Ann Inst Stat Math (2012) 64:193–211
DOI 10.1007/s10463-010-0307-4

Local asymptotic mixed normality for discretely observed non-recurrent Ornstein–Uhlenbeck processes

Yasutaka Shimizu

Received: 26 May 2009 / Revised: 5 November 2009 / Published online: 14 July 2010

Abstract Consider non-recurrent Ornstein–Uhlenbeck processes with unknown drift and diffusion parameters. Our purpose is to estimate the parameters jointly from discrete observations with a certain asymptotics. We show that the likelihood ratio of the discrete samples has the uniform LAMN property, and that some kind of approximated MLE is asymptotically optimal in a sense of asymptotic maximum concentration probability. The estimator is also asymptotically efficient in ergodic cases.

Keywords Ornstein–Uhlenbeck processes · Non-recurrency · ULAMN property · Discrete observations · Joint estimation · Asymptotic optimality

1 Introduction

On a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0})$, we consider 1-dim Ornstein–Uhlenbeck (OU) processes X given by the following SDEs:

$$\mathrm{d}X_t = \mu X_t\,\mathrm{d}t + \sqrt{\sigma}\mathrm{d}W_t, \quad X_0 = x, \tag{1}$$

where W is a Wiener process, $\vartheta = (\mu, \sigma)$ is a parameter which values as $\mu \in int(\Xi)$ and $\sigma \in int(\Pi)$, where Ξ and Π are compact convex subsets of $\mathbb{R}$ and $(0, \infty)$, respectively. We denote by $\Theta = \Xi \times \Pi$. The properties of OU processes have been well

Y. Shimizu (✉)
Graduate School of Engineering Science, Osaka University, 1-3, Machikaneyama-cho, Toyonaka, Osaka 560-8531, Japan
e-mail: yasutaka@sigmath.es.osaka-u.ac.jp

Y. Shimizu
Japan Science and Technology Agency, CREST, Sanbancho, Chiyoda-ku, Tokyo 102-0075, Japan

studied by many authors, and it is well known that X is positive recurrent (ergodic) if $\mu \in (-\infty, 0)$; null-recurrent if $\mu = 0$; non-recurrent if $\mu \in (0, \infty)$.

When we observe X at discrete time points $\{t_i\}_{i=0}^n$ with $t_i = i\Delta$ for some $\Delta > 0$, estimation of ϑ from observations $\{X_i := X_{t_i}\}_{i=0}^n$ is one of the most fundamental problems in statistical inference for SDEs. Actually, there exist many works on inference for ergodic diffusion processes from discrete samples. However, for non-ergodic diffusions, there are only a few works for discretely observed cases; see Kasonga (1990), Jacod (2006) and Shimizu (2009a), except for continuously observed cases; see e.g., a monograph by Prakasa Rao (1999a,b) and Kutoyants (2004). The aim of this paper is to construct an optimal estimator of ϑ from $\{X_i\}_{i=0}^n$ under the following assumption (H): *a high-frequent sampling in a long term*.

Assumption (H) $\Xi \subset (0, \infty)$; $\Delta \to 0$ and $n\Delta \to \infty$ as $n \to \infty$.

In the OU-case, the transition probability from X_{i-1} to X_i is explicitly known: under P_ϑ,

$$X_i = e^{\mu\Delta} X_{i-1} + \sqrt{\sigma}\epsilon_i^\Delta(\mu) \quad (i = 1, 2, \ldots, n),$$
$$\text{where } \epsilon_i^\Delta(s) := e^{t_i\mu} \int_{t_{i-1}}^{t_i} e^{-\mu s}\, \mathrm{d}W_s \sim N\left(0, \frac{1}{2\mu}(e^{2\Delta\mu} - 1)\right).$$

When $\Delta > 0$ is fixed, the sequence $\{X_i\}_{i=0}^n$ is regarded as an AR(1)-time series:

$$X_i = \alpha X_{i-1} + u_i, \quad \text{where } \alpha := e^{\mu\Delta}$$

and u_i's are I.I.D. normal innovations. A little different point in OU series from the context of usual time series is that u_i also depends on α. Unless u_i depends on α: the usual AR(1), the rigorous MLE of α is given by

$$\hat{\alpha}_n := \arg\min_{\alpha \in \mathbb{R}} \sum_{i=1}^n \left|X_{t_i} - \alpha X_{i-1}\right|^2 = \frac{\sum_{i=1}^n X_i X_{i-1}}{\sum_{i=1}^n X_{i-1}^2}. \tag{2}$$

Parameter α determines the stability of $\{X_i\}$. The non-recurrent OU cases correspond to the cases where $\alpha > 1$: non-stationary AR(1). In this case, Anderson (1959) studied the rate of convergence of $\hat{\alpha}_n$, which is α^n-order. On the other hand, when $\Delta \to 0$ as $n \to \infty$: $\alpha \to 1$, the situation is similar to an AR(1) model with a root near unity, discussed by Phillips (1987). He considered the situation such as $\alpha = e^{\mu/n}$. Therefore his argument corresponds to the case where $\Delta = 1/n$: $n\Delta$ is fixed in our notation. He showed that $n(\hat{\alpha}_n - \alpha) \xrightarrow{\mathcal{D}} \beta(\mu)$ as $n \to \infty$ for some $\beta(\mu)$, and considered the asymptotic behavior of $\beta(\mu)$ as $|\mu| \to \infty$ in order. However we are interested in the case where $\Delta \to 0$ and $n\Delta \to \infty$, which corresponds to the case where $n \to \infty$ and, at the same time, $\mu \to \infty$ in terms of Phillips (1987) setting.

Suppose that the process X is observed time-continuously on $[0, T]$, where only μ is the target of estimation since σ is estimated consistently by computing the quadratic variation of X in a local time interval. The MLE of μ is given by

$$\hat{\mu}_T^{\mathrm{MLE}} := \frac{\int_0^T X_s\,\mathrm{d}X_s}{\int_0^T X_s^2\,\mathrm{d}s} = \frac{X_T^2 - x^2 - T}{2\int_0^T X_s^2\,\mathrm{d}s},$$

which is the maximizer of the log-likelihood function $-\frac{1}{2}\int_0^T \mu^2 X_s^2\,\mathrm{d}s + \int_0^T \mu X_s\,\mathrm{d}X_s$. As is well known, if $\mu > 0$, this estimator is asymptotically mixed normal with exponential rate of convergence; see Feigin (1976), or Dietz and Kutoyants (2003): as $T \to \infty$,

$$e^{\mu T}(\hat{\mu} - \mu) \xrightarrow{\mathcal{D}} v_\mu^{-1/2} Z \quad \text{with } v_\mu := \frac{1}{2\mu\sigma}\left(x + \sqrt{\sigma}\xi_\mu\right)^2, \tag{3}$$

where $\xi_\mu = \int_0^\infty e^{-\mu s}\,\mathrm{d}W_s$ and Z is a standard normal random variable independent of ξ_μ. On the other hand, we have only discrete samples $\{X_i\}_{i=1}^n$, and the following approximated MLEs are often used:

$$\hat{\mu}_n^{\mathrm{AMLE}} := \frac{\sum_{i=1}^n X_i(X_i - X_{i-1})}{\Delta\sum_{i=1}^n X_{i-1}^2}, \text{ or } \frac{X_n^2 - x^2 - n\Delta}{2\Delta\sum_{i=1}^n X_{i-1}^2}.$$

These are discrete versions of $\hat{\mu}_T^{\mathrm{MLE}}$ replaced $\mathrm{d}X_s$ with $(X_i - X_{i-1})$, and $\int_0^T X_s^2\,\mathrm{d}s$ with $\Delta\sum_{i=1}^n X_{i-1}^2$. The former estimator is also obtained via the least squares method or a contrast function due to a discretization of the continuous time likelihood function; see, e.g., Le Breton (1975); Kasonga (1988) and Prakasa Rao (1999b). If X is ergodic, it is known that, under the asymptotics such as $\Delta \to 0$, $n\Delta \to \infty$ and $n\Delta^3 \to 0$, both of $\hat{\mu}_n^{\mathrm{AMLE}}$'s are asymptotically efficient in the sense of the minimal asymptotic variance with $\sqrt{n\Delta}$-rate of convergence; see Shimizu (2009b). See also Kessler (1997) and Gobet (2002) for more general results for diffusion processes. If X is non-recurrent, one may also expect that either estimator attains $e^{\mu n\Delta}$-rate of convergence. However the answer is negative: $e^{\mu n\Delta}(\hat{\mu}_n^{\mathrm{AMLE}} - \mu) \to \infty$ although $\sqrt{n\Delta}(\hat{\mu}_n^{\mathrm{AMLE}} - \mu)$ is tight. This is due to the rough approximations of $\mathrm{d}X_s$ and $\int_0^T X_s^2\,\mathrm{d}s$, and it indicates that the often used contrast functions based on the local Gauss approximation of the likelihoods is inadequate in non-ergodic cases: see Shimizu (2009b).

Note that the likelihood function of $(X_1, \ldots, X_n)$ is written explicitly:

$$\exp\left(-\sum_{i=1}^n \frac{\mu\left(X_i - e^{\mu\Delta}X_{i-1}\right)^2}{\sigma(e^{2\mu\Delta} - 1)} - \sum_{i=1}^n \frac{1}{2}\log\left(\frac{e^{2\mu\Delta} - 1}{2\mu}\sigma\right)\right).$$

However, the rigorous MLE of μ cannot be written explicitly. In view of (2), it may be better to use the following estimator:

$$\hat{\mu}_n := \arg\min_{\nu\in\Xi}\sum_{i=1}^n \left|X_i - e^{\nu\Delta}X_{i-1}\right|^2, \tag{4}$$

which is well-defined since the parameter space Ξ is compact. In particular, if $\hat{\mu}_n$ is a local minimum in Ξ, the explicit form is as follows:

$$\hat{\mu}_n = \frac{1}{\Delta} \log\left(\frac{\sum_{i=1}^n X_i X_{i-1}}{\sum_{i=1}^n X_{i-1}^2} \right). \tag{5}$$

This estimator is not new but a trajectory-fitting estimator (TFE) proposed by Kasonga (1988), and the weak consistency has already been shown. We further show that it attains $e^{\mu n \Delta}$-rate of convergence with the same asymptotic distribution as in (3) without any restriction except for (H).

In the discretely observed cases, the diffusion coefficient σ is also the target of statistical estimation, unlike the continuously observed cases. In OU cases, we can directly compute the maximum likelihood estimator for σ. We shall show the rate of convergence is $\sqrt{n}$ which is the same as in the ergodic cases. The rate would be natural since the diffusion coefficient is due to a local characteristic of the quadratics of X which is nothing to do with the ergodicity which is a global property. We also show that our estimators of μ and σ are optimal in some class of estimators in the sense of the maximum concentration at all $\vartheta \in \Theta$ in probability by showing that the likelihood ratio of $\{X_i\}_{i=0}^n$ is *locally asymptotically mixed normal (LAMN).*

Of course, we also remark that the proposed estimator is also asymptotically efficient if X is ergodic: see (11) and Corollary 1 below. Therefore, one should use our estimator even if X is ergodic or non-ergodic under the discrete sampling.

In the next section, we describe the main results. All the proofs are presented in Sect. 3.

2 Main results

2.1 Notation

Let us prepare further notation:

1. P_ϑ: the induced measure of $\mathbb{P}$ by X with $\vartheta = (\mu, \sigma)$; E_ϑ: the expectation w.r.t P_ϑ;
2. $\nabla := (\partial_\mu, \partial_\sigma)^\top$, where $\partial_x := \partial/\partial x$, and $\top$ is the transpose;
3. For matrices A_n and B_n, $A_n \sim B_n$ means that $B_n^{-1} A_n$ tends to the identity matrix;
4. For a matrix $A = (a_{ij})_{i,j=1}^d$, $\mathrm{diag}(A)$ stands for $\mathrm{diag}(a_{11}, a_{22}, \ldots, a_{dd})$;
5. $E_s^\Delta := e^{2s\Delta} - 1$; $D_i^\Delta(s) := X_i - e^{s\Delta} X_{i-1}$. Note that $D_i^\Delta(\mu) = \sqrt{\sigma}\epsilon_i^\Delta(\mu)$ under P_ϑ:

$$\sigma^{-1/2} D_i^\Delta(\mu) = \epsilon_i^\Delta(\mu) \sim N\left(0, E_\mu^\Delta / 2\mu\right);$$

6. $\ell_n(\vartheta)$ is the exact log-likelihood function of $\{X_i\}_{i=1}^n$, which is written as

$$\ell_n(\vartheta) := -\sum_{i=1}^{n}\left[\frac{\mu|D_i^\Delta(\mu)|^2}{\sigma E_\mu^\Delta} + \frac{1}{2}\log\left(\frac{\sigma E_\mu^\Delta}{2\mu}\right)\right];$$

7. $\hat{\vartheta}_n := \left(\hat{\mu}_n, \hat{\sigma}_n\right)^\top$, where $\hat{\mu}_n$ is given in (4), and

$$\hat{\sigma}_n := \hat{\sigma}_n(\hat{\mu}_n) \quad \text{with } \hat{\sigma}_n(s) := \frac{2s}{nE_s^\Delta}\sum_{i=1}^{n}|D_i^\Delta(s)|^2.$$

   $\hat{\sigma}_n(\mu)$ is the MLE given the true μ: $\hat{\sigma}_n(\mu) = \arg\max_{\sigma\in\Pi}\ell_n(\mu,\sigma)$.
8. $S_n(\vartheta) := \nabla_\vartheta \ell_n(\vartheta)$: *score vectors*; $B_n(\vartheta) := -\nabla_\vartheta\nabla_\vartheta^\top \ell_n(\vartheta)$: *observed information matrices*; $I_n(\vartheta) := E_\vartheta\left[\mathrm{diag}(B_n(\vartheta))\right]$: diagonalize *expected information matrices*;
9. $L_n(\vartheta) := I_n^{-1/2}(\vartheta)S_n(\vartheta)$ and $G_n(\vartheta) := I_n^{-1/2}(\vartheta)B_n(\vartheta)I_n^{-1/2}(\vartheta)$: the normalized versions of scores and observed information, respectively;
10. Let $\vartheta_n^*(h) := \vartheta + I_n^{-1/2}(\vartheta)h$ for $h \in \mathbb{R}^2$, and denote by

$$\Lambda_n^\vartheta(h) := \ell_n(\vartheta_n^*(h)) - \ell_n(\vartheta);$$
$$r_n(h,\vartheta) := \Lambda_n^\vartheta(h) - \left\{h^\top L_n(\vartheta) - \frac{1}{2}h^\top G_n(\vartheta)h\right\}$$

   $\Lambda_n^\vartheta := (\Lambda_n^\vartheta(h))_{h\in\mathbb{R}^2}$ is the log-likelihood ratio random field, and $r(h,\vartheta)$ is the remainder of a quadratic approximation of Λ_n^ϑ.

The following LAMN property for the likelihood ratio process is important in non-ergodic statistics; see Basawa and Scott (1983), Jeganathan (1982) and Luschgy (1992) for details. The definition of *uniformity* of the LAMN below is due to Basawa and Scott (1983).

Definition 1 The random field Λ_n^ϑ is locally asymptotically mixed normal (LAMN) at $\vartheta \in \Theta$ if the following two conditions are satisfied: as $n \to \infty$,

(A.1) there exists an almost-surely positive definite random matrix $G(\vartheta)$ such that

$$(L_n(\vartheta), G_n(\vartheta)) \xrightarrow{\mathcal{D}} (G^{1/2}(\vartheta)Z, G(\vartheta)) \quad \text{under } P_\vartheta,$$

where Z is a standard normal vector independent of $G(\vartheta)$;

(A.2) $r_n(h,\vartheta) \xrightarrow{P} 0$ under P_ϑ for any $h \in \mathbb{R}^2$.

If (A.2) holds for any bounded sequence $(h_n) \subset \mathbb{R}^2$: $r_n(h_n,\vartheta) \xrightarrow{P} 0$, then we say that the random field Λ_n^ϑ is "uniformly" LAMN (ULAMN).

If the matrix $G(\vartheta)$ is deterministic, then Λ_n^ϑ is called *locally asymptotically normal* (*LAN*). For details on the LAN theory, see, e.g., Ibragimov and Has'minskii (1981).

Definition 2 A sequence of estimators $\{T_n\}$ is "regular at ϑ" if there exists a random variable $T(\vartheta)$ such that, for every $h \in \mathbb{R}^2$,

$$\left(I_n^{1/2}(\vartheta)(T_n - \vartheta_n^*(h)), G_n(\vartheta)\right) \xrightarrow{\mathcal{D}} (T(\vartheta), G(\vartheta)) \quad \text{under } P_{\vartheta_n^*(h)} \tag{6}$$

as $n \to \infty$.

2.2 Main theorems

Theorem 1 *Under Assumption (H), the random field Λ_n^ϑ is ULAMN at all $\vartheta \in \Theta$ with $G(\vartheta) := \mathrm{diag}\left(g_{x,\vartheta}^{-1} v_\mu, 1/2\right)$, where v_μ is given in (3), and*

$$g_{x,\vartheta} := \frac{x^2}{2\mu\sigma} + \frac{1}{4\mu^2}.$$

The LAMN condition for all $\vartheta \in \Theta$ can yield the upper bound of the concentration probability for estimators: if $\{T_n\}$ is a *regular* sequence of estimators, then

$$\lim_{n\to\infty} P_\vartheta\{I_n^{1/2}(\vartheta)(T_n - \vartheta) \in C\} \le P_\vartheta\{G^{-1/2}(\vartheta)Z \in C\} \tag{7}$$

for any convex symmetric set $C \subset \mathbb{R}^2$ and every $\vartheta \in \Theta$; see Basawa and Scott (1983), Theorem 2.2.1. Hence an estimator T_n^* is *asymptotically optimal* in the regular class of estimators in the sense of *asymptotic maximum concentration probability at ϑ* if

$$\lim_{n\to\infty} P_\vartheta\{I_n^{1/2}(\vartheta)(T_n^* - \vartheta) \in C\} = P_\vartheta\{G^{-1/2}(\vartheta)Z \in C\} \tag{8}$$

for every $\vartheta \in \Theta$. The optimality is sometimes called *asymptotic efficiency in Wolfowitz sense*; see Weiss and Wolfowitz (1974).

Theorem 2 *Under Assumption (H),*

$$\left(e^{\mu n\Delta}(\hat{\mu}_n - \mu), \sqrt{n}(\hat{\sigma}_n - \sigma), G_n(\vartheta)\right) \xrightarrow{\mathcal{D}} \left(v_\mu^{-1/2} Z_1, \sqrt{2}\sigma Z_2, G(\vartheta)\right) \quad \textit{under } P_\vartheta \tag{9}$$

for all $\vartheta \in \Theta$, where (Z_1, Z_2) is a standard normal vector independent of v_μ.

Remark 1 In the proof of Theorem 1, we will show that

$$I_n^{1/2}(\vartheta) \sim \mathrm{diag}\left(g_{x,\vartheta}^{1/2} e^{\mu n\Delta}, \sigma^{-1}\sqrt{n}\right); \tag{10}$$

see (24) below. Thus, by (9) and (10), we have

$$\left(I_n^{1/2}(\vartheta)(\hat{\vartheta}_n - \vartheta), G_n(\vartheta)\right) \xrightarrow{\mathcal{D}} \left(G^{-1/2}(\vartheta)Z, G(\vartheta)\right) \quad \text{under } P_\vartheta$$

for every $\vartheta \in \Theta$. Moreover, it is easy to see the regularity of $\{\hat{\vartheta}_n\}$ from the above convergence with LAMN condition for every $\vartheta \in \Theta$:

$$\left(I_n^{1/2}(\vartheta)(\hat{\vartheta}_n - \vartheta_n^*(h)), G_n(\vartheta)\right) \xrightarrow{\mathcal{D}} \left(G^{-1/2}(\vartheta)Z, G(\vartheta)\right) \quad \text{under } P_{\vartheta_n^*(h)}$$

for every $h \in \mathbb{R}^2$. For the proof, use e.g., Basawa and Scott (1983), Lemma 1.3.2, (b), and check the convergence of the corresponding characteristic functions. Therefore, our estimator $\hat{\vartheta}_n$ is asymptotically efficient in Wolfowitz sense.

Remark 2 Without showing the regularity of $\hat{\vartheta}_n$, we see that $\hat{\mu}_n$ itself is asymptotically efficient since $\hat{\mu}_n$ has the same asymptotic distribution as for the continuous version MLE: $\hat{\mu}_T^{\mathrm{MLE}}$, which is also asymptotically efficient in Wolfowitz sense; see Basawa and Scott (1983) and Luschgy (1992).

Appropriate norming matrices for $\hat{\vartheta}_n$ to yield the asymptotically normal limit without a random mixture are $G_n^{1/2}(\vartheta)I_n^{1/2}(\vartheta)$, and we can show that

$$G_n^{1/2}(\vartheta)I_n^{1/2}(\vartheta) \sim \Phi_n^{1/2}(\vartheta) := \mathrm{diag}\left(\frac{\Delta}{\sigma}\sum_{i=1}^{n} X_{i-1}^2, \ \frac{n}{2\sigma^2}\right)^{1/2};$$

see (24) and (25) in the proof of Theorem 1. If X is ergodic: $\Xi \subset (-\infty, 0)$, then it is easy to see by the law of large numbers; see, e.g. Lemma 8 by Kessler (1997), that

$$\mathrm{diag}\left((n\Delta)^{-1}, n^{-1}\right)\Phi(\vartheta) \xrightarrow{P} \mathrm{diag}\left(-(2\mu)^{-1}, (2\sigma^2)^{-1}\right).$$

Hence we obtain the following well known result:

$$\left(\sqrt{n\Delta}(\hat{\mu}_n - \mu), \sqrt{n}(\hat{\sigma}_n - \sigma)\right) \xrightarrow{\mathcal{D}} \left(\sqrt{-2\mu}Z_1, \sqrt{2}\sigma Z_2\right), \tag{11}$$

which is asymptotically efficient in the sense of minimal asymptotic variance in the regular class of estimators; see Kessler (1997). See Sect. 3.4 for the proof of (11). Consequently, we have the following result.

Corollary 1 *Suppose that $\mu \neq 0$. Then, as $\Delta \to 0$ and $n\Delta \to \infty$,*

$$\Phi_n^{1/2}(\hat{\vartheta}_n)(\hat{\vartheta}_n - \vartheta) \xrightarrow{\mathcal{D}} Z,$$

where Z is a standard bivariate normal vector.

Remark 3 If $\mu = 0$, then the likelihood ratio would not possess the LAMN property. In this case, $\hat{\mu}_n$ would not be asymptotically mixed normal, but have a singular asymptotic distribution. The fact is well known in continuously observed case; see Feigin (1979).

3 Proofs

3.1 Preliminaries for proofs of main theorems

In what follows, we use the following notation:

$$V_n := \sum_{i=1}^n X_{i-1}^2; \quad U_n(s) := \frac{1}{\sqrt{\sigma}} \sum_{i=1}^n X_{i-1} D_i^\Delta(s);$$
$$W_n(s) := \sum_{i=1}^n \left(\left| \frac{D_i^\Delta(s)}{\sqrt{\sigma\Delta}} \right|^2 - \frac{E_s^\Delta}{2s\Delta} \right).$$

In particular, it follows under P_ϑ that

$$U_n := U_n(\mu) = \sum_{i=1}^n \epsilon_i^\Delta(\mu) X_{i-1}; \quad W_n := W_n(\mu) = \left(\left| \frac{\epsilon_i^\Delta(\mu)}{\sqrt{\Delta}} \right|^2 - \frac{E_\mu^\Delta}{2\mu\Delta} \right).$$

Moreover we denote by

$$\tilde{U}_n := \sum_{i=1}^n e^{-\mu(n-i)\Delta} \epsilon_i^\Delta(\mu),$$

and by $\tilde{X}_t := e^{-\mu t} X_t$; $\tilde{X}_i := e^{-\mu t_i} X_i$; $\tilde{X}^* := \sup_{t \geq 0} \tilde{X}_t$.

The following lemma is a version of Toeplitz lemma for triangular arrays. The proof is similar to the usual version's, so we omit it.

Lemma 1 *Let $\{a_i^n\}_{i=1}^n$ be a positive bounded sequence, and put $b_n := \sum_{i=1}^n a_i^n$. Suppose that a sequence $\{x_i^n\}_{i=1}^n$ satisfies the following conditions:*

$$\sup_{n \in \mathbb{N}} |x_i^n| < \infty \quad \textit{for each fixed } i; \tag{12}$$
$$\lim_{k \to \infty} \sup_{j,n: k \leq j \leq n} |x_j^n - x| = 0 \quad \textit{for some } x \in \mathbb{R}; \tag{13}$$

Then, for any sequence A_n with $A_n \sim b_n^{-1}$, $A_n \sum_{i=1}^n a_i^n x_i^n \to x$ as $n \to \infty$.

The next lemma is a corollary of Lemma 1 in the paper by Dietz and Kutoyants (2003).

Lemma 2 *$\tilde{X}^*$ is bounded in $L^p(P_\vartheta)$ for any $p > 0$. Moreover, as $n \to \infty$,*

$$\tilde{X}_n \to x + \sqrt{\sigma} \int_0^\infty e^{-\mu t}\, \mathrm{d}W_t \ (= \sqrt{2\mu\sigma v_\mu}) \tag{14}$$

almost surely. It also holds in the L^p-sense for any $p > 0$.

Proof Since

$$\tilde{X}_t = x + \sqrt{\sigma} \int_0^t e^{-\mu t} \, \mathrm{d}W_t \sim N\left(x, \frac{\sigma}{2\mu}(1 - e^{-2\mu t})\right)$$

is a Gaussian martingale, we can deduce the almost sure convergence for (14) by the martingale convergence theorem; see also Feigin (1976). Moreover it follows from Doob's inequality that, for any $T \geq 0$,

$$E_\vartheta\left[\sup_{t \leq T} |\tilde{X}_t|^{2p}\right] \leq c_p E_\vartheta |\tilde{X}_T|^{2p} \leq C_p \left(\frac{1 - e^{-2\mu T}}{2\mu}\right)^p,$$

where c_p and C_p are constants depending on p. Hence $E_\vartheta |\tilde{X}^*|^{2p} < \infty$. Therefore we can also deduce the L^p-convergence for (14) by the uniform integrability of $\tilde{X}_n$. □

Lemma 3 *Under Assumption (H),*

$$\Delta e^{-2\mu n \Delta} V_n \xrightarrow{P} \sigma v_\mu \quad \textit{under } P_\vartheta. \tag{15}$$

Proof Let $r_\Delta := e^{2\mu\Delta}$. Noticing that $V_n = \sum_{i=1}^n \tilde{X}_{i-1}^2 r_\Delta^{i-1}$, we have

$$\Delta e_n^{-2\mu n \Delta} V_n = \frac{\Delta}{r_\Delta - 1}\left(\tilde{X}_{n-1}^2 - \sum_{i=2}^n (\tilde{X}_{i-1}^2 - \tilde{X}_{t_{i-2}}^2) r_\Delta^{-n+i-1} - \tilde{X}_0^2 r_\Delta^{-n}\right).$$

Since $r_\Delta - 1 = 2\mu\Delta + O(\Delta^2)$ and $e^{\mu n \Delta} \tilde{X}_0^2 r_\Delta^{-n} \to 0$ as $n \to \infty$, it follows that

$$\Delta e^{-2\mu n \Delta} V_n - \frac{\Delta}{r_\Delta - 1} \tilde{X}_{n-1}^2 = \left(-\frac{1}{2\mu} + O(\Delta)\right) R_n + o(1),$$

where $R_n := \sum_{i=2}^n (\tilde{X}_{i-1}^2 - \tilde{X}_{i-2}^2) r_\Delta^{-n+i-1}$. Therefore, we shall show that $R_n = o_p(1)$.

$$\begin{aligned}
E_\vartheta |R_n| &\leq \sum_{i=1}^{n-1} \left\| \tilde{X}_i + \tilde{X}_{i-1} \right\|_{L^2(P_\vartheta)} \|\epsilon_i^\Delta(\mu)\|_{L^2(P_\vartheta)} r_\Delta^{-n+\frac{i}{2}} \\
&\leq 2\|\tilde{X}^*\|_{L^2(P_\vartheta)} \sqrt{\frac{1}{2\mu}(e^{2\mu\Delta} - 1)} \sum_{i=1}^{n-1} e^{-\mu\Delta(2n-i)} \\
&\leq 2\|\tilde{X}^*\|_{L^2(P_\vartheta)} e^{-\mu n \Delta} \sqrt{\frac{1}{2\mu} \frac{e^{2\mu\Delta} - 1}{(e^{\mu\Delta} - 1)^2}} \xrightarrow{P} 0,
\end{aligned}$$

thanks to Lemma 2. Thus we have

$$\Delta e^{-2\mu n\Delta} V_n = \frac{\Delta}{r_\Delta - 1}\tilde{X}_{n-1}^2 + o_p(1). \tag{16}$$

Therefore we obtain (15) from (14) and (16). □

Lemma 4 *Under Assumption (H),*

$$\left(\tilde{X}_n, \tilde{U}_n, W_n/\sqrt{n}\right) \xrightarrow{\mathcal{D}} \left(\sqrt{2\mu\sigma v_\mu}, Z_1/\sqrt{2\mu}, \sqrt{2}Z_2\right) \quad \text{under } P_\vartheta,$$

where (Z_1, Z_2) is a standard bivariate normal variable independent of v_μ.

Proof We note the following expressions:

$$\begin{aligned}
\tilde{X}_n - x &= \sum_{i=1}^n \tilde{x}_i^n, \quad \text{where } \tilde{x}_i^n := \sqrt{\sigma} e^{-\mu i\Delta}\epsilon_i^\Delta(\mu);\\
\tilde{U}_n &:= \sum_{i=1}^n \tilde{u}_i^n, \quad \text{where } \tilde{u}_i^n := e^{-\mu(n-i)\Delta}\epsilon_i^\Delta(\mu);\\
W_n/\sqrt{n} &=: \sum_{i=1}^n w_i^n, \quad \text{where } w_i^n := \frac{1}{\sqrt{n}}\left(\left|\frac{\epsilon_i^\Delta(\mu)}{\sqrt{\Delta}}\right|^2 - \frac{E_\mu^\Delta}{2\mu\Delta}\right),
\end{aligned}$$

which are martingale arrays. Therefore it suffices to show from the CLT for martingale arrays; see, e.g., Hall and Heyde (1980), Chapter 3, and the Markovian property of X that

$$\begin{aligned}
&\sum_{i=1}^n E_\vartheta[(\tilde{x}_i^n)^2|X_{i-1}] \xrightarrow{P} \frac{\sigma}{2\mu}; \quad \sum_{i=1}^n E_\vartheta[(\tilde{x}_i^n)^4|X_{i-1}] \xrightarrow{P} 0;\\
&\sum_{i=1}^n E_\vartheta[(\tilde{u}_i^n)^2|X_{i-1}] \xrightarrow{P} \frac{1}{2\mu}; \quad \sum_{i=1}^n E_\vartheta[(\tilde{u}_i^n)^4|X_{i-1}] \xrightarrow{P} 0;\\
&\sum_{i=1}^n E_\vartheta[(w_i^n)^2|X_{i-1}] \xrightarrow{P} 2; \quad \sum_{i=1}^n E_\vartheta[(w_i^n)^4|X_{i-1}] \xrightarrow{P} 0;\\
&\sum_{i=1}^n E_\vartheta[\tilde{u}_i^n\tilde{x}_i^n|X_{i-1}] \xrightarrow{P} 0; \quad \sum_{i=1}^n E_\vartheta[\tilde{u}_i^n w_i^n|X_{i-1}] \xrightarrow{P} 0;\\
&\sum_{i=1}^n E_\vartheta[\tilde{x}_i^n w_i^n|X_{i-1}] \xrightarrow{P} 0
\end{aligned}$$

under P_ϑ: the last three convergences are clear since $\epsilon_i^\Delta(\mu)$ is normal variable with mean zero and variance $(2\mu)^{-1}(e^{2\Delta\mu} - 1)$, which tends to zero as $n \to \infty$. Moreover,

$$\sum_{i=1}^{n} E_\vartheta[(\tilde{x}_i^n)^2|X_{i-1}] = \frac{\sigma}{2\mu}(e^{2\mu\Delta}-1)\sum_{i=1}^{n} e^{-2\mu i\Delta} \to \frac{\sigma}{2\mu};$$
$$\sum_{i=1}^{n} E_\vartheta[(\tilde{x}_i^n)^4|X_{i-1}] = 3\sigma^2\left(\frac{E_\mu^\Delta}{2\mu}\right)^2\sum_{i=1}^{n} e^{-4\mu i\Delta} \to 0;$$
$$\sum_{i=1}^{n} E_\vartheta[(\tilde{u}_i^n)^2|X_{i-1}] = \frac{E_\mu^\Delta}{2\mu}\frac{1-e^{-2\mu n\Delta}}{1-e^{-2\mu\Delta}} \to \frac{1}{2\mu};$$
$$\sum_{i=1}^{n} E_\vartheta[(\tilde{u}_i^n)^4|X_{i-1}] = 3\left(\frac{E_\mu^\Delta}{2\mu}\right)^2\frac{1-e^{-4\mu n\Delta}}{1-e^{-4\mu\Delta}} = O(\Delta);$$
$$\sum_{i=1}^{n} E_\vartheta[(w_i^n)^2|X_{i-1}] = 3\left(\frac{E_\mu^\Delta}{2\mu\Delta}\right)^2 - \left(\frac{E_\mu^\Delta}{2\mu\Delta}\right)^2 \to 2;$$
$$\sum_{i=1}^{n} E_\vartheta[(w_i^n)^4|X_{i-1}] = O(n^{-1}) \to 0.$$

The last equality is due to the boundedness of $E_\vartheta[|\epsilon_i^\Delta(\mu)/\sqrt{\Delta}|^8|X_{i-1}]$. □

Corollary 2 *Under Assumption (H),*

$$\left(\Delta e^{-2\mu n\Delta}V_n, e^{-\mu n\Delta}U_n, W_n/\sqrt{n}\right) \xrightarrow{\mathcal{D}} \left(\sigma v_\mu, \sqrt{\sigma v_\mu}Z_1, \sqrt{2}Z_2\right) \quad \textit{under } P_\vartheta,$$

where (Z_1, Z_2) is a standard bivariate normal variable independent of v_μ.

Proof We shall show that

$$e^{-\mu n\Delta}U_n - e^{-\mu\Delta}\tilde{X}_n\tilde{U}_n \xrightarrow{P} 0. \tag{17}$$

Then, it follows from (16) that $\left(\Delta e^{-2\mu n\Delta}V_n, e^{-\mu n\Delta}U_n, W_n/\sqrt{n}\right)$ is asymptotically equivalent to $\left(\tilde{X}_n^2/(2\mu), e^{-\mu\Delta}\tilde{X}_n\tilde{U}_n, W_n/\sqrt{n}\right)$. Hence Lemma 4 yields the consequence.

Let us show (17). By Schwarz's inequality,

$$E\left|e^{-\mu n\Delta}U_n - e^{-\mu\Delta}\tilde{X}_n\tilde{U}_n\right| \le \sum_{i=1}^{n}\left\|e^{-\mu(n-i+1)\Delta}\epsilon_i^\Delta(\mu)\right\|_{L^2(P_\vartheta)}\left\|\tilde{X}_{i-1}-\tilde{X}_n\right\|_{L^2(P_\vartheta)}$$
$$= \frac{E_\mu^\Delta}{2\mu}\sum_{i=1}^{n} e^{-\mu(n-i+1)\Delta}\left\|\tilde{X}_{i-1}-\tilde{X}_n\right\|_{L^2(P_\vartheta)},$$

which tends to zero by Lemma 1 with $a_i^n = e^{-\mu(n-i+1)\Delta}$, $A_n = E_\mu^\Delta/(2\mu)$, and $x_i^n = \|\tilde{X}_{i-1} - \tilde{X}_n\|_{L^2(P_\vartheta)}$. The conditions in Lemma 1 are easily checked using Lemma 2. This ends the proof. □

The next lemma is useful to simplify the computation below.

Lemma 5 *There exists a deterministic C^3-function γ_1^Δ satisfying that*

$$\sup_{\mu\in\Xi}\left|\partial_\mu^k \gamma_1^\Delta(\mu)\right| = O(\Delta) \quad (k=0,1,2,3) \tag{18}$$

as $\Delta \to 0$ such that $2\mu\Delta/E_\mu^\Delta = 1+\gamma_1^\Delta(\mu)$.

Proof Define a function f as $f(x) = x(e^x-1)^{-1}$ if $x \neq 0$, and $f(x)=0$ if $x=0$. Then it is easy to check by the direct computation that $f \in C^4$. By Taylor's formula,

$$2\mu\Delta/E_\mu^\Delta - 1 = f(2\mu\Delta) - 1 = 2\mu\Delta\int_0^1 f'(2\mu\Delta u)\,\mathrm{d}u =: \gamma_1^\Delta(\mu).$$

Then, for $k=0,1,2,3$,

$$\partial_\mu^k\gamma_1^\Delta(\mu) = 2\Delta\int_0^1 \partial_\mu^k\left(\mu f'(2\mu\Delta u)\right)\,\mathrm{d}u.$$

Since $f \in C^4$ and Ξ is compact, the last integral is bounded as $\Delta \to 0$ for any k, which implies (18). □

Lemma 6 *Under Assumption (H),*

$$\Delta e^{-2\mu n\Delta}\left(\sup_{s\in\Xi}|U_n(s)| + \sup_{s\in\Xi}|\partial_s^k W_n(s)|\right) \xrightarrow{P} 0 \quad (k=0,1);$$
$$e^{-2\mu n\Delta}\sup_{s\in\Xi}|\partial_s U_n(s)| \xrightarrow{P} 0$$

under P_ϑ.

Proof First, we show the case with $k=0$. Note that, for each $s\in\Theta$,

$$U_n(s) = U_n + \sigma^{-1/2}(e^{\mu\Delta}-e^{s\Delta})V_n; \tag{19}$$

$$\begin{aligned} W_n(s) = W_n &+ \frac{2}{\sqrt{\sigma}\Delta}(e^{\mu\Delta}-e^{s\Delta})U_n \\ &+\frac{1}{\sigma\Delta}(e^{\mu\Delta}-e^{s\Delta})^2V_n + n\left(\frac{E_\mu^\Delta}{2\mu\Delta}-\frac{E_s^\Delta}{2s\Delta}\right), \end{aligned} \tag{20}$$

which implies that, $\Delta e^{-2\mu n\Delta}(U_n(s), W_n(s)) \xrightarrow{P} 0$; the convergence of finite dimensional distribution of process (U_n, W_n) on $C(\Theta)$. Therefore the tightness of $\{\Delta e^{-2\mu n\Delta}(U_n, W_n)\}$ in $C(\Theta)$ yields the consequence. We shall check the simple tightness criterion:

$$\sup_{n\in\mathbb{N}}\Delta e^{-2\mu n\Delta}E_\vartheta\left[\sup_{s\in\Theta}|\partial_s U_n(s)| + \sup_{s\in\Theta}|\partial_s W_n(s)|\right] < \infty. \tag{21}$$

From the expressions (19) and (20), we have

$$\partial_s U_n(s) = -\sigma^{-1/2}\Delta e^{s\Delta} V_n; \tag{22}$$

$$\partial_s W_n(s) = -\frac{2}{\sqrt{\sigma}} e^{s\Delta} U_n - \frac{2e^{s\Delta}}{\sigma}(e^{\mu\Delta} - e^{s\Delta})V_n - n\partial_s\left(\frac{1}{1+\gamma_1^\Delta(s)}\right). \tag{23}$$

Hence (21) can be easily checked since Θ is bounded. This completes the case with $k=0$. Moreover (22), (23), and the case with $k=1$ easily yield the case with $k=1$. □

3.2 Proof of Theorem 1

Note that

$$\partial_s D_i^\Delta(s) = -\Delta e^{s\Delta} X_{i-1}, \quad D_i^\Delta(\mu) = \sqrt{\sigma}\epsilon_i^\Delta(\mu).$$

Using Lemma 5, we obtain that

$$\ell_n(\vartheta) = -\frac{1}{2\sigma\Delta}\sum_{i=1}^n |D_i^\Delta(\mu)|^2(1+\gamma_1^\Delta(\mu)) - \frac{n}{2}\log(1+\gamma_1^\Delta(\mu)) - \frac{n}{2}\log(\sigma\Delta);$$

$$\partial_\mu \ell_n(\vartheta) = \frac{1}{\sigma}(1+\gamma_1^\Delta(\mu))\sum_{i=1}^n e^{\mu\Delta} D_i^\Delta(\mu) X_{i-1} - \frac{\partial_\mu\gamma_1^\Delta(\mu)}{\sigma\Delta}\sum_{i=1}^n |D_i^\Delta(\mu)|^2 - \frac{n}{2(1+\gamma_1^\Delta(\mu))};$$

$$\partial_\sigma \ell_n(\vartheta) = \frac{1}{2\sigma^2\Delta}\sum_{i=1}^n |D_i^\Delta(\mu)|^2(1+\gamma_1^\Delta(\mu)) - \frac{n}{2\sigma} = \frac{\mu\Delta}{\sigma E_\mu^\Delta} W_n;$$

$$\partial_\mu^2 \ell_n(\vartheta) = \frac{\partial_\mu\gamma_1^\Delta(\mu) e^{\mu\Delta}}{\sqrt{\sigma}} U_n + \frac{\Delta e^{\mu\Delta}}{\sigma}(1+\gamma_1^\Delta(\mu))\left(\sqrt{\sigma}U_n - e^{\mu\Delta}V_n\right) - \frac{\partial_\mu^2\gamma_1^\Delta(\mu)}{\Delta}\sum_{i=1}^n |\epsilon_i^\Delta(\mu)|^2 + \frac{2e^{\mu\Delta}\partial_\mu\gamma_1^\Delta(\mu)}{\sqrt{\sigma}} U_n + \frac{\partial_\mu\gamma_1^\Delta(\mu) n}{2(1+\gamma_1^\Delta(\mu))^2};$$

$$\partial_\sigma^2 \ell_n(\vartheta) = -\frac{1}{\sigma^3\Delta}\sum_{i=1}^n |D_i^\Delta(\mu)|^2(1+\gamma_1^\Delta(\mu)) + \frac{n}{2\sigma^2};$$

$$\partial_\mu\partial_\sigma \ell_n(\vartheta) = -\frac{e^{\mu\Delta}}{\sigma\sqrt{\sigma}}(1+\gamma_1^\Delta(\mu))U_n + \frac{\partial_\mu\gamma_1^\Delta(\mu)}{2\sigma^2\Delta}\sum_{i=1}^n |D_i^\Delta(\mu)|^2.$$

Step.1: Computation of $I_n := -E_\vartheta\left[\mathrm{diag}\left(\nabla_\vartheta\nabla_\vartheta^\top \ell_n(\vartheta)\right)\right]$.

Note that

$$E_\vartheta[U_n] = 0; \quad E_\vartheta[|D_i^\Delta(\mu)|^2] = \frac{E_\mu^\Delta}{2\mu}\sigma; \quad E_\vartheta[X_i^2] = e^{2\mu i\Delta}x^2 + \frac{\sigma}{2\mu}(e^{2\mu i\Delta} - 1),$$

and

$$E_\vartheta[V_n] = \frac{e^{2\mu\Delta}}{E_\mu^\Delta}(e^{2\mu n\Delta} - 1)\left(x^2 + \frac{\sigma}{2\mu}\right) - \frac{n\sigma}{2\mu}.$$

Thus we have

$$\begin{aligned}
-E_\vartheta[\partial_\mu^2 \ell_n(\vartheta)] &= \frac{\Delta e^{2\mu\Delta}}{\sigma}(1 + \gamma_1^\Delta(\mu))E[V_n] + \frac{\partial_\mu^2 \gamma_1^\Delta(\mu) E_\mu^\Delta}{2\mu\Delta\sigma} n - \frac{\partial_\mu \gamma_1^\Delta(\mu) n}{2(1 + \gamma_1^\Delta(\mu))^2};\\
&= \frac{e^{2\mu n\Delta}}{2\mu}\left(\frac{x^2}{\sigma} + \frac{1}{2\mu}\right) + O(n\Delta);\\
-E_\vartheta[\partial_\sigma^2 \ell_n(\vartheta)] &= \frac{E_\mu^\Delta}{2\mu\sigma^2\Delta} n(1 + \gamma_1^\Delta(\mu)) + \frac{1}{2\sigma^2}\\
&= \frac{n}{\sigma^2}(1 + O(\Delta)) + \frac{1}{2\sigma^2}.
\end{aligned}$$

Hence we have

$$I_n(\vartheta) \sim \operatorname{diag}\left(\left(\frac{x^2}{2\mu\sigma} + \frac{1}{4\mu^2}\right)e^{2\mu n\Delta}, n\sigma^{-2}\right) = \operatorname{diag}(g_{x,\vartheta} e^{2\mu n\Delta}, n\sigma^{-2}). \tag{24}$$

Step.2: Checking Condition (A.1).
We see that

$$\begin{aligned}
-e^{-2\mu n\Delta}\partial_\mu^2 \ell_n(\vartheta) &= \sigma^{-1}\Delta e^{-2\mu n\Delta} V_n + O_p(\Delta e^{-\mu n\Delta}) \xrightarrow{P} v_\mu;\\
-n^{-1}\partial_\sigma^2 \ell_n(\vartheta) &= \frac{1}{n\sigma^2} W_n - \frac{E_\mu^\Delta}{2\sigma^2\mu\Delta} - \frac{1}{2\sigma^2} \xrightarrow{P} \frac{1}{2\sigma^2};\\
-n^{-1/2} e^{-\mu n\Delta}\partial_\mu\partial_\sigma \ell_n(\vartheta) &= \frac{1}{\sigma\sqrt{\sigma n}} e^{-\mu n\Delta} U_n + O_p(\Delta) \xrightarrow{P} 0.
\end{aligned}$$

Therefore it follows from (24) that, for $G_n(\vartheta) := I_n^{-1/2}(\vartheta)B_n(\vartheta)I_n^{-1/2}(\vartheta)$,

$$G_n(\vartheta) \xrightarrow{P} \operatorname{diag}\left(g_{x,\vartheta}^{-1} v_\mu, \frac{1}{2}\right) = G(\vartheta). \tag{25}$$

Moreover it also follows from (24) and Corollary 2 that

$$\begin{aligned}
L_n(\vartheta) &\sim \operatorname{diag}\left(g_{x,\vartheta}^{-1/2} e^{-\mu n\Delta}, \sigma/\sqrt{n}\right)\begin{pmatrix} \frac{e^{\mu\Delta}}{\sqrt{\sigma}} U_n + O_p(n\Delta); \\ \frac{\mu\Delta}{\sigma E_\mu^\Delta} W_n \end{pmatrix}\\
&\xrightarrow{\mathcal{D}} \operatorname{diag}\left(\sqrt{g_{x,\vartheta}^{-1} v_\mu}, \frac{1}{\sqrt{2}}\right)\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = G^{1/2}(\vartheta) Z \quad \text{under } P_\vartheta.
\end{aligned}$$

Since (L_n, G_n) are written by (V_n, U_n, W_n), we have $(L_n(\vartheta), G_n(\vartheta)) \xrightarrow{\mathcal{D}} (G^{1/2}(\vartheta)Z, G(\vartheta))$.

Step.3: Checking Condition (A.2) for any bounded (h_n).

By Taylor's formula, it follows for any bounded sequence $h_n \in \mathbb{R}^2$ with $h_n \neq 0$ and for n large enough that

$$\begin{aligned}
2|h_n|^{-2}|r_n(h_n,\vartheta)| &\leq \left| I_n^{-1/2}(\vartheta)\left(B_n(\vartheta_n^*(h_n)) - B_n(\vartheta)\right) I_n^{-1/2}(\vartheta)\right| \\
&\leq \Bigg[e^{-3\mu n\Delta} \sup_{\vartheta\in\Theta} |\partial_\mu^3 \ell_n(\vartheta)| + n^{-1/2} e^{-2\mu n\Delta} \sup_{\vartheta\in\Theta} |\partial_\mu^2 \partial_\sigma \ell_n(\vartheta)| \\
&\quad + n^{-1} e^{-\mu n\Delta} \sup_{\vartheta\in\Theta} |\partial_\mu \partial_\sigma^2 \ell_n(\vartheta)| + n^{-1} |\partial_\sigma^2 \ell_n(\vartheta_n^*(h_n)) \\
&\quad - \partial_\sigma^2 \ell_n(\vartheta)| \Bigg] O_p(1).
\end{aligned}$$

Noticing that $\partial_s U_n(s) = -\Delta e^{s\Delta} V_n$, and that

$$\sum_{i=1}^n |D_i^\Delta(s)|^2 = \sigma\Delta W_n(s); \quad \partial_s \left(\sum_{i=1}^n |D_i^\Delta(s)|^2 \right) = -2\Delta e^{s\Delta} U_n(s), \tag{26}$$

we have

$$\begin{aligned}
\sup_{s\in\Theta} |\partial_\mu^3 \ell_n(\vartheta)| &= \left(V_n + \sup_{s\in\Xi} |U_n(s)| \right) O(\Delta^2) + \sup_{s\in\Xi} |W_n(s)| O(\Delta) + O(n\Delta); \\
\sup_{\vartheta\in\Theta} |\partial_\mu^2 \partial_\sigma \ell_n(\vartheta)| &= \left(V_n + \sup_{s\in\Xi} |U_n(s)| + \sup_{s\in\Xi} |W_n(s)| \right) O(\Delta); \\
\sup_{\vartheta\in\Theta} |\partial_\mu \partial_\sigma^2 \ell_n(\vartheta)| &= \sup_{s\in\Xi} |U_n(s)| O(1) + \sup_{s\in\Xi} |W_n(s)| O(\Delta).
\end{aligned}$$

Thus Lemma 6 yields that

$$\begin{aligned}
&e^{-3\mu n\Delta} \sup_{\vartheta\in\Theta} |\partial_\mu^3 \ell_n(\vartheta)| + n^{-1/2} e^{-2\mu n\Delta} \sup_{\vartheta\in\Theta} |\partial_\mu^2 \partial_\sigma \ell_n(\vartheta)| \\
&\quad + n^{-1} e^{-\mu n\Delta} \sup_{\vartheta\in\Theta} |\partial_\mu \partial_\sigma^2 \ell_n(\vartheta)| \xrightarrow{P} 0.
\end{aligned}$$

Moreover

$$\left| \partial_\sigma^2 \ell_n(\vartheta_n^*(h_n)) - \partial_\sigma^2 \ell_n(\vartheta) \right| = \left| W_n(\mu + e^{-\mu n\Delta} h_n) - W_n \right| O(1) + O(1).$$

Thanks to the expression of (20) in the proof of Lemma 6 and Corollary 2, we see that

$$W_n(\mu + e^{-\mu n\Delta} h_n) - W_n = O\left(\frac{1}{\Delta} + n\Delta \right),$$

which implies that

$$n^{-1}\left|\partial_\sigma^2\ell_n(\vartheta_n^*(h_n)) - \partial_\sigma^2\ell_n(\vartheta)\right| \xrightarrow{P} 0.$$

This completes the proof. □

3.3 Proof of Theorem 2

We shall show the following three convergences under P_ϑ for any $\vartheta \in \Theta$:

$$\left(e^{\mu n\Delta}\left(\hat{\mu}_n-\mu\right), \sqrt{n}\left(\hat{\sigma}_n(\mu)-\sigma\right), G_n(\vartheta)\right) \xrightarrow{\mathcal{D}} \left(v_\mu^{-1/2}Z_1, \sqrt{2}\sigma Z_2, G(\vartheta)\right); \quad (27)$$

$$\sqrt{n}\left(\hat{\sigma}_n(\hat{\mu}_n) - \hat{\sigma}_n(\mu)\right) \xrightarrow{P} 0, \quad (28)$$

which implies the joint convergence of them, and yield the consequence.

Let $\Psi_n(\nu) := \sum_{i=1}^n \left|X_i - e^{\nu\Delta}X_{i-1}\right|^2$: the contrast function given in (4). Using the equality $X_{t_i} = e^{\mu\Delta}X_{i-1} + \epsilon_i^\Delta(\mu)$, we have

$$\Psi_n(\nu) = (e^{\mu\Delta} - e^{\nu\Delta})^2 V_n + 2\sqrt{\sigma}(e^{\mu\Delta} - e^{\nu\Delta})U_n.$$

Then, it follows from Corollary 2 that

$$\sup_{\nu\in\Xi}|\Delta e^{-2\mu n\Delta}\Psi_n(\nu) - L(\nu)| \xrightarrow{P} 0,$$

where $L(\nu) := (e^{\mu\Delta} - e^{\nu\Delta})^2 v_\mu$. The limit $L(\nu)$ satisfies that $\inf_{\nu:|\nu-\mu|>\epsilon} L(\nu) > L(\mu)\ a.s.$ for any $\epsilon > 0$. Thanks to the standard argument for consistency of M-estimator, we can conclude that $\hat{\mu}_n \xrightarrow{P} \mu$, which implies that $P_\vartheta\left\{\hat{\mu}_n \in int(\Xi)\right\} \to 1$ since $\mu \in int(\Xi)$. Therefore, by a classical routine, we can assume that $\hat{\mu}_n$ is a local minimum in Ξ for n large enough: $P_\vartheta\left\{\partial_\mu\Psi_n(\hat{\mu}_n) = 0\right\} \to 1$, which implies that the explicit form (5) is well-defined with probability tending to one. Using this expression, we have

$$\begin{aligned} e^{\mu n\Delta}(\hat{\mu}_n - \mu) &= \frac{e^{\mu n\Delta}}{\Delta}\log\left(1 + \frac{\sum_{i=1}^n X_{i-1}\sqrt{\sigma}\epsilon_i^\Delta(\mu)}{e^{\mu\Delta}\sum_{i=1}^n X_{t_{i-1}}^2}\right) \\ &= \frac{e^{\mu n\Delta}}{\Delta}\log\left(1 + \Delta e^{-\mu(n+1)\Delta}\frac{e^{-\mu n\Delta}\sqrt{\sigma}U_n}{\Delta e^{-2\mu n\Delta}V_n}\right) \\ &= \frac{e^{-\mu n\Delta}\sqrt{\sigma}U_n}{\Delta e^{-2\mu n\Delta}V_n} + o_p(1). \end{aligned} \quad (29)$$

Next, by the definition of $\hat{\sigma}_n$, we have

$$\sqrt{n}(\hat{\sigma}_n(\mu) - \sigma) = \sum_{i=1}^{n} \frac{2\mu\sigma}{\sqrt{n}E_\mu^\Delta}\left(|\epsilon_i^\Delta(\mu)|^2 - \frac{E_\mu^\Delta}{2\mu}\right) = \frac{2\mu\Delta\sigma}{\sqrt{n}E_\mu^\Delta} W_n. \tag{30}$$

Therefore, from (25), (29) and (30), the tensor

$$\left(e^{\mu n\Delta}\left(\hat{\mu}_n - \mu\right), \sqrt{n}\left(\hat{\sigma}_n(\mu) - \sigma\right), G_n(\vartheta)\right)$$

is written by $\left(\Delta e^{-2\mu n\Delta} V_n, e^{-\mu n\Delta} U_n, W_n/\sqrt{n}\right)$ via a continuous function, plus $o_p(1)$. Then Corollary 2 yields (27).

It remains to show (28). Thanks to Taylor's formula, we have

$$\left|\sqrt{n}\left(\hat{\sigma}_n(\hat{\mu}_n) - \hat{\sigma}_n(\mu)\right)\right| \le \sqrt{n}e^{-\mu n\Delta}\left|\partial_\mu \hat{\sigma}_n(\tilde{\mu}_n^u)\right|\left\{e^{\mu n\Delta}\left(\hat{\mu}_n - \mu\right)\right\} \tag{31}$$

where $\tilde{\mu}_n^u := \mu + u(\hat{\mu}_n - \mu)$ for some random $u \in (0, 1)$. It follows from Lemma 5 and (26) that

$$\begin{aligned}\partial_s \hat{\sigma}_n(s) &= \partial_s\left((1 + \gamma_1^\Delta(s))\frac{1}{n\Delta}\sum_{i=1}^{n}|D_i^\Delta(s)|^2\right)\\ &= \frac{\partial_s\gamma_1^\Delta(s)}{n}\sigma W_n(s) - \frac{2e^{s\Delta}}{n}U_n(s).\end{aligned}$$

By Taylor's formula, we have

$$\begin{aligned}|\partial_s \hat{\sigma}_n(\tilde{\mu}_n^u)| \le &\left[\frac{\partial_s\gamma_1^\Delta(\tilde{\mu}_n^u)}{n}\sigma \sup_{s\in\Xi}|\partial_s W_n(s)| + \frac{2e^{\tilde{\mu}_n^u\Delta}}{n}\sup_{s\in\Xi}|\partial_s U_n(s)|(1 + \gamma_1^\Delta(\tilde{\mu}_n^u))\right]\\ &\times(\hat{\mu}_n - \mu).\end{aligned}$$

Therefore, applying Lemma 6, we see that

$$\sqrt{n}e^{-\mu n\Delta}\partial_\mu\hat{\sigma}_n(\tilde{\mu}_n^u) = O_p(n^{-1/2}).$$

This completes the proof. □

3.4 Proof of (11)

The route of the proof is the same as one of Theorem 2. Hence we shall show only

$$\sqrt{n\Delta}(\hat{\mu}_n - \mu) \xrightarrow{\mathcal{D}} \sqrt{-2\mu}Z_1. \tag{32}$$

If X is ergodic: $\Xi \subset (-\infty, 0)$, then it is easy to see that, as $\Delta \to 0$ and $n\Delta \to \infty$,

$$n^{-1}V_n \xrightarrow{P} -\frac{\sigma}{2\mu} \quad \text{under } P_\vartheta. \tag{33}$$

For the proof, note Lemma 8 by Kessler (1997), and that the stationary distribution of X_t is $N\left(0, -\sigma(2\mu)^{-1}\right)$. Moreover we can show by the same argument as in the proof of Lemma 4 with (33) that

$$\left((n\Delta)^{-1/2}U_n, W_n/\sqrt{n}\right) \xrightarrow{\mathcal{D}} \left(\sqrt{-\sigma/2\mu}Z_1, \sqrt{2}Z_2\right) \quad \text{under } P_\vartheta. \tag{34}$$

Here we use again the expression (5) by assuming that $\hat{\mu}_n$ is a local minimum in Ξ:

$$\begin{aligned}\sqrt{n\Delta}(\hat{\mu}_n - \mu) &= \frac{\sqrt{n\Delta}}{\Delta}\log\left(1 + \frac{\sum_{i=1}^n X_{i-1}\sqrt{\sigma}\epsilon_i^\Delta(\mu)}{e^{\mu\Delta}\sum_{i=1}^n X_{t_{i-1}}^2}\right)\\ &= \frac{\sqrt{n\Delta}}{\Delta}\log\left(1 + \frac{\sqrt{\sigma}e^{-\mu\Delta}}{\sqrt{n\Delta}}\frac{(n\Delta)^{-1/2}U_n}{n^{-1}V_n}\right)\\ &= \frac{(n\Delta)^{-1/2}\sqrt{\sigma}U_n}{n^{-1}V_n} + o_p(1).\end{aligned}$$

This yields (32). □

Acknowledgments The author expresses many thanks to anonymous referees for accurate comments and instructions, which improve the paper so much. The author is also grateful to Dr. Hiroki Masuda for his helpful comments. This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B), 2009, no. 21740073; Japan Science and Technology Agency, CREST; Cooperative Research Meetings, the Institute of Statistical Mathematics: no. 21-2053–21-2055.

References

Aldous, D. J., Eagleson, G. K. (1978). On mixing and stability of limit theorems. *The Annals Probability, 6*, 325–333.
Anderson, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *The Annals of Mathematical Statistics, 30*, 676–687.
Basawa, I. V., Scott, D. J. (1983). *Asymptotic optimal inference for non-ergodic models*. New York: Springer.
Dietz, H. M., Kutoyants, Yu. A. (2003). Parameter estimation for some non-recurrent solutions of SDE. *Statistics & Decisions, 21*(1), 29–45.
Feigin, P. D. (1976). Maximum likelihood estimation for continuous-time stochastic processes. *Advances in Applied Probability, 8*(4), 712–736.
Feigin, P. D. (1979). Some comments concerning a curious singularity. *Journal of Applied Probability, 16*(2), 440–444.
Gobet, E. (2002). LAN property for ergodic diffusions with discrete observations, *Annales de l'Institut Henri Poincaré Probabilités et Statistiques, 38*(5), 711–737.
Hall, P., Heyde, C. (1980). *Martingale limit theory and its applications*. New York: Academic Press.
Ibragimov, I. A., Has'minskii, R. Z. (1981). *Statistical estimation*. Berlin: Springer.
Jacod, J. (2006). Parametric inference for discretely observed non-ergodic diffusions. *Bernoulli, 12*(3), 383–401.
Jeganathan, P. (1982). On the asymptotic theory of estimation when the limit of the loglikelihood ratio is mixed normal. *Sankhya Series A, 44*, 173–212.

Kasonga, R. A. (1988). The consistency of a nonlinear least squares estimator from diffusion processes. *Stochastic Processes and their Applications, 30*, 263–275.

Kasonga, R. A. (1990). Parameter estimation by deterministic approximation of a solution of a stochastic differential equation. *Communications in Statistics. Stochastic Models, 6*, 59–67.

Kessler, M. (1997). Estimation of Diffusion Processes From Discrete Observations. *Scandinavian Journal of Statistics, 24*, 211–229.

Kutoyants, Y. A. (2004). *Statistical inference for ergodic diffusion processes*. London: Springer.

Le Breton, A. (1975). On continuous and discrete sampling for parameter estimation in diffusion type processes. Stochastic systems: Modeling, identification and optimization. In *Proceedings of Symposium of University of Kentucky, Lexington, KY*.

Le Breton, A. (1976). *Mathematical programming study* (Vol. 5, p. 124–144).

Luschgy, H. (1992). Local asymptotic mixed normality for semimartingale experiments. *Probability Theory and Related Fields, 92*, 151–176.

Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika, 74*(3), 535–547.

Prakasa Rao, B. L. S. (1999a). *Semimartingales and their statistical inference, Monographs on Statistics and Applied Probability, 83*. Boca Raton: Chapman & Hall/CRC.

Prakasa Rao, B. L. S. (1999b). *Statistical inference for diffusion type processes, Kendall's Library of statistics* (Vol. 8). London/New York: Edward Arnold/Oxford University Press.

Shimizu, Y. (2009a). *Quadratic type contrast functions for discretely observed non-ergodic diffusion processes* (Research Report Series, 09-04). Department of Engineering Science, Osaka University.

Shimizu, Y. (2009b). Notes on drift estimation for non-recurrent Ornstein–Uhlenbeck processes from sampled data. *Statistics & Probability Letters, 79*(20), 2200–2207.

Weiss, L., Wolfowitz, J. (1974). *Maximum probability estimators and related topics*. Springer Lecture Notes in Mathematics (Vol. 424). Berlin: Springer.