Bernoulli **27**(4), 2021, 2532–2555
https://doi.org/10.3150/20-BEJ1321

Scalable Monte Carlo inference and rescaled local asymptotic normality

NING NING*, EDWARD L. IONIDES† and YA'ACOV RITOV‡

Department of Statistics, University of Michigan, Ann Arbor.
E-mail: *patning@umich.edu; †ionides@umich.edu; ‡yritov@umich.edu

In this paper, we generalize the property of local asymptotic normality (LAN) to an enlarged neighborhood, under the name of rescaled local asymptotic normality (RLAN). We obtain sufficient conditions for a regular parametric model to satisfy RLAN. We show that RLAN supports the construction of a statistically efficient estimator which maximizes a cubic approximation to the log-likelihood on this enlarged neighborhood. In the context of Monte Carlo inference, we find that this maximum cubic likelihood estimator can maintain its statistical efficiency in the presence of asymptotically increasing Monte Carlo error in likelihood evaluation.

Keywords: Monte Carlo; local asymptotic normality; big data; scalability

1. Introduction

We firstly give the background and motivation in Section 1.1 and then state our contributions in Section 1.2, followed with the organization of the paper in Section 1.3.

1.1. Background and motivation

The classical theory of asymptotics in statistics relies heavily on certain local approximations to the logarithms of likelihood ratios, where "local" is meant to indicate that one looks at parameter values close to a point [12]. The classic theory of local asymptotic normality (LAN) of [11] concerns a "local neighborhood" $\{\theta + t_n n^{-1/2}\}$ around a true parameter θ in an open subset $\Theta \subset \mathbb{R}$, where t_n is a bounded constant and n is the number of observations. We suppose the data are modeled as a real-valued sample $(Y_1, \dots, Y_n)$, for $n \in \mathbb{N}$, from the probability distribution P_θ on the probability space $(\Omega, \mathcal{A}, \mu)$ where μ is a fixed σ-finite measure dominating P_θ. Let

$$p(\theta) = p(\cdot;\theta) = \frac{dP_\theta}{d\mu}(\cdot), \qquad l(\theta) = \log p(\theta), \tag{1.1}$$

be the density and log-likelihood of P_θ, respectively. Define the log-likelihood of $(Y_1, \dots, Y_n)$ by

$$\mathbf{l}(\theta) = \sum_{i=1}^{n} l(Y_i;\theta). \tag{1.2}$$

The LAN states as follows:

$$\mathbf{l}\big(\theta + t_n n^{-1/2}\big) - \mathbf{l}(\theta) = t_n S_n(\theta) - \frac{1}{2} t_n^2 \mathcal{I}(\theta) + o(1), \tag{1.3}$$

where $\mathcal{I}(\theta)$ is a finite positive constant, $S_n(\theta) \to N[0, \mathcal{I}(\theta)]$ in distribution, and $o(1)$ is an error term goes to zero in P_θ probability as n goes to infinity.

1350-7265

We form a grid of equally spaced points with separation $n^{-1/2}$ over $\mathbb{R}$ and define θ_n^* as the midpoint of the interval into which $\widetilde{\theta}_n$ has fallen, where $\widetilde{\theta}_n$ is a uniformly $\sqrt{n}$-consistent estimator whose existence is established in Theorem 1 on page 42 of [2]. Then θ_n^* is also uniformly $\sqrt{n}$-consistent. In practice, for $\theta_{j,n}^* = \theta_n^* + jn^{-1/2}$ where $j \in \{-1, 0, 1\}$, the quadratic polynomial of equation (1.3) can be interpolated by $(\theta_{j,n}^*, \mathbf{l}(\theta_{j,n}^*))_j$, and $S_n(\theta_n^*)$ and $\mathcal{I}(\theta_n^*)$ can be estimated. Here and in the sequel, we set $\theta_{0,n}^* = \theta_n^*$. The one-step estimator $\widehat{\theta}_n^A$ defined in [2], using the estimated $S_n(\theta_n^*)$ and $\mathcal{I}(\theta_n^*)$,

$$\widehat{\theta}_n^A = \theta_n^* + \sqrt{n} \times \frac{S_n(\theta_n^*)}{\mathcal{I}(\theta_n^*)}, \tag{1.4}$$

maximizes the interpolated quadratic approximation to the log-likelihood.

The one-step estimator $\widehat{\theta}_n^A$ can be generalized to $\widehat{\theta}_n^B$, using the estimated $S_n(\theta_n^*)$ and $\mathcal{I}(\theta_n^*)$ generated through a quadratic fit to $(\theta_{j,n}^*, \mathbf{l}(\theta_{j,n}^*))_{j\in\{-J,\dots,J\}}$ with $J \geq 1$. When the likelihood can be computed perfectly, there may be little reason to use $\widehat{\theta}_n^B$ over $\widehat{\theta}_n^A$. However, when there is Monte Carlo uncertainty or other numerical error in the likelihood evaluation, then $\widehat{\theta}_n^B$ with $J > 1$ may be preferred. Taking this idea a step further, we can construct a maximum smoothed likelihood estimator (MSLE), proposed in [8], by maximizing a smooth curve fitted to the grid of log-likelihood evaluations $(\theta_{j,n}^*, \mathbf{l}(\theta_{j,n}^*))_{j\in\{-J,\dots,J\}}$. If the smoothing algorithm preserved quadratic functions then the MSLE is asymptotically equivalent to a one-step estimator under the LAN property, while behaving reasonably when the log-likelihood has a substantial deviation from a quadratic.

The motivations of a rescaled LAN property arise from both the methodological side and the theoretical side:

- Monte Carlo likelihood evaluations calculated by simulating from a model, are useful for constructing likelihood-based parameter estimates and confidence intervals, for complex models [5]. However, for large datasets any reasonable level of computational effort may result in a non-negligible numerical error in these likelihood evaluations, and the Monte Carlo methods come at the expense of "poor scalability". Specifically, on the classical scale of $n^{-1/2}$, when the number of observations n is large, the statistical signal in the likelihood function is asymptotically overcome by Monte Carlo noise in the likelihood evaluations, as the Monte Carlo variance growing with n. In line with [9] (page 4), let the statistical error SE_{stat} stand for the uncertainty resulting from randomness in the data, viewed as a draw from a statistical model, and let the Monte Carlo error SE_{MC} stand for the uncertainty resulting from implementing a Monte Carlo estimator of a statistical model. That is, in the context of Monte Carlo inference, we desire a general methodology to suffice $\text{SE}_{\text{MC}}^2 \,/\, \text{SE}_{\text{stat}}^2 \to 0$.
- From a theoretical point of view, LAN has been found to hold in various situations other than regular independent identically distributed (i.i.d.) parametric models, including semiparametric models [3], positive-recurrent Markov chains [7], stationary hidden Markov models [3], stochastic block models [1], and regression models with long memory dependence [6]. There are very close linkages between LAN established in regular i.i.d. parametric models and LAN established in models such as hidden Markov models and stochastic processes, for example Theorem 1.1 in [3]. We anticipate that comparable results could be derived for RLAN. To motivate future investigations of contexts where RLAN arises, it is necessary to rigorously establish RLAN as a worthwhile statistical property.

An idea on rescaling is to consider the $n^{-1/4}$ local neighborhoods instead, since the existence of a uniformly $\sqrt{n}$-consistent estimator implies the existence of a uniformly $n^{1/4}$-consistent estimator and then in the $n^{-1/4}$ local neighborhoods we have θ_n^* uniformly $n^{1/4}$-consistent. Figure 1 verifies and

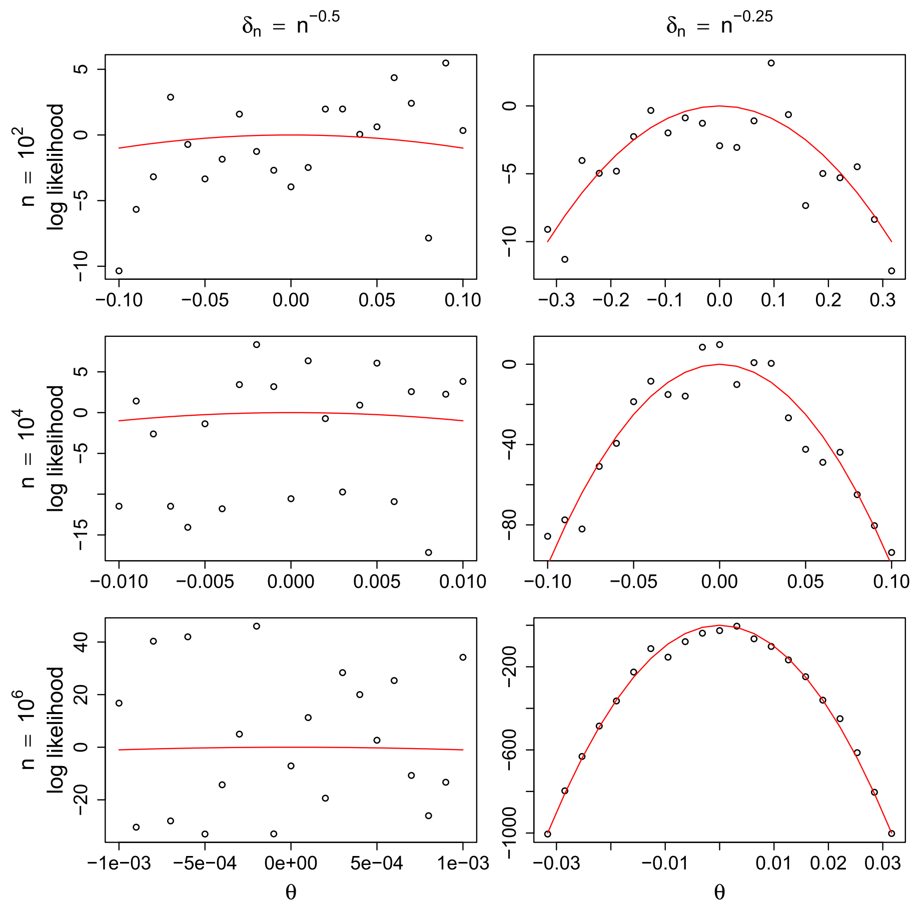

Figure 1. Illustration on the Monte Carlo estimation effects via different values of n and δ_n. The columns correspond to $\delta_n = n^{-0.5}, n^{-0.25}$ and the rows correspond to $n = 10^2, 10^4, 10^6$. The red solid curve is the log-likelihood given by $\mathbf{l}(\theta) = -n\theta^2$. The black circles are Monte Carlo log-likelihood evaluations, $\bar{\mathbf{l}}(\theta^*_{j,n}) \sim N[\mathbf{l}(\theta^*_{j,n}), n/m]$ with a sample size $m = n^{1/2}$, evaluated at 21 equally spaced values $\{\theta^*_{j,n}\}_j$ in the range $[-\delta_n, \delta_n]$.

illustrates the enlarged neighborhood idea, by showing a Gaussian likelihood function for n observations that is evaluated by a Monte Carlo estimator having variance scaling linearly with n/m, where $m = \sqrt{n}$ is the number of Monte Carlo simulations per observation. Monte Carlo evaluations of the log-likelihood were conducted at a grid of points $\{\theta^*_{j,n} = \theta^*_n + j\delta_n\}_{j\in\{-10,-9,\ldots,0,\ldots,9,10\}}$, evenly spaced on $[-\delta_n, \delta_n]$, for $\delta_n = n^{-1/2}$ and $\delta_n = n^{-1/4}$, respectively. We see from Figure 1 that on the classical scale of $n^{-1/2}$, when the number of observations n is large, the statistical signal in the likelihood function is asymptotically overcome by Monte Carlo noise in the likelihood evaluations, as the Monte Carlo variance growing with n even though there is just modest growth in the Monte Carlo effort ($m = n^{1/2}$). However, on the $n^{-1/4}$ scale, the form of the likelihood surface is evident despite the growing Monte Carlo uncertainty. That is, the classical local $n^{-1/2}$ neighborhood does not provide a useful estimate in this limit, but the rescaled local $n^{-1/4}$ neighborhood enables a quadratic likelihood approximation to be successfully fitted. Then, we aim to establish the LAN property in the rescaled local $n^{-1/4}$ neighborhood, based on which, an estimator extended the classical one-step estimator in [11] and the MSLE in [8] can be designed.

1.2. Our contributions

The contributions of the paper are two-fold:

(1) Our first contribution is the rescaled local asymptotic normality (RLAN) property defined as follows:

Definition 1.1. *Let* $\mathbf{P} := \{P_\theta : \theta \in \Theta\}$ *be a regular parametric model in the probability space* $(\Omega, \mathcal{A}, \mu)$. *We say that* $\mathbf{P}$ *has RLAN if uniformly in* $\theta \in K$ *compact* $\subset \Theta$ *and* $|t_n| \le M$,

$$\mathbf{l}\big(\theta + t_n n^{-1/4}\big) - \mathbf{l}(\theta) = n^{1/4} t_n S_n(\theta) - \frac{1}{2} n^{1/2} t_n^2 \mathcal{I}(\theta) + n^{1/4} t_n^3 \mathcal{W}(\theta) + \mathcal{O}(1) \tag{1.5}$$

and

$$\mathbf{l}\big(\theta + t_n n^{-1/2}\big) - \mathbf{l}(\theta) = t_n S_n(\theta) - \frac{1}{2} t_n^2 \mathcal{I}(\theta) + o(1), \tag{1.6}$$

under P_θ, *with* $\mathcal{I}(\theta)$ *being a finite positive constant,* $\mathcal{W}(\theta)$ *being a finite constant,* $S_n \to N[0, \mathcal{I}(\theta)]$ *in distribution,* $\mathcal{O}(1)$ *denoting an error term bounded in* P_θ *probability, and* $o(1)$ *denoting an error term converging to zero in* P_θ *probability.*

To develop the key ideas, we work in a one-dimensional parameter space. However, the ideas naturally generalize to $\Theta \subset \mathbb{R}^d$ for $d \ge 1$. The widely studied property of LAN [11,12] is defined by (1.6), so RLAN implies LAN. The LAN property asserts a quadratic approximation to the log-likelihood function in a neighborhood with scale $n^{-1/2}$, whereas RLAN additionally asserts a cubic approximation on a $n^{-1/4}$ scale. In Section 2, we present sufficient conditions for a sequence of i.i.d. random variables to satisfy RLAN. Complex dependence structures fall outside the i.i.d. theory of Section 2, except in the situation where there is also replication. Panel time series analysis via mechanistic models is one situation where i.i.d. replication arises together with model complexity requiring Monte Carlo approaches [4,16,17].

(2) Our second contribution is the scalable Monte Carlo inference. Suppose RLAN (Definition 1.1) holds. Then, with $\theta^*_{j,n} = \theta^*_n + j n^{-1/4}$ for j in a finite set $\mathcal{J}$, we can write

$$\mathbf{l}\big(\theta^*_{j,n}\big) = \beta_0 + \beta_1\big(j n^{-1/4}\big) + \beta_2\big(j n^{-1/4}\big)^2 + \beta_3\big(j n^{-1/4}\big)^3 + \epsilon_{j,n}, \tag{1.7}$$

where $\beta_1 = \mathcal{O}(n^{1/2})$, $\beta_2 = \mathcal{O}(n)$ and $\beta_3 = \mathcal{O}(n)$, and $\epsilon_{j,n} = \mathcal{O}(1)$. In practice, the cubic polynomial of equation (1.7) can be interpolated by $(\theta^*_{j,n}, \mathbf{l}(\theta^*_{j,n}))_{j \in \mathcal{J}}$, and $\{\beta_\iota\}_{\iota \in \{1,2,3\}}$ can be estimated. Based on the linear least squares estimated $\{\beta_\iota\}_{\iota \in \{1,2,3\}}$, we can define the maximum cubic log-likelihood estimator (MCLE) when it is finite, as

$$\widehat{\theta}_n^{\text{MCLE}} = \theta^*_n + n^{-1/4} \operatorname*{arg\,max}_{\chi \in \mathbb{R}} \big\{\beta_1\big(\chi n^{-1/4}\big) + \beta_2\big(\chi n^{-1/4}\big)^2 + \beta_3\big(\chi n^{-1/4}\big)^3\big\}. \tag{1.8}$$

The MCLE defined above is general, while in this paper we apply it in the context of Monte Carlo inference with i.i.d. data samples, under the situation that one does not have access to the likelihood evaluation $\mathbf{l}(\theta^*_{j,n})$ but instead can obtain the Monte Carlo likelihood evaluation $\bar{\mathbf{l}}(\theta^*_{j,n})$.

We firstly illustrate how $\bar{\mathbf{l}}(\theta^*_{j,n})$ may be generated. We suppose the data are modeled as an i.i.d. sequence $Y_1, \dots, Y_n$ drawn from a density

$$p(y;\theta) = p_Y(y;\theta) = \int p_{Y|X}(y|x;\theta)p_X(x;\theta)\,dx. \tag{1.9}$$

For each Y_i and $\theta^*_{j,n}$, independent Monte Carlo samples $(X^{(1)}_{i,j}, \dots, X^{(m)}_{i,j})$ for $m \in \mathbb{N}$ are generated from an appropriate probability density function $q(\cdot;\theta^*_{j,n})$. Then, we approximate $p(Y_i;\theta^*_{j,n})$ with $\overline{p}(Y_i;\theta^*_{j,n})$ using an importance sampling evaluator,

$$\overline{p}(Y_i;\theta^*_{j,n}) = \frac{1}{m}\sum_{\tau=1}^{m} p_{Y|X}\big(Y_i|X^{(\tau)}_{i,j};\theta^*_{j,n}\big)\frac{p_X(X^{(\tau)}_{i,j};\theta^*_{j,n})}{q(X^{(\tau)}_{i,j};\theta^*_{j,n})},$$

which is unbiased by construction. We construct

$$\bar{\mathbf{l}}(\theta^*_{j,n}) = \sum_{i=1}^{n} \ln \overline{p}\big(Y_i;\theta^*_{j,n}\big) \tag{1.10}$$

as the estimated log-likelihood.

Recalling that n is the number of observations, suppose that $m = m(n)$ is the number of Monte Carlo simulations per observation, and take $\mathcal{O}(\sqrt{n}) \ll m(n) \ll \mathcal{O}(n)$. Then the Monte Carlo log-likelihood theory gives that

$$\bar{\mathbf{l}}(\theta^*_{j,n}) = \mathbf{l}(\theta^*_{j,n}) + \gamma(\theta^*_{j,n}) + \widetilde{\epsilon}_{j,n}, \tag{1.11}$$

where $\widetilde{\epsilon}_{j,n}$ is i.i.d. such that $\frac{m}{n}\widetilde{\epsilon}_{j,n}$ convergences in distribution to a normal distribution with mean zero and positive finite variance, and the bias term $\gamma(\theta^*_{j,n})$ satisfies

$$\gamma(\theta^*_{j,n}) = \gamma(\theta^*_n) + C_\gamma \frac{n}{m} jn^{-1/4}\big(1 + o(1)\big), \tag{1.12}$$

where C_γ is a finite constant. Plugging equation (1.7) in equation (1.11), we can obtain that

$$\bar{\mathbf{l}}(\theta^*_{j,n}) = \beta_0 + \beta_1\big(jn^{-1/4}\big) + \beta_2\big(jn^{-1/4}\big)^2 + \beta_3\big(jn^{-1/4}\big)^3 + \epsilon_{j,n} + \gamma(\theta^*_{j,n}) + \widetilde{\epsilon}_{j,n}.$$

Organizing the terms in the above equation, we have the Monte Carlo meta model

$$\bar{\mathbf{l}}(\theta^*_{j,n}) = \overline{\beta}_0 + \overline{\beta}_1\big(jn^{-1/4}\big) + \overline{\beta}_2\big(jn^{-1/4}\big)^2 + \overline{\beta}_3\big(jn^{-1/4}\big)^3 + \overline{\epsilon}_{j,n}, \tag{1.13}$$

where $\theta^*_{j,n} = \theta^*_n + jn^{-1/4}$ for j in a finite set $\mathcal{J}$, $\overline{\beta}_1 = \mathcal{O}(n^{1/2})$, $\overline{\beta}_2 = \mathcal{O}(n)$, $\overline{\beta}_3 = \mathcal{O}(n)$, and $\overline{\epsilon}_{j,n}$ is i.i.d. such that $\frac{m}{n}\overline{\epsilon}_{j,n}$ convergences in distribution to a normal distribution having mean zero and positive finite variance.

The proposed general methodology takes advantage of asymptotic properties of the likelihood function in an $n^{-1/4}$ neighborhood of the true parameter value. In Section 4, we will see that $\widehat{\theta}^{\mathrm{MCLE}}_n$ is efficient with the desired property that $\mathrm{SE}^2_{\mathrm{MC}} / \mathrm{SE}^2_{\mathrm{stat}} \to 0$ as the number of data samples $n \to \infty$. The statistically efficient simulation-based likelihood inference is achieved with a computational budget of size essentially $n^{3/2}$ for a dataset of n observations and $\sqrt{n}$ Monte Carlo simulations per observation. In sum, despite substantial Monte Carlo uncertainties

Table 1. Table of Notation

$s(\theta) = s(\cdot;\theta) := \sqrt{p(\cdot;\theta)}$	Square root of density $p(\cdot;\theta)$, Eqn. (2.1).
$\mathbf{l}(\theta) = \sum_{i=1}^{n} l(Y_i;\theta)$	Log-likelihood of samples $(Y_1,\ldots,Y_n)$, Eqn. (1.2).
$\dot{\mathbf{l}}(\theta) = 2\frac{\dot{s}(\theta)}{s(\theta)}\mathbb{1}_{\{s(\theta)>0\}}$	Variable defined in Definition 2.1.
$\ddot{\mathbf{l}}(\theta) = 2\frac{\ddot{s}(\theta)}{s(\theta)}\mathbb{1}_{\{s(\theta)>0\}}$	Variable defined in Definition 2.1.
$\dddot{\mathbf{l}}(\theta) = 2\frac{\dddot{s}(\theta)}{s(\theta)}\mathbb{1}_{\{s(\theta)>0\}}$	Variable defined in Assumption 2.2.
$\ddddot{\mathbf{l}}(\theta) = 2\frac{\ddddot{s}(\theta)}{s(\theta)}\mathbb{1}_{\{s(\theta)>0\}}$	Variable defined in Assumption 2.2.
$\mathcal{I}(\theta) = E_\theta[\dot{\mathbf{l}}(\theta)]^2$	The second moment of $\dot{\mathbf{l}}(\theta)$, Definition 2.1.
$T_n = \{\frac{s(\theta+\delta_n t)}{s(\theta)} - 1 - \frac{\delta_n \dot{s}(\theta)t}{s(\theta)}\}\mathbb{1}_{\{s(\theta)>0\}}$	Key variable in this paper, Eqn. (3.1).
$A_n = \{\max_{1\le i\le n} \lvert T_{ni} + \frac{1}{2}\delta_n t_n \dot{\mathbf{l}}(\theta)\rvert < \eta\}$	Truncated variable, Eqn. (3.2).
$\widetilde{A} = \widetilde{A}_n(\epsilon) = \{\max_{1\le i\le n} \lvert T_{ni}\rvert < \epsilon\}$	Truncated variable, Eqn. (3.11).
$S_n(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \dot{\mathbf{l}}(Y_i;\theta)$	Variable defined in Eqn. (2.2).
$V_n(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} [\ddot{\mathbf{l}}(Y_i;\theta) - E_\theta[\ddot{\mathbf{l}}(\theta)]]$	Variable defined in Eqn. (2.3).
$U_n(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} [\dot{\mathbf{l}}^2(Y_i;\theta) - \mathcal{I}(\theta)]$	Variable defined in Eqn. (2.4).
$\widehat{\theta}_n^A = \theta_n^* + \sqrt{n} \times \frac{S_n(\theta_n^*)}{\mathcal{I}(\theta_n^*)}$	Estimator defined in Eqn. (1.4).
$\widehat{\theta}_n^C = \theta_n^* + \delta_n \times \frac{\sqrt{n}\delta_n S_n(\theta_n^*)}{n\delta_n^2 \mathcal{I}(\theta_n^*)}$	Estimator defined in Eqn. (4.3).
$\widehat{\theta}_n^{\text{MCLE}}$	Estimator defined in Eqn. (1.8).

involved in the proposed general Monte Carlo based method, MCLE is efficient and able to scale properly. The proposed methodology sheds light on tackling “poor scalability” issues in related Monte Carlo based approaches, such as the Monte Carlo adjusted profile methodology of [9] which has been used in various scientific studies [15–17,19]. However, the extension to profile likelihood estimation, analogous to the LAN-based profile likelihood theory of [13], is beyond the scope of this paper.

1.3. Organization of the paper

The rest of the paper proceeds as follows: Section 2 derives the RLAN property in the context of a regular parametric model, leading to a theorem which is proved in Section 3; In Section 4, based on the RLAN property, we investigate the performance of the proposed MCLE in the context of Monte Carlo inference. The notations used throughout this paper are listed in Table 1.

2. RLAN for regular parametric models

In this section, we show that parametric models with sufficient regularity enjoy the RLAN property, for n i.i.d. observations.

2.1. Model setup

We suppose the data are modeled as a real-valued i.i.d. sample $(Y_1,\ldots,Y_n)$, for $n \in \mathbb{N}$, from the probability distribution P_θ on the probability space $(\Omega, \mathcal{A}, \mu)$ where μ is a fixed σ-finite measure dominating

P_θ. We seek to infer the unknown "true" parameter θ which is situated in an open subset $\Theta \subset \mathbb{R}$. We suppose the parameterization $\theta \to P_\theta$ has a density and log-likelihood which can be written as

$$p(\theta) = p(\cdot;\theta) = \frac{dP_\theta}{d\mu}(\cdot), \qquad l(\theta) = \log p(\theta).$$

For $\mathbf{P} = \{P_\theta : \theta \in \Theta\}$ being the set of all the probability measures induced by the parameter θ in the whole parameter set Θ, we metrize $\mathbf{P}$ with the variational distance. Let $v : \mathbf{P} \to \mathbb{R}$ be a Euclidean parameter, and suppose that v can be identified with the parametric function $q : \Theta \to \mathbb{R}$ defined by

$$q(\theta) = v(P_\theta).$$

Let $\|\cdot\|$ stand for the Hilbert norm in $L_2(\mu)$, i.e., $\|f\|^2 = \int f^2\, d\mu$. It is convenient to view $\mathbf{P}$ as a subset of $L_2(\mu)$ via the embedding

$$p(\cdot;\theta) \to s(\cdot;\theta) := \sqrt{p(\cdot;\theta)}. \tag{2.1}$$

The generalization from the classical theory of LAN in the scale $\mathcal{O}(n^{-1/2})$ to RLAN in the scale $\mathcal{O}(n^{-1/4})$, requires additional smoothness assumptions. We start with the following definition.

Definition 2.1. *We say that θ_0 is a fourth-order regular point of the parametrization $\theta \to P_\theta$, if θ_0 is an interior point of Θ, and*

(1) *The map $\theta \to s(\theta)$ from Θ to $\mathcal{L}_2(\mu)$ is fourth-order differentiable at θ_0: there exist first-order derivative $\dot{s}(\theta_0)$, second-order derivative $\ddot{s}(\theta_0)$, third-order derivative $\dddot{s}(\theta_0)$, and fourth-order derivative $\ddddot{s}(\theta_0)$ of elements of $\mathcal{L}_2(\mu)$ such that*

$$\left\| \frac{s(\theta_0 + \delta_n t_n) - s(\theta_0) - \dot{s}(\theta_0)\delta_n t_n}{\delta_n^2} - \frac{1}{2}\ddot{s}(\theta_0)t_n^2 \right\| \to 0,$$

$$\left\| \frac{s(\theta_0 + \delta_n t_n) - s(\theta_0) - \dot{s}(\theta_0)\delta_n t_n - \frac{1}{2}\ddot{s}(\theta_0)\delta_n^2 t_n^2}{\delta_n^3} - \frac{1}{3!}\dddot{s}(\theta_0)t_n^3 \right\| \to 0,$$

$$\left\| \frac{s(\theta_0 + \delta_n t_n) - s(\theta_0) - \dot{s}(\theta_0)\delta_n t_n - \frac{1}{2}\ddot{s}(\theta_0)\delta_n^2 t_n^2 - \frac{1}{3!}\dddot{s}(\theta_0)\delta_n^3 t_n^3}{\delta_n^4} - \frac{1}{4!}\ddddot{s}(\theta_0)t_n^4 \right\| \to 0,$$

for any $\delta_n \to 0$ and t_n bounded.

(2) *The variable $\dot{\mathbf{I}}(\theta) := 2\frac{\dot{s}(\theta)}{s(\theta)}\mathbb{1}_{\{s(\theta)>0\}}$ has non-zero second moment $\mathcal{I}(\theta) := E_\theta[\dot{\mathbf{I}}(\theta)]^2$ and non-zero fourth moment.*

(3) *The variable $\ddot{\mathbf{I}}(\theta) := 2\frac{\ddot{s}(\theta)}{s(\theta)}\mathbb{1}_{\{s(\theta)>0\}}$ has non-zero second moment.*

Assumption 2.2. *We assume the following*:

(1) *Every point of Θ is a fourth-order regular point.*
(2) *The map $\theta \to \ddddot{s}(\theta)$ is continuous from Θ to $\mathcal{L}_2(\mu)$.*
(3) *Define $\dddot{\mathbf{I}}(\theta) := 2\frac{\dddot{s}(\theta)}{s(\theta)}\mathbb{1}_{\{s(\theta)>0\}}$ and $\ddddot{\mathbf{I}}(\theta) := 2\frac{\ddddot{s}(\theta)}{s(\theta)}\mathbb{1}_{\{s(\theta)>0\}}$. We have*

$$E_\theta\big|\dot{\mathbf{I}}(\theta)\big|^6 < \infty, \qquad E_\theta\big|\ddot{\mathbf{I}}(\theta)\big|^3 < \infty, \qquad E_\theta\big|\dddot{\mathbf{I}}(\theta)\big|^2 < \infty, \qquad E_\theta\big|\ddddot{\mathbf{I}}(\theta)\big| < \infty.$$

Remark 2.1. *We have the following comments regarding Assumption 2.2*:

(1) *The conditions* $E_\theta[\dot{\mathbf{l}}(\theta)]^2 \neq 0$, $E_\theta[\dot{\mathbf{l}}(\theta)]^4 \neq 0$ *and* $E_\theta[\ddot{\mathbf{l}}(\theta)]^2 \neq 0$ *hold unless for random variables that are zero almost sure.*
(2) *The condition* (3) *in Assumption* 2.2 *holds for all bounded random variables, such as truncated normal distributed random variables and finite discrete distributed random variables.*
(3) *The condition* (3) *in Assumption* 2.2 *holds for some unbounded random variables at least, such as the centered normal distributed random variable with* $\theta \in \{1, \sqrt{2}, \sqrt{3}\}$, *whose probability density function is given by* $p(y;\theta) = \frac{1}{\sqrt{2\pi\theta^2}} e^{-\frac{y^2}{2\theta^2}}$.
(4) *In this section, we suppose the data are modeled as a real-valued i.i.d. sample* $(Y_1, \dots, Y_n)$. *Assumption* 2.2 *still may apply on stochastic process with very desired conditions, such as the basic Ornstein–Uhlenbeck process* ([20]) *that is stationary, Gaussian, and Markovian, evolving as* $dX_t = -\rho X_t \, dt + \sigma \, dW_t$, *where* ρ *and* σ *are finite constants, and* W_t *is the standard Brownian motion with unit variance parameter. Its stationary distribution is the normal distribution with mean* 0 *and variance* $\theta^2 = \frac{\sigma^2}{2\rho}$.

2.2. The main result

Define

$$S_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{\mathbf{l}}(Y_i;\theta), \tag{2.2}$$

$$V_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big[\ddot{\mathbf{l}}(Y_i;\theta) - E_\theta \ddot{\mathbf{l}}(\theta)\big], \tag{2.3}$$

$$U_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big[\dot{\mathbf{l}}^2(Y_i;\theta) - \mathcal{I}(\theta)\big]. \tag{2.4}$$

We have the following theorem for RLAN, whose rigorous proof is provided in Section 3.3.

Theorem 2.3. *Suppose that* $\mathbf{P} = \{P_\theta : \theta \in \Theta\}$ *is a regular parametric model satisfying Assumption* 2.2. *When* $\delta_n = \mathcal{O}(n^{-1/4})$, *write*

$$\begin{aligned}
&\mathbf{l}(\theta + \delta_n t_n) - \mathbf{l}(\theta) \\
&\quad = t_n\big\{\sqrt{n}\delta_n S_n(\theta)\big\} + t_n^2\left\{\sqrt{n}\delta_n^2\left[\frac{1}{2}V_n(\theta) - \frac{1}{4}U_n(\theta)\right] - \frac{1}{2}n\delta_n^2\mathcal{I}(\theta)\right\} \\
&\qquad + t_n^3\left\{n\delta_n^3\left[\frac{1}{12}E_\theta\big[\dot{\mathbf{l}}^3(\theta)\big] - \frac{1}{8}E_\theta\big[\ddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)\big] + \frac{1}{6}E_\theta\big[\dddot{\mathbf{l}}(\theta)\big]\right]\right\} \\
&\qquad + t_n^4\left\{n\delta_n^4\left[-\frac{1}{32}E_\theta\big[\dot{\mathbf{l}}^4(\theta)\big] - \frac{1}{16}E_\theta\big[\ddot{\mathbf{l}}(\theta)\big]^2 - \frac{1}{12}E_\theta\big[\dddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)\big] + \frac{1}{24}E_\theta\big[\ddddot{\mathbf{l}}(\theta)\big]\right]\right\} \\
&\qquad + R_n(\theta, t_n).
\end{aligned}$$

Then uniformly in $\theta \in K$ *compact* $\subset \Theta$ *and* $|t_n| \le M$, *one has* $R_n(\theta, t_n) \xrightarrow{p} 0$ *in* P_θ *probability, and in the weak topology*

$$\mathbf{L}_\theta\big(S_n(\theta)\big) \to N\big(0, \mathcal{I}(\theta)\big),$$

$$\mathbf{L}_\theta\big(V_n(\theta)\big) \to N\big(0, \operatorname{Var}_\theta\big[\ddot{\mathbf{l}}(\theta)\big]\big),$$
$$\mathbf{L}_\theta\big(U_n(\theta)\big) \to N\big(0, \operatorname{Var}_\theta\big[\dot{\mathbf{l}}^2(\theta)\big]\big),$$

where $N(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2, and $\mathbf{L}_\theta$ is the law under θ.

Remark 2.2. *The equation in Theorem 2.3 with $\delta_n = \mathcal{O}(n^{-1/2})$ instead of $\delta_n = \mathcal{O}(n^{-1/4})$ implies the classical LAN result (Proposition 2 on page 16 of [2]), which can be seen as follows:*

(1) *For the t_n term,*

$$\big\{\sqrt{n}\delta_n S_n(\theta)\big\} = \frac{2}{\sqrt{n}} \sum_{i=1}^n \frac{\dot{s}(Y_i;\theta)}{s(Y_i;\theta)} \mathbb{1}_{\{s(Y_i;\theta)>0\}}.$$

(2) *For the t_n^2 term,*

$$\left\{\sqrt{n}\delta_n^2\left[\frac{1}{2}V_n(\theta) - \frac{1}{4}U_n(\theta)\right] - \frac{1}{2}n\delta_n^2\mathcal{I}(\theta)\right\}$$
$$= \frac{1}{2n}\sum_{i=1}^n\big[\ddot{\mathbf{l}}(Y_i;\theta) - E_\theta\ddot{\mathbf{l}}(\theta)\big] - \frac{1}{4n}\sum_{i=1}^n\big[\dot{\mathbf{l}}^2(Y_i;\theta) - \mathcal{I}(\theta)\big] - \frac{1}{2}\mathcal{I}(\theta).$$

By Chung's uniform strong law of large number and Lemma 3.2, one can obtain that

$$\frac{1}{n}\sum_{i=1}^n\big[\ddot{\mathbf{l}}(Y_i;\theta) - E_\theta\big[\ddot{\mathbf{l}}(\theta)\big]\big] \xrightarrow{a.s.} 0$$

and

$$\frac{1}{n}\sum_{i=1}^n\big[\dot{\mathbf{l}}^2(Y_i;\theta) - \mathcal{I}(\theta)\big] \xrightarrow{a.s.} 0,$$

uniformly in $\theta \in K$ compact $\subset \Theta$ and $|t_n| \le M$. Then the t_n^2 coefficient is asymptotically equivalent to

$$-\frac{1}{2}\mathcal{I}(\theta) = -2E_\theta\left[\frac{\dot{s}(Y_i;\theta)}{s(Y_i;\theta)}\mathbb{1}_{\{s(Y_i;\theta)>0\}}\right]^2.$$

(3) *For the t_n^3 term, since $E_\theta[\dot{\mathbf{l}}^3(\theta)]$, $E_\theta[\ddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)]$ and $E_\theta[\dddot{\mathbf{l}}(\theta)]$ are finite constants, and $n\delta_n^3 \to 0$ as $n \to \infty$,*

$$\left\{n\delta_n^3\left[\frac{1}{12}E_\theta\big[\dot{\mathbf{l}}^3(\theta)\big] - \frac{1}{8}E_\theta\big[\ddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)\big] + \frac{1}{6}E_\theta\big[\dddot{\mathbf{l}}(\theta)\big]\right]\right\} \to 0.$$

(4) *Similarly, for the t_n^4 term, since $E_\theta[\dot{\mathbf{l}}^4(\theta)]$, $E_\theta[\ddot{\mathbf{l}}(\theta)]^2$, $E_\theta[\ddddot{\mathbf{l}}(\theta)]$ and $E_\theta[\dddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)]$ are finite constants, and $n\delta_n^4 \to 0$ as $n \to \infty$,*

$$\left\{n\delta_n^4\left[-\frac{1}{32}E_\theta\big[\dot{\mathbf{l}}^4(\theta)\big] - \frac{1}{16}E_\theta\big[\ddot{\mathbf{l}}(\theta)\big]^2 - \frac{1}{12}E_\theta\big[\dddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)\big] + \frac{1}{24}E_\theta\big[\ddddot{\mathbf{l}}(\theta)\big]\right]\right\} \to 0.$$

Remark 2.3. *Theorem 2.3 implies the RLAN property in Definition 1.1. The case $\delta_n = n^{-1/2}$ is already covered in Remark 2.2. When $\delta_n = n^{-1/4}$, the terms*

$$t_n^2\left\{\sqrt{n}\delta_n^2\left[\frac{1}{2}V_n(\theta) - \frac{1}{4}U_n(\theta)\right]\right\} = \mathcal{O}(1)$$

and

$$t_n^4\left\{n\delta_n^4\left[-\frac{1}{32}E_\theta\left[\dot{\mathbf{l}}^4(\theta)\right] - \frac{1}{16}E_\theta\left[\ddot{\mathbf{l}}(\theta)\right]^2 - \frac{1}{12}E_\theta\left[\dddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)\right] + \frac{1}{24}E_\theta\left[\ddddot{\mathbf{l}}(\theta)\right]\right]\right\} = \mathcal{O}(1).$$

Hence, we have, uniformly in $\theta \in K$ compact $\subset \Theta$ and $|t_n| \le M$,

$$\mathbf{l}(\theta + \delta_n t_n) - \mathbf{l}(\theta) = \sqrt{n}\delta_n t_n S_n - \frac{1}{2}n\delta_n^2 t_n^2 \mathcal{I} + n\delta_n^3 t_n^3 \mathcal{W} + \mathcal{O}(1), \tag{2.5}$$

where $\mathcal{I} = \mathcal{I}(\theta)$ is a finite positive constant, $S_n \to N[0, \mathcal{I}]$ in distribution, and

$$\mathcal{W} = \left[\frac{1}{12}E_\theta\left[\dot{\mathbf{l}}^3(\theta)\right] - \frac{1}{8}E_\theta\left[\ddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)\right] + \frac{1}{6}E_\theta\left[\dddot{\mathbf{l}}(\theta)\right]\right]$$

is a finite constant.

3. Developing a proof of Theorem 2.3

In this section, we work toward a proof of Theorem 2.3. Throughout this section, we suppose Assumption 2.2 holds and consider $\delta_n = \mathcal{O}(n^{-1/4})$. We first use a truncation method on a Taylor series expansion of $\mathbf{l}(\theta + \delta_n t_n) - \mathbf{l}(\theta)$ in Section 3.1, and then conduct preliminary analysis in bounding $\mathbf{l}(\theta + \delta_n t_n) - \mathbf{l}(\theta)$ in Section 3.2, both of which prepare for the proof of Theorem 2.3 in Section 3.3.

3.1. A truncatated Taylor series remainder

Set

$$T_n = \left\{\frac{s(\theta + \delta_n t_n)}{s(\theta)} - 1 - \frac{1}{2}\delta_n t_n \dot{\mathbf{l}}(\theta)\right\}\mathbb{1}_{\{s(\theta)>0\}}. \tag{3.1}$$

Let $\{T_{ni}\}_{i=1,\dots,n}$ denote the n i.i.d. copies of T_n corresponding to $Y_1, \dots, Y_n$, and for $\eta \in (0, 1)$ define

$$A_n = \left\{\max_{1\le i\le n}\left|T_{ni} + \frac{1}{2}\delta_n t_n \dot{\mathbf{l}}(\theta)\right| < \eta\right\}. \tag{3.2}$$

In the following, we use a truncation method similar to [2], but our definition of T_n differs from that in [2] (page 509). Compared to the corresponding one in [2], here T_n additionally incorporates the first-order derivative of $s(\theta)$, since we have to resort to a higher order derivative of $s(\theta)$ for analysis on the $\delta_n = \mathcal{O}(n^{-1/4})$ scale. By Proposition 3.4 following, uniformly in $\theta \in K \subset \Theta$ for K compact and $|t_n| \le M$, $P_\theta(A_n^c) \to 0$ where A_n^c is the complement of A_n. On the event A_n, we have

$$\mathbf{l}(\theta + \delta_n t_n) - \mathbf{l}(\theta) = \sum_{i=1}^n \log\left\{\frac{p(Y_i; \theta + \delta_n t_n)}{p(Y_i; \theta)}\right\} = 2\sum_{i=1}^n \log\left(T_{ni} + 1 + \frac{1}{2}\delta_n t_n \dot{\mathbf{l}}(Y_i; \theta)\right).$$

By a Taylor expansion,

$$
\begin{aligned}
&\mathbf{l}(\theta+\delta_n t_n)-\mathbf{l}(\theta)\\
&\quad =2\sum_{i=1}^n\left(T_{ni}+\frac{1}{2}\delta_n t_n\dot{\mathbf{l}}(Y_i;\theta)\right)-\sum_{i=1}^n\left(T_{ni}+\frac{1}{2}\delta_n t_n\dot{\mathbf{l}}(Y_i;\theta)\right)^2\\
&\qquad +\frac{2}{3}\sum_{i=1}^n\left(T_{ni}+\frac{1}{2}\delta_n t_n\dot{\mathbf{l}}(Y_i;\theta)\right)^3-\frac{1}{2}\sum_{i=1}^n\left(T_{ni}+\frac{1}{2}\delta_n t_n\dot{\mathbf{l}}(Y_i;\theta)\right)^4+R_n,
\end{aligned}
\tag{3.3}
$$

where

$$
|R_n|\le\frac{2C(\eta)}{5}\sum_{i=1}^n\left(T_{ni}+\frac{1}{2}\delta_n t_n\dot{\mathbf{l}}(Y_i;\theta)\right)^5,
$$

for $C(\eta)<(1-\eta)^{-5}$ a finite constant that depends on η only.

3.2. Preliminary analysis

In this subsection, we develop a sequence of lemmas and propositions needed for the proof in Section 3.3. Specifically, we conduct a series of preliminary analyses to bound the quantities $(T_{ni})^\alpha(\delta_n t_n\dot{\mathbf{l}}(Y_i;\theta))^\beta$ for $\alpha,\beta\in\{0,1,2,3,4,5\}$ such that $\alpha+\beta=5$.

Lemma 3.1. *One has*

$$
E_\theta\left|T_n-\frac{1}{4}(\delta_n t_n)^2\ddot{\mathbf{l}}(\theta)\right|^2=o\big(\delta_n^4\big),
\tag{3.4}
$$

$$
E_\theta\left|T_n-\frac{1}{4}(\delta_n t_n)^2\ddot{\mathbf{l}}(\theta)-\frac{1}{12}(\delta_n t_n)^3\dddot{\mathbf{l}}(\theta)\right|^2=o\big(\delta_n^6\big),
\tag{3.5}
$$

$$
E_\theta\left|T_n-\frac{1}{4}(\delta_n t_n)^2\ddot{\mathbf{l}}(\theta)-\frac{1}{12}(\delta_n t_n)^3\dddot{\mathbf{l}}(\theta)-\frac{1}{48}(\delta_n t_n)^4\ddddot{\mathbf{l}}(\theta)\right|^2=o\big(\delta_n^8\big),
\tag{3.6}
$$

as $\delta_n\to 0$, uniformly in $\theta\in K\subset\Theta$ for K compact and $|t_n|\le M$,

Proof. See the supplementary material [14]. □

Lemma 3.2. *One has*

$$
\lim_{\lambda\to\infty}\sup_{\theta\in K}E_\theta\big[\big|\ddot{\mathbf{l}}(\theta)\big|^2\mathbb{1}_{\{|\ddot{\mathbf{l}}(\theta)|\ge\lambda\}}\big]=0,
\tag{3.7}
$$

$$
\lim_{\lambda\to\infty}\sup_{\theta\in K}E_\theta\big[\big|\dot{\mathbf{l}}(\theta)\big|^4\mathbb{1}_{\{|\dot{\mathbf{l}}(\theta)|\ge\lambda\}}\big]=0,
\tag{3.8}
$$

$$
\lim_{\lambda\to\infty}\sup_{\theta\in K}E_\theta\big[\big|\ddddot{\mathbf{l}}(\theta)\big|\mathbb{1}_{\{|\ddddot{\mathbf{l}}(\theta)|\ge\lambda\}}\big]=0,
\tag{3.9}
$$

$$
\lim_{\lambda\to\infty}\sup_{\theta\in K}E_\theta\big[\big|\dddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)\big|\mathbb{1}_{\{|\dddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)|\ge\lambda\}}\big]=0.
\tag{3.10}
$$

Proof. See the supplementary material [14]. □

Recall that $\{T_{ni}\}_{i=1,\dots,n}$ denote the n i.i.d. copies of T_n corresponding to $Y_1, \dots, Y_n$. Define

$$\widetilde{A}_n(\epsilon) = \Big\{ \max_{1 \le i \le n} |T_{ni}| < \epsilon \Big\}, \tag{3.11}$$

for every $\epsilon > 0$.

Proposition 3.3. *Uniformly in $\theta \in K \subset \Theta$ for K compact and $|t_n| \le M$,*

$$P_\theta\big(\widetilde{A}_n^c\big) \to 0,$$

where $\widetilde{A}_n^c$ is the complement of $\widetilde{A}_n$.

Proof. We firstly note that

$$P_\theta\big(\widetilde{A}_n^c\big) \le \sum_{i=1}^n P_\theta\big(|T_{ni}| \ge \epsilon\big) = n P_\theta\big(|T_n| \ge \epsilon\big).$$

Then, it suffices to show that

$$P_\theta\big(|T_n| \ge \epsilon\big) = o(1/n).$$

But

$$\begin{aligned} P_\theta\big(|T_n| \ge \epsilon\big) &\le P_\theta\bigg(\bigg|T_n - \frac{1}{4}(\delta_n t_n)^2 \ddot{\mathbf{I}}(\theta)\bigg| \ge \frac{1}{2}\epsilon\bigg) + P_\theta\bigg(\bigg|\frac{1}{4}(\delta_n t_n)^2 \ddot{\mathbf{I}}(\theta)\bigg| \ge \frac{1}{2}\epsilon\bigg) \\ &\le \frac{4}{\epsilon^2} E_\theta \bigg|T_n - \frac{1}{4}(\delta_n t_n)^2 \ddot{\mathbf{I}}(\theta)\bigg|^2 + \frac{4}{\epsilon^2} E_\theta \bigg|\frac{1}{4}(\delta_n t_n)^2 \ddot{\mathbf{I}}(\theta)\bigg|^2 \mathbb{1}_{\{|(\delta_n t_n)^2 \ddot{\mathbf{I}}(\theta)| \ge \epsilon\}} \\ &\le o\big(\delta_n^4\big) + \frac{1}{4\epsilon^2}(\delta_n t_n)^4 E_\theta \big|\ddot{\mathbf{I}}(\theta)\big|^2 \mathbb{1}_{\{|(\delta_n t_n)^2 \ddot{\mathbf{I}}(\theta)| \ge \epsilon\}} \\ &= o\big(\delta_n^4\big), \end{aligned} \tag{3.12}$$

where the second to the last step is by Lemma 3.1, and the last step is by Lemma 3.2. □

Recall that for $\eta \in (0, 1)$

$$A_n = \bigg\{ \max_{1 \le i \le n} \bigg| T_{ni} + \frac{1}{2}\delta_n t_n \dot{\mathbf{I}}(\theta) \bigg| < \eta \bigg\}. \tag{3.13}$$

Proposition 3.4. *Uniformly in $\theta \in K \subset \Theta$ for K compact and $|t_n| \le M$,*

$$P_\theta\big(A_n^c\big) \to 0,$$

where A_n^c is the complement of A_n.

Proof. We first note that

$$P_\theta\big(A_n^c\big) \le \sum_{i=1}^n P_\theta\bigg(\bigg|T_{ni} + \frac{1}{2}\delta_n t_n \dot{\mathbf{I}}(\theta)\bigg| \ge \eta\bigg) = n P_\theta\bigg(\bigg|T_n + \frac{1}{2}\delta_n t_n \dot{\mathbf{I}}(\theta)\bigg| \ge \eta\bigg).$$

Then, it suffices to show that

$$P_\theta\left(\left|T_n + \frac{1}{2}\delta_n t_n \dot{\mathbf{I}}(\theta)\right| \geq \eta\right) = o(1/n).$$

But

$$\begin{aligned} P_\theta\left(\left|T_n + \frac{1}{2}\delta_n t_n \dot{\mathbf{I}}(\theta)\right| \geq \eta\right) &\leq P_\theta\left(|T_n| \geq \frac{1}{2}\eta\right) + P_\theta\left(\left|\frac{1}{2}\delta_n t_n \dot{\mathbf{I}}(\theta)\right| \geq \frac{1}{2}\eta\right) \\ &\leq o\left(\delta_n^4\right) + \frac{1}{\eta^2}(\delta_n t_n)^4 E_\theta\left|\dot{\mathbf{I}}(\theta)\right|^4 \mathbb{1}_{\{|\delta_n t_n \dot{\mathbf{I}}(\theta)| \geq \epsilon\}} \\ &= o\left(\delta_n^4\right), \end{aligned}$$

where the second to the last step is by equation (3.12) and the last step is by Lemma 3.2. □

Proposition 3.5. *For any $r \geq 0$, we have*

$$\sum_{i=1}^n \left|T_{ni}^2 - \frac{1}{16}\delta_n^4 t_n^4 E_\theta\left[\ddot{\mathbf{I}}(\theta)\right]^2\right|^{1+r} \xrightarrow{p} 0,$$

uniformly in $|t_n| \leq M$ and in $\theta \in K \subset \Theta$ for K compact, in P_θ probability.

Proof. See the supplementary material [14]. □

Proposition 3.6. *We have, for any $k \geq 1$,*

$$\sum_{i=1}^n |T_{ni}|^{2+k} \xrightarrow{p} 0,$$

uniformly in $|t_n| \leq M$ and in $\theta \in K \subset \Theta$ for K compact, in P_θ probability.

Proof. Let $a(\epsilon')$ be a real valued function on any $\epsilon' > 0$ satisfying $(\frac{\epsilon'}{a(\epsilon') + \frac{1}{16}n\delta_n^4 t_n^4 E_\theta[\ddot{\mathbf{I}}(\theta)]^2}) \in (0, 1)$. The proof can be completed by noting that by Propositions 3.3 and 3.5, uniformly in $|t_n| \leq M$ and in $\theta \in K \subset \Theta$ for K compact,

$$P_\theta\left(\sum_{i=1}^n T_{ni}^2 > a\left(\epsilon'\right) + \frac{1}{16}n\delta_n^4 t_n^4 E_\theta\left[\ddot{\mathbf{I}}(\theta)\right]^2\right) \to 0,$$

$$P_\theta\left(\max_{1 \leq i \leq n} |T_{ni}| > \left(\frac{\epsilon'}{a(\epsilon') + \frac{1}{16}n\delta_n^4 t_n^4 E_\theta[\ddot{\mathbf{I}}(\theta)]^2}\right)^{1/k}\right) \to 0.$$

□

Proposition 3.7. *We have, for $m \in \{5, 6\}$,*

$$\sum_{i=1}^n \left|\left(\delta_n t_n \dot{\mathbf{I}}(Y_i; \theta)\right)^m\right| \xrightarrow{p} 0,$$

uniformly in $|t_n| \leq M$ and in $\theta \in K \subset \Theta$ for K compact, in P_θ probability.

Proof. By Assumption 2.2 and Markov inequality, for $|t_n| \leq M$, $\epsilon' > 0$, and $m \in \{5, 6\}$,

$$P_\theta\left(\sum_{i=1}^{n}\left|\delta_n t_n \dot{\mathbf{I}}(Y_i;\theta)\right|^m > \epsilon'\right) \leq \frac{1}{\epsilon'} E_\theta\left(\sum_{i=1}^{n}\left|\delta_n t_n \dot{\mathbf{I}}(Y_i;\theta)\right|^m\right)$$
$$\leq \frac{n(\delta_n)^m M^m}{\epsilon'} E_\theta\left|\dot{\mathbf{I}}(Y_i;\theta)\right|^m$$
$$\to 0.$$

□

Proposition 3.8. *We have, for any $l \geq 2$ and any $k \geq 1$,*

$$\sum_{i=1}^{n}\left|T_{ni}^{l}\left(\delta_n t_n \dot{\mathbf{I}}(Y_i;\theta)\right)^k\right| \xrightarrow{p} 0,$$

uniformly in $|t_n| \leq M$ and in $\theta \in K \subset \Theta$ for K compact, in P_θ probability.

Proof. See the supplementary material [14]. □

Proposition 3.9. *We have, uniformly in $|t_n| \leq M$ and in $\theta \in K \subset \Theta$ for K compact,*

$$\sum_{i=1}^{n}\left(\delta_n t_n \dot{\mathbf{I}}(Y_i;\theta)\right)^4 - n\delta_n^4 t_n^4 E_\theta\left[\dot{\mathbf{I}}(\theta)\right]^4 \xrightarrow{a.s.} 0,$$
$$\sum_{i=1}^{n}(\delta_n t_n)^4 \ddddot{\mathbf{I}}(Y_i;\theta) - n\delta_n^4 t_n^4 E_\theta\left[\ddddot{\mathbf{I}}(\theta)\right] \xrightarrow{a.s.} 0,$$
$$\sum_{i=1}^{n}(\delta_n t_n)^4 \dddot{\mathbf{I}}(Y_i;\theta)\dot{\mathbf{I}}(Y_i;\theta) - n\delta_n^4 t_n^4 E_\theta\left[\dddot{\mathbf{I}}(\theta)\dot{\mathbf{I}}(\theta)\right] \xrightarrow{a.s.} 0,$$
$$\sum_{i=1}^{n}(\delta_n t_n)^4 \ddot{\mathbf{I}}(Y_i;\theta)\dot{\mathbf{I}}^2(Y_i;\theta) - n\delta_n^4 t_n^4 E_\theta\left[\ddot{\mathbf{I}}(\theta)\dot{\mathbf{I}}^2(\theta)\right] \xrightarrow{a.s.} 0.$$

Proof. We complete the proof by noting that, by Chung's uniform strong law of large number which can be seen in Theorem A.7.3 in [2] and Lemma 3.2, one has

$$\left(\frac{1}{n}\sum_{i=1}^{n}\left(\dot{\mathbf{I}}(Y_i;\theta)\right)^4 - E_\theta\left[\dot{\mathbf{I}}(\theta)\right]^4\right) \xrightarrow{a.s.} 0,$$
$$\left(\frac{1}{n}\sum_{i=1}^{n}\ddddot{\mathbf{I}}(Y_i;\theta) - E_\theta\left[\ddddot{\mathbf{I}}(\theta)\right]\right) \xrightarrow{a.s.} 0,$$
$$\left(\frac{1}{n}\sum_{i=1}^{n}\dddot{\mathbf{I}}(Y_i;\theta)\dot{\mathbf{I}}(Y_i;\theta) - E_\theta\left[\dddot{\mathbf{I}}(\theta)\dot{\mathbf{I}}(\theta)\right]\right) \xrightarrow{a.s.} 0.$$

Furthermore, noting that $\ddot{\mathbf{l}}(Y_i;\theta)\dot{\mathbf{l}}^2(Y_i;\theta) \le \frac{1}{2}\ddot{\mathbf{l}}^2(Y_i;\theta) + \frac{1}{2}\dot{\mathbf{l}}^4(Y_i;\theta)$, $E_\theta \ddot{\mathbf{l}}^2(Y_i;\theta) < \infty$ and $E_\theta \dot{\mathbf{l}}^4(Y_i;\theta) < \infty$ (Assumption 2.2), by Chung's uniform strong law of large number and Lemma 3.2, one has

$$\left(\frac{1}{n}\sum_{i=1}^n \ddot{\mathbf{l}}(Y_i;\theta)\dot{\mathbf{l}}^2(Y_i;\theta) - E_\theta\big[\ddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}^2(\theta)\big]\right) \xrightarrow{a.s.} 0. \qquad \square$$

Proposition 3.10. *We have*

$$\sum_{i=1}^n \left(T_{ni} - \frac{1}{4}\delta_n^2 t_n^2 \ddot{\mathbf{l}}(Y_i;\theta) - \frac{1}{12}(\delta_n t_n)^3 \dddot{\mathbf{l}}(Y_i;\theta) - \frac{1}{48}(\delta_n t_n)^4 \ddddot{\mathbf{l}}(Y_i;\theta)\right) \xrightarrow{p} 0,$$

uniformly in $|t_n| \le M$ and in $\theta \in K \subset \Theta$ for K compact, in P_θ probability.

Proof. For any $\epsilon' > 0$, by Markov inequality, Jensen's inequality, and Lemma 3.1,

$$\begin{aligned}
&P_\theta\left\{\left|\sum_{i=1}^n \left(T_{ni} - \frac{1}{4}\delta_n^2 t_n^2 \ddot{\mathbf{l}}(Y_i;\theta) - \frac{1}{12}(\delta_n t_n)^3 \dddot{\mathbf{l}}(Y_i;\theta) - \frac{1}{48}(\delta_n t_n)^4 \ddddot{\mathbf{l}}(Y_i;\theta)\right)\right| > \epsilon'\right\} \\
&\le \frac{n}{\epsilon'} E_\theta\left|T_n - \frac{1}{4}(\delta_n t_n)^2 \ddot{\mathbf{l}}(\theta) - \frac{1}{12}(\delta_n t_n)^3 \dddot{\mathbf{l}}(\theta) - \frac{1}{48}(\delta_n t_n)^4 \ddddot{\mathbf{l}}(\theta)\right| \\
&\to 0. \qquad \square
\end{aligned}$$

Proposition 3.11. *We have*

$$\sum_{i=1}^n \left|\left(T_{ni} - \frac{1}{4}\delta_n^2 t_n^2 \ddot{\mathbf{l}}(Y_i;\theta)\right)\big(\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta)\big)^2\right| \xrightarrow{p} 0,$$

uniformly in $|t_n| \le M$ and in $\theta \in K \subset \Theta$ for K compact, in P_θ probability.

Proof. For any $\epsilon' > 0$, by Markov inequality, Hölder's inequality, and Lemma 3.1,

$$\begin{aligned}
&P_\theta\left\{\sum_{i=1}^n \left|\left(T_{ni} - \frac{1}{4}\delta_n^2 t_n^2 \ddot{\mathbf{l}}(Y_i;\theta)\right)\big(\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta)\big)^2\right| > \epsilon'\right\} \\
&\le \frac{n}{\epsilon'} E_\theta\left|\left(T_n - \frac{1}{4}\delta_n^2 t_n^2 \ddot{\mathbf{l}}(\theta)\right)\big(\delta_n t_n \dot{\mathbf{l}}(\theta)\big)^2\right| \\
&\le \frac{n}{\epsilon'}\left[E_\theta\left(T_n - \frac{1}{4}\delta_n^2 t_n^2 \ddot{\mathbf{l}}(\theta)\right)^2\right]^{1/2}\left[E_\theta\big(\delta_n t_n \dot{\mathbf{l}}(\theta)\big)^4\right]^{1/2} \\
&\to 0. \qquad \square
\end{aligned}$$

Proposition 3.12. *We have*

$$\sum_{i=1}^n \left|\left(T_{ni} - \frac{1}{4}\delta_n^2 t_n^2 \ddot{\mathbf{l}}(Y_i;\theta) - \frac{1}{12}(\delta_n t_n)^3 \dddot{\mathbf{l}}(Y_i;\theta)\right)\big(\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta)\big)\right| \xrightarrow{p} 0,$$

uniformly in $|t_n| \le M$ and in $\theta \in K \subset \Theta$ for K compact, in P_θ probability.

Proof. For any $\epsilon' > 0$, by Markov inequality, Hölder's inequality, and Lemma 3.1,

$$\begin{aligned}
&P_\theta\left\{\sum_{i=1}^n \left|\left(T_{ni} - \frac{1}{4}\delta_n^2 t_n^2 \ddot{\mathbf{I}}(Y_i;\theta) - \frac{1}{12}(\delta_n t_n)^3 \dddot{\mathbf{I}}(Y_i;\theta)\right)\left(\delta_n t_n \dot{\mathbf{I}}(Y_i;\theta)\right)\right| > \epsilon'\right\} \\
&\le \frac{n}{\epsilon'} E_\theta\left|\left(T_n - \frac{1}{4}\delta_n^2 t_n^2 \ddot{\mathbf{I}}(\theta) - \frac{1}{12}(\delta_n t_n)^3 \dddot{\mathbf{I}}(\theta)\right)\left(\delta_n t_n \dot{\mathbf{I}}(\theta)\right)\right| \\
&\le \frac{n}{\epsilon'}\left[E_\theta\left(T_n - \frac{1}{4}\delta_n^2 t_n^2 \ddot{\mathbf{I}}(\theta) - \frac{1}{12}(\delta_n t_n)^3 \dddot{\mathbf{I}}(\theta)\right)^2\right]^{1/2}\left[E_\theta\left(\delta_n t_n \dot{\mathbf{I}}(\theta)\right)^2\right]^{1/2} \\
&\to 0.
\end{aligned}$$

□

Proposition 3.13. *We have, for any $k \ge 3$,*

$$\sum_{i=1}^n \left|T_{ni}\left(\delta_n t_n \dot{\mathbf{I}}(Y_i;\theta)\right)^k\right| \xrightarrow{p} 0,$$

uniformly in $|t_n| \le M$ and in $\theta \in K \subset \Theta$ for K compact, in P_θ probability.

Proof. Note that

$$\sum_{i=1}^n \left|T_{ni}\left(\delta_n t_n \dot{\mathbf{I}}(Y_i;\theta)\right)^k\right| \le \frac{1}{2}\sum_{i=1}^n (T_{ni})^2 \left|\delta_n t_n \dot{\mathbf{I}}(Y_i;\theta)\right|^{2k-5} + \frac{1}{2}\sum_{i=1}^n \left|\delta_n t_n \dot{\mathbf{I}}(Y_i;\theta)\right|^5.$$

By Proposition 3.8 and Proposition 3.7, we complete the proof. □

Proposition 3.14. *Define*

$$U(\theta) = U(Y_i;\theta) = \dot{\mathbf{I}}^2(Y_i;\theta) - \mathcal{I}(\theta).$$

We have

$$\mathbf{L}_\theta\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n U(Y_i;\theta)\right) \to N\left(0, \mathrm{Var}_\theta\left[\dot{\mathbf{I}}^2(\theta)\right]\right), \tag{3.14}$$

uniformly in $\theta \in K$ for compact $K \in \Theta$ in the weak topology, where $N(0, \mathrm{Var}_\theta[\dot{\mathbf{I}}^2(\theta)])$ is the normal distribution with mean 0 *and covariance matrix* $\mathrm{Var}_\theta[\dot{\mathbf{I}}^2(\theta)]$.

Proof. See the supplementary material [14]. □

Proposition 3.15. *Define*

$$V(Y_i;\theta) = \ddot{\mathbf{I}}(Y_i;\theta) - E_\theta\left[\ddot{\mathbf{I}}(Y_i;\theta)\right].$$

We have

$$\mathbf{L}_\theta\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n V(Y_i;\theta)\right) \to N\left(0, \mathrm{Var}\left[\ddot{\mathbf{I}}(\theta)\right]\right), \tag{3.15}$$

uniformly in $\theta \in K$ for compact $K \in \Theta$ in the weak topology, where $N(0, \mathrm{Var}[\ddot{\mathbf{I}}(\theta)])$ is the normal distribution with mean 0 *and covariance matrix* $\mathrm{Var}[\ddot{\mathbf{I}}(\theta)]$.

Proof. See the supplementary material [14]. □

Proposition 3.16. *We have, uniformly in $|t_n| \le M$ and in $\theta \in K \subset \Theta$ for K compact, in P_θ probability,*

$$\sum_{i=1}^n \delta_n^3 t_n^3 \dot{\mathbf{l}}^3(Y_i;\theta) - \sum_{i=1}^n \delta_n^3 t_n^3 E_\theta\big[\dot{\mathbf{l}}^3(Y_i;\theta)\big] \xrightarrow{p} 0, \tag{3.16}$$

$$\sum_{i=1}^n \delta_n^3 t_n^3 \dddot{\mathbf{l}}(Y_i;\theta) - \sum_{i=1}^n \delta_n^3 t_n^3 E_\theta\big[\dddot{\mathbf{l}}(Y_i;\theta)\big] \xrightarrow{p} 0, \tag{3.17}$$

$$\sum_{i=1}^n \delta_n^3 t_n^3 \ddot{\mathbf{l}}(Y_i;\theta)\dot{\mathbf{l}}(Y_i;\theta) - \sum_{i=1}^n \delta_n^3 t_n^3 E_\theta\big[\ddot{\mathbf{l}}(Y_i;\theta)\dot{\mathbf{l}}(Y_i;\theta)\big] \xrightarrow{p} 0. \tag{3.18}$$

Proof. We firstly prove equation (3.16). Since $E_\theta(\dot{\mathbf{l}}^6(\theta))$ is finite by Assumption 2.2, we have $\mathrm{Var}_\theta(\dot{\mathbf{l}}^3(\theta)) < \infty$. By Chebyshev's inequality and independence of data samples, we have that for any $\epsilon' > 0$

$$P_\theta\left[\left|\sum_{i=1}^n \big(\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta)\big)^3 - \sum_{i=1}^n E_\theta\big(\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta)\big)^3\right| > \epsilon'\right]$$
$$\le \frac{1}{(\epsilon')^2}\mathrm{Var}_\theta\left[\sum_{i=1}^n \big(\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta)\big)^3\right] = \frac{n\delta_n^6 M^6}{(\epsilon')^2}\mathrm{Var}_\theta\big(\dot{\mathbf{l}}^3(\theta)\big) \to 0,$$

uniformly in $\theta \in K$ and $|t_n| \le M$. Equation (3.17) can be proved similarly, since $E_\theta(\dddot{\mathbf{l}}(\theta))^2$ is finite by Assumption 2.2.

Next we prove equation (3.18). By Hölder's inequality,

$$E_\theta\big(\ddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)\big)^2 \le \big(E_\theta\big[\ddot{\mathbf{l}}(\theta)\big]^{2\times 3/2}\big)^{2/3}\big(E_\theta\big[\dot{\mathbf{l}}(\theta)\big]^{2\times 3}\big)^{1/3}.$$

By Assumption 2.2 we know that $\ddot{\mathbf{l}}(\theta)$ has finite third moment and $\dot{\mathbf{l}}(\theta)$ has finite sixth moment, which implies

$$\mathrm{Var}_\theta\big(\ddot{\mathbf{l}}(Y_i;\theta)\dot{\mathbf{l}}(Y_i;\theta)\big) < \infty.$$

By Chebyshev's inequality and independence of data samples, we have that $\epsilon' > 0$

$$P_\theta\left[\left|\sum_{i=1}^n \delta_n^3 t_n^3 \ddot{\mathbf{l}}(Y_i;\theta)\dot{\mathbf{l}}(Y_i;\theta) - \sum_{i=1}^n \delta_n^3 t_n^3 E_\theta\big[\ddot{\mathbf{l}}(Y_i;\theta)\dot{\mathbf{l}}(Y_i;\theta)\big]\right| > \epsilon'\right]$$
$$\le \frac{1}{(\epsilon')^2}\mathrm{Var}_\theta\left[\sum_{i=1}^n \delta_n^3 t_n^3 \ddot{\mathbf{l}}(Y_i;\theta)\dot{\mathbf{l}}(Y_i;\theta)\right] = \frac{n\delta_n^6 M^6}{(\epsilon')^2}\mathrm{Var}_\theta\big(\ddot{\mathbf{l}}(Y_i;\theta)\dot{\mathbf{l}}(Y_i;\theta)\big) \to 0,$$

uniformly in $\theta \in K$ and $|t_n| \le M$. □

3.3. Proof of Theorem 2.3

By Propositions 3.6, 3.7, 3.8, and 3.13, equation (3.3) can be rewritten as

$$\begin{aligned}
&\mathbf{l}(\theta+\delta_n t_n)-\mathbf{l}(\theta)\\
&\quad=2\sum_{i=1}^{n}T_{ni}+\sum_{i=1}^{n}\delta_n t_n\dot{\mathbf{l}}(Y_i;\theta)-\sum_{i=1}^{n}T_{ni}^2-\sum_{i=1}^{n}\big(T_{ni}\delta_n t_n\dot{\mathbf{l}}(Y_i;\theta)\big)\\
&\qquad-\frac{1}{4}\sum_{i=1}^{n}\big(\delta_n t_n\dot{\mathbf{l}}(Y_i;\theta)\big)^2+\frac{1}{12}\sum_{i=1}^{n}\big(\delta_n t_n\dot{\mathbf{l}}(Y_i;\theta)\big)^3+\frac{1}{2}\sum_{i=1}^{n}T_{ni}\big(\delta_n t_n\dot{\mathbf{l}}(Y_i;\theta)\big)^2\\
&\qquad-\frac{1}{32}\sum_{i=1}^{n}\big(\delta_n t_n\dot{\mathbf{l}}(Y_i;\theta)\big)^4+R_n(\theta,t_n).
\end{aligned}\tag{3.19}$$

Here and in the sequel, $R_n(\theta,t_n)\xrightarrow{p}0$ in P_θ probability, uniformly in $\theta\in K$ compact $\subset\Theta$ and $|t_n|\le M$, while the explicit expression of $R_n(\theta,t_n)$ may change line by line. The proof proceeds by tackling the terms in equation (3.19) one by one, which all hold uniformly in $\theta\in K$ compact $\subset\Theta$ and $|t_n|\le M$, as follows:

(1) For the term $2\sum_{i=1}^n T_{ni}$, by the result of Proposition 3.10,

$$\sum_{i=1}^{n}\bigg(T_{ni}-\frac{1}{4}\delta_n^2t_n^2\ddot{\mathbf{l}}(Y_i;\theta)-\frac{1}{12}(\delta_n t_n)^3\dddot{\mathbf{l}}(Y_i;\theta)-\frac{1}{48}(\delta_n t_n)^4\ddddot{\mathbf{l}}(Y_i;\theta)\bigg)\xrightarrow{p}0.$$

Then by Propositions 3.16 and 3.9,

$$\sum_{i=1}^{n}\bigg(T_{ni}-\frac{1}{4}\delta_n^2t_n^2\ddot{\mathbf{l}}(Y_i;\theta)-\frac{1}{12}(\delta_n t_n)^3E_\theta\big[\dddot{\mathbf{l}}(\theta)\big]-\frac{1}{48}(\delta_n t_n)^4E_\theta\big[\ddddot{\mathbf{l}}(\theta)\big]\bigg)\xrightarrow{p}0.$$

By the result of Proposition 3.15, we have,

$$\begin{aligned}
2\sum_{i=1}^{n}T_{ni}={}&\frac{1}{2}t_n^2\sqrt{n}\delta_n^2V_n(\theta)+\frac{1}{2}\delta_n^2t_n^2nE_\theta\big[\ddot{\mathbf{l}}(\theta)\big]+\frac{1}{6}(\delta_n t_n)^3nE_\theta\big[\dddot{\mathbf{l}}(\theta)\big]\\
&+\frac{1}{24}(\delta_n t_n)^4nE_\theta\big[\ddddot{\mathbf{l}}(\theta)\big]+R_n(\theta,t_n),
\end{aligned}$$

where $V_n(\theta)$ is defined in equation (2.3) and distributed as

$$\mathbf{L}_\theta\big(V_n(\theta)\big)\to N\big(0,\mathrm{Var}_\theta\big(\ddot{\mathbf{l}}(\theta)\big)\big).$$

(2) For the term $\sum_{i=1}^n\delta_n t_n\dot{\mathbf{l}}(Y_i;\theta)$, we have

$$\sum_{i=1}^{n}\delta_n t_n\dot{\mathbf{l}}(Y_i;\theta)=t_n\sqrt{n}\delta_nS_n(\theta),$$

where $S_n(\theta)$ is defined in equation (2.2) and distributed as

$$\mathbf{L}_\theta\big(S_n(\theta)\big)\to N\big(0,\mathcal{I}(\theta)\big),$$

in the weak topology, by Proposition 2.2 of [2].

(3) For the term $-\sum_{i=1}^n T_{ni}^2$, by Proposition 3.5, we have

$$-\sum_{i=1}^n T_{ni}^2 = -\sum_{i=1}^n \left(T_{ni}^2 - \frac{1}{16}\delta_n^4 t_n^4 E_\theta\left[\ddot{\mathbf{l}}(\theta)\right]^2\right) - \sum_{i=1}^n \frac{1}{16}\delta_n^4 t_n^4 E_\theta\left[\ddot{\mathbf{l}}(\theta)\right]^2$$
$$= -\frac{1}{16} n\delta_n^4 t_n^4 E_\theta\left[\ddot{\mathbf{l}}(\theta)\right]^2 + R_n(\theta, t_n).$$

(4) For the term $-\sum_{i=1}^n (T_{ni}\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta))$, we have

$$-\sum_{i=1}^n \left(T_{ni}\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta)\right)$$
$$= -\sum_{i=1}^n \left(T_{ni} - \frac{1}{4}\delta_n^2 t_n^2 \ddot{\mathbf{l}}(Y_i;\theta) - \frac{1}{12}\delta_n^3 t_n^3 \dddot{\mathbf{l}}(Y_i;\theta)\right)\left(\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta)\right)$$
$$- \frac{1}{4}\sum_{i=1}^n \delta_n^3 t_n^3 \ddot{\mathbf{l}}(Y_i;\theta)\dot{\mathbf{l}}(Y_i;\theta) - \frac{1}{12}\sum_{i=1}^n \delta_n^4 t_n^4 \dddot{\mathbf{l}}(Y_i;\theta)\dot{\mathbf{l}}(Y_i;\theta).$$

By Propositions 3.9, 3.12 and 3.16, we have

$$-\sum_{i=1}^n \left(T_{ni}\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta)\right)$$
$$= -\frac{1}{4} n\delta_n^3 t_n^3 E_\theta\left[\ddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)\right] - \frac{1}{12} n\delta_n^4 t_n^4 E_\theta\left[\dddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)\right] + R_n(\theta, t_n).$$

(5) For the term $-\frac{1}{4}\sum_{i=1}^n (\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta))^2$, we have

$$-\frac{1}{4}\sum_{i=1}^n \left(\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta)\right)^2 = -\frac{1}{4}\delta_n^2 t_n^2 \sum_{i=1}^n \left[\dot{\mathbf{l}}^2(Y_i;\theta) - \mathcal{I}(\theta)\right] - \frac{1}{4}\delta_n^2 t_n^2 n\mathcal{I}(\theta)$$
$$= -\frac{1}{4} t_n^2 \delta_n^2 \sqrt{n} U_n(\theta) - \frac{1}{4}\delta_n^2 t_n^2 n\mathcal{I}(\theta),$$

where $U_n(\theta)$ is defined in equation (2.4) and is distributed (Proposition 3.14) as

$$\mathbf{L}_\theta\left(U_n(\theta)\right) \to N\left(0, \mathrm{Var}_\theta\left[\dot{\mathbf{l}}^2(\theta)\right]\right).$$

(6) For the term $\frac{1}{12}\sum_{i=1}^n (\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta))^3$, by Proposition 3.16, we have

$$\frac{1}{12}\sum_{i=1}^n \left(\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta)\right)^3 = \frac{1}{12}\sum_{i=1}^n \delta_n^3 t_n^3 \left(\dot{\mathbf{l}}^3(Y_i;\theta) - E_\theta\left[\dot{\mathbf{l}}^3(Y_i;\theta)\right]\right)$$
$$+ \frac{1}{12} n\delta_n^3 t_n^3 E_\theta\left[\dot{\mathbf{l}}^3(\theta)\right]$$
$$= \frac{1}{12} n\delta_n^3 t_n^3 E_\theta\left[\dot{\mathbf{l}}^3(\theta)\right] + R_n(\theta, t_n).$$

(7) For the term $\frac{1}{2}\sum_{i=1}^{n} T_{ni}(\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta))^2$, by Proposition 3.11, we have

$$\begin{aligned}\frac{1}{2}\sum_{i=1}^{n} T_{ni}\big(\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta)\big)^2 &= \frac{1}{2}\sum_{i=1}^{n}\bigg(T_{ni} - \frac{1}{4}\delta_n^2 t_n^2 \ddot{\mathbf{l}}(Y_i;\theta)\bigg)\big(\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta)\big)^2 \\ &\quad + \frac{1}{8}\sum_{i=1}^{n}\delta_n^3 t_n^3 \ddot{\mathbf{l}}(Y_i;\theta)\dot{\mathbf{l}}(Y_i;\theta) \\ &= \frac{1}{8} n\delta_n^3 t_n^3 E_\theta\big[\ddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)\big] + R_n(\theta, t_n).\end{aligned}$$

(8) For the term $-\frac{1}{32}\sum_{i=1}^{n}(\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta))^4$, by Proposition 3.9, we have

$$\begin{aligned}&-\frac{1}{32}\sum_{i=1}^{n}\big(\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta)\big)^4 \\ &= -\frac{1}{32}\sum_{i=1}^{n}\big(\delta_n t_n \dot{\mathbf{l}}(Y_i;\theta)\big)^4 + \frac{1}{32} t_n^4 n\delta_n^4 E_\theta\big[\dot{\mathbf{l}}^4(\theta)\big] - \frac{1}{32} t_n^4 n\delta_n^4 E_\theta\big[\dot{\mathbf{l}}^4(\theta)\big] \\ &= -\frac{1}{32} t_n^4 n\delta_n^4 E_\theta\big[\dot{\mathbf{l}}^4(\theta)\big] + R_n(\theta, t_n).\end{aligned}$$

Now, we can rewrite equation (3.19) as

$$\begin{aligned}&\mathbf{l}(\theta+\delta_n t_n) - \mathbf{l}(\theta) \\ &= \frac{1}{2} t_n^2 \sqrt{n}\delta_n^2 V_n(\theta) + \frac{1}{2}\delta_n^2 t_n^2 n E_\theta\big[\ddot{\mathbf{l}}(\theta)\big] + \frac{1}{6}(\delta_n t_n)^3 n E_\theta\big[\dddot{\mathbf{l}}(\theta)\big] + \frac{1}{24}(\delta_n t_n)^4 n E_\theta\big[\ddddot{\mathbf{l}}(\theta)\big] \\ &\quad + t_n\sqrt{n}\delta_n S_n(\theta) - \frac{1}{16} n\delta_n^4 t_n^4 E_\theta\big[\ddot{\mathbf{l}}(\theta)\big]^2 - \frac{1}{4} n\delta_n^3 t_n^3 E_\theta\big[\ddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)\big] - \frac{1}{12} n\delta_n^4 t_n^4 E_\theta\big[\dddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)\big] \\ &\quad - \frac{1}{4} t_n^2\delta_n^2\sqrt{n}U_n(\theta) - \frac{1}{4}\delta_n^2 t_n^2 n\mathcal{I}(\theta) + \frac{1}{12} n\delta_n^3 t_n^3 E_\theta\big[\dot{\mathbf{l}}^3(\theta)\big] + \frac{1}{8} n\delta_n^3 t_n^3 E_\theta\big[\ddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)\big] \\ &\quad - \frac{1}{32} t_n^4 n\delta_n^4 E_\theta\big[\dot{\mathbf{l}}^4(\theta)\big] + R_n(\theta, t_n).\end{aligned}$$

Reorganizing the terms, we have

$$\begin{aligned}&\mathbf{l}(\theta+\delta_n t_n) - \mathbf{l}(\theta) \\ &= t_n\big\{\sqrt{n}\delta_n S_n(\theta)\big\} + t_n^2\bigg\{\sqrt{n}\delta_n^2\bigg[\frac{1}{2}V_n(\theta) - \frac{1}{4}U_n(\theta)\bigg] + n\delta_n^2\bigg[\frac{1}{2}E_\theta\big[\ddot{\mathbf{l}}(\theta)\big] - \frac{1}{4}\mathcal{I}(\theta)\bigg]\bigg\} \\ &\quad + t_n^3\bigg\{n\delta_n^3\bigg[\frac{1}{12}E_\theta\big[\dot{\mathbf{l}}^3(\theta)\big] - \frac{1}{8}E_\theta\big[\ddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)\big] + \frac{1}{6}E_\theta\big[\dddot{\mathbf{l}}(\theta)\big]\bigg]\bigg\} \\ &\quad + t_n^4\bigg\{n\delta_n^4\bigg[-\frac{1}{32}E_\theta\big[\dot{\mathbf{l}}^4(\theta)\big] - \frac{1}{16}E_\theta\big[\ddot{\mathbf{l}}(\theta)\big]^2 - \frac{1}{12}E_\theta\big[\dddot{\mathbf{l}}(\theta)\dot{\mathbf{l}}(\theta)\big] + \frac{1}{24}E_\theta\big[\ddddot{\mathbf{l}}(\theta)\big]\bigg]\bigg\} \\ &\quad + R_n(\theta, t_n).\end{aligned}$$

We complete the proof by noting that differentiating $\int s^2(\theta)\,d\mu = 1$ with respect to θ yields $\int \dot{s}(\theta)s(\theta)\,d\mu = 0$, and further differentiating with respect to θ yields

$$\int \ddot{s}(\theta)s(\theta)\,d\mu + \int \dot{s}^2(\theta)\,d\mu = 0,$$

which gives

$$2E_\theta\big[\ddot{\mathbf{l}}(\theta)\big] = 4\int \ddot{s}(\theta)s(\theta)\,d\mu = -4\int \dot{s}^2(\theta)\,d\mu = -\mathcal{I}(\theta).$$

4. Properties of the MCLE methodology

In this section, we firstly elaborate how $\{\overline{\beta}_\iota\}_{\iota\in\{1,2,3\}}$ in equation (1.13) may be obtained in practice. Recalling that $\theta^*_{j,n} - \theta^*_n = jn^{-1/4}$ for j in a finite set $\mathcal{J}$. We take the classical setting that $\mathcal{J} = \{-J, -J+1, \dots, 0, \dots, J-1, J\}$ for J being an integer greater than 1, where we exclude $J = 1$ since we need at least 4 values to interpolate a cubic polynomial curve. Write

$$\overline{\mathbf{Y}} = \begin{pmatrix} \bar{\mathbf{l}}(\theta^*_{-J,n}) \\ \vdots \\ \bar{\mathbf{l}}(\theta^*_n) \\ \vdots \\ \bar{\mathbf{l}}(\theta^*_{J,n}) \end{pmatrix}, \qquad \overline{\mathbf{X}} = \begin{pmatrix} 1 & (-Jn^{-1/4}) & (-Jn^{-1/4})^2 & (-Jn^{-1/4})^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (Jn^{-1/4}) & (Jn^{-1/4})^2 & (Jn^{-1/4})^3 \end{pmatrix},$$

$\overline{\beta} = (\overline{\beta}_0, \overline{\beta}_1, \overline{\beta}_2, \overline{\beta}_3)^T$, and $\overline{\epsilon} = (\overline{\epsilon}_{-J,n}, \dots, \overline{\epsilon}_0, \dots, \overline{\epsilon}_{J,n})^T$, where the superscript "T" stands for the transpose operation. Now we fit "data" $\overline{\mathbf{Y}}$ to $\overline{\mathbf{X}}$ by linear regression

$$\overline{\mathbf{Y}} = \overline{\mathbf{X}}\overline{\beta} + \overline{\epsilon}.$$

By least-squares estimation, we obtain the estimation of regression coefficients as

$$\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3)^T = \big(\overline{\mathbf{X}}^T\overline{\mathbf{X}}\big)^{-1}\overline{\mathbf{X}}^T\overline{\mathbf{Y}}, \tag{4.1}$$

where

$$\overline{\mathbf{X}}^T\overline{\mathbf{X}} = \begin{pmatrix} J & \sum_{j=-J}^{J}(jn^{-1/4}) & \sum_{j=-J}^{J}(jn^{-1/4})^2 & \sum_{j=-J}^{J}(jn^{-1/4})^3 \\ \sum_{j=-J}^{J}(jn^{-1/4}) & \sum_{j=-J}^{J}(jn^{-1/4})^2 & \sum_{j=-J}^{J}(jn^{-1/4})^3 & \sum_{j=-J}^{J}(jn^{-1/4})^4 \\ \sum_{j=-J}^{J}(jn^{-1/4})^2 & \sum_{j=-J}^{J}(jn^{-1/4})^3 & \sum_{j=-J}^{J}(jn^{-1/4})^4 & \sum_{j=-J}^{J}(jn^{-1/4})^5 \\ \sum_{j=-J}^{J}(jn^{-1/4})^3 & \sum_{j=-J}^{J}(jn^{-1/4})^4 & \sum_{j=-J}^{J}(jn^{-1/4})^5 & \sum_{j=-J}^{J}(jn^{-1/4})^6 \end{pmatrix}.$$

Before we investigate the order of $\widehat{\beta}$, let us firstly explore the orders of the determinant and adjugate matrix of $\overline{\mathbf{X}}^T\overline{\mathbf{X}}$ in the following two lemmas.

Lemma 4.1. *For J being a fixed integer greater than* 1, *the determinant of* $\overline{\mathbf{X}}^T\overline{\mathbf{X}}$, *denoted as* $\det(\overline{\mathbf{X}}^T\overline{\mathbf{X}})$, *is of order* $\mathcal{O}(n^{-3})$.

Proof. See the supplementary material [14]. □

Lemma 4.2. *For J being a fixed integer greater than* 1, *the adjugate of* $\overline{\mathbf{X}}^T\overline{\mathbf{X}}$, *denoted as* $\operatorname{adj}(\overline{\mathbf{X}}^T\overline{\mathbf{X}})$, *has that*

$$\operatorname{adj}\big(\overline{\mathbf{X}}^T\overline{\mathbf{X}}\big)_{22} = \mathcal{O}\big(n^{-5/2}\big) \quad \text{and} \quad \operatorname{adj}\big(\overline{\mathbf{X}}^T\overline{\mathbf{X}}\big)_{33} = \mathcal{O}\big(n^{-2}\big).$$

Proof. See the supplementary material [14]. □

Recall that in equation (1.13), we have $\overline{\beta}_2 = \mathcal{O}(n)$ and $\overline{\beta}_3 = \mathcal{O}(n)$, and then we can see that the contribution from the term $\overline{\beta}_3(jn^{-1/4})^3$ in finding the desired maximizer in the interval $[-Jn^{-1/4}, Jn^{-1/4}]$ is asymptotically negligible on the $\mathcal{O}(n^{-1/2})$ scale. The MCLE is therefore close to the maximizer of the quadratic approximation. Given that the coefficient of the quadratic term is negative by the RLAN property in equation (1.5), we have

$$\widehat{\theta}_n^{\text{MCLE}} = \theta_n^* + \frac{\widehat{\beta}_1}{-2\widehat{\beta}_2} + o\big(n^{-1/2}\big), \tag{4.2}$$

where $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are given in equation (4.1). In the following theorem, we compare the performance of $\widehat{\theta}_n^{\text{MCLE}}$ with the generalized estimator $\widehat{\theta}_n^C$, which is defined on the $\delta_n = n^{-1/4}$ scale as follows:

$$\widehat{\theta}_n^C = \theta_n^* + \delta_n \times \frac{\sqrt{n}\delta_n S_n(\theta_n^*)}{n\delta_n^2 \mathcal{I}(\theta_n^*)}. \tag{4.3}$$

Note that, $\widehat{\theta}_n^C$ is the generalization of $\widehat{\theta}_n^B$ (Section 1.1), for the reason that equation (4.3) with $\delta_n = n^{-1/2}$ instead of $\delta_n = n^{-1/4}$, gives $\widehat{\theta}_n^B$.

Theorem 4.3. *Suppose that the data are modeled as an i.i.d. sequence $Y_1, \dots, Y_n$ drawn from a regular parametric model $\mathbf{P} = \{P_\theta : \theta \in \Theta\}$ satisfying Assumption* 2.2. *Take $m = m(n)$ Monte Carlo simulations per observation where $\mathcal{O}(\sqrt{n}) \ll m \ll \mathcal{O}(n)$, and take $\mathcal{J} = \{-J, -J+1, \dots, 0, \dots, J-1, J\}$ for J being a fixed integer greater than* 1. *Then the maximum cubic log-likelihood estimator $\widehat{\theta}_n^{\text{MCLE}}$ is efficient.*

Proof. By the classical results of linear regression (see, e.g., equation (2.13) on page 12 of [18]), we know that $\mathbb{E}(\widehat{\beta}_1) = \overline{\beta}_1$ and $\mathbb{E}(\widehat{\beta}_2) = \overline{\beta}_2$. By the RLAN property (Definition 1.1) we can see that the coefficients $\overline{\beta}_1$ and $\overline{\beta}_2$ are of order $\sqrt{n}$ and n, respectively (see Section 1.2 (2) for illustration). Recall that by least-squares estimation, we have

$$\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3)^T = \big(\overline{\mathbf{X}}^T\overline{\mathbf{X}}\big)^{-1}\overline{\mathbf{X}}^T\overline{\mathbf{Y}}.$$

By Lemma 4.1, we have $\det(\overline{\mathbf{X}}^T\overline{\mathbf{X}}) = \mathcal{O}(n^{-3})$, and by Lemma 4.2 we can see that $\operatorname{adj}(\overline{\mathbf{X}}^T\overline{\mathbf{X}})_{22} = \mathcal{O}(n^{-5/2})$ and $\operatorname{adj}(\overline{\mathbf{X}}^T\overline{\mathbf{X}})_{33} = \mathcal{O}(n^{-2})$. Hence, by the formula that

$$\big(\overline{\mathbf{X}}^T\overline{\mathbf{X}}\big)_{ij}^{-1} = \left(\frac{1}{\det(\overline{\mathbf{X}}^T\overline{\mathbf{X}})} \operatorname{adj}\big(\overline{\mathbf{X}}^T\overline{\mathbf{X}}\big)\right)_{ij},$$

we have

$$\big(\overline{\mathbf{X}}^T\overline{\mathbf{X}}\big)^{-1}_{22} = \mathcal{O}\big(n^{1/2}\big) \quad \text{and} \quad \big(\overline{\mathbf{X}}^T\overline{\mathbf{X}}\big)^{-1}_{33} = \mathcal{O}(n). \tag{4.4}$$

By the classical results of linear regression (see, e.g., equation (2.15) on page 12 of [18]), we have that

$$\operatorname{Var}(\widehat{\beta}_1) = \big(\overline{\mathbf{X}}^T\overline{\mathbf{X}}\big)^{-1}_{22} \operatorname{Var}(\overline{\epsilon}_{j,n}) \quad \text{and} \quad \operatorname{Var}(\widehat{\beta}_2) = \big(\overline{\mathbf{X}}^T\overline{\mathbf{X}}\big)^{-1}_{33} \operatorname{Var}(\overline{\epsilon}_{j,n}).$$

Then by the delta method (see, Proposition 9.32 in [10]) and by equation (4.4), together with the fact that $\frac{m}{n}\overline{\epsilon}_{j,n}$ convergences in distribution to a normal distribution having mean zero and positive finite variance, one obtains that

$$\operatorname{Var}(\widehat{\beta}_1) = \mathcal{O}\bigg(n^{1/2} \times \frac{n}{m}\bigg) \quad \text{and} \quad \operatorname{Var}(\widehat{\beta}_2) = \mathcal{O}\bigg(n \times \frac{n}{m}\bigg).$$

By the covariance inequality that for two random variables ζ_1 and ζ_2 their covariance

$$\operatorname{Cov}(\zeta_1, \zeta_2) \le \sqrt{\operatorname{Var}(\zeta_1)\operatorname{Var}(\zeta_2)},$$

we have

$$\frac{\overline{\beta}_1^2}{4\overline{\beta}_2^2}\bigg[\frac{\operatorname{Var}(\widehat{\beta}_1)}{\overline{\beta}_1^2} - 2\frac{\operatorname{Cov}(\widehat{\beta}_1, \widehat{\beta}_2)}{\overline{\beta}_1\overline{\beta}_2} + \frac{\operatorname{Var}(\widehat{\beta}_2)}{\overline{\beta}_2^2}\bigg] = \mathcal{O}\bigg(\frac{n}{n^2}\bigg[\frac{n^{1/2} \times \frac{n}{m}}{n}\bigg]\bigg) = \mathcal{O}\bigg(\frac{n^{1/2} \times \frac{1}{m}}{n}\bigg).$$

By the delta method, one has that as $n \to \infty$, $\frac{n}{n^{1/2} \times \frac{1}{m}}(\frac{\widehat{\beta}_1}{-2\widehat{\beta}_2})$ convergences in distribution to a normal distribution having positive finite variance. Recall that under the conditions imposed in Theorem 4.3, we have equation (4.2). Then, in the context of Monte Carlo inference under investigation, we have

$$\mathrm{SE}^2_{\mathrm{MC}} = \mathcal{O}\bigg(\operatorname{Var}\bigg(\frac{\widehat{\beta}_1}{-2\widehat{\beta}_2}\bigg)\bigg) = \mathcal{O}\bigg(\frac{n^{1/2} \times \frac{1}{m}}{n}\bigg).$$

In the context of Monte Carlo inference under investigation, the statistical standard error $\mathrm{SE}_{\mathrm{stat}}$ is the standard deviation of the MCLE constructed with no Monte Carlo error. Thus $\mathrm{SE}^2_{\mathrm{stat}}$ is given by the variance of the classical efficient estimator $\widehat{\theta}^C_n$, that is,

$$\mathrm{SE}^2_{\mathrm{stat}} = \operatorname{Var}\big[\widehat{\theta}^C_n\big] = \frac{1}{n\mathcal{I}(\theta)} = \mathcal{O}\bigg(\frac{1}{n}\bigg). \tag{4.5}$$

Hence, given $\mathcal{O}(\sqrt{n}) \ll m(n) \ll \mathcal{O}(n)$, we have $\mathrm{SE}^2_{\mathrm{MC}} / \mathrm{SE}^2_{\mathrm{stat}} \to 0$. Furthermore, the gradient of the bias in the likelihood evaluation, $C_\gamma n/m$, leads to a bias of order $1/m$ in finding the location of θ that gives the maximum of the metamodel. Taking $\mathcal{O}(\sqrt{n}) \ll m(n) \ll \mathcal{O}(n)$ ensures the asymptotic bias in the estimator is negligible compared $\mathrm{SE}_{\mathrm{stat}}$. □

Acknowledgements

The authors would like to thank the anonymous reviewers, the Associate Editor, and the Editor-in-Chief for their constructive comments that greatly improved the quality of this paper. This research project was supported by NSF grant DMS-1761603.

Supplementary Material

Supplement: Proofs of some of the results (DOI: 10.3150/20-BEJ1321SUPP; .pdf). The supplementary material contains the proofs of Lemmas 3.1 and 3.2, Propositions 3.5, 3.8, 3.14 and 3.15, and Lemmas 4.1 and 4.2.

References

[1] Bickel, P., Choi, D., Chang, X. and Zhang, H. (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Ann. Statist.* **41** 1922–1943. MR3127853 https://doi.org/10.1214/13-AOS1124

[2] Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. *Johns Hopkins Series in the Mathematical Sciences*. Baltimore, MD: Johns Hopkins Univ. Press. MR1245941

[3] Bickel, P.J. and Ritov, Y. (1996). Inference in hidden Markov models. I. Local asymptotic normality in the stationary case. *Bernoulli* **2** 199–228. MR1416863 https://doi.org/10.2307/3318520

[4] Bretó, C., Ionides, E.L. and King, A.A. (2020). Panel data analysis via mechanistic models. *J. Amer. Statist. Assoc.* **115** 1178–1188. MR4143457 https://doi.org/10.1080/01621459.2019.1604367

[5] Diggle, P.J. and Gratton, R.J. (1984). Monte Carlo methods of inference for implicit statistical models (with discussion). *J. Roy. Statist. Soc. Ser. B* **46** 193–227. MR0781880

[6] Hallin, M., Taniguchi, M., Serroukh, A. and Choy, K. (1999). Local asymptotic normality for regression models with long-memory disturbance. *Ann. Statist.* **27** 2054–2080. MR1765628 https://doi.org/10.1214/aos/1017939250

[7] Höpfner, R., Jacod, J. and Ladelli, L. (1990). Local asymptotic normality and mixed normality for Markov statistical models. *Probab. Theory Related Fields* **86** 105–129. MR1061951 https://doi.org/10.1007/BF01207516

[8] Ionides, E.L. (2005). Maximum smoothed likelihood estimation. *Statist. Sinica* **15** 1003–1014. MR2234410

[9] Ionides, E.L., Breto, C., Park, J., Smith, R.A. and King, A.A. (2017). Monte Carlo profile confidence intervals for dynamic systems. *J. R. Soc. Interface* **14** 1–10.

[10] Keener, R.W. (2010). *Theoretical Statistics*: *Topics for a Core Course*. *Springer Texts in Statistics*. New York: Springer. MR2683126 https://doi.org/10.1007/978-0-387-93839-4

[11] Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. *Springer Series in Statistics*. New York: Springer. MR0856411 https://doi.org/10.1007/978-1-4612-4946-7

[12] Le Cam, L. and Yang, G.L. (2012). *Asymptotics in Statistics*: *Some Basic Concepts*. New York: Springer.

[13] Murphy, S.A. and van der Vaart, A.W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95** 449–485. MR1803168 https://doi.org/10.2307/2669386

[14] Ning, N., Ionides, E.L. and Ritov, Y. (2021). Supplement to "Scalable Monte Carlo inference and rescaled local asymptotic normality." https://doi.org/10.3150/20-BEJ1321SUPP

[15] Pons-Salort, M. and Grassly, N.C. (2018). Serotype-specific immunity explains the incidence of diseases caused by human enteroviruses. *Science* **361** 800–803. https://doi.org/10.1126/science.aat6777

[16] Ranjeva, S., Subramanian, R., Fang, V.J., Leung, G.M., Ip, D.K., Perera, R.A., Peiris, J.S.M., Cowling, B.J. and Cobey, S. (2019). Age-specific differences in the dynamics of protective immunity to influenza. *Nat. Commun.* **10** 1660.

[17] Ranjeva, S.L., Baskerville, E.B., Dukic, V., Villa, L.L., Lazcano-Ponce, E., Giuliano, A.R., Dwyer, G. and Cobey, S. (2017). Recurring infection with ecologically distinct HPV types can explain high prevalence and diversity. *Proc. Natl. Acad. Sci. USA* **114** 13573–13578.

[18] Rao, C.R., Toutenburg, H., Shalabh and Heumann, C. (2008). *Linear Models and Generalizations*. *Least Squares and Alternatives*, 3rd ed. *Springer Series in Statistics*. Berlin: Springer. MR2370506

[19] Smith, R.A., Ionides, E.L. and King, A.A. (2017). Infectious disease dynamics inferred from genetic data via sequential Monte Carlo. *Mol. Biol. Evol.* **34** 2065–2084. https://doi.org/10.1093/molbev/msx124

[20] Uhlenbeck, G.E. and Ornstein, L.S. (1930). On the theory of the Brownian motion. *Phys. Rev.* **36** 823.

Received July 2020 and revised November 2020