CLT-RELATED LARGE DEVIATION BOUNDS BASED ON STEIN'S METHOD

MARTIN RAIČ,* *University of Ljubljana*

Abstract

Large deviation estimates are derived for sums of random variables with certain dependence structures, including finite population statistics and random graphs. The argument is based on Stein's method, but with a novel modification of Stein's equation inspired by the Cramér transform.

Keywords: Large deviations; central limit theorem; random graphs; finite population statistics

2000 Mathematics Subject Classification: Primary 60F10
Secondary 60F05; 05C80; 62D05

1. Introduction

It has been well known for a long time that under suitable conditions, sums of independent, and also certain dependent random variables, tend to have an approximately normal distribution. One of the classical problems in probability is to bound the error in the normal approximation. A particularly interesting issue is the approximation of *large deviation probabilities*. More precisely, if W is a random variable with $\mathrm{E}\,W = 0$ and $\mathrm{var}(W) = 1$, the problem is to approximate tail probabilities $\mathrm{P}(W \geq x)$, where x can be substantially greater than 1. As the corresponding normal tail probabilities, $1 - \Phi(x)$, where, as usual,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-z^2/2}\,\mathrm{d}z,$$

are rather small, it is desirable to bound the *relative error*.

One of the crucial breakthroughs on this topic is due to Cramér [11], who succeeded in transforming the problem of approximating large deviation probabilities into the problem of approximating tail probabilities $\mathrm{P}(W \geq x)$ with x close to 0. This can be accomplished by multiplying the original probability measure by $\mathrm{e}^{\lambda W} / \mathrm{E}\,\mathrm{e}^{\lambda W}$ for a suitable λ (*Cramér's transformation*).

Cramér's result was extended and refined in many directions. The following result is a simplified version of Theorem 2 of [34].

Theorem 1.1. *Suppose that $W = (\xi_1 + \cdots + \xi_n)/\sqrt{n}$ is a sum of independent and identically distributed (i.i.d.) variables, with $\mathrm{E}\,\xi_1 = 0$ and $\mathrm{var}(\xi_1) = 1$, satisfying Cramér's condition $\mathrm{E}\,\mathrm{e}^{H\xi_1} < \infty$ for some $H > 0$. Then, for all $0 \leq x \leq C_0\sqrt{n}$, we have*

$$\frac{\mathrm{P}(W \geq x)}{1 - \Phi(x)} = \exp\left(\frac{x^3}{\sqrt{n}}\lambda\left(\frac{x}{\sqrt{n}}\right)\right)\left(1 + \frac{C_1\theta(1+x)}{\sqrt{n}}\right)$$

for some $\theta \in [-1, 1]$, *where* $\lambda(z)$ *is a power series in* z *with coefficients depending on the moments of* ξ_1 *and* C_0 *and* C_1 *are constants depending only on* H *and* $\mathrm{E}\,\mathrm{e}^{H\xi_1}$.

Remark. Of course, an analogous result to Theorem 1.1 holds for negative x.

The series $\lambda(z)$ is called the *Cramér series*. Its coefficients can be explicitly bounded in terms of H and $\mathrm{E}\,\mathrm{e}^{H\xi_1}$; see estimate (5) in [34]. Therefore, Theorem 1.1 yields the following weaker version which may be useful in the absence of information about higher moments.

Corollary 1.1. *For* ξ_i *and* W *as given in Theorem 1.1 and for* $0 \le x \le C_0\sqrt{n}$, *we have*

$$\exp\left(-\frac{C_2 x^3}{\sqrt{n}}\right)\left(1 - \frac{C_1(1+x)}{\sqrt{n}}\right) \le \frac{\mathrm{P}(W \ge x)}{1 - \Phi(x)} \le \exp\left(\frac{C_2 x^3}{\sqrt{n}}\right)\left(1 + \frac{C_1(1+x)}{\sqrt{n}}\right), \quad (1.1)$$

where, again, C_0, C_1, *and* C_2 *depend only on* H *and* $\mathrm{E}\,\mathrm{e}^{H\xi_1}$.

Theorem 1.1 has been extended in many ways to sums of dependent random variables. An important general approach is based on cumulant estimation. A survey with several applications, including mixing and (weighted) U-statistics, is given in [32].

The aim of this paper is to prove a version of Corollary 1.1 for certain dependence structures where *Stein's method* can be applied. This method was introduced in 1970 to derive uniform bounds for the error in the normal approximation; see [35]. The original approach was gradually extended in various directions, including approximations by other distributions, asymptotic expansions, and multivariate and functional settings. Apart from other settings, Stein's method can be applied to various sums of dependent random variables, particularly where no natural ordering of the summands is present. For a detailed survey with numerous applications see [4] and [5]. Many applications arise from random discrete structures such as permutations and graphs, both of which will also be considered in the present paper.

Stein's method works very well when approximating expectations of Lipschitz or smooth test functions. With additional effort and some loss of generality, we can also derive uniform (Berry–Esséen type) bounds of the correct order; see Section 4. Recently, Stein's method was refined to yield nonuniform bounds, where the absolute error for the tail probability $\mathrm{P}(W \ge x)$ is bounded by $\varepsilon/(1 + |x|^3)$; see [8]–[10]. Surprisingly, in the context of Stein's method, little effort has been put into deriving large deviation probabilities in the domain of normal approximation. A heuristic treatment is given in [36], with no explicit result given.

The main idea of Stein's method to approximate the distribution of a random variable W is first to show that for a certain (usually differential or difference) linear operator $\mathcal{A}$, the expectations $\mathrm{E}\,\mathcal{A} f(W)$ are small. The next step is then to solve the equation

$$\mathcal{A}f = h - c_h$$

for a suitable constant c_h depending on h; if $\mathrm{E}\,\mathcal{A} f(W)$ is small, then we have $\mathrm{E}\,h(W) \approx c_h$ (with small absolute error). For the normal approximation, we take $\mathcal{A}f(w) = f'(w) - f(w)w$ (it can be easily checked that in this case, $\mathrm{E}\,\mathcal{A} f(Z)$ vanishes if Z is a standard normal variate).

In the present paper we modify the second step in such a way that means c_h need no longer be a constant, but a function whose expectation is easy to derive or estimate. Inspired by Cramér's transformation, c_h is taken to be a multiple of $\mathrm{e}^{\lambda w}$, so that we can write

$$f'(w) - f(w)w = h(w) - N_\lambda h\mathrm{e}^{\lambda w}, \quad (1.2)$$

where $N_\lambda h$ is a constant factor which is chosen in such a way that the expectation of the right-hand side with respect to the standard normal distribution vanishes. As in the classical setting when $\lambda = 0$, this allows us to conclude that $\mathrm{E}\, h(W) \approx N_\lambda h\, \mathrm{E}\, \mathrm{e}^{\lambda W}$ with small *absolute* error. However, the right-hand side of (1.2) is flexible enough to yield an approximation with small *relative* error for a suitably chosen λ. Finally, it turns out that the relative error can also be controlled in the approximation of $\mathrm{E}\, \mathrm{e}^{\lambda W}$. This allows us to derive our main result, Theorem 2.2.

For ξ_i and W as given in Theorem 1.1 and with Cramér's condition strengthened to $\mathrm{E}\, \mathrm{e}^{H|\xi_1|} < \infty$, Theorem 2.2 reduces to

$$\exp\left(-\frac{C_2 x^3}{\sqrt{n}}\right)\left(1 - \frac{C_1(1 + x^2)}{\sqrt{n}}\right) \leq \frac{\mathrm{P}(W \geq x)}{1 - \Phi(x)} \leq \exp\left(\frac{C_2 x^3}{\sqrt{n}}\right)\left(1 + \frac{C_1(1 + x^2)}{\sqrt{n}}\right). \quad (1.3)$$

Apart from constants, the upper bound in (1.3) is of the same quality as the one in (1.1) and is applicable in the range $x = O(n^{1/2})$; beyond $x = O(n^{1/2})$, the upper bound on $\mathrm{P}(W \geq x)$ begins to increase in x, while the tail probability $\mathrm{P}(W \geq x)$ trivially vanishes if the summands ξ_i are appropriately bounded (remember that W is scaled so that $\mathrm{var}(W) = 1$). Conversely, the lower bound in (1.3) is comparable to the one in (1.1) only for $x = O(n^{1/4})$; beyond this range, it becomes negative and therefore useless. Improvements on the lower bound will be the subject of a forthcoming work.

Theorem 2.2 can be applied in various cases where dependence plays a part (see Section 3), although in all of them we assume *boundedness* of the summands X_i. More precisely, our construction requires certain 'independently bounded' components (see condition (iii) in Section 2), which, for sums of independent random variables, allow us to derive the Berry–Esséen theorem (with a weaker constant; see (2.7)). For dependent random variables, our result does not typically yield uniform bounds expressed in terms of third absolute moments. For the sake of simplicity, we assume boundedness, though our techniques may be applied to certain unbounded random variables.

2. Main results

In this section we shall state the precise formulation of our results, proved in Section 5. We shall focus on random variables which can be decomposed as in [6] (see below). In [6], this elegant and powerful approach was applied to random graphs, but can actually be used in many other cases in which Stein's method for normal approximation has been applied; see [26].

According to [6], let I be an index set and

$$W = \sum_{i \in I} X_i, \qquad \mathrm{E}\, X_i = 0, \mathrm{var}(W) = 1. \quad (2.1)$$

For every $i \in I$, suppose that the following *first-order decomposition* is given:

$$W = W_i + Z_i, \quad (2.2)$$

where W_i is independent of X_i. Next, suppose that

$$Z_i = \sum_{k \in K_i} Z_{ik} \quad (2.3)$$

and consider a σ-algebra $\mathcal{H}_{ik}$, such that X_i and Z_{ik} are $\mathcal{H}_{ik}$-measurable.

Remark. The index sets I and K_i can also be infinite, provided that

$$\sum_{i \in I} (\mathrm{E}\, X_i^2)^{1/2} < \infty, \qquad \sum_{i \in I} \sum_{k \in K_i} \mathrm{E}|X_i Z_{ik}| < \infty. \tag{2.4}$$

The estimation of the error in normal approximation is based on the distance between the distribution of W (which will be denoted by $\mathcal{L}(W)$) and the conditional distributions of W and W_i given $\mathcal{H}_{ik}$ (which will be denoted by $\mathcal{L}(W \mid \mathcal{H}_{ik})$ and $\mathcal{L}(W_i \mid \mathcal{H}_{ik})$; see (4.2) and the text below). This is the reason why Z_i should be written as a sum: Z_{ik} usually contain considerably less information than Z_i itself, so that $\mathcal{L}(W)$ is substantially closer to $\mathcal{L}(W \mid X_i, Z_{ik})$ than to $\mathcal{L}(W \mid X_i, Z_i)$.

To estimate the error between $\mathcal{L}(W)$ and $\mathcal{L}(W \mid \mathcal{H}_{ik})$ or $\mathcal{L}(W_i \mid \mathcal{H}_{ik})$, Barbour *et al.* [6] suggest *second-order decompositions*. In order to obtain their main result (i.e. an estimation of the error in normal approximation in the Wasserstein metrics), it suffices to construct random variables W_{ik} independent of $\mathcal{H}_{ik}$, and to bound the conditional expectations of $|W - W_{ik}|$ and $|W_i - W_{ik}|$ given $\mathcal{H}_{ik}$. In our case, this is not enough and we shall need a more sophisticated construction. We need to construct random variables W_{ik}, W_{ik}^*, U_{ik}, V_{ik}, and U_{ik}^* such that

(i) W_{ik}^* is independent of $\mathcal{H}_{ik}$ and has the same distribution as W,

(ii) U_{ik} and V_{ik} are $\mathcal{H}_{ik}$-measurable,

(iii) U_{ik}^* is conditionally independent of W_{ik} given $\mathcal{H}_{ik}$,

(iv) $|W - W_{ik}| \le U_{ik}$, $|W_i - W_{ik}| \le V_{ik}$, and $|W_{ik}^* - W_{ik}| \le U_{ik}^*$.

We shall need U_{ik}^* to bound the difference between $\mathcal{L}(W) = \mathcal{L}(W^* \mid \mathcal{H}_{ik})$ and $\mathcal{L}(W_{ik} \mid \mathcal{H}_{ik})$; the random variable U_{ik} will be used to estimate the difference between $\mathcal{L}(W_{ik} \mid \mathcal{H}_{ik})$ and $\mathcal{L}(W \mid \mathcal{H}_{ik})$ and the random variable V_{ik} will be used to estimate the difference between $\mathcal{L}(W_{ik} \mid \mathcal{H}_{ik})$ and $\mathcal{L}(W_i \mid \mathcal{H}_{ik})$.

Remark. If $|W_{ik}^* - W|$ and $|W_{ik}^* - W_i|$ can be bounded by sufficiently small $\mathcal{H}_{ik}$-measurable random variables, we can simplify the whole construction by setting $W_{ik} = W_{ik}^*$ and $U_{ik}^* = 0$. However, although in all applications given in Section 3, the summands X_i are assumed to be bounded, unboundedness can nevertheless appear in the first-order decompositions, but may be controlled by U_{ik}^*. This occurs in the case of random graphs; see Subsection 3.2.

The following theorem provides uniform bounds in the normal approximation.

Theorem 2.1. *For every $x \in \mathbb{R}$, we have*

$$|\mathrm{P}(W \le x) - \Phi(x)| \le \frac{6}{\sqrt{2\pi}} \sum_{i \in I} \sum_{k \in K_i} \mathrm{E}[|X_i Z_{ik}|(U_{ik} + V_{ik} + 16 U_{ik}^*)].$$

Now we turn to large deviation results. For every $\lambda \ge 0$, define

$$\beta(\lambda) := \sum_{i \in I} \sum_{k \in K_i} \mathrm{E}|X_i Z_{ik}|[\mathrm{e}^{\lambda U_{ik}} U_{ik} + \mathrm{e}^{\lambda V_{ik}} V_{ik}$$

$$+ \exp(2\lambda \mathrm{E}(U_{ik}^* \mid \mathcal{H}_{ik}))(6\mathrm{e}^{\lambda U_{ik}} + 6\mathrm{e}^{\lambda V_{ik}} + 4\mathrm{e}^{\lambda U_{ik}^*}) U_{ik}^*]. \tag{2.5}$$

Remark. If W is a sum of n random variables, which are, along with the random variables Z_{ik}, U_{ik}, V_{ik}, and U_{ik}^* uniformly bounded by B, and if the size of the set K_i does not grow with n, we have $\beta(\lambda) = O(nB^3 \mathrm{e}^{3\lambda B})$. In this case, the variance of W is typically of order nB^2. So if $\mathrm{var}(W) = 1$, we typically have $B = O(n^{-1/2})$ and also $\beta(\lambda) = O(n^{-1/2})$ for $\lambda = O(n^{1/2})$.

The following theorem is our main result.

Theorem 2.2. *For every $x \geq 0$ with $\beta(x) < \infty$, we have*

$$e^{-x^3\beta(x)/6}[1 - Q(x)\beta(x)] \leq \frac{P(W \geq x)}{1 - \Phi(x)} \leq e^{x^3\beta(x)/6}[1 + Q(x)\beta(x)] \qquad (2.6)$$

and

$$e^{-x^3\beta(x)/6}[1 - Q(x)\beta(x)] \leq \frac{P(W \leq -x)}{\Phi(-x)} \leq e^{x^3\beta(x)/6}[1 + Q(x)\beta(x)],$$

where

$$Q(x) = \frac{12}{\sqrt{2\pi}} + \frac{23}{2}x + \frac{11\sqrt{2\pi}}{2}x^2.$$

To see how all this works, consider first the easiest case, where the summands are i.i.d.

Example 2.1. Let $\xi_1, \xi_2, \ldots$ be i.i.d. random variables with $E\,\xi_i = 0$ and $\mathrm{var}(\xi_1) = 1$. Then we can write

$$W = \frac{\xi_1 + \cdots + \xi_n}{\sqrt{n}} = \sum_{i=1}^{n} X_i,$$

where $X_i = \xi_i/\sqrt{n}$. Set $I = \{1, 2, \ldots, n\}$, $K_i := \{0\}$, $Z_i := Z_{i0} := X_i$, $W_i := W_{i0} := W - X_i$, and define $\mathcal{H}_{i0}$ to be the σ-algebra generated by X_i. Furthermore, for each $i \in I$, let X_i^* be an independent copy of X_i. Now put $W_{i0}^* := W_{i0} + X_i^*$, $U_{i0} := |X_i| = |W - W_{i0}|$, $V_{i0} := 0 = |W_i - W_{i0}|$, and $U_{i0}^* := |X_i^*| = |W_{i0}^* - W_{i0}|$. We can easily check that these random variables satisfy conditions (ii)–(iv). Applying Theorem 2.1, after some calculation, we obtain

$$|P(W \leq x) - \Phi(x)| \leq \frac{6}{\sqrt{2\pi n}}[E|\xi_1|^3 + 16\,E\,\xi_1^2\,E|\xi_1|].$$

Using the well-known fact that, for any two increasing functions f and g, the random variables $f(Y)$ and $g(Y)$ are positively correlated for any Y such that $E(f(Y))^2$ and $E(g(Y))^2$ are finite, we find that

$$|P(W \leq x) - \Phi(x)| \leq \frac{41E|\xi_1^3|}{\sqrt{n}}, \qquad (2.7)$$

which is the Berry–Esséen theorem with a weaker constant.

Now we turn to large deviation estimates. Recalling (2.5), a straightforward calculation shows that

$$\beta(\lambda) \leq \frac{1}{\sqrt{n}}[E|\xi_1|^3 e^{\lambda|\xi_1|/\sqrt{n}} + 6e^{2\lambda E|\xi_1|/\sqrt{n}}\,E\,\xi_1^2 e^{\lambda|\xi_1|/\sqrt{n}}\,E|\xi_1|$$

$$+ 6e^{2\lambda E|\xi_1|/\sqrt{n}}\,E\,\xi_1^2 E|\xi_1| + 4e^{2\lambda E|\xi_1|/\sqrt{n}}\,E\,\xi_1^2 E|\xi_1|e^{\lambda|\xi_1|/\sqrt{n}}].$$

Clearly, $E|\xi_1| \leq (E\,\xi_1^2)^{1/2} = 1$. Again, using the fact that any two increasing functions of the same random variable are positively correlated, we obtain

$$\beta(\lambda) \leq \frac{17}{\sqrt{n}}e^{2\lambda/\sqrt{n}}\,E|\xi_1|^3 e^{\lambda|\xi_1|/\sqrt{n}},$$

which together with Theorem 2.2 implies (1.3).

In the next section we give two nontrivial examples where our approach can be used to advantage: finite population statistics and random graphs. Another important case where the decompositions of [6] (and Theorem 2.2) can be applied is the case of *local dependence*. However, the latter has already been worked out by Gorchakov [14], [15], using the cumulant estimation argument (more precisely, a large deviation result can be deduced from Theorem 2 of [15], and Lemma 1 of [31]).

3. Applications

3.1. Finite population statistics

Consider a statistic W based on a simple random sample of size n, drawn from a population of size $N \geq n$. Any such statistic can be expressed in terms of a uniformly distributed random injection $\pi \colon \mathbb{N}_n \to \mathbb{N}_N$, where $\mathbb{N}_n := \{1, \ldots, n\}$. Suppose that W is decomposed in the following way:

$$W = \sum_{\substack{\alpha \subseteq \mathbb{N}_n \\ 0 < |\alpha| \leq r}} h_\alpha(\pi(\alpha)),$$

where $0 < r \leq n$, $|\cdot|$ denotes cardinality and where for each $\alpha \subseteq \mathbb{N}_n$, h_α is a function defined on all subsets of $\mathbb{N}_N$ with the same cardinality as α.

Large deviations of finite population statistics have been considered in many special cases. The case in which $r = 1$ has been well elaborated. Robinson [30] proved a result similar to Theorem 1.1 with the error terms bounded in terms of supremum norm. Assuming that $\mathrm{E}\, W = 0$ and $\mathrm{var}(W) = 1$, Robinson's result typically covers tail probabilities $\mathrm{P}(W \geq x)$ in the zone $x = o(n^{1/2})$ (with a somewhat weaker result given for $x = O(n^{1/2})$, too). In the range $x = o(n^{1/6})$ (i.e. the range where we can generally expect $\mathrm{P}(W \geq x) \sim \Phi(-x)$), it was shown, for certain special cases, that the supremum norm can be replaced by more refined moment-type quantities; see Kallenberg [22]. Kallenberg's result has also been extended in other directions; see [17] and the references therein.

Another interesting case is the one where the functions h_α are the same for all α with $|\alpha| = r$ and $h_\alpha = 0$ for $|\alpha| < r$. In this case, we obtain U-statistics of finite populations. Large deviations of U-statistics were considered by [23]. Their result covered the case in which $r = 2$ and $x = o(n^{1/6})$.

Theorem 3.1. *Suppose that* $\mathrm{E}\, X_\alpha = 0$ *for all* α *and that* $\mathrm{var}(W) = 1$. *Furthermore, suppose that there are constants* $b_1, \ldots, b_r$ *such that*

$$\max_{\substack{K \subseteq \mathbb{N}_N \\ |K| = |\alpha|}} h_\alpha(K) - \min_{\substack{K \subseteq \mathbb{N}_N \\ |K| = |\alpha|}} h_\alpha(K) \leq b_{|\alpha|}$$

for all $\alpha \subseteq \mathbb{N}_n$ *with* $0 < |\alpha| \leq r$. *Then Theorem 2.2 holds with*

$$\beta(x) \leq nrb^3(8\mathrm{e}^{4rbx} + 10\mathrm{e}^{5rbx}), \quad \text{where } b := \sum_{s=1}^{r} \binom{n-1}{s-1} b_s. \tag{3.1}$$

Remark. For $r = 1$, Theorem 3.1 is more or less a special case of Robinson's [30] result, except that the constants here are completely explicit. For $r = 2$, our result covers a wider range ($x = O(n^{1/2})$) than the result of Kokic and Weber [23], which covers the range $x = o(n^{1/6})$ and also does not give any explicit constants. Conversely, the bounds given by Kokic and Weber are sharper in the sense that they are based on moments instead of the supremum norm. Finally, our result seems to be entirely new for $r > 2$.

To see the typical behavior of the bound in (3.1), we focus on U-statistics in more detail. Consider the following sequence of U-statistics:

$$W_k := \sigma_k^{-1} \sum_{\substack{\alpha \subseteq \mathbb{N}_{n(k)} \\ |\alpha|=r}} H_k(\pi_k(\alpha)),$$

where, again, for any $k \in \mathbb{N}$, $\pi_k \colon \mathbb{N}_{n(k)} \to \mathbb{N}_{N(k)}$ is a uniformly distributed random injection. We assume that

$$\sum_{K \subseteq \mathbb{N}_{N(k)}, |K|=r} H_k(K) = 0$$

(so that $\mathrm{E}\, W = 0$) and choose σ_k so that $\mathrm{var}(W_k) = 1$. In addition, we assume that

- $\lim_{k \to \infty} n(k) = \infty$ and $n(k) \le q N(k)$ for some $q < 1$;

- there exists a constant B such that $|H_k(K)| \le B$ for all k and $K \subseteq \mathbb{N}_{N(k)}$ with $|K| = r$;

- $\lim_{k \to \infty} \dfrac{l!\,[(r-l)!]^2 [N(k) - 2r + l]!}{[N(k)]!} \displaystyle\sum_{\substack{K,L \subseteq \mathbb{N}_{N(k)} \\ |K|=|L|=r \\ |K \cap L|=l}} H_k(K) H_k(L) = \Sigma_l$ for some $\Sigma_1, \dots,$

  Σ_r with $\Sigma_1 > 0$.

In this case, we have $\sigma_k^2 \ge C N(k)^{2r-1}$ for some $C > 0$, provided that k is sufficiently large; see equations (3.7) and (3.8) of [24]. Thus, after some calculation, the following assertion can be deduced from Theorem 3.1.

Corollary 3.1. *For W_k as above and sufficiently large k, Theorem 2.2 holds with*

$$\beta(x) \le \frac{C_1}{\sqrt{n(k)}} \mathrm{e}^{C_2 x / \sqrt{n(k)}}$$

for some C_1 and C_2 depending only on B, q, and $\Sigma_1, \dots, \Sigma_r$.

Now we turn to the proof of Theorem 3.1. The decompositions required by Theorem 3.1 will be based on the standard 'rearrangement' argument; see, e.g. Theorem 6 of [3]. Here we shall continue in the following way: first consider a family of random permutations τ_K of the set $\mathbb{N}_N$, indexed by all subsets $K \subseteq \mathbb{N}_N$, such that, for each K, the images of the elements of K are chosen randomly, with the uniform distribution over all variations of $|K|$ elements. Given $\tau_K|_K$, the elements of $\tau_K(K) \setminus K$ are then mapped into the elements of $K \setminus \tau_K(K)$, again chosen uniformly at random. The other elements of $\mathbb{N}_N$ remain fixed. Thus, τ_K alters at most $2|K|$ elements. The following assertion is straightforward and is therefore left without proof.

Lemma 3.1. *Let π be a uniformly distributed random injection $\mathbb{N}_n \to \mathbb{N}_N$. Generate a family of random permutations, $\{\tau_K \mid K \subseteq \mathbb{N}_N\}$, satisfying the conditions above and independent of π. Then, for any subset $\alpha \subseteq \mathbb{N}_n$, the random mapping $\tau_{\pi(\alpha)} \circ \pi$ is independent of the restriction $\pi|_\alpha$, has the same distribution as π, and coincides with π on $\mathbb{N}_n \setminus S(\alpha)$, where $S(\alpha) := \alpha \cup \pi^{-1}(\tau_{\pi(\alpha)}(\pi(\alpha)))$.*

Proof of Theorem 3.1. Take a family $T = \{\tau_K \mid K \subseteq \mathbb{N}_N\}$ as above, which is independent of π, and define

$$X_\alpha := h_\alpha(\pi(\alpha)), \qquad X_{\alpha\beta} := h_\beta(\pi_\alpha(\beta)), \qquad Z_{\alpha\beta} := X_\beta - X_{\alpha\beta},$$

$$\pi_\alpha := \tau_{\pi(\alpha)} \circ \pi, \qquad W_\alpha := \sum_{\beta \subseteq \mathbb{N}_n} X_{\alpha\beta}, \qquad Z_\alpha := \sum_{\beta \subseteq \mathbb{N}_n} Z_{\alpha\beta}.$$

By Lemma 3.1, π_α is independent of X_α; the same holds for W_α. Next, define $\mathcal{H}_{\alpha\beta}$ to be the σ-algebra generated by $\pi|_{\alpha\cup\beta}$ and T, and notice that X_α, X_β, and $X_{\alpha\beta}$ are $\mathcal{H}_{\alpha\beta}$-measurable. Thus, to complete the decompositions, generate a copy $T' = \{\tau'_K \mid K \subseteq \mathbb{N}_N\}$ of the family T, which is independent of the pair (π, T). Set

$$\pi_{\alpha\beta} := \tau'_{\pi(\alpha)} \circ \pi, \qquad X_{\alpha\beta\gamma} := h_\gamma(\pi_{\alpha\beta}(\gamma)), \qquad W_{\alpha\beta} := W^*_{\alpha\beta} := \sum_{\gamma \subseteq \mathbb{N}_n} X_{\alpha\beta\gamma}.$$

Again by Lemma 3.1, $W_{\alpha\beta}$ is independent of $\mathcal{H}_{\alpha\beta}$ and has the same distribution as W. Therefore, these decompositions satisfy the independence conditions given in Section 2. Moreover, we can estimate

$$|Z_{\alpha\beta}| \leq b_{|\beta|} \mathbf{1}[S(\alpha) \cap \beta \neq \varnothing],$$
$$|X_{\alpha\beta\gamma} - X_\gamma| \leq b_{|\gamma|} \mathbf{1}[S'(\alpha \cup \beta) \cap \gamma \neq \varnothing],$$
$$|X_{\alpha\beta\gamma} - X_{\alpha\gamma}| \leq b_{|\gamma|} \mathbf{1}[(S(\alpha) \cup S'(\alpha \cup \beta)) \cap \gamma \neq \varnothing],$$

where $S(\alpha)$ is as given in Lemma 3.1 and where $S'(\alpha)$ is equivalent to $S(\alpha)$ for the family $\{\tau'_K \mid K \subseteq \mathbb{N}_N\}$. Consequently,

$$|W_{\alpha\beta} - W| \leq 4rb =: U_{\alpha\beta}, \qquad |W_{\alpha\beta} - W_\alpha| \leq 5rb =: V_{\alpha\beta},$$

and, of course, we can set $U^*_{\alpha\beta} := 0$. Noting that

$$\sum_\beta b_{|\beta|} \mathbf{1}[S(\alpha) \cap \beta \neq \varnothing] \leq 2|\alpha|b, \qquad \sum_\alpha |\alpha|b_{|\alpha|} = \sum_{s=1}^r s\binom{n}{s} b_s = nb,$$

and collecting terms, the result follows.

3.2. Random graph degree statistics

Let $\Gamma = K(n, p)$ be a random graph on n vertices such that, for any two distinct vertices, the event that they are adjacent has probability p and all such events are independent. An important class of problems is to consider random variables based on the so called *semi-induced properties* of certain subsets of the vertex set, that is, determined by those edges with at least one endpoint in the subset. The most natural examples are 'being an isolated tree' or 'having a given degree'. We consider a number of subsets with the given property. Normal approximation of such statistics by Stein's method was studied by [2], [6], and [13]. Large deviations were studied by [19], [20], [21], and [25].

In the present paper we consider statistics based on degrees. Take a bounded measurable function $H\colon I \to \mathbb{R}$, where $I = \{1, \ldots, n\}$ denotes the vertex set of Γ, and define

$$S := \sum_{i \in I} H(\delta_i), \qquad W := \frac{S - \mu}{\sigma},$$

where δ_i denotes the degree of the vertex i with respect to Γ, $\mu = \mathrm{E}\, S$, and $\sigma^2 = \mathrm{var}(S)$.

Theorem 3.2. *If* $\max_{k \in I} H(k) - \min_{k \in I} H(k) \leq B$, *then Theorem 2.2 holds with*

$$\beta(x) \leq \frac{n B^3}{\sigma^3}(13 + 43np + 27(np)^2) \exp\left\{\frac{Bx(8 + 4np)}{\sigma} + 2np(e^{Bx/\sigma} - 1)\right\}. \qquad (3.2)$$

If $n \to \infty$ *and* $np \to c$, *where c is fixed and where* $H : \mathbb{Z}_+ \to \mathbb{R}$ *is also a fixed nonconstant bounded function then, for sufficiently large n, Theorem 2.2 holds with*

$$\beta(x) = \frac{C_1}{\sqrt{n}} \exp\left\{\frac{C_2 x}{\sqrt{n}} + C_3(e^{C_4 x/\sqrt{n}} - 1)\right\}, \qquad (3.3)$$

where $C_1, \ldots, C_4$ *depend only on H and c.*

Remark. For the number of *isolated vertices*, i.e. the case where $H(k) = \mathbf{1}(k = 0)$, O'Connell [25] derived a large deviation principle. Janson [20] and Janson *et al.* [21] considered statistics with locally dependent summands (which is not the case for degree statistics). Applied for $P(W \leq t\sqrt{n})$, where $n \to \infty$, t is fixed, and W is a suitable statistic on $K(n, p)$ scaled in such a way that $E\,W = 0$ and $\text{var}(W) = 1$, the results of [20] and [21] yield bounds as in Corollary 1.1; O'Connell's result yields something between the set-up of Corollary 1.1 and Theorem 1.1; however, none of these results provides explicit constants.

Proof of Theorem 3.2. To prove (3.2), first rewrite W in the following way:

$$W = \sum_{i \in I} X_i, \quad \text{where } X_i = h(\delta_i) \text{ and } h(k) = \frac{1}{\sigma}\left(H(k) - \frac{\mu}{n}\right).$$

Next, for $J \subseteq I$, denote by δ_k^J the degree of the vertex k with respect to the graph obtained from Γ by removing all edges with an endpoint in J. Since $\delta_k^{\{i\}}$ is independent of X_i, we can set

$$W_i := \sum_{k \in I} h(\delta_k^{\{i\}}) \quad \text{and} \quad Z_{ik} := h(\delta_k) - h(\delta_k^{\{i\}}).$$

To construct the second order decompositions, denote by $[j \sim l]$ the event that j is adjacent to l and let $\mathcal{H}_{ik}$ be the σ-algebra generated by all events $[j \sim l]$, where $j \in \{i, k\}$ and $l \in I \setminus \{j\}$. Furthermore, take an independent copy Γ^* of the graph Γ and denote by δ_k^{J*} the degree of the vertex k with respect to the graph with vertex set I and with edge set consisting of the edges in Γ^* with an endpoint in J and the edges in Γ with no endpoint in J. Define

$$W_{ik} := \sum_{l \in I} h(\delta_l^{\{i,k\}}) \quad \text{and} \quad W_{ik}^* := \sum_{l \in I} h(\delta_l^{\{i,k\}*}).$$

Now, since $\sup X_i - \inf X_i \leq b := B/\sigma$ and $E\,X_i = 0$, we have $|X_i| \leq b$. Clearly, $|Z_{ii}| \leq b$. Next, for $i \neq k$, observe that

$$Z_{ik} = \mathbf{1}[i \sim k](h(\delta_k) - h(\delta_k^{\{i\}}))$$

and thus $|Z_{ik}| \leq b\,\mathbf{1}[i \sim k]$. Furthermore,

$$|W - W_{ii}| = \left|\sum_{l \in I}(h(\delta_l) - h(\delta_l^{\{i\}}))\right| \leq b(\delta_i + 1) =: U_{ii},$$

$$|W_i - W_{ii}| = 0 =: V_{ii},$$

$$|W_{ii}^* - W_{ii}| = \left|\sum_{l \in I}(h(\delta_l^{\{i\}*}) - h(\delta_l^{\{i\}}))\right| \leq b(\delta_i^* + 1) =: U_{ii}^*,$$

where δ_i^* denotes the degree of the vertex i with respect to Γ^*. For $i \neq k$, we have

$$|W - W_{ik}| \leq b(\delta_i + \delta_k + 2) =: U_{ik},$$
$$|W_i - W_{ik}| \leq b(\delta_k + 1) =: V_{ik},$$
$$|W_{ik}^* - W_{ik}| \leq b(\delta_i^* + \delta_k^* + 2) =: U_{ik}^*.$$

Recalling (2.5) and using independence, we now find that

$$\beta(x) \leq b^3 \left[\sum_{i \in I} A_i(x) + \sum_{i \in I} \sum_{k \in I \setminus \{I\}} p_{ik} A_{ik}(x) \right],$$

where

$$A_i(x) := \mathrm{E}[\mathrm{e}^{bx(\delta_i + 1)}(\delta_i + 1)$$
$$+ \mathrm{e}^{2bx(\mathrm{E}\,\delta_i + 1)}(6\mathrm{e}^{bx(\delta_i + 1)}(\mathrm{E}\,\delta_i + 1) + 4\,\mathrm{E}\,\mathrm{e}^{bx(\delta_i + 1)}(\delta_i + 1))]$$

and

$$A_{ik}(x) := \mathrm{E}[\mathrm{e}^{bx(\delta_i + \delta_k + 2)}(\delta_i + \delta_k + 2) + \mathrm{e}^{2bx(\mathrm{E}\,\delta_i + \mathrm{E}\,\delta_k + 2)}$$
$$\times \{6(\mathrm{e}^{bx(\delta_i + \delta_k + 2)} + \mathrm{e}^{bx(\delta_k + 1)})(\mathrm{E}\,\delta_i + \mathrm{E}\,\delta_k + 2)$$
$$+ 4\,\mathrm{E}\,\mathrm{e}^{bx(\delta_i + \delta_k + 2)}(\delta_i + \delta_k + 2)\} \mid i \sim k].$$

Now observe that $A_i(x)$ and $A_{ik}(x)$ can be expressed entirely in terms of a few expectations of type $\mathrm{E}\,S$, $\mathrm{E}\,\mathrm{e}^{bxS}$, and $\mathrm{E}\,\mathrm{e}^{bxS}\,S$; the choices of S may be different, but any such S is a sum of independent Bernoulli random variables, such that $\mathrm{E}\,S \leq np$.

Now we use the fact that the Bernoulli distribution $\mathrm{Be}(p)$ precedes the Poisson distribution $\mathrm{Po}(p)$ in the convex order sense, i.e. for every convex function f, we have

$$\int f \, \mathrm{dBe}(p) \leq \int f \, \mathrm{dPo}(p).$$

To prove this, observe that $\mathrm{Be}(p)$ precedes $\mathrm{Be}(p/2) * \mathrm{Be}(p/2)$ and that our relation is invariant under convolution; then apply the Poisson law of small numbers. Applying the invariance under convolution again, it follows that S precedes $\mathrm{Po}(\mathrm{E}\,S)$. Thus, if f is convex and increasing, we have $\mathrm{E}\,f(S) \leq \int f \, \mathrm{dPo}(np)$. Hence,

$$\mathrm{E}\,\mathrm{e}^{bxS} \leq \sum_{k=0}^{\infty} \mathrm{e}^{kbx} \frac{(np)^k \mathrm{e}^{-np}}{k!} = \mathrm{e}^{np(\mathrm{e}^{bx} - 1)},$$

$$\mathrm{E}\,\mathrm{e}^{bxS}\,S \leq \sum_{k=0}^{\infty} k \mathrm{e}^{kbx} \frac{(np)^k \, \mathrm{e}^{-np}}{k!} = np\,\mathrm{e}^{bx + np(\mathrm{e}^{bx} - 1)}.$$

The estimate (3.2) is now completed by routine calculation.

To derive (3.3), it only remains to bound the variance from below. Using independence, we may write

$$\sigma^2 = \sum_{i \in I} \left[\mathrm{var}(H(\delta_i)) + \sum_{k \in I \setminus \{i\}} \mathrm{E}(H(\delta_i) - \mathrm{E}\,H(\delta_i))(H(\delta_k) - H(\delta_k^{(i)})) \right]$$
$$= n\,\mathrm{var}(H(N_{n-1})) + n^2 p(\mathrm{E}\,H(1 + N_{n-2}) - \mathrm{E}\,H(N_{n-1}))$$
$$\times (\mathrm{E}\,H(1 + N_{n-2}) - \mathrm{E}\,H(N_{n-2})),$$

where $N_k \sim \text{Bi}(k, p)$ ($\text{Bi}(\cdot, \cdot)$ denotes the binomial distribution). Noting that the point probabilities of N_{n-1} and N_{n-2} converge to the point probabilities of $\text{Po}(c)$, we conclude that for sufficiently large n, $\sigma_n^2 \geq Cn$ for some $C > 0$ and the proof is complete.

4. A bound on the Stein expectation

The aim of this section is to prove an auxiliary result which handles the dependence structure and the boundedness conditions from Section 2. This is one of the key steps in the proof of the main result and is independent of the rest of the proof, which is given in Section 5.

The main idea of Stein's method is to express the error in the normal approximation for a test function h, i. e. $\text{E}\,h(W) - \text{N}(0, 1)\{h\}$ in terms of the *Stein expectation*,

$$\text{E}[f'(W) - f(W)W], \tag{4.1}$$

where f solves the *Stein equation* $f'(w) - f(w)w = h(w) - \text{N}(0, 1)\{h\}$ and where $\text{N}(\mu, \sigma)\{h\}$ denotes the expectation of h with respect to the appropriate normal distribution. Taylor's expansion then typically allows us, roughly speaking, to express (4.1) in terms of expectations of the form $\text{E}\,f''(\tilde{W})$, multiplied by small quantities; see (4.2). If h is Lipschitz, we can then estimate the supremum norm of (a suitable version of) f'' in terms of the Lipschitz constant of h; thus, it suffices simply to bound $|\text{E}\,f''(\tilde{W})|$ in terms of the supremum norm of f''. However, when deriving Berry–Esséen type bounds, the indicators of half-lines are not even continuous, let alone Lipschitz. Of course, they can be approximated by Lipschitz functions, but it is then too crude to bound $|\text{E}\,f''(\tilde{W})|$ by the *supremum* norm of f'', because $|f''|$ is large only on a small set. In some cases, the probability that $\tilde{W}$ actually belongs to this set can be directly estimated. This is the so called *concentration inequality approach*; see [8]–[10] and [18].

Since $|f''|$ is large only on a small set, $|\text{E}\,f''(Z)|$ can be controlled if Z has a bounded density, in particular if Z is normal. Thus, a sufficiently sharp bound on $|\text{E}\,f''(\tilde{W})|$ can be derived if we succeed to derive a sharp bound for the error in the normal approximation of $\tilde{W}$. In order to do this, we may again apply Stein's method similarly as for W. Thus, a Berry–Esséen type bound for W can be derived by induction. This type of argument was introduced by Bolthausen [7], where it was applied to random permutations. It was also extended by Schneller [33] to asymptotic expansions of order 1 and by Götze [16] to the multivariate case.

Although the induction argument is very flexible, extending it to a more general setting seems to be a difficult job. However, instead of deriving a separate CLT for $\tilde{W}$, we may refer back to the original random variable W: to bound $|\text{E}\,f''(\tilde{W})|$, it actually suffices to show that the law of $\tilde{W}$ is close to the law of W. This 'bootstrapping' argument was introduced by Stein (see [36, Chapter IX]) to derive the classical Berry–Esséen theorem and subsequently extended to many cases of dependence; see [2], [12], [27], [28], and [29]. The 'bootstrapping' argument can be written in quite a general setting, but it usually requires *boundedness*. To some extent, it also allows for some unbounded components, but in this case with some additional assumptions on independence.

Lemma 4.1, below, bounds the Stein expectation in terms of certain expectations related to W in such a way that the bootstrapping argument can then be applied to bound large deviation probabilities; see Section 5.

Lemma 4.1. *Let W be a random variable decomposed as defined in (2.1)–(2.4), along with a construction satisfying conditions (i)–(iv) from Section 2, and let $h \colon \mathbb{R} \to \mathbb{R}$ be a differentiable function with absolutely continuous derivative. Then, for each $\lambda \geq 0$, the Stein expectation*

satisfies

$$\mathrm{E}[f'(W) - f(W)W] \leq \frac{\beta(\lambda)}{2} \sup_{t \in \mathbb{R}} \mathrm{E}|f''(W + t)| \min\{1, \mathrm{e}^{-\lambda t}\},$$

where $\beta(\lambda)$ is as given in (2.5).

Before proving Lemma 4.1, we need a couple of technical results. The first one deals with 'independent unboundedness'. The result is easy and therefore left without proof.

Lemma 4.2. *Let W_0, W^*, and U^* be random variables such that U^* is independent of W_0 and $|W^* - W_0| \leq U^*$. Then, for any constant u^* such that $\mathrm{P}(U^* \leq u^*) > 0$, any measurable function $f : \mathbb{R} \to [0, \infty)$, and any $a \leq b$, we have*

$$\int_a^b \mathrm{E}\, f(W_0 + t)\, \mathrm{d}t \leq \frac{1}{\mathrm{P}(U^* \leq u^*)} \int_{a-u^*}^{b+u^*} \mathrm{E}\, f(W^* + t)\, \mathrm{d}t.$$

The following assertion is the key step in proving Lemma 4.1.

Lemma 4.3. *Let $f : \mathbb{R} \to [0, \infty)$ be a measurable function and suppose that W^* is a random variable such that, for all $t \in \mathbb{R}$,*

$$\mathrm{E}\, f(W^* + t) \leq C \max\{1, \mathrm{e}^{\lambda t}\} \quad \text{for some } C, \lambda \geq 0.$$

In addition, suppose that there are random variables W_0, W', and U^* and constants u' and u^* such that

$$|W' - W_0| \leq u', \qquad |W^* - W_0| \leq U^*, \quad \text{and} \quad \mathrm{P}(U^* \leq u^*) > 0$$

and such that U^* is independent of W_0. Then the following inequality holds:

$$\left| \mathrm{E} \int_{W^*}^{W'} f(s)\, \mathrm{d}s \right| \leq C \left\{ \mathrm{e}^{\lambda u'} u' + \frac{\mathrm{e}^{\lambda u^*}}{\mathrm{P}(U^* \leq u^*)} [2\mathrm{e}^{\lambda u'} u^* + \mathrm{e}^{\lambda u'}\, \mathrm{E}\, U^* + \mathrm{E}\, \mathrm{e}^{\lambda U^*} U^*] \right\}.$$

Taking $u^* := 2\,\mathrm{E}\, U^*$ and making use of Markov's inequality, we obtain the following result.

Corollary 4.1. *With f, C, λ, W^*, W_0, W', U^*, and u' as given in Lemma 4.3, we have*

$$\left| \mathrm{E} \int_{W^*}^{W'} f(s)\, \mathrm{d}s \right| \leq C \left\{ \mathrm{e}^{\lambda u'} u' + \mathrm{e}^{2\lambda\, \mathrm{E}\, U^*}\, \mathrm{E}[6\mathrm{e}^{\lambda u'} U^* + \mathrm{e}^{\lambda u'}\, \mathrm{E}\, U^* + 2\,\mathrm{E}\, \mathrm{e}^{\lambda U^*} U^*] \right\}.$$

Proof of Lemma 4.3. Putting $J(a, b) := \int_a^b f(s)\, \mathrm{d}s$, we first estimate

$$
\begin{aligned}
\mathrm{E}\, J(W^*, W') &\leq \mathrm{E}\, J(W^*, W_0 + u') \\
&= \mathrm{E}\, J(W^*, W^* + u') + \mathrm{E}\, J(W^* + u', W_0 + u') \\
&\leq \mathrm{E}\, J(W^*, W^* + u') + \mathrm{E}\, J(W_0 - U^* + u', W_0 + u') \\
&= J_1 + J_2,
\end{aligned}
$$

where

$$J_1 = \int_0^{u'} \mathrm{E}\, f(W^* + t)\, \mathrm{d}t \quad \text{and} \quad J_2 = \mathrm{E} \int_{-U^*+u'}^{u'} f(W_0 + t)\, \mathrm{d}t.$$

Clearly, $J_1 \leq e^{\lambda u'} u'$. Using independence and the conditional version of Lemma 4.2 given U^*, we estimate

$$J_2 \leq \frac{1}{P(U^* \leq u^*)} E \int_{-U^*-u^*+u'}^{u^*+u'} E f(W^* + t) \, dt \leq C e^{\lambda(u^*+u')}(2u^* + E U^*).$$

Similarly,

$$
\begin{aligned}
E J(W', W^*) &\leq E J(W_0 - u', W^*) \\
&= E J(W_0 - u', W^* - u') + E J(W^* - u', W^*) \\
&\leq E J(W_0 - u', W_0 + u^* - u') + E J(W^* - u', W^*) \\
&= J_3 + J_4,
\end{aligned}
$$

where

$$J_3 = \int_{-u'}^{0} E f(W^* + t) \, dt \quad \text{and} \quad J_4 = E \int_{-u'}^{U^*-u'} f(W_0 + t) \, dt.$$

Now we have $J_3 \leq u'$ and again, we use independence and a suitable conditional version of Lemma 4.2 to estimate J_4:

$$J_4 \leq \frac{1}{P(U^* \leq u^*)} E \int_{-u^*-u'}^{U^*+u^*-u'} E f(W^* + t) \, dt \leq C(E e^{\lambda(U^*+u^*-u')} U^* + 2e^{\lambda(u^*-u')} u^*).$$

Noting that $|E \int_{W^*}^{W'} f(s) \, ds| \leq \max\{J_1 + J_2, J_3 + J_4\}$ and collecting all the estimates together, the desired result follows.

Proof of Lemma 4.1. Observe that

$$
\begin{aligned}
E[f'(W) - f(W)W] &= \sum_{i \in I} E[f'(W) E X_i W - f(W)X_i] \\
&= \sum_{i \in I} E[f'(W) E X_i W_i + f'(W) E X_i Z_i \\
&\qquad\qquad - f(W_i)X_i - f'(W_i + \theta Z_i)X_i Z_i],
\end{aligned}
$$

where θ is uniformly distributed over $[0, 1]$ and independent of all other variates. By independence, the first and the third term vanish; the others can be rewritten in the following way:

$$
\begin{aligned}
E[f'(W) - f(W)W] &= \sum_{i \in I} \sum_{k \in K_i} E[(f'(W) E X_i Z_{ik} - f'(W_i + \theta Z_i))X_i Z_{ik}] \\
&= \sum_{i \in I} \sum_{k \in K_i} E[(f'(W_{ik}^*) - f'(W_i + \theta Z_i))X_i Z_{ik}] \\
&= -\sum_{i \in I} \sum_{k \in K_i} E\left[\int_{W_{ik}^*}^{W_i + \theta Z_i} f''(s) \, ds \, X_i Z_{ik}\right]. \tag{4.2}
\end{aligned}
$$

The proof is now completed by making use of the conditional version of Corollary 4.1 given $\mathcal{H}_{ik}$ and θ, with f'' in place of f, W_{ik}^* in place of W^*, W_{ik} in place of W_0, $W_i + \theta Z_i = (1-\theta)W_i + \theta W$ in place of W', and by making use of Jensen's inequality for the convex functions $x \mapsto e^{\lambda x}$ and $x \mapsto x e^{\lambda x}$.

5. Proofs of main results

First, we turn to the study of the behavior of the solutions of the Stein equation

$$f'(w) - f(w)w = h(w) - \mathrm{N}(0, 1)\{h\},\tag{5.1}$$

where $\mathrm{N}(\mu, \sigma)\{h\}$ denotes the expectation of h with respect to the appropriate normal density. Some of the results stated here are classical and some of them seem to be new.

As we can easily check, one of the solutions to (5.1) can be expressed in the following form:

$$f(w) = \frac{1}{\sqrt{2\pi}}\psi(-w)\int_{-\infty}^{w} h(x)\,\mathrm{e}^{-x^2/2}\,\mathrm{d}x - \frac{1}{\sqrt{2\pi}}\psi(w)\int_{w}^{\infty} h(x)\,\mathrm{e}^{-x^2/2}\,\mathrm{d}x,\tag{5.2}$$

where ψ denotes the *Mills ratio*, given by

$$\psi(x) := \sqrt{2\pi}\,\mathrm{e}^{x^2/2}\Phi(x),$$

and

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-z^2/2}\,\mathrm{d}z.$$

Note that (5.2) is the only 'nice' solution to (5.1). More precisely, Lemmas 5.1 and 5.2 show that f or its derivatives are bounded if h is nice enough. This cannot be true for any other solution because the latter differs from f by a multiple of $\mathrm{e}^{w^2/2}$.

We can write $\psi(x) = \int_0^{\infty} \mathrm{e}^{tx-t^2/2}\,\mathrm{d}t$ and differentiation under the integral sign yields the important fact that the derivatives of ψ satisfy

$$\psi^{(r)}(x) = \int_0^{\infty} t^r \mathrm{e}^{tx-t^2/2}\,\mathrm{d}t > 0.\tag{5.3}$$

Lemma 5.1. *The function f given by (5.2) satisfies*

$$\|f\|_\infty \le \|h\|_1, \qquad \|f'\|_\infty \le \sup h - \inf h, \quad and \quad \|f''\|_\infty \le \sup h' - \inf h',\tag{5.4}$$

where $\|\cdot\|_\infty$ denotes the supremum norm and $\|\cdot\|_1$ denotes the L^1-norm.

Remark. We shall also need 'weak' derivatives in the Lebesgue sense: for an absolutely continuous function f, $\sup f'$ will be considered to be the essential supremum of f'; similarly, we shall take $\|f'\|_\infty$ to be the essential supremum of $|f'|$. Moreover, $V(f')$ will be considered to be the infimum of the total variations of all versions of f'; see Lemma 5.2, below.

Proof of Lemma 5.1. The second and the third estimate are modifications of the estimates stated in Lemma 3 of [36, p. 25]. The second estimate, which is also stated in Lemma 2.3 of [1], can be proved in much the same way as the corresponding estimate in Stein [36]. The third estimate actually follows from the corresponding Stein's estimate by noting that f'' does not change if $h(w)$ is replaced by $h(w) + kw$ for some $k \in \mathbb{R}$.

To derive the first estimate, we use the fact that ψ is increasing to estimate

$$|f(w)| \le \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{w} |h(x)|\psi(-x)\,\mathrm{e}^{-x^2/2}\,\mathrm{d}x + \frac{1}{\sqrt{2\pi}}\int_{w}^{\infty} |h(x)|\psi(x)\,\mathrm{e}^{-x^2/2}\,\mathrm{d}x.$$

Now, since

$$\frac{\psi(-x)\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}} = \Phi(-x) \leq 1 \quad \text{and} \quad \frac{\psi(x)\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}} = \Phi(x) \leq 1,$$

the desired result obviously follows.

When estimating tail probabilities, the following test functions are of special importance: for any $a < b$, define

$$h_{a,b}(w) = \frac{1}{b-a} \int_a^b \mathbf{1}(w \geq x)\,\mathrm{d}x = \begin{cases} 0, & w \leq a, \\ \dfrac{w-a}{b-a}, & a \leq w \leq b, \\ 1, & w \geq b. \end{cases}$$

The functions $h_{a,b}$ are standard smoothings of the indicators of half-lines, already used by [36]. Denote by $f_{a,b}$ the corresponding solution to (5.1) given by (5.2).

Lemma 5.2. *The functions $f_{a,b}$ satisfy*

$$\|f_{a,b}\|_\infty \leq \frac{\sqrt{2\pi}}{4}, \qquad \|f'_{a,b}\|_\infty \leq 1, \qquad V(f'_{a,b}) \leq 2, \quad \text{and} \quad V(f''_{a,b}) \leq \frac{4}{b-a},$$

where $V(f)$ denotes the total variation of f.

Proof. The first estimate is an immediate consequence of Lemma 3 of [36, p. 25] and the second one follows from the second estimate in (5.4). To derive the third estimate, first check that

$$f'_{a,b}(w) = \frac{1}{b-a} \int_a^b G_x(w)\,\mathrm{d}x, \quad \text{where } G_x(w) = \begin{cases} -\psi'(w)\Phi(-x), & w \leq x, \\ \psi'(-w)\Phi(x), & w > x. \end{cases}$$

Now use (5.3) to deduce that $V(G_x) = 2$ for all x. Similarly, we can check that

$$f''_{a,b}(w) = \frac{H_b(w) - H_a(w)}{b-a}, \quad \text{where } H_x(w) = \begin{cases} \psi''(w) \displaystyle\int_x^\infty \Phi(-t)\,\mathrm{d}t, & w \leq x, \\ -\psi''(-w) \displaystyle\int_{-\infty}^x \Phi(t)\,\mathrm{d}t, & w > x, \end{cases}$$

and again use (5.3) to find that $V(H_x) = 2$ for all x. This completes the proof.

The following assertion allows us to express the solutions of (1.2) in terms of the solutions to the 'classical' Stein equation. In other words, it allows us to express the solutions of the Stein equation for test functions of exponential growth with the solutions for bounded test functions. The result is immediate.

Lemma 5.3. *Suppose that $\tilde{f}$ solves*

$$\tilde{f}'(w) - \tilde{f}(w)w = h(w+\lambda) - \mathrm{N}(\lambda,1)\{h\}.$$

Then the function $f(w) := \mathrm{e}^{\lambda w}\tilde{f}(w-\lambda)$ solves

$$f'(w) - f(w)w = \mathrm{e}^{\lambda w}(h(w) - \mathrm{N}(\lambda,1)\{h\}).$$

Now we turn to the proof of Theorem 2.2. Here and hereafter, we assume that W is decomposed as in Section 2. In particular, recall the definition of $\beta(\lambda)$ from (2.5). Furthermore, denote by Z a standard normal random variable.

Lemma 5.4. *For all $x \in \mathbb{R}$, all $\lambda \geq 0$, and all h with bounded total variation, we have*

$$|\operatorname{E} \mathrm{e}^{\lambda W}[h(W) - \operatorname{E} h(Z + \lambda)]| \leq \left(\frac{6}{\sqrt{2\pi}} + \lambda + \frac{\sqrt{2\pi}}{4} \lambda^2 \right) \operatorname{E} \mathrm{e}^{\lambda W} V(h) \beta(\lambda). \qquad (5.5)$$

Remark. The term with Z in (5.5) is chosen so that the left-hand side vanishes if W is standard normal.

Proof of Lemma 5.4. Fix $\lambda \geq 0$. Clearly, we may assume that $\beta(\lambda)$ and $\operatorname{E} \mathrm{e}^{\lambda W}$ are finite. Set

$$\delta := \sup_{V(h)>0} \frac{1}{V(h)} |\operatorname{E} \mathrm{e}^{\lambda W}[h(W) - \operatorname{E} h(Z + \lambda)]|. \qquad (5.6)$$

Since Z has a density and since every function h with bounded total variation can be expressed in the form

$$h(w) = \int_{\mathbb{R}} \mathbf{1}(w \geq x) \mu(\mathrm{d}x) + \int_{\mathbb{R}} \mathbf{1}(w > x) \nu(\mathrm{d}x),$$

where $\|\mu\| + \|\nu\| = V(h)$, we can also write

$$\delta = \sup_{x \in \mathbb{R}} |\operatorname{E} \mathrm{e}^{\lambda W} \mathbf{1}(W \geq x) - \Phi(\lambda - x) \operatorname{E} \mathrm{e}^{\lambda W}|.$$

From the latter formula, it also follows that $\delta \leq \operatorname{E} \mathrm{e}^{\lambda W} < \infty$. Now let $\varepsilon > 0$. Noting that $\mathbf{1}(W \geq x) \leq h_{x-\varepsilon, x}(W)$ and

$$\begin{aligned}
\operatorname{E} h_{x-\varepsilon, x}(Z + \lambda) - \Phi(\lambda - x) &= \frac{1}{\varepsilon} \operatorname{E}[(Z + \lambda - x + \varepsilon) \mathbf{1}(x - \varepsilon \leq Z + \lambda \leq x)] \\
&\leq \frac{1}{\varepsilon \sqrt{2\pi}} \int_{x-\varepsilon}^{x} (z - x + \varepsilon) \, \mathrm{d}z \\
&= \frac{\varepsilon}{2\sqrt{2\pi}},
\end{aligned} \qquad (5.7)$$

we find that

$$\operatorname{E} \mathrm{e}^{\lambda W}[\mathbf{1}(W \geq x) - \Phi(\lambda - x)] \leq \operatorname{E} \mathrm{e}^{\lambda W} \left[h_{x-\varepsilon, x}(W) - \operatorname{E} h_{x-\varepsilon, x}(Z + \lambda) + \frac{\varepsilon}{2\sqrt{2\pi}} \right].$$

Similarly,

$$\operatorname{E} \mathrm{e}^{\lambda W}[\mathbf{1}(W \geq x) - \Phi(\lambda - x)] \geq \operatorname{E} \mathrm{e}^{\lambda W} \left[h_{x, x+\varepsilon}(W) - \operatorname{E} h_{x, x+\varepsilon}(Z + \lambda) - \frac{\varepsilon}{2\sqrt{2\pi}} \right].$$

Thus,

$$\delta \leq \sup_{x \in \mathbb{R}} |\operatorname{E} \mathrm{e}^{\lambda W}[h_{x, x+\varepsilon}(W) - \operatorname{E} h_{x, x+\varepsilon}(Z + \lambda)]| + \frac{\varepsilon}{2\sqrt{2\pi}} \operatorname{E} \mathrm{e}^{\lambda W}. \qquad (5.8)$$

Now fix $x \in \mathbb{R}$ and consider the function $f(w) := \mathrm{e}^{\lambda w} f_{x-\lambda, x-\lambda+\varepsilon}(w - \lambda)$. By Lemma 5.3, this function satisfies

$$f'(w) - f(w)w = \mathrm{e}^{\lambda w}[h_{x, x+\varepsilon}(w) - \operatorname{E} h_{x, x+\varepsilon}(Z + \lambda)]. \qquad (5.9)$$

Now we shall apply Lemma 4.1 to estimate the expectation of the left-hand side with W in place of w. By Lemma 4.1, we need to consider $E|f''(W + t)|$. Since

$$f''(w) = e^{\lambda w}[\lambda^2 f_{x-\lambda, x-\lambda+\varepsilon}(w - \lambda) + \lambda f'_{x-\lambda, x-\lambda+\varepsilon}(w - \lambda) + f''_{x-\lambda, x-\lambda+\varepsilon}(w - \lambda)],$$

we find that $E|f''(W + t)| \le e^{\lambda t} E e^{\lambda W}(R_0 + R_1 + R_2)$, where

$$R_0 = \lambda^2 |f_{x-\lambda, x-\lambda+\varepsilon}(W - \lambda)| \le \frac{\sqrt{2\pi}}{4}\lambda^2,$$
$$R_1 = \lambda |f'_{x-\lambda, x-\lambda+\varepsilon}(W - \lambda)| \le \lambda,$$
$$R_2 = |f''_{x-\lambda, x-\lambda+\varepsilon}(W - \lambda)|,$$

by Lemma 5.2. It remains to estimate $E e^{\lambda W} R_2$. First, observe that

$$E|f''_{x-\lambda, x-\lambda+\varepsilon}(Z)| \le \frac{1}{\sqrt{2\pi}}\|f''\|_1 = \frac{1}{\sqrt{2\pi}}V(f') \le \frac{2}{\sqrt{2\pi}}, \tag{5.10}$$

again, by Lemma 5.2. Recalling (5.6) and applying Lemma 5.2 once more, we find that

$$\begin{aligned} E e^{\lambda W}[|f''_{x-\lambda, x-\lambda+\varepsilon}(W - \lambda)| - |f''_{x-\lambda, x-\lambda+\varepsilon}(Z)|] &\le \delta V(|f''_{x-\lambda, x-\lambda+\varepsilon}|) \\ &= \delta V(f''_{x-\lambda, x-\lambda+\varepsilon}) \\ &\le \frac{4\delta}{\varepsilon}. \end{aligned} \tag{5.11}$$

From (5.10) and (5.11), we deduce that

$$E e^{\lambda W} R_2 \le \frac{2}{\sqrt{2\pi}} E e^{\lambda W} + \frac{4\delta}{\varepsilon}. \tag{5.12}$$

Putting everything together, we conclude that

$$E|f''(W + t)| \le \left(\frac{\sqrt{2\pi}}{4}\lambda^2 + \lambda + \frac{2}{\sqrt{2\pi}}\right) E e^{\lambda W} + \frac{4\delta}{\varepsilon}.$$

From Lemma 4.1 and the equality (5.9), we now find that

$$|E e^{\lambda W}[h_{x,x+\varepsilon}(W) - E h_{x,x+\varepsilon}(Z + \lambda)]| \le \left[\left(\frac{\sqrt{2\pi}}{8}\lambda^2 + \frac{\lambda}{2} + \frac{1}{\sqrt{2\pi}}\right) E e^{\lambda W} + \frac{2\delta}{\varepsilon}\right]\beta(\lambda).$$

Taking the supremum over $x \in \mathbb{R}$ and making use of (5.8), it follows that

$$\delta \le \left[\left(\frac{\sqrt{2\pi}}{8}\lambda^2 + \frac{\lambda}{2} + \frac{1}{\sqrt{2\pi}}\right) E e^{\lambda W} + \frac{2\delta}{\varepsilon}\right]\beta(\lambda) + \frac{\varepsilon}{2\sqrt{2\pi}} E e^{\lambda W}.$$

Choosing $\varepsilon := 4\beta(\lambda)$, we obtain

$$\delta \le \left(\frac{\sqrt{2\pi}}{8}\lambda^2 + \frac{\lambda}{2} + \frac{3}{\sqrt{2\pi}}\right) E e^{\lambda W} \beta(\lambda) + \frac{\delta}{2},$$

which, together with the observation that δ is finite, yields the desired result.

Lemma 5.5. *For all $x \in \mathbb{R}$ and all $\lambda \geq 0$, we have*

$$|P(W \geq x) - e^{-\lambda^2/2} \, E \, e^{\lambda W} \Phi(-x)| \leq e^{-\lambda x} \, E \, e^{\lambda W} \left(\frac{6}{\sqrt{2\pi}} + \frac{11}{4}\lambda \right) \beta(\lambda). \qquad (5.13)$$

Proof. Fix $x \in \mathbb{R}$, $\lambda \geq 0$, and $\varepsilon > 0$. First, write

$$e^{\lambda(w-x-\varepsilon)} h_{x,x+\varepsilon}(w) \leq \mathbf{1}(w \geq x) \leq e^{\lambda(w-x)} h_{x-\varepsilon,x}(w), \qquad (5.14)$$

where, for $a < b$,

$$h_{a,b,\lambda}(w) = \begin{cases} 0, & w \leq a, \\ \dfrac{w-a}{b-a}, & a \leq w \leq b, \\ e^{-\lambda(w-b)}, & w \geq b. \end{cases}$$

Next, observe that, similarly as in (5.7),

$$E \, h_{x-\varepsilon,x,\lambda}(Z + \lambda) - e^{\lambda x - \lambda^2/2} \Phi(-x) = E[h_{x-\varepsilon,x,\lambda}(Z + \lambda) - e^{-\lambda(Z+\lambda-x)} \mathbf{1}(Z + \lambda \geq x)]$$

$$\leq \frac{\varepsilon}{2\sqrt{2\pi}}. \qquad (5.15)$$

Moreover,

$$e^{\lambda x - \lambda^2/2} \Phi(-x) - e^{-\lambda\varepsilon} \, E \, h_{x,x+\varepsilon,\lambda}(Z + \lambda)$$

$$= E\left\{ \left[e^{-\lambda(Z+\lambda-x)} - \frac{e^{-\lambda\varepsilon}}{\varepsilon}(Z + \lambda - x) \right] \mathbf{1}(x \leq Z + \lambda \leq x + \varepsilon) \right\}$$

$$\leq \frac{1}{\sqrt{2\pi}} \int_x^{x+\varepsilon} \left[e^{-\lambda(z-x)} - \frac{e^{-\lambda\varepsilon}}{\varepsilon}(z - x) \right] dz$$

$$\leq \frac{1}{\sqrt{2\pi}} \int_x^{x+\varepsilon} \left[1 - \frac{1 - e^{-\lambda\varepsilon}}{\varepsilon}(z - x) - \frac{e^{-\lambda\varepsilon}}{\varepsilon}(z - x) \right] dz$$

$$= \frac{\varepsilon}{2\sqrt{2\pi}} \qquad (5.16)$$

(the last inequality was obtained by convexity of the exponential function). Combining (5.14), (5.15), and (5.16), we find that the quantity

$$\delta^* := \sup_x e^{\lambda x} |P(W \geq x) - e^{-\lambda^2/2} \, E \, e^{\lambda W} \Phi(-x)|$$

satisfies

$$\delta^* \leq \sup_x e^{\lambda x} | E \, e^{\lambda W} (h_{x,x+\varepsilon,\lambda}(W) - E \, h_{x,x+\varepsilon,\lambda}(Z + \lambda))| + \frac{\varepsilon}{2\sqrt{2\pi}} \, E \, e^{\lambda W}. \qquad (5.17)$$

Now consider the function $f(w) := e^{\lambda w} f_{x-\lambda,x-\lambda+\varepsilon}(w - \lambda)$, where $f_{x-\lambda,x-\lambda+\varepsilon}$ is the solution of the Stein equation given in (5.2), with $h_{x-\lambda,x-\lambda+\varepsilon}$ in place of h. By Lemma 5.3, the function f satisfies

$$f'(w) - f(w)w = e^{\lambda w}[h_{x,x+\varepsilon,\lambda}(w) - E \, h_{x,x+\varepsilon,\lambda}(Z + \lambda)].$$

Differentiating f, we find that $\mathrm{E}|f''(W+t)| \leq e^{\lambda t} \mathrm{E}\, e^{\lambda W}(R_0 + R_1 + R_2 + R_2^*)$, where

$$R_{0,\lambda} = \lambda^2 |f_{x-\lambda, x-\lambda+\varepsilon, \lambda}(W-\lambda)| \leq \frac{\varepsilon \lambda^2}{2} + \lambda,$$

$$R_{1,\lambda} = \lambda |f'_{x-\lambda, x-\lambda+\varepsilon, \lambda}(W-\lambda)| \leq \lambda,$$

$$R_2 = |f''_{x-\lambda, x-\lambda+\varepsilon}(W-\lambda)|,$$

$$R_{2,\lambda}^* = |f''_{x-\lambda, x-\lambda+\varepsilon}(W-\lambda) - f''_{x-\lambda, x-\lambda+\varepsilon, \lambda}(W-\lambda)| \leq \lambda,$$

by Lemma 5.1 (notice that the function $h_{x-\lambda, x-\lambda+\varepsilon} - h_{x-\lambda, x-\lambda+\varepsilon, \lambda}$ is increasing with the Lipschitz constant λ). Putting everything together and applying Lemma 4.1 and (5.17), we conclude that

$$\delta^* \leq \mathrm{E}\, e^{\lambda W}\left[\left(\frac{\varepsilon \lambda^2}{4} + \frac{3}{2}\lambda + \frac{1}{2}R_2\right)\beta(\lambda) + \frac{\varepsilon}{2\sqrt{2\pi}}\right]. \tag{5.18}$$

Now choose

$$\varepsilon := \frac{8 + 2\sqrt{2\pi}\lambda}{2 + \sqrt{2\pi}\lambda^2 \beta(\lambda)} \beta(\lambda),$$

which reduces (5.18) to

$$\delta^* \leq \mathrm{E}\, e^{\lambda W}\left(\frac{2}{\sqrt{2\pi}} + 2\lambda + \frac{1}{2}R_2\right)\beta(\lambda).$$

The term with R_2 will be estimated in two ways. Trivially, from the last estimate in (5.4), it follows that $R_2 \leq 1/\varepsilon$. Conversely, we can use (5.12) and then estimate δ (which is defined in (5.6)) by Lemma 5.4. Combining both estimates, we obtain

$$\mathrm{E}\, e^{\lambda W} R_2 \leq \mathrm{E}\, e^{\lambda W}\left[\frac{2}{\sqrt{2\pi}} + \frac{1}{\varepsilon}\min\{1, H(\lambda)\}\right],$$

where $H(\lambda) := (24/\sqrt{2\pi} + 4\lambda + \sqrt{2\pi}\lambda^2)\beta(\lambda)$. Estimating

$$\frac{1}{\varepsilon}\min\{1, H(\lambda)\} \leq \frac{2H(\lambda) + \sqrt{2\pi}\lambda^2 \beta(\lambda)}{(8 + 2\sqrt{2\pi}\lambda)\beta(\lambda)},$$

we find, after some calculation, that

$$\mathrm{E}\, e^{\lambda W} R_2 \leq \mathrm{E}\, e^{\lambda W}\left[\frac{2}{\sqrt{2\pi}} + \frac{48/\sqrt{2\pi} + 8\lambda + 3\sqrt{2\pi}\lambda^2}{8 + 2\sqrt{2\pi}\lambda}\right]$$

$$\leq \mathrm{E}\, e^{\lambda W}\left[\frac{8}{\sqrt{2\pi}} + \frac{3}{2}\lambda\right].$$

Substituting this into (5.18), the result follows.

Lemma 5.6. *For every* $\lambda \in \mathbb{R}$, *we have*

$$\exp\left(\frac{\lambda^2}{2} - \frac{\beta(|\lambda|)}{6}\right) \leq \mathrm{E}\, e^{\lambda W} \leq \exp\left(\frac{\lambda^2}{2} + \frac{\beta(|\lambda|)}{6}\right).$$

Proof. Noting that $-W$ can be decomposed in the same way as W (with the same function $\beta(\lambda)$), we find that it suffices to prove the result for $\lambda \geq 0$. Define

$$F(\lambda) = \ln \mathrm{E}\, \mathrm{e}^{\lambda W}$$

and note that

$$F'(\lambda) = \frac{\mathrm{E}\, \mathrm{e}^{\lambda W}\, W}{\mathrm{E}\, \mathrm{e}^{\lambda W}}$$

so that

$$F'(\lambda) - \lambda = -\frac{1}{\mathrm{E}\, \mathrm{e}^{\lambda W}}\, \mathrm{E}[f_\lambda'(W) - f_\lambda(W)W],$$

where $f_\lambda(x) := \mathrm{e}^{\lambda x}$. Again, we shall apply Lemma 4.1. Noting that $\mathrm{E}\, f_\lambda''(W + t) = \lambda^2 \mathrm{e}^{\lambda t}\, \mathrm{E}\, \mathrm{e}^{\lambda W}$, it follows that

$$|F'(\lambda) - \lambda| \leq \frac{\lambda^2}{2}\beta(\lambda). \tag{5.19}$$

Now write

$$F(\lambda) = \int_0^\lambda F'(t)\, \mathrm{d}t = \frac{\lambda^2}{2} + \int_0^\lambda (F'(t) - t)\, \mathrm{d}t,$$

which together with (5.19) and the fact that β is nondecreasing yields $|F(\lambda) - \frac{1}{2}\lambda^2| \leq \beta(\lambda)/6$. This completes the proof.

Proofs of Theorems 2.1 and 2.2. Theorem 2.1 is an immediate consequence of Lemma 5.4 for $\lambda = 0$. To prove Theorem 2.2, first observe that, similarly to the proof of Lemma 5.6, it suffices to prove (2.6). We shall use the fact that, for all $x \geq 0$,

$$\Phi(-x) \geq \frac{\mathrm{e}^{-x^2/2}}{2 + \sqrt{2\pi}x} \tag{5.20}$$

(first, note that we have an equality for $x = 0$ and $x \to \infty$; then examine the derivatives of both sides). Applying inequality (5.20) to the right-hand side of (5.13) and dividing by $\mathrm{e}^{-\lambda^2/2}\, \mathrm{E}\, \mathrm{e}^{\lambda W}\Phi(-x)$, we find that

$$\left| \frac{\mathrm{P}(W \geq x)}{\mathrm{e}^{-\lambda^2/2}\, \mathrm{E}\, \mathrm{e}^{\lambda W}\Phi(-x)} - 1 \right| \leq \left(\frac{6}{\sqrt{2\pi}} + \frac{11}{4}\lambda \right)(2 + \sqrt{2\pi}x)\mathrm{e}^{(\lambda-x)^2/2}\beta(\lambda).$$

The proof is now completed by choosing $\lambda = x$ and applying Lemma 5.6.

References

[1] 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*, Academic Press, Boston, MA, pp. 59–81.

[2] BARBOUR, A. D. (1982). Poisson convergence and random graphs. *Math. Proc. Camb. Philos. Soc.* **92,** 349–359.

[3] BARBOUR, A. D. (1990). Stein's method for diffusion approximations. *Prob. Theory Relat. Fields* **84,** 297–322.

[4] BARBOUR, A. D. AND CHEN, L. H. Y. (eds) (2005). *An Introduction to Stein's Method*. World Scientific, Singapore.

[5] BARBOUR, A. D. AND CHEN, L. H. Y. (2005). *Stein's Method and Applications*. World Scientific, Singapore.

[6] 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.

[7] BOLTHAUSEN, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrscheinlichkeitsth.* **66,** 379–386.

[8] CHEN, L. H. Y. AND SHAO, Q.-M. (2001). A non-uniform Berry–Esseen bound via Stein's method. *Prob. Theory Relat. Fields* **120,** 236–254.

[9] CHEN, L. H. Y. AND SHAO, Q.-M. (2007). Normal approximation for nonlinear statistics using a concentration inequality approach. *Bernoulli* **13,** 581–599.

[10] CHEN, L. H. Y. AND SHAO, Q.-M. (2004). Normal approximation under local dependence. *Ann. Prob.* **32,** 1985–2028.

[11] CRAMÉR, H. (1938). Sur un nouveau théorème limite de la théorie des probabilités. In *Colloque Consacre à la Théorie des Probabilités* **736**, Hermann, Paris, pp. 5–23.

[12] DEMBO, A. AND RINOTT, Y. (1996). Some examples of normal approximations by Stein's method. In *Random Discrete Structures* (IMA Vol. Math. Appl. **76**). Springer, New York, pp. 25–44.

[13] GOLDSTEIN, L. AND RINOTT, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Prob.* **33,** 1–17.

[14] GORCHAKOV, A. B. (1995). Upper bounds for cumulants of the sum of multi-indexed random variables. *Discrete Math Appl.* **5,** 317–331.

[15] GORCHAKOV, A. B. (1997). Explicit bounds for probabilities of large deviations of sums of random vectors with a given graph of dependencies. In *Probabilistic Methods in Discrete Mathematics* (Petrozavodsk, 1996), VSP, Utrecht, pp. 219–230.

[16] GÖTZE, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Prob.* **19,** 724–739.

[17] HALLIN, M. AND PURI, M. L. (1992). Some asymptotic results for a broad class of nonparametric statistics. *J. Statist. Planning Infer.* **32,** 165–196.

[18] HO, S.-T. AND CHEN, L. H. Y. (1978). An L_p bound for the remainder in a combinatorial central limit theorem. *Ann. Prob.* **6,** 231–249.

[19] JANSON, S. (1990). Poisson approximation for large deviations. *Random Structures Algorithms* **1,** 221–229.

[20] JANSON, S. (2004). Large deviations for sums of partly dependent random variables. *Random Structures Algorithms* **24,** 234–248.

[21] JANSON, S., OLESZKIEWICZ, K. AND RUCIŃSKI, A. (2004). Upper tails for subgraph counts in random graphs. *Israel J. Math.* **142,** 61–92.

[22] KALLENBERG, W. C. M. (1982). Cramér type large deviations for simple linear rank statistics. *Z. Wahrscheinlichkeitsth.* **60,** 403–409.

[23] KOKIC, P. N. AND WEBER, N. C. (1995). Large deviation probabilities for U-statistics based on samples from finite populations. In *Exploring Stochastic Laws*, VSP, Utrecht, pp. 175–181.

[24] NANDI, H. K. AND SEN, P. K. (1963). On the properties of U-statistics when the observations are not independent. II. Unbiased estimation of the parameters of a finite population. *Calcutta Statist. Assoc. Bull.* **12,** 124–148.

[25] O'CONNELL, N. (1998). Some large deviation results for sparse random graphs. *Prob. Theory Relat. Fields* **110,** 277–285.

[26] RAIČ, M. (2004). A multivariate CLT for decomposable random vectors with finite second moments. *J. Theoret. Prob.* **17,** 573–603.

[27] RINOTT, Y. (1994). On normal approximation rates for certain sums of dependent random variables. *J. Comput. Appl. Math.* **55,** 135–143.

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

[29] 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. Prob.* **7,** 1080–1105.

[30] ROBINSON, J. (1977). Large deviation probabilities for samples from a finite population. *Ann. Prob.* **5,** 913–925.

[31] RUDZKIS, R., SAULIS, L. AND STATULEVIČIUS, V. (1978). A general lemma on probabilities of large deviations. *Lithuanian Math. J.* **18,** 226–238.

[32] SAULIS, L. AND STATULEVIČIUS, V. (1991). *Limit Theorems for Large Deviations* (*Math. Appl.* (Soviet Ser.) **73**). Kluwer, Dordrecht.

[33] SCHNELLER, W. (1989). Edgeworth expansions for linear rank statistics. *Ann. Statist*. **17,** 1103–1123.

[34] STATULEVIČIUS, V. A. (1966). On large deviations. *Z. Wahrscheinlichkeitsth*. **6,** 133–144.

[35] STEIN, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* (University of California, Berkeley, 1970/1971), Vol. II, *Probability Theory*, University of California Press, Berkeley, pp. 583–602.

[36] STEIN, C. (1986). *Approximate Computation of Expectations* (Instit. Math. Statist. Lecture Notes Monogr. Ser. **7**). Institute of Mathematical Statistics, Hayward, CA.