n u) du \right)^{-1} \int u \mathbb{Z}_n(u) \varpi(\theta^* + a_n u) du.$$

Then the convergence (1.3) suggests the joint convergence

$$\begin{aligned} & \left(\int \mathbb{Z}_n(u) \varpi(\theta^* + a_n u) du, \int u \mathbb{Z}_n(u) \varpi(\theta^* + a_n u) du \right) \\ & \rightarrow^d \left(\int \mathbb{Z}(u) \varpi(\theta^*) du, \int u \mathbb{Z}(u) \varpi(\theta^*) du \right), \end{aligned} \quad (1.5)$$

and hence

$$\tilde{u}_n \rightarrow^d \left(\int \mathbb{Z}(u) du \right)^{-1} \int u \mathbb{Z}(u) du = I(\theta^*)^{-1} \Delta$$

if (1.4) holds. One crucial point we should pay attention to is that the integrals appearing in (1.5) are essentially integrals over non-compact space since the domain tends to $\mathbb{R}^p$ as $n \rightarrow \infty$ even if Θ is bounded. To control these integrals, we need fast decay of the random field $\mathbb{Z}_n$.

The Ibragimov–Khasminskii theory features the large deviation inequality

$$\sup_n P \left[\sup_{u: |u| \geq r} \mathbb{Z}_n(u) \geq e^{-r^\alpha} \right] \leq \epsilon(r) \quad (r > 0) \quad (1.6)$$

for the likelihood ratio random field $\mathbb{Z}_n$, where α is a positive constant and $\epsilon(r)$ is a function of the form $c_0 e^{-c_1 r^{c_2}}$ or $c_0 r^{-L}$. Then the L^p -boundedness of $\hat{u}_n$ is a consequence of (1.6).

The inequality (1.6) is extremely important since it quantitatively estimates the tail of $\mathbb{Z}_n$. As well as the maximum likelihood estimator, the L^p -boundedness of $\tilde{u}_n$ also follows from (1.6).

Kutoyants (1984, 1994, 2004, 2012) successfully applied the Ibragimov–Khasminskii theory to semimartingales. Motivated by his pioneering works, the author tried to approach inference for stochastic processes by means of a quasi-likelihood (Yoshida 1990, 2011, 2021). In the applications of this article, the statistical models are differentiable and the used limit theorems are standard. So our interest is in the large deviation inequality (1.6). The aim of this paper is to give an overview of the quasi-likelihood analysis¹ and its applications.

2 Quasi-likelihood analysis

In this section, we will recall a simplified version of the quasi-likelihood analysis. We refer the interested reader to Yoshida (2021, 2011) for details.

2.1 Polynomial type large deviation inequality

We will work with a sequence of random fields $\mathbb{H}_T : \Omega \times \overline{\Theta} \rightarrow \mathbb{R}$ ($T \in \mathbb{T}$) for a probability space $(\Omega, \mathcal{F}, P)$ and a bounded open set Θ in $\mathbb{R}^p$. The set $\mathbb{T}$ is driving the asymptotic theory, supposed to satisfy $\mathbb{T} \subset \mathbb{R}$ and $\sup \mathbb{T} = \infty$. As for the regularity of the random field $\mathbb{H}_T$, since most basic is the case where the map $\Theta \ni \theta \mapsto \mathbb{H}_T(\omega, \theta) \in \mathbb{R}$ is of class $C^2(\Theta)$ for a.s. ω , we suppose this regularity and also that the map $\overline{\Theta} \ni \theta \mapsto \mathbb{H}_T(\omega, \theta) \in \mathbb{R}$ is continuous for a.s. ω . Though this smoothness assumption is much stronger than that assumed in the Ibragimov–Khasminskii theory and Le Cam’s LAN theory, it simplifies our theory and still applies to many applications in practice.

The targeted value of θ is denoted by $\theta^* \in \Theta$. The limit of the observed information is denoted by a $p \times p$ random symmetric matrix Γ . The minimum eigenvalues of a symmetric matrix M is denoted by $\lambda_{\min}(M)$. We need identifiability of θ^* expressed in terms of a random field $\mathbb{Y} : \Omega \times \Theta \rightarrow \mathbb{R}$, that will be related with $\mathbb{Y}_T$ at (2.4), and non-degeneracy of Γ as follows.

[T1] (i) There exists a positive random variable χ_0 and the following conditions are fulfilled.

(i-1) $\mathbb{Y}(\theta) = \mathbb{Y}(\theta) - \mathbb{Y}(\theta^*) \leq -\chi_0|\theta - \theta^*|^2$ for all $\theta \in \Theta$.

(i-2) For every $L > 0$, there exists a constant C such that

$$P[\chi_0 \leq r^{-1}] \leq \frac{C}{r^L} \quad (r > 0). \quad (2.1)$$

(ii) For every $L > 0$, there exists a constant C such that

$$P[\lambda_{\min}(\Gamma) < r^{-1}] \leq \frac{C}{r^L} \quad (r > 0)$$

Condition [T1] is almost trivial in ergodic statistics because χ_0 is a constant and Γ is a deterministic matrix. We remark that $\chi_0^{-1} \in L^{\infty-} = \cap_{p>1} L^p$ under [T1] (i-2). Moreover,

¹ The quasi-likelihood is not in the sense of the GLM. This term comes from statistics for stochastic processes, where one inevitably needs to resort to a quasi-likelihood method since the exact likelihood is not available for nonlinear discretely sampled processes.

$|\Gamma^{-1}| \in L^{\infty-}$ under [T1] (ii) since $(\lambda_{\min}(\Gamma))^{-1} \in L^{\infty-}$ and $|\Gamma^{-1}| \leq C_p(\lambda_{\min}(\Gamma))^{-1}$ for a constant only depending on p .

Let a_T be a $p \times p$ regular matrix such that $|a_T| \rightarrow 0$ as $T \rightarrow \infty$. The matrix a_T will specify the rate of convergence of the QLA estimators. Define b_T by $b_T = \{\lambda_{\min}(a_T^* a_T)\}^{-1}$. Obviously, $b_T \rightarrow \infty$ as $T \rightarrow \infty$. It is assumed that

$$b_T^{-1} \leq \lambda_{\max}(a_T^* a_T) \leq C_0 b_T^{-1} \quad (2.2)$$

for all $T \in \mathbb{T}$, for some constant $C_0 \in [1, \infty)$. A typical example is n for b_T and $n^{-1/2} I_p$ for a_T , I_p being the identity matrix. Define a p -dimensional random variable Δ_T and a $p \times p$ random matrix $\Gamma_T(\theta)$ by

$$\Delta_T = \partial_\theta \mathbb{H}_T(\theta^*) a_T \text{ and } \Gamma_T(\theta) = -a_T^* \partial_\theta^2 \mathbb{H}_T(\theta) a_T \quad (\theta \in \Theta), \quad (2.3)$$

respectively. Consistency of an estimator based on the random field $\mathbb{H}_T$ is established when $\mathbb{Y}$ is associated with the random field

$$\mathbb{Y}_T(\theta) = b_T^{-1} (\mathbb{H}_T(\theta) - \mathbb{H}_T(\theta^*)) \quad (\theta \in \Theta). \quad (2.4)$$

We assume the following set of conditions.

[T2] There exist positive numbers ϵ_1 and ϵ_2 such that the following conditions are satisfied for all $p > 1$:

- (i) $\sup_{T \in \mathbb{T}} \|\Delta_T\|_p < \infty$.
- (ii) $\sup_{T \in \mathbb{T}} \left\| \sup_{\theta \in \Theta} b_T^{\epsilon_1} |\mathbb{Y}_T(\theta) - \mathbb{Y}(\theta)| \right\|_p < \infty$.
- (iii) $\sup_{T \in \mathbb{T}} \left\| \sup_{u \in \delta^{-1}(\Theta - \theta^*), |u| \leq 1} |\Gamma_T(\theta^* + \delta u) - \Gamma_T(\theta^*)| \right\|_p = O(\delta) \quad (\delta \downarrow 0)$.
- (iv) $\sup_{T \in \mathbb{T}} \|b_T^{\epsilon_2} |\Gamma_T(\theta^*) - \Gamma| \|_p < \infty$.

To verify Conditions [T2] (ii) and (iii), one can apply Sobolev's embedding inequality, the Garsia-Rodemich-Rumsey inequality, or Kolmogorov's continuity theorem.

Let $\mathbb{U}_T = \{u \in \mathbb{R}^p; \theta^* + a_T u \in \Theta\}$. We define the random field $\mathbb{Z}_T$ on $\mathbb{U}_T$ by

$$\mathbb{Z}_T(u) = \exp(\mathbb{H}_T(\theta^* + a_T u) - \mathbb{H}_T(\theta^*))$$

for $u \in \mathbb{U}_T$. When $\mathbb{H}_T$ is a log likelihood function, the random field $\mathbb{Z}_T(u)$ is the likelihood ratio between $\theta = \theta^* + a_T u$ and θ^* .

For $r_T(u)$ defined by

$$r_T(u) = \begin{cases} \mathbb{H}_T(\theta^* + a_T u) - \mathbb{H}_T(\theta^*) - \left(\Delta_T[u] - \frac{1}{2} \Gamma[u^{\otimes 2}] \right) & (u \in \mathbb{U}_T) \\ 1 & (u \notin \mathbb{U}_T), \end{cases} \quad (2.5)$$

under [T2], the random field $\mathbb{Z}_T$ admits the representation

$$\mathbb{Z}_T(u) = \exp \left(\Delta_T[u] - \frac{1}{2} \Gamma[u^{\otimes 2}] + r_T(u) \right) \quad (u \in \mathbb{U}_T) \quad (2.6)$$

with $r_T(u) \rightarrow^p 0$ as $T \rightarrow \infty$ for every $u \in \mathbb{R}^p$, i.e., the random field $\mathbb{Z}_T$ is locally asymptotically quadratic (LAQ) at θ^* . The property (2.6) suggests that the tail of the random field $\mathbb{Z}_T$ is light like a Gaussian kernel. This fact is stated as a polynomial type large deviation inequality. Write $\mathbb{V}_T(r) = \{u \in \mathbb{U}_T; |u| \geq r\}$ for $r > 0$.

Theorem 2.1 Suppose that [T1] and [T2] are fulfilled. Let $\xi \in (1, 2)$. Then, for any $L > 0$, there exists a constant C such that

$$P \left[\sup_{u \in \mathbb{V}_T(r)} \mathbb{Z}_T(u) \geq \exp(-r^\xi) \right] \leq \frac{C}{r^L} \quad (2.7)$$

for all $r > 0$ and $T \in \mathbb{T}$. The supremum of the empty set should read $-\infty$.

The polynomial type large deviation inequality (2.7) ensures the L^p -boundedness of the scaled error of the QLA estimators. Usually, a common L^p -inequality based on some kind of orthogonality such as a martingale or mixing property can verify Condition [T2], combined with an uniform estimate like Sobolev's inequality. This enables flexible applications of the scheme to nonlinear stochastic processes, as we will see in various applications later.

2.2 Quasi-maximum likelihood estimator

Suppose that $\mathcal{G}$ is a σ -field satisfying $\sigma[\Gamma] \subset \mathcal{G} \subset \mathcal{F}$. For a sequence $(V_T)_{T \in \mathbb{T}}$ of random elements on $(\Omega, \mathcal{F}, P)$, we say that V_T converges $\mathcal{G}$ -stably to a random element V_∞ defined on an extension of $(\Omega, \mathcal{F}, P)$ if $(V_T, \Psi) \xrightarrow{d} (V_\infty, \Psi)$ as $T \rightarrow \infty$ for any $\mathcal{G}$ -measurable random variable Ψ . The $\mathcal{G}$ -stable convergence is denoted by $\xrightarrow{d_s(\mathcal{G})}$. We assume the $\mathcal{G}$ -stable convergence

$$\Delta_T \xrightarrow{d_s(\mathcal{G})} \Delta \quad (2.8)$$

as $T \rightarrow \infty$, for some p -dimensional random vector Δ on some extension of $(\Omega, \mathcal{F}, P)$.

Any measurable mapping $\hat{\theta}_T^M : \Omega \rightarrow \bar{\Theta}$ is called a quasi-maximum likelihood estimator (QMLE) for $\mathbb{H}_T$ if

$$\mathbb{H}_T(\hat{\theta}_T^M) = \max_{\theta \in \bar{\Theta}} \mathbb{H}_T(\theta). \quad (2.9)$$

Such a measurable function always exists, which is ensured by the measurable selection theorem applied to $\mathbb{H}_T$, a continuous random field on the compact $\bar{\Theta}$. We do not assume uniqueness of $\hat{\theta}_T^M$. Let $\hat{u}_T^M = a_T^{-1}(\hat{\theta}_T^M - \theta^*)$ for the QMLE $\hat{\theta}_T^M$. Let $\hat{u} = \Gamma^{-1}\Delta$.

Theorem 2.2 Suppose that Conditions [T1] and [T2] are satisfied and that the convergence (2.8) holds as $T \rightarrow \infty$. Then

- (a) $\hat{u}_T^M - \Gamma^{-1}\Delta_T \xrightarrow{p} 0$ as $T \rightarrow \infty$.
- (b) $E[f(\hat{u}_T^M)\Phi] \rightarrow E[f(\hat{u})\Phi]$ as $T \rightarrow \infty$ for any $f \in C_p(\mathbb{R}^p)$ and any $\mathcal{G}$ -measurable random variable $\Phi \in \cup_{p \geq 1} L^p$.

Remark 2.3 Define a random field $\mathbb{Z}$ on $\mathbb{R}^p$ by

$$\mathbb{Z}(u) = \exp \left(\Delta[u] - \frac{1}{2} \Gamma[u^{\otimes 2}] \right) \quad (2.10)$$

for $u \in \mathbb{R}^p$. From (2.7) and (2.8), we obtain the convergence

$$\mathbb{Z}_T \xrightarrow{d_s(\mathcal{G})} \mathbb{Z} \text{ in } \widehat{C}(\mathbb{R}^p) \quad (2.11)$$

if $\mathbb{Z}_T$ is adequately, measurably extended as a random variable taking values in $\widehat{C}(\mathbb{R}^p)$, where $\widehat{C}(\mathbb{R}^p)$ is the space of continuous functions f on $\mathbb{R}^p$ satisfying $\lim_{|u| \rightarrow \infty} f(u) = 0$. The space $\widehat{C}(\mathbb{R}^p)$ is equipped with the supremum norm. Roughly speaking, by applying the *argmax*-operator to the both sides of (2.11), we obtain the convergence $\hat{u}_T^M \xrightarrow{d_s(\mathcal{G})} \hat{u} = \Gamma^{-1}\Delta$. This is a smart way but one can bypass the discussion on the space $\widehat{C}(\mathbb{R}^p)$. See e.g. Yoshida (2021).

2.3 Quasi-Bayesian estimator

The mapping

$$\hat{\theta}_T^B = \left(\int_{\Theta} \exp(\mathbb{H}_T(\theta)) \varpi(\theta) d\theta \right)^{-1} \int_{\Theta} \theta \exp(\mathbb{H}_T(\theta)) \varpi(\theta) d\theta \quad (2.12)$$

is called a quasi-Bayesian estimator (QBE) with respect to the prior density ϖ . The QBE $\hat{\theta}_T^B$ is the Bayesian estimator with respect to the quadratic loss function when $\mathbb{H}_T$ is the log likelihood function. The QBE $\hat{\theta}_T^B$ takes values in the convex-hull of Θ , therefore the values are bounded but may be outside of Θ . It is assumed that ϖ is continuous and $0 < \inf_{\theta \in \Theta} \varpi(\theta) \leq \sup_{\theta \in \Theta} \varpi(\theta) < \infty$.

Theorem 2.4 *Suppose that Conditions [T1] and [T2] are satisfied and that the convergence (2.8) holds as $T \rightarrow \infty$. Moreover, suppose that $|\Gamma| \in L^q$ for some $q > p$. Then*

- (a) $\hat{u}_T^B - \Gamma^{-1} \Delta_T \rightarrow^p 0$ as $T \rightarrow \infty$.
- (b) $E[f(\hat{u}_T^B) \Phi] \rightarrow E[f(\hat{u}) \Phi]$ as $T \rightarrow \infty$ for any $f \in C_p(\mathbb{R}^p)$ and any $\mathcal{G}$ -measurable random variable $\Phi \in \cup_{p>1} L^p$.

As seen above, the QLA is constructed in an abstract way, and Conditions [T1] and [T2] are easy to verify. By this reason, the QLA theory has been widely applied, in particular, to nonlinear stochastic processes. We will discuss several applications in the following sections.

3 Diffusion process

3.1 Quasi-likelihood analysis for ergodic diffusion processes

Suppose that a d -dimensional stationary mixing diffusion process $X = (X_t)_{t \in \mathbb{R}_+}$ satisfies the stochastic differential equation

$$dX_t = a(X_t, \theta_2) dt + b(X_t, \theta_1) dw_t \quad (3.1)$$

on a stochastic basis $(\Omega, \mathcal{F}, \mathbf{F}, P)$ with a filtration $\mathbf{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$. Here $w = (w_t)_{t \in \mathbb{R}_+}$ is an r -dimensional $\mathbf{F}$ -Wiener process, $a : \mathbb{R}^d \times \Theta_2 \rightarrow \mathbb{R}^d$ and $b : \mathbb{R}^d \times \Theta_1 \rightarrow \mathbb{R}^d \otimes \mathbb{R}^r$ are given functions. The parameter space Θ_i of θ_i is a bounded open set in $\mathbb{R}^{p_i}$ for $i = 1, 2$. A standard Sobolev embedding inequality $W^{1, p_i}(\Theta_i) \hookrightarrow C(\Theta_i)$ is assumed for $p_i > p_i$, for each $i = 1, 2$. We also assume that the functions a and b are continuously extended to the boundaries $\mathbb{R}^d \times \partial\Theta_i$, respectively, and that b is uniformly non-degenerate.

The process X is observed at discrete times $t_j = jh$ ($j \in \{0, 1, \dots, n\}$) for a positive value $h = h_n$ depending on n . We assume that $h \rightarrow 0$, $nh \rightarrow \infty$ and $nh^2 \rightarrow 0$ as $n \rightarrow \infty$, that is, we have high frequency and long-run data. For estimation of the parameter $\theta = (\theta_1, \theta_2)$, we consider a random field $\mathcal{H}_n$ given by

$$\begin{aligned} \mathcal{H}_n(\theta) = & -\frac{1}{2} \sum_{j=1}^n \left\{ \log \det B(X_{t_{j-1}}, \theta_1) \right. \\ & \left. + h^{-1} B(X_{t_{j-1}}, \theta_1)^{-1} [(\Delta_j X - ha(X_{t_{j-1}}, \theta_2))^{\otimes 2}] \right\} \end{aligned} \quad (3.2)$$

for $B = bb^*$.

The asymptotic properties of the QMLE $\hat{\theta}_n^M = (\hat{\theta}_{n,1}^M, \hat{\theta}_{n,2}^M)$ with respect to $\mathcal{H}_n$ of (3.2) can be shown by the QLA approach recalled in Sect. 2.2, as was done in Yoshida (2011). Denote by $\theta^* = (\theta_1^*, \theta_2^*)$ the true value of θ . To obtain the asymptotic properties for $\hat{\theta}_{1,n}^M$ for θ_1 , we can use the random field

$$\mathbb{H}_n^{(1)}(\theta_1) = \mathcal{H}_n(\theta_1, \hat{\theta}_{2,n}^M)$$

in the proof, for $\mathbb{H}_T$ in the general theory. In the second step, the random field

$$\mathbb{H}_n^{(2)}(\theta_2) = \mathcal{H}_n(\hat{\theta}_{1,n}^M, \theta_2)$$

is used for $\mathbb{H}_T$ in the proof of the asymptotic properties of $\hat{\theta}_{2,n}^M$. Since the relation (a) of Theorem 2.2 has been obtained for each component $\hat{\theta}_{i,n}$, we have the joint convergence of these components. As a matter of fact, Yoshida (2011) gave the convergence

$$E \left[f \left(n^{1/2} (\hat{\theta}_{1,n}^M - \theta_1^*), (nh)^{1/2} (\hat{\theta}_{2,n}^M - \theta_2^*) \right) \right] \rightarrow \mathbb{E}[f(\xi_1, \xi_2)] \quad (3.3)$$

as $n \rightarrow \infty$ for any continuous function f of at most polynomial growth, where ξ_i is a $\mathbf{p}_i$ -dimensional centered Gaussian random vector with covariance matrix Γ_i^{-1} for each $i = 1, 2$, and ξ_1 and ξ_2 are independent. More precisely,

$$\Gamma_1[u_1^{\otimes 2}] = \frac{1}{2} \int_{\mathbb{R}^d} \text{Tr} \{ B^{-1}(\partial_{\theta_1} B[u_1]) B^{-1}(\partial_{\theta_1} B[u_1]) \} (x, \theta_1^*) \nu(dx) \quad (u_1 \in \mathbb{R}^{p_1}) \quad (3.4)$$

where ν is the stationary probability measure of X , and

$$\Gamma_2[u_2^{\otimes 2}] = \int_{\mathbb{R}^d} B(x, \theta_1^*)^{-1} [(\partial_{\theta_2} a(x, \theta_2^*)[u_2])^{\otimes 2}] \nu(dx) \quad (u_2 \in \mathbb{R}^{p_2}). \quad (3.5)$$

The condition (2.2) is not restrictive since one can choose a suitable random field $\mathbb{H}_n$ with a single scale at each step of the proof of asymptotic properties, although there are two different scales of estimators. Condition [T1] is trivial in the present situation, and Condition [T2] can be checked with the help of Sobolev's inequality and a Rosenthal type inequality. This task is easy.

For quasi-Bayesian inference, Yoshida (2011) proposed the adaptive Bayesian method (adaBayes). Though the adaptive Bayesian estimator was defined for the general $\mathbb{H}_T$ having k scales, it becomes in the present situation as follows. For an arbitrarily fixed value θ_2^0 of θ_2 , we define the quasi-Bayesian estimator $\hat{\theta}_{1,n}^B$ for θ_1 by

$$\begin{aligned} \hat{\theta}_{1,n}^B &= \left(\int_{\Theta_1} \exp(\mathcal{H}_n(\theta_1, \theta_2^0)) \varpi_1(\theta_1) d\theta_1 \right)^{-1} \\ &\quad \times \int_{\Theta_1} \theta_1 \exp(\mathcal{H}_n(\theta_1, \theta_2^0)) \varpi_1(\theta_1) d\theta_1 \end{aligned} \quad (3.6)$$

for a prior density ϖ_1 for θ_1 , and next define the quasi-Bayesian estimator $\hat{\theta}_{2,n}^B$ for θ_2 by

$$\begin{aligned} \hat{\theta}_{2,n}^B &= \left(\int_{\Theta_2} \exp(\mathcal{H}_n(\hat{\theta}_{1,n}^B, \theta_2)) \varpi_2(\theta_2) d\theta_2 \right)^{-1} \\ &\quad \times \int_{\Theta_2} \theta_2 \exp(\mathcal{H}_n(\hat{\theta}_{1,n}^B, \theta_2)) \varpi_2(\theta_2) d\theta_2 \end{aligned} \quad (3.7)$$

for a prior density ϖ_2 for θ_2 . Then, for the adaptive Bayesian estimator $\hat{\theta}_n^B = (\hat{\theta}_{1,n}^B, \hat{\theta}_{2,n}^B)$, we can apply the scheme in Sect. 2.3 twice to obtain the convergence

$$E \left[f \left(n^{1/2} (\hat{\theta}_{1,n}^B - \theta_1^*), (nh)^{1/2} (\hat{\theta}_{2,n}^B - \theta_2^*) \right) \right] \rightarrow \mathbb{E}[f(\xi_1, \xi_2)] \quad (3.8)$$

as $n \rightarrow \infty$ for f and ξ_i ($i = 1, 2$) described at (3.3).

The YUIMA is an R package for simulation and statistical analysis for stochastic processes. It constructs a *yuima* object from the user's data of a stochastic differential equation. The YUIMA function “*qml*” implemented the QMLE, and the function “*adaBaye*” the adaptive Bayesian estimator. For example, the “*qml*” applied to a *yuima* object returns the estimated value and the standard error of the QMLE. See Brouste et al. (2014) and Iacus and Yoshida (2018).

3.2 Adaptive methods

An advantage of the adaptive method is that it can suppress the dimension of the integral in computation of the Bayesian estimate. The maximum likelihood type estimator also enjoys this merit by the idea of adaptive estimation. Adaptive methods for diffusion models were studied by Yoshida (1992) and Kessler (1995). The condition $nh^2 \rightarrow 0$ is called the condition for rapidly increasing experimental design (Prakasa Rao 1983; Prakasa Rao 1988). The relaxation of this condition to $nh^3 \rightarrow 0$ was by Yoshida (1992) with a higher-order expansion of the transition probability of the diffusion process, and this was extended by Kessler (1997) to achieve $nh^p \rightarrow 0$ for any $p > 0$.

Uchida and Yoshida (2012) proposed various adaptive methods of the maximum likelihood type for the stochastic differential equation (3.1). They introduced a ladder of annealed random fields for $\mathbb{Z}_T$, and applied the QLA to prove that their adaptive schemes gave the same convergence as (3.3) at the last stage of the algorithm under the assumption $nh^p \rightarrow 0$, as explained below. Consider a sequence of estimating functions $U_{p,n}$ ($p = 1, 2, \dots$) as follows. Let

$$U_n^{(0)}(\theta) = -\frac{1}{2} \sum_{j=1}^n \left\{ h^{-1} B(X_{t_{j-1}}, \theta_1)^{-1} [(\Delta_j X)^{\otimes 2}] + \log \det B(X_{t_{j-1}}, \theta_1) \right\}.$$

Denote by $\hat{\theta}_{1,n}^{(0)}$ a maximum likelihood (ML) type estimator for θ_1 :

$$U_n^{(0)}(\hat{\theta}_{1,n}^{(0)}) = \max_{\theta_1 \in \Theta_1} U_n^{(0)}(\theta_1).$$

For $p \geq 2$, let $k_0 = \lfloor p/2 \rfloor$ and $l_0 = \lfloor (p-1)/2 \rfloor$. Then the function $U_{p,n}$ for $p \geq 2$ is defined by

$$U_{p,n}(\theta) = -\frac{1}{2} \sum_{j=1}^n \left\{ h^{-1} \left(\sum_{k=0}^{k_0} h^k D^{(k)}(X_{t_{j-1}}, \theta) \right) [(X_{t_j} - r^{(k_0)}(h, X_{t_{j-1}}, \theta))^{\otimes 2}] + \sum_{k=0}^{k_0} h^k E^{(k)}(X_{t_{j-1}}, \theta) \right\},$$

where the functions $r^{(k_0)}(h, x, \theta)$, $D^{(k)}(x, \theta)$ and $E^{(k)}(x, \theta)$ are coming from an expansion of the semigroup of the diffusion process satisfying (3.1). The adaptive ML type estimators

$\widehat{\theta}_{1,p,n}^{(l_0)}$ for θ_1 and $\widehat{\theta}_{2,p,n}^{(k_0)}$ are characterized by $\widehat{\theta}_{1,p,n}^{(0)} = \widehat{\theta}_{1,n}^{(0)}$ and

$$\begin{aligned} U_{p,n}(\widehat{\theta}_{1,p,n}^{(k-1)}, \widehat{\theta}_{2,p,n}^{(k)}) &= \max_{\theta_2 \in \Theta_2} U_{p,n}(\widehat{\theta}_{1,p,n}^{(k-1)}, \theta_2), \\ U_{p,n}(\widehat{\theta}_{1,p,n}^{(k)}, \widehat{\theta}_{2,p,n}^{(k)}) &= \max_{\theta_1 \in \Theta_1} U_{p,n}(\theta_1, \widehat{\theta}_{2,p,n}^{(k)}) \end{aligned}$$

for $k = 1, 2, \dots, k_0$. Remark that $l_0 \leq k_0 \leq l_0 + 1$. In Uchida and Yoshida (2012), the QLA provided $L^{\infty-} = \cap_{q>1} L^q$ -boundedness of $(nh)^{k/(p-1)}(\widehat{\theta}_{2,p,n}^{(k)} - \theta_2^*)$ when $p \geq 2k + 1$, and of $n^{(k+1)/p}(\widehat{\theta}_{1,p,n}^{(k)} - \theta_1^*)$ when $p \geq 2(k + 1)$. Climbing up the ladder, under the balance condition $nh^p \rightarrow 0$, they proved the convergence

$$\mathbb{E} \left[f \left(n^{1/2}(\widehat{\theta}_{1,p,n}^{(l_0)} - \theta_1^*), (nh)^{1/2}(\widehat{\theta}_{2,p,n}^{(k_0)} - \theta_2^*) \right) \right] \rightarrow \mathbb{E}[f(\xi_1, \xi_2)]$$

as $n \rightarrow \infty$ for any continuous function f of at most polynomial growth, where ξ_1 and ξ_2 are given in Sect. 3.1.

Uchida and Yoshida (2014) provided three types of algorithms for adaptive quasi-Bayesian estimation. The QLA worked effectively to show that the spiral of the estimators based on the annealed quasi-likelihood functions attains the convergence (3.3). To reduce computational load for the adaptive estimators, Kamatani and Uchida (2014) proposed hybrid multi-step estimators for diffusion processes. According to their numerical studies, the hybrid multi-step estimator with an initial QBE gives stable estimates.

Recently, Kutoyants (2017) proposed a multi-step MLE for ergodic diffusion, and Kutoyants (2014) presented approximation of the solution of the backward stochastic differential equation with a multi-step method. Dabye et al. (2018) gave moments estimators and multi-step MLE for Poisson Processes.

4 Stochastic differential equation with jumps

4.1 Jump diffusion process

Let us consider a d -dimensional ergodic process $X = (X_t)_{t \in \mathbb{R}_+}$ satisfying the stochastic differential equation

$$\begin{aligned} dX_t &= a(X_t, \theta_2)dt + b(X_t, \theta_1)dw_t + \int_{\mathbb{E}} c(X_{t-}, z, \theta_2)p(dt, dz), \\ X_0 &= x_0, \end{aligned} \quad (4.1)$$

$\mathbb{E} = \mathbb{R}^d \setminus \{0\}$, on a stochastic basis $(\Omega, \mathcal{F}, \mathbf{F}, P)$ with filtration $\mathbf{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$. For each $i \in \{1, 2\}$, the unknown parameter θ_i is in a bounded open set Θ_i of $\mathbb{R}^{p_i}$. The coefficients $a : \mathbb{R}^d \times \Theta_2 \rightarrow \mathbb{R}^d$, $b : \mathbb{R}^d \times \Theta_1 \rightarrow \mathbb{R}^d \otimes \mathbb{R}^r$ and $c : \mathbb{R}^d \times \mathbb{E} \times \Theta_2 \rightarrow \mathbb{R}^d$ are supposed to satisfy some mild regularity conditions. In (4.1), $w = (w_t)_{t \in \mathbb{R}_+}$ is an r -dimensional $\mathbf{F}$ -Wiener process, and $p(dt, dz)$ is a Poisson random measure on $\mathbb{R}^+ \times \mathbb{E}$ with the deterministic $\mathbf{F}$ -compensator $q^{\theta_2}(dt, dz)$. We want to estimate the true value $\theta^* = (\theta_1^*, \theta_2^*)$ of the unknown parameter $\theta = (\theta_1, \theta_2)$ by observing the data $(X_{t_j})_{j=0,1,\dots,n}$ with $t_j = jh$.

Estimation of a semimartingale with jumps has a different technical aspect than that of a continuous semimartingale. A natural idea for constructing an estimator is to apply a Gaussian type likelihood to the continuous part of X and a Poissonian likelihood to the jump component of X . However, this idea is naive. It is impossible to tell whether an increment

$\Delta_j X = X_{t_j} - X_{t_{j-1}}$ of X has jumps since only temporally discrete observations are available. We need some filter that detects jumps from the increments $(\Delta_j X)_{j=1,\dots,n}$. Shimizu and Yoshida (2006) proposed an estimator for θ by a threshold method and showed its asymptotic normality. Threshold method is a standard technique going back to studies of Lévy processes at latest.

The QLA was presented by Ogihara and Yoshida (2011) for the jump-diffusion process (4.1) satisfying a mixing condition. They considered the time-discretization step-size $h = h_n$ satisfying $n^{-3/5} \lesssim h \lesssim n^{-4/7}$, where $a_n \lesssim b_n$ means $s_n \leq Cb_n$ for all $n \in \mathbb{N}$, for some constant C , for sequences (a_n) and (b_n) of numbers. This balance condition is equivalent to $n^{2/5} \lesssim nh \lesssim n^{3/7}$ and $n^{-1/5} \lesssim nh^2 \lesssim n^{-1/7}$. Suppose that for each (θ_2, x) , the mapping $z \mapsto y = c(x, z, \theta_2)$ is an injection from E into E and has an inverse $z = c^{-1}(x, y, \theta_2)$ from the image of c onto E . We assume that $q^{\theta_2}(dt, dz) = f_{\theta_2}(z)dzdt$ with a density $f_{\theta_2}(z)$ (possibly $\int f_{\theta_2}(z)dz \neq 1$). Let

$$\Psi_{\theta_2}(y, x) = f_{\theta_2}(c^{-1}(x, y, \theta_2)) |\det \partial_y c^{-1}(x, y, \theta_2)|.$$

We suppose that $B(x) = \text{Im}(c(x, \cdot, \theta_2))$ is independent of $\theta_2 \in \Theta_2$. With $B = bb^*$, a positive constant ρ less than $1/2$ and a positive constant D , Ogihara and Yoshida (2011) defined the random field $\mathcal{H}_n(\theta)$ by

$$\begin{aligned} \mathcal{H}_n(\theta) = & -\frac{1}{2} \sum_{j=1}^n \left\{ h^{-1} B(X_{t_{j-1}}, \theta_1)^{-1} [(\overline{\Delta_j X})^{\otimes 2}] + \log \det B(X_{t_{j-1}}, \theta_1) \right\} \\ & \times 1_{\{|\Delta_j X| \leq Dh^\rho\}} \\ & + \sum_{j=1}^n \left\{ (\log \Psi_{\theta_2}(\Delta_j X, X_{t_{j-1}})) \varphi_n(X_{t_{j-1}}, \Delta_j X) 1_{\{|\Delta_j X| > Dh^\rho\}} \right. \\ & \left. - h \int_{B(X_{t_{j-1}})} \Psi_{\theta_2}(y, X_{t_{j-1}}) \varphi_n(X_{t_{j-1}}, y) dy \right\}, \end{aligned} \quad (4.2)$$

where $\overline{\Delta_j X} = \Delta_j X - ha(X_{t_{j-1}}, \theta_2)$ and φ_n is a truncation function that removes extremely small or extremely large increments. Denote by ν the invariant probability measure of the jump diffusion process X . By applying the QLA theory, Ogihara and Yoshida (2011) obtained the convergence (3.3) of the QMLE $\hat{\theta}_n^M = (\hat{\theta}_{1,n}^M, \hat{\theta}_{2,n}^M)$ with respect to $\mathcal{H}_n$ of (4.2), Γ_1 given by (3.4) and Γ_2 by

$$\begin{aligned} \Gamma_2[u_2^{\otimes 2}] = & \int_{\mathbb{R}^d} B(x, \theta_1^*)^{-1} [(\partial_{\theta_2} a(x, \theta_2^*)[u_2])^{\otimes 2}] \nu(dx) \\ & + \int \int_{A(x)} \Psi_{\theta_2^*}(y, x)^{-1} (\partial_{\theta_2} \Psi_{\theta_2^*}(y, x)[u_2])^2 dy \nu(dx) \quad (u_2 \in \mathbb{R}^{p_2}) \end{aligned} \quad (4.3)$$

instead of (3.5), where $A(x) := \{y \in B(x); \Psi_{\theta_2}(y, x) \neq 0\}$ is supposed to be independent of $\theta_2 \in \Theta_2$. The adaptive QBE $\hat{\theta}_n^B = (\hat{\theta}_{1,n}^B, \hat{\theta}_{2,n}^B)$ is defined by (3.6) and (3.7) with $\mathcal{H}_n$ of (4.2). The QLA ensures the convergence (3.8) with Γ_1 of (3.4) and Γ_2 of (4.3).

4.2 Gaussian quasi-likelihood to Lévy driven stochastic differential equation

Given a stochastic basis $(\Omega, \mathcal{F}, \mathbf{F}, P)$ with a filtration $\mathbf{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$, we consider an $\mathbf{F}$ -adapted process $X = (X_t)_{t \in \mathbb{R}_+}$ satisfying the stochastic differential equation

$$dX_t = a(X_t, \theta_2)dt + b(X_t, \theta_1)dw_t + c(X_{t-}, \theta_1)dJ_t$$

where $w = (w_t)_{t \in \mathbb{R}_+}$ is an r -dimensional Wiener process, and $J = (J_t)_{t \in \mathbb{R}_+}$ is an r_1 -dimensional pure-jump Lévy process with Lévy measure λ . The functions $a : \mathbb{R}^d \times \overline{\Theta}_2 \rightarrow \mathbb{R}^d$, $b : \mathbb{R}^d \times \overline{\Theta}_1 \rightarrow \mathbb{R}^d \otimes \mathbb{R}^r$ and $c : \mathbb{R}^d \times \overline{\Theta}_1 \rightarrow \mathbb{R}^d \otimes \mathbb{R}^{r_1}$ are supposed to satisfy certain regularity conditions. The parameter spaces Θ_1 and Θ_2 are bounded open sets in $\mathbb{R}^{p_1}$ and $\mathbb{R}^{p_2}$, respectively, each having a nice boundary.

In this section, our interest is in the phenomena when the Gaussian quasi-likelihood is applied to the Lévy driven stochastic differential equation. Nothing to say, this is a quasi-likelihood approach. This problem is of practical importance because implementation is easy with the Gaussian quasi-likelihood, as YUIMA (Brouste et al. 2014, Iacus and Yoshida 2018).

We assume that $|J_1| \in L^\infty$, $E[J_1] = 0$ and $E[J_1^{\otimes 2}] = I_{r_1}$, the r_1 -dimensional identity matrix. Let $V = bb^* + cc^*$ and assume the non-degeneracy of V . The fact that b and c share a common parameter θ_1 and the use of the function V suggest that only the variance structure of the increments will be paid attention in what follows.

We assume high frequency long-run data, that is, the data consists of $(X_{t_j})_{j=0,1,\dots,n}$ with $t_j = jh$, $h = h_n$, and $nh \rightarrow \infty$ and $nh^2 \rightarrow 0$ as $n \rightarrow \infty$. Let $\Theta = \Theta_1 \times \Theta_2$, $\theta = (\theta_1, \theta_2) \in \Theta$ and $p = p_1 + p_2$. For estimation of θ , we will work with the Gaussian quasi-likelihood

$$\begin{aligned} \mathbb{Q}_n(\theta) = & - \sum_{j=1}^n \left\{ \log \det V(X_{t_{j-1}}, \theta_1) \right. \\ & \left. + h^{-1} V(X_{t_{j-1}}, \theta_1)^{-1} [(\Delta_j X - ha(X_{t_{j-1}}, \theta_2))^{\otimes}] \right\}. \end{aligned}$$

A measurable mapping $\hat{\theta}_n = (\hat{\theta}_{1,n}, \hat{\theta}_{2,n})$ is called a Gaussian quasi-maximum likelihood estimator (GQMLE) if $\mathbb{Q}_n(\hat{\theta}_n) = \max_{\theta \in \overline{\Theta}} \mathbb{Q}_n(\theta)$. By a not straightforward application of the QLA theory, under ergodicity, Masuda (2013) showed the convergence

$$E[f((nh)^{1/2}(\hat{\theta}_n - \theta^*))] \rightarrow E[f(\xi^*)]$$

as $n \rightarrow \infty$ for any continuous function $f : \mathbb{R}^p \rightarrow \mathbb{R}$ of at most polynomial growth, where θ^* denotes the true value of θ , $\xi^* \sim N_p(0, \Sigma)$, and Σ is a $p \times p$ positive-definite matrix. Matrix Σ is not necessarily a block diagonal matrix, in other words, θ_1 and θ_2 are not necessarily orthogonal.

5 Quasi-likelihood analysis for volatility

5.1 Estimation of volatility

We treated inference for ergodic processes in Sects. 3 and 4. If the time horizon is finite, two probability measures for different values of the drift parameter can be absolutely continuous, and this means the drift parameter cannot be estimated consistently. Therefore, the diffusion parameter is only targeted by asymptotic statistics. Let us go back to the estimation problem for the unknown parameter θ in the m -dimensional semimartingale $Y = (Y_t)_{t \in [0, T]}$ having

the decomposition (1.1). We observe the data $(X_{t_j}, Y_{t_j})_{j=0,1,\dots,n}$ for $t_j = t_j^n = jT/n$ with fixed T , while the process b is unobservable. The parameter space Θ is supposed to be a p -dimensional bounded open set having a good boundary to admit Sobolev's embedding inequality.

Uchida and Yoshida (2013) constructed the QLA for the model (1.1). The quasi-log likelihood function $\mathbb{H}_n$ is defined by (1.2). Based on $\mathbb{H}_n$ of (1.2), the QMLE $\hat{\theta}_n^M$ is characterized by (2.9) with n in place of T , and the QBE $\hat{\theta}_n^B$ is defined by (2.12) with n for T . The true value of θ is denoted by θ^* . The information matrix at θ^* is defined by

$$\Gamma[u^{\otimes 2}] = \frac{1}{2T} \int_0^T \text{Tr}(S^{-1}(\partial_\theta S[u])S^{-1}(\partial_\theta S[u]))(X_t, \theta^*) dt \quad (5.1)$$

for $u \in \mathbb{R}^p$. The matrix Γ is symmetric and random because it involves the process X . We prepare the random field $\mathbb{Y} : \Theta \rightarrow \mathbb{R}$ as

$$\mathbb{Y}(\theta) = -\frac{1}{2T} \int_0^T \left\{ \log \frac{\det S(X_t, \theta)}{\det S(X_t, \theta^*)} + \text{Tr}(S(X_t, \theta)^{-1}S(X_t, \theta^*) - I_m) \right\} dt,$$

where I_m is the m -dimensional identity matrix. The key index to estimation is

$$\chi_0 = \inf_{\substack{\theta \in \Theta \\ \theta \neq \theta^*}} \frac{-\mathbb{Y}(\theta)}{|\theta - \theta^*|^2}.$$

If the key index χ_0 satisfies the non-degeneracy (2.1), i.e. $[T1]$ (i-2), then under mild regularity conditions,

$$E[f(\sqrt{n}(\hat{\theta}_n^A - \theta^*))\Phi] \rightarrow E[f(\Gamma^{-1/2}\zeta)\Phi] \quad (A \in \{M, B\}) \quad (5.2)$$

as $n \rightarrow \infty$ for any continuous function $f : \mathbb{R}^p \rightarrow \mathbb{R}$ of at most polynomial growth and any $\mathcal{F}$ -measurable random variable $\Phi \in \cup_{p>1} L^p$, where ζ is a p -dimensional standard Gaussian random vector independent of $\mathcal{F}$. This statistics is non-ergodic. Handy criteria for Condition (2.1) are available. See Uchida and Yoshida (2013) for details. The asymptotic mixed normality of the QMLE was presented by Genon-Catalot and Jacod (1993).

5.2 Jump filters

It is necessary to modify the estimating function (1.2) when the process Y has a jump component. Instead of (1.1), we will consider a semimartingale Y having the decomposition

$$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma(X_s, \theta) dw_s + J_t, \quad t \in [0, T] \quad (5.3)$$

and estimate θ from the data $(X_{t_j}, Y_{t_j})_{j=0,1,\dots,n}$. The jump component $J = (J_t)_{t \in [0, T]}$ is a random step process. As discussed in Sect. 4.1, we need some jump filter. The classical filter $|\Delta_j Y| > Dh^\rho$ is a possibility. However, it is known that the performance of the classical filter strongly depends on the tuning parameters; see e.g. Iacus and Yoshida (2018). Recently, Inatsugu and Yoshida (2021a) (updated by arXiv:1806.10706 (2018)) proposed global jump filters to enable stable and precise estimation of the volatility parameter θ . The global filter uses the order statistics of the increments. Let $V_j = |(\bar{S}_{n,j-1})^{-1/2} \Delta_j Y|$ with an initial estimator $\bar{S}_{n,j-1}$ for the spot volatility $S(X_{t_{j-1}}, \theta^*)$ up to a possibly unknown scaling constant. The r -th order statistic of $\{V_j\}_{j=1,\dots,n}$ is denoted by $V_{(r)}$. For a preset constant

$\alpha \in (0, 1)$, the global jump filter is specified by the index set

$$\mathcal{J}_n(\alpha) = \{j \in \{1, \dots, n\}; V_j < V_{(s_n(\alpha))}\}$$

with $s_n(\alpha) = \lfloor n(1 - \alpha) \rfloor$. Then the α -quasi-log likelihood function is given by

$$\begin{aligned} \mathbb{H}_n(\theta; \alpha) = & -\frac{1}{2} \sum_{j \in \mathcal{J}_n(\alpha)} \left\{ q(\alpha)^{-1} h^{-1} S(X_{t_{j-1}}, \theta)^{-1} [(\Delta_j Y)^{\otimes 2}] K_{n,j} \right. \\ & \left. + p(\alpha)^{-1} \log \det S(X_{t_{j-1}}, \theta) \right\}. \end{aligned} \quad (5.4)$$

In Formula (5.4), $p(\alpha) = 1 - \alpha$, and the constant $q(\alpha)$ is determined by $q(\alpha) = m^{-1} E[V 1_{\{V \leq c(\alpha)\}}]$, where the constant $c(\alpha)$ is the upper α -quantile of the chi-square distribution of degree m , i.e., $P[V \leq c(\alpha)] = 1 - \alpha$ for a random variable $V \sim \chi^2(m)$. The cap function $K_{n,j} = 1_{\{|\Delta_j Y| < C_* n^{-1/4}\}}$ with positive constant C_* . This function is a very loose filter only for removing technical assumptions on the distribution of the jumps ΔJ_t . Practically $K_{n,j}$ will be almost always 1. A measurable mapping $\hat{\theta}_n^{M,\alpha}$ maximizing $\mathbb{H}_n(\theta; \alpha)$ in $\theta \in \Theta$ is called an α -quasi-maximum likelihood estimator (α -QMLE). Inatsugu and Yoshida (2021a) (updated by arXiv:1806.10706 (2018)) gave a rate of convergence in L^p -norm of the α -QMLE $\hat{\theta}_n^{M,\alpha}$ with the help of an annealed quasi-likelihood ratio process and a resulting polynomial type large deviation inequality. A quasi-Bayesian estimator is also treated there. Moreover, they introduced the QMLE $\hat{\theta}_n^{M,\alpha_n}$ and the QBE $\hat{\theta}_n^{B,\alpha_n}$ with a shrinking α_n to obtain the same convergence as (5.2), that is,

$$E[f(\sqrt{n}(\hat{\theta}_n^{A,\alpha_n} - \theta^*))\Phi] \rightarrow E[f(\Gamma^{-1/2}\zeta)\Phi] \quad (A \in \{M, B\}) \quad (5.5)$$

as $n \rightarrow \infty$ for any continuous function $f: \mathbb{R}^p \rightarrow \mathbb{R}$ of at most polynomial growth and any $\mathcal{F}$ -measurable random variable $\Phi \in \cup_{p \geq 1} L^p$, where ζ is a p -dimensional standard Gaussian random vector independent of $\mathcal{F}$. The global filter is applied to the realized volatility by Inatsugu and Yoshida (2021b). The realized volatility with a global jump filter outperforms the bipower variation and the minimum realized volatility, that were regarded as estimators robust against jumps.

5.3 Non-synchronous observations

Different components of high frequency data often have asynchronous timestamps. A seemingly natural idea for estimation of the covariance between two components is to use the ordinary realized co-volatility after synchronizing the data by some interpolation method. However, it is known as the Epps effect (Epps 1979) that any such interpolation causes a severe bias and the estimated correlation disappears when the frequency of the observations diverges. This problem of measurement of multivariate volatilities was solved by Malliavin and Mancino (2002) with a Fourier series method and by Hayashi and Yoshida (2005) with an association kernel. These estimators are nonparametric.

Ogihara and Yoshida (2014) studied parametric estimation of volatility with non-synchronous data. Consider the model (1.1). Suppose that Y is two-dimensional, X is possibly multi-dimensional and different components in (Y, X) are observed in a non-synchronous manner. The quasi-likelihood function is based on local Gaussian approximation but a global quadratic form having many off-diagonal nonzero elements appears due to non-synchronicity. Though theoretical treatment is fairly complicated, it is possible to construct a QLA, and consequently to prove asymptotic mixed normality and moments convergence of QMLE and

QBE. The nonsynchronous covariance estimator (H-Y estimator) has a central limit theorem (Hayashi and Yoshida 2008, 2011). It is said that the H-Y estimator attains the minimum asymptotic variance among nonparametric estimators. If we consider a simple model having a constant diffusion matrix, then we can compare the two estimates of covariance between the two components of Y , one is obtained as a product of estimates by the QMLE and another one from the H-Y estimator. It is shown that the parametric estimator achieves better precision, as expected since it uses information about the structure of the model. Ogihara (2015) proved the local asymptotic mixed normality for a non-synchronously observed diffusion process, and concluded that the QMLE and QBE are asymptotically optimal.

6 Model selection

Eguchi and Masuda (2018) applied the QLA to a Schwarz type model selection criterion for stochastic processes. In short, they considered the integral

$$I_n = \int_{\Theta} \exp(\mathbb{H}_n(\theta)) \varpi_n(\theta) d\theta \quad (6.1)$$

for a quasi-log likelihood random field $\mathbb{H}_n : \Omega \times \Theta \rightarrow \mathbb{R}$, $n \in \mathbb{N}$, given a probability space $(\Omega, \mathcal{F}, P)$. The parameter space Θ is a bounded open set in $\mathbb{R}^p$, and ϖ_n is a prior density on Θ . The technical essence of the quasi-Bayesian information criterion (QBIC) is to validate an L^1 -approximation of $F_n = -2 \log I_n$ by the statistic $\text{QBIC}_n^\natural$ given by

$$\text{QBIC}_n^\natural = -2\mathbb{H}_n(\hat{\theta}_n) + \log \det(-\partial_{\theta}^2 \mathbb{H}_n(\hat{\theta}_n)) - 2 \log \varpi_n(\hat{\theta}_n) - p \log(2\pi),$$

where $\hat{\theta}_n$ is a QMLE for θ . The expression (6.1) of I_n is useful for the estimate since

$$F_n = -2\mathbb{H}_n(\theta^*) - 2 \log |\det a_n| - 2 \log \int_{\mathbb{U}_n} \mathbb{Z}_n(u) \varpi_n(\theta^* + a_n u) du.$$

Then it is possible to estimate the L^1 -norm of $F_n - \text{QBIC}_n^\natural$ by using a polynomial type large deviation inequality for $\mathbb{Z}_n$. Finally, they reached the QBIC_n defined by

$$\text{QBIC}_n = -2\mathbb{H}_n(\hat{\theta}_n) + \log \det(-\partial_{\theta}^2 \mathbb{H}_n(\hat{\theta}_n)).$$

The QBIC is valid for non-ergodic models, not only for ergodic models. Eguchi and Masuda (2018) showed that QBIC performs well for volatility model selection and selection of ergodic diffusion models. The results are practically promising as well as showing effectiveness of the QLA theory. Further considerations may be possible in some philosophical and technical aspects.

In the context of information criteria for semimartingales, the QLA was first applied in Uchida (2010) by Uchida, who treated the exact likelihood of a sampled diffusion process by means of the Malliavin calculus to derive the contrast based information criterion (CIC).

The QLA found another application in Umezū et al. (2019) to AIC for the non-concave penalized likelihood method. Sparse estimation is a new direction of the QLA theory. The polynomial type large deviation (PLD) inequality for a random field $\mathbb{H}_T$ having the LAQ property can be transferred to a PLD inequality for the penalized random field $\mathbb{H}_T^\sharp = \mathbb{H}_T -$ (penalty terms) (Kinoshita and Yoshida 2019). Therefore, the PLD inequality is basic even in the theory of sparse estimation for stochastic processes. Related papers are Masuda and Shimizu (2017) and Suzuki and Yoshida (2020). Partial quasi-likelihood analysis is another direction of extension of the theory (Yoshida 2018).

7 Point processes

Recently, modeling with point processes is attracting attention in applications of ultra-high frequency financial data (cf. Abergel and Jedidi 2015). Multi-dimensional point processes are used to model the limit order book.

On a stochastic basis $(\Omega, \mathcal{F}, \mathbf{F}, P)$, $\mathbf{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$, we consider a d -dimensional point process $N = (N_t)_{t \in \mathbb{R}_+} = (N_t^\alpha)_{\alpha \in \mathbb{I}, t \in \mathbb{R}_+}$, $\mathbb{I} = \{1, \dots, d\}$. More precisely, $N_0 = 0$ and each $N^\alpha = (N^\alpha_t)_{t \in \mathbb{R}_+}$ is a simple counting process. We will assume that the components N^α ($\alpha \in \mathbb{I}$) have not common jump. Further, we assume the $\mathbf{F}$ -compensator of N has an intensity process $\lambda^* = (\lambda^*_t)_{t \in \mathbb{R}_+}$. That is, λ^* is a nonnegative locally integrable predictable process such that $N - \int_0^\cdot \lambda^*_s ds$ is a d -dimensional local martingale with respect to $\mathbf{F}$.

For statistical modeling of this point process, we consider a random field $\lambda : \Omega \times \mathbb{R}_+ \times \overline{\Theta} \rightarrow \mathbb{R}_+$. The set Θ is a bounded open set in $\mathbb{R}^p$ as usual. Suppose that the map $\mathbb{R}_+ \ni t \mapsto \lambda(t, \theta) = (\lambda^\alpha(t, \theta))_{\alpha \in \mathbb{I}}$ is an $\mathbb{R}_+^d$ -valued left-continuous adapted process for every $\theta \in \overline{\Theta}$, and the map $\Theta \ni \theta \mapsto \lambda(s, \theta)$ is of class $C^3(\Theta)$ for every $s \in \mathbb{R}_+$. We refer the reader to Clinet and Yoshida (2017) for other regularity conditions and details of this section. We assume that the $\lambda^*_t = \lambda(t, \theta^*)$ for some $\theta^* \in \Theta$. Statistically, the value θ^* is unknown and to be estimated from the data $(X_t)_{t \in [0, T]}$. The quasi-log likelihood function for estimation of θ is

$$\ell_T(\theta) = \sum_{\alpha \in \mathbb{I}} \left\{ \int_0^T \log \lambda^\alpha(t, \theta) dN_t^\alpha - \int_0^T \lambda^\alpha(t, \theta) dt \right\}.$$

Let $\mathbf{E} = \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^p$. For ergodicity, we assume that for each $\alpha \in \mathbb{I}$, there exists a mapping $\pi_\alpha : C_b(\mathbf{E}) \times \overline{\Theta} \rightarrow \mathbb{R}$ such that

$$\frac{1}{T} \int_0^T \psi(\lambda^\alpha(t, \theta^*), \lambda^\alpha(t, \theta), \partial_\theta \lambda^\alpha(t, \theta)) dt \xrightarrow{P} \pi_\alpha(\psi, \theta) \quad (T \rightarrow \infty)$$

for every $(\psi, \theta) \in C_b(\mathbf{E}) \times \overline{\Theta}$. Then, under an identifiability condition, Clinet and Yoshida (2017) showed asymptotic normality and moments convergence for QMLE and QBE with respect to ℓ_T . These results were applied to a multivariate Hawkes process, for which ergodicity can be verified. Today the Hawkes process is one of standard models in analysis of limit order books.

Muni Toke and Yoshida discussed about modeling intensities of order flows in a limit order book in Muni Toke and Yoshida (2017), analysis of order flows in limit order books with ratios of Cox-type intensities in Muni Toke and Yoshida (2019), and marked point processes and intensity ratios for limit order book modeling in Muni Toke and Yoshida (2020). QLA and information criteria for stochastic processes were applied. Flexible modeling is possible thanks to the formulation of the QLA. The resulting models incorporate various effective covariates. This approach enables us to predict the next market order more precisely than the traditional models.

Acknowledgements The author thanks Professor Yuri Kutoyants for the invitation to the special issue and his valuable comments. This work was in part supported by Japan Science and Technology Agency CREST JPMJCR14D7, JPMJCR2115; Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research No. 17H01702 (Scientific Research); and by a Cooperative Research Program of the Institute of Statistical Mathematics.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence,

and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abergel F, Jedidi A (2015) Long-time behavior of a Hawkes process-based limit order book. *SIAM J Financ Math* 6(1):1026–1043
- Brouste A, Fukasawa M, Hino H, Iacus S, Kamatani K, Koike Y, Masuda H, Nomura R, Ogihara T, Shimuzu Y, Uchida M, Yoshida N (2014) Statistical inference for stochastic processes: overview and prospects. *J Stat Softw* 57(4):1–51
- Clinet S, Yoshida N (2017) Statistical inference for ergodic point processes and application to limit order book. *Stoch Process Their Appl* 127(6):1800–1839
- Dabye AS, Gounoung AA, Kutoyants YA (2018) Method of moments estimators and multi-step MLE for Poisson processes. *J Contemp Math Anal (Armenian Acad Sci)* 53(4):237–246
- Eguchi S, Masuda H (2018) Schwarz type model comparison for LAQ models. *Bernoulli* 24(3):2278–2327
- Epps T (1979) Comovements in stock prices in the very short run. *JASA* 74:291–298
- Genon-Catalot V, Jacod J (1993) On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann Inst H Poincaré Probab Statist* 29(1):119–151
- Hayashi T, Yoshida N (2005) On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* 11(2):359–379
- Hayashi T, Yoshida N (2008) Asymptotic normality of a covariance estimator for nonsynchronously observed diffusion processes. *Ann Inst Stat Math* 60(2):367–406. <https://doi.org/10.1007/s10463-007-0138-0>
- Hayashi T, Yoshida N (2011) Nonsynchronous covariation process and limit theorems. *Stoch Process Appl* 121(10):2416–2454. <https://doi.org/10.1016/j.spa.2010.12.005>
- Iacus SM, Yoshida N (2018) Simulation and inference for stochastic processes with YUIMA. Springer, Berlin
- Ibragimov I, Khas' minskii R (1973) Asymptotic behavior of statistical estimators in the smooth case. I. Study of the likelihood ratio. *Theory Probab Appl* 17(3):445–462
- Ibragimov I, Khas' minskii R (1973) Asymptotic behavior of some statistical estimators II. limit theorems for the a posteriori density and Bayes' estimators. *Theory Probab Appl* 18(1):76–91
- Ibragimov IA, Has' minskii RZ (1981) Statistical estimation, Applications of Mathematics, vol. 16. Springer-Verlag, New York. Asymptotic theory, Translated from the Russian by Samuel Kotz
- Inatsugu H, Yoshida N (2021a) Global jump filters and quasi likelihood analysis for volatility. *Annals of the Institute of Statistical Mathematics, on-line*
- Inatsugu H, Yoshida N (2021b) Global jump filters and realized volatility. [arXiv:2102.05307](https://arxiv.org/abs/2102.05307)
- Kamatani K, Uchida M (2014) Hybrid multi-step estimators for stochastic differential equations based on sampled data. *Stat Inference Stoch Process* 18(2):177–204
- Kessler M (1995) Estimation des paramètres d'une diffusion par des contrastes corrigés. *C R Acad Sci Paris Sér I Math* 320(3):359–362
- Kessler M (1997) Estimation of an ergodic diffusion from discrete observations. *Scand J Statist* 24(2):211–229
- Kinoshita Y, Yoshida N (2019) Penalized quasi likelihood estimation for variable selection. *arXiv preprint arXiv:1910.12871*
- Kutoyants YA (1984) Parameter estimation for stochastic processes. vol. 6. Heldermann. Translated from the Russian and edited by B. L. S. Prakasa Rao
- Kutoyants YA (1994) Identification of dynamical systems with small noise, Mathematics and its Applications, vol 300. Kluwer Academic Publishers Group, Dordrecht
- Kutoyants YA (2004) Statistical inference for ergodic diffusion processes. Springer Series in Statistics. Springer-Verlag London Ltd., London
- Kutoyants YA (2012) Statistical inference for spatial Poisson processes, vol 134. Springer Science & Business Media, Berlin
- Kutoyants YA (2014) Approximation of the solution of the backward stochastic differential equation. Small noise, large sample and high frequency cases. *Proc Steklov Inst Math* 287(1):133–154
- Kutoyants YA (2017) On the multi-step MLE-process for ergodic diffusion. *Stoch Process Their Appl* 127(7):2243–2261
- Le Cam L (1960) Locally asymptotically normal families of distributions. *Univ California Publ Statist* 3:37–98

- Malliavin P, Mancino ME (2002) Fourier series method for measurement of multivariate volatilities. *Finance Stoch* 6(1):49–61
- Masuda H (2013) Convergence of Gaussian quasi-likelihood random fields for ergodic Lévy driven sde observed at high frequency. *Ann Stat* 41(3):1593–1641
- Masuda H, Shimizu Y (2017) Moment convergence in regularized estimation under multiple and mixed-rates asymptotics. *Math Methods Stat* 26(2):81–110
- Muni Toke I, Yoshida N (2017) Modelling intensities of order flows in a limit order book. *Quant Finance* 17(5):683–701
- Muni Toke I, Yoshida N (2019) Analyzing order flows in limit order books with ratios of Cox-type intensities. *Quantitative Finance* pp 1–18
- Muni Toke I, Yoshida N (2020) Marked point processes and intensity ratios for limit order book modeling. [arXiv:2001.08442](https://arxiv.org/abs/2001.08442)
- Nualart D, Yoshida N (2019) Asymptotic expansion of Skorohod integrals. *Electron J Probab* 24:1–64
- Ogihara T (2015) Local asymptotic mixed normality property for nonsynchronously observed diffusion processes. *Bernoulli* 21(4):2024–2072
- Ogihara T, Yoshida N (2011) Quasi-likelihood analysis for the stochastic differential equation with jumps. *Stat Inference Stoch Process* 14(3):189–229. <https://doi.org/10.1007/s11203-011-9057-z>
- Ogihara T, Yoshida N (2014) Quasi-likelihood analysis for nonsynchronously observed diffusion processes. *Stoch Process Their Appl* 124(9):2954–3008
- Prakasa Rao B (1988) Statistical inference from sampled data for stochastic processes. *Statistical inference from stochastic processes* (Ithaca, NY, 1987) **80**, 249–284
- Prakasa Rao BLS (1983) Asymptotic theory for nonlinear least squares estimator for diffusion processes. *Math Operationsforsch Stat Ser Stat* 14(2):195–209
- Shimizu Y, Yoshida N (2006) Estimation of parameters for diffusion processes with jumps from discrete observations. *Stat Inference Stoch Process* 9(3):227–277. <https://doi.org/10.1007/s11203-005-8114-x>
- Suzuki T, Yoshida N (2020) Penalized least squares approximation methods and their applications to stochastic processes. *Japanese J Stat Data Sci* 3(2):513–541
- Uchida M (2010) Contrast-based information criterion for ergodic diffusion processes from discrete observations. *Ann Inst Stat Math* 62(1):161–187
- Uchida M, Yoshida N (2012) Adaptive estimation of an ergodic diffusion process based on sampled data. *Stoch Process Their Appl* 122(8):2885–2924
- Uchida M, Yoshida N (2013) Quasi likelihood analysis of volatility and nondegeneracy of statistical random field. *Stoch Process Appl* 123(7):2851–2876. <https://doi.org/10.1016/j.spa.2013.04.008>
- Uchida M, Yoshida N (2014) Adaptive Bayes type estimators of ergodic diffusion processes from discrete observations. *Stat Inference Stoch Process* 17(2):181–219
- Umezū Y, Shimizu Y, Masuda H, Ninomiya Y (2019) AIC for the non-concave penalized likelihood method. *Ann Inst Stat Math* 71(2):247–274
- Yoshida N (1990) Asymptotic behavior of M-estimator and related random field for diffusion process. *Ann Inst Stat Math* 42(2):221–251
- Yoshida N (1992) Estimation for diffusion processes from discrete observation. *J Multivariate Anal* 41(2):220–242
- Yoshida N (2011) Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Ann Inst Stat Math* 63(3):431–479
- Yoshida N (2013) Martingale expansion in mixed normal limit. *Stoch Process Their Appl* 123(3):887–933
- Yoshida N (2018) Partial quasi-likelihood analysis. *Japanese J Stat Data Sci* 1(1):157–189
- Yoshida N (2020) Asymptotic expansion of a variation with anticipative weights. [arXiv:2101.00089](https://arxiv.org/abs/2101.00089)
- Yoshida N (2021) Simplified quasi-likelihood analysis for a locally asymptotically quadratic random field. [arXiv:2102.12460](https://arxiv.org/abs/2102.12460)