Statistics and Probability Letters 129 (2017) 171–181

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Bounds for the normal approximation of the maximum likelihood estimator from m-dependent random variables

Andreas Anastasiou

London School of Economics and Political Science, London, UK

A R T I C L E   I N F O

Article history:
Received 20 September 2016
Received in revised form 24 April 2017
Accepted 24 April 2017
Available online 9 June 2017

Keywords:
Maximum likelihood estimator
Dependent random variables
Normal approximation
Stein's method

A B S T R A C T

The asymptotic normality of the Maximum Likelihood Estimator (MLE) is a long established result. Explicit bounds for the distributional distance between the distribution of the MLE and the normal distribution have recently been obtained for the case of independent random variables. In this paper, a local dependence structure is introduced between the random variables and we give upper bounds which are specified for the Wasserstein metric.

1. Introduction

The asymptotic normality of Maximum Likelihood Estimators (MLEs) was first discussed in Fisher (1925). It is a fundamental qualitative result and a cornerstone in mathematical statistics. The aim of assessing the quality of this normal approximation in the case of independent random variables has recently been accomplished in Anastasiou and Reinert (2017) where the case of single-parameter distributions has been treated and the result is so general that no restrictions on the form of the MLE are required. Anastasiou and Ley (2017), provide an alternative approach to the problem based partly on the Delta method for cases where the MLE can be expressed as a function of the sum of independent terms. For such sum structures of the MLE, Pinelis and Molzon (2016), through also a Delta method approach provide results on the closeness of the distribution of the MLE to the normal distribution in the Kolmogorov distance under different conditions than those used in Anastasiou and Ley (2017). Bounds related to the normal approximation of interest for the case of high-dimensional and heterogeneous data from multi-parameter distributions have already been obtained in Anastasiou (2016).

In this paper, the independence assumption is relaxed and we assess the normal approximation of the MLE under the presence of a local dependence structure between the random variables; for limit theorems for sums of m-dependent random variables see Heinrich (1982), Berk (1973) and Orey (1958). For our purpose, we partly employ a powerful probabilistic technique called Stein's method, first introduced by Charles Stein in Stein (1972), while the monograph (Stein, 1986) explains in detail the method and it is in our opinion the most notable contribution. Stein's method is used to assess whether a random variable, W, has a distribution close to a target distribution. In this paper, the normal approximation related to the MLE is assessed in terms of the Wasserstein distance. If F, G are two random variables with values in $\mathbb{R}$ and

$$H_W = \{h : \mathbb{R} \to \mathbb{R} : |h(x) - h(y)| \leq |x - y|\} \tag{1}$$

E-mail address: A.Anastasiou@lse.ac.uk.

http://dx.doi.org/10.1016/j.spl.2017.04.022
0167-7152/

is the set of Lipschitz functions with constant equal to one, then the Wasserstein distance between the laws of F and G is

$$d_W(F, G) = \sup\{|\mathrm{E}[h(F)] - \mathrm{E}[h(G)]| : h \in H_W\}.$$

We need to mention that Berry–Esseen type bounds for the Kolmogorov distance between the distribution of the sum of m-dependent random variables and the normal distribution are given in Erickson (1974). However, the results of our paper, which are related to the MLE, are much more general in the sense that they can be applied whatever the form of the MLE is (not necessarily a sum). In addition, our results are given in terms of the aforementioned Wasserstein distance which allows someone to obtain bounds on the Kolmogorov distance too since

$$d_W(F, Z) \leq 2\sqrt{d_K(F, Z)}$$

where Z follows the standard normal distribution and $d_K(F, Z)$ is the Kolmogorov distance between the distribution of F and the standard normal distribution. For a proof of this result see Theorem 3.3 of Chen et al. (2011).

A general approach is first developed to get upper bounds on the Wasserstein distance between the distribution of the suitably scaled MLE and the standard normal distribution; here Stein's method is used for some results. The special case of independent random variables is briefly discussed, while an example of normally distributed locally dependent random variables serves as an illustration of the main results.

The notion of local dependence is introduced before the Stein's method result that is used in the case of locally dependent random variables is given. An m-dependent sequence of random variables $\{X_i, i \in \mathbb{N}\}$ is such that for each $i \in \mathbb{N}$ the sets of random variables $\{X_j, j \leq i\}$ and $\{X_j, j > i + m\}$ are independent. The Stein's method result for the case of locally dependent random variables is based on the local dependence condition (LD); for a set of random variables $\{\xi_i, i = 1, 2, \ldots, n\}$, for any $A \subset \{1, 2, \ldots, n\}$ we define

$$A^c = \{i \in \{1, 2, \ldots, n\} : i \notin A\}, \qquad \xi_A = \{\xi_i : i \in A\}.$$

Then,

(LD) For each $i \in \{1, 2, \ldots, n\}$ there exist $A_i \subset B_i \subset \{1, 2, \ldots, n\}$ such that ξ_i is independent of $\xi_{A_i^c}$ and ξ_{A_i} is independent of $\xi_{B_i^c}$.

Whenever this condition holds,

$$\eta_i = \sum_{j \in A_i} \xi_j, \qquad \tau_i = \sum_{j \in B_i} \xi_j. \tag{2}$$

Lemma 1.1 gives an upper bound on the Wasserstein distance between the distribution of a sum of m-dependent random variables satisfying (LD) and the normal distribution. The random variables are assumed to have mean zero with the variance of their sum being equal to one. The proof of the lemma is beyond the scope of the paper and can be found in Chen et al. (2011, p. 134).

Lemma 1.1. *Let $\{\xi_i, i = 1, 2, \ldots, n\}$ be a set of random variables with mean zero and* $\mathrm{Var}(W) = 1$, *where $W = \sum_{i=1}^{n} \xi_i$. If* (LD) *holds, then with η_i and τ_i as in (2),*

$$d_W(W, Z) \leq 2\sum_{i=1}^{n} \left(\mathrm{E}|\xi_i\eta_i\tau_i| + |\mathrm{E}(\xi_i\eta_i)|\,\mathrm{E}|\tau_i|\right) + \sum_{i=1}^{n} \mathrm{E}\left|\xi_i\eta_i^2\right|. \tag{3}$$

Now the notation used throughout the paper is explained. First of all, θ is a scalar unknown parameter found in a parametric statistical model. Let θ_0 be the true (still unknown) value of the parameter θ and let $\Theta \subset \mathbb{R}$ denote the parameter space, while $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ for $\{X_i, i = 1, 2, \ldots, n\}$ an m-dependent sequence of identically distributed random variables. The joint density function of $X_1, X_2, \ldots, X_n$ is

$$\begin{aligned} f(\boldsymbol{x}|\theta) = L(\theta; \boldsymbol{x}) &= f(x_1; \theta) f(x_2|x_1; \theta) \ldots f(x_n|x_{n-1}, \ldots, x_{n-m}; \theta) \\ &= f(x_1; \theta) \prod_{i=2}^{n} f(x_i|x_{i-1}, \ldots, x_{m_i^*}; \theta), \end{aligned}$$

where $m_i^* = \max\{i - m, 1\}$. The likelihood function is $L(\theta; \boldsymbol{x}) = f(\boldsymbol{x}|\theta)$. Its natural logarithm, called the log-likelihood function is denoted by $l(\theta; \boldsymbol{x})$. Derivatives of the log-likelihood function, with respect to θ, are denoted by $l'(\theta; \boldsymbol{x}), l''(\theta; \boldsymbol{x}), \ldots, l^{(j)}(\theta; \boldsymbol{x})$, for j any integer greater than 2. The MLE is denoted by $\hat{\theta}_n(\boldsymbol{X})$. For many models the MLE exists and is unique; this is known as the 'regular' case. For a number of statistical models, however, uniqueness or even existence of the MLE is not secured; see Billingsley (1961) for an example of non-uniqueness. Assumptions that ensure existence and uniqueness of the MLE are given in Mäkeläinen et al. (1981).

In Section 2 we explain, for locally dependent random variables, the process of finding an upper bound on the Wasserstein distance between the distribution of the suitably scaled MLE and the standard normal distribution. The quantity we are interested in is split into two terms with the one being bounded using Stein's method and the other using alternative techniques based mainly on Taylor expansions. After obtaining the general upper bound, we comment on how our bound behaves for i.i.d. ($m = 0$) random variables and this specific result is compared to already existing bounds for i.i.d. random variables as given in Anastasiou and Reinert (2017). The main result of this paper is applied in Section 3 to the case of 1-dependent normally distributed random variables.

2. The general bound

The purpose is to obtain an upper bound on the Wasserstein distance between the distribution of an appropriately scaled MLE and the standard normal distribution. The results of Lemma 1.1 will be applied to a sequence $\{\xi_i, i = 1, 2, \ldots, n\}$ of $2m$-dependent random variables. We denote by

$$M_{1j} := \max\{1, j-2m\} \qquad M_{2j} := \min\{n, j+2m\}$$
$$K_{1j} := \max\{1, j-4m\} \qquad K_{2j} := \min\{n, j+4m\}.$$

In addition, the dependency neighbourhoods, A_j and B_j, as defined in (LD) are

$$A_j = \left\{M_{1j}, M_{1j}+1, \ldots, M_{2j}-1, M_{2j}\right\}, \qquad B_j = \left\{K_{1j}, K_{1j}+1, \ldots, K_{2j}-1, K_{2j}\right\}. \tag{4}$$

Having that $\forall i \in \{1, 2, \ldots, n\}$, $|A_i|$ and $|B_i|$ denote the number of elements in the sets A_i and B_i, respectively, then

$$|A_i| \leq 4m+1, \qquad |B_i| \leq 8m+1.$$

We work under the following assumptions:

(A.D.1) The log-likelihood function is three times differentiable with uniformly bounded third derivative in $\theta \in \Theta$, $(x_1, x_2, \ldots, x_n) \in S$. The supremum is denoted by

$$S_d(n) := \sup_{\substack{\theta\in\Theta \\ \boldsymbol{x}\in S}} \left|l^{(3)}(\theta; \boldsymbol{x})\right| < \infty. \tag{5}$$

(A.D.2) $\mathrm{E}\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta)\right] = \mathrm{E}\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_i|X_{i-1}, \ldots, X_{i-m}; \theta)\right] = 0$, for $i = 2, 3, \ldots, n$.

(A.D.3) With θ_0, as usual, denoting the true value of the unknown parameter,

$$\sqrt{n}\mathrm{E}\left[\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right] \xrightarrow[n\to\infty]{} 0.$$

(A.D.4) The limit of the reciprocal of $n\mathrm{Var}\left(\hat{\theta}_n(\boldsymbol{X})\right)$ exists and from now on, unless otherwise stated,

$$0 < i_2(\theta_0) = \lim_{n\to\infty} \frac{1}{n\mathrm{Var}(\hat{\theta}_n(\boldsymbol{X}))}.$$

The following theorem gives the bound.

Theorem 2.1. *Let $\{X_i, i = 1, 2, \ldots, n\}$ be an m-dependent sequence of identically distributed random variables with probability density (or mass) function $f(x_i|x_{i-1}, \ldots, x_{i-m}; \theta)$, where $\theta \in \Theta$ and $(x_1, x_2, \ldots, x_n) \in S$, where S is the support of the joint probability density (or mass) function. Assume that $\hat{\theta}_n(\boldsymbol{X})$ exists and is unique. In addition, assume that* (A.D.1)–(A.D.4) *hold and that* $\mathrm{Var}\left[l'(\theta_0; \boldsymbol{X})\right] > 0$. *Let*

$$\alpha := \alpha(\theta_0, n) := \sqrt{\frac{\mathrm{Var}\left(l'(\theta_0; \boldsymbol{X})\right)}{\mathrm{Var}\left(\hat{\theta}_n(\boldsymbol{X})\right)}}, \tag{6}$$

which is assumed to be finite and not equal to zero. In addition, we denote by

$$\xi_1 = \frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta)\Big|_{\theta=\theta_0}\sqrt{\frac{n}{\mathrm{Var}(l'(\theta_0; \boldsymbol{X}))}}$$

and for $i = 2, 3, \ldots, n$,

$$\xi_i = \frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_i|X_{i-1}, \ldots, X_{i-m}; \theta)\Big|_{\theta=\theta_0}\sqrt{\frac{n}{\mathrm{Var}(l'(\theta_0; \boldsymbol{X}))}}.$$

Then, for $Z \sim \mathrm{N}(0, 1)$,

$$d_W\left(\sqrt{n\, i_2(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right), Z\right) \leq \frac{2}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j\in A_i} \sum_{k\in B_i} \left[\mathrm{E}\left((\xi_i)^4\right)\mathrm{E}\left((\xi_j)^4\right)\mathrm{E}\left((\xi_k)^4\right)\right]^{\frac{1}{4}}$$

$$+\frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left((\xi_i)^2\right)\mathrm{E}\left(\left(\xi_j\right)^2\right)\mathrm{E}\left((\xi_k)^2\right)\right]^{\frac{1}{2}}+\frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}|A_i|\sum_{j\in A_i}\left[\mathrm{E}\left((\xi_i)^2\right)\mathrm{E}\left(\left(\xi_j\right)^4\right)\right]^{\frac{1}{2}}$$
$$+\left|\frac{\sqrt{n\,i_2(\theta_0)\mathrm{Var}[l'(\theta_0;\boldsymbol{X})]}}{\alpha}-1\right|+\frac{S_d(n)\sqrt{n\,i_2(\theta_0)}}{2\alpha}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]$$
$$+\frac{\sqrt{n\,i_2(\theta_0)}}{\alpha}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]}\sqrt{\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X})+\alpha\right)^2\right]}. \tag{7}$$

Proof. By the definition of the MLE and (A.D.1), $l'\left(\hat{\theta}_n(\boldsymbol{x});\boldsymbol{x}\right)=0$. A second order Taylor expansion gives that

$$\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)l''(\theta_0;\boldsymbol{X})=-l'(\theta_0;\boldsymbol{X})-R_1(\theta_0;\boldsymbol{X})$$
$$\Rightarrow -\alpha\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)=-l'(\theta_0;\boldsymbol{X})-R_1(\theta_0;\boldsymbol{X})-\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\left(l''(\theta_0;\boldsymbol{X})+\alpha\right),$$

where

$$R_1(\theta_0;\boldsymbol{X})=\frac{1}{2}\left(\hat{\theta}_n(\boldsymbol{x})-\theta_0\right)^2 l^{(3)}(\theta^*;\boldsymbol{x})$$

is the remainder term with θ^* lying between $\hat{\theta}_n(\boldsymbol{x})$ and θ_0. Multiplying both sides by $-\frac{\sqrt{n\,i_2(\theta_0)}}{\alpha}$,

$$\sqrt{n\,i_2(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)=\frac{\sqrt{n\,i_2(\theta_0)}}{\alpha}\left[l'(\theta_0;\boldsymbol{X})+R_1(\theta_0;\boldsymbol{X})+\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\left(l''(\theta_0;\boldsymbol{X})+\alpha\right)\right]. \tag{8}$$

Applying the triangle inequality,

$$\left|\mathrm{E}\left[h\left(\sqrt{n\,i_2(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\right)\right]-\mathrm{E}[h(Z)]\right|$$
$$\leq\left|\mathrm{E}\left[h\left(\frac{\sqrt{n\,i_2(\theta_0)}l'(\theta_0;\boldsymbol{X})}{\alpha}\right)\right]-\mathrm{E}[h(Z)]\right| \tag{9}$$
$$+\left|\mathrm{E}\left[h\left(\sqrt{n\,i_2(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\right)-h\left(\frac{\sqrt{n\,i_2(\theta_0)}l'(\theta_0;\boldsymbol{X})}{\alpha}\right)\right]\right|. \tag{10}$$

Step 1: Bound for (9). Let, for ease of presentation $l'(\theta_0;\boldsymbol{X})=\sum_{i=1}^{n}\tilde{\xi}_i$, where

$$\tilde{\xi}_1=\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta)\Big|_{\theta=\theta_0},\qquad \tilde{\xi}_i=\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_i|X_{i-1},\ldots,X_{i-m};\theta)\Big|_{\theta=\theta_0}\quad\text{for }\ i=2,3,\ldots,n.$$

Assumption (A.D.2) ensures that $\tilde{\xi}_i$, $i=1,2,\ldots,n$ have mean zero. Furthermore, for some function $g:\mathbb{R}^{m+1}\to\mathbb{R}$, it holds that $\tilde{\xi}_i=g(X_i,X_{i-1},\ldots,X_{i-m})$ and taking into account that $\{X_i,i=1,2,\ldots,n\}$ is an m-dependent sequence, we conclude that $\left\{\tilde{\xi}_i,i=1,2,\ldots,n\right\}$ forms a $2m$-dependent sequence. Define now

$$W:=\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{\mathrm{Var}\left[l'(\theta_0;\boldsymbol{X})\right]}}=\sum_{i=1}^{n}\left(\frac{\xi_i}{\sqrt{n}}\right), \tag{11}$$

with

$$\xi_i=\tilde{\xi}_i\sqrt{\frac{n}{\mathrm{Var}\left[l'(\theta_0;\boldsymbol{X})\right]}},\quad\forall i\in\{1,2,\ldots,n\}.$$

It follows that $\left\{\frac{\xi_i}{\sqrt{n}},i=1,2,\ldots,n\right\}$ is a random $2m$-dependent sequence with mean zero and also $\mathrm{Var}(W)=1$. In addition, (LD) is satisfied with A_j and B_j as in (4). A simple triangle inequality gives that

$$(9)\leq|\mathrm{E}[h(W)]-\mathrm{E}[h(Z)]| \tag{12}$$
$$+\left|\mathrm{E}\left[h\left(\frac{\sqrt{n\,i_2(\theta_0)}l'(\theta_0;\boldsymbol{X})}{\alpha}\right)-h(W)\right]\right|. \tag{13}$$

Since the assumptions of Lemma 1.1 are satisfied for W as in (11), one can directly use (3) in order to find an upper bound for (12). For (13), a first order Taylor expansion of $h\left(\frac{\sqrt{n\,i_2(\theta_0)}l'(\theta_0;\boldsymbol{X})}{\alpha}\right)$ about W yields

$$\begin{aligned}&h\left(\frac{\sqrt{n\,i_2(\theta_0)}l'(\theta_0;\boldsymbol{X})}{\alpha}\right)-h\left(\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}\right)\\&\quad=\left(\frac{\sqrt{n\,i_2(\theta_0)}l'(\theta_0;\boldsymbol{X})}{\alpha}-\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}\right)h'(t_1(\boldsymbol{X})),\end{aligned}$$

where $t_1(\boldsymbol{X})$ is between $\frac{\sqrt{n\,i_2(\theta_0)}l'(\theta_0;\boldsymbol{X})}{\alpha}$ and $\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}$. Therefore,

$$(13)\le\|h'\|\left|\frac{\sqrt{n\,i_2(\theta_0)}}{\alpha}-\frac{1}{\sqrt{\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}\right|\mathrm{E}\left|l'(\theta_0;\boldsymbol{X})\right|\le\|h'\|\left|\frac{\sqrt{n\,i_2(\theta_0)\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}{\alpha}-1\right|. \tag{14}$$

For $h\in H_W$ as in (1), then $\|h'\|\le 1$, which yields

$$\begin{aligned}(9)\le{}&\frac{2}{n^{\frac{3}{2}}}\left[\sum_{i=1}^{n}(\mathrm{E}|\xi_i\eta_i\tau_i|)+\sum_{i=1}^{n}(|\mathrm{E}(\xi_i\eta_i)|\,\mathrm{E}|\tau_i|)\right]+\frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\mathrm{E}\left|\xi_i\eta_i^2\right|\\&+\left|\frac{\sqrt{n\,i_2(\theta_0)\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}{\alpha}-1\right|,\end{aligned} \tag{15}$$

with η_i and τ_i as in (2). The absolute expectations in (15) can be difficult to bound and the first three quantities of the above bound are therefore expressed in terms of more easily calculable terms. For the first term in (15), using Hölder's inequality

$$\begin{aligned}\mathrm{E}|\xi_i\eta_i\tau_i|&=\mathrm{E}\left|\xi_i\sum_{j\in A_i}\xi_j\sum_{k\in B_i}\xi_k\right|\le\sum_{j\in A_i}\sum_{k\in B_i}\mathrm{E}\left|\xi_i\xi_j\xi_k\right|\le\sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left(|\xi_i|^3\right)\mathrm{E}\left(\left|\xi_j\right|^3\right)\mathrm{E}\left(|\xi_k|^3\right)\right]^{\frac{1}{3}}\\&\le\sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left((\xi_i)^4\right)\mathrm{E}\left(\left(\xi_j\right)^4\right)\mathrm{E}\left((\xi_k)^4\right)\right]^{\frac{1}{4}}.\end{aligned} \tag{16}$$

For the second term of the bound in (15), the Cauchy–Schwarz inequality yields

$$\begin{aligned}|\mathrm{E}\left(\xi_i\eta_i\right)|\,\mathrm{E}\left|\tau_i\right|&=\left|\mathrm{E}\left(\xi_i\sum_{j\in A_i}\xi_j\right)\right|\mathrm{E}\left|\sum_{k\in B_i}\xi_k\right|\le\sum_{j\in A_i}\mathrm{E}\left|\xi_i\xi_j\right|\sum_{k\in B_i}\mathrm{E}\left|\xi_k\right|\\&\le\sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left((\xi_i)^2\right)\mathrm{E}\left(\left(\xi_j\right)^2\right)\mathrm{E}\left((\xi_k)^2\right)\right]^{\frac{1}{2}}.\end{aligned} \tag{17}$$

For the third term, Jensen's inequality is employed to get that

$$\left(\sum_{i\in J}|a_i|\right)^z\le J^{z-1}\sum_{i\in J}|a_i|^z,\quad\forall a_i\in\mathbb{R}\ \text{ and }\ z\in\mathbb{N}$$

and therefore

$$\begin{aligned}\mathrm{E}\left|\xi_i\eta_i^2\right|&=\mathrm{E}\left|\xi_i\left(\sum_{j\in A_i}\xi_j\right)^2\right|\le|A_i|\,\mathrm{E}\left|\xi_i\sum_{j\in A_i}\xi_j^2\right|\le|A_i|\sum_{j\in A_i}\mathrm{E}\left|\xi_i\xi_j^2\right|\\&\le|A_i|\sum_{j\in A_i}\left[\mathrm{E}\left((\xi_i)^2\right)\mathrm{E}\left(\left(\xi_j\right)^4\right)\right]^{\frac{1}{2}}.\end{aligned} \tag{18}$$

The results in (16), (17) and (18) yield

$$
\begin{aligned}
(12) &\leq \frac{2}{n^{\frac{3}{2}}}\left[\sum_{i=1}^{n}\left(\mathrm{E}|\xi_i\eta_i\tau_i|\right) + \sum_{i=1}^{n}\left(|\mathrm{E}(\xi_i\eta_i)|\,\mathrm{E}|\tau_i|\right)\right] + \frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\mathrm{E}\left|\xi_i\eta_i^2\right| \\
&\leq \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left((\xi_i)^4\right)\mathrm{E}\left((\xi_j^4)\right)\mathrm{E}\left((\xi_k)^4\right)\right]^{\frac{1}{4}} \\
&\quad + \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left((\xi_i)^2\right)\mathrm{E}\left(\left(\xi_j\right)^2\right)\mathrm{E}\left((\xi_k)^2\right)\right]^{\frac{1}{2}} + \frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}|A_i|\sum_{j\in A_i}\left[\mathrm{E}\left((\xi_i)^2\right)\mathrm{E}\left(\left(\xi_j\right)^4\right)\right]^{\frac{1}{2}}.
\end{aligned}
\tag{19}
$$

The bound for (12) is now obviously a function only of $\mathrm{E}\left(\xi_i^2\right)$ and $\mathrm{E}\left(\xi_i^4\right)$.

Step 2: Bound for (10). The main tool used here is Taylor expansions. For ease of presentation, let

$$
\begin{aligned}
\tilde{C}(\theta_0) = \tilde{C}(h,\theta_0;\boldsymbol{X}) &:= h\left(\sqrt{n\,i_2(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\right) - h\left(\frac{\sqrt{n\,i_2(\theta_0)}\,l'(\theta_0;\boldsymbol{X})}{\alpha}\right) \\
&= h\left(\frac{\sqrt{n\,i_2(\theta_0)}\left[l'(\theta_0;\boldsymbol{X}) + R_1(\theta_0;\boldsymbol{X}) + \left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\left(l''(\theta_0;\boldsymbol{X})+\alpha\right)\right]}{\alpha}\right) \\
&\quad - h\left(\frac{\sqrt{n\,i_2(\theta_0)}\,l'(\theta_0;\boldsymbol{X})}{\alpha}\right)
\end{aligned}
$$

using (8). A first order Taylor expansion of $h\left(\frac{\sqrt{n\,i_2(\theta_0)}\left[l'(\theta_0;\boldsymbol{X})+R_1(\theta_0;\boldsymbol{X})+\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\left(l''(\theta_0;\boldsymbol{X})+\alpha\right)\right]}{\alpha}\right)$ about $\frac{\sqrt{n\,i_2(\theta_0)}\,l'(\theta_0;\boldsymbol{X})}{\alpha}$ yields

$$
\begin{aligned}
(10) = \left|\mathrm{E}\left[\tilde{C}(\theta_0)\right]\right| &\leq \frac{\sqrt{n\,i_2(\theta_0)}}{\alpha}\|h'\|\left(\mathrm{E}\left[\frac{1}{2}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\left|l^{(3)}(\theta^*;\boldsymbol{X})\right|\right] + \mathrm{E}\left|\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\left(l''(\theta_0;\boldsymbol{X})+\alpha\right)\right|\right) \\
&\leq \frac{\sqrt{n\,i_2(\theta_0)}}{\alpha}\|h'\|\left(\frac{S_d(n)}{2}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right] + \sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]}\sqrt{\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X})+\alpha\right)^2\right]}\right),
\end{aligned}
\tag{20}
$$

where for the last step Cauchy–Schwarz inequality has been used while $S_d(n)$ is as in (5). We conclude that (14), (19) and (20) yield, for $h \in H_W$, the assertion of the theorem as expressed in (7). ∎

The following corollary specifies the result of Theorem 2.1 for the simple scenario of i.i.d. random variables. This allows for a comparison with the bound given in Anastasiou and Reinert (2017), which is for i.i.d. random variables. The proof of the corollary is a result of simple steps and therefore only an outline is provided.

Corollary 2.1. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with probability density (or mass) function $f(x|\theta)$. Assume that $\hat{\theta}_n(\boldsymbol{X})$ exists and is unique and that* (A.D.1)–(A.D.4) *hold. In addition,* $\mathrm{Var}\left[l'(\theta_0;\boldsymbol{X})\right] > 0$. *For α as in* (6) *and $Z \sim \mathrm{N}(0,1)$,*

$$
\begin{aligned}
d_W\left(\sqrt{n\,i_2(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right), Z\right) &\leq \frac{5\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f\left(X_1|\theta_0\right)\right|^3}{\sqrt{n}\left[\mathrm{Var}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log f\left(X_1|\theta_0\right)\right)\right]^{\frac{3}{2}}} \\
&\quad + \left|\frac{n\sqrt{i_2(\theta_0)\mathrm{Var}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log f\left(X_1|\theta_0\right)\right)}}{\alpha} - 1\right| + \frac{S_d(n)\sqrt{n\,i_2(\theta_0)}}{2\alpha}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right] \\
&\quad + \frac{\sqrt{n\,i_2(\theta_0)}}{\alpha}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]}\sqrt{\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X})+\alpha\right)^2\right]}.
\end{aligned}
\tag{21}
$$

Outline of the proof. A similar process as the one followed in the proof of Theorem 2.1 shows that a bound is obtained by bounding the terms (12), (13) and (10). For independent random variables, applying Hölder's inequality to the bound in (3), where now

$$
W = \sum_{i=1}^{n}\left(\frac{\xi_i}{\sqrt{n}}\right), \qquad \xi_i = \frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_i|\theta)\Big|_{\theta=\theta_0}\sqrt{\frac{n}{\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}},
$$

leads to

$$
(12) \leq \frac{5}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\mathrm{E}|\xi_i|^3 = \frac{5\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3}{\sqrt{n}\left[\mathrm{Var}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right)\right]^{\frac{3}{2}}}.
$$

The second term of the bound in (21) is the special form of (14) for the case of i.i.d. random variables, while the last two terms are as in the result of Theorem 2.1.

Remark 2.1. The bound in (21) is not as simple and sharp as the bound given in Theorem 2.1 of Anastasiou and Reinert (2017). This is expected since Corollary 2.1 is a special application of a result which was originally obtained to satisfy the assumption of local dependence for our random variables, while Anastasiou and Reinert (2017) used directly results of Stein's method for independent random variables. In addition, the assumption (A.D.1) used for the result of Corollary 2.1 is stronger than the condition (R3) of Anastasiou and Reinert (2017). Using uniform boundedness of the third derivative of the log-likelihood function in (A.D.1) allows us to get bounds on the Wasserstein distance related to the MLE. On the other hand, Anastasiou and Reinert (2017) relaxed this condition and assumed that the third derivative of the log-likelihood function is bounded in an area of θ_0. This leads to bounds on the bounded Wasserstein (or Fortet–Mourier) distance; see Nourdin and Peccati (2012) for a definition of this metric.

3. Example: 1-dependent normal random variables

To illustrate the general results, as an example assume that we have a sequence $\{S_1, S_2, \ldots, S_n\}$ of random variables where for $k \in \mathbb{Z}^+$ and $\forall j \in \{1, 2, \ldots, n\}$,

$$S_j = \sum_{i=(j-1)k}^{jk} X_i,$$

for $X_i, i = 0, 1, 2, \ldots, nk$ i.i.d. random variables from the $\mathrm{N}(\mu, \sigma^2)$ distribution with $\mu = \theta \in \mathbb{R}$ being the unknown parameter and σ^2 is known. Hence S_j and S_{j+1} share one summand, X_{jk}. For $\delta \in \mathbb{Z} \setminus \{0\}$, we have that

$$\mathrm{Cov}(S_i, S_{i+\delta}) = \begin{cases} \mathrm{Var}(X_1) = \sigma^2, & \text{if } |\delta| = 1 \\ 0, & \text{if } |\delta| > 1. \end{cases}$$

Therefore, $\{S_i\}_{i=1,2,\ldots,n}$ is a 1-dependent sequence of random variables. Furthermore,

$$S_i \sim \mathrm{N}((k+1)\theta, (k+1)\sigma^2) \tag{22}$$

as it is a sum of $k+1$ independent normally distributed random variables with mean θ and variance σ^2. As

$$\rho = \frac{\mathrm{Cov}(S_{i-1}, S_i)}{\sqrt{\mathrm{Var}(S_{i-1})\mathrm{Var}(S_i)}} = \frac{\sigma^2}{(k+1)\sigma^2} = \frac{1}{k+1}, \ \forall i \in \{2, 3, \ldots, n\},$$

it is standard, see Casella and Berger (2002, p. 177), that for $i = 2, 3, \ldots, n$

$$(S_i | S_{i-1} = s_{i-1}) \sim \mathrm{N}\left((k+1)\theta + \frac{1}{k+1}(s_{i-1} - (k+1)\theta), \frac{k(k+2)}{k+1}\sigma^2\right). \tag{23}$$

After basic steps, the likelihood function for the parameter θ under $\boldsymbol{S} = (S_1, S_2, \ldots, S_n)$ is

$$\begin{aligned} L(\theta; \boldsymbol{S}) &= f(S_1|\theta) \prod_{i=2}^{n} f(S_i | S_{i-1}; \theta) \\ &= \frac{(k+1)^{\frac{n-1}{2}}}{\sqrt{2\pi(k+1)\sigma^2}(2\pi k(k+2)\sigma^2)^{\frac{n-1}{2}}} \exp\left\{ -\frac{(S_1 - (k+1)\theta)^2}{2(k+1)\sigma^2} \right. \\ &\quad \left. - \frac{k+1}{2k(k+2)\sigma^2} \sum_{i=2}^{n} \left(S_i - \left((k+1)\theta + \frac{1}{k+1}(S_{i-1} - (k+1)\theta) \right) \right)^2 \right\}. \end{aligned}$$

Having this closed-form expression for the likelihood allows us to derive the MLE under this local dependence structure. The unique MLE for θ is

$$\hat{\theta}_n(\boldsymbol{S}) = \frac{k\sum_{i=1}^{n} S_i + S_1 + S_n}{(nk+2)(k+1)}. \tag{24}$$

In addition, the first three derivatives of the log-likelihood function are

$$\begin{aligned} l'(\theta; \boldsymbol{S}) &= \frac{1}{(k+2)\sigma^2} \left\{ k\sum_{i=1}^{n} S_i + S_1 + S_n - (k+1)(nk+2)\theta \right\} \\ l''(\theta; \boldsymbol{S}) &= -\frac{(nk+2)(k+1)}{(k+2)\sigma^2} \\ l^{(3)}(\theta; \boldsymbol{S}) &= 0. \end{aligned} \tag{25}$$

The following corollary gives the upper bound on the Wasserstein distance between the distribution of $\hat{\theta}_n(\boldsymbol{S})$ and the normal distribution.

Corollary 3.1. *Let $S_1, S_2, \ldots, S_n$ be a 1-dependent sequence of random variables with $S_i \sim \mathrm{N}((k+1)\theta, (k+1)\sigma^2)$. The conditions (A.D.1)–(A.D.4) hold. For $Z \sim \mathrm{N}(0,1)$ and $i_2(\theta_0) = \frac{(k+1)^2}{(k+3)\sigma^2}$,*

$$d_W\left(\sqrt{n\, i_2(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{S}) - \theta_0\right), Z\right) \leq 339(n-5)\left[\frac{k(k+1)(k+2)}{(nk^3+(3n+2)k^2+10k+2)}\right]^{\frac{3}{2}}$$
$$+ \frac{(k+1)^{\frac{3}{2}}(k+2)^{\frac{3}{2}}}{(nk^3+(3n+2)k^2+10k+2)^{\frac{3}{2}}}\left\{\left(1+3^{\frac{3}{4}}\right)\left(2\sqrt{k+2}(37k+2)+4\sqrt{k}(61k+8)\right)\right.$$
$$\left.+\sqrt{3}\left(3\sqrt{k+2}(k+1)+\sqrt{k}(91k+18)\right)\right\}$$
$$+\left|\left(1-\frac{2}{nk+2}\right)\left[\frac{k+3+\frac{2}{n}+\frac{10}{nk}+\frac{2}{nk^2}}{k+3}\right]^{\frac{1}{2}} - 1\right|.$$

Remark 3.1. The order of the bound with respect to the sample size is $\frac{1}{\sqrt{n}}$.

Proof. We first check that the assumptions (A.D.1)–(A.D.4) are satisfied. The first assumption is satisfied from (25) with $S_d(n) = 0$. From (22) and (23), simple steps yield $\mathrm{E}\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(S_1|\theta)\right] = \mathrm{E}\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(S_i|S_{i-1};\theta)\right] = 0$ and thus (A.D.2) holds. The (A.D.3) is also satisfied since, using (22) and (24),

$$\mathrm{E}\left[\hat{\theta}_n(\boldsymbol{S})\right] = \frac{nk(k+1)\theta_0 + 2(k+1)\theta_0}{(nk+2)(k+1)} = \theta_0.$$

To show that (A.D.4) holds, we first calculate

$$\begin{aligned}\mathrm{Var}\left[\hat{\theta}_n(\boldsymbol{S})\right] &= \frac{1}{(nk+2)^2(k+1)^2}\mathrm{Var}\left(k\sum_{i=1}^n S_i + S_1 + S_n\right)\\ &= \frac{1}{(nk+2)^2(k+1)^2}\left\{k^2\mathrm{Var}\left(\sum_{i=1}^n S_i\right) + \mathrm{Var}(S_1) + \mathrm{Var}(S_n) + 2k\mathrm{Cov}\left(S_1, \sum_{i=1}^n S_i\right)\right.\\ &\qquad \left. + 2k\mathrm{Cov}\left(S_n, \sum_{i=1}^n S_i\right)\right\}. \end{aligned} \tag{26}$$

From (22), $\mathrm{Var}(S_i) = (k+1)\sigma^2, \forall i \in \{1,2,\ldots,n\}$. In addition, since $\{S_i\}_{i=1,2,\ldots,n}$ is a 1-dependent sequence of random variables,

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^n S_i\right) &= n\mathrm{Var}(S_1) + 2(n-1)\mathrm{Cov}(S_1,S_2) = n(k+1)\sigma^2 + 2(n-1)\sigma^2\\ \mathrm{Cov}\left(S_1, \sum_{i=1}^n S_i\right) &= \mathrm{Var}(S_1) + \mathrm{Cov}(S_1,S_2) = (k+2)\sigma^2. \end{aligned} \tag{27}$$

Applying the above results of (27) to (26) gives that

$$\mathrm{Var}\left[\hat{\theta}_n(\boldsymbol{S})\right] = \frac{\sigma^2\left(nk^3+3nk^2+2k^2+10k+2\right)}{(nk+2)^2(k+1)^2}. \tag{28}$$

Therefore,

$$\begin{aligned} i_2(\theta_0) &= \lim_{n\to\infty}\frac{1}{n\mathrm{Var}\left(\hat{\theta}_n(\boldsymbol{S})\right)} = \lim_{n\to\infty}\frac{(nk+2)^2(k+1)^2}{n\sigma^2(nk^3+3nk^2+2k^2+10k+2)}\\ &= \lim_{n\to\infty}\frac{n^2(k+1)^2\left(k^2+\frac{4k}{n}+\frac{4}{n^2}\right)}{n^2\sigma^2\left(k^3+3k^2+\frac{2k^2}{n}+\frac{10k}{n}+\frac{2}{n}\right)} = \frac{(k+1)^2}{(k+3)\sigma^2} > 0, \end{aligned} \tag{29}$$

which shows that (A.D.4) is satisfied. To obtain α as defined in (6), the variance of the score function is calculated which, after simple steps and using (27), is

$$\mathrm{Var}\left[l'(\theta_0;\mathbf{S})\right] = \frac{1}{(k+2)^2\sigma^2}\left[nk^3+3nk^2+2k^2+10k+2\right]. \tag{30}$$

The above result and (28) yield

$$\alpha = \sqrt{\frac{\mathrm{Var}\left[l'(\theta_0;\mathbf{S})\right]}{\mathrm{Var}\left[\hat{\theta}_n(\mathbf{S})\right]}} = \sqrt{\frac{(nk+2)^2(k+1)^2}{\sigma^4(k+2)^2}} = \frac{(nk+2)(k+1)}{(k+2)\sigma^2}. \tag{31}$$

For $\xi_1 = \frac{\mathrm{d}}{\mathrm{d}\theta}\log f(S_1|\theta)\Big|_{\theta=\theta_0}\sqrt{\frac{n}{\mathrm{Var}\left[l'(\theta_0;\mathbf{S})\right]}}$, $\xi_i = \frac{\mathrm{d}}{\mathrm{d}\theta}\log f(S_i|S_{i-1};\theta)\Big|_{\theta=\theta_0}\sqrt{\frac{n}{\mathrm{Var}\left[l'(\theta_0;\mathbf{S})\right]}}$, $i=2,3,\ldots,n$, using (30) and (22), we get that

$$\xi_1 = \frac{\sqrt{n}(k+2)[S_1-(k+1)\theta_0]}{\sigma\sqrt{nk^3+(3n+2)k^2+10k+2}}$$

and therefore

$$\begin{aligned}
\mathrm{E}\left(\xi_1^2\right) &= \frac{n(k+2)^2\mathrm{E}(S_1-(k+1)\theta)^2}{(nk^3+(3n+2)k^2+10k+2)\sigma^2} = \frac{n(k+2)^2(k+1)}{nk^3+(3n+2)k^2+10k+2}\\
\mathrm{E}\left(\xi_1^4\right) &= \frac{n^2(k+2)^4\mathrm{E}(S_1-(k+1)\theta)^4}{(nk^3+(3n+2)k^2+10k+2)^2\sigma^4} = \frac{3n^2(k+2)^4(k+1)^2}{(nk^3+(3n+2)k^2+10k+2)^2}.
\end{aligned} \tag{32}$$

Furthermore, for $i=2,3,\ldots,n$, the results in (30) and (23) yield

$$\xi_i = \frac{\sqrt{n}(k+1)\left[S_i-\left((k+1)\theta+\frac{1}{k+1}\left(S_{i-1}-(k+1)\theta\right)\right)\right]}{\sigma\sqrt{nk^3+(3n+2)k^2+10k+2}}$$

so that

$$\begin{aligned}
\mathrm{E}\left(\xi_i^2\right) &= \frac{n(k+1)^2\mathrm{E}\left[S_i-\left((k+1)\theta+\frac{1}{k+1}\left(S_{i-1}-(k+1)\theta\right)\right)\right]^2}{\sigma^2(nk^3+(3n+2)k^2+10k+2)} = \frac{nk(k+1)(k+2)}{nk^3+(3n+2)k^2+10k+2}\\
\mathrm{E}\left(\xi_i^4\right) &= \frac{n^2(k+1)^4\mathrm{E}\left[S_i-\left((k+1)\theta+\frac{1}{k+1}\left(S_{i-1}-(k+1)\theta\right)\right)\right]^4}{\sigma^4(nk^3+(3n+2)k^2+10k+2)^2}\\
&= \frac{3n^2k^2(k+1)^2(k+2)^2}{(nk^3+(3n+2)k^2+10k+2)^2}.
\end{aligned} \tag{33}$$

The first three terms of the general bound (7) are now calculated. These are denoted from now on by

$$\begin{aligned}
Q_v = Q_v(k,n) &:= \frac{1}{n^{\frac{3}{2}}}\left\{2\sum_{j\in A_v}\sum_{l\in B_v}\left[\mathrm{E}\left((\xi_v)^2\right)\mathrm{E}\left((\xi_j)^2\right)\mathrm{E}\left((\xi_l)^2\right)\right]^{\frac{1}{2}}\right.\\
&\left.+2\sum_{j\in A_v}\sum_{l\in B_v}\left[\mathrm{E}\left((\xi_v)^4\right)\mathrm{E}\left((\xi_j)^4\right)\mathrm{E}\left((\xi_l)^4\right)\right]^{\frac{1}{4}} + |A_v|\sum_{j\in A_v}\left[\mathrm{E}\left((\xi_v)^2\right)\mathrm{E}\left((\xi_j)^4\right)\right]^{\frac{1}{2}}\right\}.
\end{aligned}$$

Our approach is split depending on whether 1 is an element of either A_i or B_i as defined in (4) for $i\in\{1,2,\ldots,n\}$.
Case 1: $i=6,7,\ldots,n$. Using the results in (33) and since $|A_i|\le 5$, $|B_i|\le 9$, $\forall i\in\{6,7,\ldots,n\}$,

$$\begin{aligned}
Q_i &\le \frac{1}{n^{\frac{3}{2}}}\left\{90\left[\mathrm{E}\left((\xi_2)^4\right)\right]^{\frac{3}{4}} + 90\left[\mathrm{E}\left((\xi_2)^2\right)\right]^{\frac{3}{2}} + 25\left[\mathrm{E}\left((\xi_2)^2\right)\mathrm{E}\left((\xi_2)^4\right)\right]^{\frac{1}{2}}\right\}\\
&= \left[\frac{k(k+1)(k+2)}{(nk^3+(3n+2)k^2+10k+2)}\right]^{\frac{3}{2}}\left(90(3)^{\frac{3}{4}}+90+25\sqrt{3}\right)\\
&< 339\left[\frac{k(k+1)(k+2)}{(nk^3+(3n+2)k^2+10k+2)}\right]^{\frac{3}{2}}.
\end{aligned} \tag{34}$$

Issues arise due to ξ_1 not having the same distribution as ξ_i for $i\in\{2,3,\ldots,n\}$. There are hence five more special cases corresponding to $i=1,2,\ldots,5$. These cases are treated separately.

Case 2: $i = 1$. For $A_1 = \{1, 2, 3\}$ and $B_1 = \{1, 2, \dots, 5\}$, the results in (32) and (33) yield

$$Q_1 = \frac{1}{n^{\frac{3}{2}}}\Bigg\{2\left[\left[\mathrm{E}\left((\xi_1)^2\right)\right]^{\frac{3}{2}} + 6\mathrm{E}\left((\xi_1)^2\right)\left[\mathrm{E}\left((\xi_2)^2\right)\right]^{\frac{1}{2}} + 8\mathrm{E}\left((\xi_2)^2\right)\left[\mathrm{E}\left((\xi_1)^2\right)\right]^{\frac{1}{2}}\right]$$
$$+2\left[\left[\mathrm{E}\left((\xi_1)^4\right)\right]^{\frac{3}{4}} + 6\left[\mathrm{E}\left((\xi_1)^4\right)\right]^{\frac{1}{2}}\left[\mathrm{E}\left((\xi_2)^4\right)\right]^{\frac{1}{4}} + 8\left[\mathrm{E}\left((\xi_1)^4\right)\right]^{\frac{1}{4}}\left[\mathrm{E}\left((\xi_2)^4\right)\right]^{\frac{1}{2}}\right]$$
$$+3\left[\left[\mathrm{E}\left((\xi_1)^2\right)\right]^{\frac{1}{2}}\left(\left[\mathrm{E}\left((\xi_1)^4\right)\right]^{\frac{1}{2}} + 2\left[\mathrm{E}\left((\xi_2)^4\right)\right]^{\frac{1}{2}}\right)\right]\Bigg\}$$
$$= \frac{2(k+1)^{\frac{3}{2}}(k+2)^2}{(nk^3+(3n+2)k^2+10k+2)^{\frac{3}{2}}}\left\{\left(9k+2+6\sqrt{k(k+2)}\right)\left(1+3^{\frac{3}{4}}\right)+3\sqrt{3}(k+1)\right\}. \tag{35}$$

Case 3: $i = 2$. For $A_2 = \{1, 2, 3, 4\}$ and $B_2 = \{1, 2, \dots, 6\}$, a similar approach as the one in Case 2 yields

$$Q_2 = \frac{4\sqrt{k}(k+1)^{\frac{3}{2}}(k+2)^{\frac{3}{2}}}{(nk^3+(3n+2)k^2+10k+2)^{\frac{3}{2}}}\left\{\left(8k+1+4\sqrt{k(k+2)}\right)\left(1+3^{\frac{3}{4}}\right)+2\sqrt{3}(2k+1)\right\}. \tag{36}$$

Case 4: $i = 3$. Following the same steps as in Case 3, now for $A_3 = \{1, 2, \dots, 5\}$ and $B_3 = \{1, 2, \dots, 7\}$, the results in (32) and (33) give that

$$Q_3 = \frac{\sqrt{k}(k+1)^{\frac{3}{2}}(k+2)^{\frac{3}{2}}}{(nk^3+(3n+2)k^2+10k+2)^{\frac{3}{2}}}\left\{2\left(25k+2+10\sqrt{k(k+2)}\right)\left(1+3^{\frac{3}{4}}\right)+5\sqrt{3}(5k+2)\right\}. \tag{37}$$

Case 5: $i = 4$. In this case, $A_4 = \{2, 3, \dots, 6\}$, $B_4 = \{1, 2, \dots, 8\}$, which lead to

$$Q_4 = \frac{5k(k+1)^{\frac{3}{2}}(k+2)^{\frac{3}{2}}}{(nk^3+(3n+2)k^2+10k+2)^{\frac{3}{2}}}\left\{2\left[\sqrt{k+2}+7\sqrt{k}\right]\left(1+3^{\frac{3}{4}}\right)+5\sqrt{3k}\right\}. \tag{38}$$

Case 6: $i = 5$. Now $A_5 = \{3, 4, \dots, 7\}$ and $B_5 = \{1, 2, \dots, 9\}$ to obtain that

$$Q_5 = \frac{5k(k+1)^{\frac{3}{2}}(k+2)^{\frac{3}{2}}}{(nk^3+(3n+2)k^2+10k+2)^{\frac{3}{2}}}\left\{2\left[\sqrt{k+2}+8\sqrt{k}\right]\left(1+3^{\frac{3}{4}}\right)+5\sqrt{3k}\right\}. \tag{39}$$

The sum of the results of (35), (36), (37), (38) and (39) with $(n-5)$ times the bound in (34) consists an upper bound for the first three terms of the general upper bound as expressed in (7). For the fourth term of the general upper bound, (29), (30) and (31) yield

$$\left|\frac{\sqrt{n\,i_2(\theta_0)\mathrm{Var}[l'(\theta_0;\boldsymbol{X})]}}{\alpha}-1\right| = \left|\frac{nk}{nk+2}\left[\frac{k+3+\frac{2}{n}+\frac{10}{nk}+\frac{2}{nk^2}}{k+3}\right]^{\frac{1}{2}}-1\right|$$
$$= \left|\left(1-\frac{2}{nk+2}\right)\left[\frac{k+3+\frac{2}{n}+\frac{10}{nk}+\frac{2}{nk^2}}{k+3}\right]^{\frac{1}{2}}-1\right|. \tag{40}$$

The fifth term of the bound in (7) involves the calculation of $S_d(n)$, which is equal to zero from (25). Therefore, the fifth term of the general upper bound vanishes for this example. For the last term we have from (25) that $\mathrm{E}\left[l''(\theta_0;\boldsymbol{S})\right] = -\frac{(nk+2)(k+1)}{(k+2)\sigma^2} = -\alpha$ and therefore

$$\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{S})-\theta_0\right)^2\right]\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{S})+\alpha\right)^2\right]} = \sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{S})-\theta_0\right)^2\right]\mathrm{Var}\left[l''(\theta_0;\boldsymbol{S})\right]} = 0.$$

The results of Case 1–Case 6 and (40) give the assertion of the corollary. ■

Remarks. Several exciting paths lead from the work explained in this paper. Firstly, treating the case of a vector parameter is the next reasonable step to go. Furthermore, other types of dependence structure between the random variables (or vectors) could be investigated to get bounds for the distributional distance of interest. In addition, our theoretical results can be very useful when it comes to applications that satisfy the assumed dependence structure for the data.

Acknowledgements

This research occurred whilst Andreas Anastasiou was studying for a D.Phil. at the University of Oxford, supported by a Teaching Assistantship Bursary from the Department of Statistics, University of Oxford, and the Engineering and Physical Sciences Research Council (EPSRC) grant EP/K503113/1. The author would like to thank Gesine Reinert for insightful comments and suggestions.

References

Anastasiou, A., 2016. Assessing the multivariate normal approximation of the maximum likelihood estimator from high-dimensional, heterogeneous data. ArXiv:1510.03679.

Anastasiou, A., Ley, C., 2017. Bounds for the asymptotic normality of the maximum likelihood estimator using the delta method. ALEA, Lat. Am. J. Probab. Math. Stat. 14, 153–171.

Anastasiou, A., Reinert, G., 2017. Bounds for the normal approximation of the maximum likelihood estimator. Bernoulli 23 (1), 191–218.

Berk, K.N., 1973. A Central Limit Theorem for m-dependent random variables with unbounded m. Ann. Probab. 1 (2), 352–354.

Billingsley, P., 1961. Statistical methods in Markov chains. Ann. Math. Stat. 32, 12–40.

Casella, G., Berger, R.L., 2002. Statistical Inference, second ed. Brooks/Cole, Cengage Learning, Duxbury, Pacific Grove.

Chen, L.H.Y., Goldstein, L., Shao, Q.M., 2011. Normal Approximation by Stein's Method. Springer-Verlag, Berlin, Heidelberg.

Erickson, R.V., 1974. L_1 bounds for asymptotic normality of m-dependent sums using Stein's technique. Ann. Probab. 2, 522–529.

Fisher, R.A., 1925. Theory of statistical estimation. Math. Proc. Cambridge Philos. Soc. 22, 700–725.

Heinrich, L., 1982. A method for the derivation of limit theorems for sums of m-dependent random variables. Z. Wahrscheinlichkeitstheor. Verwandte Geb. 60, 501–515.

Mäkeläinen, T., Schmidt, K., Styan, G.P.H., 1981. On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples. Ann. Statist. 9, 758–767.

Nourdin, I., Peccati, G., 2012. Normal Approximations with Malliavin Calculus. From Stein's Method to Universality. In: Cambridge Tracts in Mathematics, no. 192, Cambridge University Press.

Orey, S.A., 1958. Central Limit Theorem for m-dependent random variables. Duke Math. J. 25, 543–546.

Pinelis, I., Molzon, R., 2016. Optimal-order bounds on the rate of convergence to normality in the multivariate delta method. Electron. J. Stat. 10, 1001–1063.

Stein, C., 1972. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 2. University of California Press, Berkeley, pp. 586–602.

Stein, C., 1986. Approximate Computation of Expectations. In: Lecture Notes-Monograph Series, Institute of Mathematical Statistics, Hayward, California.