Bernoulli **30**(2), 2024, 983–1006
https://doi.org/10.3150/23-BEJ1621

Malliavin calculus techniques for local asymptotic mixed normality and their application to hypoelliptic diffusions

MASAAKI FUKASAWA[1,a] and TEPPEI OGIHARA[2,3,b]

[1]*Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560–8531, Japan,* [a]*fukasawa@sigmath.es.osaka-u.ac.jp*
[2]*Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113–8656, Japan,* [b]*ogihara@mist.i.u-tokyo.ac.jp*
[3]*Risk Analysis Research Center, The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan*

We study sufficient conditions for a local asymptotic mixed normality property of statistical models. We accommodate the framework of Jeganathan [*Sankhyā Ser. A* **44** (1982) 173–212] to a triangular array of variable dimension to, in particular, treat high-frequency observations of stochastic processes. When observations are smooth in the Malliavin sense, with the aid of Malliavin calculus techniques by Gobet [*Bernoulli* **7** (2001) 899–912], we further give tractable sufficient conditions which do not require Aronson-type estimates of the transition density function. The transition density function is even allowed to have zeros. For an application, we prove the local asymptotic mixed normality property of hypoelliptic diffusion models under high-frequency observations, in both complete and partial observation frameworks. The former and the latter extend previous results for elliptic diffusions and for integrated diffusions, respectively.

Keywords: Hypoelliptic diffusion processes; integrated diffusion processes; local asymptotic mixed normality; L^2 regularity condition; Malliavin calculus; partial observations

1. Introduction

In the study of statistical inference for parametric models, *asymptotic efficiency* plays a key role when we consider the asymptotic optimality of estimators. This notion was first studied for models that satisfy *local asymptotic normality* (LAN); Hájek [8] showed the convolution theorem, and Hájek [9] showed the minimax theorem under the LAN property. Both theorems give different concepts of asymptotic efficiency. For statistical models with the extended notion of *local asymptotic mixed normality* (LAMN), Jeganathan [11,12] showed the convolution theorem and the minimax theorem.

The LAMN property for discretely observed elliptic diffusion processes on a fixed interval was shown by Gobet [6], which in particular implies the asymptotic efficiency of the maximum-likelihood-type estimator proposed by Genon-Catalot and Jacod [2]. For further results related to diffusion processes on a fixed interval, see Gloter and Jacod [4] (LAN for noisy observations of diffusion processes with deterministic diffusion coefficients), Gloter and Gobet [3] (LAMN for integrated diffusion processes), Ogihara [17] (LAMN for nonsynchronously observed diffusion processes), and Ogihara [18] (LAN for noisy, nonsynchronous observations of diffusion processes with deterministic diffusion coefficients). For the proof of the LAMN property, Gobet [6] introduced Malliavin calculus techniques that were effective for elliptic diffusions. One of the key ingredients of Gobet's scheme was to control the asymptotics of log-likelihood ratios by using that the transition density functions of elliptic diffusions are estimated from below and above by Gaussian density functions. Such estimates are known as Aronson's estimate. A key ingredient in Gloter and Gobet [3] to apply Gobet's scheme to one-dimensional

1350-7265

integrated diffusion models (that are two-dimensional hypoelliptic diffusions) was indeed to prove Aronson's estimate under those models. Aronson's estimate for a more general multi-dimensional hypoelliptic diffusions has recently been shown by Menozzi [15]. In general, an Aronson-type estimate is however difficult to obtain, which has been an obstacle for an application of Gobet's scheme to more general high-frequency observation models.

The original approach by Jeganathan [11] to show a LAMN property in an abstract framework, in constrast, did not need estimates for the transition density functions and instead, assumed the so-called L^2 regularity condition. Our idea in this paper is to accommodate Jaganathan's framework to high-frequency observations and then to provide an alternative scheme to Gobet's which works without Aronson-type estimates. The results in [11] are not directly applicable to high-frequency observations because they require a framework of triangular arrays. Further, for integrated diffusions, to follow an idea in [3] to deal with their hidden Markov structure, we need to consider a triangular array of expanding data blocks.

This paper studies four topics. First, we extend Theorem 1 in [11] so that it can be applied to statistical models with triangular array observations appearing in the above diffusion models with high-frequency observations. Second, we show that the new scheme based on the L^2 regularity condition can be applied under several conditions described via notions of Malliavin calculus. The new scheme is highly compatible with Gobet's scheme. Indeed, the L^2 regularity condition is satisfied when observations are smooth in the Malliavin sense, and the inverse of Malliavin matrix and its derivatives have moments (see (B1), (B2), and Theorem 2.2). Moreover, if observations admit an Euler–Maruyama approximation, then the sufficient condition for the LAMN property is simplified (Theorem 2.3). Third, by using these schemes, we prove the LAMN property for diffusion processes with degenerate diffusion coefficient (hypoelliptic diffusions). Finally, we deal with partial observations of hypoelliptic diffusion processes.

Our new schemes can be applied to general statistical models without transition density estimates. In particular, they can potentially be applied even in situations where the transition density function may have zero points. First, this scheme allows a simplified proof of the results in Gobet [6]. Moreover, this scheme yields two interesting results. The first one is an extension of the results in Gobet [6] to a wider class including hypoelliptic diffusion processes. The second one is an extension of the LAMN property for one-dimensional integrated diffusion processes in Gloter and Gobet [3] to the multi-dimensional case. We deal with the integrated diffusion process model in the general framework of partial observations for hypoelliptic diffusion processes. We find that efficient asymptotic variance is the same for an integrated diffusion process model and for a diffusion process model, which is exactly twice as large as for the complete observations of both the diffusion and its integrated processes (see Remark 2.11). Because our scheme does not require transition density estimates, we expect these ideas to be useful also for jump-diffusion process models or Lévy driven stochastic differential equation models. However, we left this for future work.

Our study of integrated diffusion models is motivated by experimental observations of single molecules (see e.g. Li et al. [20]), behind which are Langevin-type molecular dynamics

$$\ddot{Y} = b(\dot{Y}, Y) + a(\dot{Y})\dot{W}.$$

Here Y represents the position of a molecule (or a particle) and $\dot{W}$ is white noise. When $a = 0$ this reduces to the Newtonian equation of classical dynamics. The system can be written as an integrated diffusion

$$\begin{aligned} dY_t &= X_t\,dt, \\ dX_t &= b(X_t, Y_t)\,dt + a(X_t)\,dW_t. \end{aligned} \tag{1.1}$$

These models have recently attracted much attention in various contexts, and parametric inference from high-frequency observations of (X,Y) under ergodic assumptions has been studied in Melnykova [14] and Gloter and Yoshida [5]. Our LAMN property enables us to discuss optimality in estimating the coefficient a based on high-frequency observations of the position Y, without any ergodicity assumption.

The rest of this paper is organized as follows. In Section 2, we introduce our main results, namely, the extended scheme using the L^2 regularity condition, the scheme via Malliavin calculus techniques, and the LAMN property of degenerate diffusion processes. Section 3 contains details of Malliavin calculus techniques. We combine the extended scheme of the L^2 regularity condition with the approaches of Gobet [6] and Gloter and Gobet [3].

Notations

- $\bar{S}$ denotes the closure for a set S in a topological space.
- $\mathcal{B}(X)$ denotes the Borel σ-algebra of a topological space X.
- $[V]_l$ denotes the l-th element of a vector V.
- For a $k \times l$ matrix A,
  - $[A]_{ij}$ denotes (i,j) element of A,
  - $\|A\|_{\mathrm{op}}$ denotes the operator norm of A,
  - $\mathrm{Ker}(A) = \{x \in \mathbb{R}^l; Ax = 0\}$,
  - $\mathrm{Im}(A) = \{Ax; x \in \mathbb{R}^l\}$,
  - A^+ denotes the Moore–Penrose inverse,
  - $|A|$ denotes the Frobenius norm, $|A| = \sqrt{\sum_{ij} |[A]_{ij}|^2}$,
  - $A^\top$ denotes the transpose matrix of A.
- We often regard a p-dimensional vector v as a $p \times 1$ matrix.
- I_k denotes the unit matrix of size k.
- $O_{k,l}$ denotes a $k \times l$ matrix with each element equal to zero.
- For a vector $x = (x_1, \cdots, x_k)$, $\partial_x^l = (\frac{\partial^l}{\partial x_{i_1} \cdots \partial x_{i_l}})_{i_1,\cdots,i_l=1}^k$.
- We regard $\partial_x v = (\partial_{x_i} v_j)_{i,j}$ as a matrix for vectors $x = (x_i)_i$ and $v = (v_j)_j$.

2. Main results

2.1. The LAMN property via the L^2 regularity condition

In this subsection, we extend Theorem 1 in Jeganathan [11] to statistical models of triangular array observations so that it can be applied to high-frequency observations of stochastic processes.

For each $n \geq 1$, let $\{P_{\theta,n}\}_{\theta \in \Theta}$ be a family of probability measures defined on a measurable space $(\mathfrak{R}_n, \mathcal{F}_n)$, where Θ is an open subset of $\mathbb{R}^d$.

Condition (L). The following two conditions are satisfied for $\{P_{\theta,n}\}_{\theta \in \Theta}$.

1. There exists a sequence $\{V_n(\theta_0)\}$ of $\mathcal{F}_n$-measurable d-dimensional vectors and a sequence $\{T_n(\theta_0)\}$ of $\mathcal{F}_n$-measurable $d \times d$ symmetric matrices such that

$$T_n(\theta_0) \text{ is nonnegative definite } P_{\theta_0,n}\text{-almost surely} \tag{2.1}$$

for any $n \geq 1$, and

$$\log \frac{dP_{\theta_0+r_n h,n}}{dP_{\theta_0,n}} - h^\top V_n(\theta_0) + \frac{1}{2} h^\top T_n(\theta_0) h \to 0 \tag{2.2}$$

in $P_{\theta_0,n}$-probability for any $h \in \mathbb{R}^d$, where $\{r_n\}_{n=1}^{\infty}$ is a sequence of non-random, symmetric, and positive definite (p.d. in short) matrices.

2. There exists an almost surely nonnegative definite random matrix $T(\theta_0)$ such that

$$\mathcal{L}(V_n(\theta_0), T_n(\theta_0)|P_{\theta_0,n}) \to \mathcal{L}(T^{1/2}(\theta_0)W, T(\theta_0)),$$

where W is a d-dimensional standard normal random variable independent of $T(\theta_0)$.

The following definition of the LAMN property is Definition 1 in [11].

Definition 2.1. The sequence of the families $\{P_{\theta,n}\}_{\theta\in\Theta}$ $(n \geq 1)$ satisfies the LAMN condition at $\theta = \theta_0 \in \Theta$ if Condition (L) is satisfied, $T_n(\theta_0)$ is p.d. $P_{\theta_0,n}$-almost surely for any $n \geq 1$, and $T(\theta_0)$ is p.d. almost surely.

For proving the LAMN property for diffusion processes using a localization technique such as Lemma 4.1 in [6], Condition (L) is useful because (L) for the localized model often implies (L) for the original model. See the proofs of Theorems 2.4 and 2.5 for the details.

Remark 2.1. When Condition (L) is satisfied and $T(\theta_0)$ is p.d. almost surely, by setting

$$\tilde{T}_n(\theta_0) = T_n(\theta_0)1_{\{T_n(\theta_0) \text{ is p.d.}\}} + I_d 1_{\{T_n(\theta_0) \text{ is not p.d.}\}}, \tag{2.3}$$

the LAMN property holds with $\tilde{T}_n(\theta_0)$ and $V_n(\theta_0)$.

Let $(m_n)_{n=1}^{\infty}$ be a sequence of positive integers. Let $\{\mathfrak{R}_{n,j}\}_{j=1}^{m_n}$ be a sequence of complete, separable metric spaces, and let Θ be an open subset of $\mathbb{R}^d$. Let $\mathfrak{R}_n = \mathfrak{R}_{n,1} \times \cdots \times \mathfrak{R}_{n,m_n}$. We consider statistical experiments $(\mathfrak{R}_n, \mathcal{B}(\mathfrak{R}_n), \{P_{\theta,n}\}_{\theta\in\Theta})$. Let $X_j = X_{n,j} : \mathfrak{R}_n \to \mathfrak{R}_{n,j}$ be the natural projection, $\acute{X}_j = \acute{X}_{n,j} = (X_1, \dots, X_j)$, $\mathcal{F}_0 = \mathcal{F}_{n,0} = \{\emptyset, \mathfrak{R}_n\}$, and $\mathcal{F}_j = \mathcal{F}_{n,j} = \sigma(\acute{X}_j)$ for $1 \leq j \leq m_n$. Suppose that there exists a σ-finite measure $\mu_j = \mu_{n,j}$ on $\mathfrak{R}_{n,j}$ such that $P_{\theta,n}(X_1 \in \cdot) \ll \mu_1$ and $P_{\theta,n}(X_j \in \cdot|\acute{X}_{j-1} = \acute{x}_{j-1}) \ll \mu_j$ for all $\acute{x}_{j-1} \in \mathfrak{R}_{n,1} \times \cdots \times \mathfrak{R}_{n,j-1}$, $2 \leq j \leq m_n$. Let $E_\theta = E_{\theta,n}$ denote the expectation with respect to $P_{\theta,n}$, and let $p_j = p_{n,j}$ be the conditional density functions defined by

$$p_1(\theta) = \frac{dP_{\theta,n}(X_1 \in \cdot)}{d\mu_1} : \mathfrak{R}_{n,1} \to \mathbb{R}, \quad p_j(\theta) = \frac{dP_{\theta,n}(X_j \in \cdot|\acute{X}_{j-1})}{d\mu_j} : \mathfrak{R}_{n,j} \to \mathbb{R}$$

for $2 \leq j \leq m_n$. Then we can see that for $g : \mathfrak{R}_{n,1} \times \dots \mathfrak{R}_{n,j} \to \mathbb{R}$,

$$\int_{\mathfrak{R}_{n,j}} p_j(\theta) g(\acute{X}_{j-1}, x_j) d\mu_j(x_j) = E_\theta[g(\acute{X}_{j-1}, X_j)|\mathcal{F}_{j-1}]. \tag{2.4}$$

Remark 2.2. In this section, we consider the general framework of non-Markovian models such that the conditional density functions p_j depend on $\acute{X}_{j-1}$. Although diffusion processes are Markov processes, there appear some non-Markovian models in statistics for diffusion processes, for example, the partial observation models for hypoelliptic diffusion processes in Section 2.4 of this paper, and the model for diffusion processes with market microstructure noise in Gloter and Jacod [4]. The partial observation models become non-Markovian due to hidden components. To show the LAMN property for the partial observation models in this paper, we consider Markovian augmented models obtained by adding some observations, and we apply the results in this section to the augmented model, and therefore, a Markovian setting is sufficient in this case. On the other hand, in Gloter and Jacod [4], they consider ‘superexperiments’ obtained by adding some observations and ‘subexperiments’ obtained by reducing observations, to show the LAN property for diffusion process model with market microstructure noise. They show the LAN property of the original model by showing the LAN property for both ‘superexmeriments’ and ‘subexperiments’. ‘Subexperiments’ are non-Markovian, and therefore, the non-Markovian scheme to show the LAN property is required in such a setting.

Assumption (A1). There are a sequence $\{r_n\}_{n=1}^\infty$ of $d\times d$ symmetric and p.d. matrices and a measurable function $\dot{\xi}_{n,j}(\theta_0,\cdot):\Re_{n,1}\times\cdots\times\Re_{n,j}\to\mathbb{R}^d$ for each $1\le j\le m_n$ and $n\in\mathbb{N}$ such that for every $h\in\mathbb{R}^d$,

$$\sum_{j=1}^{m_n} E_{\theta_0}\left[\int [\xi_{n,j}(\theta_0,h)-\frac{1}{2}h^\top r_n\dot{\xi}_j(\theta_0)]^2 d\mu_j\right]\to 0 \tag{2.5}$$

as $n\to\infty$, where $\xi_{n,j}(\theta_0,h)=\sqrt{p_j(\theta_0+r_nh)}-\sqrt{p_j(\theta_0)}$ and $\dot{\xi}_j(\theta_0)=\dot{\xi}_{n,j}(\theta_0,\acute{X}_{j-1},\cdot)$: $\Re_{n,j}\to\mathbb{R}^d$.

Condition (A1) is the L^2 regularity condition. Jeganathan [11] established a scheme (Theorem 1) with the L^2 regularity condition to show the LAMN property, that has the advantage of not requiring estimates for the transition density functions. To illustrate this, we first review the conventional approach for diffusion processes in Gobet [6]. If p_j is smooth with respect to θ and $p_j\neq 0$, then the log-likelihood ratio is rewritten as

$$\log\frac{dP_{\theta',n}}{dP_{\theta,n}}=\sum_{j=1}^{m_n}\log\frac{p_j(\theta')}{p_j(\theta)}=\sum_{j=1}^{m_n}(\theta'-\theta)^\top\int_0^1\frac{\partial_\theta p_j}{p_j}(t\theta'+(1-t)\theta)dt.$$

To show the LAMN property, we must identify the limit distribution of this function under $P_{\theta,n}$. Doing so requires estimates for density ratios with different probability measures, which are not easy to obtain for stochastic processes in general. Gobet [6] dealt with this problem for discretely observed diffusion processes by using estimates from below and above by Gaussian density functions and show the LAMN property of that model.

On the other hand, by setting $\dot{\xi}_j(\theta_0)=\partial_\theta p_j(\theta_0)p_j(\theta_0)^{-1/2}$ and $\theta_h=\theta_0+r_nh$ for $h\in\mathbb{R}^d$, if $p_j\in C^2(\Theta)$ and $p_j(\theta_{th})\neq 0$ for any $t\in[0,1]$ μ_j-a.e., then we obtain

$$\begin{aligned}
&\int [\xi_{n,j}(\theta_0,h)-\frac{1}{2}h^\top r_n\dot{\xi}_j(\theta_0)]^2 d\mu_j\\
&=\int\left[h^\top r_n\int_0^1\frac{\partial_\theta p_j(\theta_{th})}{2\sqrt{p_j(\theta_{th})}}dt-\frac{1}{2}h^\top r_n\frac{\partial_\theta p_j(\theta_0)}{\sqrt{p_j(\theta_0)}}\right]^2 d\mu_j\\
&=\int\left[\frac{h^\top r_n}{2}\int_0^1\int_0^t\left(\frac{\partial_\theta^2 p_j(\theta_{sh})}{\sqrt{p_j(\theta_{sh})}}-\frac{\partial_\theta p_j(\partial_\theta p_j)^\top}{2p_j^{3/2}}(\theta_{sh})\right)dsdtr_nh\right]^2 d\mu_j\\
&\le\frac{1}{4}\sup_{0\le s\le 1}E_{\theta_{sh}}\left[\left\{h^\top r_n\left(\frac{\partial_\theta^2 p_j(\theta_{sh})}{p_j(\theta_{sh})}-\frac{\partial_\theta p_j(\partial_\theta p_j)^\top}{2p_j^2}(\theta_{sh})\right)r_nh\right\}^2\middle|\mathcal{F}_{j-1}\right].
\end{aligned}\tag{2.6}$$

In the right-hand side of the above inequality, the value θ_{sh} of the parameter is the same for the probability measure of expectation and p_j in the integrand, and therefore we do not need estimates for the transition density ratios. Thus, a scheme with the L^2 regularity condition does not require estimates for the transition density function. This is an advantage, and this scheme can be widely applicable to degenerate diffusion processes including partial observation models.

Define

$$\eta_j = \left(\frac{\dot{\xi}_j(\theta_0)}{\sqrt{p_j(\theta_0)}} 1_{\{p_j(\theta_0)\neq 0\}} \right)(X_j).$$

We typically set $\dot{\xi}_j(\theta_0) = \partial_\theta p_j p_j^{-1/2}(\theta_0)$ if $p_j \in C^2(\Theta)$ and $p_j(\theta_0) \neq 0$ as seen above. In this case, η_j can be simplified as $\eta_j = \partial_\theta p_j / p_j(\theta_0)$.

Assumption (A2). $E_{\theta_0}[|\eta_j|^2|\mathcal{F}_{j-1}] < \infty$ and $E_{\theta_0}[\eta_j|\mathcal{F}_{j-1}] = 0$, $P_{\theta_0,n}$-almost surely for every $j \geq 1$.
Assumption (A3). For every $\epsilon > 0$ and $h \in \mathbb{R}^d$, $\sum_{j=1}^{m_n} E_{\theta_0}[|h^\top r_n \eta_j|^2 1_{\{|h^\top r_n \eta_j|>\epsilon\}}] \to 0$.
Assumption (A4). For every $h \in \mathbb{R}^d$, there exists a constant $K > 0$ such that

$$\sup_{n\geq 1} \sum_{j=1}^{m_n} E_{\theta_0}[|h^\top r_n \eta_j|^2] \leq K.$$

Let

$$T_n = r_n \sum_{j=1}^{m_n} E_{\theta_0}[\eta_j \eta_j^\top |\mathcal{F}_{j-1}] r_n \quad \text{and} \quad V_n = r_n \sum_{j=1}^{m_n} \eta_j. \tag{2.7}$$

Assumption (A5). There exists a random $d \times d$ symmetric matrix T such that T is nonnegative definite almost surely and $\mathcal{L}((V_n,T_n)|P_{\theta_0,n}) \to \mathcal{L}(T^{1/2}W,T)$, where W is a d-dimensional standard normal random variable, and W and T are independent.
Assumption (P). T in (A5) is p.d. almost surely.

Conditions (A1)–(A4) correspond to (2.A.1), (2.A.2), (2.A.4), and (2.A.5) in [11], respectively. Condition (A5) ensures Point 2 of Condition (L). For the sequential observations in [11], convergence of $r_n \sum_{j=1}^{m_n} \eta_j$ always holds by virtue of Hall [10] (see (2.3) in [11]). However, the results of [10] cannot be applied to the triangular array observations, so we instead assume (A5) for our scheme. To check (A5), the results in Sweeting [19] are useful. Moreover, it is not difficult to check (A5) for statistical models of discretely observed diffusion processes by using a martingale central limit theorem. See, for example, Theorems 2.4 and 2.5 and their proofs.

Theorem 2.1. *Assume* (A1)–(A5). *Then (L) holds true with T_n and V_n in (2.7) for the family $\{P_{\theta,n}\}_{\theta,n}$ of probability measures. If further (P) is satisfied, then $\{P_{\theta,n}\}_{\theta,n}$ satisfies the LAMN condition at $\theta = \theta_0$ with $\tilde{T}_n$ in Remark 2.1.*

The proof is given in Section A of the supplementary material [1].

Remark 2.3. We assumed that r_n is symmetric and p.d. because this assumption is made in the definition of the LAMN property in Jeganathan [11] (Definition 1). However, we can see that Theorem 2.1 holds even if r_n is a nondegenerate asymmetric matrix. In that case, even though the assumptions of convolution theorem (Corollary 1) in [11] are not satisfied, the convolution theorem in Hájek [8] is satisfied when *local asymptotic normality* is satisfied (i.e., T in (A5) is non-random) and the operator norm of $r_n r_n^\top$ converges to zero.

2.2. The LAMN property via Malliavin calculus techniques

Gobet [6,7] used Malliavin calculus techniques to show the LAMN property for discretely observed diffusion processes. Gloter and Gobet [3] developed Gobet's scheme into a more general one and showed the LAMN property for a one-dimensional integrated diffusion process. These approaches require estimates for transition density functions by Gaussian density functions. Our alternative approach introduces tractable sufficient conditions to show the LAMN property for smooth observations in the Malliavin sense, by combining with a scheme with the L^2 regularity condition in Section 2.1. In particular, we can show the L^2 regularity condition under (B1) and (B2), which are related to the smoothness of observations and estimates for the inverse of the Malliavin matrix (Theorem 2.2). If further, observations admit a Gaussian approximation like the Euler–Maruyama approximation, then the sufficient conditions for the LAMN property are simplified as in Theorem 2.3.

We assume that Θ is convex and that m_n in Section 2.1 satisfies that $m_n \to \infty$ as $n \to \infty$. Let $(\epsilon_n)_{n=1}^\infty$ be a sequence of positive numbers and $(\Omega, \mathcal{F}, P)$ be a probability space. We set $r_n = \epsilon_n I_d$ for r_n in Section 2.1. Let $(k_j^n)_{j=0}^{m_n}$ be an increasing sequence of nonnegative integers such that $k_0^n = 0$. Hereinafter, we abbreviate k_j^n as simply k_j. Let $N_n = k_{m_n}$ and $X_j^{n,\theta}$ be an $\mathbb{R}^{k_j - k_{j-1}}$-valued random variable on $(\Omega, \mathcal{F}, P)$ for $1 \le j \le m_n$. Let $P_{\theta,n}$ be the induced probability measure by $\{X_j^{n,\theta}\}_{j=1}^{m_n}$ on $(\mathbb{R}^{N_n}, \mathcal{B}(\mathbb{R}^{N_n}))$ and $\mathcal{F}_{j,n} = \{A \times \mathbb{R}^{N_n - k_j} | A \in \mathcal{B}(\mathbb{R}^{k_j})\} \subset \mathcal{B}(\mathbb{R}^{N_n})$. For each $1 \le j \le m_n$, we adopt the notation of Nualart [16]. Specifically, let H_j be a real separable Hilbert space and $W_j = \{W_j(h), h \in H_j\}$ be an isonormal Gaussian process defined on a complete probability space $(\Omega_j, \mathcal{G}_j, Q_j)$. We assume that $\mathcal{G}_j$ is generated by W_j. Even though these objects possibly depend on n, we omit the dependence in our notation. Let δ_j be the Hitsuda–Skorokhod integral (the divergence operator), D_j be the Malliavin–Shigekawa derivative, and $\mathcal{S}_j = \{f(W_j(h_1), \cdots, W_j(h_k)); k \ge 1, h_i \in H_j \ (i = 1, \cdots, k), f \in C^\infty(\mathbb{R}^k)\}$. For a nonnegative integer k and $p \ge 1$, $\|\cdot\|_{k,p}$ denotes the operator on $\mathcal{S}_j$ defined by

$$\|F\|_{k,p} = \left[E_j[|F|^p] + \sum_{l=1}^k E_j[\|D_j^l F\|_{H_j^{\otimes l}}^p] \right]^{1/p},$$

where E_j denotes the expectation with respect to Q_j. Let $\mathbb{D}_j^{k,p}$ be the completion of $\mathcal{S}_j$ with respect to the distance $d(F, G) := \|F - G\|_{k,p}$. For general properties of W_j, D_j, and δ_j, see Nualart [16]. Let $F_{n,\theta,j,\acute{x}_{j-1}}$ be an $\mathbb{R}^{k_j - k_{j-1}}$-valued random variable on $(\Omega_j, \mathcal{G}_j)$ such that $Q_j F_{n,\theta,j,\acute{x}_{j-1}}^{-1} = P(X_j^{n,\theta} \in \cdot | \acute{X}_{j-1}^{n,\theta} = \acute{x}_{j-1})$, where $\acute{X}_{j-1}^{n,\theta} = \{X_l^{n,\theta}\}_{l=1}^{j-1}$ and $\acute{x}_{j-1} \in \mathbb{R}^{k_{j-1}}$. We assume that $F_{n,\theta,j,\acute{x}_{j-1}}$ is Fréchet differentiable with respect to θ on $L^p(\Omega_j)$ for any $p > 1$ and denote its derivative by $\partial_\theta F_{n,\theta,j,\acute{x}_{j-1}} = (\partial_{\theta_1} F_{n,\theta,j,\acute{x}_{j-1}}, \cdots, \partial_{\theta_d} F_{n,\theta,j,\acute{x}_{j-1}})^\top$. We often omit the parameter $\acute{x}_{j-1}$ in $F_{n,\theta,j,\acute{x}_{j-1}}$ and write $F_{n,\theta,j}$. Let $\bar{k}_n = \max_j (k_j - k_{j-1})$.

Remark 2.4. The dimension $k_j - k_{j-1}$ of state space depends on n because we apply Theorem 2.3 to a sequence of block observations (defined in (C.1) (C.2)) in Section C of the supplementary material [1], to obtain the LAMN property for partial observation models. In these models, $k_j - k_{j-1}$ depends on a slowly divergent sequence $(e_n)_{n=1}^\infty$.

We assume the following conditions.

Assumption (B1). $\partial_\theta^l [F_{n,\theta,j}]_i \in \cap_{p>1} \mathbb{D}_j^{4-l,p}$ for any $n, \theta, j, i, 0 \le l \le 3$, and

$$\sup_{n,i,j,\acute{x}_{j-1},\theta} \|\partial_\theta^l [F_{n,\theta,j}]_i\|_{4-l,p} < \infty$$

for $p > 1$.

Assumption (B2). The matrix $K_j(\theta) = (\langle D_j[F_{n,\theta,j}]_k, D_j[F_{n,\theta,j}]_l\rangle_{H_j})_{k,l}$ is invertible almost surely for any $j, \acute{x}_{j-1}$ and θ, and there exists a sequence $\{\alpha_n\}_{n=1}^\infty$ of positive numbers such that $\alpha_n \geq 1$ for $n \in \mathbb{N}$,

$$\sup_{i,l,j,\acute{x}_{j-1},\theta} \|[K_j^{-1}(\theta)]_{il}\|_{2,8} \leq \alpha_n, \quad \text{and} \quad \epsilon_n^2 \bar{k}_n^4 \sqrt{m_n}\alpha_n^2 \to 0 \tag{2.8}$$

as $n \to \infty$.

In (B2), m_n corresponds to the number of (block) observations, $\bar{k}_n$ corresponds to the maximum of the dimensions of the observations, ϵ_n corresponds to the convergence rate, and α_n controls the nondegeneracy of the matrix $K_j(\theta)$. The rate $\epsilon_n^2 \bar{k}_n^4 \sqrt{m_n}\alpha_n^2$ in (2.8) appears in the calculation of the left-hand side of (2.5). Therefore, the convergence in (2.8) is required to demonstrate the convergence in (A1). In the statistical model of discrete observations $(X_{k/n})_{k=0}^n$ of the m-dimensional nondegenerate diffusion process $(X_t)_{t\in[0,1]}$ in Gobet [6], we set $m_n = n$, $\bar{k}_n = m$, $\epsilon_n = 1/\sqrt{n}$, and $\alpha_n = \text{const.}$. In this case, we can check that

$$\epsilon_n^2 \bar{k}_n^4 \sqrt{m_n}\alpha_n^2 = const./\sqrt{n} \to 0$$

as $n \to \infty$. In Section 2.4, we cannot take $\bar{k}_n$ and α_n to be constants because we consider block observations, however, we can choose these sequences so that (2.8) is satisfied. We list in Table 1 at the end of this section how we choose these sequences for the model of degenerate diffusion processes in Section 2.3 and the partial observation models of the degenerate diffusion processes in Section 2.4.

Fix $\theta_0 \in \Theta$. We will see later in Proposition 3.1 that $F_{n,\theta,j}$ admits a density $p_{j,\acute{x}_{j-1}}(x_j, \theta)$ that satisfies $p_{j,\acute{x}_{j-1}}(x_j, \cdot) \in C^2(\Theta)$ almost everywhere in $x_j \in \mathbb{R}^{k_j - k_{j-1}}$ under (B1) and (B2). Let $N_j = \{x_j \in \mathbb{R}^{k_j-k_{j-1}} \,|\, \sup_{\theta\in\Theta} p_{j,\acute{x}_{j-1}}(x_j,\theta) > \inf_{\theta\in\Theta} p_{j,\acute{x}_{j-1}}(x_j,\theta) = 0\}$. We further assume the following condition.

Assumption (N1). For any $h \in \mathbb{R}^d$,

$$E_{\theta_0}\left[\sum_{j=1}^{m_n} \int_{N_j} p_{j,\acute{x}_{j-1}}(x_j, \theta_0 + r_n h)dx_j\right] \to 0$$

as $n \to \infty$.

If $\sup_{\theta\in\Theta} p_j(x_j,\theta) = 0$ or $\inf_{\theta\in\Theta} p_j(x_j,\theta) > 0$, we have $x_j \in N_j^c$. Condition (N1) says that the probability of other cases is asymptotically negligible. This condition is used to validate an estimate such as (2.6). However, if $F_{n,\theta,j}$ is approximated by a Gaussian random variable and satisfies (B3) and (N2) below, then we can check (A1) without (N1) (see Lemma 3.3).

With these definitions, the following theorem shows that the L^2 regularity condition is automatically satisfied under (B1), (B2), and (N1). Let

$$\dot{\xi}_j(\theta) = \frac{\partial_\theta p_j}{\sqrt{p_j}} 1_{\{p_j\neq 0\}}(x_j,\theta), \quad \eta_j = \frac{\partial_\theta p_j}{p_j} 1_{\{p_j \neq 0\}}(x_j, \theta_0). \tag{2.9}$$

Theorem 2.2. *Assume (B1), (B2), (N1), (A4), and (A5) with $\dot{\xi}_j(\theta)$ and η_j defined in (2.9). Then (L) holds true for $\{P_{\theta,n}\}_{\theta,n}$ at $\theta = \theta_0$ with $r_n = \epsilon_n I_d$. If further (P) is satisfied, then $\{P_{\theta,n}\}_{\theta,n}$ satisfies the LAMN condition at $\theta = \theta_0$.*

In the following, we give sufficient conditions for (A4) and (A5) when $F_{n,\theta,j}$ has a Gaussian approximation $\tilde{F}_{n,\theta,j}$.

Assumption (B3). There exist a matrix $B_{j,i,\theta} = B_{j,i,\theta,\acute{x}_{j-1},n}$ and $h_{j,l} = h_{n,\theta,j,l,\acute{x}_{j-1}} \in H_j$ $(1 \le l \le k_j - k_{j-1})$ such that $\tilde{F}_{n,\theta,j,\acute{x}_{j-1}} = (W_j(h_{j,l}))_{l=1}^{k_j - k_{j-1}}$ is Fréchet differentiable with respect to θ on L^p space, and

$$\partial_{\theta_i} \tilde{F}_{n,\theta,j} = B_{j,i,\theta} \tilde{F}_{n,\theta,j}. \tag{2.10}$$

Moreover, $\partial_\theta B_{j,i,\theta}$ exists and is continuous with respect to θ, and there exists a constant C_p and a sequence $(\rho_n)_{n \ge 1}$ of positive numbers such that

$$\sup_{n,i,j,\acute{x}_{j-1},\theta,l_1,l_2} |[\partial_\theta^l B_{j,i,\theta}]_{l_1,l_2}| < \infty,$$

and

$$\|[F_{n,\theta,j} - \tilde{F}_{n,\theta,j}]_{i'}\|_{3,p} + \|\partial_{\theta_i}[F_{n,\theta,j} - \tilde{F}_{n,\theta,j}]_{i'}\|_{2,p} \le C_p \rho_n$$

for $p > 1$, $1 \le j \le m_n$, $1 \le i \le d$, $1 \le i' \le k_j - k_{j-1}$, $\theta \in \Theta$ and $\acute{x}_{j-1}$.

Equation (2.10) is required for the following reasons. In Proposition 3.1, we obtain an expression for $\partial_\theta p_{j,\acute{x}_{j-1}}(x_j, \theta)$ using a conditional expectation. Under (B3), we replace $F_{n,\theta,j}$ and $\partial_\theta F_{n,\theta,j}$ with $\tilde{F}_{n,\theta,j}$ and $\partial_\theta \tilde{F}_{n,\theta,j}$ in the integrand of the conditional expectation, respectively. Then, the integrand is approximated by a functional of $\tilde{F}_{n,\theta,j}$ and $\partial_\theta \tilde{F}_{n,\theta,j}$. Thereafter, we can remove conditional expectation by using (2.10). As a result, η_j can be approximated by a simple quadratic form in Proposition 3.3, which is crucial to deduce the LAMN property. For a statistical model of diffusion processes, $F_{n,\theta,j}$ corresponds to normalized increments of discrete observations and $\tilde{F}_{n,\theta,j}$ corresponds their Euler–Maruyama approximations. In this case, we obtain (2.10) if the diffusion coefficient is a square matrix and nondegenerate. For the case of degenerate diffusion processes, (2.10) is satisfied under Assumption (C2) (see Section B of the supplementary material [1]).

Let $\tilde{K}_j(\theta) = (\langle h_{j,l_1}, h_{j,l_2} \rangle_{H_j})_{l_1,l_2}$. Then, we will see that for sufficiently large n, $\tilde{K}_j(\theta)$ is invertible almost surely under (B1)–(B3) and that $\alpha_n \rho_n \bar{k}_n^2 \to 0$ in Lemma 3.1 of Section 3. Let

$$\mathcal{L}_{j,i,\acute{x}_{j-1}}(u, \theta) = u^\top B_{j,i,\theta}^\top \tilde{K}_j^{-1}(\theta) u - \operatorname{tr}(B_{j,i,\theta}).$$

Let $\Phi_{j,i} = (B_{j,i,\theta_0}^\top \tilde{K}_j^{-1}(\theta_0) + \tilde{K}_j^{-1}(\theta_0) B_{j,i,\theta_0})/2$, and let

$$\gamma_j(\acute{x}_{j-1}) = (2\operatorname{tr}(\Phi_{j,i} \tilde{K}_j(\theta_0) \Phi_{j,i'} \tilde{K}_j(\theta_0)))_{i,i'=1}^d.$$

Assumption (B4). There exist $\mathbb{R}^d$-valued random variables $\{G_j^n\}_{1 \le j \le m_n, n, \theta}$ and a filtration $\{\mathcal{G}_j\}_{j=1}^{m_n}$ on $(\Omega, \mathcal{F}, P)$ such that $(X_j^{n,\theta_0}, G_j^n)$ is $\mathcal{G}_j$-measurable, $E[G_j^n | \mathcal{G}_{j-1}] = 0$, and

$$\begin{aligned} &Q_j((F_{n,\theta_0,j}, (\mathcal{L}_{j,i,\acute{x}_{j-1}}(\tilde{F}_{n,\theta_0,j}, \theta_0))_{i=1}^d) \in A)|_{\acute{x}_{j-1} = \acute{X}_{j-1}^{n,\theta_0}} \\ &= P((X_j^{n,\theta_0}, G_j^n) \in A | \mathcal{G}_{j-1}) \end{aligned} \tag{2.11}$$

for $A \in \mathcal{B}(\mathbb{R}^{k_j - k_{j-1}} \times \mathbb{R}^d)$ and sufficiently large n. Moreover,

$$\sup_n \left(\epsilon_n^2 \sum_{j=1}^{m_n} E[|\gamma_j(\acute{X}_{j-1}^{n,\theta_0})|] \right) < \infty,$$

$$\alpha_n \rho_n \bar{k}_n^2 \to 0 \quad \text{and} \quad \epsilon_n^2 m_n \alpha_n^3 \rho_n \bar{k}_n^6 \to 0 \tag{2.12}$$

as $n \to \infty$, and there exist a random $d \times d$ matrix Γ and a d-dimensional standard normal random variable $\mathcal{N}$ such that $\mathcal{N}$ and Γ are independent and

$$\left(\epsilon_n \sum_{j=1}^{m_n} G_j^n, \epsilon_n^2 \sum_{j=1}^{m_n} \gamma_j(\acute{X}_{j-1}^{n,\theta_0})\right) \xrightarrow{d} (\Gamma^{1/2}\mathcal{N}, \Gamma). \tag{2.13}$$

Assumption (N2). $[F_{n,\theta,j}]_i \in \cap_{p>1,r\geq 1} \mathbb{D}^{r,p}$ for $n,\theta,j,i,\acute{x}_{j-1}$ and $\sup_{\theta\in\Theta} \|[F_{n,\theta,j}]_i\|_{r,p} < \infty$ for any $n,i,j,\acute{x}_{j-1}$, $r \geq 1$ and $p > 1$. Also,

$$\partial_{\theta_i} D_j[\tilde{F}_{n,\theta,j}]_k = \sum_{l=1}^{k_j - k_{j-1}} [B_{j,i,\theta}]_{k,l} D_j[\tilde{F}_{n,\theta,j}]_l$$

and $\sup_{\theta\in\Theta} E_j[|\det K_j^{-1}(\theta)|^p] < \infty$ for any $p > 1$, n, j, k, $\acute{x}_{j-1}$ and $1 \leq i \leq d$.
Assumption (B5). (N1) or (N2) holds true.
Assumption (P′). Γ in (B4) is p.d. almost surely.

The condition (2.12) is required to evaluate the error in approximating η_j by a function of $\tilde{F}_{n,\theta,j}$. The rate ρ_n controls the difference between $F_{n,\theta,j}$ and $\tilde{F}_{n,\theta,j}$. For the nondegenerate diffusion process model in Gobet [6], we can set $\rho_n = 1/\sqrt{n}$. In this case, we obtain $\alpha_n \rho_n \bar{k}_n^2 = const./\sqrt{n} \to 0$, and $\epsilon_n^2 m_n \alpha_n^3 \rho_n \bar{k}_n^6 = const./\sqrt{n} \to 0$ as $n \to \infty$.

Remark 2.5. For $x = (x_1, \cdots, x_d) \in \mathbb{R}^d$, let $X_j = \sum_{i=1}^d x_i \Phi_{j,i}$. Then because

$$x^\top \gamma_j x = 2\mathrm{tr}(X_j \tilde{K}_j X_j \tilde{K}_j) = 2\mathrm{tr}(\tilde{K}_j^{1/2} X_j \tilde{K}_j X_j \tilde{K}_j^{1/2}) \geq 0,$$

γ_j is symmetric and nonnegative definite. Hence Γ is also symmetric and nonnegative definite almost surely under (B1)–(B3) and the assumption that $\epsilon_n^2 \sum_{j=1}^{m_n} \gamma_j(\acute{X}_{j-1}^{n,\theta_0}) \xrightarrow{d} \Gamma$ as $n \to \infty$.

Theorem 2.3. *Assume (B1)–(B5). Then $\{P_{\theta,n}\}_{\theta,n}$ satisfies (L) with $T(\theta_0)$ equal to Γ in (B4), $r_n = \epsilon_n I_d$, V_n defined in (2.7) and (2.9), and*

$$T_n(\theta_0) = \epsilon_n^2 \sum_{j=1}^{m_n} \gamma_j(\acute{X}_{j-1}^{n,\theta_0}). \tag{2.14}$$

Moreover,

$$\epsilon_n \sum_{j=1}^{m_n} \eta_j(\acute{X}_j^{n,\theta_0}) - \epsilon_n \sum_{j=1}^{m_n} G_j^n \xrightarrow{P} 0 \tag{2.15}$$

as $n \to \infty$. If further (P′) is satisfied, then $\{P_{\theta,n}\}_{\theta,n}$ satisfies the LAMN property at $\theta = \theta_0$ with $\tilde{T}_n$ in Remark 2.1.

Note that Theorem 2.3 works without having to identify zero points of the density function p_j, unlike in previous studies. This is useful because it is often not an easy task to show either that p_j has no zero points or that zero points are common for every θ. In the following section, we see that Theorem 2.3 can be applied to this model.

The following lemma is useful when we check (2.13) by using a martingale central limit theorem. The proof is given in Section D of the supplementary material [1].

Table 1. List of the notations related to Theorem 2.3 for each statistical model

statistical model	m_n	$\bar{k}_n$	ϵ_n	α_n	ρ_n
The nondegenerate diffusion model in Gobet [6]	n	m	$1/\sqrt{n}$	C	$1/\sqrt{n}$
The degenerate diffusion model in Section 2.3	n	m	$1/\sqrt{n}$	C	$1/\sqrt{n}$
The partial observation model in Section 2.4	$\le Cne_n^{-1}$	$\le Ce_n$	$1/\sqrt{n}$	$\iota(e_n)$	$e_n/\sqrt{n}$

∗ C denotes a generic positive constant.
∗ $(e_n)_{n=1}^{\infty}$ is a sequence of positive integers converging to infinity very slowly (defined in Section C of the supplementary material [1]).
∗ $(\iota(e_n))_{n=1}^{\infty}$ is a sequence of positive numbers that can be taken arbitrarily slowly converging to infinity by suitably choosing $(e_n)_{n=1}^{\infty}$.
∗ For the partial observation model, Theorem 2.3 is applied to an augmented model.
∗ See Sections B and C of the supplementary material [1] for the detailed discussions.

Lemma 2.1. *Assume (B1)–(B3), that $\alpha_n\rho_n\bar{k}_n^2 \to 0$, and that (2.11) is satisfied for any $A \in \mathcal{B}(\mathbb{R}^{k_j-k_{j-1}} \times \mathbb{R}^d)$. Then,*

1. $E[G_j^n(G_j^n)^\top|\mathcal{G}_{j-1}] = \gamma_j(\acute{X}_{j-1}^{n,\theta_0})$ *for any* $1 \le j \le m_n$, *and*
2. $\epsilon_n^4\sum_{j=1}^{m_n} E[|G_j^n|^4|\mathcal{G}_{j-1}] \xrightarrow{P} 0$ *as* $n \to \infty$ *if*

$$\epsilon_n^4 m_n\alpha_n^8\bar{k}_n^{12} \sup_{j,i,\acute{x}_{j-1}} \|B_{j,i,\theta_0}\tilde{K}_j(\theta_0) + \tilde{K}_j(\theta_0)B_{j,i,\theta_0}^\top\|_{\mathrm{op}}^4 \to 0. \tag{2.16}$$

2.3. The LAMN property for degenerate diffusion models

In this section, we show the LAMN property for degenerate diffusion processes by applying the results in Sections 2.1 and 2.2.

Let $r \ge 1$, and let $(\Omega,\mathcal{F},P)$ be the canonical probability space associated with an r-dimensional Wiener process $W = \{W_t\}_{t\in[0,1]}$, that is, $\Omega = C([0,1];\mathbb{R}^r)$, P is the r-dimensional Wiener measure, $W_t(\omega) = \omega(t)$ for $\omega \in \Omega$, and $\mathcal{F}$ is the completion of the Borel σ-field of Ω with respect to P. Let D be the Malliavin–Shigekawa derivative related to the underlying Hilbert space $H = L^2([0,1];\mathbb{R}^r)$. Let Θ be a bounded open convex set in $\mathbb{R}^d$.

For $\theta \in \Theta$, let $Y^\theta = (Y_t^\theta)_{t\in[0,1]}$ be an m-dimensional diffusion process satisfying $Y_0^\theta = z_{\mathrm{ini}}$, and

$$dY_t^\theta = \begin{pmatrix} \tilde{b}(Y_t^\theta,\theta) \\ \check{b}(Y_t^\theta) \end{pmatrix} dt + \begin{pmatrix} \tilde{a}(Y_t^\theta,\theta) \\ O_{m-\kappa,r} \end{pmatrix} dW_t. \tag{2.17}$$

where $z_{\mathrm{ini}} \in \mathbb{R}^m$, $m/2 \le \kappa < m$, and $\tilde{a}$, $\tilde{b}$ and $\check{b}$ are $\mathbb{R}^\kappa \otimes \mathbb{R}^r$-, $\mathbb{R}^\kappa$-, and $\mathbb{R}^{m-\kappa}$-valued Borel functions, respectively. We consider a statistical model with observations $(Y_{j/n}^\theta)_{j=0}^n$.

We assume the following conditions.

Assumption (C1). The derivatives $\partial_z^i\partial_\theta^j\tilde{a}(z,\theta)$, $\partial_z^i\partial_\theta^j\tilde{b}(z,\theta)$, and $\partial_z^i\check{b}(z)$ exist on $(z,\theta) \in \mathbb{R}^m \times \Theta$ and can be extended to continuous functions on $(z,\theta) \in \mathbb{R}^m \times \bar{\Theta}$ for $i \ge 0$ and $0 \le j \le 3$. Moreover, $\sup_{z,\theta}(|\partial_z\tilde{a}(z,\theta)| \vee |\partial_z\tilde{b}(z,\theta)| \vee |\partial_z\check{b}(z)|) < \infty$, and $\tilde{a}\tilde{a}^\top(z,\theta)$ is p.d. for any $(z,\theta) \in \mathbb{R}^m \times \bar{\Theta}$.

There exists a unique strong solution $(Y_t^\theta)_{t\in[0,1]}$ of (2.17) under (C1). Let $P_{\theta,n}$ be the distribution of $(Y_{k/n}^\theta)_{k=0}^n$, and let θ_0 be the true value of θ. We denote $Y_t = Y_t^{\theta_0}$.

We denote $z = (x, y)$ for $x \in \mathbb{R}^\kappa$ and $y \in \mathbb{R}^{m-\kappa}$, $\nabla_1 = (\partial_{z_1}, \cdots, \partial_{z_\kappa})$, and $\nabla_2 = (\partial_{z_{\kappa+1}}, \cdots, \partial_{z_m})$.

Assumption (C2). The derivative $\partial_z^i \check{b}(z)$ is bounded for $i \geq 1$, $((\nabla_1 \check{b})^\top \nabla_1 \check{b})(z)$ is invertible for $z \in \mathbb{R}^m$ and

$$\sup_{z \in \mathbb{R}^m} \|((\nabla_1 \check{b})^\top \nabla_1 \check{b})^{-1}(z)\|_{\mathrm{op}} < \infty.$$

Moreover,

$$\begin{aligned} &\mathrm{Ker}(\tilde{a}(z,\theta)) \subset \mathrm{Ker}(\partial_{\theta_i} \tilde{a}(z,\theta)), \\ &\mathrm{Ker}((\nabla_1 \check{b})^\top(z)) \subset \mathrm{Ker}((\nabla_1 \check{b})^\top(z) \partial_{\theta_i} \tilde{a}\tilde{a}^+(z',\theta)) \end{aligned} \tag{2.18}$$

for any $z, z' \in \mathbb{R}^m$, $1 \leq i \leq d$, and $\theta \in \Theta$. Furthermore, at least one of the following two conditions is satisfied;

1. $\check{b}$ is bounded;
2. $\nabla_2 \tilde{a}(z,\theta) = 0$ and $\nabla_2 \check{b}(z) = 0$ for any $z \in \mathbb{R}^m$ and $\theta \in \Theta$.

We need (C2) to satisfy $\partial_\theta \tilde{F}_{n,\theta,j} = B_{j,i,\theta} \tilde{F}_{n,\theta,j}$ in (B3). See Section B of the supplementary material [1] for the details.

We can write $\tilde{a}^+ = \tilde{a}^\top (\tilde{a}\tilde{a}^\top)^{-1}$ because $\tilde{a}\tilde{a}^\top$ is invertible. If $r = \kappa$ and (C1) is satisfied, then we can easily check $\mathrm{Ker}(\tilde{a}(z,\theta)) \subset \mathrm{Ker}(\partial_{\theta_i} \tilde{a}(z,\theta))$ because $\tilde{a}(z,\theta)$ is invertible. Similarly, we can easily check $\mathrm{Ker}((\nabla_1 \check{b})^\top(z)) \subset \mathrm{Ker}((\nabla_1 \check{b})^\top(z) \partial_{\theta_i} \tilde{a}\tilde{a}^+(z',\theta))$ if $m - \kappa = \kappa$ and $(\nabla_1 \check{b})^\top \nabla_1 \check{b}(z)$ is p.d.

Let $\Psi_{t,\theta} = (\nabla_1 \check{b})^\top \tilde{a}\tilde{a}^\top \nabla_1 \check{b}(Y_t, \theta)$, and let

$$\begin{aligned} \Gamma = \Bigg(&\frac{1}{2} \int_0^1 \mathrm{tr}((\tilde{a}\tilde{a}^\top)^{-1} \partial_{\theta_i} (\tilde{a}\tilde{a}^\top)(\tilde{a}\tilde{a}^\top)^{-1} \partial_{\theta_j} (\tilde{a}\tilde{a}^\top))(Y_t, \theta_0) dt \\ &+ \frac{1}{2} \int_0^1 \mathrm{tr}(\Psi_{t,\theta_0}^{-1} \partial_{\theta_i} \Psi_{t,\theta_0} \Psi_{t,\theta_0}^{-1} \partial_{\theta_j} \Psi_{t,\theta_0}) dt \Bigg)_{1 \leq i,j \leq d}. \end{aligned} \tag{2.19}$$

Assumption (C3). Γ is p.d. almost surely.

Theorem 2.4. *Assume (C1)–(C3). Then $\{P_{\theta,n}\}_{\theta,n}$ satisfies the LAMN property at $\theta = \theta_0$ with $r_n = n^{-1/2} I_d$ and $T_n(\theta_0) = \Gamma$ where Γ is defined in (2.19).*

Remark 2.6. If $\kappa < m/2$, (C2) is not satisfied because $(\nabla_1 \check{b})^\top \nabla_1 \check{b}$ is not invertible ($\nabla_1 \check{b}$ is $\kappa \times (m-\kappa)$ matrix and $m - \kappa > \kappa$). In that case, $\Psi_{t,\theta}$ appearing the definition of Γ is not invertible either. Therefore, we need to assume $\kappa \geq m/2$.

Remark 2.7. The proof of Theorem 2.4 in Section B of the supplementary file [1] shows that we obtain similar results when $\kappa = m$ by ignoring $\check{b}$ and $\Psi_{t,\theta}$. This approach allows another proof of the LAMN property for nondegenerate diffusion processes by Gobet [6].

Remark 2.8. The first term in the right-hand side of (2.19) is equal to Γ in Gobet [6] for nondegenerate diffusion processes when we observe $\{[Y_{j/n}]_l\}_{0 \leq j \leq n, 1 \leq l \leq \kappa}$. Then, the second term in the right-hand side of (2.19) corresponds to additional information obtained by observation $\{[Y_{j/n}]_l\}_{0 \leq j \leq n, \kappa+1 \leq l \leq m}$ for the degenerate process. Assumption (C3) is satisfied if the first term in the right-hand side of (2.19) is p.d. almost surely, and therefore, we can reduce (C3) to almost sure positive definiteness of Γ for the statistical model with the nondegenerate process. In fact, in Examples 2.1–2.4, the second term in the right-hand side of (2.19) is a scalar multiple of the first term, and therefore, the positive definiteness of Γ is derived from the positive definiteness of the first term.

Remark 2.9. Menozzi [15] showed Aronson-type estimates for models of degenerate diffusion processes including the settings of this section. So it may be possible to show the LAMN property following Gobet's approach. On the other hand, as discussed in the introduction of Kohatsu, Nualart, and Tran [13], Aronson-type estimates do not hold with fat-tailed noise like compound Poisson processes. Our approach is robust enough to show the LAMN property when asymptotically negligible noise which is independent of θ and violates Aronson-type estimates is added to (2.17). Moreover, our approach is also valid when the density functions of the noise are not specified or do not exist.

Example 2.1. Let $\kappa \geq 1$. Let $X = (X_t^\theta)_{t\in[0,1]}$ and $\bar{X} = (\bar{X}_t^\theta)_{t\in[0,1]}$ be κ-dimensional diffusion processes satisfying

$$dX_t^\theta = d(X_t^\theta, \bar{X}_t^\theta, \theta)dt + c(X_t^\theta, \theta)dW_t, \quad d\bar{X}_t^\theta = X_t^\theta dt, \quad t \in [0,1], \tag{2.20}$$

where $\theta \in \Theta \subset \mathbb{R}^d$ and $(W_t)_{t\in[0,1]}$ is a κ-dimensional standard Wiener process. We assume that $cc^\top(x,\theta)$ is p.d. and $c(x,\theta)$ and $d(z,\theta)$ are smooth functions with bounded derivatives $\partial_x c$ and $\partial_z d$. Then, (C1) and (C2) are satisfied. Γ is given by

$$\Gamma = \left(\int_0^1 \mathrm{tr}((cc^\top)^{-1}\partial_{\theta_i}(cc^\top)(cc^\top)^{-1}\partial_{\theta_j}(cc^\top))(X_t^{\theta_0}, \theta_0)dt\right)_{1\leq i,j\leq d}. \tag{2.21}$$

If further Γ is p.d. almost surely, then we obtain the LAMN property of this model by Theorem 2.4.

Example 2.1 pertains to Langevin-type molecular dynamics (1.1). Here we assumed that the position X_t^θ and velocity $\bar{X}_t^\theta$ of a molecule are observed at discrete time points. In Example 2.5 of Section 2.4, we deal with the case where we observe only the position $\bar{X}_t^\theta$.

With some restriction on the diffusion coefficient c, we can extend Example 2.1 to the case where $\dim(X_t^\theta) > \dim(\bar{X}_t^\theta)$.

Example 2.2. Let $\kappa' \leq \kappa$. Let $(X_t^\theta)_{t\in[0,1]}$ be the same as in Example 2.1, and let $(\bar{X}_t^\theta)_{t\in[0,1]}$ be a κ'-dimensional stochastic process satisfying $[\bar{X}_t^\theta]_i = \int_0^t [X_s^\theta]_i ds$ for $1 \leq i \leq \kappa'$. Moreover, the structure of the diffusion coefficient c is specific, proportional to a scalar function, that is, let $c(x,\theta) = f(x,\theta)A$ for some $\mathbb{R}$-valued function f and matrix A independent of x and θ. We assume that $AA^\top$ is p.d., f is positive-valued, and $f(x,\theta)$ and $d(z,\theta)$ are smooth functions with bounded derivatives $\partial_x f$ and $\partial_z d$. Then, (C1) and (C2) are satisfied because $(\nabla_1 \check{b})^\top = (I_{\kappa'} \; O_{\kappa',\kappa-\kappa'})$ and $\partial_\theta cc^{-1}(x,\theta) = \partial_\theta f f^{-1}(x,\theta) I_\kappa$. We have $\Psi_{t,\theta} = f^2(X_t^\theta,\theta)([AA^\top]_{ij})_{1\leq i,j\leq\kappa'}$, and hence we have

$$\begin{aligned}[\Gamma]_{ij} &= \frac{1}{2}\int_0^1 \left\{\frac{2\partial_{\theta_i} f}{f}\frac{2\partial_{\theta_j} f}{f}(X_t,\theta_0)\cdot\kappa + \frac{2\partial_{\theta_i} f}{f}\frac{2\partial_{\theta_j} f}{f}(X_t,\theta_0)\cdot\kappa'\right\}dt \\ &= 2(\kappa+\kappa')\int_0^1 \frac{\partial_{\theta_i} f \partial_{\theta_j} f}{f^2}(X_t,\theta_0)dt.\end{aligned}$$

If we only observe $(X_{k/n}^\theta)_{k=0}^n$, then Γ in Gobet [6] is calculated as

$$\Gamma = \left(2\kappa\int_0^1 \frac{\partial_{\theta_i} f \partial_{\theta_j} f}{f^2}(X_t,\theta_0)dt\right)_{1\leq i,j\leq d}.$$

Therefore, we conclude that Γ for observations X_t^θ and $\bar{X}_t^\theta$ is $(\kappa+\kappa')/\kappa$ times as much as the one for observations X_t^θ.

Remark 2.10. In the above example, even if the structure of $c(x,\theta) = f(x,\theta)A$ does not hold, we can apply Theorem 2.4 if (2.18), (C1), and (C3) hold. In that case, however, the expression of Γ is not generally expected to be equal to $(\kappa + \kappa')/\kappa$ times as much as the one for observations X_t^θ. The second term on the right-hand side of (2.19) is expressed by using $\Psi_{t,\theta} = (I_{\kappa'}\ O_{\kappa',\kappa-\kappa'})cc^\top(X_t,\theta)(I_{\kappa'}\ O_{\kappa',\kappa-\kappa'})^\top$.

Example 2.3. Let $(X_t^\theta)_{t\in[0,1]} = ((X_t^{\theta,1}, X_t^{\theta,2})^\top)_{t\in[0,1]}$ be a two-dimensional diffusion process satisfying

$$\begin{cases} dX_t^{\theta,1} = (d(X_t^{\theta,1}, X_t^{\theta,2},\theta) + e(X_t^{\theta,1}, X_t^{\theta,2}))dt + c(X_t^{\theta,1}, X_t^{\theta,2},\theta)dW_t, \\ dX_t^{\theta,2} = d(X_t^{\theta,1}, X_t^{\theta,2},\theta)dt + c(X_t^{\theta,1}, X_t^{\theta,2},\theta)dW_t, \end{cases} \tag{2.22}$$

where $\theta \in \Theta \subset \mathbb{R}^d$ and $(W_t)_{t\in[0,1]}$ is a one-dimensional standard Wiener process. That is, the diffusion coefficients of $X_t^{\theta,1}$ and $X_t^{\theta,2}$ are the same. We assume that c is positive-valued, $\sup_{x,y}|\partial_x e(x,y)|^{-1} < \infty$, and $c(x,y,\theta)$, $d(x,y,\theta)$, and $e(x,y)$ are smooth functions with bounded derivatives $\partial_z c$, $\partial_z d$, and $\partial_z^i e$ for $i \geq 1$ $(z = (x,y))$. Moreover, we assume that at least one of the following two conditions holds true:

1. e is bounded;
2. $e(x,y) = \tilde{e}(x+y)$ and $c(x,y,\theta) = \tilde{c}(x+y,\theta)$ for some functions $\tilde{e}$ and $\tilde{c}$.

Then (C1) and (C2) are satisfied by setting $Y_t^\theta = UX_t^\theta$, where

$$U = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

and Γ is given by (2.21). If further Γ is p.d. almost surely, then we obtain the LAMN property of this model by Theorem 2.4. For the statistical model with observations $(X_{j/n}^{\theta,2})_{j=0}^n$, Γ is equal to half of the one in (2.21). The above result shows that the efficient asymptotic variance for estimators does not depend on e and is equal to just half of the one when we observe $(X_{j/n}^{\theta,2})_{j=0}^n$.

Example 2.4. Let $m/2 \leq \kappa < m$. Let $(X_t^\theta)_{t\in[0,1]}$ be an m-dimensional diffusion process satisfying

$$dX_t^\theta = e(X_t^\theta)dt + f(X_t^\theta,\theta)AdW_t, \tag{2.23}$$

where $W = (W_t)_{t\in[0,1]}$ is a κ-dimensional standard Wiener process, $f(z,\theta)$ is an $\mathbb{R}$-valued function, and A is an $m \times \kappa$ matrix independent of z and θ. Let $U^\top(\Lambda\ O_{\kappa,m-\kappa})^\top V$ be the singular value decomposition of A for a $\kappa \times \kappa$ diagonal matrix Λ, and orthogonal matrices U and V of size m and κ, respectively. Then we have

$$Uf(z,\theta)A = \begin{pmatrix} \tilde{c}(Uz,\theta) \\ O_{m-\kappa,\kappa} \end{pmatrix}, \quad Ue(z) = \begin{pmatrix} \tilde{e}(Uz) \\ \check{e}(Uz) \end{pmatrix},$$

where $\tilde{c}(z,\theta) = f(U^\top z,\theta)\Lambda V$, and $\tilde{e}(z)$ and $\check{e}(z)$ are suitable functions.

We assume that $\operatorname{rank}(A) = \kappa$ (that is, Λ is invertible), f is positive–valued, $f(z,\theta)$ and $e(z)$ are smooth functions, and $\partial_z f$, $\partial_z^i e$, and $\|((\nabla_1\check{e})^\top\nabla_1\check{e})^{-1}\|_{\mathrm{op}}$ are bounded for $i \geq 0$. Then we obtain $\partial_\theta\tilde{c}\tilde{c}^{-1}(z,\theta) = \partial_\theta f f^{-1}(U^\top z,\theta)I_\kappa$, and consequently (C1) and (C2) hold with $Y_t^\theta = UX_t^\theta$. Moreover, we have

$$\Psi_{t,\theta} = f^2(X_t^{\theta_0},\theta)((\nabla_1\check{e})^\top\Lambda^2\nabla_1\check{e})(UX_t^{\theta_0})$$

and hence

$$\Gamma = \left(2m\int_0^1 \frac{\partial_{\theta_i} f \partial_{\theta_j} f}{f^2}(X_t^{\theta_0},\theta_0)dt\right)_{1\le i,j\le d}.$$

If Γ is p.d. almost surely, then we have the LAMN property of this model.

We can regard (2.23) as a multi-factor model for stock prices, where each component of W is regarded as a factor that influences the stock prices, A comprises the contributions of each factor to each stock, and f is a scalar that depends on the stock prices. The above results show that we obtain the LAMN property of such a degenerate model if the number κ of factors is in $[m/2,m)$.

2.4. The LAMN property for partial observations

In this section, we show the LAMN property for degenerate diffusion processes with partial observations. This setting includes the one-dimensional integrated diffusion model by Gloter and Gobet [3] which is a similar model to the one in Example 2.1 but the observations are only the integrated process $\bar{X}_t^{\theta_0}$ (partial observations). We extend their results of the LAMN property to multi-dimensional processes and combinatorial observations of $X_t^{\theta_0}$ and $\bar{X}_t^{\theta_0}$. Our setting also includes an interesting example of a stock process and integrated volatility observations (Example 2.6).

Let $m\ge 1$, and let $(\Omega,\mathcal{F},P)$, W, D, H, κ, and Θ be the same as in Section 2.3. We consider a process $Y_t^\theta = ((\tilde{Y}_t^\theta)^\top,(\check{Y}_t^\theta)^\top)^\top$ that satisfies a slightly restricted version of the stochastic differential equation (2.17): $\tilde{Y}_0^\theta = \tilde{z}_{\mathrm{ini}}$, $\check{Y}_0^\theta = \check{z}_{\mathrm{ini}}$, and

$$\begin{aligned} d\tilde{Y}_t^\theta &= \tilde{b}(\tilde{Y}_t^\theta,\check{Y}_t^\theta,\theta)dt + \tilde{a}(\tilde{Y}_t^\theta,\theta)dW_t, \\ d\check{Y}_t^\theta &= B\tilde{Y}_t^\theta dt, \end{aligned} \tag{2.24}$$

where B is an $(m-\kappa)\times\kappa$ matrix such that B is independent of θ, and $BB^\top$ is p.d.. Let $Q:\mathbb{R}^\kappa\to\mathbb{R}^\kappa$ be a projection. We assume that $(Q\tilde{Y}_{k/n}^{\theta_0})_{k=0}^n$ and $(\check{Y}_{k/n}^{\theta_0})_{k=0}^n$ are observed. Let $q_1=\mathrm{rank}(Q)$, $q_2=m-\kappa$, and let $q=q_1+q_2$. We assume that $0\le q_1<\kappa$.

Assumption (C2′). The derivatives $\partial_x^i\partial_\theta^j\tilde{a}(x,\theta)$ and $\partial_z^i\partial_\theta^j\tilde{b}(z,\theta)$ exist on $\mathbb{R}^m\times\Theta$ and can be extended to continuous functions on $\mathbb{R}^m\times\bar{\Theta}$ for $i\ge 0$ and $0\le j\le 3$. Moreover, $\sup_{z,\theta}(|\partial_z\tilde{b}(z,\theta)|\vee|\partial_x\tilde{a}(x,\theta)|)<\infty$, $\tilde{a}\tilde{a}^\top(x,\theta)$ is p.d., $\mathrm{Ker}(\tilde{a}(x,\theta))\subset\mathrm{Ker}(\partial_{\theta_i}\tilde{a}(x,\theta))$, and $\mathrm{Ker}(B)\subset\mathrm{Ker}(B\partial_{\theta_i}\tilde{a}\tilde{a}^+(x,\theta))$ for any $x\in\mathbb{R}^\kappa$, $1\le i\le d$, and $\theta\in\bar{\Theta}$.

Assumption (C4). For any $1\le i\le d$, $x\in\mathbb{R}^\kappa$, and $\theta\in\Theta$,

$$\mathrm{Ker}(B)\subset\mathrm{Im}(Q), \quad \text{and} \quad Q\partial_{\theta_i}\tilde{a}\tilde{a}^+(x,\theta) = \partial_{\theta_i}\tilde{a}\tilde{a}^+(x,\theta)Q. \tag{2.25}$$

By (2.25), we have

$$q_1 = \dim\mathrm{Im}(Q) \ge \dim\mathrm{Ker}(B) = \kappa - \dim\mathrm{Im}(B) = \kappa - q_2, \tag{2.26}$$

which implies $q\ge\kappa$.

Let $R_1:\mathrm{Im}(Q)\to\mathbb{R}^{q_1}$ and $R_3:\mathrm{Im}(I_\kappa - Q)\to\mathbb{R}^{\kappa-q_1}$ be any isomorphism on vector spaces. We denote $\tilde{Q}_1 = R_1Q$, $\tilde{Q}_2 = B$, and $\tilde{Q}_3 = R_3(I_\kappa - Q)$. For a $\kappa\times\kappa$ matrix A, we denote $\Upsilon_{i,j}(A) = \tilde{Q}_iA\tilde{Q}_j^\top$ for $1\le i,j\le 3$,

$$\Xi_1(A) = \begin{pmatrix} \Upsilon_{1,1} & \Upsilon_{1,2}/2 \\ \Upsilon_{2,1}/2 & \Upsilon_{2,2}/3 \end{pmatrix}(A), \quad \Xi_2(A) = \begin{pmatrix} O_{q_1,q_1} & \Upsilon_{1,2}/2 \\ O_{q_2,q_1} & \Upsilon_{2,2}/6 \end{pmatrix}(A),$$

$$\Xi_3(A) = \begin{pmatrix} \Upsilon_{1,1} & \Upsilon_{1,2}/2 \\ \Upsilon_{2,1}/2 & 2\Upsilon_{2,2}/3 \end{pmatrix}(A),$$

and for $L \geq 1$, and $1 \leq k, l \leq 2$, we define an $(Lq + (\kappa - q_1)(k-1)) \times (Lq + (\kappa - q_1)(l-1))$ matrix $\psi_L^{k,l}(A)$ by

$$\psi_L^{1,1}(A) = \begin{pmatrix} \Xi_1 & \Xi_2 & O_{q,q} & \cdots & O_{q,q} \\ \Xi_2^\top & \Xi_3 & \ddots & \ddots & \vdots \\ O_{q,q} & \ddots & \ddots & \ddots & O_{q,q} \\ \vdots & \ddots & \ddots & \Xi_3 & \Xi_2 \\ O_{q,q} & \cdots & O_{q,q} & \Xi_2^\top & \Xi_3 \end{pmatrix}(A),$$

$$\psi_L^{1,2}(A) = \begin{pmatrix} & O_{(L-1)q,\kappa-q_1} \\ \psi_L^{1,1} & \Upsilon_{1,3}/2 \\ & \Upsilon_{2,3}/6 \end{pmatrix}(A), \quad \psi_L^{2,1}(A) = (\psi_L^{1,2})^\top(A),$$

$$\psi_L^{2,2}(A) = \begin{pmatrix} & \psi_L^{1,2} \\ O_{\kappa-q_1,(L-1)q} & \Upsilon_{3,1}/2 \ \Upsilon_{3,2}/6 \ \Upsilon_{3,3}/3 \end{pmatrix}(A).$$

Here we ignore $\Upsilon_{i,j}$ if $\mathrm{rank}(\tilde{Q}_i) = 0$ or $\mathrm{rank}(\tilde{Q}_j) = 0$. Let

$$\mathcal{T}_{k,l,L}(x) = \Big(\mathrm{tr}(\partial_{\theta_i}(\psi_L^{k,k}(\tilde{a}\tilde{a}^\top)^{-1})(x,\theta_0)\psi_L^{k,l}(\tilde{a}\tilde{a}^\top)(x,\theta_0) \\ \times \partial_{\theta_j}(\psi_L^{l,l}(\tilde{a}\tilde{a}^\top)^{-1})(x,\theta_0)\psi_L^{l,k}(\tilde{a}\tilde{a}^\top)(x,\theta_0))\Big)_{1\leq i,j\leq d}.$$

Assumption (C5). There exists an $\mathbb{R}^d \otimes \mathbb{R}^d$-valued continuous function $g(x)$ such that

$$L^{-1}\mathcal{T}_{k,l,L}(x) \to g(x) \tag{2.27}$$

as $L \to \infty$ uniformly in x on compact sets for $1 \leq k, l \leq 2$.

Let

$$\Gamma' = \frac{1}{2}\int_0^1 g(\tilde{Y}_t)dt. \tag{2.28}$$

Assumption (C6). Γ' is p.d. almost surely.

The intuition for Assumption (C5) is as follows. In the partial observation model, the observation sequence is non-Markovian in general. To show the LAMN property, we consider an augmented model generated by block observations with some observations of $(I_\kappa - Q)\tilde{Y}$, following the idea of Gloter and Gobet [3]. This observation sequence for the augmented model is Markovian, and this model becomes a good approximation of the original model. Therefore, we show the LAMN property of the original model by showing the LAMN property of the augmented model. The number of observations in each block is controlled by e_n, where $(e_n)_{n=1}^\infty$ is a sequence of positive integers converging to infinity very slowly. The matrix $\psi_L^{k,l}$ corresponds to the covariance matrix of the block observations with the number of observations equal to L, and $\mathcal{T}_{k,l,L}$ is quantity corresponding to Γ for the block observations. Then, (C5) is required when we show the LAMN property for the block observations. See Sections C.2

and C.3 of the supplementary material [1] for the details. In Examples 2.5 and 2.6 below, (C5) is confirmed by using Lemma E.1 of the supplementary material [1].

Let $P_{\theta,n}$ be the distribution of partial observations $(Q\tilde{Y}^{\theta}_{k/n})_{k=0}^{n}$ and $(\check{Y}^{\theta}_{k/n})_{k=0}^{n}$.

Theorem 2.5. *Assume (C2′) and (C4)–(C6). Then $\{P_{\theta,n}\}_{\theta,n}$ satisfies the LAMN property at $\theta=\theta_0$ with $r_n=n^{-1/2}I_d$ and $T_n(\theta_0)=\Gamma'$ where Γ' is defined in (2.28).*

Example 2.5 (Integral observations). Let $(X^{\theta}_t)_{t\in[0,1]}$ and $(\bar{X}^{\theta}_t)_{t\in[0,1]}$ be the same as in Example 2.1. We consider a statistical model with observations $(\bar{X}^{\theta_0}_{k/n})_{k=0}^{n}$. In this case, we have $m=2\kappa$, $B=I_\kappa$, $Q=O_{\kappa,\kappa}$. We assume that $cc^\top(x,\theta)$ is p.d. and that $c(x,\theta)$ and $d(z,\theta)$ are smooth functions with bounded derivatives $\partial_x c$ and $\partial_z d$. As in Example 2.1, we have (C2′). Moreover, we can check (C4).

We can see that $\psi^{2,2}_L(A)=V_L\otimes A$, where $\otimes$ denotes the Kronecker product, and V_L is an $(L+1)\times(L+1)$ matrix satisfying

$$[V_L]_{ij}=(2/3)1_{\{i=j\}}+(1/6)1_{\{|i-j|=1\}}-(1/3)1_{\{i=j\text{ and }i\in\{1,L+1\}\}}. \tag{2.29}$$

Because we obtain similar equations for $\psi^{1,1}_L(A)$ and $\psi^{1,2}_L(A)$, together with Lemma E.1 in the supplementary material [1], we have (2.27) for

$$g(x)=\left(\mathrm{tr}\left((cc^\top)^{-1}\partial_{\theta_i}(cc^\top)(cc^\top)^{-1}\partial_{\theta_j}(cc^\top)\right)(x,\theta_0)\right)_{1\le i,j\le d}. \tag{2.30}$$

Therefore, we have the LAMN property of this model if Γ' in (2.28) is p.d. almost surely.

This result is an extension of Gloter and Gobet [3] to multi-dimensional processes. Moreover, the result can be applied to the Langevin-type molecular dynamics in (1.1) with positional observations.

Remark 2.11. If we observe $(X^{\theta_0}_{k/n})_{k=0}^{n}$ instead of $(\bar{X}^{\theta_0}_{k/n})_{k=0}^{n}$, then Gobet [6] shows the LAMN property for this model with Γ the same as (2.28) and (2.30). On the other hand, if we observe both $(X^{\theta_0}_{k/n})_{k=0}^{n}$ and $(\bar{X}^{\theta_0}_{k/n})_{k=0}^{n}$, then Example 2.1 shows the LAMN property with Γ twice that in (2.28). Therefore, we can say that the efficient asymptotic variance with observations $(X^{\theta_0}_{k/n})_{k=0}^{n}$ and $(\bar{X}^{\theta_0}_{k/n})_{k=0}^{n}$ is half of that with observations $(X^{\theta_0}_{k/n})_{k=0}^{n}$ or half of that with $(\bar{X}^{\theta_0}_{k/n})_{k=0}^{n}$.

Example 2.6 (Observations of a stock process and integrated volatility). Let $(W_t)_{t\in[0,1]}$ be a two-dimensional standard Wiener process, and let c be an $\mathbb{R}^2\otimes\mathbb{R}^2$-valued function with the j-th row vector c^j for $j\in\{1,2\}$. Let $(X_t)_{t\in[0,1]}=((X^1_t,X^2_t,X^3_t)^\top)_{t\in[0,1]}$ be a three-dimensional process satisfying

$$\begin{aligned} dX^1_t&=d^1(X_t,\theta)dt+c^1(X^1_t,X^2_t,\theta)dW_t,\\ dX^2_t&=d^2(X_t,\theta)dt+c^2(X^1_t,X^2_t,\theta)dW_t,\\ dX^3_t&=X^2_t dt. \end{aligned} \tag{2.31}$$

We assume that we observe $((X^1_{k/n},X^3_{k/n})^\top)_{k=0}^{n}$. In this case, we have $m=3$, $\kappa=2$, $r=2$,

$$Q=\begin{pmatrix}1&0\\0&0\end{pmatrix},\quad B=(0\ 1). \tag{2.32}$$

We assume that $cc^\top(x,\theta)$ is p.d. for each (x,θ), and $c(x,\theta)$, $d^1(z,\theta)$, and $d^2(z,\theta)$ are smooth functions with bounded derivatives $\partial_x c$, $\partial_z d^1$, and $\partial_z d^2$. We can check (2.25). We regard X^1 as a stock process and X^3 as the integrated volatility process. If we observe daily stock prices and realized volatility calculated from high-frequency data, then we can regard it as an approximation of the integrated volatility process.

We consider the following two cases.

1. The case where $c(x_1,x_2,\theta) = f(x_1,x_2,\theta)A$ for a matrix A and a positive–valued function $f(x_1,x_2,\theta)$:

   We have that $AA^\top$ is p.d. and $\partial_{\theta_i}\tilde{a}\tilde{a}^+ = \partial_{\theta_i}cc^{-1} = \partial_{\theta_i}ff^{-1}I_2$. Then (C2′) and (C4) are satisfied. Moreover, we obtain

$$\psi_L^{k,l}(\partial_{\theta_i}(\tilde{a}\tilde{a}^\top)) = \frac{2\partial_{\theta_i}f}{f}\psi_L^{k,l}(\tilde{a}\tilde{a}^\top). \tag{2.33}$$

   Together with Lemma E.1 in the supplementary material [1], we have (C5) with

$$g(x_1,x_2) = \left(\frac{8\partial_{\theta_i}f\partial_{\theta_j}f}{f^2}(x_1,x_2,\theta_0)\right)_{i,j}.$$

   Therefore, we have the LAMN property if Γ' in (2.28) is p.d. almost surely.

2. The case where $c(x_1,x_2,\theta)$ is a diagonal matrix for any (x_1,x_2,θ):

   Because $\partial_{\theta_i}\tilde{a}\tilde{a}^+$ also becomes a diagonal matrix, (C2′) and (C4) are satisfied. Moreover, we have $\Upsilon_{1,2} = 0$ and $\Upsilon_{1,3} = 0$. Then, by rearranging the rows and columns of $\psi_L^{2,2}$ by using an orthogonal matrix $\mathcal{V}_L$ of size $2L+1$, we have

$$\mathcal{V}_L\psi_L^{2,2}(M)\mathcal{V}_L^\top = \begin{pmatrix} [M]_{11}I_L & O_{L,L} \\ O_{L,L} & [M]_{22}V_L \end{pmatrix} \tag{2.34}$$

   for any diagonal matrix M of size 2, where V_L is defined in (2.29). Together with Lemma E.1 in the supplementary material [1] and similar equations for $\psi_L^{1,2}$ and $\psi_L^{1,1}$, we have (2.27) with

$$g(x_1,x_2) = \left(\left(\frac{4\partial_{\theta_i}[c]_{11}\partial_{\theta_j}[c]_{11}}{[c]_{11}^2} + \frac{4\partial_{\theta_i}[c]_{22}\partial_{\theta_j}[c]_{22}}{[c]_{22}^2}\right)(x_1,x_2,\theta_0)\right)_{1\le i,j\le d}. \tag{2.35}$$

   Then we have the LAMN property if Γ' in (2.28) is p.d. almost surely.

3. Malliavin calculus and the L^2 regularity condition

In this section, we show how to check (A1)–(A5) in Section 2.1 under (B1)–(B5). The equations for density derivatives in Proposition 3.1 are crucial for the proof. From these equations, we obtain Proposition 3.2 and Lemma 3.2, which are necessary for checking (A1).

Let

$$L^\theta(V) = \sum_{k,k'}[K_j^{-1}(\theta)]_{k,k'}D_j[F_{n,\theta,j}]_{k'}[V]_k$$

for a vector $V \in \mathbb{R}^{k_j - k_{j-1}}$.

The following proposition is essentially from Proposition 4.1 in [6] and Theorem 5 in [3]. To check (A1), we need an equation for $\partial_\theta^2 p_j$. The proof is given in Section D of the supplementary material [1].

Proposition 3.1. *Assume (B1) and (B2). Then $F_{n,\theta,j}$ admits a density denoted by $p_{j,\acute{x}_{j-1}}(x_j,\theta)$. Moreover, $p_{j,\acute{x}_{j-1}}(x_j,\cdot)\in C^2(\Theta)$,*

$$\partial_\theta p_{j,\acute{x}_{j-1}}(x_j,\theta)=p_{j,\acute{x}_{j-1}}(x_j,\theta)E_j\left[\delta_j(L^\theta(\partial_\theta F_{n,\theta,j}))\middle|F_{n,\theta,j}=x_j\right], \tag{3.1}$$

and

$$\partial_\theta^2 p_{j,\acute{x}_{j-1}}(x_j,\theta)=p_{j,\acute{x}_{j-1}}(x_j,\theta)E_j\left[\delta_j(L^\theta(\partial_\theta^2 F_{n,\theta,j}))+\delta_j(L^\theta(\mathfrak{A}_j))\middle|F_{n,\theta,j}=x_j\right] \tag{3.2}$$

almost everywhere in $x_j\in\mathbb{R}^{k_j-k_{j-1}}$, where $\mathfrak{A}_j=(\delta_j(L^\theta(\partial_\theta F_{n,\theta,j}\partial_\theta[F_{n,\theta,j}]_k)))_k$.

The proof of the following proposition is given in Section D of the supplementary material [1].

Proposition 3.2. *Assume (B1) and (B2). Then*

$$\sup_{i,j,\acute{x}_{j-1},\theta} E_j[|\partial_{\theta_i}p_{j,\acute{x}_{j-1}}/p_{j,\acute{x}_{j-1}}|^4 1_{\{p_{j,\acute{x}_{j-1}}\neq 0\}}(F_{n,\theta,j},\theta)]^{1/4}\le C\alpha_n\bar{k}_n^2, \tag{3.3}$$

$$\sup_{i,l,j,\acute{x}_{j-1},\theta} E_j[|\partial_{\theta_i}\partial_{\theta_l}p_{j,\acute{x}_{j-1}}/p_{j,\acute{x}_{j-1}}|^2 1_{\{p_{j,\acute{x}_{j-1}}\neq 0\}}(F_{n,\theta,j},\theta)]^{1/2}\le C\alpha_n^2\bar{k}_n^4. \tag{3.4}$$

Let $\theta_h=\theta_0+\epsilon_n h$ for $h\in\mathbb{R}^d$,

$$\mathcal{E}_j^1(x_j,\theta)=\mathcal{E}_j^1(x_j,\theta,\acute{x}_{j-1})=(E_j[\delta_j(L^\theta(\partial_{\theta_i}F_{n,\theta,j,\acute{x}_{j-1}}))|F_{n,\theta,j,\acute{x}_{j-1}}=x_j])_{i=1}^d,$$

and $\mathcal{E}_j^2(x_j,\theta)$ be a $d\times d$ random matrix with elements

$$\begin{aligned}[\mathcal{E}_j^2(x_j,\theta)]_{il}&=E_j[\delta_j(L^\theta(\partial_{\theta_i}\partial_{\theta_l}F_{n,\theta,j}))|F_{n,\theta,j}=x_j]\\&\quad+E_j\left[\delta_j\big(L^\theta\big((\delta_j(L^\theta(\partial_{\theta_i}F_{n,\theta,j}\partial_{\theta_l}[F_{n,\theta,j}]_k)))_k\big)\big)\middle|F_{n,\theta,j}=x_j\right].\end{aligned}$$

We set the conditional expectations equal to zero when $p_j(x_j,\theta)=0$. Then $\mathcal{E}_j^1(x_j,\theta)$ and $\mathcal{E}_j^2(x_j,\theta)$ are measurable with respect to θ almost everywhere in x_j because Proposition 3.1 yields

$$[\mathcal{E}_j^1(x_j,\theta)]_i=(\partial_{\theta_i}p_j/p_j)1_{\{p_j\neq 0\}}\quad\text{and}\quad[\mathcal{E}_j^2(x_j,\theta)]_{il}=(\partial_{\theta_i}\partial_{\theta_l}p_j/p_j)1_{\{p_j\neq 0\}}. \tag{3.5}$$

Proof of Theorem 2.2. We check (A1)–(A3) in Theorem 2.1 by setting $p_j(\theta)=p_j(x_j,\theta)=p_{j,\acute{x}_{j-1}}(x_j,\theta)$.

For sufficiently large n, we have $\{\theta_{th}\}_{t\in[0,1]} \subset \Theta$. (N1), Proposition 3.2, and the Cauchy-Schwarz inequality yield

$$E_{\theta_0}\left[\sum_{j=1}^{m_n}\int_{N_j}\left[\sqrt{p_j}(x_j,\theta_h)-\sqrt{p_j}(x_j,\theta_0)-\frac{\epsilon_n}{2}h^\top\frac{\partial_\theta p_j}{\sqrt{p_j}}1_{\{p_j\neq 0\}}(x_j,\theta_0)\right]^2 dx_j\right]$$

$$\leq CE_{\theta_0}\left[\sum_{j=1}^{m_n}\int_{N_j}\left[p_j(x_j,\theta_h)+p_j(x_j,\theta_0)+\frac{\epsilon_n^2}{4}\left(\frac{h^\top\partial_\theta p_j}{p_j}\right)^2 p_j 1_{\{p_j\neq 0\}}(x_j,\theta_0)\right]dx_j\right]$$

$$\leq o(1)+C\epsilon_n^2\sqrt{E_{\theta_0}\left[\sum_{j=1}^{m_n}\int\left|\frac{\partial_\theta p_j}{p_j}\right|^4 p_j 1_{\{p_j\neq 0\}}(x_j,\theta_0)dx_j\right]}\sqrt{E_{\theta_0}\left[\sum_{j=1}^{m_n}\int_{N_j}p_j(x_j,\theta_0)dx_j\right]}$$

$$\leq o(1)+C\epsilon_n^2\sqrt{m_n}\alpha_n^2\bar{k}_n^4\times o(1)$$

as $n\to\infty$. The right-hand side of the above inequality converges to zero by (2.8).

Moreover, similarly to (2.6), we have

$$\int_{N_j^c}\left\{\sqrt{p_j}(x_j,\theta_h)-\sqrt{p_j}(x_j,\theta_0)-\frac{\epsilon_n}{2}h^\top\dot{\xi}_j(\theta_0)\right\}^2 dx_j$$

$$\leq \frac{\epsilon_n^4|h|^4}{4}\int_0^1 E\left[\left\|\frac{\partial_\theta^2 p_j}{p_j}-\frac{\partial_\theta p_j\partial_\theta p_j^\top}{2p_j^2}\right\|_{\mathrm{op}}^2 1_{\{p_j\neq 0\}}(F_{n,\theta_{sh},j},\theta_{sh})\right]ds.$$

Together with Proposition 3.2 and (2.8), we have

$$\sum_{j=1}^{m_n}E\left[\int_{N_j^c}\left\{\sqrt{p_j}(x_j,\theta_h)-\sqrt{p_j}(x_j,\theta_0)-\frac{\epsilon_n}{2}h^\top\dot{\xi}_j(\theta_0)\right\}^2 dx_j\right]\to 0, \tag{3.6}$$

which implies (A1).

Moreover, we have (A2) because

$$E_{\theta_0}\left[\frac{\partial_\theta p_j}{p_j}1_{\{p_j\neq 0\}}(x_j,\theta_0)\middle|\mathcal{F}_{j-1}\right]=\int\partial_\theta p_j(x_j,\theta_0)dx_j=0,$$

by Proposition 3.1, where $\mathcal{F}_{j-1}$ is the one in Section 2.1.

Further, Proposition 3.2 yields (A3). □

In the following, we prove Theorem 2.3. To show (A5), we replace $\mathcal{E}_j^1(F_{n,\theta,j},\theta)$ by $(\mathcal{L}_{j,i,\acute{x}_{j-1}}(\tilde{F}_{n,\theta,j}))_{i=1}^d$, and then we apply (B4). For that purpose, we first estimate the difference between K_j and $\tilde{K}_j$.

Lemma 3.1. *Assume (B1)–(B3) and that $\alpha_n\rho_n\bar{k}_n^2\to 0$. Then, for any $1\leq j\leq m_n$ and $p>1$, $\tilde{K}_j(\theta)$ is an invertible matrix almost surely and satisfies*

$$\sup_{i,l,j,\acute{x}_{j-1},\theta}\|[K_j(\theta)-\tilde{K}_j(\theta)]_{il}\|_{2,p}\leq C_p\rho_n,\qquad \sup_{j,\acute{x}_{j-1},\theta}\|\tilde{K}_j^{-1}(\theta)\|_{\mathrm{op}}\leq C\alpha_n\bar{k}_n \tag{3.7}$$

for sufficiently large n.

The proof is given in Section D of the supplementary material [1].

Proposition 3.3. *Assume (B1)–(B3) and that $\alpha_n\rho_n\bar{k}_n^2 \to 0$ as $n \to \infty$. Then there exists a positive constant C such that*

$$\sup_{i,j,\acute{x}_{j-1},\theta} E_j\left[\left|\frac{\partial_{\theta_i} p_{j,\acute{x}_{j-1}}}{p_{j,\acute{x}_{j-1}}} 1_{\{p_j\neq 0\}}(F_{n,\theta,j},\theta) - \mathcal{L}_{j,i,\acute{x}_{j-1}}(\tilde{F}_{n,\theta,j},\theta)\right|^2\right]^{1/2} \leq C\alpha_n^2\rho_n\bar{k}_n^4 \tag{3.8}$$

for sufficiently large n.

Proof. For $V \in (\mathbb{D}_j^{1,p})^{k_j-k_{j-1}}$, we regard $D_jV = (D_j[V]_l)_l$ as a vector of size $k_j - k_{j-1}$. Let $\mathbf{L}_{j,i}^\theta = \partial_{\theta_i}\tilde{F}_{n,\theta,j}^\top \tilde{K}_j^{-1}(\theta) D_j\tilde{F}_{n,\theta,j}$. First, we show that

$$\sup_{i,j,\acute{x}_{j-1},\theta} \|L^\theta(\partial_{\theta_i}F_{n,\theta,j}) - \mathbf{L}_{j,i}^\theta\|_{\mathbb{D}^{1,4}(H_j)} \leq C\alpha_n^2\rho_n\bar{k}_n^4. \tag{3.9}$$

Conditions (B1)–(B3) and Lemma 3.1 yield estimates for

$$(\partial_{\theta_i}F_{n,\theta,j} - \partial_{\theta_i}\tilde{F}_{n,\theta,j})^\top K_j^{-1} D_jF_{n,\theta,j} \quad \text{and} \quad \partial_{\theta_i}\tilde{F}_{n,\theta,j}^\top \tilde{K}_j^{-1}(D_jF_{n,\theta,j} - D_j\tilde{F}_{n,\theta,j}).$$

Because $K_j^{-1} - \tilde{K}_j^{-1} = \tilde{K}_j^{-1}(\tilde{K}_j - K_j)K_j^{-1}$, we also obtain an estimate for $\partial_{\theta_i}\tilde{F}_{n,\theta,j}^\top(K_j^{-1} - \tilde{K}_j^{-1})D_jF_{n,\theta,j}$. Then we have (3.9).

Moreover, Proposition 1.3.3 in Nualart [16] and (B3) yield

$$\begin{aligned} \delta_j(\mathbf{L}_{j,i}^\theta) &= \partial_{\theta_i}\tilde{F}_{n,\theta,j}^\top \tilde{K}_j^{-1}\delta_j(D_j\tilde{F}_{n,\theta,j}) - \mathrm{tr}(\tilde{K}_j^{-1}\langle D_j\partial_{\theta_i}\tilde{F}_{n,\theta,j}, D_j\tilde{F}_{n,\theta,j}\rangle_{H_j}) \\ &= \tilde{F}_{n,\theta,j}^\top B_{j,i,\theta}^\top \tilde{K}_j^{-1}\tilde{F}_{n,\theta,j} - \mathrm{tr}(\tilde{K}_j^{-1}B_{j,i,\theta}\tilde{K}_j) = \mathcal{L}_{j,i,\acute{x}_{j-1}}(\tilde{F}_{n,\theta,j},\theta). \end{aligned} \tag{3.10}$$

Together with Proposition 3.1, we have

$$\begin{aligned} &\frac{\partial_{\theta_i}p_j}{p_j} 1_{\{p_j\neq 0\}}(F_{n,\theta,j}) - \mathcal{L}_{j,i,\acute{x}_{j-1}}(\tilde{F}_{n,\theta,j},\theta) \\ &\quad = E_j[\delta_j(L^\theta(\partial_{\theta_i}F_{n,\theta,j}) - \mathbf{L}_{j,i}^\theta)|F_{n,\theta,j}] + E_j[r_i|F_{n,\theta,j}] - r_i \end{aligned}$$

almost surely, where $r_i = \mathcal{L}_{j,i,\acute{x}_{j-1}}(\tilde{F}_{n,\theta,j},\theta) - \mathcal{L}_{j,i,\acute{x}_{j-1}}(F_{n,\theta,j},\theta)$. Then we obtain (3.8) by (B3), (3.9), and the fact that

$$E_j[|r_i|^p]^{1/p} \leq C_p\alpha_n\rho_n\bar{k}_n^3 \tag{3.11}$$

for any $p \geq 1$. □

Lemma 3.2. *Assume (B1), (B2), and (N2). Then, for any $n \geq 1$, $1 \leq j \leq m_n$, and $h \in \mathbb{R}^d$ satisfying $\{\theta_{th}\}_{t\in[0,1]} \subset \Theta$, the function $\sqrt{p_{j,\acute{x}_{j-1}}}(x_j,\theta_{th})$ is absolutely continuous on $t \in [0,1]$ almost everywhere in x_j.*

The proof is given in Section D of the supplementary material [1].

Lemma 3.3. *Assume (B1)–(B3), (B5), and (2.12). Then (A1) holds true.*

The proof is given in Section D of the supplementary material [1].

Proof of Theorem 2.3. Thanks to Remark 2.1, Lemma 3.3, and the proof of Theorem 2.2, to show (L), it is sufficient to check (A4) and (A5) under (B1)–(B5). Let $X_j = X_j^{n,\theta_0}$, $\acute{X}_{j-1} = (X_1, \cdots, X_{j-1})$, and $\mathcal{H}_j = E[\mathcal{E}_j^1(\mathcal{E}_j^1)^\top(X_j, \theta_0, \acute{X}_{j-1})|\sigma(\acute{X}_{j-1})]$. Then by (3.5), it suffices to show that

$$\sup_n \left(\epsilon_n^2 \sum_{j=1}^{m_n} E[|\mathcal{H}_j|]\right) < \infty \tag{3.12}$$

and

$$\sum_{j=1}^{m_n} (\epsilon_n \mathcal{E}_j^1(X_j, \theta_0, \acute{X}_{j-1}), \epsilon_n^2 \mathcal{H}_j) \xrightarrow{d} (\Gamma^{1/2}\mathcal{N}, \Gamma). \tag{3.13}$$

For sufficiently large n, (2.11) and Proposition 3.3 yield

$$\begin{aligned} &E[|\mathcal{E}_j^1(X_j, \theta_0, \acute{X}_{j-1}) - G_j^n|^2 | \mathcal{G}_{j-1}] \\ &\quad = E_j[|\mathcal{E}_j^1(F_{n,\theta_0,j}, \theta_0, \acute{x}_{j-1}) - (\mathcal{L}_{j,i,\acute{x}_{j-1}}(\tilde{F}_{n,\theta_0,j}, \theta_0))_{i=1}^d|^2]|_{\acute{x}_{j-1} = \acute{X}_{j-1}} \\ &\quad \leq C\alpha_n^4 \rho_n^2 \bar{k}_n^8. \end{aligned}$$

Together with (2.12) and (B4), we obtain

$$\begin{aligned} E\left[\left|\epsilon_n \sum_{j=1}^{m_n} \mathcal{E}_j^1(X_j, \theta_0, \acute{X}_{j-1}) - \epsilon_n \sum_{j=1}^{m_n} G_j^n\right|^2\right] &\leq \epsilon_n^2 \sum_{j=1}^{m_n} E\left[E[|\mathcal{E}_j^1(X_j, \theta_0, \acute{X}_{j-1}) - G_j^n|^2 | \mathcal{G}_{j-1}]\right] \\ &= O(\epsilon_n^2 m_n \alpha_n^4 \rho_n^2 \bar{k}_n^8) \to 0 \end{aligned} \tag{3.14}$$

as $n \to \infty$.

Let $\mathfrak{F}_j = (\mathcal{L}_{j,i,\acute{x}_{j-1}}(\tilde{F}_{n,\theta_0,j}))_{i=1}^d$, then we have

$$\gamma_j(\acute{x}_{j-1}) = E_j[\mathfrak{F}_j \mathfrak{F}_j^\top], \tag{3.15}$$

and $\sup_{\acute{x}_{j-1}} E_j[|\mathfrak{F}_j|^2]^{1/2} = O(\alpha_n \bar{k}_n^2)$ by (D.7) of the supplementary material [1]. Together with (2.12) and Propositions 3.2 and 3.3, we have

$$\begin{aligned} &\sup_{\acute{x}_{j-1}} |E_j[\mathcal{E}_j^1(\mathcal{E}_j^1)^\top(F_{n,\theta_0,j}, \theta_0, \acute{x}_{j-1})] - \gamma_j(\acute{x}_{j-1})| \\ &\quad \leq C \sup_{\acute{x}_{j-1}} \left(E_j[|\mathcal{E}_j^1|^2]^{\frac{1}{2}} E_j[|\mathcal{E}_j^1 - \mathfrak{F}_j|^2]^{\frac{1}{2}} + E_j[|\mathfrak{F}_j|^2]^{\frac{1}{2}} E_j[|\mathcal{E}_j^1 - \mathfrak{F}_j|^2]^{\frac{1}{2}}\right) \\ &\quad = O(\alpha_n \bar{k}_n^2 \cdot \alpha_n^2 \rho_n \bar{k}_n^4) = o(\epsilon_n^{-2} m_n^{-1}). \end{aligned} \tag{3.16}$$

Then, (3.14), (3.16), and (B4) yield

$$\sup_n \left(\epsilon_n^2 \sum_{j=1}^{m_n} E[|\mathcal{H}_j|]\right) = \sup_n \left(\epsilon_n^2 \sum_{j=1}^{m_n} E[|\gamma_j(\acute{X}_{j-1})|] + O(1)\right) < \infty$$

and

$$\sum_{j=1}^{m_n} (\epsilon_n \mathcal{E}_j^1(X_j, \theta_0, \acute{X}_{j-1}), \epsilon_n^2 \mathcal{H}_j) = \sum_{j=1}^{m_n} (\epsilon_n G_j^n, \epsilon_n^2 \gamma_j(\acute{X}_{j-1})) + o_p(1) \xrightarrow{d} (\Gamma^{1/2}\mathcal{N}, \Gamma). \tag{3.17}$$

Moreover, we can define $T_n(\theta_0)$ by (2.14), and (2.15) holds because of (3.17). □

Acknowledgements

The authors would like to thank the anonymous referees, an Associate Editor and the Editor for their constructive comments that improved the quality of this paper.

Funding

The second author was supported partially by Japan Science and Technology Agency, PRESTO Grant Number JPMJPR15E2, and partially by Japan Society for he Promotion of Science KAKENHI Grant Number 19K14604.

Supplementary Material

Supplement to "Malliavin calculus techniques for local asymptotic mixed normality and their application to degenerate diffusions" (DOI: 10.3150/23-BEJ1621SUPP; .pdf). The supplementary material gives the proofs of Theorem 2.1 and the results in Sections 2.3, 2.4, and 3.

References

[1] Fukasawa, M. and Ogihara, T. (2024). Supplement to "Malliavin calculus techniques for local asymptotic mixed normality and their application to hypoelliptic diffusions." https://doi.org/10.3150/23-BEJ1621SUPP

[2] Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **29** 119–151. MR1204521

[3] Gloter, A. and Gobet, E. (2008). LAMN property for hidden processes: The case of integrated diffusions. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 104–128. MR2451573 https://doi.org/10.1214/07-AIHP111

[4] Gloter, A. and Jacod, J. (2001). Diffusions with measurement errors. I. Local asymptotic normality. *ESAIM Probab. Stat.* **5** 225–242. MR1875672 https://doi.org/10.1051/ps:2001110

[5] Gloter, A. and Yoshida, N. (2021). Adaptive estimation for degenerate diffusion processes. *Electron. J. Stat.* **15** 1424–1472. MR4255288 https://doi.org/10.1214/20-ejs1777

[6] Gobet, E. (2001). Local asymptotic mixed normality property for elliptic diffusion: A Malliavin calculus approach. *Bernoulli* **7** 899–912. MR1873834 https://doi.org/10.2307/3318625

[7] Gobet, E. (2002). LAN property for ergodic diffusions with discrete observations. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 711–737. MR1931584 https://doi.org/10.1016/S0246-0203(02)01107-X

[8] Hájek, J. (1969/70). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. Verw. Gebiete* **14** 323–330. MR0283911 https://doi.org/10.1007/BF00533669

[9] Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. I: Theory of Statistics* 175–194. Berkeley, CA: Univ. California Press. MR0400513

[10] Hall, P. (1977). Martingale invariance principles. *Ann. Probab.* **5** 875–887. MR0517471 https://doi.org/10.1214/aop/1176995657

[11] Jeganathan, P. (1982). On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal. *Sankhyā Ser. A* **44** 173–212. MR0688800

[12] Jeganathan, P. (1983). Some asymptotic properties of risk functions when the limit of the experiment is mixed normal. *Sankhyā Ser. A* **45** 66–87. MR0749355

[13] Kohatsu-Higa, A., Nualart, E. and Tran, N.K. (2017). LAN property for an ergodic diffusion with jumps. *Statistics* **51** 419–454. MR3609328 https://doi.org/10.1080/02331888.2016.1239727

[14] Melnykova, A. (2020). Parametric inference for hypoelliptic ergodic diffusions with full observations. *Stat. Inference Stoch. Process.* **23** 595–635. MR4136706 https://doi.org/10.1007/s11203-020-09222-4

[15] Menozzi, S. (2011). Parametrix techniques and martingale problems for some degenerate Kolmogorov equations. *Electron. Commun. Probab.* **16** 234–250. MR2802040 https://doi.org/10.1214/ECP.v16-1619

[16] Nualart, D. (2006). *The Malliavin Calculus and Related Topics*, 2nd ed. *Probability and Its Applications (New York)*. Berlin: Springer. MR2200233

[17] Ogihara, T. (2015). Local asymptotic mixed normality property for nonsynchronously observed diffusion processes. *Bernoulli* **21** 2024–2072. MR3378458 https://doi.org/10.3150/14-BEJ634

[18] Ogihara, T. (2018). Parametric inference for nonsynchronously observed diffusion processes in the presence of market microstructure noise. *Bernoulli* **24** 3318–3383. MR3788175 https://doi.org/10.3150/17-BEJ962

[19] Sweeting, T.J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Statist.* **8** 1375–1381. MR0594652

[20] Tongcang, L., Simon, K., David, M. and Mark, G.R. (2010). Measurement of the instantaneous velocity of a Brownian particle. *Science* **328** 1673–1675.

Received November 2020 and revised March 2023