The Annals of Probability
2013, Vol. 41, No. 1, 262–293
DOI: 10.1214/12-AOP746

FROM STEIN IDENTITIES TO MODERATE DEVIATIONS

BY LOUIS H. Y. CHEN[1], XIAO FANG[1] AND QI-MAN SHAO[2]

National University of Singapore, National University of Singapore and Hong Kong University of Science and Technology

Stein's method is applied to obtain a general Cramér-type moderate deviation result for dependent random variables whose dependence is defined in terms of a Stein identity. A corollary for zero-bias coupling is deduced. The result is also applied to a combinatorial central limit theorem, a general system of binary codes, the anti-voter model on a complete graph, and the Curie–Weiss model. A general moderate deviation result for independent random variables is also proved.

1. Introduction. Moderate deviations date back to Cramér (1938) who obtained expansions for tail probabilities for sums of independent random variables about the normal distribution. For independent and identically distributed random variables $X_1, \dots, X_n$ with $EX_i = 0$ and $\operatorname{Var}(X_i) = 1$ such that $Ee^{t_0|X_1|} \le c < \infty$ for some $t_0 > 0$, it follows from Petrov [(1975), Chapter 8, equation (2.41)] that

$$\frac{P(W_n > x)}{1 - \Phi(x)} = 1 + O(1)\bigl(1 + x^3\bigr)/\sqrt{n} \tag{1.1}$$

for $0 \le x \le a_0 n^{1/6}$, where $W_n = (X_1 + \cdots + X_n)/\sqrt{n}$ and Φ is the standard normal distribution function, $a_0 > 0$ depends on c and t_0 and $O(1)$ is bounded by a constant depending on c and t_0. The range $0 \le x \le a_0 n^{1/6}$ and the order of the error term $O(1)(1 + x^3)/\sqrt{n}$ are optimal.

The proof of (1.1) depends on the conjugate method and a Berry–Esseen bound, while the classical proof of Berry–Esseen bound for independent random variables uses the Fourier transform. However, for dependent random variables, Stein's method performs much better than the method of Fourier transform. Stein's method was introduced by Charles Stein in 1972 and further developed by him in 1986. Extensive applications of Stein's method to obtain Berry–Esseen-type bounds for dependent random variables can be found in, for example, Diaconis (1977), Baldi, Rinott and Stein (1989), Barbour (1990), Dembo and Rinott (1996),

Received November 2009; revised February 2012.

[1]Supported in part by Grant C-389-000-010-101 from the National University of Singapore.

[2]Supported in part by Hong Kong RGC CERG—602608 and 603710.

MSC2010 subject classifications. Primary 60F10; secondary 60F05.

Key words and phrases. Stein's method, Stein identity, moderate deviations, Berry–Esseen bounds, zero-bias coupling, exchangeable pairs, dependent random variables, combinatorial central limit theorem, general system of binary codes, anti-voter model, Curie–Weiss model.

Goldstein and Reinert (1997), Chen and Shao (2004), Chatterjee (2008) and Nourdin and Peccati (2009). Recent applications to concentration of measures and large deviations can be found in, for example, Chatterjee (2007) and Chatterjee and Dey (2010). Expositions of Stein's method and its applications in normal and other distributional approximations can be found in Diaconis and Holmes (2004) and Barbour and Chen (2005).

In this paper we apply Stein's method to obtain a Cramér-type moderate deviation result for dependent random variables whose dependence is defined in terms of an identity, called Stein identity, which plays a central role in Stein's method. A corollary for zero-bias coupling is deduced. The result is then applied to a combinatorial central limit theorem, the anti-voter model, a general system of binary codes and the Curie–Weiss model. The bounds obtained in these examples are as in (1.1) and therefore may be optimal (see Remark 4.1). It is noted that Raič (2007) also used Stein's method to obtain moderate deviation results for dependent random variables. However, the dependence structure he considered is related to local dependence and is of a different nature from what we assume through the Stein identity.

This paper is organized as follows. Section 2 is devoted to a description of Stein's method and to the construction of Stein identities using zero-bias coupling and exchangeable pairs. Section 3 presents a general Cramér-type moderate deviation result and a corollary for zero-bias coupling. The result is applied to the four examples mentioned above in Section 4. Although the general Cramér-type moderate deviation result cannot be applied directly to unbounded independent random variables, the proof of the general result can be adapted to prove (1.1) under less stringent conditions, thereby extending a result of Linnik (1961). These are also presented in Section 4. The rest of the paper is devoted to proofs.

2. Stein's method and Stein's identity. Let W be the random variable of interest and Z be another random variable. In approximating $\mathcal{L}(W)$ by $\mathcal{L}(Z)$ using Stein's method, the difference between $Eh(W)$ and $Eh(Z)$ for a class of functions h is expressed as

$$Eh(W) - Eh(Z) = E\{Lf_h(W)\}, \tag{2.1}$$

where L is a linear operator and f_h a bounded solution of the equation $Lf = h - Eh(Z)$. It is known that for $N(0,1)$, $Lf(w) = f'(w) - wf(w)$ [see Stein (1972)] and for Poisson(λ), $Lf(w) = \lambda f(w+1) - wf(w)$; see Chen (1975). However, L is not unique. For example, for normal approximation L can also be the generator of the Ornstein–Uhlenbeck process, and for Poisson approximation L, the generator of an immigration-death process. The solution f_h will then be expressed in terms of a Markov process. This generator approach to Stein's method is due to Barbour (1988, 1990).

By (2.1), bounding $Eh(W) - Eh(Z)$ is equivalent to bounding $E\{Lf_h(W)\}$. To bound the latter one finds another operator $\tilde{L}$ such that $E\{\tilde{L}f(W)\} = 0$, for a class of functions f including f_h, and write $\tilde{L} = L - R$ for a suitable operator R. The error term $E\{Lf_h(W)\}$ is then expressed as $E\{Rf_h(W)\}$. The equation

$$E\{\tilde{L}f(W)\} = 0 \tag{2.2}$$

for a class of functions f including f_h, is called a Stein identity for $\mathcal{L}(W)$. For normal approximation, there are four methods for constructing a Stein identity: the direct method [Stein (1972)], zero-bias coupling [Goldstein and Reinert (1997) and Goldstein (2005)], exchangeable pairs [Stein (1986)] and Stein coupling [Chen and Röllin (2010)]. We discuss below the construction of Stein identities using zero-bias coupling and exchangeable pairs. As proved in Goldstein and Reinert (1997), for W with $EW = 0$ and $\mathrm{Var}(W) = 1$, there always exists W^* such that

$$E\bigl(Wf(W)\bigr) = Ef'\bigl(W^*\bigr) \tag{2.3}$$

for all bounded absolutely continuous f with bounded derivative f'. The distribution of W^* is called W-zero-biased. If W and W^* are defined on the same probability space (zero-bias coupling), we may write $\Delta = W^* - W$. Then by (2.3), we obtain the Stein identity

$$E\bigl(Wf(W)\bigr) = Ef'(W + \Delta) = E\int_{-\infty}^{\infty} f'(W + t)\, d\mu(t|W), \tag{2.4}$$

where $\mu(\cdot|W)$ is the conditional distribution of Δ given W. Here $\tilde{L}(w) = \int_{-\infty}^{\infty} f'(w + t)\, d\mu(t|W = w) - wf(w)$.

The method of exchangeable pairs [Stein (1986)] consists of constructing W' such that (W, W') is exchangeable. Then for any anti-symmetric function $F(\cdot,\cdot)$, that is, $F(w, w') = -F(w', w)$,

$$EF\bigl(W, W'\bigr) = 0,$$

if the expectation exists. Suppose that there exist a constant λ $(0 < \lambda < 1)$ and a random variable R such that

$$E\bigl(W - W'|W\bigr) = \lambda\bigl(W - E(R|W)\bigr). \tag{2.5}$$

Then for all f,

$$E\bigl\{\bigl(W - W'\bigr)\bigl(f(W) + f\bigl(W'\bigr)\bigr)\bigr\} = 0,$$

provided the expectation exists. This gives the Stein identity

$$\begin{aligned} E\bigl(Wf(W)\bigr) &= -\frac{1}{2\lambda}E\bigl\{\bigl(W - W'\bigr)\bigl(f\bigl(W'\bigr) - f(W)\bigr)\bigr\} + E\bigl(Rf(W)\bigr) \\ &= E\int_{-\infty}^{\infty} f'(W + t)\hat{K}(t)\, dt + E\bigl(Rf(W)\bigr) \end{aligned} \tag{2.6}$$

for all absolutely continuous functions f for which expectations exist, where $\hat{K}(t) = \frac{1}{2\lambda}\Delta(I(0 \le t \le \Delta) - I(\Delta \le t < 0))$ and $\Delta = W' - W$. In this case, $\tilde{L}(w) = \int_{-\infty}^{\infty} f'(w + t)E(\hat{K}(t)|W = w)\, dt + E(R|W = w)f(w) - wf(w)$.

Both Stein identities (2.4) and (2.6) are special cases of

$$E\big(Wf(W)\big) = E\int_{-\infty}^{\infty} f'(W+t)\,d\hat{\mu}(t) + E\big(Rf(W)\big), \tag{2.7}$$

where $\hat{\mu}$ is a random measure. We will prove a moderate deviation result by assuming that W satisfies the Stein identity (2.7).

3. A Cramér-type moderate deviation theorem. Let W be a random variable of interest. Assume that there exist a deterministic positive constant δ, a random positive measure $\hat{\mu}$ with support $[-\delta, \delta]$ and a random variable R such that

$$E\big(Wf(W)\big) = E\int_{|t|\le\delta} f'(W+t)\,d\hat{\mu}(t) + E\big(Rf(W)\big) \tag{3.1}$$

for all absolutely continuous function f for which the expectation of either side exists. Let

$$D = \int_{|t|\le\delta} d\hat{\mu}(t). \tag{3.2}$$

THEOREM 3.1. *Suppose that there exist constants δ_1, δ_2 and $\theta \ge 1$ such that*

$$\big|E(D|W) - 1\big| \le \delta_1\big(1+|W|\big), \tag{3.3}$$

$$\begin{aligned} &\big|E(R|W)\big| \le \delta_2\big(1+|W|\big) \quad \textit{or} \\ &\big|E(R|W)\big| \le \delta_2\big(1+W^2\big) \quad \textit{and} \quad \delta_2|W| \le \alpha < 1 \end{aligned} \tag{3.4}$$

and

$$E(D|W) \le \theta. \tag{3.5}$$

Then

$$\frac{P(W>x)}{1-\Phi(x)} = 1 + O_\alpha(1)\theta^3\big(1+x^3\big)(\delta+\delta_1+\delta_2) \tag{3.6}$$

for $0 \le x \le \theta^{-1}\min(\delta^{-1/3}, \delta_1^{-1/3}, \delta_2^{-1/3})$, *where* $O_\alpha(1)$ *denotes a quantity whose absolute value is bounded by a universal constant which depends on* α *only under the second alternative of* (3.4).

REMARK 3.1. Theorem 3.1 is intended for bounded random variables but with very general dependence assumptions. For this reason, the support of the random measure $\hat{\mu}$ is assumed to be within $[-\delta, \delta]$ where δ is typically of the order of $1/\sqrt{n}$ due to standardization. In order for the normal approximation to work, $E(D|W)$ should be close to 1 and $E(R|W)$ small. This is reflected in δ_1 and δ_2 which are assumed to be small.

For zero-bias coupling, $D = 1$ and $R = 0$, so conditions (3.3), (3.4) and (3.5) are satisfied with $\delta_1 = \delta_2 = 0$ and $\theta = 1$. Therefore, we have:

COROLLARY 3.1. *Let W and W^* be defined on the same probability space satisfying* (2.3). *Assume that $EW = 0$, $EW^2 = 1$ and $|W - W^*| \le \delta$ for some constant δ. Then*

$$\frac{P(W \ge x)}{1 - \Phi(x)} = 1 + O(1)\big(1 + x^3\big)\delta$$

for $0 \le x \le \delta^{-1/3}$.

REMARK 3.2. For an exchangeable pair (W, W') satisfying (2.5) and $|\Delta| \le \delta$, (3.1) is satisfied with $D = \Delta^2/(2\lambda)$.

REMARK 3.3. Although one cannot apply Theorem 3.1 directly to unbounded random variables, one can adapt the proof of Theorem 3.1 to give a proof of (1.1) for independent random variables assuming the existence of the moment generating functions of $|X_i|^{1/2}$ thereby extending a result of Linnik (1961). This result is given in Proposition 4.6. The proof also suggests the possibility of extending Theorem 3.1 to the case where the support of $\hat{\mu}$ may not be bounded.

4. Applications. In this section we apply Theorem 3.1 to four cases of dependent random variables, namely, a combinatorial central limit theorem, the anti-voter model on a complete graph, a general system of binary codes, and the Curie–Weiss model. The proofs of the results for the third and the fourth example will be given in the last section. At the end of this section, we will present a moderate deviation result for sums of independent random variables and the proof will also be given in the last section.

4.1. *Combinatorial central limit theorem.* Let $\{a_{ij}\}_{i,j=1}^n$ be an array of real numbers satisfying $\sum_{j=1}^n a_{ij} = 0$ for all i and $\sum_{i=1}^n a_{ij} = 0$ for all j. Set $c_0 = \max_{i,j} |a_{ij}|$ and $W = \sum_{i=1}^n a_{i\pi(i)}/\sigma$, where π is a uniform random permutation of $\{1, 2, \dots, n\}$ and $\sigma^2 = E(\sum_{i=1}^n a_{i\pi(i)})^2$. In Goldstein (2005) W is coupled with the zero-biased W^* in such a way that $|\Delta| = |W^* - W| \le 8c_0/\sigma$. Therefore, by Corollary 3.1 with $\delta = 8c_0/\sigma$, we have

$$\frac{P(W \ge x)}{1 - \Phi(x)} = 1 + O(1)\big(1 + x^3\big)c_0/\sigma \tag{4.1}$$

for $0 \le x \le (\sigma/c_0)^{1/3}$.

4.2. *Anti-voter model on a complete graph.* Consider the anti-voter model on a complete graph with n vertices, $1, \dots, n$ and $(n-1)n/2$ edges. Let X_i be a random variable taking value 1 or -1 at the vertex i, $i = 1, \dots, n$.

Let $X = (X_1, \dots, X_n)$, where X_i takes values 1 or -1. The anti-voter model in discrete time is described as the following Markov chain: in each step, uniformly pick a vertex I and an edge connecting it to J, and then change X_I to $-X_J$. Let $U = \sum_{i=1}^n X_i$ and $W = U/\sigma$, where $\sigma^2 = \mathrm{Var}(U)$. Let $W' = (U - X_I - X_J)/\sigma$, where I is uniformly distributed on $\{1, 2, \dots, n\}$ independent of other random variables. Consider the case where the distribution of X is the stationary distribution. Then as shown in Rinott and Rotar (1997), (W, W') is an exchangeable pair and

$$E(W - W'|W) = \frac{2}{n} W. \tag{4.2}$$

According to (2.6), (3.1) is satisfied with $\delta = 2/\sigma$ and $R = 0$. To check conditions (3.3) and (3.5), let T denote the number of 1's among $X_1, \dots, X_n$, a be the number of edges connecting two 1's, b be the number of edges connecting two -1's and c be the number of edges connecting 1 and -1. Since it is a complete graph, $a = \frac{T(T-1)}{2}$, $b = \frac{(n-T)(n-T-1)}{2}$. Therefore [see, e.g., Rinott and Rotar (1997)]

$$\begin{aligned} E\big[(W - W')^2|X\big] &= \frac{1}{\sigma^2} E\big[(U' - U)^2|X\big] = \frac{4}{\sigma^2}\frac{2a + 2b}{n(n-1)} \\ &= \frac{1}{\sigma^2}\frac{2U^2 + 2n^2 - 4n}{n(n-1)} = \frac{2\sigma^2 W^2 + 2n^2 - 4n}{\sigma^2 n(n-1)}, \end{aligned} \tag{4.3}$$

$$\begin{aligned} E(D|W) - 1 &= \frac{n}{4} E\big((W' - W)^2|W\big) - 1 \\ &= \frac{W^2}{2(n-1)} - \frac{2\sigma^2(n-1) - (n^2 - 2n)}{2\sigma^2(n-1)}. \end{aligned} \tag{4.4}$$

Noting that $E(E(D|W) - 1) = 0$ and $EW^2 = 1$, we have $\sigma^2 = \frac{n^2-2n}{2n-3}$. Hence

$$E(D|W) - 1 = \frac{W^2}{2(n-1)} - \frac{1}{2(n-1)}, \tag{4.5}$$

which means that (3.3) is satisfied with $\delta_1 = O(n^{-1/2})$. Thus, we have the following moderate deviation result.

PROPOSITION 4.1. *We have*

$$\frac{P(W \ge x)}{1 - \Phi(x)} = 1 + O(1)\big(1 + x^3\big)/\sqrt{n}$$

for $0 \le x \le n^{1/6}$.

4.3. *A general system of binary codes.* In Chen, Hwang and Zacharovas (2011), a general system of binary codes is defined as follows. Suppose each nonnegative integer x is coded by a binary string consisting of 0's and 1's. Let $\tilde{S}(x)$ denote the number of 1's in the resulting coding string of x, and let

$$\tilde{\mathbf{S}} = \big(\tilde{S}(0), \tilde{S}(1), \dots\big). \tag{4.6}$$

For each nonnegative integer n, define $\tilde{S}_n = \tilde{S}(X)$, where X is a random integer uniformly distributed over the set $\{0, 1, \dots, n\}$. The general system of binary codes introduced by Chen, Hwang and Zacharovas (2011) is one in which

$$\tilde{S}_{2m-1} = \tilde{S}_{m-1} + \mathcal{I} \qquad \text{in distribution} \qquad \text{for all } m \geq 1, \tag{4.7}$$

where $\mathcal{I}$ is an independent Bernoulli$(1/2)$ random variable. Chen, Hwang and Zacharovas (2011) proved the asymptotic normality of $\tilde{S}_n$. Here, we apply Theorem 3.1 to obtain the following Cramér moderate deviation result. For $n \geq 1$, let integer k be such that $2^{k-1} - 1 < n \leq 2^k - 1$, and let $\tilde{W}_n = (\tilde{S}_n - k/2)/\sqrt{k/4}$.

PROPOSITION 4.2. *Under the assumption* (4.7), *we have*

$$\frac{P(\tilde{W}_n \geq x)}{1 - \Phi(x)} = 1 + O(1)\big(1 + x^3\big)\frac{1}{\sqrt{k}} \tag{4.8}$$

for $0 \leq x \leq k^{1/6}$.

As an example of this system of binary codes, we consider the binary expansion of a random integer X uniformly distributed over $\{0, 1, \dots, n\}$. For $2^{k-1} - 1 < n \leq 2^k - 1$, write X as

$$X = \sum_{i=1}^{k} X_i 2^{k-i},$$

and let $S_n = X_1 + \cdots + X_k$. Set $W_n = (S_n - k/2)/\sqrt{k/4}$. It is easy to verify that S_n satisfies (4.7). A Berry–Esseen bound for W_n was first obtained by Diaconis (1977). Proposition 4.2 provides a Cramér moderate deviation result for W_n. Other examples of this system of binary codes include the binary reflected Gray code and a coding system using translation and complementation. Detailed descriptions of these codes are given in Chen, Hwang and Zacharovas (2011).

4.4. *Curie–Weiss model.* Consider the Curie–Weiss model for n spins $\Sigma = (\sigma_1, \sigma_2, \dots, \sigma_n) \in \{-1, 1\}^n$. The joint distribution of Σ is given by

$$Z_{\beta,h}^{-1} \exp\Bigg(\frac{\beta}{n} \sum_{1 \leq i < j \leq n} \sigma_i \sigma_j + \beta h \sum_{i=1}^{n} \sigma_i\Bigg),$$

where $Z_{\beta,h}$ is the normalizing constant, and $\beta > 0, h \in \mathbb{R}$ are called the inverse of temperature and the external field, respectively. We are interested in the total magnetization $S = \sum_{i=1}^{n} \sigma_i$. We divide the region $\beta > 0, h \in \mathbb{R}$ into three parts, and for each part, we list the concentration property and the limiting distribution of S under proper standardization. Consider the solution(s) to the equation

$$m = \tanh\bigl(\beta(m+h)\bigr). \tag{4.9}$$

Case 1. $0 < \beta < 1, h \in \mathbb{R}$ or $\beta \geq 1, h \neq 0$. There is a unique solution m_0 to (4.9) such that $m_0 h \geq 0$. In this case, S/n is concentrated around m_0 and has a Gaussian limit under proper standardization.

Case 2. $\beta > 1, h = 0$. There are two nonzero solutions to (4.9), $m_1 < 0 < m_2$, where $m_1 = -m_2$. Given condition on $S < 0$ ($S > 0$, resp.), S/n is concentrated around m_1 (m_2, resp.) and has a Gaussian limit under proper standardization.

Case 3. $\beta = 1, h = 0$. S/n is concentrated around 0, but the limit distribution is not Gaussian.

We refer to Ellis (1985) for the concentration of measure results, Ellis and Newman (1978a, 1978b) for the results on limiting distributions. See also Chatterjee and Shao (2011) for a Berry–Esseen-type bound when the limiting distribution is not Gaussian. Here we focus on the Gaussian case and prove the following two Cramér moderate deviation results for cases 1 and 2.

PROPOSITION 4.3. *In case* 1, *define*

$$W = \frac{S - nm_0}{\sigma}, \tag{4.10}$$

where

$$\sigma^2 = \frac{n(1-m_0^2)}{1-(1-m_0^2)\beta}. \tag{4.11}$$

Then we have

$$\frac{P(W \geq x)}{1-\Phi(x)} = 1 + O(1)\bigl(1+x^3\bigr)/\sqrt{n} \tag{4.12}$$

for $0 \leq x \leq n^{1/6}$.

PROPOSITION 4.4. *In case* 2, *define*

$$W_1 = \frac{S - nm_1}{\sigma_1}, \qquad W_2 = \frac{S - nm_2}{\sigma_2}, \tag{4.13}$$

where

$$\sigma_1^2 = \frac{n(1-m_1^2)}{1-(1-m_1^2)\beta}, \qquad \sigma_2^2 = \frac{n(1-m_2^2)}{1-(1-m_2^2)\beta}. \tag{4.14}$$

Then we have

$$\frac{P(W_1 \geq x | S < 0)}{1-\Phi(x)} = 1 + O(1)\bigl(1+x^3\bigr)/\sqrt{n} \tag{4.15}$$

and

$$\frac{P(W_2 \geq x|S>0)}{1-\Phi(x)} = 1 + O(1)\big(1+x^3\big)/\sqrt{n} \tag{4.16}$$

for $0 \leq x \leq n^{1/6}$.

4.5. *Independent random variables*. Moderate deviation for independent random variables has been extensively studied in literature [see, e.g., Petrov (1975), Chapter 8] based on the conjugated method. Here, we will adapt the proof of Theorem 3.1 to prove the following moderate deviation result, which is a variant of those in the literature [see again Petrov (1975), Chapter 8].

PROPOSITION 4.5. *Let* $\xi_i, 1 \leq i \leq n$ *be independent random variables with* $E\xi_i = 0$ *and* $Ee^{t_n|\xi_i|} < \infty$ *for some* t_n *and for each* $1 \leq i \leq n$. *Assume that*

$$\sum_{i=1}^{n} E\xi_i^2 = 1. \tag{4.17}$$

Then

$$\frac{P(W \geq x)}{1-\Phi(x)} = 1 + O(1)\big(1+x^3\big)\gamma e^{4x^3\gamma} \tag{4.18}$$

for $0 \leq x \leq t_n$, *where* $\gamma = \sum_{i=1}^{n} E|\xi_i|^3 e^{x|\xi_i|}$.

We deduce (1.1) under less stringent conditions from Proposition 4.5 and extend a result of Linnik (1961) to independent but not necessarily identically distributed random variables.

PROPOSITION 4.6. *Let* $X_i, 1 \leq i \leq n$ *be a sequence of independent random variables with* $EX_i = 0$. *Put* $S_n = \sum_{i=1}^{n} X_i$ *and* $B_n^2 = \sum_{i=1}^{n} EX_i^2$. *Assume that there exists positive constants* c_1, c_2 *and* t_0 *such that*

$$B_n^2 \geq c_1^2 n, \qquad Ee^{t_0\sqrt{|X_i|}} \leq c_2 \qquad \text{for } 1 \leq i \leq n. \tag{4.19}$$

Then

$$\frac{P(S_n/B_n \geq x)}{1-\Phi(x)} = 1 + O(1)\big(1+x^3\big)/\sqrt{n} \tag{4.20}$$

for $0 \leq x \leq (c_1 t_0^2)^{1/3} n^{1/6}$, *where* $O(1)$ *is bounded by a constant depending on* c_2 *and* $c_1 t_0^2$. *In particular, we have*

$$\frac{P(S_n/B_n \geq x)}{1-\Phi(x)} \to 1 \tag{4.21}$$

uniformly in $0 \leq x \leq o(n^{1/6})$.

PROOF. The main idea is first truncating X_i and then applying Proposition 4.5 to the truncated sequence. Let

$$\tau_n = (c_1^2 t_0 n)^{1/3} 2^{-2/3}, \qquad \bar{X}_i = X_i \mathbf{1}(|X_i| \le \tau_n^2), \qquad \bar{S}_n = \sum_{i=1}^n \bar{X}_i.$$

Observe that

$$\begin{aligned}
&\big|P(S_n/B_n \ge x) - P(\bar{S}_n/B_n \ge x)\big| \\
&\quad \le \sum_{i=1}^n P(|X_i| \ge \tau_n^2) \\
&\quad \le \sum_{i=1}^n e^{-t_0\tau_n} E e^{t_0\sqrt{|X_i|}} \le c_2 n e^{-t_0\tau_n} \\
&\quad = O(1)\big(1-\Phi(x)\big)(1+x^3)/\sqrt{n}
\end{aligned}$$

for $0 \le x \le (c_1 t_0^2)^{1/3} n^{1/6}$; here we used the fact that

$$t_0\tau_n = (c_1 t_0^2)^{2/3} n^{1/3} 2^{-2/3}.$$

Now let $\xi_i = (\bar{X}_i - E\bar{X}_i)/\bar{B}_n$, where $\bar{B}_n^2 = \sum_{i=1}^n \operatorname{Var}(\bar{X}_i)$. It is easy to see that

$$\begin{aligned}
\sum_{i=1}^n |E\bar{X}_i| &\le \sum_{i=1}^n E|X_i|\mathbf{1}(|X_i| \ge \tau_n^2) \\
&\le \sum_{i=1}^n \sup_{s\ge\tau_n} (s^2 e^{-t_0 s}) E e^{t_0\sqrt{|X_i|}} \qquad (4.22)\\
&\le c_2 n c_1 (c_1 t_0^2)^{-1} \sup_{s \ge t_0\tau_n} (s^2 e^{-s}) = c_1 o(n^{-2})
\end{aligned}$$

and similarly, $\bar{B}_n = B_n(1+o(n^{-2}))$. Thus, for $0 \le x \le (c_1 t_0^2)^{1/3} n^{1/6}$

$$\begin{aligned}
x|\xi_i| &\le \frac{2^{1/3}x}{c_1 n^{1/2}} |X_i| \mathbf{1}(|X_i| \le \tau_n^2) + o(1) \le \frac{2^{1/3} x \tau_n}{c_1 n^{1/2}} \sqrt{|X_i|} + o(1) \\
&\le \frac{t_0}{2^{1/3}} \sqrt{|X_i|} + o(1)
\end{aligned}$$

and hence $\gamma = O(n^{-1/2})$. Applying Proposition 4.5 to $\{\xi_i, 1 \le i \le n\}$ gives (4.20). □

REMARK 4.1. As stated previously for (1.1) in the Introduction, the range $0 \le x \le (c_1 t_0^2)^{1/3} n^{1/6}$ and the order of the error term $O(1)(1+x^3)/\sqrt{n}$ in Proposition 4.6 are optimal. By comparing with (1.1) the results in the four examples discussed above may be optimal.

5. Preliminary lemmas. To prove Theorem 3.1, we first need to develop two preliminary lemmas. Our first lemma gives a bound for the moment generating function of W.

LEMMA 5.1. *Let W be a random variable with $E|W| \le C$. Assume that there exist $\delta > 0$, $\delta_1 \ge 0, 0 \le \delta_2 \le 1/4$ and $\theta \ge 1$ such that* (3.1) *and* (3.3)–(3.5) *are satisfied. Then for all $0 < t \le 1/(2\delta)$ satisfying*

$$t\delta_1 + C_\alpha t\theta\delta_2 \le 1/2, \tag{5.1}$$

where

$$C_\alpha = \begin{cases} 12, & \text{under the first alternative of (3.4)}, \\ \dfrac{2(3+\alpha)}{1-\alpha}, & \text{under the second alternative of (3.4)}, \end{cases} \tag{5.2}$$

we have

$$Ee^{tW} \le \exp\bigl(t^2/2 + c_0(t)\bigr), \tag{5.3}$$

where

$$c_0(t) = c_1(C, C_\alpha)\theta\bigl\{\delta_2 t + \delta_1 t^2 + (\delta + \delta_1 + \delta_2)t^3\bigr\}, \tag{5.4}$$

where $c_1(C, C_\alpha)$ is a constant depending only on C and C_α.

PROOF. Fix $a > 0$, $t \in (0, 1/(2\delta)]$ and $s \in (0, t]$, and let $f(w) = e^{s(w\wedge a)}$. Letting $h(s) = Ee^{s(W\wedge a)}$, firstly we prove that $h'(s)$ can be bounded by $sh(s)$ and $EW^2 f(W)$. By (3.1),

$$\begin{aligned}
h'(s) &= E(W\wedge a)e^{s(W\wedge a)} \le E\bigl(Wf(W)\bigr) \\
&= E\int f'(W+t)\,d\hat\mu(t) + E\bigl(Rf(W)\bigr) \\
&= sE\int e^{s(W+t)} I(W+t\le a)\,d\hat\mu(t) + E\bigl(e^{s(W\wedge a)}E(R|W)\bigr) \\
&\le sE\int e^{s[(W+t)\wedge a]}\,d\hat\mu(t) + E\bigl(e^{s(W\wedge a)}E(R|W)\bigr) \\
&\le sE\int e^{s(W\wedge a+\delta)}\,d\hat\mu(t) + E\bigl(e^{s(W\wedge a)}E(R|W)\bigr) \\
&= sE\int e^{s(W\wedge a)}\,d\hat\mu(t) + sE\int e^{s(W\wedge a)}\bigl(e^{s\delta}-1\bigr)\,d\hat\mu(t) \\
&\quad + E\bigl(e^{s(W\wedge a)}E(R|W)\bigr) \\
&\le sEe^{s(W\wedge a)}D + sEe^{s(W\wedge a)}\bigl|e^{s\delta}-1\bigr|D + 2\delta_2 E\bigl((1+W^2)e^{s(W\wedge a)}\bigr),
\end{aligned}$$

where we have applied (3.2) and (3.4) to obtain the last inequality. Now, applying the simple inequality

$$|e^x - 1| \le 2|x| \qquad \text{for } |x| \le 1,$$

and then (3.3), we find that

$$\begin{aligned} E\big(Wf(W)\big) &\le sEe^{s(W\wedge a)}D + sEe^{s(W\wedge a)}2s\delta D + 2\delta_2 E\big((1+W^2)e^{s(W\wedge a)}\big) \\ &\le sEe^{s(W\wedge a)}E(D|W) + 2s^2\theta\delta Ee^{s(W\wedge a)} + 2\delta_2 E\big((1+W^2)e^{s(W\wedge a)}\big) \\ &= sEe^{s(W\wedge a)} + sEe^{s(W\wedge a)}\big[E(D|W)-1\big] \\ &\quad + 2s^2\theta\delta Ee^{s(W\wedge a)} + 2\delta_2 E\big((1+W^2)e^{s(W\wedge a)}\big) \\ &\le sEe^{s(W\wedge a)} + s\delta_1 Ee^{s(W\wedge a)}\big(1+|W|\big) + 2s^2\theta\delta Ee^{s(W\wedge a)} \\ &\quad + 2\delta_2 E\big((1+W^2)e^{s(W\wedge a)}\big). \end{aligned}$$

Note that

$$\begin{aligned} E|W|e^{s(W\wedge a)} &= EWe^{s(W\wedge a)} + 2EW^- e^{s(W\wedge a)} \\ &\le E\big(Wf(W)\big) + 2E|W| \le 2C + E\big(Wf(W)\big). \end{aligned} \tag{5.5}$$

Collecting terms, we obtain

$$\begin{aligned} h'(s) &\le E\big(Wf(W)\big) \\ &\le \big\{\big(s(1+\delta_1+2t\theta\delta)+2\delta_2\big)h(s) + 2\delta_2 EW^2 f(W) + 2Cs\delta_1\big\} \\ &\quad /(1-s\delta_1). \end{aligned} \tag{5.6}$$

Secondly, we show that $EW^2f(W)$ can be bounded by a function of $h(s)$ and $h'(s)$. Letting $g(w) = we^{s(w\wedge a)}$, and then arguing as for (5.6),

$$\begin{aligned} EW^2f(W) &= EWg(W) \\ &= E\int \big(e^{s[(W+t)\wedge a]} + s(W+t)e^{s[(W+t)\wedge a]}I(W+t\le a)\big)\,d\hat{\mu}(t) \\ &\quad + E\big(RWf(W)\big) \\ &\le E\int \big(e^{s(W\wedge a)}e^{s\delta} + s\big[(W+t)\wedge a\big]e^{s(W\wedge a)}e^{s\delta}\big)\,d\hat{\mu}(t) \\ &\quad + E\big(RWf(W)\big) \\ &= e^{s\delta}E\big(f(W) + sf(W)\big((W\wedge a)+\delta\big)\big)D + E\big(RWf(W)\big) \\ &\le \theta e^{0.5}(1+0.5)Ef(W) + s\theta e^{s\delta}E(W\wedge a)f(W) + E\big(RWf(W)\big) \\ &\le 1.5e^{0.5}\theta h(s) + 2s\theta h'(s) + E\big(RWf(W)\big). \end{aligned} \tag{5.7}$$

Note that under the first alternative of (3.4),

$$|E(RWf(W))| \le \delta_2 Ef(W) + 2\delta_2 EW^2 f(W), \tag{5.8}$$

and under the second alternative of (3.4),

$$|E(RWf(W))| \le \alpha Ef(W) + \alpha EW^2 f(W). \tag{5.9}$$

Thus, recalling $\delta_2 \le 1/4$ and $\alpha < 1$, we have

$$EW^2 f(W) \le \frac{C_\alpha}{2}\big(\theta h(s) + s\theta h'(s)\big), \tag{5.10}$$

where C_α is defined in (5.2).

We are now ready to prove (5.3). Substituting (5.10) into (5.6) yields

$$\begin{aligned}(1 - s\delta_1)h'(s) &\le \big(s(1 + \delta_1 + 2t\theta\delta) + 2\delta_2\big)h(s)\\ &\quad + \delta_2 C_\alpha\big(\theta h(s) + s\theta h'(s)\big) + 2Cs\delta_1\\ &= \big(s(1 + \delta_1 + 2t\theta\delta) + 2\delta_2(1 + C_\alpha\theta)\big)h(s)\\ &\quad + C_\alpha s\theta\delta_2 h'(s) + 2Cs\delta_1\\ &\le \big(s(1 + \delta_1 + 2t\theta\delta) + 2\delta_2(1 + C_\alpha\theta)\big)h(s)\\ &\quad + C_\alpha t\theta\delta_2 h'(s) + 2Cs\delta_1.\end{aligned} \tag{5.11}$$

Solving for $h'(s)$, we obtain

$$h'(s) \le \big(sc_1(t) + c_2(t)\big)h(s) + \frac{2Cs\delta_1}{1 - c_3(t)}, \tag{5.12}$$

where

$$\begin{aligned} c_1(t) &= \frac{1 + \delta_1 + 2t\theta\delta}{1 - c_3(t)},\\ c_2(t) &= \frac{2\delta_2(1 + C_\alpha\theta)}{1 - c_3(t)},\\ c_3(t) &= t\delta_1 + C_\alpha t\theta\delta_2.\end{aligned}$$

Now taking t to satisfy (5.1) yields $c_3(t) \le 1/2$, so in particular, $c_i(t)$ is nonnegative for $i = 1, 2$ and $1/(1 - c_3(t)) \le 1 + 2c_3(t)$.

Solving (5.12), we have

$$h(s) \le \exp\bigg(\frac{t^2}{2}c_1(t) + tc_2(t) + 2C\delta_1 t^2\bigg). \tag{5.13}$$

Note that $c_3(t) \le 1/2$, $\delta_2 \le 1/4$ and $\theta \ge 1$. Elementary calculations now give

$$\begin{aligned}
&\frac{t^2}{2}\bigl(c_1(t) - 1\bigr) + tc_2(t) + 2C\delta_1 t^2 \\
&\qquad = \frac{t^2}{2}\frac{\delta_1 + 2t\theta\delta + c_3(t)}{1 - c_3(t)} + \frac{2t\delta_2(1 + C_\alpha\theta)}{1 - c_3(t)} + 2C\delta_1 t^2 \\
&\qquad \le t^2(\delta_1 + 2t\theta\delta + t\delta_1 + C_\alpha t\theta\delta_2) + 4t\delta_2(1 + C_\alpha) + 2C\delta_1 t^2 \\
&\qquad \le c_0(t),
\end{aligned}$$

and hence

$$t^2 c_1(t)/2 + tc_2(t) + 2C\delta_1 t^2 \le t^2/2 + c_0(t),$$

thus proving (5.3) by letting $a \to \infty$. $\square$

LEMMA 5.2. *Suppose that for some nonnegative* δ, δ_1 *and* δ_2, *satisfying* $\max(\delta, \delta_1, \delta_2) \le 1$ *and* $\theta \ge 1$, (5.3) *is satisfied, with* $c_0(t)$ *as in* (5.4), *for all*

$$t \in \bigl[0, \theta^{-1}\min\bigl(\delta^{-1/3}, \delta_1^{-1/3}, \delta_2^{-1/3}\bigr)\bigr]. \tag{5.14}$$

Then for integers $k \ge 1$,

$$\int_0^t u^k e^{u^2/2} P(W \ge u)\,du \le c_2(C, C_\alpha)t^k, \tag{5.15}$$

where $c_2(C, C_\alpha)$ *is a constant depending only on* C *and* C_α *defined in Lemma* 5.1.

PROOF. For t satisfying (5.14), it is easy to see that $c_0(t) \le 5c_1(C, C_\alpha)$, where $c_1(C, C_\alpha)$ is as in Lemma 5.1, and (5.1) is satisfied. Write

$$\begin{aligned}
&\int_0^t u^k e^{u^2/2} P(W \ge u)\,du \\
&\qquad = \int_0^{[t]} u^k e^{u^2/2} P(W \ge u)\,du + \int_{[t]}^t u^k e^{u^2/2} P(W \ge u)\,du,
\end{aligned}$$

where $[t]$ denotes the integer part of t. For the first integral, noting that $\sup_{j-1 \le u \le j} e^{u^2/2 - ju} = e^{(j-1)^2/2 - j(j-1)}$, we have

$$\begin{aligned}
&\int_0^{[t]} u^k e^{u^2/2} P(W \ge u)\,du \\
&\qquad \le \sum_{j=1}^{[t]} j^k \int_{j-1}^j e^{u^2/2 - ju} e^{ju} P(W \ge u)\,du \\
&\qquad \le \sum_{j=1}^{[t]} j^k e^{(j-1)^2/2 - j(j-1)} \int_{j-1}^j e^{ju} P(W \ge u)\,du
\end{aligned} \tag{5.16}$$

$$
\begin{aligned}
&\le 2\sum_{j=1}^{[t]} j^k e^{-j^2/2} \int_{-\infty}^{\infty} e^{ju} P(W \ge u)\, du \\
&= 2\sum_{j=1}^{[t]} j^k e^{-j^2/2} (1/j) E e^{jW} \\
&\le 2\sum_{j=1}^{[t]} j^{k-1} \exp\bigl(-j^2/2 + j^2/2 + c_0(j)\bigr) \\
&\le 2e^{c_0(t)} \sum_{j=1}^{[t]} j^{k-1} \\
&\le c_2(C, C_\alpha) t^k.
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
&\int_{[t]}^{t} u^k e^{u^2/2} P(W \ge u)\, du \\
&\qquad \le t^k \int_{[t]}^{t} e^{u^2/2 - tu} e^{tu} P(W \ge u)\, du \\
&\qquad \le t^k e^{[t]^2/2 - t[t]} \int_{[t]}^{t} e^{tu} P(W \ge u)\, du \\
&\qquad \le 2t^k e^{-t^2/2} \int_{-\infty}^{\infty} e^{tu} P(W \ge u)\, du \\
&\qquad \le c_2(C, C_\alpha) t^k.
\end{aligned}
$$

This completes the proof. $\square$

6. Proofs of results. In this section, let $O_\alpha(1)$ denote universal constants which depend on α only under the second alternative of (3.4).

6.1. *Proof of Theorem* 3.1. If $\theta^{-1} \min(\delta^{-1/3}, \delta_1^{-1/3}, \delta_2^{-1/3}) \le O_\alpha(1)$, then $1/(1 - \Phi(x)) \le 1/(1 - \Phi(O_\alpha(1)))$ for $0 \le x \le O_\alpha(1)$. Moreover, $\theta^3(\delta + \delta_1 + \delta_2) \ge O_\alpha(1)$. Therefore, (3.6) is trivial. Hence, we can assume

$$
\theta^{-1} \min\bigl(\delta^{-1/3}, \delta_1^{-1/3}, \delta_2^{-1/3}\bigr) \ge O_\alpha(1) \tag{6.1}
$$

so that $\delta \le 1, \delta_2 \le 1/4, \delta_1 + 2\delta_2 < 1$, and moreover, $\delta_1 + \delta_2 + \alpha < 1$ under the second alternative of (3.4). Our proof is based on Stein's method. Let $f = f_x$ be the solution to the Stein equation

$$wf(w) - f'(w) = I(w \geq x) - \bigl(1 - \Phi(x)\bigr). \tag{6.2}$$

It is known that

$$\begin{aligned} f(w) &= \begin{cases} \sqrt{2\pi} e^{w^2/2}\bigl(1 - \Phi(w)\bigr)\Phi(x), & w \geq x, \\ \sqrt{2\pi} e^{w^2/2}\bigl(1 - \Phi(x)\bigr)\Phi(w), & w < x, \end{cases} \\ &\leq \frac{4}{1+w}\mathbf{1}(w \geq x) + 3\bigl(1 - \Phi(x)\bigr)e^{w^2/2}\mathbf{1}(0 < w < x) \\ &\quad + 4\bigl(1 - \Phi(x)\bigr)\frac{1}{1+|w|}\mathbf{1}(w \leq 0) \end{aligned} \tag{6.3}$$

by using the following well-known inequality:

$$\bigl(1 - \Phi(w)\bigr)e^{w^2/2} \leq \min\biggl(\frac{1}{2}, \frac{1}{w\sqrt{2\pi}}\biggr), \qquad w > 0.$$

It is also known that $wf(w)$ is an increasing function; see Chen and Shao (2005), Lemma 2.2. By (3.1) we have

$$E\bigl(Wf(W)\bigr) - E\bigl(Rf(W)\bigr) = E\int f'(W+t)\,d\hat{\mu}(t), \tag{6.4}$$

and monotonicity of $wf(w)$ and equation (6.2) imply that

$$f'(W+t) \leq (W+\delta)f(W+\delta) + 1 - \Phi(x) - \mathbf{1}(W \geq x+\delta). \tag{6.5}$$

Recall that $\int d\hat{\mu}(t) = D$. Thus using nonnegativity of $\hat{\mu}$ and combining (6.4) and (6.5), we have

$$\begin{aligned} &E\bigl(Wf(W)\bigr) - E\bigl(Rf(W)\bigr) \\ &\quad \leq E\int \bigl((W+\delta)f(W+\delta) - Wf(W)\bigr)\,d\hat{\mu}(t) + EWf(W)D \\ &\qquad + E\int \bigl\{1 - \Phi(x) - \mathbf{1}(W > x+\delta)\bigr\}\,d\hat{\mu}(t). \end{aligned} \tag{6.6}$$

Now, by (3.2), the expression above can be written

$$\begin{aligned} &E\bigl((W+\delta)f(W+\delta) - Wf(W)\bigr)D \\ &\quad + EWf(W)D + E\bigl\{1 - \Phi(x) - \mathbf{1}(W > x+\delta)\bigr\}D \\ &\quad\quad = 1 - \Phi(x) - P(W > x+\delta) \\ &\quad\quad\quad + E\bigl((W+\delta)f(W+\delta) - Wf(W)\bigr)D + EWf(W)D \\ &\quad\quad\quad + E\bigl\{1 - \Phi(x) - \mathbf{1}(W > x+\delta)\bigr\}(D-1). \end{aligned} \tag{6.7}$$

Therefore, we have

$$
\begin{aligned}
&P(W > x+\delta) - \bigl(1-\Phi(x)\bigr)\\
&\quad\le E\bigl((W+\delta)f(W+\delta) - Wf(W)\bigr)D + EWf(W)(D-1)\\
&\qquad + E\bigl\{1-\Phi(x) - \mathbf{1}(W > x+\delta)\bigr\}(D-1) + ERf(W)\\
&\quad\le \theta E\bigl((W+\delta)f(W+\delta) - Wf(W)\bigr) + \delta_1 E\bigl(|W|\bigl(1+|W|\bigr)f(W)\bigr)\\
&\qquad + \delta_1 E\bigl|1-\Phi(x) - \mathbf{1}(W > x+\delta)\bigr|\bigl(1+|W|\bigr) + \delta_2 E\bigl(2+W^2\bigr)f(W),
\end{aligned}
\tag{6.8}
$$

where we have again applied the monotonicity of $wf(w)$ as well as (3.5), (3.3) and (3.4). Hence we have that

$$
P(W > x+\delta) - \bigl(1-\Phi(x)\bigr) \le \theta I_1 + \delta_1 I_2 + \delta_1 I_3 + \delta_2 I_4, \tag{6.9}
$$

where

$$
\begin{aligned}
I_1 &= E\bigl((W+\delta)f(W+\delta) - Wf(W)\bigr),\\
I_2 &= E\bigl(|W|\bigl(1+|W|\bigr)f(W)\bigr),\\
I_3 &= E\bigl|1-\Phi(x) - \mathbf{1}(W > x+\delta)\bigr|\bigl(1+|W|\bigr)
\end{aligned}
$$

and

$$
I_4 = E\bigl(2+W^2\bigr)f(W).
$$

By (6.3) we have

$$
\begin{aligned}
Ef(W) &\le 4P(W > x) + 4\bigl(1-\Phi(x)\bigr)\\
&\quad + 3\bigl(1-\Phi(x)\bigr)Ee^{W^2/2}\mathbf{1}(0 < W \le x).
\end{aligned}
\tag{6.10}
$$

Note that by (3.1) with $f(w) = w$,

$$
\begin{aligned}
EW^2 &= E\int d\hat{\mu}(t) + E(RW)\\
&= ED + E(RW).
\end{aligned}
$$

Therefore, under the first alternative of (3.4), $EW^2 \le (1+2\delta_1+\delta_2) + (\delta_1 + 2\delta_2)EW^2$, and under the second alternative of (3.4), $EW^2 \le (1+2\delta_1+\delta_2) + (\delta_1 + \delta_2 + \alpha)EW^2$. This shows $EW^2 \le O_\alpha(1)$. Hence the hypotheses of Lemma 5.1 is satisfied with $C = O_\alpha(1)$, and therefore also the conclusion of Lemma 5.2. In particular,

$$
\begin{aligned}
Ee^{W^2/2}\mathbf{1}(0 < W \le x) &\le P(0 < W \le x) + \int_0^x ye^{y^2/2}P(W > y)\,dy\\
&\le O_\alpha(1)(1+x).
\end{aligned}
\tag{6.11}
$$

Similarly, by (6.3) again,

$$EW^2 f(W) \le 4E|W|\mathbf{1}(W > x) + 4\big(1 - \Phi(x)\big)E|W| \\ + 3\big(1 - \Phi(x)\big)EW^2 e^{W^2/2}\mathbf{1}(0 < W \le x)$$

and by Lemma 5.2,

$$\begin{aligned} EW^2 e^{W^2/2}\mathbf{1}(0 < W \le x) &\le \int_0^x (y^3 + 2y)e^{y^2/2} P(W > y)\,dy \\ &\le O_\alpha(1)\big(1 + x^3\big). \end{aligned} \tag{6.12}$$

As to

$$E|W|\mathbf{1}(W > x) \le P(W > x) + EW^2 I(W > x),$$

it follows from Lemma 5.1 that

$$P(W > x) \le e^{-x^2} Ee^{xW} = O_\alpha(1)e^{-x^2/2} \tag{6.13}$$

and

$$\begin{aligned} \int_x^\infty tP(W \ge t)\,dt &\le Ee^{xW} \int_x^\infty te^{-xt}\,dt \\ &= Ee^{xW} x^{-2}(1 + x^2)e^{-x^2} \\ &\le O_\alpha(1)e^{-x^2/2} x^{-2}\big(1 + x^2\big) \\ &\le O_\alpha(1)e^{-x^2/2} \end{aligned} \tag{6.14}$$

for $x \ge 1$. Thus we have for $x > 1$,

$$\begin{aligned} EW^2\mathbf{1}(W > x) &= x^2 P(W > x) + \int_x^\infty 2yP(W > y)\,dy \\ &\le O_\alpha(1)\big(1 + x^2\big)e^{-x^2/2} \le O_\alpha(1)\big(1 + x^3\big)\big(1 - \Phi(x)\big). \end{aligned} \tag{6.15}$$

Clearly, (6.15) remains valid for $0 \le x \le 1$ by the fact that $EW^2\mathbf{1}(W > x) \le EW^2 \le 2$. Combining (6.11)–(6.15), we have

$$I_2 \le O_\alpha(1)\big(1 + x^3\big)\big(1 - \Phi(x)\big). \tag{6.16}$$

Similarly,

$$I_4 \le O_\alpha(1)\big(1 + x^3\big)\big(1 - \Phi(x)\big) \tag{6.17}$$

and

$$\begin{aligned} I_3 &\le \big(1 - \Phi(x)\big)E\big(2 + W^2\big) + E\big(2 + W^2\big)\mathbf{1}(W \ge \delta + x) \\ &\le O_\alpha(1)\big(1 + x^3\big)\big(1 - \Phi(x)\big). \end{aligned} \tag{6.18}$$

Let $g(w) = (wf(w))'$. Then $I_1 = \int_0^\delta Eg(W+t)\,dt$. It is easy to see that [e.g., Chen and Shao (2001)]

$$g(w) = \begin{cases} \big(\sqrt{2\pi}(1+w^2)e^{w^2/2}\big(1-\Phi(w)\big) - w\big)\Phi(x), & w \ge x, \\ \big(\sqrt{2\pi}(1+w^2)e^{w^2/2}\Phi(w) + w\big)\big(1-\Phi(x)\big), & w < x, \end{cases} \tag{6.19}$$

and

$$0 \le \sqrt{2\pi}(1+w^2)e^{w^2/2}\big(1-\Phi(w)\big) - w \le \frac{2}{1+w^3}, \tag{6.20}$$

and we have for $0 \le t \le \delta$,

$$\begin{aligned} &Eg(W+t) \\ &\quad = Eg(W+t)\mathbf{1}\{W+t \ge x\} + Eg(W+t)\mathbf{1}\{W+t \le 0\} \\ &\qquad + Eg(W+t)\mathbf{1}\{0 < W+t < x\} \\ &\quad \le \frac{2}{1+x^3}P(W+t \ge x) + 2\big(1-\Phi(x)\big)P(W+t \le 0) \qquad (6.21) \\ &\qquad + \sqrt{2\pi}\big(1-\Phi(x)\big) \\ &\qquad\quad \times E\big\{\big(1+(W+t)^2 + (W+t)\big)e^{(W+t)^2/2}\mathbf{1}\{0 < W+t < x\}\big\} \\ &\quad = O_\alpha(1)\big(1+x^3\big)\big(1-\Phi(x)\big) \end{aligned}$$

and hence

$$I_1 = O_\alpha(1)\delta\big(1+x^3\big)\big(1-\Phi(x)\big). \tag{6.22}$$

Putting (6.9), (6.16), (6.17), (6.18) and (6.22) together gives

$$P(W \ge x+\delta) - \big(1-\Phi(x)\big) \le O_\alpha(1)\big(1-\Phi(x)\big)\theta\big(1+x^3\big)(\delta+\delta_1+\delta_2),$$

and therefore

$$P(W \ge x) - \big(1-\Phi(x)\big) \le O_\alpha(1)\big(1-\Phi(x)\big)\theta\big(1+x^3\big)(\delta+\delta_1+\delta_2). \tag{6.23}$$

As to the lower bound, similarly to (6.5) and (6.8), we have

$$f'(W+t) \ge (W-\delta)f(W-\delta) + 1 - \Phi(x) - \mathbf{1}(W \ge x-\delta)$$

and

$$\begin{aligned} &P(W > x-\delta) - \big(1-\Phi(x)\big) \\ &\quad \ge \theta E\big((W-\delta)f(W-\delta) - Wf(W)\big) - \delta_1 E\big(|W|\big(1+|W|\big)f(W)\big) \\ &\qquad - \delta_1 E\big|1-\Phi(x) - \mathbf{1}(W > x-\delta)\big|\big(1+|W|\big) - \delta_2 E\big(2+W^2\big)f(W). \end{aligned}$$

Now follwoing the same proof of (6.23) leads to

$$P(W \ge x) - \big(1-\Phi(x)\big) \ge -O_\alpha(1)\big(1-\Phi(x)\big)\theta\big(1+x^3\big)(\delta+\delta_1+\delta_2).$$

This completes the proof of Theorem 3.1.

6.2. *Proof of Proposition* 4.2. For $n \geq 2$, $X \sim U\{0, 1, \ldots, n\}$, let $\tilde{S}_n = \tilde{S}(X)$ be the number of 1's in the binary string of X generated in any system of binary codes satisfying (4.7). Without loss of generality, assume that

$$\tilde{S}(0) = 0. \tag{6.24}$$

Condition (4.7) allows $\tilde{S}(X)$ to be represented in terms of the labels of the nodes in a binary tree described as follows. Let $\tilde{T}$ be an infinite binary tree. For $k \geq 0$, the nodes of $\tilde{T}$ in the kth generation are denoted by (from left to right) $(V_{k,0}, \ldots, V_{k,2^k-1})$. Each node is labeled by 0 or 1. Assume $\tilde{T}$ satisfies:

(C1) the root is labeled by 0;
(C2) the labels of two siblings are different;
(C3) infinite binary subtrees of $\tilde{T}$ with roots $\{V_{k,0} : k \geq 0\}$ are the same as $\tilde{T}$.

For $2^{k-1} - 1 < n \leq 2^k - 1$, represent $0, \ldots, n$ by the nodes $V_{k,0}, \ldots, V_{k,n}$, respectively. Then $\tilde{S}(X)$ is the sum of 1's in the shortest path from $V_{k,X}$ to the root of the tree. Condition (C3) implies that $\tilde{S}(X)$ does not depend on k so that the representation is well defined.

We consider two extreme cases. Define a binary tree T by always assigning 0 to the left sibling and 1 to the right sibling. Then the number of 1's in the binary string of X is that in the binary expansion of X. Denote it by $S_n (= S(X))$. Next, define a binary tree $\bar{T}$ by assigning $V_{k,0} = 0$, $V_{k,1} = 1$ for all k and assigning 1 to the left sibling and 0 to the right sibling for all other nodes. Let the number of 1's in the binary string of X on $\bar{T}$ be $\bar{S}_n (= \bar{S}(X))$. Both T and $\bar{T}$ are infinity binary trees satisfying C1, C2 and C3, and both S_n and $\bar{S}_n$ satisfy (4.7). It is easy to see that for all integers $n \geq 0$,

$$S_n \leq_{st} \tilde{S}_n \leq_{st} \bar{S}_n, \tag{6.25}$$

where $\leq_{st}$ denotes stochastic ordering. Therefore, it suffices to prove Cramér moderate deviation results for W_n and $\bar{W}_n$ where $W_n = (S_n - \frac{k}{2})/\sqrt{\frac{k}{4}}$ and $\bar{W}_n = (\bar{S}_n - \frac{k}{2})/\sqrt{\frac{k}{4}}$. We suppress the subscript n in the following and follow Diaconis (1977) in constructing the exchangeable pair (W, W'). Let I be a random variable uniformly distributed over the set $\{1, 2, \ldots, k\}$ and independent of X, and let the random variable X' be defined by

$$X' = \sum_{i=1}^{k} X_i' 2^{k-i},$$

where

$$X_i' = \begin{cases} X_i, & \text{if } i \neq I, \\ 1, & \text{if } i = I, X_I = 0 \text{ and } X + 2^{k-I} \leq n, \\ 0, & \text{else.} \end{cases} \tag{6.26}$$

Let $S' = S - X_I + X_I'$, $W' = (S' - k/2)/\sqrt{k/4}$. As proved in Diaconis (1977), (W, W') is an exchangeable pair and

$$E(W - W'|W) = \lambda\Big(W - \Big(-\frac{E(Q|W)}{\sqrt{k}}\Big)\Big), \tag{6.27}$$

$$\frac{1}{2\lambda}E\big((W - W')^2|W\big) - 1 = -\frac{E(Q|W)}{k}, \tag{6.28}$$

where $\lambda = 2/k$ and $Q = \sum_{i=1}^{k} I(X_i = 0, X + 2^{k-i} > n)$. From Lemma 6.1 and Theorem 3.1 [with $\delta = O(k^{-1/2})$, $\delta_1 = O(k^{-1})$, $\delta_2 = O(k^{-1/2})$],

$$\frac{P(W \ge x)}{1 - \Phi(x)} = 1 + O(1)\big(1 + x^3\big)\frac{1}{\sqrt{k}}$$

for $0 \le x \le k^{1/6}$. Repeat the above argument for $-W$, and we have

$$\frac{P(W \le -x)}{1 - \Phi(x)} = 1 + O(1)\big(1 + x^3\big)\frac{1}{\sqrt{k}}$$

for $0 \le x \le k^{1/6}$.

Next, we notice that S and $\bar{S}$ can be written as, with $X \sim U\{0, 1, \dots, n\}$,

$$S = I\big(0 \le X \le 2^{k-1} - 1\big)S + I\big(2^{k-1} \le X \le n\big)S$$

and

$$\bar{S} = I\big(0 \le X \le 2^{k-1} - 1\big)\bar{S} + I\big(2^{k-1} \le X \le n\big)\bar{S}.$$

Therefore,

$$\begin{aligned} -W - \frac{1}{\sqrt{k/4}} &\\ = \Big(-\frac{1}{2} &+ I\big(0 \le X \le 2^{k-1} - 1\big)\Big(\frac{k-1}{2} - S\Big) \\ &+ I\big(2^{k-1} \le X \le n\big)\Big(\frac{k-1}{2} - S\Big)\Big)\Big/\sqrt{k/4} \end{aligned}$$

and

$$\begin{aligned} \bar{W} = \Big(-\frac{1}{2} &+ I\big(0 \le X \le 2^{k-1} - 1\big)\Big(\bar{S} - \frac{k-1}{2}\Big) \\ &+ I\big(2^{k-1} \le X \le n\big)\Big(\bar{S} - \frac{k-1}{2}\Big)\Big)\Big/\sqrt{k/4}. \end{aligned}$$

Conditioning on $0 \le X \le 2^{k-1} - 1$, both the distributions of $S(X)$ and $\bar{S}(X)$ are Binomial$(k - 1, 1/2)$, which yields

$$\mathcal{L}\Big(\frac{k-1}{2} - S\Big|0 \le X \le 2^{k-1} - 1\Big) = \mathcal{L}\Big(\bar{S} - \frac{k-1}{2}\Big|0 \le X \le 2^{k-1} - 1\Big).$$

On the other hand, when $2^{k-1} \le X \le n$, $\bar{S}(X) = k - 1 - S(X)$. Therefore, $\bar{W}$ has the same distribution as $-W - 1/\sqrt{\frac{k}{4}}$, which implies Cramér moderate deviation results also holds for $\bar{W}$. Thus finishes the proof of Proposition 4.2.

LEMMA 6.1. *We have* $E(Q|S) = O(1)(1 + |W|)$.

PROOF. Write

$$n = \sum_{i \ge 1} 2^{k-p_i}$$

with $1 = p_1 < p_2 < \cdots \le p_{k_1}$ the positions of the ones in the binary expansion of n, where $k_1 \le k$. Recall that X is uniformly distributed over $\{0, 1, \dots, n\}$, and that

$$X = \sum_{i=1}^{k} X_i 2^{k-i}$$

with exactly S of the indicator variables $X_1, \dots, X_k$ equal to 1.

We say that X falls in category i, $i = 1, \dots, k_1$, when

$$X_{p_1} = 1, \qquad X_{p_2} = 1, \dots, X_{p_{i-1}} = 1 \quad \text{and} \quad X_{p_i} = 0. \tag{6.29}$$

We say that X falls in category $k_1 + 1$ if $X = n$. This special category is nonempty only when $S = k_1$, and in this case, $Q = k - k_1$, which gives the last term in (6.30).

Note that if X is in category i for $i \le k_1$, then, since X can be no greater than n, the digits of X and n match up to the p_ith, except for the digit in place p_i, where n has a one, and X a zero. Further, up to this digit, n has $p_i - i$ zeros, and so X has $a_i = p_i - i + 1$ zeros. Changing any of these a_i zeros, except the zero in position p_i to ones, results in a number n or greater, while changing any other zeros, since digit p_i of n is one and of X zero, does not. Hence Q is at most a_i when X falls in category i. Since X has S ones in its expansion, $i - 1$ of which are accounted for by (6.29), the remaining $S - (i - 1)$ are uniformly distributed over the $k - p_i = k - (i - 1) - a_i$ remaining digits $\{X_{p_i+1}, \dots, X_k\}$. Thus, we have the inequality

$$E(Q|S) \le \frac{1}{A} \sum_{i \ge 1} \binom{k - (i-1) - a_i}{S - (i-1)} a_i + \frac{I(S = k_1)}{A}(k - k_1), \tag{6.30}$$

where

$$A = \sum_{i \ge 1} \binom{k - (i-1) - a_i}{S - (i-1)} + I(S = k_1)$$

and $1 = a_1 \le a_2 \le a_3 \le \cdots$.

Note that if $k_1 = k$, the last term of (6.30) equals 0. When $k_1 < k$, we have

$$\frac{I(S=k_1)}{A}(k-k_1) \le \binom{k-1}{k_1}^{-1}(k-k_1) \le 1, \tag{6.31}$$

so we omit this term in the following argument.

We consider two cases.

Case 1: $S \ge k/2$. As $a_i \ge 1$ for all i, there are at most $k+1$ nonzero terms in the sum (6.30). Divide the summands into two groups, those for which $a_i \le 2\log_2 k$ and those with $a_i > 2\log_2 k$. The first group can sum to no more than $2\log_2 k$ because the sum is like weighted average of a_i.

For the second group, note that

$$\begin{aligned}
&\binom{k-(i-1)-a_i}{S-(i-1)}\Big/A \\
&\quad\le \binom{k-(i-1)-a_i}{S-(i-1)}\Big/\binom{k-1}{S} \\
&\quad= \prod_{j=1}^{a_i-1}\left(\frac{k-S-j}{k-j}\right)\prod_{j=0}^{i-2}\left(\frac{S-j}{k-(a_i-1)-1-j}\right) \\
&\quad\le \frac{1}{2^{a_i-1}} \le \frac{1}{k^2},
\end{aligned} \tag{6.32}$$

where the second inequality follows from $S \ge k/2$, and the last inequality from $a_i > 2\log_2 k$. Therefore, the sum of the second group of terms is bounded by 1.

Case 2: $S < k/2$. Divide the sum on the right-hand side into two groups according to whether $i \le 2\log_2 k$ or $i > 2\log_2 k$. Clearly,

$$\begin{aligned}
&\binom{k-(i-1)-a_i}{S-(i-1)}\Big/A \\
&\quad\le \prod_{j=0}^{i-2}\left(\frac{S-j}{k-1-j}\right)\prod_{j=1}^{a_i-1}\left(\frac{k-S-j}{k-(i-1)-j}\right) \\
&\quad\le 1/2^{i-1}
\end{aligned}$$

using the assumption $S < k/2$ and the fact that $S \ge i-1$. The above inequality is true for all i, so the summation for the part where $i > 2\log_2 k$ is bounded by 1.

Next we consider $i \le 2\log_2 k$. When $S \ge k(\frac{\log a_i}{a_i-1}) + 2\log_2 k$, we have $a_i(\frac{k-S-1}{k-(i-1)-1})^{a_i-1} \le 1$. Solving S from the inequality $a_i(\frac{k-S-1}{k-(i-1)-1})^{a_i-1} \le 1$, we see that it is equivalent to the inequality $S \ge (1-e^{-(\log a_i)/(a_i-1)})k - 1 + e^{-(\log a_i)/(a_i-1)}i$, which is a result of the above assumption on S when $i < 2\log_2 k$.

Now we have

$$
\begin{aligned}
&a_i \binom{k-(i-1)-a_i}{S-(i-1)} \Big/ A \\
&\quad \le a_i \binom{k-(i-1)-a_i}{S-(i-1)} \Big/ \binom{k-1}{S} \\
&\quad = a_i \prod_{j=0}^{i-2} \left(\frac{S-j}{k-1-j}\right) \prod_{j=1}^{a_i-1} \left(\frac{k-S-j}{k-(i-1)-j}\right) \\
&\quad \le a_i \frac{1}{2^{i-1}} \left(\frac{k-S-1}{k-(i-1)-1}\right)^{a_i-1} \le \frac{1}{2^{i-1}}
\end{aligned} \tag{6.33}
$$

using the fact that $a_i(\frac{k-S-1}{k-(i-1)-1})^{a_i-1} \le 1$.

On the other hand, if $S < k(\frac{\log a_i}{a_i-1}) + 2\log_2 k$, then $a_i S/(k-1) = O(1)\log_2 k$, which implies

$$
\begin{aligned}
&a_i \binom{k-(i-1)-a_i}{S-(i-1)} \Big/ A \\
&\quad \le \frac{a_i S}{k-1} \prod_{j=1}^{i-2} \left(\frac{S-j}{k-1-j}\right) \prod_{j=1}^{a_i-1} \left(\frac{k-S-j}{k-(i-1)-j}\right) \\
&\quad = O(1)\log_2 k/2^{i-2}.
\end{aligned}
$$

This proves that the right-hand side of (6.30) is bounded by $O(1)\log_2 k$.

To complete the proof of the lemma, that is, to prove $E(Q|W) \le C(1+|W|)$, we only need to show that $E(Q|S) \le C$ for some universal constant C when $|W| \le \log_2 k$, that is, when $k/2 - \sqrt{k/4}\log_2 k \le S \le k/2 + \sqrt{k/4}\log_2 k$. Following the argument in case 2 above, we only need to consider the summands where $i \le 2\log_2 k$ because the other part where $i > 2\log_2 k$ is bounded by 1 as proved in case 2.

When a_i, k are bigger than some universal constant, $k/2 - \sqrt{k/4}\log_2 k \ge \frac{\log a_i}{a_i-1} \times k + 2\log_2 k$, which implies $(\frac{k-S-1}{k-(i-1)-1})^{a_i-1} \times a_i \le 1$ and $\binom{k-(i-1)-a_i}{S-(i-1)} \times a_i/A \le 1/2^{i-1}$. Since both parts for $i \le 2\log_2 k$ and $i > 2\log_2 k$ are bounded by some constant, $E(Q|S) \le C$ when $|W| \le \log_2 k$, and hence the lemma is proved. □

6.3. *Proof of Propositions* 4.3 *and* 4.4. Let $\tilde{W}$ have the conditional distribution of W (W_1, W_2, resp.) given $|W| \le c_1\sqrt{n}$ ($|W_1|, |W_2| \le c_1\sqrt{n}$, resp.) where c_1 is to be determined. If we can prove that

$$
\frac{P(\tilde{W} \ge x)}{1-\Phi(x)} = 1 + O(1)\big(1+x^3\big)/\sqrt{n} \tag{6.34}
$$

for $0 \le x \le n^{1/6}$, then from the fact that [Ellis (1985)]

$$P(|W| > K\sqrt{n}) \le e^{-nC(K)} \tag{6.35}$$

and

$$P(|W_1| > K\sqrt{n}|S < 0) \le e^{-nC(K)}, \qquad P(|W_2| > K\sqrt{n}|S > 0) \le e^{-nC(K)}$$

for any positive number K where $C(K)$ is a positive constant depending only on K, we have, with $\delta_2 = O(1/\sqrt{n})$,

$$\begin{aligned}\frac{P(W \ge x)}{1-\Phi(x)} &\le \frac{P(\tilde{W} \ge x) + P(\delta_2|W| > 1/2)}{1-\Phi(x)} \\ &= 1 + O(1)(1+x^3)/\sqrt{n}\end{aligned}$$

for $0 \le x \le n^{1/6}$. Similarly, (4.15) and (4.16) are also true. Therefore, we prove Cramér moderate deviation for $\tilde{W}$ (still denoted by W in the following) defined below. Assume the state space of the spins is $\Sigma = (\sigma_1, \sigma_2, \dots, \sigma_n) \in \{-1, 1\}^n$ such that $\sum_{i=1}^n \sigma_i/n \in [a, b]$ where $[a, b]$ is any interval within which there is only one solution m to (4.9). Let $S = \sum_{i=1}^n \sigma_i$, $W = \frac{S-nm}{\sigma}$ and $\sigma^2 = n\frac{1-m^2}{1-(1-m^2)\beta}$. Note that in cases 1 and 2, $1-(1-m^2)\beta > 0$, thus σ^2 is well defined. Moreover, $[a, b]$ is chosen such that $|W| \le c_1\sqrt{n}$. The joint distribution of the spins is

$$Z_{\beta,h}^{-1}\exp\left(\frac{\beta\sum_{1\le i<j\le n}\sigma_i\sigma_j}{n} + \beta h\sum_{i=1}^n \sigma_i\right).$$

Let I be a random variable uniformly distributed over $\{1, \dots, n\}$ independent of $\{\sigma_i, 1 \le i \le n\}$. Let σ_i' be a random sample from the conditional distribution of σ_i given $\{\sigma_j, j \ne i, 1 \le j \le n\}$. Define $W' = W - (\sigma_I - \sigma_I')/\sigma$. Then (W, W') is an exchangeable pair. Let

$$A(w) = \frac{\exp(-\beta(m+h) - \beta\sigma w/n + \beta/n)}{\exp(-\beta(m+h) - \beta\sigma w/n + \beta/n) + \exp(\beta(m+h) + \beta\sigma w/n - \beta/n)}$$

and

$$B(w) = \frac{\exp(\beta(m+h) + \beta\sigma w/n + \beta/n)}{\exp(\beta(m+h) + \beta\sigma w/n + \beta/n) + \exp(-\beta(m+h) - \beta\sigma w/n - \beta/n)}.$$

It is easy to see that

$$\begin{aligned}&\frac{e^{-\beta(m+h)-\beta\sigma w/n}}{e^{-\beta(m+h)-\beta\sigma w/n} + e^{\beta(m+h)+\beta\sigma w/n}} \\ &\quad\le A(w) = \frac{\exp(-\beta(m+h) - \beta\sigma w/n)}{\exp(-\beta(m+h) - \beta\sigma w/n) + \exp(\beta(m+h) + \beta\sigma w/n - 2\beta/n)} \\ &\quad\le \frac{e^{-\beta(m+h)-\beta\sigma w/n}}{e^{-\beta(m+h)-\beta\sigma w/n} + e^{\beta(m+h)+\beta\sigma w/n}}e^{2\beta/n}\end{aligned}$$

and

$$\begin{aligned}
&\frac{e^{\beta(m+h)+\beta\sigma w/n}}{e^{\beta(m+h)+\beta\sigma w/n}+e^{-\beta(m+h)-\beta\sigma w/n}}\\
&\quad\le B(w)=\frac{\exp(\beta(m+h)+\beta\sigma w/n)}{\exp(\beta(m+h)+\beta\sigma w/n)+\exp(-\beta(m+h)-\beta\sigma w/n-2\beta/n)}\\
&\quad\le \frac{e^{\beta(m+h)+\beta\sigma w/n}}{e^{\beta(m+h)+\beta\sigma w/n}+e^{-\beta(m+h)-\beta\sigma w/n}}e^{2\beta/n}.
\end{aligned}$$

Therefore

$$A(W)+B(W)=1+O(1)\frac{1}{n}$$

and

$$A(W)-B(W)=-\tanh\bigl(\beta(m+h)+\beta\sigma W/n\bigr)+O(1)\frac{1}{n}.$$

Note that

$$\begin{aligned}
&E\bigl(W-W'|\Sigma\bigr)\\
&\quad=\frac{1}{\sigma}E(\sigma_I-\sigma_I|\Sigma)\\
&\quad=\frac{2}{\sigma}E\bigl(I\bigl(\sigma_I=1,\sigma_I'=-1\bigr)-I\bigl(\sigma_I=-1,\sigma_I'=1\bigr)|\Sigma\bigr)\\
&\quad=\frac{2}{\sigma}\frac{\sigma W+nm+n}{2n}A(W)I(S-2\ge an)\\
&\qquad-\frac{2}{\sigma}\frac{n-\sigma W-nm}{2n}B(W)I(S+2\le bn)\\
&\quad=\bigl(A(W)+B(W)\bigr)\left(\frac{W}{n}+\frac{m}{\sigma}\right)+\frac{1}{\sigma}\bigl(A(W)-B(W)\bigr)\\
&\qquad-\frac{\sigma W+nm+n}{\sigma n}A(W)I(S-2<an)\\
&\qquad+\frac{n-\sigma W-nm}{\sigma n}B(W)I(S+2>bn)\\
&\quad=\left(\frac{W}{n}+\frac{m}{\sigma}\right)\left(1+O\left(\frac{1}{n}\right)\right)-\frac{1}{\sigma}\left(\tanh\left(\beta(m+h)+\frac{\beta\sigma W}{n}\right)+O\left(\frac{1}{n}\right)\right)\\
&\qquad-\frac{S+n}{\sigma n}A(W)I(S-2<an)+\frac{n-S}{\sigma n}B(W)I(S+2>bn)\\
&\quad=\lambda(W-R),
\end{aligned}$$

where

$$\lambda = \frac{1-(1-m^2)\beta}{n} > 0$$

and

$$R = \frac{1}{\lambda}\frac{\tanh''(\beta(m+h)+\xi)\beta^2\sigma}{2n^2}W^2 + \frac{1}{\lambda}\frac{S+n}{\sigma n}A(W)I(S-2<an)$$
$$- \frac{1}{\lambda}\frac{n-S}{\sigma n}B(W)I(S+2>bn) + O(1)\left(\frac{W}{n}+\frac{1}{\sigma}\right),$$

where ξ is between 0 and $\beta\sigma W/n$. Similarly,

$$\begin{aligned}
&E\big((W-W')^2|\Sigma\big)\\
&\quad= \frac{4}{\sigma^2}E\big(I(\sigma_I=1,\sigma_I'=-1)+I(\sigma_I=-1,\sigma_I'=1)|\Sigma\big)\\
&\quad= \frac{2(1-m^2)}{\sigma^2} + O(1)\frac{W}{n\sigma} + O\left(\frac{1}{n\sigma^2}\right) + O\left(\frac{I(S-2<an \text{ or } S+2>bn)}{\sigma^2}\right).
\end{aligned}$$

Therefore, recall that $\sigma^2 = n\frac{1-m^2}{1-(1-m^2)\beta}$,

$$\big|E(D|W)-1\big| \le O\left(\frac{1}{\sqrt{n}}\right)(1+|W|).$$

For R, with $\delta_2 = O(1/\sqrt{n})$,

$$\big|E(R|W)\big| \le \delta_2(1+W^2),$$

and if c_1 is chosen such that $\delta_2|W| \le 1/2$, the second alternative of (3.4) is satisfied with $\alpha = 1/2$. Thus from Theorem 3.1, we have the following moderate deviation result for W:

$$\frac{P(W\ge x)}{1-\Phi(x)} = 1 + O(1)(1+x^3)\frac{1}{\sqrt{n}}$$

for $0 \le x \le n^{1/6}$. This completes the proof of (4.12) and (4.15).

6.4. *Proof of Proposition* 4.5. Since $(1-\Phi(x)) \ge \frac{1}{2(1+x)}e^{-x^2/2}$ for $x \ge 0$, (4.18) becomes trivial if $x\gamma \ge 1/8$. Thus we can assume

$$x\gamma \le 1/8. \tag{6.36}$$

Let $f = f_x$ be the Stein solution to equation (6.2). Let $W^{(i)} = W - \xi_i$ and $K_i(t) = E\xi_i(I\{0\le t\le \xi_i\} - I\{\xi_i \le t \le 0\})$. It is known that [see, e.g., (2.18) in Chen and Shao (2005)]

$$EWf(W) = \sum_{i=1}^n E\int_{-\infty}^{\infty} f'(W^{(i)}+t)K_i(t)\,dt.$$

Since $\int_{-\infty}^{\infty} K_i(t)\,dt = E\xi_i^2$, we have

$$
\begin{aligned}
&P(W \ge x) - \big(1 - \Phi(x)\big)\\
&\quad = EWf(W) - Ef'(W)\\
&\quad = \sum_{i=1}^{n} E\int_{-\infty}^{\infty}\big(f'(W^{(i)}+t) - f'(W)\big)K_i(t)\,dt\\
&\quad = \sum_{i=1}^{n} E\int_{-\infty}^{\infty}\big((W^{(i)}+t)f(W^{(i)}+t) - Wf(W)\big)K_i(t)\,dt\\
&\qquad + \sum_{i=1}^{n} E\int_{-\infty}^{\infty}\big(I\{W^{(i)}+t \ge x\} - I\{W \ge x\}\big)K_i(t)\,dt\\
&\quad := R_1 + R_2.
\end{aligned}
\tag{6.37}
$$

It suffices to show that

$$
|R_1| \le C(1+x^3)\gamma\big(1-\Phi(x)\big)e^{x^3\gamma}
\tag{6.38}
$$

and

$$
|R_2| \le C(1+x^2)\gamma\big(1-\Phi(x)\big)e^{x^3\gamma}.
\tag{6.39}
$$

To estimate R_1, let $g(w) = (wf(w))'$. It is easy to see that

$$
R_1 = \sum_{i=1}^{n} E\int\int_{\xi_i}^{t} g(W^{(i)}+s)\,ds\,K_i(t)\,dt.
\tag{6.40}
$$

By (6.19) and (6.20), following the proof of (6.21), we have

$$
\begin{aligned}
&Eg(W^{(i)}+s)\\
&\quad = Eg(W^{(i)}+s)I\{W^{(i)}+s \ge x\} + Eg(W^{(i)}+s)I\{W^{(i)}+s \le 0\}\\
&\qquad + Eg(W^{(i)}+s)I\{0 < W^{(i)}+s < x\}\\
&\quad \le \frac{2}{1+x^3}P(W^{(i)}+s \ge x) + 2\big(1-\Phi(x)\big)P(W^{(i)}+s \le 0)\\
&\qquad + \sqrt{2\pi}\big(1-\Phi(x)\big)\\
&\qquad \times E\big\{\big(1+(W^{(i)}+s)^2\big)e^{(W^{(i)}+s)^2/2}I\{0 < W^{(i)}+s < x\}\big\}\\
&\quad \le \frac{2}{1+x^3}P(W^{(i)} \ge x-s) + 2\big(1-\Phi(x)\big)P(W^{(i)}+s \le 0)\\
&\qquad - \sqrt{2\pi}\big(1-\Phi(x)\big)\int_0^x (1+y^2)e^{y^2/2}\,dP(W^{(i)}+s > y)\\
&\quad \le \frac{2}{1+x^3}P(W^{(i)} \ge x-s) + 2\big(1-\Phi(x)\big)P(W^{(i)}+s \le 0)
\end{aligned}
\tag{6.41}
$$

$$
\begin{aligned}
&+\sqrt{2\pi}\big(1-\Phi(x)\big)P\big(W^{(i)}+s>0\big)+\sqrt{2\pi}\big(1-\Phi(x)\big)J(s)\\
&\le \frac{2}{1+x^3}P\big(W^{(i)}\ge x-s\big)+\sqrt{2\pi}\big(1-\Phi(x)\big)+\sqrt{2\pi}\big(1-\Phi(x)\big)J(s),
\end{aligned}
$$

where

$$
J(s)=\int_0^x (3y+y^3)e^{y^2/2}P\big(W^{(i)}+s>y\big)\,dy. \tag{6.42}
$$

Clearly, for $0<t\le x$

$$
\begin{aligned}
Ee^{t\xi_j} &= 1+t^2E\xi_j^2/2+\sum_{k=3}^{\infty}\frac{(t\xi_j)^k}{k!}\\
&\le 1+t^2E\xi_j^2/2+\frac{t^3}{6}E|\xi_j|^3e^{t|\xi_j|}\\
&\le \exp\biggl(t^2E\xi_j^2/2+\frac{x^3}{6}E|\xi_j|^3e^{x|\xi_j|}\biggr)
\end{aligned}
$$

and hence

$$
Ee^{t(W^{(i)}+s)}\le \exp\biggl(t^2/2+x|s|+\frac{x^3}{6}\gamma\biggr)\qquad \text{for } 0\le t\le x. \tag{6.43}
$$

By (6.43), following the proof of Lemma 5.2 yields

$$
J(s)\le C\big(1+x^3\big)e^{x^3\gamma+x|s|}. \tag{6.44}
$$

Noting that (6.43) also implies that

$$
\begin{aligned}
P\big(W^{(i)}\ge x-s\big)&\le e^{-x^2}Ee^{x(W^{(i)}+s)}\le \exp\big(-x^2/2+x|s|+x^3\gamma\big)\\
&\le (1+x)\big(1-\Phi(x)\big)\exp\big(x|s|+x^3\gamma\big),
\end{aligned}
$$

we have

$$
Eg\big(W^{(i)}+s\big)\le C\big(1+x^3\big)\big(1-\Phi(x)\big)e^{x^3\gamma+x|s|}
$$

and therefore by (6.40),

$$
\begin{aligned}
|R_1| &\le \sum_{i=1}^n E\int_{-\infty}^{\infty}\biggl|\int_{\xi_i}^t g\big(W^{(i)}+s\big)\,ds\biggr|K_i(t)\,dt\\
&\le C\big(1+x^3\big)\big(1-\Phi(x)\big)e^{x^3\gamma}\sum_{i=1}^n E\int_{-\infty}^{\infty}\big(|t|e^{x|t|}+|\xi_i|e^{x|\xi_i|}\big)K_i(t)\,dt \qquad (6.45)\\
&\le C\big(1+x^3\big)\gamma\big(1-\Phi(x)\big)e^{x^3\gamma}.
\end{aligned}
$$

This proves (6.38).

As to R_2, we apply an exponential concentration inequality of Shao (2010) [see Theorem 2.7 in Shao (2010)]: for $a \geq 0$ and $b \geq 0$,

$$
\begin{aligned}
&P(x - a \leq W^{(i)} \leq x + b) \\
&\quad \leq Ce^{x\gamma + xa - x^2}\bigl((\gamma + b + a)E|W^{(i)}|e^{xW^{(i)}} + (Ee^{2xW^{(i)}})^{1/2}\exp(-\gamma^{-2}/32)\bigr) \\
&\quad \leq Ce^{x\gamma + xa - x^2}\bigl((\gamma + b + a)(EW^{(i)}e^{xW^{(i)}} + 1) \\
&\qquad\qquad + (Ee^{2xW^{(i)}})^{1/2}\exp(-\gamma^{-2}/32)\bigr) \\
&\quad \leq Ce^{x\gamma + xa - x^2}\bigl((\gamma + b + a)(1 + x)e^{x^2/2 + x^3\gamma} + e^{x^2 + x^3\gamma}\exp(-\gamma^{-2}/32)\bigr) \\
&\quad \leq Ce^{x^3\gamma + xa - x^2/2}\bigl((\gamma + b + a)(1 + x) + \exp(x^2/2 - \gamma^{-2}/32)\bigr) \\
&\quad \leq C(1 - \Phi(x))e^{x^3\gamma + xa}\bigl((\gamma + b + a)(1 + x^2) + \exp(x^2 - \gamma^{-2}/32)\bigr).
\end{aligned}
$$

Here we use the fact that $EW^{(i)}e^{xW^{(i)}} \leq xe^{x^2/2 + x^3\gamma}$, by following the proof of (6.43). Therefore

$$
\begin{aligned}
R_2 &\leq \sum_{i=1}^{n} E\int_{-\infty}^{\infty} P(x - \xi_i \leq W^{(i)} \leq x - t|\xi_i)K_i(t)\,dt \\
&\leq C(1 - \Phi(x))e^{x^3\gamma}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\{(1 + x^2)E(\gamma + |t| + |\xi_i|)e^{x|\xi_i|} \\
&\qquad\qquad + \exp(x^2 - \gamma^{-2}/32)\}K_i(t)\,dt \\
&\leq C(1 - \Phi(x))e^{x^3\gamma}\bigl((1 + x^2)\gamma + \exp(x^2 - \gamma^{-2}/32)\bigr) \\
&\leq C\gamma(1 + x^2)(1 - \Phi(x))e^{x^3\gamma}
\end{aligned}
$$

by (6.36). Similarly, the above bound holds for $-R_2$. This proves (6.39).

Acknowledgments. We would like to thank the Associate Editor and the referees for their helpful comments and suggestions which have significantly improved the presentation of this paper.

REFERENCES

BALDI, P., RINOTT, Y. and STEIN, C. (1989). A normal approximation for the number of local maxima of a random function on a graph. In *Probability, Statistics, and Mathematics* 59–81. Academic Press, Boston, MA. MR1031278

BARBOUR, A. D. (1988). Stein's method and Poisson process convergence. *J. Appl. Probab.* **25A** 175–184. MR0974580

BARBOUR, A. D. (1990). Stein's method for diffusion approximations. *Probab. Theory Related Fields* **84** 297–322. MR1035659

BARBOUR, A. D. and CHEN, L. H. Y. (2005). *An Introduction to Stein Method. Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **4**. Singapore Univ. Press and World Scientific, Singapore.

CHATTERJEE, S. (2007). Stein's method for concentration inequalities. *Probab. Theory Related Fields* **138** 305–321. MR2288072

CHATTERJEE, S. (2008). A new method of normal approximation. *Ann. Probab.* **36** 1584–1610. MR2435859

CHATTERJEE, S. and DEY, P. S. (2010). Applications of Stein's method for concentration inequalities. *Ann. Probab.* **38** 2443–2485. MR2683635

CHATTERJEE, S. and SHAO, Q.-M. (2011). Nonnormal approximation by Stein's method of exchangeable pairs with application to the Curie–Weiss model. *Ann. Appl. Probab.* **21** 464–483. MR2807964

CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545. MR0428387

CHEN, L. H. Y., HWANG, H. K. and ZACHAROVAS, V. (2011). Distribution of the sum-of-digits function of random integers: A survey. Unpublished manuscript.

CHEN, L. H. Y. and RÖLLIN, A. (2010). Stein couplings for normal approximation. Preprint. Available at http://arxiv.org/abs/1003.6039.

CHEN, L. H. Y. and SHAO, Q.-M. (2001). A non-uniform Berry–Esseen bound via Stein's method. *Probab. Theory Related Fields* **120** 236–254. MR1841329

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

CHEN, L. H. Y. and SHAO, Q.-M. (2005). Stein's method for normal approximation. In *An Introduction to Stein's Method* (A. D. Barbour and L. H. Y. Chen, eds.). *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **4** 1–59. Singapore Univ. Press and World Scientific, Singapore. MR2235448

CRAMÉR, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Scientifiques et Industrielles* **736** 5–23.

DEMBO, A. and RINOTT, Y. (1996). Some examples of normal approximations by Stein's method. In *Random Discrete Structures* (*Minneapolis*, *MN*, 1993). *The IMA Volumes in Mathematics and its Applications* **76** 25–44. Springer, New York. MR1395606

DIACONIS, P. (1977). The distribution of leading digits and uniform distribution mod 1. *Ann. Probab.* **5** 72–81. MR0422186

DIACONIS, P. and HOLMES, S. (2004). *Stein's Method*: *Expository Lectures and Applications*. *Institute of Mathematical Statistics Lecture Notes* **46**. IMS, Hayward, CA.

ELLIS, R. S. (1985). *Entropy*, *Large Deviations*, *and Statistical Mechanics*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **271**. Springer, New York. MR0793553

ELLIS, R. S. and NEWMAN, C. M. (1978a). The statistics of Curie–Weiss models. *J. Stat. Phys.* **19** 149–161. MR0503332

ELLIS, R. S. and NEWMAN, C. M. (1978b). Limit theorems for sums of dependent random variables occurring in statistical mechanics. *Z. Wahrsch. Verw. Gebiete* **44** 117–139. MR0503333

GOLDSTEIN, L. (2005). Berry–Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. *J. Appl. Probab.* **42** 661–683. MR2157512

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

LINNIK, J. V. (1961). On the probability of large deviations for the sums of independent variables. In *Proc.* 4*th Berkeley Sympos. Math. Statist. and Prob.*, *Vol. II* 289–306. Univ. California Press, Berkeley, CA. MR0137142

NOURDIN, I. and PECCATI, G. (2009). Stein's method on Wiener chaos. *Probab. Theory Related Fields* **145** 75–118. MR2520122

PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, New York. MR0388499

RAIČ, M. (2007). CLT-related large deviation bounds based on Stein's method. *Adv. in Appl. Probab.* **39** 731–752. MR2357379

RINOTT, Y. and ROTAR, V. (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann. Appl. Probab.* **7** 1080–1105. MR1484798

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. Hindustan Book Agency, New Delhi. MR2827974

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. Univ. California Press, Berkeley, CA. MR0402873

STEIN, C. (1986). *Approximate Computation of Expectations. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **7**. IMS, Hayward, CA. MR0882007

L. H. Y. CHEN
DEPARTMENT OF MATHEMATICS
NATIONAL UNIVERSITY OF SINGAPORE
10 LOWER KENT RIDGE ROAD
SINGAPORE 119076
REPUBLIC OF SINGAPORE
E-MAIL: matchyl@nus.edu.sg

X. FANG
DEPARTMENT OF STATISTICS
AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
6 SCIENCE DRIVE 2
SINGAPORE 117546
REPUBLIC OF SINGAPORE
E-MAIL: fangwd2003@gmail.com

Q.-M. SHAO
DEPARTMENT OF MATHEMATICS
HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY
CLEAR WATER BAY, KOWLOON, HONG KONG
CHINA
E-MAIL: maqmshao@ust.hk
AND
DEPARTMENT OF STATISTICS
CHINESE UNIVERSITY OF HONG KONG
SHATIN, N. T., HONG KONG
CHINA
(AFTER SEPTEMBER 1, 2012)