

A refined Cramér-type moderate deviation for sums of local statistics

XIAO FANG^{1,*}, LI LUO^{1,†} and QI-MAN SHAO^{1,2}

¹Department of Statistics, The Chinese University of Hong Kong, Hong Kong.
E-mail: *xfang@sta.cuhk.edu.hk; †luoli@link.cuhk.edu.hk

²Department of Statistics and Data Science, Southern University of Science and Technology, China.
E-mail: qmshao@sta.cuhk.edu.hk

We prove a refined Cramér-type moderate deviation result by taking into account of the skewness in normal approximation for sums of local statistics of independent random variables. We apply the main result to k -runs, U-statistics and subgraph counts in the Erdős–Rényi random graph. To prove our main result, we develop exponential concentration inequalities and higher-order tail probability expansions via Stein’s method.

Keywords: Cramér-type moderate deviation; Erdős–Rényi random graph; k -runs; local dependence; skewness correction; Stein’s method; U-statistic

1. Introduction

Moderate deviations date back to Cramér [17] who obtained expansions for tail probabilities for sums of independent random variables about the normal distribution. Let $W_n = \sum_{i=1}^n X_i / \sqrt{n}$ where $\{X_1, X_2, \dots\}$ are independent and identically distributed (i.i.d.) with $EX_1 = 0$, $EX_1^2 = 1$, and $Ee^{t_0|X_1|} < \infty$ for a constant $t_0 > 0$. It is known that (cf. Petrov [21], Chapter 8, Theorem 1)

$$\left| \frac{P(W_n > x)}{1 - \Phi(x)} - 1 \right| \leq \frac{C(1 + x^3)}{\sqrt{n}} \quad \text{for } 0 \leq x \leq C_0 n^{1/6} \quad (1.1)$$

and

$$\left| \frac{P(W_n > x)}{(1 - \Phi(x))e^{\gamma x^3/6}} - 1 \right| \leq C \left(\frac{1 + x}{\sqrt{n}} + \frac{x^4}{n} \right) \quad \text{for } 0 \leq x \leq C_0 n^{1/4}, \quad (1.2)$$

where C_0 is any fixed constant, Φ denotes the standard normal distribution function, $\gamma = EW_n^3 = EX_1^3 / \sqrt{n}$ and C is a positive constant depending only on t_0 , C_0 , and the moment generating function $Ee^{t_0|X_1|}$. The range $0 \leq x = o(n^{1/6})$ ($0 \leq x = o(n^{1/4})$, resp.) for the relative error in (1.1) ((1.2), resp.) to vanish is optimal. We refer to results such as (1.1) as Cramér-type moderate deviations. We refer to the modification of the normal distribution function in (1.2) as skewness correction, and results such as (1.2) as *refined* Cramér-type moderate deviations.

In statistical inference problems, it is crucial to compute the p -value of a test statistic, which is defined to be the probability that the statistic is greater than or equal to the actual observed value under the null hypothesis. Such a p -value is typically small. When the test statistic is

asymptotically normal, Cramér-type moderate deviations provide theoretical justification of using the normal tail probability to approximate the p -value with a small relative error. When the sample size is relatively small and the underlying distribution is not symmetric, it has been observed empirically that making a skewness correction can greatly improve the accuracy of the tail probability approximation. See, for example, Tu and Siegmund [30] and Tang and Siegmund [29] for sums of independent random variables and Chen and Zhang [10] for dependent random variables. See also Table 1 below. Refined Cramér-type moderate deviations provide theoretical justification of such a skewness correction.

The main purpose of this paper is to prove a refined Cramér-type moderate deviation for sums of local statistics of independent random variables. For a positive integer m , let $\{X_1, \dots, X_m\}$ be a sequence of independent random variables in a general space. Let

$$W = \sum_{i=1}^n \xi_i,$$

where for each $i \in \{1, \dots, n\}$, ξ_i is a real-valued function of a small subset of $\{X_1, \dots, X_m\}$.

Absolute-error bounds in normal approximation for such W are well studied in the literature. See, for example, Chen and Shao [14] for results under a more general local dependence setting. However, the accuracy of tail probability approximations for such W is less well understood. Recently, Zhang [31] considered Cramér-type moderate deviations as in (1.1) for such W .

Assume $EW = 0$, $EW^2 = 1$ and let $\gamma = EW^3$. Our main result is a general relative-error bound (cf. (2.3)) for

$$\left| \frac{P(W > x)}{(1 - \Phi(x))e^{\gamma x^3/6}} - 1 \right|$$

under certain boundedness conditions (cf. (2.1)). For standardized sums of i.i.d., bounded random variables, our bound vanishes for the optimal range $0 \leq x = o(n^{1/4})$. We apply our main result to k -runs, U-statistics, and subgraph counts in the Erdős–Rényi random graph. In each application, our bound vanishes for presumably the optimal range of x in terms of the system size.

Our proof is based on Stein's method, which was introduced by Stein [28] for normal approximation. We refer to Chen, Goldstein and Shao [13] for an introduction to the method and a survey of its recent developments. Chen, Fang and Shao [11] developed the method to prove Cramér-type moderate deviation results in normal approximation without skewness correction for dependent random variables under a boundedness condition. Chen, Fang and Shao [12] and Shao, Zhang and Zhang [27] considered Poisson approximation and nonnormal approximations, respectively. Zhang [31] refined the results in Chen, Fang and Shao [11] by relaxing the boundedness condition. Braverman and Dai [7], Chapter 4, obtained a Cramér-type moderate deviation result in a higher-order approximation for the Erlang-C queuing model. His proof relies heavily on explicit expressions of certain conditional expectations in the model. To prove our general bound, we develop Stein's method for exponential concentration inequalities (cf. Proposition 3.2) and for higher-order tail probability expansions. For the latter, we use $P(Z_\gamma > x)$ in place of $(1 - \Phi(x))e^{\gamma x^3/6}$ for an intermediate approximation, where Z_γ follows a suitable standardized Poisson distribution.

Related results are available in the literature. (a) Asymptotic expansions in the central limit theorem have been extensively studied. See, for example, Petrov [21] for the classical Edgeworth expansion and Barbour [1] and Rinott and Rotar [22] for related expansions using Stein’s method. These expansions require either a continuity condition on the random variable or a smoothness condition on certain test functions. The $O(1/\sqrt{n})$ rate of convergence in the absolute-error bound for normal approximation for sums of n independent discrete random variables generally can not be improved. Nevertheless, (1.2), as well as our main result, shows that it is still possible to improve the accuracy in terms of the relative error in tail probability approximations using an appropriate expansion. (b) In the proof of our main result, we use a standardized Poisson distribution for an intermediate approximation. Translated Poisson distributions have been proposed as alternatives to normal distributions to approximate lattice random variables in the total variation distance. See, for example, Röllin [24], Röllin [25], Barbour, Luczak and Xia [4], Barbour, Luczak and Xia [5], and Barbour and Xia [6]. Instead of matching the support of random variables as in these results, we use standardized Poisson distributions to correct for skewness. See Rio [23] for a similar use of standardized Poisson distributions.

The remainder of this paper is organized as follows. In Section 2, we state the general relative-error bound in normal approximation with skewness correction for sums of local statistics of independent random variables and discuss applications to k -runs, U-statistics, and subgraph counts in the Erdős–Rényi random graph. In Section 3, we prove an exponential concentration inequality, which is crucial to the proof of the general bound. In Section 4, we prove the general bound.

2. Main results

2.1. A general relative-error bound

For a positive integer N , denote $[N] := \{1, \dots, N\}$. Let m and n be positive integers. Let $\{X_\alpha : \alpha \in [m]\}$ be a sequence of independent random variables. Let $W = \sum_{i=1}^n \xi_i$, where each ξ_i is a function of $\{X_\alpha : \alpha \in \mathcal{I}_i\}$ for some $\mathcal{I}_i \subset [m]$. For $\alpha \in [m]$, let $N_\alpha = \{i \in [n] : \alpha \in \mathcal{I}_i\}$.

Theorem 2.1. *Under the above setting, assume that $E\xi_i = 0$ for each $i \in [n]$ and $\text{Var}(W) = 1$. Assume further that*

$$|\xi_i| \leq \delta, \quad |\mathcal{I}_i| \leq s, \quad |N_\alpha| \leq d, \tag{2.1}$$

where $|\cdot|$ denotes the cardinality when applied to a set. Denote $\gamma := EW^3$. Let C_0 be any fixed constant. For

$$0 \leq x \leq C_0(mns^4d^4\delta^5)^{-1/2}, \tag{2.2}$$

we have

$$\left| \frac{P(W > x)}{(1 - \Phi(x))e^{\gamma x^3/6}} - 1 \right| \leq Cmns^4d^4\delta^5(1 + x^2), \tag{2.3}$$

where C is a positive constant depending only on C_0 .

Clearly, applying the above result to $-W$ yields

$$\left| \frac{P(W < -x)}{\Phi(-x)e^{-\gamma x^3/6}} - 1 \right| \leq C m n s^4 d^4 \delta^5 (1 + x^2).$$

To illustrate that the range of x for the relative error in our approximation to vanish is correct, we first consider the standardized sums of i.i.d., bounded random variables. Let X_1, X_2, \dots be i.i.d. with $EX_i = 0, EX_i^2 = 1, |X_i| \leq C_1 < \infty$. For an integer $n \geq 1$, let $\xi_i = X_i/\sqrt{n}$ for each $i \in [n]$ and let $W = \sum_{i=1}^n \xi_i$. This satisfies the assumptions in Theorem 2.1 with

$$m = n, \quad \delta = \frac{C_1}{\sqrt{n}}, \quad s = 1, \quad d = 1, \quad \gamma = \frac{EX_1^3}{\sqrt{n}}.$$

Let C_0 be any fixed constant. From (2.3), we have, for $0 \leq x \leq C_0 n^{1/4}$,

$$\left| \frac{P(W > x)}{(1 - \Phi(x))e^{\gamma x^3/6}} - 1 \right| \leq \frac{C}{\sqrt{n}}(1 + x^2), \tag{2.4}$$

where C is a positive constant depending only on C_0 and C_1 .

The roadmap for proving Theorem 2.1. First, we prove an absolute-error bound in normal approximation for W using standard techniques in Stein’s method and argue that (2.3) holds for bounded x . Next, to circumvent the problem that $(1 - \Phi(x))e^{\gamma x^3/6}$ is in general not a probability distribution function, we use a standardized Poisson distribution, which has right-tail probabilities equivalent to $(1 - \Phi(x))e^{\gamma x^3/6}$ up to the range of x of interest (cf. (4.4)), for an intermediate approximation. In applying Stein’s method for Poisson approximations to our problem, we face with two difficulties: (1) the random variable of interest W may not have the same support as the Poisson random variable; (2) we need an error bound that depends on x in an optimal way. We overcome the first difficulty by extending the solution to the Stein equation for the standardized Poisson distribution to the whole real line and control the error introduced by the extension. To overcome the second difficulty, we need optimal upper bounds for $P(W > x), P(x \leq W \leq x + \varepsilon)$ and related quantities. We do it by proving moment generating function bounds and concentration inequalities for W in Section 3. The local dependence structure assumed above Theorem 2.1 is crucial in obtaining these bounds.

Remark 2.1. For any integer $k \geq 3$, corrections to the normal distribution function can be formally expressed as

$$(1 - \Phi(x)) \exp\left(\sum_{v=1}^{k-2} q_v x^{v+2}\right),$$

where q_v is a constant that depends on the cumulants of W up to the $(v + 2)$ th order. For a sum of n i.i.d. random variables with finite moment generating functions, this approximation is accurate up to $x = o(n^{\frac{k-1}{2(k+1)}})$ (cf. Petrov [21], Chapter 8, Theorem 1). We only considered $k = 3$ in this paper. One obstacle to obtaining a complete proof for even higher-order expansions using our

approach is that the exponential concentration inequality (cf. Proposition 3.2) is only useful in the range $x = o(n^{1/4})$.

2.2. Applications

In this subsection, we apply Theorem 2.1 to three examples: k -runs, U-statistics and subgraph counts in the Erdős–Rényi random graph. All the examples fit exactly into the local dependence structure stated in Theorem 2.1. There are, however, examples possessing a local dependence structure but may not fit into our framework. For example, the so-called m -dependent sequence $\{\xi_i, 1 \leq i \leq n\}$ only requires $\{\xi_k : k \leq j\}$ to be independent of $\{\xi_k : k \geq j + m + 1\}$ for any j ; the ξ_s may not be functions of an underlying independent sequence.

2.2.1. k -Runs

Let $n > k > 1$ be integers. Let $p \in (0, 1)$. Let X_1, \dots, X_n be i.i.d. and $P(X_i = 1) = 1 - P(X_i = 0) = p$. Let

$$W = \sum_{i=1}^n \xi_i, \quad \xi_i = \frac{X_i X_{i+1} \cdots X_{i+k-1} - p^k}{\sigma},$$

where σ is the normalizing constant such that $\text{Var}(W) = 1$, and $X_{n+i} := X_i$ for $i \geq 1$. It satisfies the assumptions in Theorem 2.1 with

$$m = n, \quad \delta = \frac{1}{\sigma}, \quad s = k, \quad d = k.$$

Therefore, we obtain the following.

Proposition 2.1. *Let $\gamma = EW^3$ with the W above. Let C_0 be any fixed constant. We have, for $0 \leq x \leq C_0(\sigma^5/n^2k^8)^{1/2}$,*

$$\max \left\{ \left| \frac{P(W < -x)}{\Phi(-x)e^{-\gamma x^3/6}} - 1 \right|, \left| \frac{P(W > x)}{(1 - \Phi(x))e^{\gamma x^3/6}} - 1 \right| \right\} \leq \frac{Cn^2k^8}{\sigma^5}(1 + x^2),$$

where C is a positive constant depending only on C_0 .

In Propositions 2.1 and 2.3, the formulation of the problem is not symmetric; therefore, we state the bound for both the left and right tail probabilities. The computation of σ^2 and γ is not central to our study and is omitted from this and the next two examples. If k and p are fixed, then the range of x for the relative-error bound to vanish is $0 \leq x = o(n^{1/4})$, which is presumably optimal in comparison to the i.i.d. case.

In the following, we provide empirical evidence of the advantage of skewness correction. Consider $k = 2$. It can be computed that

$$\sigma^2 = n(p^2 + 2p^3 - 3p^4)$$

Table 1. $n = 1500, p = 0.25, \gamma \approx 0.138$. Values of $L_N(L_{\text{skew}})$ and $R_N(R_{\text{skew}})$ based on 10^6 repetitions

x	L_N	L_{skew}	R_N	R_{skew}
2	-0.195	-0.032	0.262	0.050
2.5	-0.238	0.093	0.344	-0.063
3	-0.538	-0.138	0.476	-0.208
3.5	-0.811	-0.491	1.201	-0.182
4	-0.968	-0.862	1.810	-0.358

and

$$\gamma = \frac{n}{\sigma^3}(p^2 + 6p^3 - 3p^4 - 24p^5 + 20p^6).$$

In Table 1, we provide simulated values (based on 10^6 repetitions) for

$$L_N := \frac{P(W < -x)}{\Phi(-x)} - 1, \quad L_{\text{skew}} := \frac{P(W < -x)}{\Phi(-x)e^{-\gamma x^3/6}} - 1,$$

and

$$R_N := \frac{P(W > x)}{1 - \Phi(x)} - 1, \quad R_{\text{skew}} := \frac{P(W > x)}{(1 - \Phi(x))e^{\gamma x^3/6}} - 1,$$

for the case $n = 1500$ and $p = 0.25$ and various values of x . Table 1 clearly shows that the tail probability approximations with skewness correction is much more accurate.

2.2.2. U-statistics

Let X_1, X_2, \dots be a sequence of i.i.d. random variables. Let $s \geq 2$ be a fixed integer. Let $h : \mathbb{R}^s \rightarrow \mathbb{R}$ be a symmetric, Borel-measurable function. We consider the Hoeffding [20] U-statistic

$$\sum_{1 \leq i_1 < \dots < i_s \leq m} h(X_{i_1}, \dots, X_{i_s}).$$

Assume that

$$Eh(X_1, \dots, X_s) = 0, \quad |h(X_1, \dots, X_s)| \leq C_1 < \infty$$

and the U-statistic is nondegenerate, namely, $Eg^2(X_1) > 0$, where

$$g(x) := E(h(X_1, \dots, X_s) | X_1 = x).$$

Applying Theorem 2.1 to the U-statistic above yields the following result.

Proposition 2.2. *In the above setting, let*

$$W = \frac{1}{\sigma} \sum_{1 \leq i_1 < \dots < i_s \leq m} h(X_{i_1}, \dots, X_{i_s}),$$

where

$$\sigma^2 = \text{Var} \left[\sum_{1 \leq i_1 < \dots < i_s \leq m} h(X_{i_1}, \dots, X_{i_s}) \right].$$

Let $\gamma = EW^3$. Let C_0 be any fixed constant. We have, for $0 \leq x \leq C_0m^{1/4}$,

$$\left| \frac{P(W > x)}{(1 - \Phi(x))e^{\gamma x^3/6}} - 1 \right| \leq \frac{C}{\sqrt{m}}(1 + x^2)$$

where C is a positive constant depending only on C_0 , h and the distribution of X_1 .

Proof. The above W satisfies the assumptions in Theorem 2.1 with

$$n = \binom{m}{s}, \quad \delta = \frac{C_1}{\sigma}, \quad d \leq m^{s-1}.$$

From (5.13) of Hoeffding [20] and the nondegeneracy condition, we have

$$\sigma^2 \geq \binom{m}{s} s \binom{m-s}{s-1} E g^2(X_1) \asymp m^{2s-1}.$$

The proposition then follows from (2.3). □

Remark 2.2. Chen and Shao [15] obtained a bound on the Kolmogorov distance in normal approximation for nondegenerate U-statistics. The references therein comprise a large body of literature on the rate of convergence in normal approximation for U-statistics. Our relative error bound for the skewness corrected tail probability approximation for U-statistics seems to be new.

2.2.3. Subgraph counts in the Erdős–Rényi random graph

Let $K(N, p)$ be the Erdős–Rényi random graph with N vertices. Each pair of vertices is connected with probability p and remains disconnected with probability $1 - p$, independent of all else. Let G be a given fixed graph. For any graph H , let $v(H)$ and $e(H)$ denote the number of its vertices and edges, respectively. Let $v = v(G)$, $e = e(G)$. Theorem 2.1 leads to the following result.

Proposition 2.3. *Let S be the number of copies (not necessarily induced) of G in $K(N, p)$, and let $W = (S - ES)/\sqrt{\text{Var}(S)}$ be the standardized version. Let $\gamma = EW^3$. Let C_0 be any fixed constant. We have, for $0 \leq x \leq C_0[N^6(1 - p)^{5/2}p^{5e}/\psi^{5/2}]^{1/2}$,*

$$\max \left\{ \left| \frac{P(W < -x)}{\Phi(-x)e^{-\gamma x^3/6}} - 1 \right|, \left| \frac{P(W > x)}{(1 - \Phi(x))e^{\gamma x^3/6}} - 1 \right| \right\} \leq \frac{C(G)\psi^{5/2}}{(1 - p)^{5/2}p^{5e}N^6}(1 + x^2),$$

where $C(G)$ is a constant depending only on C_0 and G , and

$$\psi = \min_{H \subset G, e(H) > 0} \{N^{v(H)} p^{e(H)}\}.$$

Proof. In this proof, C denotes positive constants that are allowed to depend on C_0 and the given fixed graph G . Let the potential edges of $K(N, p)$ be denoted by $(e_1, \dots, e_{\binom{N}{2}})$. In applying Theorem 2.1, let $W = \sum_{i \in I} X_i$, where the index set is

$$I = \left\{ i = (i_1, \dots, i_e) : 1 \leq i_1 < \dots < i_e \leq \binom{N}{2}, G_i := (e_{i_1}, \dots, e_{i_e}) \text{ is a copy of } G \right\},$$

$$X_i = \sigma^{-1}(Y_i - p^e), \quad \sigma^2 := \text{Var}(S), \quad Y_i = \prod_{l=1}^e E_{i_l},$$

and E_{i_l} is the indicator of the event that the edge e_{i_l} is connected in $K(N, p)$. The above W satisfies the assumptions in Theorem 2.1 with

$$n := |I| \leq N^v, \quad m = \binom{N}{2}, \quad \delta = \frac{1}{\sigma}, \quad s \leq C, \quad d \leq CN^{v-2}.$$

It is known that (cf. (3.7) of Barbour, Karoński and Ruciński [3])

$$\sigma^2 \geq C(1 - p)N^{2v} p^{2e} \psi^{-1}.$$

The proposition then follows from (2.3). □

Remark 2.3. Barbour, Karoński and Ruciński [3] first studied normal approximation for the above W using Stein’s method. Because $\psi \leq N^2 p$, if p is fixed, then the range of x for the relative error to vanish is $o(N^{1/2})$. It is presumably optimal and larger than the range of $o(N^{1/3})$, for which Zhang [31] proved that the relative error in normal approximation vanishes.

3. Exponential concentration inequality

3.1. Preliminaries

Let $\{X'_\alpha : \alpha \in [m]\}$ be an independent copy of $\{X_\alpha : \alpha \in [m]\}$. For each $\alpha \in [m]$, let $W^{(\alpha)}$ be defined as for W at the beginning of Section 2.1, except by changing X_α to X'_α . We have

$$\mathcal{L}(W, W^{(\alpha)}) = \mathcal{L}(W^{(\alpha)}, W). \tag{3.1}$$

From (2.1), we have

$$|W - W^{(\alpha)}| \leq 2d\delta. \tag{3.2}$$

By the Efron–Stein inequality, we have

$$C_2 := \sum_{\alpha=1}^m E(W - W^{(\alpha)})^2 \geq 2 \operatorname{Var}(W) = 2. \tag{3.3}$$

Moreover, it is straightforward to verify that $1 \leq n\delta$ and

$$1 = \operatorname{Var}(W) \leq nsd\delta^2, \quad \sum_{\alpha=1}^m E(W - W^{(\alpha)})^2 \leq 4md^2\delta^2, \quad |\gamma| \leq 4ns^2d^2\delta^3. \tag{3.4}$$

We have the following local dependence structure for W . **(LD1)**: For $i \in [n]$, let $A_i = \{j \in [n] : \mathcal{I}_j \cap \mathcal{I}_i \neq \emptyset\}$; hence, ξ_i is independent of $\{\xi_j : j \notin A_i\}$. **(LD2)**: For $i \in [n]$ and $j \in A_i$, let $A_{ij} = \{k \in [n] : \mathcal{I}_k \cap (\mathcal{I}_i \cup \mathcal{I}_j) \neq \emptyset\}$; hence, $\{\xi_i, \xi_j\}$ is independent of $\{\xi_k : k \notin A_{ij}\}$. **(LD3)**: For $i \in [n]$, $j \in A_i$ and $k \in A_{ij}$, let $A_{ijk} = \{l \in [n] : \mathcal{I}_l \cap (\mathcal{I}_i \cup \mathcal{I}_j \cup \mathcal{I}_k) \neq \emptyset\}$; hence, $\{\xi_i, \xi_j, \xi_k\}$ is independent of $\{\xi_l : l \notin A_{ijk}\}$. From (2.1), we have

$$|A_i|, |A_{ij}|, |A_{ijk}| \leq 3sd. \tag{3.5}$$

For $A \subset [n]$, denote $\xi_A = \sum_{i \in A} \xi_i$ and $\xi_i := \xi_{\{i\}}$. We have

$$\gamma = EW^3 = 2 \sum_{i=1}^n \sum_{j \in A_i} \sum_{k \in A_{ij}} E\xi_i \xi_j \xi_k - \sum_{i=1}^n \sum_{j, k \in A_i} E\xi_i \xi_j \xi_k. \tag{3.6}$$

Let

$$V_1 = \sum_{\alpha=1}^m (W - W^{(\alpha)}), \quad V_2 = \sum_{\alpha=1}^m (W - W^{(\alpha)})^2.$$

Lemma 3.1. *Regard V_1 and V_2 as functions of the independent random variables $\{X_\alpha : \alpha \in [m]\} \cup \{X'_\alpha : \alpha \in [m]\}$. For some $\beta \in [m]$, if we change X_β or X'_β to another independent copy, V_1 is changed by at most $2sd\delta$, and V_2 is changed by at most $4sd^2\delta^2$.*

Proof of Lemma 3.1. For each $\alpha \in [m]$, define $\{\xi_i^{(\alpha)} : i \in [n]\}$ as for $\{\xi_i : i \in [n]\}$, except by changing X_α to X'_α . We have

$$W^{(\alpha)} = \sum_{i=1}^n \xi_i^{(\alpha)}.$$

From the definition of N_α and \mathcal{I}_i , we have

$$V_1 = \sum_{\alpha=1}^m \sum_{i \in [n]} (\xi_i - \xi_i^{(\alpha)}) = \sum_{\alpha=1}^m \sum_{i \in N_\alpha} (\xi_i - \xi_i^{(\alpha)}) = \sum_{i=1}^n \sum_{\alpha \in \mathcal{I}_i} (\xi_i - \xi_i^{(\alpha)}).$$

Changing X_β or X'_β only affects $(\xi_i - \xi_i^{(\alpha)})$ if $i \in N_\beta$, which has cardinality at most d by (2.1). From $|\mathcal{I}_i| \leq s$ and $|\xi_i - \xi_i^{(\alpha)}| \leq 2\delta$ (cf. (2.1)), V_1 is changed by at most $2sd\delta$.

Now we turn to V_2 . We have

$$\begin{aligned} V_2 &= \sum_{\alpha=1}^m \left[\sum_{i \in N_\alpha} (\xi_i - \xi_i^{(\alpha)}) \right]^2 = \sum_{\alpha=1}^m \sum_{i, j \in N_\alpha} (\xi_i - \xi_i^{(\alpha)}) (\xi_j - \xi_j^{(\alpha)}) \\ &= \sum_{\substack{i, j \in N_\alpha \\ \mathcal{I}_i \cap \mathcal{I}_j \neq \emptyset}} \sum_{\alpha \in \mathcal{I}_i \cap \mathcal{I}_j} (\xi_i - \xi_i^{(\alpha)}) (\xi_j - \xi_j^{(\alpha)}). \end{aligned}$$

Reasoning similar to that for V_1 above leads to the observation that changing X_β or X'_β changes V_2 by at most $4sd^2\delta^2$. □

3.2. Moment generating function bound

Proposition 3.1. *Let C_0 be any fixed constant. Under the assumptions in Theorem 2.1, for*

$$0 \leq t \leq C_0(ns^2d^2\delta^3)^{-1/2},$$

we have

$$Ee^{tW} \leq C \exp\left(\frac{t^2}{2} + \frac{\gamma t^3}{6}\right), \tag{3.7}$$

where C is a positive constant depending only on C_0 .

Proof. In this proof, C denotes positive constants that can depend on C_0 , $O(a)$ denotes a quantity such that $|O(a)| \leq Ca$. Let $h(t) = Ee^{tW}$. We follow the standard way of proving exponential bounds in Stein’s method by first bounding $h'(t)$ using the local dependence structure (LD1)–(LD3) of W stated in Section 3.1, then obtaining an upper bound for $(\log h(t))'$, and finally proving (3.7).

Note that from $t = O(1)(ns^2d^2\delta^3)^{-1/2}$ and $n\delta \geq 1$, we have

$$sd\delta t = O(1). \tag{3.8}$$

Because ξ_i is independent of $W - \xi_{A_i}$ by (LD1), $E\xi_i = 0$, $|A_i| \leq Csd$ from (3.5) and $|\xi_i| \leq \delta$ from (2.1), we have

$$\begin{aligned} h'(t) &= EW e^{tW} = \sum_{i=1}^n E\xi_i e^{tW} = \sum_{i=1}^n E\xi_i [e^{tW} - e^{t(W - \xi_{A_i})}] \\ &= \sum_{i=1}^n E\xi_i \left[\xi_{A_i} t e^{tW} - \frac{\xi_{A_i}^2}{2} t^2 e^{tW} + O(s^3 d^3 \delta^3 t^3 e^{tW + Csd\delta t}) \right]. \end{aligned} \tag{3.9}$$

For the first term on the right-hand side of (3.9), we have, recalling $\sum_{i=1}^n E\xi_i\xi_{A_i} = EW^2 = 1$ and using similar arguments as above for the error term,

$$\begin{aligned} & \sum_{i=1}^n E\xi_i\xi_{A_i}te^{tW} \\ &= \sum_{i=1}^n \sum_{j \in A_i} E\xi_i\xi_jtEe^{t(W-\xi_{A_{ij}})} + \sum_{i=1}^n \sum_{j \in A_i} E\xi_i\xi_jt[e^{tW} - e^{t(W-\xi_{A_{ij}})}] \\ &= th(t) + \sum_{i=1}^n \sum_{j \in A_i} E\xi_i\xi_jtE[e^{t(W-\xi_{A_{ij}})} - e^{tW}] + \sum_{i=1}^n \sum_{j \in A_i} E\xi_i\xi_jt[e^{tW} - e^{t(W-\xi_{A_{ij}})}] \\ &= th(t) + \sum_{i=1}^n \sum_{j \in A_i} E[\xi_i\xi_j - E\xi_i\xi_j]\xi_{A_{ij}}t^2e^{tW} + O(ns^3d^3\delta^4t^3e^{tW+Csd\delta t}). \end{aligned} \tag{3.10}$$

For the second terms on the right-hand of (3.9) and of (3.10), we have, by recalling (3.6),

$$\begin{aligned} & \sum_{i=1}^n E[\xi_i\xi_{A_i}\xi_{A_{ij}} - (E\xi_i\xi_{A_i})\xi_{A_{ij}} - \xi_i\xi_{A_i}^2/2]t^2e^{tW} \\ &= \sum_{i=1}^n \sum_{j \in A_i} \sum_{k \in A_{ij}} E\left[\xi_i\xi_j\xi_k - (E\xi_i\xi_j)\xi_k - \frac{\xi_i\xi_j\xi_k I(k \in A_i)}{2}\right]t^2Ee^{t(W-\xi_{A_{ijk}})} \\ &+ \sum_{i=1}^n \sum_{j \in A_i} \sum_{k \in A_{ij}} E\left[\xi_i\xi_j\xi_k - (E\xi_i\xi_j)\xi_k - \frac{\xi_i\xi_j\xi_k I(k \in A_i)}{2}\right]t^2[e^{tW} - e^{t(W-\xi_{A_{ijk}})}] \\ &= \gamma \frac{t^2h(t)}{2} + O(ns^3d^3\delta^4t^3e^{tW+Csd\delta t}). \end{aligned} \tag{3.11}$$

Combining (3.9), (3.10) and (3.11), we have

$$h'(t) = th(t) + \gamma \frac{t^2h(t)}{2} + O(ns^3d^3\delta^4t^3e^{Csd\delta t})h(t). \tag{3.12}$$

Recall (3.8). Because $h(0) = 1$, we have

$$\log h(t) = \int_0^t [u + \gamma u^2/2 + O(ns^3d^3\delta^4u^3)] du = \frac{t^2}{2} + \frac{\gamma t^3}{6} + O(ns^3d^3\delta^4t^4).$$

This implies (3.7) because from $t = O(1)(ns^2d^2\delta^3)^{-1/2}$ and (3.4), we have

$$ns^3d^3\delta^4t^4 \leq \frac{Cns^3d^3\delta^4}{n^2s^4d^4\delta^6} = \frac{C}{nsd\delta^2} \leq C. \tag{3.13}$$

□

3.3. Exponential concentration inequality

What we call a concentration inequality here is a smoothing inequality originally used in normal approximation by Esseen [18]. It was developed via Stein’s method in, for example, Ho and Chen [19] and Chen and Shao [14]. Shao [26] developed exponential concentration inequalities in normal approximation for nonlinear statistics.

Proposition 3.2. *Let C_0 be any fixed constant. Under the assumptions of Theorem 2.1, for $d\delta \leq 1/2$ and*

$$1 \leq x \leq C_0(ns^2d^2\delta^3)^{-1/2},$$

we have, for any $\varepsilon > 0$,

$$P(x \leq W \leq x + \varepsilon) \leq Cms^2d^2\delta^2(\varepsilon + d\delta)e^{\varepsilon x} \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right) + \exp\left(-\frac{1}{Cms^2d^4\delta^4}\right),$$

where C is a positive constant depending only on C_0 .

To prove Proposition 3.2, we apply the following lemma, which provides moment generating function bounds for a function of independent random variables. It is proved in a manner similar to that in Chatterjee [8]. See Chatterjee [9] and Chen and Röllin [16] for related ideas.

Lemma 3.2. *Let $V = h(Y_1, \dots, Y_N)$ where (Y_1, \dots, Y_N) are independent. Assume that $EV = 0$. Let $(\tilde{Y}_1, \dots, \tilde{Y}_N)$ be an independent copy of (Y_1, \dots, Y_N) . Suppose that for any $i \in [N]$,*

$$|h(Y_1, \dots, Y_N) - h(Y_1, \dots, Y_{i-1}, \tilde{Y}_i, Y_{i+1}, \dots, Y_N)| \leq \delta_3.$$

Then we have, for any $\theta > 0$,

$$Ee^{\theta V} \leq \exp(N\delta_3^2\theta^2/4).$$

Proof of Lemma 3.2. Let $V_0 = V$ and for $i \in [N]$, let

$$V_i = h(\tilde{Y}_1, \dots, \tilde{Y}_i, Y_{i+1}, \dots, Y_N)$$

and

$$U'_i = h(Y_1, \dots, Y_{i-1}, \tilde{Y}_i, Y_{i+1}, \dots, Y_N).$$

For $a > 0$ and $\theta \geq 0$, let $m_a(\theta) = Ee^{\theta(V \wedge a)}$. As in the proof of Proposition 3.1, we follow the standard way of obtaining the exponential bound by considering $m'_a(\theta)$. Because V_N is independent of V and $EV_N = 0$, we have

$$m'_a(\theta) = E(V \wedge a)e^{\theta(V \wedge a)} \leq EVe^{\theta(V \wedge a)} = E(V - V_N)e^{\theta(V \wedge a)} = \sum_{i=1}^N E(V_{i-1} - V_i)e^{\theta(V \wedge a)}.$$

Note that $E(V_{i-1} - V_i)e^{\theta(V \wedge a)} = E(V_i - V_{i-1})e^{\theta(U'_i \wedge a)}$, which is a consequence of the exchangeability of Y_i and \tilde{Y}_i . Therefore,

$$m'_a(\theta) \leq \frac{1}{2} \sum_{i=1}^N E(V_{i-1} - V_i)(e^{\theta(V \wedge a)} - e^{\theta(U'_i \wedge a)}).$$

From the fact that (cf. (7) of Chatterjee [8]) for any $x, y \in \mathbb{R}$,

$$\left| \frac{e^x - e^y}{x - y} \right| \leq \frac{1}{2}(e^x + e^y),$$

we have

$$\begin{aligned} m'_a(\theta) &\leq \frac{\theta}{4} \sum_{i=1}^N E|V_{i-1} - V_i| |V - U'_i| (e^{\theta(V \wedge a)} + e^{\theta(U'_i \wedge a)}) \\ &= \frac{\theta}{2} \sum_{i=1}^N E|V_{i-1} - V_i| |V - U'_i| e^{\theta(V \wedge a)}, \end{aligned}$$

again by the exchangeability of Y_i and \tilde{Y}_i . From the boundedness conditions on $|V_{i-1} - V_i|$ and $|V - U'_i|$, we have

$$m'_a(\theta) \leq \frac{\theta}{2} N \delta_3^2 m_a(\theta), \quad \forall \theta \geq 0,$$

which implies

$$m_a(\theta) \leq \exp(N \delta_3^2 \theta^2 / 4).$$

The lemma is proved by letting $a \rightarrow \infty$. □

Proof of Proposition 3.2. In this proof, we use c and C to denote positive constants that can depend only on C_0 . We follow the approach of Shao [26], who proved an exponential concentration inequality for normal approximation of nonlinear statistics. The basic idea is choosing a suitable function (3.14) in the Stein identity for W in (3.15). One side of the Stein identity has a lower bound in terms of $P(x \leq W \leq x + \varepsilon)$, and the other side has a suitable upper bound, leading to our result.

Recall $W^{(\alpha)}$ from Section 3.1. Let I be a uniform random variable on $[m]$ and independent of all else. Let $W' = W^{(I)}$. Define

$$f(w) = \begin{cases} 0, & w \leq x - 2d\delta, \\ e^{xw}(w - x + 2d\delta), & x - 2d\delta < w \leq x + \varepsilon + 2d\delta, \\ e^{x(x+\varepsilon+2d\delta)}(\varepsilon + 4d\delta), & w > x + \varepsilon + 2d\delta. \end{cases} \quad (3.14)$$

From (3.1), we have

$$\mathcal{L}(W, W') = \mathcal{L}(W', W).$$

Hence

$$E(W - W')(f(W) + f(W')) = 0.$$

Rewrite it as

$$\text{LHS} := 2E(W - W')f(W) = E(W - W')(f(W) - f(W')) =: \text{RHS}. \tag{3.15}$$

Part I: Upper bound for LHS.

Averaging over I :

$$\text{LHS} = \frac{2}{m} \sum_{\alpha=1}^m E(W - W^{(\alpha)})f(W).$$

Recall $V_1 = \sum_{\alpha=1}^m (W - W^{(\alpha)})$ and note that $x - 2d\delta \geq 0$ by the assumptions of the proposition. From the upper bound on f , we have

$$\begin{aligned} |\text{LHS}| &\leq \frac{2}{m} E|V_1|(\varepsilon + 4d\delta)e^{x(x+\varepsilon+2d\delta)} I(W \geq x - 2d\delta) \\ &\leq \frac{2}{m}(\varepsilon + 4d\delta)e^{x(x+\varepsilon+2d\delta)} [E|V_1|I(|V_1| > M(x - 2d\delta)) \\ &\quad + MEWI(W \geq x - 2d\delta)], \end{aligned} \tag{3.16}$$

where $M \geq 1$ is to be chosen above (3.20). Note that V_1 is symmetrical. For the first term on the right-hand side of (3.16), we have

$$\begin{aligned} &E|V_1|I(|V_1| > M(x - 2d\delta)) \\ &= 2EV_1I(V_1 > M(x - 2d\delta)) \\ &\leq 2M(x - 2d\delta)P(V_1 > M(x - 2d\delta)) + 2 \int_{M(x-2d\delta)}^{\infty} P(V_1 > y) dy. \end{aligned} \tag{3.17}$$

Applying Lemma 3.2 with $\theta = x$ and Lemma 3.1 to V_1 , we have

$$Ee^{xV_1} \leq \exp(Cms^2d^2\delta^2x^2).$$

Therefore,

$$\begin{aligned} &E|V_1|I(|V_1| > M(x - 2d\delta)) \\ &\leq 2M(x - 2d\delta) \frac{e^{Cms^2d^2\delta^2x^2}}{e^{xM(x-2d\delta)}} + 2 \int_{M(x-2d\delta)}^{\infty} \frac{e^{Cms^2d^2\delta^2x^2}}{e^{xy}} dy \\ &\leq Ce^{2Md\delta x} e^{-Mx^2} Mxe^{Cms^2d^2\delta^2x^2}. \end{aligned}$$

Now we consider the second term on the right-hand side of (3.16). Note that $|\gamma x| \leq Cns^2d^2\delta^3x \leq C$ (cf. (3.4)) for the range of x in the proposition to be nonempty. Following reasoning similar to that for (3.13) and (3.8), we have $ns^3d^3\delta^4x^3 \leq ns^3d^3\delta^4x^4 \leq C$ and

$$s\delta\delta x \leq C. \tag{3.18}$$

From the proof of Proposition 3.1 (cf. (3.12)), we have

$$EW e^{xW} \leq Cx \exp\left(\frac{x^2}{2} + \frac{\gamma x^3}{6}\right).$$

Therefore, from (3.18),

$$EW I(W \geq x - 2d\delta) \leq EW e^{xW} / e^{x(x-2d\delta)} \leq Cx \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right).$$

Combining the above bounds, we have

$$|\text{LHS}| \leq \frac{C}{m} e^{x^2} (\varepsilon + d\delta) e^{\varepsilon x} Mx [e^{2Md\delta x} e^{-Mx^2 + Cms^2d^2\delta^2x^2} + e^{-x^2/2 + \gamma x^3/6}]. \tag{3.19}$$

Now let $M = C(ms^2d^2\delta^2 + 1)$ for a sufficiently large C . Note that from (3.3) and (3.4), we have $1 \leq 2md^2\delta^2$. Recall $2d\delta \leq 1$ from the assumption of the proposition. The first term inside the brackets in (3.19) is dominated by the second term, and we have

$$|\text{LHS}| \leq \frac{C}{m} e^{x^2} (\varepsilon + d\delta) e^{\varepsilon x} ms^2d^2\delta^2x \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right). \tag{3.20}$$

Part II: Lower bound for RHS.

Because f is increasing and for $x - 2d\delta \leq w \leq x + \varepsilon + 2d\delta$,

$$f'(w) = x e^{xw} (w - x + 2d\delta) + e^{xw} \geq e^{x(x-2d\delta)},$$

we have, from (3.2) and (3.18),

$$\begin{aligned} \text{RHS} &= E(W - W')(f(W) - f(W')) \\ &\geq E(W - W')(f(W) - f(W')) I(x \leq W \leq x + \varepsilon) I(|W - W'| \leq 2d\delta) \\ &\geq cE(W - W')^2 e^{x^2} I(x \leq W \leq x + \varepsilon). \end{aligned}$$

Averaging over I , we have, recalling $V_2 = \sum_{\alpha=1}^m (W - W^{(\alpha)})^2$,

$$\text{RHS} \geq \frac{c}{m} e^{x^2} EI(x \leq W \leq x + \varepsilon) V_2.$$

Recall from (3.3) that $EV_2 = C_2$. We have

$$\begin{aligned} \text{RHS} &\geq \frac{cC_2}{m} e^{x^2} EI(x \leq W \leq x + \varepsilon)I(V_2 \geq C_2/2) \\ &\geq \frac{cC_2}{m} e^{x^2} (P(x \leq W \leq x + \varepsilon) - P(V_2 < C_2/2)). \end{aligned}$$

We now find an upper bound for the second probability, which equals

$$P(EV_2 - V_2 > C_2/2).$$

Applying Lemmas 3.2 and 3.1 to $EV_2 - V_2$, we have

$$P(EV_2 - V_2 > C_2/2) \leq e^{-\theta C_2/2} \exp(Cms^2d^4\delta^4\theta^2) = \exp\left(-\frac{1}{Cms^2d^4\delta^4}\right)$$

by choosing the optimal $\theta = C_2/4Cms^2d^4\delta^4$ and using $C_2 \geq 2$ from (3.3). We have arrived at:

$$\text{RHS} \geq \frac{c}{m} e^{x^2} \left[P(x \leq W \leq x + \varepsilon) - \exp\left(-\frac{1}{Cms^2d^4\delta^4}\right) \right]. \tag{3.21}$$

The proof is finished by combining (3.20) and (3.21). □

4. Proof of the main result

In this section, we prove our main result, Theorem 2.1. The lemmas stated in the proof are proved below. In this section, we use C to denote positive constants and use K to denote positive integers. They can depend only on C_0 and may differ in different expressions. We use $O(a)$ to denote a quantity such that $|O(a)| \leq Ca$.

4.1. Proof of Theorem 2.1

First, we have the following absolute-error bound in normal approximation for W , which is proved by a standard application of Stein’s method for sums of locally dependent random variables.

Lemma 4.1.

$$\sup_{x \in \mathbb{R}} |P(W \leq x) - \Phi(x)| \leq Cns^2d^2\delta^3. \tag{4.1}$$

From (3.3) and (3.4), we have

$$|\gamma|x^2 \leq Cns^2d^2\delta^3x^2 \leq Cmns^4d^4\delta^5x^2 \leq C \tag{4.2}$$

for x in (2.2). If x is bounded, from (4.2), we have

$$|(1 - \Phi(x)) - (1 - \Phi(x))e^{\gamma x^3/6}| \leq C|\gamma| \leq Cns^2d^2\delta^3. \tag{4.3}$$

From (4.1), (4.3) and (4.2), (2.3) holds for bounded x . Therefore, without loss of generality, we can assume in the following proof that x is sufficiently large and $mns^4d^4\delta^5$, and hence $|\gamma|$, is sufficiently small. These conditions may be used implicitly below.

We only prove for the case $\gamma \neq 0$. The case $\gamma = 0$ follows from a similar and simpler proof by working directly with the standard normal distribution. For $\gamma \neq 0$, $(1 - \Phi(x))e^{\gamma x^3/6}$ is no longer a distribution function. We are not aware of any version of Stein’s method that is directly applicable to signed-measure approximations. Instead, we use a standardized Poisson distribution, which has right-tail probabilities equivalent to $(1 - \Phi(x))e^{\gamma x^3/6}$ up to the range of x of interest (cf. (4.4)), for an intermediate approximation. Although other intermediate approximations might also work, we choose Poisson approximation because it is well studied in the Stein’s method literature and many relevant results are available.

Let $Z_\gamma = \gamma(Y_\gamma - \frac{1}{\gamma^2})$, where $Y_\gamma \sim \text{Poi}(\frac{1}{\gamma^2})$. We have $EZ_\gamma = 0$, $EZ_\gamma^2 = 1$, $EZ_\gamma^3 = \gamma$. From Cramér’s expansion, see, for example, Petrov [21], Chapter 8, Theorem 2, we have

$$\frac{P(Z_\gamma > x)}{(1 - \Phi(x))e^{\gamma x^3/6}} = 1 + O(|\gamma|)(1 + x) + O(\gamma^2)x^4 \quad \text{for } 0 \leq x \leq C_0|\gamma|^{-1/2} \tag{4.4}$$

for $|\gamma| \leq 1$. Therefore, it suffices to prove

$$\begin{aligned} |P(W > x) - P(Z_\gamma > x)| &\leq Cms^4d^4\delta^5(1 + x^2)(1 - \Phi(x))e^{\gamma x^3/6} \\ &\leq Cms^4d^4\delta^5x \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right). \end{aligned} \tag{4.5}$$

Denote the support of Z_γ by

$$\mathcal{S} = \left\{ \gamma\mathbb{Z}^+ - \frac{1}{\gamma} \right\}.$$

We denote

$$\alpha = ns^2d^2\delta^3$$

and use them interchangeably below. Let

$$h_\alpha^+(w) = \begin{cases} 1 & w < x, \\ 1 - 2\left(\frac{w-x}{\alpha}\right)^2 & x \leq w < x + \alpha/2, \\ 2\left(1 - \frac{w-x}{\alpha}\right)^2 & x + \alpha/2 \leq w < x + \alpha, \\ 0 & w \geq x + \alpha \end{cases}$$

and $h_{\alpha}^{-}(w) := h_{\alpha}^{+}(w + \alpha)$. Let $h_{\alpha} = h_{\alpha}^{+}$ or $h_{\alpha} = h_{\alpha}^{-}$. The following holds for either choice of h_{α} . It is straightforward to verify that h'_{α} exists and is continuous and

$$|h'_{\alpha}(w_1)| \leq \frac{2}{\alpha}, \quad \left| \frac{h'_{\alpha}(w_1) - h'_{\alpha}(w_2)}{w_1 - w_2} \right| \leq \frac{8}{\alpha^2}, \quad \forall w_1 \neq w_2.$$

Note that

$$\begin{aligned} & Eh_{\alpha}^{-}(W) - Eh_{\alpha}^{-}(Z_{\gamma}) - P(x - \alpha < Z_{\gamma} \leq x) \\ & \leq P(W \leq x) - P(Z_{\gamma} \leq x) \\ & \leq Eh_{\alpha}^{+}(W) - Eh_{\alpha}^{+}(Z_{\gamma}) + P(x < Z_{\gamma} \leq x + \alpha). \end{aligned} \tag{4.6}$$

For $w_0 \in \mathcal{S}$, $|w_0| = O(|\gamma|^{-1/2})$ and sufficiently small $|\gamma|$, applying Stirling’s approximation and Taylor’s expansion to the Poisson probability, we have

$$P(Z_{\gamma} = w_0) = P\left(Y_{\gamma} = \frac{w_0}{\gamma} + \frac{1}{\gamma^2}\right) = \frac{|\gamma|}{\sqrt{2\pi}} \exp\left(-\frac{w_0^2}{2} + \frac{\gamma w_0^3}{6} + O(1)\right). \tag{4.7}$$

Therefore, the difference between $P(W \leq x) - P(Z_{\gamma} \leq x)$ and $Eh_{\alpha}(W) - Eh_{\alpha}(Z_{\gamma})$ in (4.6) is bounded by

$$P(x - \alpha < Z_{\gamma} \leq x + \alpha) = O(\alpha) \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right), \tag{4.8}$$

which is bounded by the right-hand side of (4.5). To bound $Eh_{\alpha}(W) - Eh_{\alpha}(Z_{\gamma})$, consider the Stein equation for Z_{γ} :

$$\frac{1}{\gamma}(f(w + \gamma) - f(w)) - wf(w) = h_{\alpha}(w) - Eh_{\alpha}(Z_{\gamma}). \tag{4.9}$$

Unlike the standard application of Poisson approximation in the Stein’s method literature, our problem here has two difficulties: (1) the random variable of interest W may not have the same support as Z_{γ} ; (2) we need an error bound that depends on x in an optimal way. To overcome these difficulties, in the following, we first extend the solution to (4.9) to the whole real line and control the error introduced by the extension (from Lemma 4.2 to (4.21)). Then, we control the main error term, namely the expectation of the left-hand side of (4.9) with w replaced by W and with the extended solution f (from Lemma 4.5 to the end of the proof). To have the optimal dependence of the error bound on x , we make use the moment generating function bound and the concentration inequality established in Propositions 3.1 and 3.2, respectively.

The Stein equation (4.9) has the following solution $f := f_{h_{\alpha}}$ on \mathcal{S} : $f(-1/\gamma) = 0$ and for $w_0 \in \mathcal{S} \setminus \{-\frac{1}{\gamma}\}$,

$$\begin{aligned} f(w_0) &= \frac{1}{\frac{1}{\gamma}P(Y_{\gamma} = \frac{1}{\gamma^2} + \frac{w_0}{\gamma} - 1)} E[h_{\alpha}(Z_{\gamma}) - Eh_{\alpha}(Z_{\gamma})] I\left(Y_{\gamma} \leq \frac{1}{\gamma^2} + \frac{w_0}{\gamma} - 1\right) \\ &= -\frac{1}{\frac{1}{\gamma}P(Y_{\gamma} = \frac{1}{\gamma^2} + \frac{w_0}{\gamma})} E[h_{\alpha}(Z_{\gamma}) - Eh_{\alpha}(Z_{\gamma})] I\left(Y_{\gamma} \geq \frac{1}{\gamma^2} + \frac{w_0}{\gamma}\right), \end{aligned} \tag{4.10}$$

where we recall that $Y_\gamma \sim \text{Poi}(1/\gamma^2)$. From the expression of f in (4.10) and $|h_\alpha(Z_\gamma) - Eh_\alpha(Z_\gamma)| \leq 1$, we have

$$|f(w_0)| \leq \frac{1}{|\gamma|} \min \left\{ \frac{\gamma^2 P(Y_\gamma \leq \frac{1}{\gamma^2} + \frac{w_0}{\gamma} - 1)}{P(Y_\gamma = \frac{1}{\gamma^2} + \frac{w_0}{\gamma} - 1)}, \frac{\gamma^2 P(Y_\gamma \geq \frac{1}{\gamma^2} + \frac{w_0}{\gamma})}{P(Y_\gamma = \frac{1}{\gamma^2} + \frac{w_0}{\gamma})} \right\}.$$

From the proof of Lemma 1.1.1 of Barbour, Holst and Janson [2] (cf. (1.20) and (1.21) therein), if $w_0 \leq \gamma$, then the first term inside the minimum is bounded by $2(1 \wedge |\gamma|)$, and if $w_0 > \gamma$, then the second term inside the minimum is bounded by $2(1 \wedge |\gamma|)$. Therefore,

$$|f(w_0)| \leq 2. \tag{4.11}$$

Because in general our W has different support from \mathcal{S} , we extend f to $f : \mathbb{R} \rightarrow \mathbb{R}$ as follows. Let $f(w_0) = 0$ for $w_0 \in \{\gamma\mathbb{Z}^- - \frac{1}{\gamma}\}$. For w between w_0 and $w_0 + \gamma$ such that $w_0 \in \{\gamma\mathbb{Z}^- - \frac{1}{\gamma}\}$, we define $f(w)$ to be a fifth-order polynomial function such that it matches the discrete derivatives at w_0 and $w_0 + \gamma$ up to the second order. In more detail, let

$$\begin{aligned} f_0 &:= f(w_0), & f_1 &:= f(w_0 + \gamma), & f_2 &:= f(w_0 + 2\gamma), & f_{-1} &:= f(w_0 - \gamma), \\ f'_0 &:= \frac{f_1 - f_{-1}}{2\gamma}, & f'_1 &:= \frac{f_2 - f_0}{2\gamma}, \end{aligned} \tag{4.12}$$

$$f''_0 := \frac{f_1 - 2f_0 + f_{-1}}{\gamma^2}, \quad f''_1 := \frac{f_2 - 2f_1 + f_0}{\gamma^2}, \tag{4.13}$$

and let

$$f(w) = \sum_{i=1}^6 b_i (w - w_0)^{6-i}, \tag{4.14}$$

where

$$\begin{aligned} b_1 &= -\frac{1}{\gamma^2} \cdot \frac{f'_1 - f'_0}{\gamma}, & b_2 &= \frac{5}{2\gamma} \cdot \frac{f''_1 - f''_0}{\gamma}, \\ b_3 &= -\frac{3}{2} \cdot \frac{f''_1 - f''_0}{\gamma}, & b_4 &= \frac{f''_0}{2}, & b_5 &= f'_0, & b_6 &= f_0. \end{aligned}$$

In the following, for any $w \in \mathbb{R}$, let w_0 be such that $w_0 \in \{\gamma\mathbb{Z}^- - \frac{1}{\gamma}\}$ and $w_0 + \gamma < w \leq w_0$ if $\gamma < 0$ and $w_0 \leq w < w_0 + \gamma$ if $\gamma > 0$. For a random variable W , W_0 is defined in the same way as for w_0 .

It follows from the construction of f above that $f''(w)$ exists and is continuous and $f^{(3)}(w)$ exists for $w \notin \mathcal{S}$. For $w \in \mathcal{S}$, we define $f^{(3)}(w) = 0$ as they will not enter into consideration when we do Taylor's expansion below (cf. (4.28)). Note that

$$f(w) = O(1)(f(w_0 - \gamma) + f(w_0) + f(w_0 + \gamma) + f(w_0 + 2\gamma)). \tag{4.15}$$

Therefore, from (4.11), f is bounded. Note that after such extension, f no longer satisfies (4.9) exactly, except on S . However, we can quantify the error as in the following lemma.

Lemma 4.2. *For the above defined f , we have*

$$\begin{aligned} & \frac{1}{\gamma}(f(w + \gamma) - f(w)) - wf(w) \\ &= h_\alpha(w) - Eh_\alpha(Z_\gamma) + O(1)I(|w - x| \leq C\alpha) \\ & \quad + O(|\gamma|) \sum_{i=-K}^K |f(w_0 + i \cdot \gamma)| + O(1)I(w < -1/\gamma + \gamma)I(\gamma > 0) \\ & \quad + O(1)I(w > -1/\gamma + \gamma)I(\gamma < 0). \end{aligned} \tag{4.16}$$

By replacing w by W and w_0 by W_0 in (4.16) and taking expectations on both sides, we have

$$\begin{aligned} & Eh_\alpha(W) - Eh_\alpha(Z_\gamma) \\ &= E \left[\frac{1}{\gamma}(f(W + \gamma) - f(W)) - Wf(W) \right] + O(1)P(|W - x| \leq C\alpha) \\ & \quad + O(|\gamma|) \sum_{i=-K}^K E|f(W_0 + i \cdot \gamma)| + O(1)P(W < -1/\gamma + \gamma)I(\gamma > 0) \\ & \quad + O(1)P(W > -1/\gamma + \gamma)I(\gamma < 0) \\ &=: R_1 + R_2 + R_3 + R_4 + R_5. \end{aligned}$$

We bound these remainders in the reverse order. If $\gamma > 0$, we have, by applying Proposition 3.1 to $-W$,

$$\begin{aligned} P(W < -1/\gamma + \gamma) &\leq CEe^{-xW}/e^{x/\gamma} \\ &\leq C \exp\left(\frac{x^2}{2} - \frac{\gamma x^3}{6} - \frac{x}{|\gamma|}\right) \\ &= C|\gamma| \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right) \frac{1}{|\gamma|} \exp\left(x^2 - \frac{\gamma x^3}{3} - \frac{x}{|\gamma|}\right) \\ &\leq C|\gamma| \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right), \end{aligned} \tag{4.17}$$

where we use $1 \leq x = O(1)|\gamma|^{-1/2}$ and $|\gamma|$ is sufficiently small (cf. the arguments below (4.3)). Together with the same bound for R_5 , we have

$$|R_4| + |R_5| \leq C|\gamma| \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right).$$

To bound R_3 , we use the following lemma.

Lemma 4.3. *We have*

$$\begin{aligned} \sum_{i=-K}^K E|f(W_0 + i\gamma)| &\leq CP(W \geq x - C\alpha) + CP(Z_\gamma > x - \alpha) \\ &\quad + CEI(0 \leq W \leq x)e^{\frac{W^2}{2} - \frac{\gamma W^3}{6}} P(Z_\gamma > x - \alpha). \end{aligned} \tag{4.18}$$

For the first term on the right-hand side of (4.18), we have, by (3.7),

$$\begin{aligned} P(W \geq x - C\alpha) &\leq Ce^{-x^2} \exp\left(\frac{x^2}{2} + \frac{\gamma x^3}{6}\right) \\ &= C \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right). \end{aligned}$$

For the second term on the right-hand side of (4.18), we have, by (4.4),

$$P(Z_\gamma \geq x - \alpha) \leq \frac{C}{x} \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right).$$

For the third term on the right-hand side of (4.18), we have

$$\begin{aligned} EI(0 \leq W \leq x)e^{\frac{W^2}{2} - \frac{\gamma W^3}{6}} &\leq 1 + C \int_0^x (y + 1)e^{\frac{y^2}{2} - \frac{\gamma y^3}{6}} P(W > y) dy \\ &= O(x), \end{aligned} \tag{4.19}$$

where we use the following lemma in the last step.

Lemma 4.4. *For integer $k \geq 1$, we have*

$$\int_0^x y^k e^{\frac{y^2}{2} - \frac{\gamma y^3}{6}} P(W > y) dy = O(1)x^k.$$

Combining these bounds, we have

$$|R_3| \leq C|\gamma| \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right).$$

Next, we use Proposition 3.2 to bound R_2 as follows. Recall we assumed without loss of generality that x is sufficiently large, $mns^4 d^4 \delta^5$, and hence α , is sufficiently small (cf. (3.4)). We have,

from Proposition 3.2 and $d\delta \leq \alpha$ (cf. (3.4)),

$$\begin{aligned}
 |R_2| &\leq CP(|W - x| \leq C\alpha) \\
 &\leq Cms^2d^2\delta^2\alpha x \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right) + \exp\left(-\frac{1}{Cms^2d^4\delta^4}\right) \\
 &\leq Cms^2d^2\delta^2\alpha x \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right) \\
 &\quad + C \exp\left(-\frac{1}{Cms^2d^4\delta^4} + Cx^2\right) \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right). \tag{4.20}
 \end{aligned}$$

Note that for $x = O(mns^4d^4\delta^5)^{-1/2}$, we have (cf. (3.4))

$$ms^2d^4\delta^4x^2 = O\left(\frac{ms^2d^4\delta^4}{mns^4d^4\delta^5}\right) = O\left(\frac{ns^2d^2\delta^3}{n^2s^4d^2\delta^4}\right) = O(ns^2d^2\delta^3).$$

Therefore, the second term on the right-hand side of (4.20) is dominated by the first term and

$$|R_2| \leq Cms^4d^4\delta^5x \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right). \tag{4.21}$$

We are now left to bound R_1 . By Taylor’s expansion and exploiting the local dependence structure (LD1)–(LD3) in Section 3.1, we have the following lemma. Note that this is where we use the crucial choice of Z_γ so that it matches the moments of W up to the third order.

Lemma 4.5. *We have*

$$E\left[\frac{1}{\gamma}(f(W + \gamma) - f(W)) - Wf(W)\right] = O(\alpha^2)E|f^{(3)}(W + O(\alpha))|.$$

To bound $f^{(3)}$, we use the following lemma.

Lemma 4.6. *We have*

$$\begin{aligned}
 &E|f^{(3)}(W + O(\alpha))| \\
 &\leq \frac{C}{\alpha^2}P(|W - x| \leq C\alpha) + CE(1 + |W|^3)|f(W + O(\alpha))| \\
 &\quad + CE(1 + W^2)I(W \geq x - C\alpha) + CP(Z_\gamma > x - C\alpha) \\
 &\quad + \frac{C}{\gamma^2}P(W \leq -1/\gamma + O(\alpha))I(\gamma > 0) + \frac{C}{\gamma^2}P(W \geq -1/\gamma - O(\alpha))I(\gamma < 0). \tag{4.22}
 \end{aligned}$$

The first term on the right-hand side of (4.22) is bounded as in (4.20) and (4.21). For the second term on the right-hand side of (4.22), from the proof of Lemma 4.3, we have

$$\begin{aligned} & E(1 + |W|^3)|f(W + O(\alpha))| \\ & \leq CE(1 + W^3)I(W \geq x - C\alpha) + CP(Z_\gamma > x - \alpha)(1 + E|W^3|) \\ & \quad + CE(1 + W^3)I(0 \leq W \leq x)e^{\frac{W^2}{2} - \frac{\gamma W^3}{6}} P(Z_\gamma > x - \alpha). \end{aligned} \tag{4.23}$$

Similar to (4.19), using Lemma 4.4, we have

$$\begin{aligned} & E(1 + W^3)I(0 \leq W \leq x)e^{\frac{W^2}{2} - \frac{\gamma W^3}{6}} \\ & \leq 1 + \int_0^x \left[3y^2 + (1 + y^3) \left(y - \frac{\gamma y^2}{2} \right) \right] e^{\frac{y^2}{2} - \frac{\gamma y^3}{6}} P(W > y) dy \\ & = O(1) \int_0^x (1 + y^4) e^{\frac{y^2}{2} - \frac{\gamma y^3}{6}} P(W > y) dy \\ & = O(1)x^4. \end{aligned}$$

For the third term on the right-hand side of (4.22) and the first term on the right-hand side of (4.23), we have the following.

Lemma 4.7.

$$E(1 + W^3)I(W \geq x - C\alpha) = O(1)x^3 \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right).$$

Note that

$$E|W|^3 \leq \sqrt{EW^4} \leq C\sqrt{1 + ns^3d^3\delta^4} \leq C(1 + ns^2d^2\delta^3) \leq C.$$

The fourth term on the right-hand side of (4.22) and the second term on the right-hand side of (4.23) are bounded from (4.4) by

$$CP(Z_\gamma > x - C\alpha) \leq \frac{C}{x} \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right).$$

The fifth and sixth terms on the right-hand side of (4.22) are bounded in a manner similar as for R_4 (cf. (4.17)) by

$$\begin{aligned} & \frac{C}{\gamma^2} P(W \leq -1/\gamma + O(\alpha))I(\gamma > 0) \\ & \quad + \frac{C}{\gamma^2} P(W \geq -1/\gamma - O(\alpha))I(\gamma < 0) \\ & \leq C|\gamma| \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right). \end{aligned}$$

In summary, we have

$$\frac{|R_1|}{\exp(-x^2/2 + \gamma x^3/6)} \leq C\alpha^2 \left(\frac{1}{\alpha^2} mns^4 d^4 \delta^5 x + x^3 \right) \leq Cmns^4 d^4 \delta^5 x,$$

where we use $n^2 s^4 d^4 \delta^6 x^2 \leq Cns^2 d^2 \delta^3 \leq Cmns^4 d^4 \delta^5$ (cf. (3.4)). The bound (4.5), hence the theorem, is proved by combining (4.8) and the bounds on $|R_1| - |R_5|$.

4.2. Proofs of lemmas

In the following, we prove the lemmas stated in the proof above.

Proof of Lemma 4.1. Denote the Kolmogorov distance between two probability distributions by

$$d_K(\mathcal{L}(X), \mathcal{L}(Y)) := \sup_{x \in \mathbb{R}} |P(X \leq x) - P(Y \leq x)|.$$

For $\beta > 0$ to be chosen, let

$$g_\beta(w) = \begin{cases} 1 & w \leq x, \\ 1 + (x - w)/\beta & x < w \leq x + \beta, \\ 0 & w > x + \beta. \end{cases}$$

Let $F := F_{g_\beta}$ be the bounded solution to

$$F'(w) - wF(w) = g_\beta(w) - Eg_\beta(Z), \tag{4.24}$$

where $Z \sim N(0, 1)$. From Lemma 2.5 of Chen, Goldstein and Shao [13], we have

$$|F'(w + v) - F'(w)| \leq |v| \left(1 + |w| + \frac{1}{\beta} \int_0^1 I_{[x, x+\beta]}(w + rv) dr \right). \tag{4.25}$$

Replacing w by W and taking expectations on both sides of the equation (4.24), we have

$$\begin{aligned} P(W \leq x) - \Phi(x) &\leq Eg_\beta(W) - Eg_\beta(Z) + Eg_\beta(Z) - \Phi(x) \\ &\leq EF'(W) - EWF(W) + P(x \leq Z \leq x + \beta) \\ &\leq EF'(W) - EWF(W) + C\beta. \end{aligned} \tag{4.26}$$

Let $U \sim \text{Unif}[0, 1]$ be independent of all else. By (LD1), (LD2), $E\xi_i = 0$ and Taylor's expansion, we have

$$\begin{aligned} EWF(W) &= \sum_{i=1}^n E\xi_i F(W) = \sum_{i=1}^n E\xi_i [F(W) - F(W - \xi_{A_i})] = \sum_{i=1}^n E\xi_i \xi_{A_i} F'(W - U\xi_{A_i}) \\ &= \sum_{i=1}^n \sum_{j \in A_i} E\xi_i \xi_j E F'(W - \xi_{A_{ij}}) + \sum_{i=1}^n \sum_{j \in A_i} E\xi_i \xi_j [F'(W - U\xi_{A_i}) - F'(W - \xi_{A_{ij}})]. \end{aligned}$$

From $EW^2 = \sum_{i=1}^n \sum_{j \in A_i} E\xi_i \xi_j = 1$, we have

$$\begin{aligned} EF'(W) - EWF(W) &= \sum_{i=1}^n \sum_{j \in A_i} E\xi_i \xi_j E[F'(W) - F'(W - \xi_{A_{ij}})] \\ &\quad - \sum_{i=1}^n \sum_{j \in A_i} E\xi_i \xi_j [F'(W - U\xi_{A_i}) - F'(W - \xi_{A_{ij}})]. \end{aligned}$$

From (4.25) and the boundedness conditions in (2.1) and (3.5), we have

$$|EF'(W) - EWF(W)| \leq Cns^2 d^2 \delta^3 \left(1 + \frac{1}{\beta} P(W \in [x - Csd\delta, x + \beta + Csd\delta]) \right).$$

Using

$$P(W \in [x - Csd\delta, x + \beta + Csd\delta]) \leq 2d_K(\mathcal{L}(W), N(0, 1)) + C(sd\delta + \beta),$$

we have

$$\begin{aligned} |EF'(W) - EWF(W)| &\leq Cns^2 d^2 \delta^3 + \frac{Cns^2 d^2 \delta^3 sd\delta}{\beta} + Cns^2 d^2 \delta^3 \frac{d_K(\mathcal{L}(W), N(0, 1))}{\beta}. \end{aligned} \tag{4.27}$$

From (4.26) and (4.27), we have

$$P(W \leq x) - \Phi(x) \leq C\beta + Cns^2 d^2 \delta^3 + \frac{Cns^2 d^2 \delta^3 sd\delta}{\beta} + Cns^2 d^2 \delta^3 \frac{d_K(\mathcal{L}(W), N(0, 1))}{\beta}.$$

From a similar argument for the lower bound, we have

$$|P(W \leq x) - \Phi(x)| \leq C\beta + Cns^2 d^2 \delta^3 + \frac{Cns^2 d^2 \delta^3 sd\delta}{\beta} + Cns^2 d^2 \delta^3 \frac{d_K(\mathcal{L}(W), N(0, 1))}{\beta}.$$

Taking supremum over x , choosing $\beta = 2Cns^2 d^2 \delta^3$, solving the resulting recursive inequality for $d_K(\mathcal{L}(W), N(0, 1))$, and noting that $sd\delta \leq ns^2 d^2 \delta^3$ from (3.4), we arrive at

$$d_K(\mathcal{L}(W), N(0, 1)) \leq Cns^2 d^2 \delta^3. \quad \square$$

Proof of Lemma 4.2. We only prove for the case $\gamma > 0$. The case $\gamma < 0$ can be proved similarly. For $w < -\frac{1}{\gamma} + \gamma$, because

$$\begin{aligned} f(-1/\gamma) &= 0, & f(-1/\gamma + \gamma) &= \gamma(1 - Eh_\alpha(Z_\gamma)), \\ f(-1/\gamma + 2\gamma) &= \gamma^2 f(-1/\gamma + \gamma) + \gamma(1 - Eh_\alpha(Z_\gamma)), \\ f(-1/\gamma + 3\gamma) &= 2\gamma^2 f(-1/\gamma + 2\gamma) + \gamma(1 - Eh_\alpha(Z_\gamma)), \end{aligned}$$

we have (cf. (4.15))

$$\frac{1}{\gamma}(f(w + \gamma) - f(w)) - wf(w) = O(1).$$

For $-\frac{1}{\gamma} + \gamma \leq w$ and $w_0 \leq w < w_0 + \gamma$ such that $w_0 \in \{\gamma\mathbb{Z} - \frac{1}{\gamma}\}$, we have, from the construction of f (cf. (4.14)),

$$\begin{aligned} &\frac{1}{\gamma}(f(w + \gamma) - f(w)) - wf(w) \\ &= \frac{1}{\gamma}[f(w_0 + \gamma) - f(w_0)] - wf(w_0) + (w - w_0) \left\{ \frac{1}{\gamma}[f'(w_0 + \gamma) - f'(w_0)] - wf'(w_0) \right\} \\ &\quad + \frac{(w - w_0)^2}{2} \left\{ \frac{1}{\gamma}[f''(w_0 + \gamma) - f''(w_0)] - wf''(w_0) \right\} \\ &\quad + \left[-\frac{3(w - w_0)^3}{2\gamma} + \frac{5(w - w_0)^4}{2\gamma^2} - \frac{(w - w_0)^5}{\gamma^3} \right] \\ &\quad \times \left\{ \frac{1}{\gamma}[(f''(w_0 + 2\gamma) - f''(w_0 + \gamma)) - (f''(w_0 + \gamma) - f''(w_0))] \right. \\ &\quad \left. - w(f''(w_0 + \gamma) - f''(w_0)) \right\} \\ &=: H_1 + H_2 + H_3 + H_4. \end{aligned}$$

Note that f satisfies (4.9) on \mathcal{S} . We have

$$\begin{aligned} H_1 &= \frac{1}{\gamma}[f(w_0 + \gamma) - f(w_0)] - w_0 f(w_0) - (w - w_0)f(w_0) \\ &= h_\alpha(w_0) - Eh_\alpha(Z_\gamma) + O(|\gamma|) \sum_{i=-K}^K |f(w_0 + i \cdot \gamma)| \\ &= h_\alpha(w) - Eh_\alpha(Z_\gamma) + O(1)I(|w - x| \leq C\alpha) + O(|\gamma|) \sum_{i=-K}^K |f(w_0 + i \cdot \gamma)|. \end{aligned}$$

For H_2 , from the expression of f' on \mathcal{S} (cf. (4.12)) and using again the fact that f satisfies (4.9) on \mathcal{S} , we have

$$\begin{aligned} H_2 &= \frac{w - w_0}{2\gamma} \left\{ \frac{1}{\gamma} [(f(w_0 + 2\gamma) - f(w_0)) - (f(w_0 + \gamma) - f(w_0 - \gamma))] \right. \\ &\quad \left. - w(f(w_0 + \gamma) - f(w_0 - \gamma)) \right\} \\ &= O(1)[h_\alpha(w_0 + \gamma) - h_\alpha(w_0 - \gamma)] + O(|\gamma|) \sum_{i=-K}^K |f(w_0 + i \cdot \gamma)| \\ &= O(1)I(|w - x| \leq C\alpha) + O(|\gamma|) \sum_{i=-K}^K |f(w_0 + i \cdot \gamma)|. \end{aligned}$$

Similarly, from (4.13),

$$\begin{aligned} H_3 &= \frac{(w - w_0)^2}{2\gamma^2} \left\{ \frac{1}{\gamma} [(f(w_0 + 2\gamma) - 2f(w_0 + \gamma) + f(w_0)) \right. \\ &\quad \left. - (f(w_0 + \gamma) - 2f(w_0) + f(w_0 - \gamma))] \right. \\ &\quad \left. - w(f(w_0 + \gamma) - 2f(w_0) + f(w_0 - \gamma)) \right\} \\ &= O(1)I(|w - x| \leq C\alpha) + O(|\gamma|) \sum_{i=-K}^K |f(w_0 + i \cdot \gamma)|, \end{aligned}$$

and

$$H_4 = O(1)I(|w - x| \leq C\alpha) + O(|\gamma|) \sum_{i=-K}^K |f(w_0 + i \cdot \gamma)|.$$

Equation (4.16) is proved by combining the above estimates and observing that the right-hand side is bounded. \square

Proof of Lemma 4.3. We only prove for the case $\gamma > 0$. The case $\gamma < 0$ can be proved similarly. Recall the definition of f . If $w_0 - \gamma > x - \alpha$, then we use $|f(w_0)| \leq C$. If $w_0 \leq -1/\gamma$, then $f(w_0) = 0$. If $-1/\gamma \leq w_0 - \gamma \leq x - \alpha$, then

$$f(w_0) = \frac{\gamma P(Z_\gamma \leq w_0 - \gamma)}{P(Z_\gamma = w_0 - \gamma)} [1 - Eh_\alpha(Z_\gamma)].$$

Recall the proof of (4.11), if $-1/\gamma \leq w_0 - \gamma \leq 0$, then

$$0 \leq f(w_0) \leq 2P(Z_\gamma > x - \alpha).$$

If $0 < w_0 - \gamma \leq x - \alpha$, then by (4.7),

$$|f(w_0)| \leq C e^{\frac{w_0^2}{2} - \frac{\gamma w_0^3}{6}} P(Z_\gamma > x - \alpha).$$

The lemma is proved by combining the above bounds and noting that $|W - W_0| \leq \gamma$. □

Proof of Lemma 4.4. Similar to the proof of Lemma 5.2 of Chen, Fang and Shao [11] and use (3.7), we have, for some $\epsilon \in [0, 1]$,

$$\begin{aligned} & \int_0^{[x]} y^k e^{\frac{y^2}{2} - \frac{\gamma y^3}{6}} P(W > y) dy \\ & \leq \sum_{j=1}^{[x]} j^k \int_{j-1}^j e^{\frac{y^2}{2} - \frac{\gamma y^3}{6} - jy} e^{jy} P(W > y) dy \\ & \leq \sum_{j=1}^{[x]} j^k e^{\frac{(j-1)^2}{2} - \frac{\gamma(j-\epsilon)^3}{6} - j(j-1)} \int_{j-1}^j e^{jy} P(W > y) dy \\ & \leq 2 \sum_{j=1}^{[x]} j^k e^{-\frac{j^2}{2} - \frac{\gamma(j-\epsilon)^3}{6}} \int_{-\infty}^{\infty} e^{jy} P(W > y) dy = 2 \sum_{j=1}^{[x]} j^k e^{-\frac{j^2}{2} - \frac{\gamma(j-\epsilon)^3}{6}} \frac{1}{j} E e^{jW} \\ & = O(1) \sum_{j=1}^{[x]} j^{k-1} e^{\frac{j^3}{6} - \frac{\gamma(j-\epsilon)^3}{6}} = O(1)x^k. \end{aligned}$$

Similarly, we have

$$\begin{aligned} & \int_{[x]}^x y^k e^{\frac{y^2}{2} - \frac{\gamma y^3}{6}} P(W > y) dy \\ & \leq x^k \int_{[x]}^x e^{\frac{y^2}{2} - \frac{\gamma y^3}{6} - xy} e^{xy} P(W > y) dy \\ & \leq x^k e^{\frac{[x]^2}{2} - \frac{\gamma([x]+\epsilon)^3}{6} - x[x]} \int_{[x]}^x e^{xy} P(W > y) dy \leq 2x^k e^{-\frac{x^2}{2} - \frac{\gamma([x]+\epsilon)^3}{6}} \int_{-\infty}^{\infty} e^{xy} P(W > y) dy \\ & = O(1)x^k. \end{aligned}$$

This finishes the proof. □

Proof of Lemma 4.5. The proof is by Taylor’s expansion and by exploiting the local dependence structure (LD1)–(LD3) in Section 3.1. As we will see below, by the crucial choice of γ , Z_γ matches the moments of W up to the third order and the first and second derivative terms of f vanish.

Let U_1, U_2 be independent $\sim \text{Unif}[0, 1]$ and independent of all else. By Taylor's expansion,

$$\begin{aligned} & E \frac{1}{\gamma} (f(W + \gamma) - f(W)) \\ &= E \frac{1}{\gamma} \left[\gamma f'(W) + \frac{\gamma^2}{2} f''(W) + \gamma^3 (1 - U_2) U_2 f^{(3)}(W + \gamma U_1 U_2) \right] \\ &= E f'(W) + \frac{\gamma}{2} E f''(W) + O(\gamma^2) E |f^{(3)}(W + O(|\gamma|))|. \end{aligned} \tag{4.28}$$

By the local dependence structure (LD1)–(LD3) in Section 3.1, $E\xi_i = 0$, Taylor's expansion and the boundedness conditions in (2.1) and (3.5), we have

$$\begin{aligned} EWf(W) &= \sum_{i=1}^n E\xi_i [f(W) - f(W - \xi_{A_i})] \\ &= \sum_{i=1}^n E\xi_i \left[\xi_{A_i} f'(W - \xi_{A_i}) + \frac{\xi_{A_i}^2}{2} f''(W - \xi_{A_i}) + O(s^3 d^3 \delta^3) |f^{(3)}(W + O(sd\delta))| \right] \\ &= \sum_{i=1}^n \sum_{j \in A_i} E\xi_i \xi_j f'(W - \xi_{A_i}) + \frac{1}{2} \sum_{i=1}^n \sum_{j \in A_i} \sum_{k \in A_i} E\xi_i \xi_j \xi_k f''(W - \xi_{A_i}) \\ &\quad + O(ns^3 d^3 \delta^4) E |f^{(3)}(W + O(sd\delta))| \\ &=: B_1 + B_2 + O(ns^3 d^3 \delta^4) E |f^{(3)}(W + O(sd\delta))|. \end{aligned}$$

For B_1 , by a similar expansion as above, we have

$$\begin{aligned} & \sum_{i=1}^n \sum_{j \in A_i} E\xi_i \xi_j f'(W - \xi_{A_i}) \\ &= \sum_{i=1}^n \sum_{j \in A_i} E\xi_i \xi_j E f'(W - \xi_{A_{ij}}) + \sum_{i=1}^n \sum_{j \in A_i} E\xi_i \xi_j [f'(W - \xi_{A_i}) - f'(W - \xi_{A_{ij}})] \\ &= E f'(W) + \sum_{i=1}^n \sum_{j \in A_i} E\xi_i \xi_j E [f'(W - \xi_{A_{ij}}) - f'(W)] \\ &\quad + \sum_{i=1}^n \sum_{j \in A_i} E\xi_i \xi_j [f'(W - \xi_{A_i}) - f'(W - \xi_{A_{ij}})] \\ &= E f'(W) - \sum_{i=1}^n \sum_{j \in A_i} E\xi_i \xi_j E \xi_{A_{ij}} f''(W - \xi_{A_{ij}}) + O(ns^3 d^3 \delta^4) E |f^{(3)}(W + O(sd\delta))| \\ &\quad + \sum_{i=1}^n \sum_{j \in A_i} E\xi_i \xi_j (\xi_{A_{ij}} - \xi_{A_i}) f''(W - \xi_{A_{ij}}). \end{aligned}$$

For B_2 , we have

$$\begin{aligned}
 & \frac{1}{2} \sum_{i=1}^n \sum_{j \in A_i} \sum_{k \in A_i} E \xi_i \xi_j \xi_k f''(W - \xi_{A_i}) \\
 &= \frac{1}{2} \sum_{i=1}^n \sum_{j \in A_i} \sum_{k \in A_i} E \xi_i \xi_j \xi_k E f''(W - \xi_{A_{ijk}}) \\
 & \quad + \frac{1}{2} \sum_{i=1}^n \sum_{j \in A_i} \sum_{k \in A_i} E \xi_i \xi_j \xi_k [f''(W - \xi_{A_i}) - f''(W - \xi_{A_{ijk}})] \\
 &= \frac{1}{2} \sum_{i=1}^n \sum_{j \in A_i} \sum_{k \in A_i} E \xi_i \xi_j \xi_k E f''(W) + \frac{1}{2} \sum_{i=1}^n \sum_{j \in A_i} \sum_{k \in A_i} E \xi_i \xi_j \xi_k E [f''(W - \xi_{A_{ijk}}) - f''(W)] \\
 & \quad + \frac{1}{2} \sum_{i=1}^n \sum_{j \in A_i} \sum_{k \in A_i} E \xi_i \xi_j \xi_k [f''(W - \xi_{A_i}) - f''(W - \xi_{A_{ijk}})] \\
 &= \frac{1}{2} \sum_{i=1}^n \sum_{j \in A_i} \sum_{k \in A_i} E \xi_i \xi_j \xi_k E f''(W) + O(ns^3 d^3 \delta^4) E |f^{(3)}(W + O(sd\delta))|.
 \end{aligned}$$

Similarly,

$$\begin{aligned}
 & - \sum_{i=1}^n \sum_{j \in A_i} E \xi_i \xi_j E \xi_{A_{ij}} f''(W - \xi_{A_{ij}}) \\
 &= - \sum_{i=1}^n \sum_{j \in A_i} \sum_{k \in A_{ij}} E \xi_i \xi_j E \xi_k [f''(W - \xi_{A_{ij}}) - f''(W - \xi_{A_{ijk}})] \\
 &= O(ns^3 d^3 \delta^4) E |f^{(3)}(W + O(sd\delta))|,
 \end{aligned}$$

and

$$\begin{aligned}
 & \sum_{i=1}^n \sum_{j \in A_i} E \xi_i \xi_j (\xi_{A_{ij}} - \xi_{A_i}) f''(W - \xi_{A_{ij}}) \\
 &= \sum_{i=1}^n \sum_{j \in A_i} E \xi_i \xi_j (\xi_{A_{ij}} - \xi_{A_i}) f''(W) + O(ns^3 d^3 \delta^4) E |f^{(3)}(W + O(sd\delta))|.
 \end{aligned}$$

Recall γ from (3.6). Combining the above estimates, we have

$$EWf(W) = Ef'(W) + \frac{\gamma}{2} Ef''(W) + O(ns^3 d^3 \delta^4) E |f^{(3)}(W + O(sd\delta))|. \tag{4.29}$$

By (4.28) and (4.29), we conclude that

$$\begin{aligned} & E\left[\frac{1}{\gamma}(f(W + \gamma) - f(W)) - Wf(W)\right] \\ &= O(\gamma^2)Ef^{(3)}(W + O(|\gamma|)) + O(ns^3d^3\delta^4)E|f^{(3)}(W + O(sd\delta))| \\ &= O(n^2s^4d^4\delta^6)E|f^{(3)}(W + O(ns^2d^2\delta^3))|, \end{aligned}$$

where we use $|\gamma| \leq Cns^2d^2\delta^3$, $sd\delta \leq Cns^2d^2\delta^3$ and $ns^3d^3\delta^4 \leq C(ns^2d^2\delta^3)^2$ from (3.4). \square

Proof of Lemma 4.6. We only prove for the case $\gamma > 0$. The case $\gamma < 0$ can be proved similarly. Note that from the construction of f (cf. (4.14)),

$$f^{(3)}(w) = O(1)\frac{f''(w_0 + \gamma) - f''(w_0)}{\gamma}.$$

For $w_0 \leq -1/\gamma$, from the arguments at the beginning of the proof of Lemma 4.2, $\frac{f''(w_0 + \gamma) - f''(w_0)}{\gamma} = O(1/\gamma^2)$. For $w_0 \geq -1/\gamma + \gamma$, from the construction of f (cf. (4.13)) and the equation (4.9) for $w \in \mathcal{S}$, we have

$$\begin{aligned} & \frac{f''(w_0 + \gamma) - f''(w_0)}{\gamma} \\ &= \frac{[f(w_0 + 2\gamma) - 2f(w_0 + \gamma) + f(w_0)] - [f(w_0 + \gamma) - 2f(w_0) + f(w_0 - \gamma)]}{\gamma^3} \\ &= \frac{[(w_0 + \gamma)f(w_0 + \gamma) + h_\alpha(w_0 + \gamma)] - [w_0f(w_0) + h_\alpha(w_0)]}{\gamma^2} \\ & \quad - \frac{[w_0f(w_0) + h_\alpha(w_0)] - [(w_0 - \gamma)f(w_0 - \gamma) + h_\alpha(w_0 - \gamma)]}{\gamma^2}. \end{aligned} \tag{4.30}$$

Rearranging terms, using $|h''_\alpha(w)| \leq \frac{8}{\alpha^2}I(x - \alpha \leq w \leq x + \alpha)$ and that f solves (4.9) on \mathcal{S} , we have

$$\begin{aligned} (4.30) &= O\left(\frac{1}{\alpha^2}\right)I(|w_0 - x| \leq C\alpha) + \frac{1}{\gamma}[f(w_0 + \gamma) - f(w_0)] + \frac{w_0}{\gamma^2}[f(w_0 + \gamma) - f(w_0)] \\ & \quad - \frac{w_0 - \gamma}{\gamma^2}[f(w_0) - f(w_0 - \gamma)] \\ &= O\left(\frac{1}{\alpha^2}\right)I(|w_0 - x| \leq C\alpha) + [w_0f(w_0) + h_\alpha(w_0) - Eh_\alpha(Z_\gamma)]\left(1 + \frac{w_0}{\gamma}\right) \\ & \quad - \frac{w_0 - \gamma}{\gamma}[(w_0 - \gamma)f(w_0 - \gamma) + h_\alpha(w_0 - \gamma) - Eh_\alpha(Z_\gamma)]. \end{aligned}$$

Rearranging terms, using $|h'_\alpha(w)| \leq \frac{2}{\alpha}I(x - \alpha \leq w \leq x + \alpha)$ and that f solves (4.9) on \mathcal{S} , we have

$$\begin{aligned}
 (4.30) &= O\left(\frac{1}{\alpha^2}\right)I(|w_0 - x| \leq C\alpha) + w_0f(w_0) + (w_0 - \gamma)f(w_0 - \gamma) + w_0f(w_0 - \gamma) \\
 &\quad + h_\alpha(w_0) - Eh_\alpha(Z_\gamma) + h_\alpha(w_0 - \gamma) - Eh_\alpha(Z_\gamma) \\
 &\quad + \frac{w_0}{\gamma}(f(w_0) - f(w_0 - \gamma)) + \frac{w_0}{\gamma}(h_\alpha(w_0) - h_\alpha(w_0 - \gamma)) \\
 &= O\left(\frac{1}{\alpha^2}\right)I(|w_0 - x| \leq C\alpha) + w_0f(w_0) + (w_0 - \gamma)f(w_0 - \gamma) + w_0f(w_0 - \gamma) \\
 &\quad + h_\alpha(w_0) - Eh_\alpha(Z_\gamma) + h_\alpha(w_0 - \gamma) - Eh_\alpha(Z_\gamma) \\
 &\quad + O\left(\frac{1}{\alpha}\right)xI(|w_0 - x| \leq C\alpha) + w_0^2[(w_0 - \gamma)f(w_0 - \gamma) + h_\alpha(w_0 - \gamma) - Eh_\alpha(Z_\gamma)].
 \end{aligned}$$

The lemma is proved by replacing w by W , w_0 by W_0 , and taking expectations. □

Proof of Lemma 4.7. We only prove for the case $\gamma > 0$. The case $\gamma < 0$ can be proved similarly. By Proposition 3.1,

$$P(W > y) \leq \frac{Ee^{xW}}{e^{xy}} \leq C \exp\left(\frac{x^2}{2} + \frac{\gamma x^3}{6} - xy\right).$$

Therefore,

$$\begin{aligned}
 E(1 + W^3)I(W \geq x - C\alpha) &= (1 + y^3)P(W > y)|_{y=x-C\alpha} + \int_{x-C\alpha}^\infty 3y^3P(W > y) dy \\
 &= O(1)x^3 \exp\left(-\frac{x^2}{2} + \frac{\gamma x^3}{6}\right).
 \end{aligned}$$
□

Acknowledgments

We thank the two anonymous referees for their careful reading of the manuscript and for their valuable suggestions which led to many improvements. Fang X. was partially supported by Hong Kong RGC ECS 24301617 and GRF 14302418, a CUHK direct grant and a CUHK start-up grant. Shao Q. M. was partially supported by Hong Kong RGC GRF 14302515 and 14304917, and a CUHK direct grant.

References

[1] Barbour, A.D. (1986). Asymptotic expansions based on smooth functions in the central limit theorem. *Probab. Theory Related Fields* **72** 289–303. MR0836279 <https://doi.org/10.1007/BF00699108>

- [2] Barbour, A.D., Holst, L. and Janson, S. (1992). *Poisson Approximation. Oxford Studies in Probability* 2. New York: Oxford University Press. MR1163825
- [3] Barbour, A.D., Karoński, M. and Ruciński, A. (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B* 47 125–145. MR1047781 [https://doi.org/10.1016/0095-8956\(89\)90014-2](https://doi.org/10.1016/0095-8956(89)90014-2)
- [4] Barbour, A.D., Luczak, M.J. and Xia, A. (2018). Multivariate approximation in total variation, I: Equilibrium distributions of Markov jump processes. *Ann. Probab.* 46 1351–1404. MR3785590 <https://doi.org/10.1214/17-AOP1204>
- [5] Barbour, A.D., Luczak, M.J. and Xia, A. (2018). Multivariate approximation in total variation, II: Discrete normal approximation. *Ann. Probab.* 46 1405–1440. MR3785591 <https://doi.org/10.1214/17-AOP1205>
- [6] Barbour, A.D. and Xia, A. (2019). Multivariate approximation in total variation using local dependence. *Electron. J. Probab.* 24 Paper No. 27, 35. MR3933206 <https://doi.org/10.1214/19-EJP284>
- [7] Braverman, A. and Dai, J.G. (2017). Stein’s method for steady-state diffusion approximations of $M/Ph/n + M$ systems. *Ann. Appl. Probab.* 27 550–581. MR3619795 <https://doi.org/10.1214/16-AAP1211>
- [8] Chatterjee, S. (2007). Stein’s method for concentration inequalities. *Probab. Theory Related Fields* 138 305–321. MR2288072 <https://doi.org/10.1007/s00440-006-0029-y>
- [9] Chatterjee, S. (2008). A new method of normal approximation. *Ann. Probab.* 36 1584–1610. MR2435859 <https://doi.org/10.1214/07-AOP370>
- [10] Chen, H. and Zhang, N. (2015). Graph-based change-point detection. *Ann. Statist.* 43 139–176. MR3285603 <https://doi.org/10.1214/14-AOS1269>
- [11] Chen, L.H.Y., Fang, X. and Shao, Q.-M. (2013). From Stein identities to moderate deviations. *Ann. Probab.* 41 262–293. MR3059199 <https://doi.org/10.1214/12-AOP746>
- [12] Chen, L.H.Y., Fang, X. and Shao, Q.-M. (2013). Moderate deviations in Poisson approximation: A first attempt. *Statist. Sinica* 23 1523–1540. MR3222808
- [13] Chen, L.H.Y., Goldstein, L. and Shao, Q.-M. (2011). *Normal Approximation by Stein’s Method. Probability and Its Applications (New York)*. Heidelberg: Springer. MR2732624 <https://doi.org/10.1007/978-3-642-15007-4>
- [14] Chen, L.H.Y. and Shao, Q.-M. (2004). Normal approximation under local dependence. *Ann. Probab.* 32 1985–2028. MR2073183 <https://doi.org/10.1214/009117904000000450>
- [15] Chen, L.H.Y. and Shao, Q.-M. (2007). Normal approximation for nonlinear statistics using a concentration inequality approach. *Bernoulli* 13 581–599. MR2331265 <https://doi.org/10.3150/07-BEJ5164>
- [16] Chen, L.H.Y. and Röllin, A. (2010). Stein couplings for normal approximation. Available at <https://arxiv.org/abs/1003.6039>.
- [17] Cramér, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actual. Sci. Ind.* 736 5–23.
- [18] Esseen, C.-G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace–Gaussian law. *Acta Math.* 77 1–125. MR0014626 <https://doi.org/10.1007/BF02392223>
- [19] Ho, S.T. and Chen, L.H.Y. (1978). An L_p bound for the remainder in a combinatorial central limit theorem. *Ann. Probab.* 6 231–249. MR0478291 <https://doi.org/10.1214/aop/1176995570>
- [20] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* 19 293–325. MR0026294 <https://doi.org/10.1214/aoms/1177730196>
- [21] Petrov, V.V. (1975). *Sums of Independent Random Variables*. New York: Springer. Translated from the Russian by A. A. Brown, *Ergebnisse der Mathematik und ihrer Grenzgebiete*, Band 82. MR0388499
- [22] Rinott, Y. and Rotar, V. (2003). On Edgeworth expansions for dependency-neighborhoods chain structures and Stein’s method. *Probab. Theory Related Fields* 126 528–570. MR2001197 <https://doi.org/10.1007/s00440-003-0271-5>

- [23] Rio, E. (2009). Upper bounds for minimal distances in the central limit theorem. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** 802–817. MR2548505 <https://doi.org/10.1214/08-AIHP187>
- [24] Röllin, A. (2005). Approximation of sums of conditionally independent variables by the translated Poisson distribution. *Bernoulli* **11** 1115–1128. MR2189083 <https://doi.org/10.3150/bj/1137421642>
- [25] Röllin, A. (2007). Translated Poisson approximation using exchangeable pair couplings. *Ann. Appl. Probab.* **17** 1596–1614. MR2358635 <https://doi.org/10.1214/105051607000000258>
- [26] Shao, Q.-M. (2010). Stein’s method, self-normalized limit theory and applications. In *Proceedings of the International Congress of Mathematicians. Volume IV* 2325–2350. New Delhi: Hindustan Book Agency. MR2827974
- [27] Shao, Q.M., Zhang, M. and Zhang, Z.S. (2018). Cramér-type moderate deviation theorems for non-normal approximation. Available at <https://arxiv.org/abs/1809.07966>.
- [28] 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 (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability Theory* 583–602. MR0402873
- [29] Tang, H.K. and Siegmund, D. (2001). Mapping quantitative trait loci in oligogenic models. *Biostatistics* **2** 147–162.
- [30] Tu, I. and Siegmund, D. (1999). The maximum of a function of a Markov chain and application to linkage analysis. *Adv. in Appl. Probab.* **31** 510–531. MR1724565 <https://doi.org/10.1239/aap/1029955145>
- [31] Zhang, Z.S. (2019). Cramér-type moderate deviation of normal approximation for exchangeable pairs. Available at <https://arxiv.org/abs/1901.09526>.

Received August 2019 and revised January 2020