ALEA, Lat. Am. J. Probab. Math. Stat. **22**, 537–555 (2025)
DOI: 10.30757/ALEA.v22-20

Berry–Esseen Theorem for Sample Quantiles with Locally Dependent Data

Partha S. Dey and Grigory Terlov

Department of Mathematics, University of Illinois at Urbana-Champaign, IL, USA
E-mail address: psdey@illinois.edu

Department of Statistics and Operational Research, University of North Carolina at Chapel Hill, NC, USA
E-mail address: gterlov@unc.edu

Abstract. We derive a Gaussian Central Limit Theorem for the sample quantiles based on locally dependent random variables with explicit convergence rate. Our approach is based on converting the problem to a sum of indicator random variables, applying Stein's method for local dependence, and bounding the distance between two normal distributions. We also generalize this approach to the joint convergence of sample quantiles with an explicit convergence rate.

1. Introduction

The Central Limit Theorem (CLT) is one of the fundamental theorems in probability theory and statistics. In classical form, it states that the sum of i.i.d. finite variance random variables, appropriately centered and scaled, converges in distribution to the standard normal distribution. Since then, CLT has been strengthened and extended to various settings. The study of the asymptotic distribution of the sample median was first developed for the case of a continuous random variable by Sheppard in 1890 (according to Hald, 1998). Asymptotic properties of sample quantiles of independent random variables have been extensively studied for both continuous distributions (see *e.g.,* David and Nagaraja, 2003 and references therein) and discrete ones (see *e.g.,* Ma et al., 2011). The CLT for the sample median of i.i.d. random variables is derived either by converting the problem to a sum of indicators (Pollard, 2014), by the Delta method (van der Vaart, 1998), or through Bahadur representation (Bahadur, 1966) (known as Bahadur–Kiefer theorem due to its generalization Kiefer, 1967). The last one is particularly useful as it gives an almost sure representation of the empirical quantile as a sum of i.i.d. random variables and a further error term of order $n^{-3/4}\log n$. This, in turn, reduces the question of CLT and the rate of convergence to the classical setting. This method also has been generalized to tackle some of the dependent cases, which include m-dependent sequences (Sen, 1968), stationary sequences (Dembińska, 2014; Hesse, 1990), ϕ-mixing process (Sen, 1972), and auto-regressive processes (Dutta and Sen, 1971) among others.

Received by the editors August 11th, 2023; accepted February 19th, 2025.
2010 *Mathematics Subject Classification.* 60F05.
Key words and phrases. Sample median, Central Limit Theorem, Rate of convergence, Stein's method, Multivariate normal approximation.

While exact representations given by Bahadur–Kiefer type theorems yield Gaussian convergences in all these cases, the rates are not immediate, if achievable at all, as in the independent case.

In this work, we build on the elementary technique of converting sample quantiles to a sum of indicators as in Pollard (2014) and combine it with Stein's method to derive a Gaussian convergence for sample qualities with local dependencies. The class of dependencies that we consider is a direct generalization of the m-dependent sequences, where we require to know only some properties of dependency neighborhoods (see Assumption II). Hence, our result can be applied to models not only for linear dependency structures but also for more graphical ones. Another benefit of our approach is that it gives an explicit rate of convergence and directly generalizes to a multivariate version. We also discuss the optimality of the derived rate of convergence in Section 5.

1.1. *Stein's method.* Stein's method bounds the distance between the random variable of interest W and the standard normal variable Z in the following way

$$d(W,Z) \leqslant \sup_{f\in\mathcal{D}} \left|\mathbb{E}(f'(W) - Wf(W))\right| \tag{1.1}$$

for an adequately chosen class of functions $\mathcal{D}$ depending on the metric $d(\cdot,\cdot)$. When $W = \sum_{i=1}^n X_i$, it turns out that it is often easier to work with the right-hand side of (1.1), even if X_i have dependencies between each other. Depending on the structure of such dependencies, one would bound $\mathbb{E}(f'(W) - Wf(W))$ in different ways. Hence Stein's method can be used with variety of approaches such as exchangeable pairs (Stein, 1972), dependency neighborhoods or local dependencies (Baldi et al., 1989; Chen and Shao, 2004; Fang, 2016; Fang and Koike, 2021; Rinott, 1994), size-bias (Goldstein and Rinott, 1996) and zero-bias couplings (Goldstein, 2022; Goldstein and Reinert, 1997), Stein coupling (Chen and Röllin, 2010), and through Malliavin calculus (Nourdin et al., 2010) among others. In Section 3.1, we briefly discuss the basics of Stein's method for local dependence (also known as the dependency neighborhoods approach) and refer to Chen et al. (2011); Diaconis and Holmes (2004); Ross (2011) for further reading on the topic.

One particular advantage of Stein's method is that $\mathbb{E}(f'(W) - Wf(W))$ naturally preserves the additive structure of random variables. Hence, when $W = \sum_{i=1}^n X_i$, Stein's method allows working with local interactions among X_i's to derive global convergence. On the other hand, when one works with non-additive functions, such as sample quantiles, applying Stein's method requires additional effort.

1.2. *Notations.* Throughout this paper, we will use the following notations

- Z always denotes a standard normal random variable, Φ, ϕ denote the distribution and density function of Z, respectively.
- $\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_\ell)$ denotes an ℓ-dimensional standard Gaussian vector.
- $\mathcal{L}(W)$ - the law of random variable W,
- $\overline{X} := X - \mathbb{E}\,X$ denotes a centered version of a random variable X.
- W_i represents the i^{th} coordinate of a vector $\boldsymbol{W}$,
- $\mathrm{d}_{\mathrm{KS}}(W,Z) := \sup_{x\in\mathbb{R}} |\mathbb{P}(W \leqslant x) - \mathbb{P}(Z \leqslant x)|$ – Kolmogorov-Smirnov distance,
- $\mathrm{d}_{\mathcal{W}}(W,Z) := \mathrm{d}_{\mathcal{W}}(\mathcal{L}(W),\mathcal{L}(Z)) := \sup_{h:1\text{-Lip.}} |\mathbb{E}\,h(W) - \mathbb{E}\,h(Z)|$ - Wasserstein distance,
- $\mathrm{d}_{\mathrm{TV}}(W,Z) := \mathrm{d}_{\mathrm{TV}}(\mathcal{L}(W),\mathcal{L}(Z)) := \sup_{\text{Borel set } A} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)|$ - Total variation distance.
- $f \lesssim g$ if $f = O(g)$, $f \ll g$ if $f = o(g)$ and $f \approx g$ if $c'g \leqslant f \leqslant cg$ for some universal constants $c, c' > 0$,
- $x \vee y := \max\{x, y\}$ and $x \wedge y := \min\{x, y\}$.

Recall that for a matrix $V = (\!(v_{ij})\!)_{1\leqslant i\leqslant n, 1\leqslant j\leqslant m}$ the Frobenius norm of V is defined as

$$\|V\|_{\mathrm{F}} := \sqrt{\sum_{i,j} |v_{ij}|^2}$$

and the operator norm of V is defined as

$$\|V\|_{\mathrm{op}} := \sup\{\|V\mathbf{u}\|_2 \,/\, \|\mathbf{u}\|_2 \,, \mathbf{u} \neq 0\}.$$

1.3. *Setup.* Let $X_1, X_2, \dots, X_n$ be real-valued random variables (not necessarily independent) with CDF's $\{F_i(x)\}_{i\in[n]}$, respectively. Given $\alpha \in (0,1)$, we define the α^{th} sample quantile as

$$Q_n^{(\alpha)} := \sup\{x \in \mathbb{R} \mid |\{i \mid X_i \leqslant x\}| \leqslant \lfloor n\alpha \rfloor\}. \tag{1.2}$$

In particular, the sample median is defined as

$$M_n = Q_n^{(1/2)} := \sup\{x \in \mathbb{R} \mid |\{i \mid X_i \leqslant x\}| \leqslant \lfloor n/2 \rfloor\}. \tag{1.3}$$

Fix an integer $\ell \geqslant 1$ and two increasing sequences of real numbers $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_\ell < 1$ and $m_1 < m_2 < \cdots < m_\ell$. We assume the following.

Assumption I. (1) For all $k \in [\ell], i \in [n]$, the equations $F_i(x) = \alpha_k$ has a unique solution at $x = m_{\alpha_k} =: m_k$,

(2) For all $k \in [\ell], i \in [n]$, the CDF F_i is continuously twice differentiable at m_k,

(3) There exists $\varepsilon, A > 0$ such that for all $|y| \leqslant \varepsilon$ we have

$$\left|F_i''(m_k + y)\right|, \left|F_i'(m_k + y)\right| \leqslant A \text{ for all } k \in [\ell],\ i \in [n]. \tag{1.4}$$

We define

$$\begin{aligned} \boldsymbol{\mu} &:= (m_1, m_2, \cdots, m_\ell), \\ \theta_n(x) &:= \frac{1}{n}\sum_{j=1}^n F_j'(x) \quad \text{for } x \in \mathbb{R}, \\ \text{and} \quad \Theta_n &:= \operatorname{diag}\left(\theta_n(m_1), \theta_n(m_2), \dots, \theta_n(m_\ell)\right). \end{aligned} \tag{1.5}$$

When $\alpha = 1/2$, for simplicity, we may assume that $m_\alpha = 0$, and thus

$$\theta_n := \theta_n(0) := \frac{1}{n}\sum_{j=1}^n F_j'(0). \tag{1.6}$$

In this article, we consider the case when $(X_i)_{i\in[n]}$ is a locally dependent sequence of random variables. In particular, we will quantify the dependence via dependency neighborhoods. Dependency neighborhoods are present in various forms in the literature, see Barbour et al. (1989); Chen and Shao (2004); Fang (2016); Paulin (2012); Rinott and Rotar (1996) among others. For instance, one of the most common notions of generating dependency neighborhoods is via dependency graphs.

Definition 1.1 (Dependency graph). A graph $G = ([n], E)$ is called a dependency graph for random variables $(X_i)_{i\in[n]}$ if $(X_i)_{i\in A}$ and $(X_j)_{j\in B}$ are mutually independent whenever the sets of vertices $A \subseteq [n]$ and $B \subseteq [n]$ share no edges.

In this paper, we consider a more general notion of dependency neighborhoods given in the following assumption.

Assumption II. We assume that the sequence of random variables $(X_1, X_2, \dots, X_n)$ has the following structure of dependency neighborhoods, *i.e.,*

(1) $\forall i \in [n]$, $\exists N_i \subseteq [n]$ such that X_i is independent of $(X_l)_{l\in N_i^c}$.

(2) $\forall i \in [n]$ and $j \in N_i$, $\exists N_{ij} \subseteq [n]$ such that $N_i \subset N_{ij}$ and (X_i, X_j) are independent of $(X_k)_{k\in N_{ij}^c}$,

(3) $\forall i \in [n]$, $j \in N_i$, and $k \in N_{ij}$, $\exists N_{ijk} \subseteq [n]$ such that $N_{ij} \subset N_{ijk}$ and (X_i, X_j, X_k) are independent of $(X_s)_{s \in N_{ijk}^c}$.

We refer to the set $N_i = \{j \mid j \sim i\} \cup \{i\}$ as the dependency neighborhood of X_i and let

$$D_1 := \max_i |N_i|, \quad D_2 := \max_{i,j} |N_{ij}|, \quad \text{and} \quad D_3 := \max_{i,j,k} |N_{ijk}|. \tag{1.7}$$

These quantities are *the main parameters of a dependency neighborhood and may depend on* n. Notice that the dependency graph condition is stronger than Assumption II as it provides a natural choice for the dependency neighborhoods such that $D_2 \leqslant D_1^2$ and $D_3 \leqslant D_1^3$.

Given a vector $\boldsymbol{x} \in \mathbb{R}^\ell$, we consider the centered random vector

$$\boldsymbol{Y}_{i,\boldsymbol{x}} := \left(\mathbb{1}_{X_i \leqslant m_k + n^{-1/2} \cdot x_k} - F_i(m_k + n^{-1/2} \cdot x_k) \right)_{k=1}^{\ell}. \tag{1.8}$$

Notice that the random vectors $(\boldsymbol{Y}_{i,\boldsymbol{x}})_{i=1}^n$ inherit the structure of dependency neighborhoods from $(X_i)_{i=1}^n$. Let

$$\Sigma_{n,\boldsymbol{x}} = \frac{1}{n} \operatorname{Var}\left(\sum_{i=1}^n \boldsymbol{Y}_{i,\boldsymbol{x}} \right)$$

be the variance-covariance matrix of the random vector $\frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{Y}_{i,\boldsymbol{x}}$. In the univariate case, we denote it as $\sigma_{n,x}^2$. Finally, to simplify notation, we denote

$$\sigma_n := \sigma_{n,0} \quad \text{and} \quad \Sigma_n := \Sigma_{n,\mathbf{0}}. \tag{1.9}$$

We will assume that the matrix Σ is invertible.

Remark 1.1. In particular, if $\alpha = 1/2$ and $m_\alpha = 0$ and in addition to the Assumption I we know that $\mathbb{P}(X_i \vee X_j \leqslant 0) = \frac{1}{4}$ for all $i \neq j$, then $\mathbb{1}_{X_i \leqslant 0}$ and $\mathbb{1}_{X_j \leqslant 0}$ are uncorrelated for $i \neq j$ and hence $\sigma \equiv 1/2$. Similarly, when the random variables are independent, we have

$$\Sigma_n = (\!(\alpha_{i \wedge j} - \alpha_i \alpha_j)\!)_{i,j=1}^{\ell},$$

which does not depend on n.

1.4. *Main results.* In this section, we present our main results covering the univariate case (see Theorem 1.2) and its extension to the multivariate case (see Theorem 1.3). To present the univariate result, we assume that $\alpha = 1/2$ and $m_\alpha = 0$; however, our argument can be easily extended to any other quantile that satisfies Assumption I. To keep the exposition simple, we make a natural assumption that some of the parameters are uniformly bounded (1.10) and (1.12), which could be relaxed by changing the Theorems 1.2 and 1.3 to hold for n large enough, see Remark 1.5.

Theorem 1.2. *Let $(X_i)_{i \in [n]}$ be a sequence of random variables satisfying Assumption I with $\alpha = 1/2$, $m_\alpha = 0$ and Assumption II. Let M_n be as in (1.3), A be as in (1.4), θ_n be as in (1.6), D_i as in (1.7), σ_n as in (1.9), and $Z \sim \mathrm{N}(0,1)$. Assume that there are constants $0 < c < C < \infty$ such that for all $n \in \mathbb{N}$*

$$c < \sigma_n^2, \theta_n < C \quad \text{and} \quad \max\{D_1, D_2, D_3\} \leqslant C \cdot n^{\frac{1}{4}}. \tag{1.10}$$

Then

$$d_{KS}\left(\frac{\theta_n \sqrt{n}}{\sigma_n} M_n \,,\, Z \right) \lesssim D_1 \sqrt{\log n + D_1} \cdot \sqrt{\frac{\log n}{n}} + \frac{\max\{D_1, D_2, D_3\}^2}{\sqrt{n}}.$$

In particular, when $\{X_i\}_{n \in \mathbb{N}}$ are i.i.d. the bound becomes $\log n/\sqrt{n}$. In Section 5, we show that in i.i.d. case, one can actually remove the $\log n$ factor and conjecture that it is true whenever $\{D_i\}_{i \in \{1,2,3\}}$ are bounded by some constant.

Now, we present the multivariate version involving the joint distribution of sample quantiles. For two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^\ell$ we say that $\mathbf{x} \preceq \mathbf{y}$ if for all $i \in [\ell]$ we have that $x_i \leqslant y_i$.

Theorem 1.3. *Let $(X_i)_{i\in[n]}$ be a sequence of random variables satisfying Assumption I and II. Let A be as in (1.4), $\Theta_n, \theta_n, \boldsymbol{\mu}$ be as in (1.5), D_i as in (1.7), Σ_n as in (1.9),*

$$\theta_{\min} := \min_{k\in[\ell]} \theta_n(m_k), \quad \sigma^2_{\max} := \max_{k\in[\ell]} \Sigma_n(k,k) \quad \text{and} \quad \mathbf{Q}_n := \left(Q_n^{(\alpha_1)}, Q_n^{(\alpha_2)}, \ldots, Q_n^{(\alpha_\ell)}\right), \tag{1.11}$$

and $\boldsymbol{Z} \sim \mathrm{N}(0, I_\ell)$. Assume that there are constants $0 < c < C < \infty$ such that for all $n \in \mathbb{N}$

$$c < \sigma^2_{\max}, \theta_n(m_1), \ldots, \theta_n(m_\ell), \left\|\Sigma_n^{-1/2}\right\|_{\mathrm{op}} < C \quad \text{and} \quad \max\{D_1, D_2, D_3\} \leqslant C \cdot n^{\frac{1}{4}}. \tag{1.12}$$

Then

$$\sup_{\boldsymbol{x}\in\mathbb{R}^\ell} \left|\mathbb{P}\left(\Theta_n\sqrt{n}\cdot(\mathbf{Q}_n - \boldsymbol{\mu}) \preceq \boldsymbol{x}\right) - \mathbb{P}\left(\Sigma_n^{1/2}\boldsymbol{Z} \preceq \boldsymbol{x}\right)\right|$$
$$\lesssim D_1\left(\sqrt{\log n + D_1}\right)\cdot\frac{\log n}{\sqrt{n}} + \frac{\ell^{1/4}}{\sqrt{n}} D_1\left(D_2 + D_3\ell^{-1}\right).$$

Remark 1.4. Notice that letting $\ell = 1$ and assuming that $\mu = 0$ reduces the bound from Theorem 1.3 to the one in Theorem 1.2.

Remark 1.5. One can replace the bounded second derivative assumption in (1.4) by Hölder continuity assumption for the F_i'. But this will give a slower rate of convergence in Theorem 1.2. One can also relax assumption (1.10) and derive the following bound that holds for $n > N(\varepsilon, \theta_n, \sigma_n)$

$$\mathrm{d_{KS}}\left(\frac{\theta_n\sqrt{n}}{\sigma_n}M_n\,,\, Z\right) \lesssim \frac{A}{\theta_n^2}\cdot\frac{D_1\vee\sigma_n^2}{\sigma_n}\left(1 + \frac{\theta_n}{\sigma_n}\cdot\sqrt{\frac{D_1\vee\sigma_n^2}{\log n}}\right)\cdot\frac{\log n}{\sqrt{n}} + \frac{D_1(D_2+D_3)}{\sigma_n^3\sqrt{n}}.$$

Notice that in multivariate case c and C depend on ℓ, again relaxing this assumption one could derive the bound for n large enough

$$\sup_{\boldsymbol{x}\in\mathbb{R}^\ell} \left|\mathbb{P}\left(\Theta_n\sqrt{n}\cdot(\mathbf{Q}_n - \boldsymbol{\mu}) \preceq \boldsymbol{x}\right) - \mathbb{P}\left(\Sigma_n^{1/2}\boldsymbol{Z} \preceq \boldsymbol{x}\right)\right|$$
$$\lesssim \frac{A}{\theta^2_{\min}}\cdot\ell\left\|\Sigma_n^{-1/2}\right\|_{\mathrm{op}}\left(D_1\vee\sigma^2_{\max}\right)\left(1 + \left\|\Sigma_n^{-1/2}\right\|_{\mathrm{op}}\theta_{\min}\cdot\sqrt{\frac{D_1\vee\sigma^2_{\max}}{\log n}}\right)\cdot\frac{\log n}{\sqrt{n}}$$
$$+ \frac{\ell^{1/4}}{\sqrt{n}}\left\|\Sigma_n^{-1/2}\right\|^3_{\mathrm{op}} D_1\left(D_2 + D_3\ell^{-1}\right).$$

Finally, the constant factor in the bound can also be made explicit. Since we did not optimize over it at in any of the steps in the proof, we decided not to pursue this direction.

While we present Theorem 1.2 under the assumption of local dependency, our approach can be applied to other dependency structures. The first step of our argument is applying the classical linearization technique that rewrites $\mathbb{P}(M_n \leqslant x)$ as the probability that the sum of Bernoulli random variables is less than some value. This step can be done under any dependence of X_i's. However, the mean of these Bernoulli random variables Y_i depends on x. While it is not an issue in the derivation of regular CLT, it poses significant complications in bounding the convergence rate when the value of x is "far" from the true value of the median. Hence, we consider two cases when $|x| \leqslant K_n$ and $|x| > K_n$ for some threshold function K_n. The next step is applying the appropriate approach of Stein's method to treat the former case and establishing a concentration bound for Bernoulli random variables to treat the latter. In the end, we optimize over K_n to derive the result. In conclusion, our argument can be adopted whenever Stein's method can be used to derive CLT for the corresponding Bernoulli random variables Y_i, and one can derive a needed concentration inequality.

1.5. *Organization.* This paper is organized as follows. We present an application of our univariate result to the moving-average model in Section 2, and state and prove preliminary results in Section 3. The proofs of Theorems 1.2 and 1.3 are presented in Section 4. Finally, we discuss the optimality of the rate of convergence in Section 5.

2. Application to the moving-average model

In this section, we present an application of Theorem 1.2 to the moving-average model. We refer to Brockwell and Davis (2006) for an overview of moving-average models and their significance in time series analysis.

One usually considers a sequence $(\zeta_i)_{i\in\mathbb{Z}}$ of i.i.d. standard normal random variables. However, to apply our results, we assume that $\{\zeta_i\}_{i\in\mathbb{Z}}$ is a sequence of independent (not necessarily identical distributed) random variables that are symmetric around zero and satisfy Assumption I. In particular, the median of any finite linear combination of ζ_i's is zero.

Thus, one can estimate the median μ for the following linear model using sample median based on the observed data $(X_t)_{t\in[n]}$, where

$$X_t = \mu + \sum_{i=1}^{q} c_i\zeta_{t-i} + \zeta_t$$

for all $t \in [n]$. This is also known as MA(q) model with parameters μ, $c_1, c_2, \ldots, c_q$.

Notice that X_i is independent of X_j whenever $|i-j| \geqslant q+1$. Consider a natural choice of dependency neighborhoods (in this case generated by a graph), where $X_j \in N_i$ if and only if $|i-j| \leqslant q$. Hence the parameters D_i of the dependency neighborhoods, as in (1.7), satisfy $D_1 \leqslant 2q+1$, $D_2 \leqslant D_1^2$, and $D_3 \leqslant D_1^3$. Although X_t's are no longer independent, they are still identically distributed. Thus, θ and σ are some fixed but unknown constants. In fact, from (1.6) one can see that $\theta = F'_{X_1}(\mu)$. Hence, Theorem 1.2 implies the sample median $M_n(X_1, X_2, \ldots, X_n)$ obeys a central limit theorem with the following upper bound on the rate of convergence,

$$\mathrm{d}_{\mathrm{KS}}\left(\frac{\theta}{\sigma}\cdot\sqrt{n}(M_n-\mu)\,,\,Z\right) \lesssim \frac{q^2+q\log n}{\sqrt{n}}, \tag{2.1}$$

where $Z \sim \mathrm{N}(0,1)$. In particular, the median of MA(q) obeys a Gaussian CLT whenever $q \ll n^{1/4}$. Here, we note that the constant θ/σ appearing on the left-hand side of (2.1) is still an unknown quantity. However, we aim to merely highlight the Gaussian approximation and we leave the estimation of this quantity as a topic for future consideration.

3. Preliminary Results

3.1. *Local dependence.* In this section, we state the results related to the dependency neighborhoods approach that will be used to prove the main results.

As mentioned above, we will use a CLT for a sum of indicator random variables with local dependencies. Various theorems cover this case in the univariate case (see Baldi et al., 1989; Chen and Shao, 2004; Rinott, 1994 among others) and give the same order. We chose to present the multivariate version of the result.

Theorem 3.1 (Theorem 2.1 and Remark 2.2 in Fang, 2016)**.** *Suppose that $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ are ℓ-dimensional random vectors with $\mathbb{E}\,\boldsymbol{X}_i = \boldsymbol{0}$. Suppose $(\boldsymbol{X}_i)_{i\in[n]}$ have dependency neighborhoods as in Assumption II and for all $i \in [n]$, $j \in N_i$, and $k \in N_{ij}$ we have that*

$$|X_i| \leqslant \beta, \quad |N_i| \leqslant D_1, \quad |N_{ij}| \leqslant D_2, \quad \text{and} \quad |N_{ijk}| \leqslant D_3,$$

then letting $\boldsymbol{W} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ *and its variance-covariance matrix given by* Σ *we have that*

$$\sup_{A \in \mathcal{C}} \left| \mathbb{P}\left(\boldsymbol{W} \in A\right) - \mathbb{P}\left(\Sigma^{1/2} \boldsymbol{Z} \in A\right) \right| \lesssim n \ell^{1/4} \cdot \left\| \Sigma^{-1/2} \right\|_{\mathrm{op}}^{3} \cdot \beta^3 \cdot D_1 \left(D_2 + D_3/\ell\right),$$

where $\mathcal{C}$ *is the collection of all convex sets in* $\mathbb{R}^\ell$ *and* $\|\cdot\|_{\mathrm{op}}$ *denotes the operator norm.*

An important ingredient in our argument is the application of a concentration inequality for locally dependent random variables. Some such inequalities are already present in the literature (Janson, 2004; Paulin, 2012). However, they focus on different, often less general, dependency settings. On top of that, a slightly weaker statement than a Hoeffding-type inequality will be sufficient for our purposes.

Lemma 3.2. *Let* $X_1, X_2, \ldots, X_n$ *be a sequence of mean zero random variables taking values in* $[-1, 1]$. *Suppose that* $\forall i \in [n]$, $\exists N_i \subseteq [n]$ *such that* X_i *is independent of* $(X_l)_{l \in N_i^c}$ *(*i.e., *Assumption II.(1) holds). Then for all* $0 < t \ll n$, *we have*

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geqslant t\right) \leqslant \exp\left(-\frac{n}{D_1} \cdot g\left(\frac{t}{n}\right)\right),$$

where for $x \in \mathbb{R}$

$$g(x) := 1 + x + x^2 - e^x.$$

Notice that when x is small $g(x) \approx x^2/2$. Thus, Lemma 3.2 provides a local sub-Gaussian tail bound.

Proof: Consider the moment generating function of $W := \sum_{i=1}^{n} X_i$, *i.e.*, $M(r) := \mathbb{E}\left(e^{rW}\right)$. Define $W_{N_i} := \sum_{j \in N_i} X_j$. Since for every $i \in [n]$ we have that $\mathbb{E}\, X_i = 0$ and X_i is independent of $W - W_{N_i}$ we can rewrite the derivative of $M(r)$ in the following way

$$\begin{aligned} \left|M'(r)\right| &\leqslant \sum_{i=1}^{n} \left|\mathbb{E}\left(X_i e^{rW}\right)\right| \\ &= \sum_{i=1}^{n} \left|\mathbb{E}\left(X_i e^{rW}(1 - e^{-rW_{N_i}})\right)\right| \leqslant n(e^{rD_1} - 1) M(r), \end{aligned}$$

where in the last inequality we used the fact that $|X_i| \leqslant 1$ and $|N_i| \leqslant D_1$ for all i. Dividing by $M(r)$ on both sides of the inequality and integrating with respect to r yields that

$$\log M(r) \leqslant n \int_0^r (e^{sD_1} - 1)\, ds = \frac{n}{D_1}(e^{rD_1} - 1 - rD_1).$$

This bound allows us to derive the following concentration inequality for $t > 0$

$$\mathbb{P}\left(W \geqslant t\right) \leqslant \exp(-rt) \cdot M(r) \leqslant \exp\left(-rt + \frac{n}{D_1}(e^{rD_1} - 1 - rD_1)\right).$$

Picking $r = t/(nD_1)$ and recalling that $t \ll n$ we get the desired inequality

$$\mathbb{P}\left(W \geqslant t\right) \leqslant \exp\left(-\frac{n}{D_1}\left(\frac{t^2}{n^2} + 1 + \frac{t}{n} - e^{\frac{t}{n}}\right)\right).$$

This completes the proof. ∎

3.2. *Bound on the distance between two Gaussian vectors with the same mean.* One of the terms in the upper bounds in each of our main theorems is of the form

$$\sup_{\boldsymbol{x}} \left|\mathbb{P}(\boldsymbol{Z}_1 \preceq \boldsymbol{x}) - \mathbb{P}(\boldsymbol{Z}_2 \preceq \boldsymbol{x})\right|, \tag{3.1}$$

where $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are ℓ-dimensional Gaussian random vectors with the same mean and variance–covariance matrices Σ_1 and Σ_2, respectively. This section is dedicated to addressing this problem.

In the one-dimensional case, one can easily upper bound the Kolmogorov-Smirnov distance by the square root of the L^1-Wasserstein distance as follows. With $\rho \geqslant 1$, we have

$$\begin{aligned} \mathrm{d}_{\mathrm{KS}}(\mathrm{N}(0,\rho^2), \mathrm{N}(0,1)) &\leqslant \sqrt{4/\sqrt{2\pi} \cdot \mathrm{d}_{\mathcal{W}}(\mathrm{N}(0,\rho^2), \mathrm{N}(0,1))} \\ &\leqslant \sqrt{4/\sqrt{2\pi} \cdot \mathbb{E}_{Z' \sim \mathrm{N}(0,\rho^2-1)} \left|Z'\right|} \leqslant \sqrt{4/\pi} \cdot \left|\rho^2 - 1\right|^{1/4}. \end{aligned}$$

However, one can derive a better bound by applying Stein's estimates (see Proposition 3.6.1 in Nourdin and Peccati, 2012) or via direct computation as follows.

Lemma 3.3. *For $\rho \geqslant 1$, we have*

$$d_{\mathrm{TV}}(\mathrm{N}(0,\rho^2), \mathrm{N}(0,1)) \leqslant \frac{2}{\sqrt{\pi}} \left|\rho - 1\right| \sqrt{\frac{\log \rho}{\rho^2 - 1}} \cdot \exp\left(-\frac{\log \rho}{\rho^2 - 1}\right) \leqslant \sqrt{\frac{2}{\pi e}} \cdot \left|\rho - 1\right|.$$

Proof: Letting φ be the density function of standard normal distribution, we get

$$\mathrm{d}_{\mathrm{TV}}(\mathrm{N}(0,\rho^2), \mathrm{N}(0,1)) = \int_0^\infty \left|\rho^{-1}\varphi(x/\rho) - \varphi(x)\right| dx = 2\int_{x^*}^\infty (\rho^{-1}\varphi(x/\rho) - \varphi(x)) dx,$$

where $x^* > 0$ satisfies the following equality $\rho^{-1}\exp(-x^2/2\rho^2) = \exp(-x^2/2)$, *i.e.*,

$$x^* = \sqrt{2(\log \rho)/(1 - \rho^{-2})}. \tag{3.2}$$

Letting $Z \sim \mathrm{N}(0,1)$, we get that

$$\begin{aligned} \mathrm{d}_{\mathrm{TV}}(\mathrm{N}(0,\rho^2), \mathrm{N}(0,1)) &= 2\,\mathbb{P}\left(\frac{x^*}{\rho} \leqslant Z \leqslant x^*\right) \\ &\leqslant 2x^*\left(1 - \rho^{-1}\right)\varphi\left(x^*/\rho\right) = \frac{2}{\sqrt{\pi}}(\rho - 1)\sqrt{\frac{\log \rho}{\rho^2 - 1}} \cdot \exp\left(-\frac{\log \rho}{\rho^2 - 1}\right). \end{aligned}$$

It is easy to check that for all $x \geqslant 0$ we have $\sqrt{x}e^{-x} \leqslant 1/\sqrt{2e}$. Simplifying, we get the result. ∎

Remark 3.4. For general $\sigma_1 > \sigma_2$, letting $\rho = \sigma_1/\sigma_2 > 1$, by the scaling property of normal distribution, we have

$$\mathrm{d}_{\mathrm{TV}}(\mathrm{N}(0,\sigma_1^2), \mathrm{N}(0,\sigma_2^2)) = \mathrm{d}_{\mathrm{TV}}(\mathrm{N}(0,\rho^2), \mathrm{N}(0,1)).$$

In the multivariate case, bounding (3.1) becomes more complex. We refer to Devroye et al. (2018) and references therein for the background on this question. While for our purposes, we need to bound the difference between two measures over specific sets, computing that does not seem feasible. Hence, similarly to the univariate case, we derive a bound in the total variation distance. One can bound this distance in terms of the eigenvalues of the $\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}$ or, equivalently, $\Sigma_2^{-1/2}\Sigma_1\Sigma_2^{-1/2}$. We present that in Theorem 3.5, however for clarity of presentation we assume that $\Sigma_1 = I, \Sigma_2 = \Sigma$. Although our technique is simple, to our knowledge, it is not present in the literature, and we find it interesting on its own. For instance, it improves the constant in the upper bound of Theorem 1.1 in Devroye et al. (2018).

Theorem 3.5. *Let $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ be mean zero ℓ-dimensional normal random vectors with variance-covariance matrices I and Σ, respectively. Suppose that $\lambda_1, \lambda_2, \ldots, \lambda_\ell$ are the eigenvalues of Σ, then*

$$d_{\mathrm{TV}}(\boldsymbol{Z}_1, \boldsymbol{Z}_2)^2 \leqslant \sum_{i=1}^{\ell} \frac{\left(\sqrt{\lambda_i} - 1\right)^2}{\lambda_i + 1} \leqslant \|\Sigma - I\|_{\mathrm{F}}^2 .$$

The following bound will be helpful when none of the variance–covariance matrices are identity matrices.

Lemma 3.6. *Let Σ_1 be a $\ell \times \ell$ invertible symmetric matrix and Σ_2 be another $\ell \times \ell$ matrix. Then*

$$\left\|\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}\right\|_{\mathrm{F}} \leqslant \left\|\Sigma_1^{-1}\right\|_{\mathrm{op}} \cdot \|\Sigma_2\|_{\mathrm{F}} .$$

Proof: Let $\lambda_1, \lambda_2, \ldots, \lambda_\ell$ be the eigenvalues of Σ_1, and $\{u_i \mid 1 \leqslant i \leqslant \ell\}$ be the corresponding orthonormal eigenbasis. Then Σ_1^{-1} can be written as $\Sigma_1^{-1} = \sum_{i=1}^{d} \lambda_i^{-1} u_i u_i^{\intercal}$. Using this, we get that

$$\left\|\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}\right\|_{\mathrm{F}}^2 = \sum_{i,j=1}^{\ell} (\lambda_i\lambda_j)^{-1}(u_j^{\intercal}\Sigma_2 u_i)^2 \leqslant \left\|\Sigma_1^{-1}\right\|_{\mathrm{op}}^2 \cdot \sum_{i,j=1}^{\ell} (u_j^{\intercal}\Sigma_2 u_i)^2 = \left\|\Sigma_1^{-1}\right\|_{\mathrm{op}}^2 \cdot \|\Sigma_2\|_{\mathrm{F}}^2 .$$

This completes the proof. ■

Let f, g be the density of the random variables X, Y, respectively, and recall the definitions of the Hellinger distance and affinity (also referred to as the Bhattacharyya coefficient) defined as

$$\mathrm{d_H}(X, Y) := \left(\frac{1}{2}\int \left(\sqrt{f} - \sqrt{g}\right)^2\right)^{\frac{1}{2}} \quad \text{and} \quad \alpha(X, Y) := \int \sqrt{fg}.$$

It is well known (Strasser, 1985) that the total variation distance can be upper bounded in terms of these distances as follows

$$\mathrm{d_{TV}}(X, Y) \leqslant \sqrt{2}\mathrm{d_H}(X, Y) = \sqrt{2} \cdot \sqrt{1 - \alpha(X, Y)}. \tag{3.3}$$

In fact, a simple computation yields a slightly stronger inequality, namely

$$\mathrm{d_{TV}}(X, Y) \leqslant \sqrt{1 - \alpha(X, Y)^2}. \tag{3.4}$$

Indeed,

$$\begin{aligned} \mathrm{d_{TV}}(X, Y)^2 &= \left(\int f \vee g - 1\right)\left(1 - \int f \wedge g\right) = 1 - \int f \vee g \cdot \int f \wedge g \\ &\leqslant 1 - \left(\sqrt{(f \vee g)(f \wedge g)}\right)^2 = 1 - \alpha(X, Y)^2. \end{aligned}$$

Proof of Theorem 3.5: It is enough to bound $\int \sqrt{fg}$, where

$$f(\boldsymbol{x}) = (2\pi)^{-\ell/2} \cdot \exp\left(-\boldsymbol{x}^{\intercal}\boldsymbol{x}/2\right) \text{ and } g(\boldsymbol{x}) = (2\pi)^{-\ell/2}\det(\Sigma)^{-1/2} \cdot \exp\left(-\boldsymbol{x}^{\intercal}\Sigma^{-1}\boldsymbol{x}/2\right).$$

Then

$$\begin{aligned} \sqrt{f(\boldsymbol{x})g(\boldsymbol{x})} &= (2\pi)^{-\ell/2} \cdot \det(\Sigma)^{-1/4} \cdot \exp\left(-\boldsymbol{x}^{\intercal}(I + \Sigma^{-1})\boldsymbol{x}/4\right) \\ &= \frac{\det(\Sigma)^{1/4}}{\det((I+\Sigma)/2)^{1/2}} \cdot (2\pi)^{-\ell/2} \cdot \frac{\det((I+\Sigma)/2)^{1/2}}{\det(\Sigma)^{1/2}} \cdot \exp\left(-\boldsymbol{x}^{\intercal}(I + \Sigma^{-1})\boldsymbol{x}/4\right). \end{aligned}$$

Recognizing the density function of a normal distribution with mean zero and variance-covariance matrix $(I + \Sigma^{-1})/2$, we conclude that

$$\alpha(\boldsymbol{Z}_1, \boldsymbol{Z}_2) = \int \sqrt{fg} = \left(\frac{\det(\Sigma)^{1/2}}{\det((I+\Sigma)/2)}\right)^{\frac{1}{2}} = \left(\prod_{i=1}^{\ell} \frac{2\sqrt{\lambda_i}}{1+\lambda_i}\right)^{\frac{1}{2}}.$$

Hence, by (3.4) we get that

$$\mathrm{d_{TV}}(\boldsymbol{Z}_1, \boldsymbol{Z}_2)^2 \leqslant 1 - \left(\prod_{i=1}^{\ell} \frac{2\sqrt{\lambda_i}}{1+\lambda_i}\right) \leqslant \sum_{i=1}^{\ell}\left(1 - \frac{2\sqrt{\lambda_i}}{1+\lambda_i}\right) = \sum_{i=1}^{\ell} \frac{\left(\sqrt{\lambda_i}-1\right)^2}{\lambda_i+1}.$$

The conclusion follows from the fact that $\|\Sigma - I\|_{\mathrm{F}}^2 = \sum_{i=1}^{\ell}(\lambda_i - 1)^2$. ∎

3.3. *Variance Control.* We present our main results with the implicit variance $\sigma = \sigma_{n,0}$. To compute $\sigma_{n,0}$ explicitly, one would have to rely on particular properties of the model. For example, as we mentioned before, suppose in addition to Assumptions I, we have that $\mathbb{P}(X_i \vee X_j \leqslant 0) = 1/4$ for all $i \neq j$, then $\sigma \equiv \frac{1}{2}$.

Lemma 3.7. *In the situation of Theorem 1.3, let θ_n be as in (1.6), and ε, A are as in Assumption I.3. Then for any $\boldsymbol{x} \in \mathbb{R}^{\ell}$ such that $\max_{k\in[\ell]} |x_k/\theta_n(m_k)| \leqslant \varepsilon\sqrt{n}$ we have that*

$$d_{\mathrm{TV}}\left(\mathbf{N}_{\ell}\left(0, \Sigma_{n,\boldsymbol{x}}\right), \mathbf{N}_{\ell}\left(0, \Sigma_0\right)\right) \leqslant \frac{3\sqrt{2}A\ell}{\theta_{\min}} \cdot \left\|\Sigma_0^{-1}\right\|_{\mathrm{op}} \cdot \frac{D_1 \left\|\boldsymbol{x}\right\|_{\infty}}{\sqrt{n}}.$$

Proof: Applying Theorem 3.5 and Lemma 3.6 gives us that

$$\begin{aligned}\mathrm{d_{TV}}\left(\mathbf{N}_{\ell}\left(0, \Sigma_{n,\boldsymbol{x}}\right), \mathbf{N}_{\ell}\left(0, \Sigma_0\right)\right) &\leqslant \left\|\Sigma_0^{-1/2}\sigma_{nn,\boldsymbol{x}}\Sigma_0^{-1/2} - I\right\|_{\mathrm{F}} \\ &\leqslant \left\|\Sigma_0^{-1}\right\|_{\mathrm{op}} \cdot \left\|\Sigma_{n,\boldsymbol{x}} - \Sigma_0\right\|_{\mathrm{F}}.\end{aligned}$$

To compute $\left\|\Sigma_{n,\boldsymbol{x}} - \Sigma_0\right\|_{\mathrm{F}}$, we bound the difference in each coordinate. Define the function

$$f_{k,\boldsymbol{x}}(u) := \mathbb{1}_{u\in(m_k, m_k+n^{-1/2}\cdot x_k/\theta_n(m_k))} - \mathbb{1}_{u\in(m_k+n^{-1/2}\cdot x_k/\theta_n(m_k), m_k)}$$

for $k \in [\ell]$. Then

$$\mathbb{1}_{u\leqslant m_k+n^{-1/2}\cdot x_k/\theta_n(m_k)} = \mathbb{1}_{u\leqslant m_k} + f_{k,\boldsymbol{x}}(u).$$

In particular, using Assumption I.3 together with MVT we derive that

$$\begin{aligned}\mathbb{E}\left|f_{k,\boldsymbol{x}}(X_j)\right| &\leqslant \mathbb{P}\left(X_j \in (m_k, m_t + n^{-1/2}\cdot x_k/\theta_n(m_k))\right) + \mathbb{P}\left(X_j \in (m_k + n^{-1/2}\cdot x_k/\theta_n(m_k), m_k)\right) \\ &\leqslant A \cdot \frac{D_1}{\sqrt{n}}\left|\frac{x_k}{\theta_n(m_k)}\right|. \qquad (3.5)\end{aligned}$$

Thus for s and t in $[\ell]$, we get

$$\begin{aligned}&\left|(\Sigma_{n,\boldsymbol{x}})_{st} - \Sigma_{st}\right| \\ &= \frac{1}{n}\left|\sum_{i,\,j\in N_i} \mathrm{Cov}\left(\mathbb{1}_{X_i\leqslant m_s}, f_{t,\boldsymbol{x}}(X_j)\right) + \mathrm{Cov}\left(f_{s,\boldsymbol{x}}(X_i), \mathbb{1}_{X_t\leqslant m_t}\right) + \mathrm{Cov}\left(f_{s,\boldsymbol{x}}(X_i), f_{t,\boldsymbol{x}}(X_j)\right)\right| \\ &\leqslant 2A\cdot\frac{D_1}{\sqrt{n}}\cdot\left(\left|\frac{x_s}{\theta_s}\right| + \left|\frac{x_t}{\theta_t}\right| + \sqrt{\left|\frac{x_s x_t}{\theta_s\theta_t}\right|}\right) \leqslant 3A\cdot\frac{D_1}{\sqrt{n}}\cdot\left(\left|\frac{x_s}{\theta_s}\right| + \left|\frac{x_t}{\theta_t}\right|\right),\end{aligned}$$

where in the first inequality, we bounded each indicator $\mathbb{1}_{X_i\leqslant m_k}$ by 1, used (3.5), and applied the Cauchy–Schwarz inequality for the third term within the sum. Hence

$$\left\|\Sigma_{n,\boldsymbol{x}} - \Sigma_0\right\|_{\mathrm{F}} \leqslant 3A\cdot\frac{D_1}{\sqrt{n}}\cdot\sqrt{\sum_{s,t}\left(\left|\frac{x_s}{\theta_s}\right| + \left|\frac{x_t}{\theta_t}\right|\right)^2} \leqslant 3\sqrt{2}\ell\cdot\frac{A}{\theta_{\min}}\cdot\frac{D_1}{\sqrt{n}}\cdot\left\|\boldsymbol{x}\right\|_{\infty}. \qquad (3.6)$$

This completes the proof. ∎

Remark 3.8. In Lemma 3.7, D_1 can be replaced by the average size of the neighborhood instead of the maximum one.

4. Proof of Main Results

4.1. *Limiting distribution of the sample median.*

Proof of Theorem 1.2: Recall the definition of the sample median

$$M_n = \sup\{x \in \mathbb{R} : |\{i \mid X_i \leqslant x\}| \leqslant \lfloor n/2 \rfloor\}$$

and so

$$\mathbb{P}(M_n \leqslant y) = \mathbb{P}(|\{i \mid X_i \leqslant y\}| \geqslant \lfloor n/2 \rfloor + 1) = \begin{cases} \mathbb{P}\left(|\{i \mid X_i \leqslant y\}| \geqslant \frac{n+2}{2}\right) & \text{if } n \text{ is even,} \\ \mathbb{P}\left(|\{i \mid X_i \leqslant y\}| \geqslant \frac{n+1}{2}\right) & \text{if } n \text{ is odd.} \end{cases}$$

Thus fixing $x \in \mathbb{R}$. We have

$$\begin{aligned} \mathbb{P}\left(\theta_n \sqrt{n} M_n \leqslant x\right) &= \mathbb{P}\left(M_n \leqslant n^{-1/2} \cdot x/\theta_n\right) \\ &\in \left[\mathbb{P}\left(\left|\{i \mid X_i \leqslant n^{-1/2} \cdot x/\theta_n\}\right| \geqslant n/2 + 1\right), \mathbb{P}\left(\left|\{i \mid X_i \leqslant n^{-1/2} \cdot x/\theta_n\}\right| \geqslant n/2\right)\right] \\ &= \left[\mathbb{P}\left(\sum_{i=1}^n \mathbb{1}_{X_i \leqslant n^{-1/2} \cdot x/\theta_n} \geqslant n/2 + 1\right), \mathbb{P}\left(\sum_{i=1}^n \mathbb{1}_{X_i \leqslant n^{-1/2} \cdot x/\theta_n} \geqslant n/2\right)\right] \end{aligned} \tag{4.1}$$

By subtracting $\sum_{i=1}^n F_i\left(n^{-1/2} \cdot x/\theta_n\right)$ to each side of the inequality inside of the probability we get

$$\mathbb{P}\left(\sum_{i=1}^n \mathbb{1}_{X_i \leqslant n^{-1/2} \cdot x/\theta_n} \geqslant n/2\right) = \mathbb{P}\left(\sum_{i=1}^n \left(\mathbb{1}_{X_i \leqslant n^{-1/2} \cdot x/\theta_n} - F_i(n^{-1/2} \cdot x/\theta_n)\right) \geqslant -\sqrt{n} \cdot x_n\right) \tag{4.2}$$

where

$$x_n = n^{-1/2} \cdot \left(\sum_{i=1}^n F_i(n^{-1/2} \cdot x/\theta_n) - n/2\right).$$

By Assumption I, F_i's are twice continuously differentiable at 0. Thus, using Taylor expansion for each $F_i(x)$ at 0, x_n could be rewritten as

$$x_n = \frac{1}{\sqrt{n}}(n/2 - n/2) + \frac{x}{\theta_n n} \sum_{i=1}^n F_i'(0) + \frac{x^2}{2\theta_n^2 n^{3/2}} \sum_{i=1}^n F_i''(y_i),$$

where $y_i \in (0, x/(\theta_n \sqrt{n}))$. Recall constants ε, A from Assumption I.3 and constants c, C from (1.10).

Define

$$K_n^2 = \frac{3}{2} \cdot \frac{\sigma_n^2}{\min(1, c)^2} \cdot D_1 \cdot \log n,$$

so that

$$K_n^2/(3D_1) \geqslant \frac{1}{2} \log n, \quad K_n^2/(2\sigma_n^2) \geqslant \frac{1}{2} \log n \tag{4.3}$$

and

$$K_n^2 \leqslant \frac{3C^3}{2c^2} \cdot n^{\frac{1}{4}} \log n \leqslant \frac{3C^3}{2c^3} \cdot \theta_n \cdot n^{\frac{1}{4}} \log n \leqslant \varepsilon \cdot \theta_n \cdot \sqrt{n}$$

holds for

$$n \geqslant N(\varepsilon, c, C), \text{ where } N(\varepsilon, c, C) \text{ is the smallest integer solution to } \frac{n^{1/4}}{\log n} \geqslant \frac{3C^3}{2\varepsilon c^3}. \tag{4.4}$$

Thus, when $|x| \leqslant K_n$ we have $\max_{i \in [n]} |y_i| \leqslant \varepsilon$ and from Assumption I.3 it follows that for all $n \geqslant N(\varepsilon, c, C)$

$$|x_n - x| \leqslant \frac{K_n^2}{2\sqrt{n}} \cdot \frac{A}{\theta_n^2}. \tag{4.5}$$

On the other hand, the left-hand side of the inequality inside of probability in (4.2) is a scaled sum of centered Bernoulli random variables

$$Y_{i,x} := \mathbb{1}_{X_i \leqslant n^{-1/2} \cdot x/\theta_n} - F_i(n^{-1/2} \cdot x/\theta_n), \quad i \in [n].$$

We treat the left bound of the interval from (4.1) similarly. This allows us to rewrite the original quantity of interest in the following way.

$$\mathbb{P}\left(\theta_n \sqrt{n} M_n \leqslant x\right) \in \left[\mathbb{P}\left(\sum_{i=1}^n Y_{i,x} \geqslant 1 - \sqrt{n} \cdot x_n\right), \mathbb{P}\left(\sum_{i=1}^n Y_{i,x} \geqslant -\sqrt{n} \cdot x_n\right)\right]. \tag{4.6}$$

Moreover, for each $x \in \mathbb{R}$ random variables $\{Y_{i,x}\}$ have the same dependency structure as $\{X_i\}$. We will now consider two cases when $|x| > K_n$, in which we apply a concentration inequality, and $|x| \leqslant K_n$, where we use Stein's method for normal approximation.

The Kolmogorov–Smirnov distance between $(\theta_n \sqrt{n}/\sigma_n) M_n$ and the standard normal random variable Z can be approximated in terms of $\{Y_{i,x}\}_{i \leqslant n}$ as follows

$$\begin{aligned} &\mathrm{d_{KS}}\left(\frac{\theta_n \sqrt{n}}{\sigma_n} \cdot M_n\,,\, Z\right) = \sup_{x \in \mathbb{R}} \left|\mathbb{P}\left(\theta_n \sqrt{n} \cdot M_n \leqslant x\right) - \Phi(x/\sigma_n)\right| \\ &\leqslant \sup_{x \in \mathbb{R}} \left\{\left|\mathbb{P}\left(\sum_{i=1}^n Y_{i,x} \geqslant 1 - \sqrt{n} \cdot x_n\right) - \Phi(x/\sigma_n)\right|, \left|\mathbb{P}\left(\sum_{i=1}^n Y_{i,x} \geqslant -\sqrt{n} \cdot x_n\right) - \Phi(x/\sigma_n)\right|\right\}. \end{aligned} \tag{4.7}$$

Focusing on the second term inside of the maximum in (4.7) we rewrite it as

$$\sup_{x \in \mathbb{R}} \left|\mathbb{P}\left(\sum_{i=1}^n Y_{i,x} \geqslant -\sqrt{n} \cdot x_n\right) - \Phi(x/\sigma_n)\right| \leqslant \mathrm{Err}_1 + \mathrm{Err}_2,$$

where

$$\mathrm{Err}_1 = \sup_{|x| \geqslant K_n} \left|\mathbb{P}\left(\sum_{i=1}^n Y_{i,x} \geqslant -\sqrt{n} \cdot x_n\right) - \Phi(x/\sigma_n)\right|, \tag{4.8}$$

$$\text{and} \quad \mathrm{Err}_2 = \sup_{|x| < K_n} \left|\mathbb{P}\left(\sum_{i=1}^n Y_{i,x} \geqslant -\sqrt{n} \cdot x_n\right) - \Phi(x/\sigma_n)\right|. \tag{4.9}$$

Bound on Err_1: We bound the first term by considering two cases: when x_n is negative and when it is positive. When $x_n < 0$ both quantities in (4.8) are negligible. Indeed, by Lemma 3.2 we have the following concentration bound

$$\mathbb{P}\left(\sum_{i=1}^n Y_{i,x} \geqslant -\sqrt{n} \cdot x_n\right) \leqslant \mathbb{P}\left(\sum_{i=1}^n Y_{i,x} \geqslant \sqrt{n} \cdot K_n\right) \lesssim \exp\left(-K_n^2/3D_1\right) \leqslant n^{-1/2}$$

and since $\max\left\{\Phi_{\sigma_n^2}(-K_n), 1 - \Phi_{\sigma_n^2}(K_n)\right\} \leqslant \exp\left(-K_n^2/(2\sigma_n^2)\right) \leqslant n^{-1/2}$ we get that

$$\sup_{x < -K_n} \left|\mathbb{P}\left(\sum_{i=1}^n Y_{i,x} \geqslant -\sqrt{n} \cdot x_n\right) - \Phi(x/\sigma_n)\right| \leqslant n^{-1/2}$$

On the other hand, when $x_n > 0$, both terms are close to 1,

$$\begin{aligned} &\sup_{x > K_n} \left|\mathbb{P}\left(\sum_{i=1}^n Y_{i,x} \geqslant -\sqrt{n} \cdot x_n\right) - \Phi(x/\sigma_n)\right| \\ &\qquad = \sup_{x > K_n} \mathbb{P}\left(\sum_{i=1}^n Y_{i,x} \leqslant -\sqrt{n} \cdot x_n\right) + \Phi(-x/\sigma_n) \lesssim n^{-1/2}. \end{aligned}$$

Bound on Err_2: We now consider the second error term with $|x| \leqslant K_n$. Recall that

$$\sigma_{n,x}^2 = \frac{1}{n} \operatorname{Var}\left(\sum_{i=1}^n Y_{i,x}\right).$$

Then Err_2 can be bounded as follows

$$\mathrm{Err}_2 \leqslant \sup_{|x| \leqslant K_n} \left| \mathbb{P}\left(\sum_{i=1}^n Y_{i,x} \geqslant -\sqrt{n} \cdot x_n\right) - \Phi(x_n/\sigma_{n,x}) \right| \tag{4.10}$$

$$+ \sup_{|x| \leqslant K_n} |\Phi(x_n/\sigma_{n,x}) - \Phi(x_n/\sigma_n)| \tag{4.11}$$

$$+ \sup_{|x| \leqslant K_n} |\Phi(x_n/\sigma_n) - \Phi(x/\sigma_n)|. \tag{4.12}$$

The terms (4.11) and (4.12) are easily bound using the previously established inequalities. First, we established that

$$(4.11) \leqslant \sup_{|x| \leqslant K_n} \frac{\left|\sigma_n^2 - \sigma_{n,x}^2\right|}{\sigma_n^2 - \max\{0, \sigma_n^2 - \sigma_{n,x}^2\}} \leqslant \sup_{|x| \leqslant K_n} \frac{6\sqrt{2}A}{\theta_n} \cdot \frac{D_1 \, |x|}{\sigma_n^2 \sqrt{n}} \lesssim \frac{D_1 \, K_n}{\sqrt{n}} \lesssim D_1^{\frac{3}{2}} \cdot \sqrt{\frac{\log n}{n}}, \tag{4.13}$$

where the first inequality follows from Lemma 3.3 and the second inequality from the variance comparison as in (3.6).

Turning our attention to the term (4.12), we notice that as long as $K_n \leqslant \varepsilon \cdot \theta_n \sqrt{n}$ we can use the bound from (4.5) to derive that

$$(4.12) \leqslant \sup_{|x| \leqslant K_n} \frac{|x_n - x|}{\sigma_n} \leqslant \frac{A}{\theta_n^2} \cdot \frac{K_n^2}{\sigma_n \sqrt{n}} \lesssim D_1 \cdot \frac{\log n}{\sqrt{n}}. \tag{4.14}$$

Finally, it remains to bound (4.10). Since all random variables $|Y_i| \leqslant 1$, by Theorem 3.1 we have a quantitative version of CLT for $\{Y_i\}_{i \geqslant 1}$. Since $Y_{i,x}$ are centered we we can rewrite (4.10) as

$$\sup_{y \in \mathbb{R}} \left| \mathbb{P}\left(\sum_{i=1}^n Y_{i,x} \leqslant \sqrt{n} \cdot y\right) - \Phi(y/\sigma_{n,x}) \right| \lesssim \frac{D_1(D_2 + D_3)}{\sigma_{n,x}^3 \sqrt{n}} \lesssim \frac{D_1(D_2 + D_3)}{\sigma_n^3 \sqrt{n}} \lesssim \frac{D_1(D_2 + D_3)}{\sqrt{n}},$$

where we again used the variance comparison (3.6) to justify the change from $\sigma_{n,x}$ to σ_n for all x such that $|x| \leqslant K_n \leqslant \varepsilon \cdot \theta_n \sqrt{n}$ and then (1.10) to bound σ_n by a universal constant. Putting these bounds together, we conclude that

$$\begin{aligned}
&\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(\sum_{i=1}^n Y_i \geqslant -\sqrt{n} \cdot x_n\right) - \Phi(x/\sigma_n) \right| \\
&\qquad \lesssim n^{-1/2} + D_1^{\frac{3}{2}} \cdot \sqrt{\frac{\log n}{n}} + D_1 \frac{\log n}{\sqrt{n}} + \frac{D_1(D_2 + D_3)}{\sqrt{n}} \\
&\qquad \lesssim D_1 \sqrt{\log n + D_1} \cdot \sqrt{\frac{\log n}{n}} + \frac{\max\{D_1, D_2, D_3\}^2}{\sqrt{n}},
\end{aligned}$$

for $n \geqslant N(\varepsilon, c, C)$. Multiplying the right-hand side by a universal constant that is large enough yields the bound for all n. ■

4.2. *Limiting joint distribution of sample quantiles.*

Proof of Theorem 1.3: Define

$$\overline{\mathbf{Q}}_n := \mathbf{Q}_n - \boldsymbol{\mu}.$$

Since each centered empirical quantile can be rewritten as

$$\overline{Q}_n^{(\alpha_k)} := \sup\{x \in \mathbb{R} : |\{i : X_i \leqslant x\}| \leqslant \lfloor n\alpha_k \rfloor\} - m_k,$$

for any $\boldsymbol{x} = (x_1, x_2, \dots, x_\ell) \in \mathbb{R}^\ell$ we rewrite the joint distribution of quantiles in the following way

$$\begin{aligned}
&\mathbb{P}(\Theta_n \sqrt{n} \cdot (\mathbf{Q}_n - \boldsymbol{\mu}) \preceq \boldsymbol{x}) \\
&\quad = \mathbb{P}\left(\theta_1 \sqrt{n}\, \overline{Q}_n^{(\alpha_1)} \leqslant x_1, \dots, \theta_\ell \sqrt{n}\, \overline{Q}_n^{(\alpha_\ell)} \leqslant x_\ell\right) \\
&\quad \in \left[\mathbb{P}\left(\sum_{i=1}^n \mathbb{1}_{X_i - m_k \leqslant (\theta_n(m_k)\sqrt{n})^{-1} x_k} \geqslant \lfloor n\alpha_k \rfloor + 1, \text{ for all } k \in [\ell]\right),\right. \qquad (4.15)\\
&\qquad\qquad \left.\mathbb{P}\left(\sum_{i=1}^n \mathbb{1}_{X_i - m_k \leqslant (\theta_n(m_k)\sqrt{n})^{-1} x_k} \geqslant \lfloor n\alpha_k \rfloor, \text{ for all } k \in [\ell]\right)\right].
\end{aligned}$$

Following the same idea as in the proof of Theorem 1.2 we treat each of the probabilities from (4.15) similarly but separately. In particular, by subtraction corresponding values of CDF functions F_i on both sides of the inequalities inside of the probabilities and then using Taylor expansion, we get an analogs equation to (4.6), namely

$$\mathbb{P}\left(\Theta_n \cdot (\mathbf{Q}_n - \boldsymbol{\mu}) \preceq \boldsymbol{x}\right) \in \left[\mathbb{P}\left(\sum_{i=1}^n \boldsymbol{Y}_{i,\boldsymbol{x}} \succeq \mathbf{1} - \sqrt{n} \cdot \boldsymbol{x}_n\right), \mathbb{P}\left(\sum_{i=1}^n \boldsymbol{Y}_{i,\boldsymbol{x}} \succeq -\sqrt{n} \cdot \boldsymbol{x}_n\right)\right], \qquad (4.16)$$

where $\mathbf{1} := (1, 1, \dots, 1)$ and

$$\boldsymbol{x}_n := \frac{1}{\sqrt{n}} \left(\sum_{i=1}^n F_i\left(m_k + n^{-1/2} \cdot x_k/\theta_n(m_k)\right) - \lfloor n\alpha_k \rfloor\right)_{k=1}^{\ell}.$$

Define K_n in a similar fashion to the univariate case replacing σ_n by $\sigma_{\max}$

$$K_n^2 = \frac{3}{2} \cdot \frac{\sigma_{\max}^2}{\min(1,c)^2} \cdot D_1 \cdot \log n,$$

then if $\|\boldsymbol{x}\|_\infty \leqslant K_n$ we have that $\boldsymbol{x}_n - \boldsymbol{x}$ satisfies the following inequality for all for all $n \geqslant N(\varepsilon, c, C)$, defined in (4.4),

$$\|\boldsymbol{x}_n - \boldsymbol{x}\|_2 \leqslant \frac{\|\boldsymbol{x}\|_2^2}{2\sqrt{n}} \cdot \frac{A}{\theta_{\min}^2} \leqslant \frac{\ell\, \|\boldsymbol{x}\|_\infty^2}{2\sqrt{n}} \cdot \frac{A}{\theta_{\min}^2}. \qquad (4.17)$$

For some function K_n, each of the terms on the right-hand side of (4.16) are treated similarly. For simplicity, we present the bound on the right bound of the interval.

$$\sup_{\boldsymbol{x}} \left|\mathbb{P}\left(\sum_{i=1}^n \boldsymbol{Y}_{i,\boldsymbol{x}} \succeq -\sqrt{n} \cdot \boldsymbol{x}_n\right) - \mathbb{P}\left(\Sigma_n^{1/2} \boldsymbol{Z} \succeq -\boldsymbol{x}\right)\right| \leqslant \widehat{\mathrm{Err}}_1 + \widehat{\mathrm{Err}}_2,$$

where

$$\widehat{\mathrm{Err}}_1 = \sup_{\|\boldsymbol{x}\|_\infty \geqslant K_n} \left|\mathbb{P}\left(\sum_{i=1}^n \boldsymbol{Y}_{i,\boldsymbol{x}} \succeq -\sqrt{n} \cdot \boldsymbol{x}_n\right) - \mathbb{P}\left(\Sigma_n^{1/2} \boldsymbol{Z} \succeq -\boldsymbol{x}\right)\right| \qquad (4.18)$$

$$\text{and} \quad \widehat{\mathrm{Err}}_2 = \sup_{\|\boldsymbol{x}\|_\infty < K_n} \left|\mathbb{P}\left(\sum_{i=1}^n \boldsymbol{Y}_{i,\boldsymbol{x}} \succeq -\sqrt{n} \cdot \boldsymbol{x}_n\right) - \mathbb{P}\left(\Sigma_n^{1/2} \boldsymbol{Z} \succeq -\boldsymbol{x}\right)\right| \qquad (4.19)$$

Bound on $\widehat{\mathrm{Err}}_1$. Recall that $\sigma_{\max}^2 := \max_{i \in [\ell]} \Sigma_n(i,i)$.

In the first term, we bound similarly to the way we bounded (4.8) in the one-dimensional case. If $x_{\min} := \min_k x_k \leqslant -K_n$ then using a union bound and Lemma 3.2,

$$\sup_{x_{\min}\leqslant -K_n}\left|\mathbb{P}\left(\sum_{i=1}^{n}\boldsymbol{Y}_{i,\boldsymbol{x}} \succeq -\sqrt{n}\cdot\boldsymbol{x}_n\right) - \mathbb{P}\left(\Sigma_n^{1/2}\boldsymbol{Z} \succeq -\boldsymbol{x}\right)\right| \leqslant 2\ell\exp\left(-K_n^2/(3D_1 \vee 2\sigma_{\max}^2)\right).$$

On the other hand if $x_{\min} > -K_n$ and $x_{\max} := \max_k x_k \geqslant K_n$, then we divide coordinates into two sets $A := \{k : x_k \in (-K_n, K_n)\}$ and $B := \{k : x_k \leqslant -K_n\}$. For simplicity let

$$E_k := \left\{\sum_{i=1}^{n}(\boldsymbol{Y}_{i,\boldsymbol{x}})_k \geqslant -\sqrt{n}\cdot(\boldsymbol{x}_n)_k\right\} \quad \text{and} \quad F_k := \left\{\Sigma_n^{1/2} Z_k \geqslant -(x)_k\right\}.$$

Thus, for all $\boldsymbol{x}$ such that $x_{\min} > -K_n$ and $x_{\max} \geqslant K_n$ we have

$$\begin{aligned}
&\left|\mathbb{P}\left(\sum_{i=1}^{n}\boldsymbol{Y}_{i,\boldsymbol{x}} \succeq -\sqrt{n}\cdot\boldsymbol{x}_n\right) - \mathbb{P}\left(\Sigma_n^{1/2}\boldsymbol{Z} \succeq -\boldsymbol{x}\right)\right| \\
&= \left|\mathbb{P}\left(\bigcap_{k\in A} E_k \bigcap_{k'\in B} E_{k'}\right) - \mathbb{P}\left(\bigcap_{k\in A} F_k \bigcap_{k'\in B} F_{k'}\right)\right| \\
&\leqslant \left|\mathbb{P}\left(\bigcap_{k=1}^{\ell} E_k\right) - \mathbb{P}\left(\bigcap_{k=1}^{\ell} F_k\right)\right| &(4.20)\\
&\quad + \left|\mathbb{P}\left(\bigcap_{k\in A} E_k \left(\bigcap_{k'\in B} E_{k'}\right)^c\right) - \mathbb{P}\left(\bigcap_{k\in A} F_k \left(\bigcap_{k'\in B} F_{k'}\right)^c\right)\right|, &(4.21)
\end{aligned}$$

where in (4.21), we switched to the compliment events. This enables the application of a concentration inequality

$$\begin{aligned}
(4.21) &\leqslant \left|\mathbb{P}\left(\bigcup_{k\in B}\left\{\sum_{i=1}^{n}\boldsymbol{Y}_{i,x_k} < -\sqrt{n}\cdot(x_n)_k\right\}\right) - \mathbb{P}\left(\bigcup_{k\in B}\left\{\Sigma_n^{1/2} Z_k < -(x)_k\right\}\right)\right| \\
&\lesssim 2\ell\exp\left(-K_n^2/(3D_1 \vee 2\sigma_{\max}^2)\right) \lesssim \ell n^{-1/2}.
\end{aligned}$$

Finally, the remaining term (4.20) can be bounded similarly to the term $\widehat{\mathrm{Err}}_2$ as presented below.

Bound on $\widehat{\mathrm{Err}}_2$. We rewrite $\widehat{\mathrm{Err}}_2$ in a similar way to how we treated Err_2 in the univariate case

$$\begin{aligned}
\widehat{\mathrm{Err}}_2 = &\sup_{\|\boldsymbol{x}\|_\infty\leqslant K_n}\left|\mathbb{P}\left(\sum_{i=1}^{n}\boldsymbol{Y}_{i,\boldsymbol{x}} \succeq -\sqrt{n}\cdot\boldsymbol{x}_n\right) - \mathbb{P}\left(\Sigma_{n,\boldsymbol{x}}^{1/2}\boldsymbol{Z} \succeq -\boldsymbol{x}_n\right)\right| &(4.22)\\
&+ \sup_{\|\boldsymbol{x}\|_\infty\leqslant K_n}\left|\mathbb{P}\left(\Sigma_{n,\boldsymbol{x}}^{1/2}\boldsymbol{Z} \succeq -\boldsymbol{x}_n\right) - \mathbb{P}\left(\Sigma_n^{1/2}\boldsymbol{Z} \succeq -\boldsymbol{x}_n\right)\right| &(4.23)\\
&+ \sup_{\|\boldsymbol{x}\|_\infty\leqslant K_n}\left|\mathbb{P}\left(\Sigma_n^{1/2}\boldsymbol{Z} \succeq -\boldsymbol{x}_n\right) - \mathbb{P}\left(\Sigma_n^{1/2}\boldsymbol{Z} \succeq -\boldsymbol{x}\right)\right|. &(4.24)
\end{aligned}$$

Applying multivariate CLT as in Theorem 3.1 we get

$$
\begin{aligned}
(4.22) &\leqslant \sup_{\|\boldsymbol{x}\|_\infty \leqslant K_n} \sup_{A \in \mathcal{C}} \left| \mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{Y}_{i,\boldsymbol{x}} \in A \right) - \mathbb{P}\left(\Sigma_{n,\boldsymbol{x}}^{1/2} \boldsymbol{Z} \in A \right) \right| \\
&\lesssim \frac{\ell^{1/4}}{\sqrt{n}} \left\| \Sigma_{n,\boldsymbol{x}}^{-1/2} \right\|_{\text{op}}^3 D_1 \left(D_2 + D_3 \ell^{-1} \right) \lesssim \frac{\ell^{1/4}}{\sqrt{n}} \left\| \Sigma_n^{-1/2} \right\|_{\text{op}}^3 D_1 \left(D_2 + D_3 \ell^{-1} \right) \\
&\lesssim \frac{\ell^{1/4}}{\sqrt{n}} D_1 \left(D_2 + D_3 \ell^{-1} \right)
\end{aligned}
$$

To bound the error term (4.23) we derive the bound on the total variation distance and invoke Lemma 3.7, which states that

$$
\begin{aligned}
(4.23) &\leqslant \sup_{\|\boldsymbol{x}\|_\infty \leqslant K_n} \max\{ \left\| \Sigma_n^{-1} \right\|_{\text{op}}, \left\| \Sigma_{n,\boldsymbol{x}}^{-1} \right\|_{\text{op}} \} \cdot \left\| \Sigma_{n,\boldsymbol{x}} - \Sigma_n \right\|_{\text{F}} \\
&\leqslant \sup_{\|\boldsymbol{x}\|_\infty \leqslant K_n} \frac{6\sqrt{2}A}{\theta_{\min}} \cdot \ell \cdot \left\| \Sigma_n^{-1} \right\|_{\text{op}} \cdot \frac{D_1 \|\boldsymbol{x}\|_\infty}{\sqrt{n}} \lesssim \frac{D_1 K_n}{\sqrt{n}} \lesssim D_1^{\frac{3}{2}} \sqrt{\frac{\log n}{n}}.
\end{aligned}
$$

Finally, the term (4.24) we bound coordinate-wise is as in the univariate case (see (4.14)).

$$
(4.24) \leqslant \frac{\ell A}{\theta_{\min}^2} \cdot \left\| \Sigma_n^{-1/2} \right\|_{\text{op}} \frac{K_n^2}{\sqrt{n}} \lesssim D_1 \frac{\log n}{\sqrt{n}}.
$$

Putting this all together and optimizing over K_n, similarly to how we did it in the univariate case yields

$$
\begin{aligned}
&\sup_{\boldsymbol{x}} \left| \mathbb{P}\left(\sum_{i=1}^n \boldsymbol{Y}_{i,\boldsymbol{x}} \succeq -\sqrt{n} \cdot \boldsymbol{x}_n \right) - \mathbb{P}\left(\Sigma_n^{1/2} \boldsymbol{Z} \succeq -\boldsymbol{x}_n \right) \right| \\
&\qquad \lesssim \ell n^{-1/2} + D_1^{\frac{3}{2}} \cdot \sqrt{\frac{\log n}{n}} + D_1 \frac{\log n}{\sqrt{n}} + \frac{\ell^{1/4}}{\sqrt{n}} D_1 \left(D_2 + D_3 \ell^{-1} \right),
\end{aligned}
$$

which holds for $n \geqslant N(\varepsilon, c, C)$. Multiplying the right-hand side by a universal constant that is large enough yields the bound for all n. ■

5. Optimal rate of convergence

In the situation of Theorem 1.2, suppose that D_1, D_2, and D_3 are also bounded by constants, Theorem 1.2 gives the bound of order $\log n / \sqrt{n}$. In this section, we analyze the i.i.d. case to motivate the following conjecture about the optimal rate.

Conjecture 5.1. In the situation of Theorem 1.2, if $\max_{i \in \{1,2,3\}} \{D_i\} < C$ for some $C > 0$, then

$$
\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(\theta_n \sqrt{n} \cdot M_n \leqslant \sigma_n x \right) - \Phi(x) \right| \approx n^{-1/2}.
$$

Lemma 5.2. *Let $n = 2m + 1$ for some integer m. Suppose $(X_i)_{i \in [n]}$ are i.i.d. continuous random variables with CDF $F(x)$ that satisfy Assumption I.1–2 with $\alpha = 1/2$ and $m_\alpha = 0$. Assume that $F''(0)$ exists. Then*

$$
n^{1/2} \cdot \sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(\theta \sqrt{2m} \cdot M_n \leqslant x/2 \right) - \Phi(x) \right| \to \frac{|F''(0)|}{4F'(0)^2} \cdot \sup_{x \in \mathbb{R}} x^2 \phi(x) = \frac{1}{\sqrt{8\pi e}} \cdot \frac{|F''(0)|}{F'(0)^2}.
$$

Recall that when X_i are uncorrelated we have that $\sigma \equiv 1/2$ and $\theta = F'(0)$.

Proof: By definition of the sample median and independence of X_i's, we have

$$\begin{aligned}\mathbb{P}\left(2\theta M_n \leqslant x/\sqrt{2m}\right) &= (2m+1)\binom{2m}{m}\int_{-\infty}^{x/(2\theta\sqrt{2m})}(1-F(t))^m F(t)^m\,dF(t)\\ &= \frac{2m+1}{2^{2m}}\binom{2m}{m}\int_{-\infty}^{x/(2\theta\sqrt{2m})}(1-(2F(t)-1)^2)^m\,dF(t)\\ &= \frac{2m+1}{\sqrt{2m}\cdot 2^{2m+1}}\binom{2m}{m}\int_{-\sqrt{2m}}^{x_m}(1-t^2/2m)^m\,dt, \end{aligned}\tag{5.1}$$

where $x_m = 2\sqrt{2m}(F(x/(2\theta\sqrt{2m}))-F(0))$. Here, in the last equality, we use the change of variable $t = 2\sqrt{2m}\cdot(F(t)-1/2)$. Define

$$a_m = \sqrt{2\pi}\cdot\frac{2m+1}{\sqrt{2m}\cdot 2^{2m+1}}\binom{2m}{m} = 1+\frac{3}{8m}+O(m^{-2}),$$

where the last equality follows by Stirling's approximation. Thus, we have

$$(5.1) = a_m\int_{-\sqrt{2m}}^{x_m}\left(g(t^2/2m)\right)^m\phi(t)\,dt$$

where $g(s) := (1-s)\exp(s)$. Notice that for all $|s|\leqslant 1$, we have

$$1-s^2 \leqslant g(s) \leqslant 1-s^2/4.$$

Thus we get,

$$\begin{aligned}\mathbb{P}\left(\theta M_n \leqslant x/\sqrt{2m}\right)-\Phi(x) &= \int_{-\sqrt{2m}}^{x_m}\left(g(t^2/2m)^m-1\right)\phi(t)\,dt+\int_x^{x_m}\phi(t)dt+O(1/m)\\ &\approx \int_{-\sqrt{2m}}^{x_m}\frac{t^4}{2m}\phi(t)\,dt+\phi(x)(x_m-x)+O(1/m)\\ &= O(1/m)+\phi(x)(x_m-x).\end{aligned}$$

Recall that $x_m = 2\sqrt{2m}\left(F\left(x/(2\theta\sqrt{2m})\right)-F(0)\right)$. Hence, using Taylor expansion, we have,

$$x_m - x = 2\sqrt{2m}\left(\frac{x}{2\theta\sqrt{2m}}\cdot F'(0)+\frac{x^2}{16\theta^2 m}F''(0)+R\right)-x = \frac{F''(0)}{4F'(0)^2}\cdot\frac{x^2}{\sqrt{n}}(1+o(1)),$$

where $R = O(1/n)$. This completes the proof. ■

Acknowledgements

We would like to thank Sabyasachi Chatterjee for many insightful conversations. We thank the anonymous referees for their careful reading and insightful comments, which resulted in an improved presentation of the article.

References

Bahadur, R. R. A note on quantiles in large samples. *Ann. Math. Statist.*, **37**, 577–580 (1966). MR189095.

Baldi, P., Rinott, Y., and Stein, C. A normal approximation for the number of local maxima of a random function on a graph. In *Probability, statistics, and mathematics*, pp. 59–81. Academic Press, Boston, MA (1989). MR1031278.

Barbour, A. D., Karoński, M., and Ruciński, A. A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B*, **47** (2), 125–145 (1989). MR1047781.

Brockwell, P. J. and Davis, R. A. *Time series: theory and methods*. Springer Series in Statistics. Springer, New York (2006). ISBN 978-1-4419-0319-8; 1-4419-0319-8. MR2839251.

Chen, L. H. Y., Goldstein, L., and Shao, Q.-M. *Normal approximation by Stein's method*. Probability and its Applications (New York). Springer, Heidelberg (2011). ISBN 978-3-642-15006-7. MR2732624.

Chen, L. H. Y. and Röllin, A. Stein couplings for normal approximation. *ArXiv Mathematics e-prints* (2010). arXiv: 1003.6039v2.

Chen, L. H. Y. and Shao, Q.-M. Normal approximation under local dependence. *Ann. Probab.*, **32** (3A), 1985–2028 (2004). MR2073183.

David, H. A. and Nagaraja, H. N. *Order statistics*. Wiley Series in Probability and Statistics. Wiley-Interscience, Hoboken, NJ, third edition (2003). ISBN 0-471-38926-9. MR1994955.

Dembińska, A. Asymptotic behavior of central order statistics from stationary processes. *Stochastic Process. Appl.*, **124** (1), 348–372 (2014). MR3131297.

Devroye, L., Mehrabian, A., and Reddad, T. The total variation distance between high-dimensional Gaussians with the same mean. *ArXiv Mathematics e-prints* (2018). arXiv: 1810.08693.

Diaconis, P. and Holmes, S., editors. *Stein's method: expository lectures and applications*, volume 46 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*. Institute of Mathematical Statistics, Beachwood, OH (2004). ISBN 0-940600-62-5. MR2118599.

Dutta, K. and Sen, P. K. On the Bahadur representation of sample quantiles in some stationary multivariate autoregressive processes. *J. Multivariate Anal.*, **1** (2), 186–198 (1971). MR301881.

Fang, X. A multivariate CLT for bounded decomposable random vectors with the best known rate. *J. Theoret. Probab.*, **29** (4), 1510–1523 (2016). MR3571252.

Fang, X. and Koike, Y. High-dimensional central limit theorems by Stein's method. *Ann. Appl. Probab.*, **31** (4), 1660–1686 (2021). MR4312842.

Goldstein, L. Zero bias enhanced Stein couplings. *Electron. Commun. Probab.*, **27**, Paper No. 62, 13 (2022). MR4529630.

Goldstein, L. and Reinert, G. Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.*, **7** (4), 935–952 (1997). MR1484792.

Goldstein, L. and Rinott, Y. Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.*, **33** (1), 1–17 (1996). MR1371949.

Hald, A. *A history of mathematical statistics from 1750 to 1930*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York (1998). ISBN 0-471-17912-4. MR1619032.

Hesse, C. H. A Bahadur-type representation for empirical quantiles of a large class of stationary, possibly infinite-variance, linear processes. *Ann. Statist.*, **18** (3), 1188–1202 (1990). MR1062705.

Janson, S. Large deviations for sums of partly dependent random variables. *Random Structures Algorithms*, **24** (3), 234–248 (2004). MR2068873.

Kiefer, J. On Bahadur's representation of sample quantiles. *Ann. Math. Statist.*, **38**, 1323–1342 (1967). MR217844.

Ma, Y., Genton, M. G., and Parzen, E. Asymptotic properties of sample quantiles of discrete distributions. *Ann. Inst. Statist. Math.*, **63** (2), 227–243 (2011). MR2780484.

Nourdin, I. and Peccati, G. *Normal approximations with Malliavin calculus. From Stein's method to universality*, volume 192 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge (2012). ISBN 978-1-107-01777-1. MR2962301.

Nourdin, I., Peccati, G., and Réveillac, A. Multivariate normal approximation using Stein's method and Malliavin calculus. *Ann. Inst. Henri Poincaré Probab. Stat.*, **46** (1), 45–58 (2010). MR2641769.

Paulin, D. Concentration inequalities in locally dependent spaces. *ArXiv Mathematics e-prints* (2012). arXiv: 1212.2013.

Pollard, D. Lecture notes in Probability Theory with Applications (2014). `http://www.stat.yale.edu/~pollard/Courses/241.fall2014/notes2014/CLT.pdf`.

Rinott, Y. On normal approximation rates for certain sums of dependent random variables. *J. Comput. Appl. Math.*, **55** (2), 135–143 (1994). MR1327369.

Rinott, Y. and Rotar, V. A multivariate CLT for local dependence with $n^{-1/2}\log n$ rate and applications to multivariate graph related statistics. *J. Multivariate Anal.*, **56** (2), 333–350 (1996). MR1379533.

Ross, N. Fundamentals of Stein's method. *Probab. Surv.*, **8**, 210–293 (2011). MR2861132.

Sen, P. K. Asymptotic normality of sample quantiles for m-dependent processes. *Ann. Math. Statist.*, **39**, 1724–1730 (1968). MR232522.

Sen, P. K. On the Bahadur representation of sample quantiles for sequences of ϕ-mixing random variables. *J. Multivariate Anal.*, **2**, 77–95 (1972). MR303581.

Stein, C. 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 (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pp. 583–602. Univ. California Press, Berkeley, CA (1972). MR402873.

Strasser, H. *Mathematical Theory of Statistics. Statistical Experiments and Asymptotic Decision Theory.* De Gruyter, Berlin, New York (1985). ISBN 9783110850826. DOI: 10.1515/9783110850826.

van der Vaart, A. W. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press, Cambridge (1998). ISBN 0-521-49603-9; 0-521-78450-6. MR1652247.