

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. Our results cover local dependence (including U -statistics and Nash equilibria), 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; Local dependence; Finite population statistics

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

1 Introduction

In 1970, Stein [38] introduced a powerful new technique for obtaining estimates for the error in the normal approximation. His approach was subsequently extended by Chen [8] to Poisson approximation. Further extensions to other distributions, asymptotic expansions, as well as to multivariate and functional settings have also been undertaken. The general approach is presented in the monograph Stein [39], along with the specializations relevant

* Postal address: Faculty of Mathematics and Physics
Jadranska 19
SI-1000 Ljubljana
Slovenia

* Email address: martin.raic@fmf.uni-lj.si

to normal and Poisson approximation. A comprehensive presentation of the Poisson approximation is given in Barbour, Holst and Janson [3]; for an overview of normal approximation by Stein's method, see Rinott and Rotar [30] and the references therein.

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 $\mathbf{E} \mathcal{A}g(W)$ are small. The next step is then to solve the equation:

$$\mathcal{A}g = f - c_f \tag{1.1}$$

for a suitable constant c_f depending on f ; if $\mathbf{E} \mathcal{A}g(W)$ is small, we then have $\mathbf{E} f(W) \approx c_f$. In the present paper, we modify the second step in the way that c_f need not be a constant, but a function whose expectation is easy to derive or estimate; for details, see (1.9).

A remarkable feature of Stein's method is that it can be applied in many circumstances where dependence plays a part. In particular, Stein's method is suitable for certain sums of dependent random variables with no natural ordering. In the context of normal approximation, numerous applications include simple random sampling (see Bolthausen [5], Schneller [35], Bolthausen and Götze [6] and Goldstein and Reinert [15]), local dependence (see Rinott [27] and Rinott and Rotar [28]), random graphs (see Barbour, Karoński and Ruciński [4]), Nash equilibria (see Rinott and Scarsini [31]) and many others.

In the first applications of Stein's method related to normal approximation, the error was expressed in terms of smooth or Lipschitz test functions; gradually, the method was refined to yield uniform bounds, i. e.:

$$|\mathbf{P}(W \leq x) - \Phi(x)| \leq \varepsilon \tag{1.2}$$

uniformly in all $x \in \mathbb{R}$, where W is a sum of weakly dependent random variables with $\mathbf{E} W = 0$ and $\text{var}(W) = 1$ and where:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}z^2} dz \tag{1.3}$$

Nevertheless, this has not been done in such a generality as for smooth and Lipschitz test functions. Apart from special cases, comparable bounds have only been derived for bounded random variables. In the present paper, we relax the boundedness condition to *conditional* boundedness, which allows

for certain unbounded, but in some sense independent components. In particular, this is useful in the study of certain properties of random graphs.

Recently, Stein's method was refined to yield bounds of the following form:

$$|\mathbf{P}(W \leq x) - \Phi(x)| \leq \frac{\varepsilon}{1 + |x|^3} \quad (1.4)$$

for sums of independent random variables (see Shen and Shao [9]), locally dependent random variables (see Chen and Shao [11]) and certain non-linear statistics of independent random variables (see Chen and Shao [10]).

However, considering the *relative* error $|\mathbf{P}(W \leq x) - \Phi(x)|/\Phi(x)$ or $|\mathbf{P}(W \geq x) - \Phi(-x)|/\Phi(-x)$, even (1.4) allows to bound it only for not too large x . For heavy-tailed random variables, this is inevitable. On the other hand, for many sums of random variables with finite exponential moments, approximations with the error of substantially faster decay as $|x| \rightarrow \infty$ can be derived, allowing us to take control over the relative error, too. The first such result is due to Cramér [12], where such bounds for independent random variables are derived. This result was extended and improved by Petrov [25] and many others (for a recent improvement, see Sakhanenko [34]). Various extensions to dependent random variables have also been undertaken, some of them are mentioned in Section 3. The following assertion is a consequence of Petrov's [25] result:

Theorem 1.1. *Suppose that:*

$$W = \frac{\xi_1 + \dots + \xi_n}{\sqrt{n}} \quad (1.5)$$

is a sum of i. i. d. variables with $\mathbf{E} \xi_1 = 0$ and $\text{var}(\xi_1) = 1$, satisfying Cramér's condition:

$$\mathbf{E} e^{H\xi_1} < \infty, \quad \mathbf{E} e^{-H\xi_1} < \infty \quad \text{for some } H > 0 \quad (1.6)$$

Then for all $0 \leq x \leq C_1\sqrt{n}$, we have:

$$\mathbf{P}(W \geq x) = \exp\left(\frac{x^3}{\sqrt{n}} \lambda\left(\frac{x}{\sqrt{n}}\right)\right) \left(\Phi(-x) + \frac{\theta^+}{\sqrt{n}}\right) \quad (1.7)$$

$$\mathbf{P}(W \leq -x) = \exp\left(-\frac{x^3}{\sqrt{n}} \lambda\left(-\frac{x}{\sqrt{n}}\right)\right) \left(\Phi(-x) + \frac{\theta^-}{\sqrt{n}}\right) \quad (1.8)$$

for some θ^+ and θ^- with $|\theta^+|, |\theta^-| \leq C_2$, where $\lambda(z)$ is a power series in z with coefficients depending on the moments of ξ_1 and where C_1 and C_2 are constants depending only on H , $\mathbf{E} e^{H\xi_1}$ and $\mathbf{E} e^{-H\xi_1}$.

Remark. Since $\lambda(z) \sim \frac{1}{6} \mathbf{E} \xi_1^3$ for $z = o(1)$, the standard normal approximation produces small relative error only for $x = o(n^{1/6})$ if $\mathbf{E} \xi_1^3 \neq 0$; for larger x , there is no general approximation with small relative error depending only on the expectation and the variance. Most approximations consider the third and higher moments of ξ_1 . Nevertheless, even without explicit data on higher moments, $\lambda(z)$ can be suitably estimated in terms of $\mathbf{E} e^{H\xi_1}$ and $\mathbf{E} e^{-H\xi_1}$ for $z = O(1)$, so that useful upper and lower bounds for probabilities of large deviations can still be derived for $x = O(n^{1/2})$.

Surprisingly, in the context of Stein's method, little effort has been put into large deviation estimates related to the central limit theorem. A heuristic treatment is given in Stein [39], but with no explicit result given. Our approach, inspired by Cramér's transformation, is based on writing the Stein equation in the following way:

$$h'(w) - h(w)w = f(w) - N_\lambda f e^{\lambda w} \quad (1.9)$$

where $N_\lambda f$ is a constant factor, chosen so that the expectation of the r. h. s. with respect to the standard normal distribution vanishes. Like in the classical setting when $\lambda = 0$, this allows us to conclude that $\mathbf{E} f(W) \approx N_\lambda f \mathbf{E} e^{\lambda W}$ with small *absolute* error. However, the r. h. s. of (1.9) is flexible enough to yield an approximation with small *relative* error for suitably chosen λ . Although $e^{\lambda W}$ is no longer a constant, its expectation can also be estimated with small relative error. For a function f with a support away from the origin, we thus get useful estimates for large deviations.

In contrast to Theorem 1.1, we shall focus on the case where only the expectation and variance are explicitly known; thus, our result can be considered as a central limit theorem with non-uniform bounds. Nevertheless, the upper and the lower bounds for the error in the normal approximation are different and lower bounds also remain non-trivial in some cases where the exact probabilities are of different order than the standard normal probabilities.

In view of Theorem 2.3, we shall be able to derive results of the following form:

$$\mathbf{P}(W \geq x) = \exp\left(\eta \frac{x^3}{6M}\right) \left[\Phi(-x) + \theta\beta(1+x)e^{-\frac{1}{2}x^2} \right] \quad ; 0 \leq x \leq M \quad (1.10)$$

$$\mathbf{P}(W \geq x) \leq \exp\left\{-\frac{1}{3}M^3 - M(x - M)\right\} \quad ; x \geq M \quad (1.11)$$

where M and β are constants (depending on W) and where $\eta, \theta \in [-1, 1]$ depend on W and x . In the easiest case (1.5), where ξ_1 satisfies:

$$\gamma = e^{2c\mathbf{E}|\xi_1|} \mathbf{E} e^{c|\xi_1|} |\xi_1|^3 < \infty \quad (1.12)$$

for some $c > 0$, we can set:

$$M = C_1\sqrt{n}, \quad \beta = C_2/\sqrt{n} \quad (1.13)$$

where C_1 and C_2 depend only on c and γ (for details, see Section 3).

As an example of dependent random variables, we consider random graph degree statistics (other applications are given in Section 3). Suppose that Γ is a random graph on n vertices where any two vertices are adjacent with probability $\Delta/(n-1)$, independently of other pairs of vertices. Take a bounded non-constant function $h: \mathbb{Z}_+ \rightarrow \mathbb{R}$ and define:

$$S := \sum_{i \in I} h(\delta_i), \quad W := \text{var}(S)^{-1/2}(S - \mathbf{E}S) \quad (1.14)$$

where δ_i denotes the degree of the vertex i . In Section 3, we prove that W satisfies (1.10) and (1.11) together with (1.13) for some C_1 and C_2 depending only on Δ and the function h . Thus, the result is of the same quality as for i. i. d. random variables.

Comparing (1.10) and (1.13) with (1.7), we find that the upper bound on $\mathbf{P}(W \geq x)$ in (1.10) obtained by substituting $\eta = \theta = 1$ has the same behavior as the upper bound obtained from (1.7). However, the lower bounds (obtained by substituting $\eta = \theta = -1$) are of the same behavior only for $x = O(n^{1/4})$; if x is greater, the lower bound in (1.10) is negative and therefore useless.

It is possible to refine the approximation (1.10) to yield small relative error and thus non-trivial lower bound for $x = o(n^{1/2})$, i. e. to derive a result comparable to Theorem 1.1. However, this requires all moments of W and Cramér's condition (1.6) no longer assures convergence of the generalized Cramér series. In fact, we have to consider correlations of all orders, which gives rise to considerable technical detail. Such refinements may be considered in future.

2 Main results

In this section, we shall state the precise formulation of our results, which will be proved in Section 5. We shall focus on random variables which can

be decomposed as in Barbour, Karoński and Ruciński [4] (see below). This elegant and powerful approach was applied to random graphs, but is, as we shall see in Section 3, relevant in most of the cases where Stein's method for normal approximation has been applied. According to Barbour, Karoński and Ruciński, let I be an index set and:

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

For every $i \in I$, suppose that:

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

where W_i is independent of X_i . Moreover, suppose that:

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

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

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

As already mentioned, we shall assume certain *conditional* boundedness. In fact, the construction described below is an extension of the argument used in Barbour [2]. We assume that for each $k \in K_i$, there exist a σ -algebra \mathcal{H}_{ik} and random variables W_{ik} , W_{ik}^* , U_{ik} , V_{ik} , U_{ik}^* and \tilde{U}_{ik}^* , such that:

$$W_{ik}^* \text{ is independent of } \mathcal{H}_{ik} \text{ and has the same distribution as } W \quad (2.5)$$

$$U_{ik}, V_{ik} \text{ and } U_{ik}^* \text{ are } \mathcal{H}_{ik}\text{-measurable} \quad (2.6)$$

$$\tilde{U}_{ik}^* \text{ is conditionally independent of } W_{ik} \text{ given } \mathcal{H}_{ik} \quad (2.7)$$

$$|W - W_{ik}| \leq U_{ik} \quad (2.8)$$

$$|W_i - W_{ik}| \leq V_{ik} \quad (2.9)$$

$$|W_{ik}^* - W_{ik}| \leq \tilde{U}_{ik}^* \quad (2.10)$$

$$\mathbf{P}(\tilde{U}_{ik}^* \leq U_{ik}^* \mid \mathcal{H}_{ik}) > 0 \quad \text{a. s.} \quad (2.11)$$

Now define:

$$\eta_{ik} := \frac{1}{\mathbf{P}(\tilde{U}_{ik}^* \leq U_{ik}^* \mid \mathcal{H}_{ik})} \quad (2.12)$$

The following theorem provides uniform bounds in the normal approximation:

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

$$|\mathbf{P}(W \leq x) - \Phi(x)| \leq 2.33 \sum_{i \in I} \sum_{k \in K_i} \mathbf{E} \left[|X_i Z_{ik}| \eta_{ik} (U_{ik} + V_{ik} + 4U_{ik}^* + 2\tilde{U}_{ik}^*) \right] \quad (2.13)$$

Remark. Although the bound on the r. h. s. of (2.13) depends roughly on third absolute moments, decompositions stated above are mostly intended for *bounded* random variables. This is because of the assumptions (2.6) and (2.7). However, (2.7) also allows for certain independent unbounded components. This is particularly important for degree statistics (see Section 3).

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

$$\begin{aligned} \beta(\lambda) := & \sum_{i \in I} \sum_{k \in K_i} \mathbf{E} |X_i Z_{ik}| \eta_{ik} \left[e^{\lambda(U_{ik}^* + \tilde{U}_{ik}^*)} (U_{ik}^* + \tilde{U}_{ik}^*) + \right. \\ & \left. + e^{\lambda(\max\{U_{ik}, V_{ik}\} + U_{ik}^*)} \left(\frac{U_{ik} + V_{ik}}{2} + U_{ik}^* \right) \right] \end{aligned} \quad (2.14)$$

$$\beta^*(\lambda) := \int_0^\lambda t^2 \beta(t) dt$$

The following theorem is our main result.

Theorem 2.2. *For every $x \in \mathbb{R}$ and every $\lambda \geq 0$, we have:*

$$\mathbf{P}(W \geq x) = e^{\eta^+ \beta^*(\lambda)} \left[\Phi(-x) + (4.66 + 8.58\lambda)\theta^+ e^{\frac{1}{2}\lambda^2 - \lambda x} \beta(\lambda) \right] \quad (2.15)$$

$$\mathbf{P}(W \leq -x) = e^{\eta^- \beta^*(\lambda)} \left[\Phi(-x) + (4.66 + 8.58\lambda)\theta^- e^{\frac{1}{2}\lambda^2 - \lambda x} \beta(\lambda) \right] \quad (2.16)$$

for some $\eta^+, \eta^-, \theta^+, \theta^- \in [-1, 1]$. Moreover,

$$\max\{\mathbf{P}(W \geq x), \mathbf{P}(W \leq -x)\} \leq \exp\left\{\frac{1}{2}\lambda^2 - \lambda x + \beta^*(\lambda)\right\} \quad (2.17)$$

Now we turn to two special cases with respect to the the growth of $\beta(\lambda)$. The simplest case is when $\beta(\lambda)$ can be uniformly bounded in some finite interval.

Theorem 2.3. *Suppose that $\beta(\lambda) \leq \beta_1$ for all $\lambda \leq M$, where $M \leq 1/(2\beta_1)$. Then for all $0 \leq x \leq M$, we have:*

$$\mathbf{P}(W \geq x) = \exp\left(\eta^+ \frac{x^3}{6M}\right) \left[\Phi(-x) + (4.66 + 8.58x)\theta^+ \beta_1 e^{-\frac{1}{2}x^2} \right] \quad (2.18)$$

$$\mathbf{P}(W \leq -x) = \exp\left(\eta^- \frac{x^3}{6M}\right) \left[\Phi(-x) + (4.66 + 8.58x)\theta^- \beta_1 e^{-\frac{1}{2}x^2} \right] \quad (2.19)$$

for some $\eta^+, \eta^-, \theta^+, \theta^- \in [-1, 1]$. In addition, the following estimate holds for $x \geq M$:

$$\max\{\mathbf{P}(W \geq x), \mathbf{P}(W \leq -x)\} \leq \exp\{-\frac{1}{3}M^2 - M(x - M)\} \quad (2.20)$$

The other special case we shall consider is when:

$$\beta(\lambda) \leq \beta_1 e^{\beta_2 \lambda} \quad (2.21)$$

for all $\lambda \geq 0$.

Theorem 2.4. *Suppose that (2.21) holds. Put:*

$$M := \frac{1}{\max\{2e\beta_1, \beta_2\}} \quad (2.22)$$

Then for all $0 \leq x \leq M$, we have:

$$\mathbf{P}(W \geq x) = \exp\left(\eta^+ \frac{x^3}{6M}\right) \left[\Phi(-x) + (4.66 + 8.58x)\theta^+ \beta_1 e^{\beta_2 x - \frac{1}{2}x^2} \right] \quad (2.23)$$

$$\mathbf{P}(W \leq -x) = \exp\left(\eta^- \frac{x^3}{6M}\right) \left[\Phi(-x) + (4.66 + 8.58x)\theta^- \beta_1 e^{\beta_2 x - \frac{1}{2}x^2} \right] \quad (2.24)$$

for some $\eta^+, \eta^-, \theta^+, \theta^- \in [-1, 1]$. In addition, the following estimate holds for $x \geq M$:

$$\max\{\mathbf{P}(W \geq x), \mathbf{P}(W \leq -x)\} \leq e^{-\frac{1}{3}M^2} \left(\frac{M}{x}\right)^{\frac{1}{3}Mx} \quad (2.25)$$

3 Applications

3.1 Independent random variables

Let $X_i, i \in I$, be independent random variables with $\mathbf{E} X_i = 0$. Put $W := \sum_{i \in I} X_i$ and assume that $\text{var}(W) = 1$. Set:

$$\begin{aligned} K_i &:= \{0\}, & Z_i &:= Z_{i0} := X_i \\ W_i &:= W_{i0} := W - X_i \\ \mathcal{H}_{i0} &:= \sigma(X_i) \end{aligned} \quad (3.1)$$

Furthermore, for each $i \in I$, let X_i^* be an independent copy of X_i . Set:

$$\begin{aligned}
W_{i0}^* &:= W_{i0} + X_i^* \\
U_{i0} &:= |X_i| = |W - W_{i0}| \\
V_{i0} &:= 0 = |W_i - W_{i0}| \\
\tilde{U}_{i0}^* &:= |X_i^*| = |W_{i0}^* - W_{i0}|
\end{aligned} \tag{3.2}$$

Letting:

$$U_{i0}^* := 2 \mathbf{E} |X_i| = 2 \mathbf{E} |X_i^*| \tag{3.3}$$

we have, by Markov's inequality:

$$P(\tilde{U}_{i0}^* > U_{i0}^* \mid \mathcal{H}_{i0}) = P(|X_i^*| > 2 \mathbf{E} |X_i^*|) \leq \frac{1}{2} \tag{3.4}$$

so that $\eta_{i0} \leq 2$ and the conditions for Theorems 2.1 and 2.2 are satisfied. For uniform bounds, Theorem 2.1 yields:

$$\begin{aligned}
|\mathbf{P}(W \leq x) - \Phi(x)| &\leq 4.66 \sum_{i \in I} \mathbf{E} X_i^2 (|X_i| + 2|X_i^*| + 8 \mathbf{E} |X_i|) = \\
&= 4.66 \sum_{i \in I} (\mathbf{E} |X_i|^3 + 10 \mathbf{E} X_i^2 \mathbf{E} |X_i|)
\end{aligned} \tag{3.5}$$

Making use of Jensen's inequality, we obtain:

$$|\mathbf{P}(W \leq x) - \Phi(x)| \leq 51.3 \sum_{i \in I} \mathbf{E} |X_i|^3 \tag{3.6}$$

which is the classical Berry–Esseen theorem, but with a constant which is far from optimal. Constants can be substantially reduced, but at the expense of added technical detail and loss of generality related to the dependence structure (see Section 4).

Now we turn to large deviation estimates. Recalling (2.14), we have:

$$\begin{aligned}
\beta(\lambda) &\leq 2 \sum_{i \in I} \mathbf{E} X_i^2 \left[e^{\lambda(|X_i| + 2 \mathbf{E} |X_i|)} \left(\frac{1}{2} |X_i| + 2 \mathbf{E} |X_i| \right) + \right. \\
&\quad \left. + e^{\lambda(|X_i^*| + 2 \mathbf{E} |X_i|)} (|X_i^*| + 2 \mathbf{E} |X_i|) \right] = \\
&= \sum_{i \in I} e^{2\lambda \mathbf{E} |X_i|} \mathbf{E} e^{\lambda |X_i|} \left[|X_i|^3 + 4 X_i^2 \mathbf{E} |X_i| + \right. \\
&\quad \left. + 2 |X_i| \mathbf{E} X_i^2 + 4 \mathbf{E} |X_i| \mathbf{E} X_i^2 \right]
\end{aligned} \tag{3.7}$$

If f is an increasing function, random variables X and $f(X)$ are positively correlated, i. e. $\mathbf{E} X \mathbf{E} f(X) \leq \mathbf{E} X f(X)$. Making use of this simple inequality, we obtain after some calculation:

$$\beta(\lambda) \leq 11 \sum_{i \in I} e^{2\lambda \mathbf{E}|X_i|} \mathbf{E} e^{\lambda|X_i|} |X_i|^3 \quad (3.8)$$

so that we can apply Theorem 2.2. In particular, if all random variables X_i have the same law as a random variable ξ/\sqrt{n} and:

$$\gamma = e^{2c \mathbf{E}|\xi|} \mathbf{E} e^{c|\xi|} |\xi|^3 < \infty \quad (3.9)$$

for some $c > 0$, we can apply Theorem 2.3 with $\beta_1 := 11n^{-1/2}\gamma$ and $M := \min\{1/(22\gamma), c\}\sqrt{n}$. Thus we have proved (1.10) and (1.11).

3.2 Local dependence

This is one of the most important cases where Stein's method can be readily applied. In this section, we shall extend Rinott's [27] Berry–Esseen type result for bounded locally dependent random variables. More precisely, Rinott [27] considers the following:

Definition. Let I be an index set. A graph Γ with the vertex set I is said to be a *dependence graph* for a collection of random variables X_i , $i \in I$, if the collections $\{X_k : k \in K\}$ and $\{X_l : l \in L\}$ are independent for any disjoint subsets $K, L \subseteq I$, such that no vertex in K is adjacent to a vertex in L .

Now suppose that Γ is a dependence graph for a collection $\{X_i : i \in I\}$. For each $i \in I$, denote by K_i the set of all vertices adjacent to i , including i itself. To construct decompositions from Section 2 for the sum $W := \sum_{i \in I} X_i$, first put:

$$Z_{ik} := X_k, \quad Z_i := \sum_{k \in K_i} Z_{ik}, \quad \mathcal{H}_{ik} := \sigma\{X_i, X_k\} \quad (3.10)$$

In order to construct second-order decompositions, we consider a conditionally independent copy $\{X_{ikl}^* : l \in I\}$ of the original family $\{X_i : i \in I\}$ given $\mathcal{G}_{ik} := \sigma\{X_l : l \in I \setminus (K_i \cup K_k)\}$. For all $l \in I \setminus (K_i \cup K_k)$, X_l is \mathcal{G}_{ik} -measurable, so that $X_l^* = X_l$ a. s. Hence the sum:

$$W_{ik} := W_{ik}^* := \sum_{l \in I \setminus (K_i \cup K_k)} X_{ikl} + \sum_{l \in K_i \cup K_k} X_{ikl}^* \quad (3.11)$$

is a conditionally independent copy of W given \mathcal{H}_{ik} .

Now suppose that the maximum degree of Γ is bounded by D and furthermore:

$$|I| = N, \quad |X_i| \leq B \text{ for all } i \in I \quad (3.12)$$

for some constant B . Thus,

$$\begin{aligned} |W - W_{ik}^*| &\leq 2|K_i \cup K_k| B =: U_{ik} \\ |W_i - W_{ik}^*| &\leq (|K_i \cup K_k| + |K_k \setminus K_i|) B =: V_{ik} \end{aligned} \quad (3.13)$$

and we can set $U_{ik}^* := \tilde{U}_{ik}^* := 0$. As $|K_i| \leq D + 1$, we have $U_{ii} \leq 2(D + 1)B$ and $V_{ii} \leq (D + 1)B$. In addition, for $k \in K_i \setminus \{i\}$, we have $|K_i \cup K_k| \leq 2D$ and $|K_k \setminus K_i| \leq D - 1$, so that $U_{ik} \leq 4DB$ and $V_{ik} \leq (3D - 1)B$. Thus, recalling (2.14), a straightforward calculation shows that $\beta(\lambda) \leq \beta_1 e^{\beta_2 \lambda}$, where:

$$\beta_1 := \frac{1}{2}(7D^2 + 2D + 3)NB^3, \quad \beta_2 := \max\{2, 4D\}B \quad (3.14)$$

and the following assertion extends Rinott's [27] result:

Theorem 3.1. *Under the conditions given above, Theorem 2.4 holds with β_1 and β_2 as in (3.14), provided that $\mathbf{E} X_i = 0$ for all i and that $\text{var}(W) = 1$. In particular, for independent random variables, one can take $\beta_1 = \frac{3}{2}NB^3$ and $\beta_2 = 2B$.*

Remark. Considering a triangular array of locally dependent random variables where D is bounded independently of N , we typically have $B = O(N^{-1/2})$, so that also $\beta_1, \beta_2 = O(N^{-1/2})$.

We shall consider two special cases of local dependence.

3.2.1 U -statistics.

Let η_1, η_2, \dots be i. i. d. random variables taking values in some measurable space (S, \mathcal{S}) . Suppose that $h: S^r \rightarrow \mathbb{R}$ is a symmetric bounded function with:

$$\mathbf{E} h(\eta_1, \dots, \eta_r) = 0, \quad \text{var}[\mathbf{E}(h(\eta_1, \dots, \eta_r) \mid \eta_1)] > 0 \quad (3.15)$$

Take $n \geq r$. For every subset $\alpha = \{i_1, \dots, i_r\} \subseteq \{1, \dots, n\}$ with exactly r elements, define:

$$Y_\alpha := h(\eta_{i_1}, \dots, \eta_{i_r}) \quad (3.16)$$

Thus,

$$U := \sum_{|\alpha|=r} Y_\alpha \quad (3.17)$$

where $|\cdot|$ denotes cardinality, is a non-degenerate U -statistic.

Large deviations of U -statistics have been considered, among others, by Rubin and Sethuraman [32], Slastnikov [37], Malevich and Abdalimov [24], Vandermaele [40], Aleshkyavichene [1], Inglot and Ledwina [19], Keener, Robinson and Weber [22] and Borovskikh and Weber [7]. Under various conditions, approximation formulae for tail probabilities $\mathbf{P}(U \geq x)$ with small relative error as $n \rightarrow \infty$ (keeping h fixed) are derived within the range up to $x = o(\sigma\sqrt{n})$, where $\sigma^2 = \text{var}(U)$ (the latter range was achieved by Aleshkyavichene [1] for $r = 2$). Below we derive an approximation with the relative error being small within the range $x = o(\sigma n^{1/6})$ (which could be extended by taking higher moments into consideration; see the example with independent random variables). However, our approach is different from the classical treatment of U -statistics, based on Höfdding's [18] decompositions, as well as from the argument of Aleshkyavichene [1], based on the method of moments. Stein's method provides an alternative way, which is in some circumstances (x not too large, $r > 2$) sharp enough, but more direct and elegant.

In order to apply Theorem 3.1, observe that the random variables Y_α are locally dependent with respect to the dependence graph where α and β are adjacent if $\alpha \cap \beta \neq \emptyset$. Writing $a_n \asymp b_n$ for sequences of positive real numbers with $0 < \liminf_{n \rightarrow \infty} a_n/b_n \leq \limsup_{n \rightarrow \infty} a_n/b_n < \infty$, the dependence graph has $N \asymp n^r$ vertices and maximum degree $D \asymp n^{r-1}$. Setting:

$$W := \sigma^{-1}U, \quad X_\alpha := \sigma^{-1}Y_\alpha \quad (3.18)$$

and noting that $\sigma \asymp n^{r-1/2}$, the random variables X_α can be bounded by $B \asymp n^{1/2-r}$. Making use of Theorem 3.1, we find that Theorem 2.4 holds for $\beta_1 = C_1 n^{-1/2}$ and $\beta_2 = C_2 n^{-1/2}$, where C_1 and C_2 only depend on the function h , the conditional variance $\text{var}[\mathbf{E}(h(\eta_1, \dots, \eta_r) \mid \eta_1)]$ and r .

3.2.2 Nash equilibria in random games.

Consider a game with p players where the k -th player chooses a pure strategy $i_k \in \{1, \dots, s\}$. Denote by $V_{\mathbf{i}}^{(k)}$ the payoff of the k -th player with respect to the vector of chosen strategies $\mathbf{i} = (i_1, \dots, i_p)$. We say that \mathbf{i} is a *Nash equilibrium*

point if $V_{\mathbf{i}}^{(k)} \geq V_{\mathbf{j}}^{(k)}$ for all $\mathbf{j} = (i_1, \dots, i_{k-1}, j, i_{k+1}, \dots, i_p)$, $j \in \{1, \dots, s\}$, and for all $k \in \{1, \dots, p\}$. Denote by S the number of Nash equilibria.

We shall consider games with random payoffs, such that the payoff vectors $V_{\mathbf{i}} = (V_{\mathbf{i}}^{(1)}, \dots, V_{\mathbf{i}}^{(p)})$, $\mathbf{i} \in \{1, \dots, s\}^p =: I$, are independent and identically distributed. Rinott and Scarsini [31] investigate asymptotic distribution of S under several conditions on the distribution of $V_{\mathbf{i}}$. In particular, they use Rinott's [27] result to obtain a Berry–Esseen type theorem.

Writing:

$$S = \sum_{\mathbf{i} \in I} Y_{\mathbf{i}} \quad (3.19)$$

where:

$$Y_{\mathbf{i}} := \begin{cases} 1 & \text{if } \mathbf{i} \text{ is a Nash equilibrium point} \\ 0 & \text{otherwise} \end{cases} \quad (3.20)$$

observe that $Y_{\mathbf{i}}$ are locally dependent with respect to the dependence graph Γ , where two strategy vectors are adjacent if they differ in at most two components. Observe that Γ is a regular graph with s^p vertices of degree $(s-1)p + \binom{p}{2}(s-1)^2 \leq (sp)^2 - 1$. Setting:

$$\sigma^2 := \text{var}(S), \quad W := \frac{S - \mathbf{E} S}{\sigma} \quad (3.21)$$

it follows from Theorem 3.1 that Theorem 2.4 holds with:

$$\beta_1 = \frac{7}{2}(sp)^4 s^p \sigma^{-3} Q, \quad \beta_2 = 4(sp)^2 \sigma^{-1} Q \quad (3.22)$$

where Q is the probability that a particular strategy vector is a Nash point.

Now suppose that $V_{\mathbf{i}}$ is a multivariate normal vector with exchangeable components, such that:

$$\rho := \frac{\text{cov}(V_{\mathbf{i}}^{(k)}, V_{\mathbf{i}}^{(l)})}{\text{var}(V_{\mathbf{i}}^{(k)})^{1/2} \text{var}(V_{\mathbf{i}}^{(l)})^{1/2}} > 0 \quad (3.23)$$

Rinott and Scarsini [31] study the asymptotic behavior of Q and σ . In particular, they show that $\sigma^2 \geq c(\rho)s^p Q$ for some $c(\rho) > 0$, provided that sp is large. Hence Theorem 2.4 holds with:

$$\beta_1 = C_1(\rho) \frac{(sp)^4}{s^{p/2} Q^{1/2}}, \quad \beta_2 = C_2(\rho) (sp)^2 \frac{Q^{1/2}}{s^{p/2}} \quad (3.24)$$

for some $C_1(\rho), C_2(\rho) > 0$. This extends Theorem 5.2 in Rinott and Scarsini [31].

3.3 Finite population statistics

Consider a statistic W based on a simple random sample of size n is drawn from a population of size $N \geq n$. Any such statistic can be written in the form:

$$W = h(\pi) \quad (3.25)$$

where, writing $\mathbb{N}_n := \{1, \dots, n\}$, $\pi: \mathbb{N}_n \rightarrow \mathbb{N}_N$ is a uniformly distributed random injection and $h: \mathbb{N}_N \rightarrow \mathbb{R}$ is a function. Suppose that h is decomposed in the following way:

$$h = \sum_{\substack{\alpha \subseteq \mathbb{N}_n \\ 0 < |\alpha| \leq r}} h_\alpha \quad (3.26)$$

for some $r \leq n$ (typically considerably smaller than n), where $|\cdot|$ denotes cardinality and where $h_\alpha(\rho)$ only depends on the images of the elements of α under ρ . In other words, for $\alpha = \{a_1, \dots, a_r\}$, h_α must be of the form:

$$h_\alpha(\rho) = g(\rho(a_1), \dots, \rho(a_r)) \quad (3.27)$$

For $r = 1$, we obtain linear rank statistics. In the context of large deviations, they were considered by Kallenberg [21] and Inogamov [20], and also by Seoh, Ralesen and Puri [36] and Puri [26], who consider a more general class of statistics, including linear signed rank statistics, which were considered by Dufour [14].

Another interesting case is the one where the functions h_α are all equal for $|\alpha| = r$ and $h_\alpha = 0$ for $|\alpha| < r$. In this case, we obtain U -statistics of finite populations. Large deviations of them were considered by Kokic and Weber [23].

To construct decompositions from Section 2, first consider random mappings $\tau_{A,B}: \mathbb{N}_N \rightarrow \mathbb{N}_N$ defined for any two subsets $A, B \subset \mathbb{N}_N$ of the same cardinality, so that they are drawn uniformly from all maps mapping $A \setminus B$ bijectively onto $B \setminus A$ and vice versa and leaving the other elements (i. e. $(A \cap B) \cup (A \cup B)^c$) unchanged. Thus $\tau_{A,B}$ maps the set A bijectively onto B and $\mathbb{N}_N \setminus A$ onto $\mathbb{N}_N \setminus B$. The following lemma is straightforward and is therefore left without proof.

Lemma 3.2. *Let π and ρ be independent and uniformly distributed random injections $\mathbb{N}_n \rightarrow \mathbb{N}_N$. Generate a family of random mappings $\{\tau_{A,B}: A, B \subseteq \mathbb{N}_N\}$, satisfying the conditions above and independent of the pair (π, ρ) . Then for any subset $\alpha \subseteq \mathbb{N}_n$, the random mapping $\tau_{\pi(\alpha), \rho(\alpha)} \circ \pi$ is independent of the family $\{\pi(a): a \in \alpha\}$ and has the same distribution as π .*

Now take an independent copy ρ of the random map π and generate a family $\{\tau_{A,B} : A, B \subseteq \mathbb{N}_N\}$ of random mappings being as in Lemma 3.2 and independent of the pair (π, ρ) . Define:

$$\begin{aligned} X_\alpha &:= h_\alpha(\pi) & \pi_\alpha &:= \tau_{\pi(\alpha), \rho(\alpha)} \circ \pi \\ X_{\alpha\beta} &:= h_\beta(\pi_\alpha) & W_\alpha &:= \sum_{\beta \subseteq \mathbb{N}_n} X_{\alpha\beta} \\ Z_{\alpha\beta} &:= X_\beta - X_{\alpha\beta} & Z_\alpha &:= \sum_{\beta \subseteq \mathbb{N}_n} Z_{\alpha\beta} \end{aligned} \quad (3.28)$$

By Lemma 3.2, π_α is independent of X_α ; the same holds for W_α . Next, let $\mathcal{H}_{\alpha\beta}$ be the σ -algebra generated by $\pi|_{\alpha \cup \beta}$, ρ and $\{\tau_{A,B} : A, B \subseteq \mathbb{N}_N\}$. Notice that X_α , X_β and $X_{\alpha\beta}$ are $\mathcal{H}_{\alpha\beta}$ -measurable. Thus, to complete the decompositions, take another random injection $\rho' : \mathbb{N}_n \rightarrow \mathbb{N}_N$ and generate another family $\{\tau'_{A,B} : A, B \subseteq \mathbb{N}_N\}$ from Lemma 3.2, such that $(\pi, \rho, \{\tau_{A,B} : A, B \subseteq \mathbb{N}_N\})$, ρ' and $\{\tau'_{A,B} : A, B \subseteq \mathbb{N}_N\}$ are independent. Put:

$$\begin{aligned} \pi_{\alpha\beta} &:= \tau'_{\pi(\alpha \cup \beta), \rho'(\alpha \cup \beta)} \circ \pi, & X_{\alpha\beta\gamma} &:= h_\gamma(\pi_{\alpha\beta}) \\ W_{\alpha\beta} &:= W_{\alpha\beta}^* := \sum_{\gamma \subseteq \mathbb{N}_n} X_{\alpha\beta\gamma} \end{aligned} \quad (3.29)$$

Again by Lemma 3.2, $W_{\alpha\beta}$ is independent of $\mathcal{H}_{\alpha\beta}$. Moreover, it has the same distribution as W . Therefore these decompositions satisfy the independence conditions given in Section 2.

Now suppose that $\mathbf{E} X_\alpha = 0$ for all α and that $\text{var}(W) = 1$. Furthermore, suppose that there are some non-negative constants B_1, \dots, B_r , such that:

$$\|h_\alpha\|_0 := \max_{j \in \mathbb{N}_N} h_\alpha(j) - \min_{j \in \mathbb{N}_N} h_\alpha(j) \leq B_{|\alpha|} \quad (3.30)$$

for all $\alpha \subset \mathbb{N}_N$ with $0 < |\alpha| \leq r$. In this case, we can estimate:

$$\begin{aligned} |Z_{\alpha\beta}| &\leq B_{|\beta|} \mathbf{1}[(\pi(\alpha) \cup \rho(\alpha)) \cap \pi(\beta) \neq \emptyset] \\ |X_{\alpha\beta\gamma} - X_\gamma| &\leq B_{|\gamma|} \mathbf{1}[(\pi(\alpha \cup \beta) \cup \rho'(\alpha \cup \beta)) \cap \pi(\gamma) \neq \emptyset] \\ |X_{\alpha\beta\gamma} - X_{\alpha\gamma}| &\leq B_{|\gamma|} \mathbf{1}[(\pi(\alpha \cup \beta) \cup \rho(\alpha) \cup \rho'(\alpha \cup \beta)) \cap \pi(\gamma) \neq \emptyset] \end{aligned} \quad (3.31)$$

Consequently,

$$\begin{aligned} |W_{\alpha\beta} - W| &\leq 4rB =: U_{\alpha\beta} \\ |W_{\alpha\beta} - W_\alpha| &\leq 6rB =: V_{\alpha\beta} \end{aligned} \quad (3.32)$$

where:

$$B := \sum_{s=1}^r \binom{n-1}{s-1} B_s \quad (3.33)$$

Clearly, we set $U_{\alpha\beta}^* := \tilde{U}_{\alpha\beta}^* := 0$. Since:

$$\begin{aligned} \sum_{\beta} B_{|\beta|} \mathbf{1}[(\pi(\alpha) \cup \rho(\alpha)) \cap \pi(\beta) \neq \emptyset] &\leq 2|\alpha|B \\ \sum_{\alpha} |\alpha| B_{|\alpha|} &= \sum_{s=1}^r s \binom{n}{s} B_s = nB \end{aligned} \quad (3.34)$$

we can finally estimate $\beta(\lambda) \leq \beta_1 e^{\beta_2 \lambda}$, where:

$$\beta_1 := \frac{9}{2}nrB^3, \quad \beta_2 := 5rB \quad (3.35)$$

and we can apply Theorem 2.4.

Remark. In the case of U -statistics, we typically have $B = O(n^{-1/2})$ and thus $\beta_1, \beta_2 = O(n^{-1/2})$ – similarly as in the case of independent observations.

3.4 Random graph degree statistics

Let Γ be a random graph on the vertex set I , where any two distinct vertices i and j are adjacent with probability p_{ij} , independently of all other unordered pairs of vertices; note that $p_{ij} = p_{ji}$. An important class of problems is to consider random variables based on the so called *semi-induced properties* of certain subsets of I , that is, determined by those edges with at least endpoint in the subset. The most natural examples are ‘being an isolated tree’ or ‘having a given degree’. In the context of normal approximation by Stein’s method, problems of this kind were considered by Barbour [2], Barbour, Karoński and Ruciński [4] and Goldstein and Rinott [16].

In the present paper, we consider statistics based on degrees: for each $i \in I$, take a bounded measurable function $h_i: \mathbb{Z}_+ \rightarrow \mathbb{R}$, and define:

$$X_i := h_i(\delta_i); \quad W := \sum_{i \in I} X_i \quad (3.36)$$

where δ_i denotes the degree of the vertex i with respect to Γ . Without loss of generality, we may (and will) assume that $\mathbf{E} X_i = 0$ for all i and that

$\text{var}(W) = 1$. The decompositions from Section 2 can be constructed in the following way: 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:

$$W_i := \sum_{k \in I} h_k(\delta_k^{\{i\}}), \quad Z_{ik} := h_k(\delta_k) - h_k(\delta_k^{\{i\}}) \quad (3.37)$$

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 all edges with an endpoint in J as in Γ^* and the others as in Γ . Define:

$$W_{ik} := \sum_{l \in I} h_l(\delta_l^{\{i, k\}}), \quad W_{ik}^* := \sum_{l \in I} h_l(\delta_l^{\{i, k\}*}) \quad (3.38)$$

The main quantity used in the bounds of large deviation probabilities, $\beta(\lambda)$ (defined in (2.14)), will be bounded in terms of:

$$n := |I|, \quad B := \max_{i \in I} \|h_i\|_0 \quad (3.39)$$

where $\|h\|_0 := \sup_{r \in \mathbb{Z}_+} h(r) - \inf_{r \in \mathbb{Z}_+} h(r)$. Since $\mathbf{E} X_i = 0$, we have $|X_i| \leq B$. Clearly, $|Z_{ii}| \leq B$. Next, observe that:

$$Z_{ik} = \mathbf{1}[i \sim k] \left(h_k(\delta_k) - h_k(\delta_k^{\{i\}}) \right) \quad (3.40)$$

and thus:

$$|Z_{ik}| \leq B \mathbf{1}[i \sim k] \quad (3.41)$$

Furthermore,

$$\begin{aligned} |W - W_{ii}| &= \left| \sum_{l \in I} \left(h_l(\delta_l) - h_l(\delta_l^{\{i\}}) \right) \right| \leq B(\delta_i + 1) =: U_{ii} \\ |W_i - W_{ii}| &= 0 =: V_{ii} \\ |W_{ii}^* - W_{ii}| &= \left| \sum_{l \in I} \left(h_l(\delta_l^{\{i\}*}) - h_l(\delta_l^{\{i\}}) \right) \right| \leq B(\delta_i^* + 1) =: \tilde{U}_{ii}^* \end{aligned} \quad (3.42)$$

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

$$\begin{aligned}
|W - W_{ik}| &= \left| \sum_{l \in I} \left(h_l(\delta_l) - h_l(\delta_l^{\{i,k\}}) \right) \right| \leq B(\delta_i + \delta_k + 2) =: U_{ik} \\
|W_i - W_{ik}| &= \left| \sum_{l \in I} \left(h_l(\delta_l^{\{i\}}) - h_l(\delta_l^{\{i,k\}}) \right) \right| \leq B(\delta_k + 1) =: V_{ik} \\
|W_{ik}^* - W_{ik}| &= \left| \sum_{l \in I} \left(h_l(\delta_l^{\{i,k\}^*}) - h_l(\delta_l^{\{i,k\}}) \right) \right| \leq B(\delta_i^* + \delta_k^* + 2) =: \tilde{U}_{ik}^*
\end{aligned} \tag{3.43}$$

Setting:

$$U_{ii}^* := B(1 + 2\Delta), \quad U_{ik}^* := B(2 + 4\Delta), \quad i \neq k \tag{3.44}$$

where:

$$\Delta := \max_{i \in I} \sum_{k \in I \setminus \{i\}} p_{ik} = \max_{i \in I} \mathbf{E} \delta_i = \max_{i \in I} \mathbf{E} \delta_i^* \tag{3.45}$$

Markov's inequality implies:

$$\begin{aligned}
\mathbf{P}(\tilde{U}_{ii}^* \leq U_{ii}^* \mid \mathcal{H}_{ii}) &= \mathbf{P}(\tilde{U}_{ii}^* \leq U_{ii}^*) = \mathbf{P}(\delta_i^* \leq 2\Delta) \geq \frac{1}{2} \\
\mathbf{P}(\tilde{U}_{ik}^* \leq U_{ik}^* \mid \mathcal{H}_{ik}) &= \mathbf{P}(\tilde{U}_{ik}^* \leq U_{ik}^*) = \mathbf{P}(\delta_i^* + \delta_k^* \leq 4\Delta) \geq \frac{1}{2}
\end{aligned} \tag{3.46}$$

Recalling (2.14) and using independence, we obtain after some calculation:

$$\begin{aligned}
\beta(\lambda) &\leq B^3 \sum_{i \in I} \mathbf{E} e^{\lambda B(2+2\Delta+\delta_i)} (7 + 8\Delta + 3\delta_i) + \\
&\quad + B^3 \sum_{i \in I} \sum_{k \in I \setminus \{i\}} \mathbf{P}(i \sim k) \mathbf{E} \left[e^{\lambda B(6+4\Delta+(\delta_i-1)+(\delta_k-1))} \times \right. \\
&\quad \quad \left. \times (10 + 4\Delta + (\delta_i - 1) + 2(\delta_k - 1)) \mid i \sim k \right] + \\
&\quad + B^3 \sum_{i \in I} \sum_{k \in I \setminus \{i\}} \mathbf{P}(i \sim k) \mathbf{E} \left[e^{\lambda B(4+4\Delta+\delta_i+\delta_k)} (8 + 8\Delta + 2\delta_i + 2\delta_k) \right]
\end{aligned} \tag{3.47}$$

In particular,

$$\beta(0) \leq nB^3(7 + 29\Delta + 19\Delta^2) \tag{3.48}$$

and Theorem 2.1 yields:

$$|\mathbf{P}(W \leq x) - \Phi(x)| \leq nB^3(33 + 136\Delta + 89\Delta^2) \quad (3.49)$$

for all $x \in \mathbb{R}$.

To derive a large deviation result, we shall apply the following assertion, which will be proved at the end of the section:

Lemma 3.3. *Let ξ_1, \dots, ξ_n be independent Bernoulli random variables with sum S satisfying $\mathbf{E}S \leq \Delta$. The following estimate holds for $0 \leq t \leq 1/(10 + 10\Delta)$:*

$$\begin{aligned} \mathbf{E}e^{tS} &\leq e^{e^{1/10}-1} \\ \mathbf{E}e^{tS}S &\leq e^{e^{1/10}-1}\Delta \end{aligned} \quad (3.50)$$

Therefore, for every $0 \leq \lambda \leq M_0 := 1/(10B(1 + \Delta))$ and every $i \in I$, we have:

$$\mathbf{E}e^{\lambda B\delta_i} \leq e^{e^{1/10}-1}, \quad \mathbf{E}e^{\lambda B\delta_i}\delta_i \leq e^{e^{1/10}-1}\Delta \quad (3.51)$$

Similarly, for all $k \neq i$,

$$\begin{aligned} \mathbf{E}(e^{\lambda B(\delta_i-1)} \mid i \sim k) &\leq e^{e^{1/10}-1} \\ \mathbf{E}(e^{\lambda B(\delta_i-1)}(\delta_i - 1) \mid i \sim k) &\leq e^{e^{1/10}-1}\Delta \end{aligned} \quad (3.52)$$

Thus, for all $0 \leq \lambda \leq M_0$, we can estimate after some calculation:

$$\beta(\lambda) \leq nB^3(10 + 53\Delta + 38\Delta^2) =: \beta_1 \quad (3.53)$$

and we can apply Theorem 2.3 with β_1 as above and with $M := \min\{M_0, 1/(2\beta_1)\}$.

Now consider the special case where $p_{ij} = \Delta/(n-1)$ and where Δ remains fixed, while $n \rightarrow \infty$. Take a bounded non-constant function $h: \mathbb{Z}_+ \rightarrow \mathbb{R}$ and define:

$$S^{(n)} := \sum_{i \in I} h(\delta_i) \quad (3.54)$$

To compute the variance of $S^{(n)}$, observe that:

$$\begin{aligned}
\text{var}(S^{(n)}) &= \sum_{i \in I} \text{var}(h(\delta_i)) + \\
&\quad + \sum_{i \in I} \sum_{k \in I \setminus \{i\}} \mathbf{E} \left(h(\delta_i) - \mathbf{E} h(\delta_i) \right) \left(h(\delta_k) - h(\delta_k^{\{i\}}) \right) = \\
&= n \text{var} \left(h(B_{n-1}^{(n-1)}) \right) + n\Delta \left(\mathbf{E} h(1 + B_{n-2}^{(n-1)}) - \mathbf{E} h(B_{n-1}^{(n-1)}) \right) \times \\
&\quad \times \left(\mathbf{E} h(1 + B_{n-2}^{(n-1)}) - \mathbf{E} h(B_{n-2}^{(n-1)}) \right)
\end{aligned} \tag{3.55}$$

where $B_k^{(n)} \sim \text{Bi}(k, \Delta/n)$. By the Poisson law of small numbers, we have:

$$\text{var}(S^{(n)}) \sim n \text{var}(h(\Pi)) + n\Delta (\mathbf{E} h(1 + \Pi) - \mathbf{E} h(\Pi))^2 \tag{3.56}$$

where $\Pi \sim \text{Po}(\Delta)$. Setting:

$$W^{(n)} := \text{var}(S^{(n)})^{-1/2} (S^{(n)} - \mathbf{E} S^{(n)}) = \sum_{i \in I} h^{(n)}(\delta_i) \tag{3.57}$$

where:

$$h^{(n)} := \text{var}(S^{(n)})^{-1/2} \left(h - \mathbf{E} h(B_{n-1}^{(n-1)}) \right) \tag{3.58}$$

we have $\|h^{(n)}\|_0 \asymp n^{-1/2}$. Hence by (3.53), the random variables $W^{(n)}$ satisfy the conditions of Theorem 2.3 with $\beta_1 \asymp n^{-1/2}$ and $M \asymp n^{1/2}$. In particular, (1.10) and (1.11) hold.

Proof of Lemma 3.3. The key observation is the fact that the Bernoulli distribution $\text{Be}(p)$ precedes the Poisson distribution $\text{Po}(p)$ in the convex order sense, i. e. for every convex function f , we have $\int f d\text{Be}(p) \leq \int f d\text{Po}(p)$. To prove this, observe that $\text{Be}(p)$ precedes $\text{Be}(p/2) * \text{Be}(p/2)$ and that our relation is invariant under convolution; then apply the Poisson law of small numbers. Applying again the invariance under convolution, it follows that S precedes $\text{Po}(\mathbf{E} S)$. Thus, if f is convex and increasing, we have $\mathbf{E} f(S) \leq \int f d\text{Po}(\Delta)$. Hence,

$$\begin{aligned}
\mathbf{E} e^{tS} &\leq \sum_{k=0}^{\infty} e^{tk} \frac{\Delta^k e^{-\Delta}}{k!} = e^{\Delta(e^t-1)} \\
\mathbf{E} e^{tS} S &\leq \sum_{k=0}^{\infty} k e^{tk} \frac{\Delta^k e^{-\Delta}}{k!} = \Delta e^{t+\Delta(e^t-1)}
\end{aligned} \tag{3.59}$$

Now since $t \leq 1/(10 + 10\Delta) \leq 1/10$ and since the function $x \mapsto (e^x - 1)/x$ is increasing for $x > 0$, we have:

$$e^t - 1 = t \frac{e^t - 1}{t} \leq \frac{1}{10 + 10\Delta} \frac{e^{1/10} - 1}{1/10} = \frac{e^{1/10} - 1}{1 + \Delta} \quad (3.60)$$

implying:

$$\begin{aligned} \Delta(e^t - 1) &\leq e^{1/10} - 1 \\ t + \Delta(e^t - 1) &\leq \frac{1}{10 + 10\Delta} \frac{e^{1/10} - 1}{1/10} + \Delta \frac{e^{1/10} - 1}{1 + \Delta} = e^{1/10} - 1 \end{aligned} \quad (3.61)$$

This completes the proof. \square

4 A bound on the Stein expectation

The aim of this section is to prove an auxiliary result which handles the dependence structure and conditional boundedness described in Section 2. This is one of 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 $\mathbf{E} f(W) - \mathbf{N}(0, 1)\{f\}$ in terms of the *Stein expectation*:

$$\mathbf{E}[h'(W) - h(W)W] \quad (4.1)$$

where h solves the *Stein equation*:

$$h'(w) - h(w)w = f(w) - \mathbf{N}(0, 1)\{f\} \quad (4.2)$$

and where $\mathbf{N}(\mu, \sigma)\{f\}$ denotes the expectation of f with respect to the appropriate normal density. The Stein expectation can usually be estimated by means of Taylor's expansion. Taking the first three moments into consideration, terms with h and h' cancel and the Stein expectation is expressed in terms of h'' , provided that h' is absolutely continuous. It is well-known (see Section 5) that h' is AC provided that f is AC and that h'' behaves similarly as f' . Thus if f is Lipschitz, the Stein expectation can simply be estimated in terms of the Lipschitz constant of f . However, in our case, the test functions are even not continuous, let alone Lipschitz. Although, say, the indicators of half-lines can be approximated in terms of Lipschitz functions,

the Lipschitz constant in the optimal approximation still remains relatively high, leading to crude bounds with incorrect rate of convergence (typically $O(n^{-1/4})$ instead of $O(n^{-1/2})$).

Thus, in our case, estimating the Stein expectation in terms of the supremum norm of h'' does not work and (4.1) must be examined more carefully. One possible approach is the *bootstrapping argument*, where we estimate (4.1) in terms of $\mathbf{E} |h''(W+t)|$, $t \in \mathbb{R}$, take advantage of the fact that $N(t, 1)\{|h''|\}$ remains bounded and refer back to the fact that W is approximately normal. This approach has been used by several authors in cases where the random variables are bounded (see Barbour [2], Rinott [27], Rinott and Rotar [28], Dembo and Rinott [13] and Rinott and Rotar [29]). The version described below also allows for certain unbounded components, which must, however, be bounded by independent random quantities. The following assertion states it precisely:

Lemma 4.1. *Let W be a random variable decomposed as in (2.1)–(2.12) and let $h: \mathbb{R} \rightarrow \mathbb{R}$ be a differentiable function with absolutely continuous derivative. Suppose that for some version of h'' ,*

$$\mathbf{E} |h''(W+t)| \leq \rho(|t|) \tag{4.3}$$

where $\rho: [0, \infty) \rightarrow [0, \infty)$ is a non-decreasing function. Then the Stein expectation satisfies:

$$\begin{aligned} \mathbf{E}[h'(W) - h(W)W] &\leq \sum_{i \in I} \sum_{k \in K_i} \mathbf{E} \left\{ |X_i Z_{ik}| \eta_{ik} \left[\rho(U_{ik}^* + \tilde{U}_{ik}^*)(U_{ik}^* + \tilde{U}_{ik}^*) + \right. \right. \\ &\quad \left. \left. + \rho(\max\{U_{ik}, V_{ik}\} + U_{ik}^*) \left(\frac{1}{2}(U_{ik} + V_{ik}) + U_{ik}^* \right) \right] \right\} \end{aligned} \tag{4.4}$$

In particular, if $\rho(t) = e^{\lambda t}$ for some $\lambda \geq 0$, we have:

$$\mathbf{E}[h'(W) - h(W)W] \leq \beta(\lambda) \tag{4.5}$$

where $\beta(\lambda)$ is as in (2.14).

One of the disadvantages of the bootstrapping argument is that it yields much too large constants (consider (3.6)). Moreover, it does not work for several important cases where the random variables are not bounded (e. g.

locally dependent random variables with finite third moments). This goes on account of estimating (4.1) in terms of $\mathbf{E} |h''(W+t)|$. In fact, in view of (4.12), it would be more natural to estimate *conditional* expectations of quantities related to W . This leads to a rather complicated *inductive* argument, which has been used by Bolthausen [5] to derive a CLT for linear rank statistics and by Götze [17] to derive a multivariate CLT for independent random vectors. For the general case, however, this argument would be rather technical and the dependence structure from Section 2 would have to be replaced by a more complicated one, making possible application of the result rather awkward. Lemma 4.1 provides a tool for estimating the Stein expectation, which is easy to use, applicable in many interesting cases and provides bounds of the correct order.

Before proving Lemma 4.1, we shall prove the following assertion, which is the key step in proving (4.4)–(4.5).

Lemma 4.2. *Let $f: \mathbb{R} \rightarrow [0, \infty)$ be a measurable function, Lebesgue integrable on finite intervals. Suppose that W is a random variable, such that for all $t \in \mathbb{R}$,*

$$\mathbf{E} f(W+t) \leq \rho(|t|) \tag{4.6}$$

where $\rho: [0, \infty) \rightarrow [0, \infty)$ is a non-decreasing function. In addition, suppose that there are constants $U, U' \geq 0$ and random variables W', W_0 and \tilde{U} , such that \tilde{U} is independent of W_0 and:

$$|W - W_0| \leq \tilde{U}, \quad |W' - W_0| \leq U', \quad \mathbf{P}(\tilde{U} \leq U) > 0 \tag{4.7}$$

Then the following inequality holds:

$$\left| \mathbf{E} \int_W^{W'} f(s) ds \right| \leq \frac{\mathbf{E}[\rho(U + \tilde{U})(U + \tilde{U})] + \rho(U + U')(U + U')}{P(\tilde{U} \leq U)} \tag{4.8}$$

Proof. Using independence, observe that for any $A, B \geq 0$,

$$\begin{aligned}
\mathbf{E} \int_{W_0-A}^{W_0+B} f(s) ds &= \frac{1}{\mathbf{P}(\tilde{U} \leq U)} \mathbf{E} \int_{W_0-A}^{W_0+B} f(s) ds \mathbf{1}(\tilde{U} \leq U) \leq \\
&\leq \frac{1}{\mathbf{P}(\tilde{U} \leq U)} \mathbf{E} \int_{W-A-U}^{W+B+U} f(s) ds = \\
&= \frac{1}{\mathbf{P}(\tilde{U} \leq U)} \left[\int_{-A-U}^0 \mathbf{E} f(W+t) dt + \int_0^{B+U} \mathbf{E} f(W+t) dt \right] \leq \\
&\leq \frac{\rho(A+U)(A+U) + \rho(B+U)(B+U)}{\mathbf{P}(\tilde{U} \leq U)}
\end{aligned} \tag{4.9}$$

Observe that (4.9) remains true if A or B is replaced by a random variable independent of W_0 and the expectation is taken on the r. h. s. Hence,

$$\begin{aligned}
\mathbf{E} \int_W^{W'} f(s) ds &\leq \mathbf{E} \int_{W_0-\tilde{U}}^{W_0+U'} f(s) ds \leq \\
&\leq \frac{\mathbf{E} \rho(U+\tilde{U})(U+\tilde{U}) + \rho(U+U')(U+U')}{\mathbf{P}(\tilde{U} \leq U)}
\end{aligned} \tag{4.10}$$

Since the same estimate also holds for $\mathbf{E} \int_{W'}^W f(s) ds$, the proof is complete. \square

Proof of Lemma 4.1. Observe that:

$$\begin{aligned}
\mathbf{E}[h'(W) - h(W)W] &= \sum_{i \in I} \mathbf{E}[h'(W) \mathbf{E} X_i W - h(W)X_i] = \\
&= \sum_{i \in I} \mathbf{E}[h'(W) \mathbf{E} X_i W_i + h'(W) \mathbf{E} X_i Z_i - \\
&\quad - h(W_i)X_i - h'(W_i + \theta Z_i)X_i Z_i]
\end{aligned} \tag{4.11}$$

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:

$$\begin{aligned}
\mathbf{E}[h'(W) - h(W)W] &= \sum_{i \in I} \sum_{k \in K_i} \mathbf{E}[h'(W) \mathbf{E} X_i Z_{ik} - h'(W_i + \theta Z_i) X_i Z_{ik}] = \\
&= \sum_{i \in I} \sum_{k \in K_i} \mathbf{E}[h'(W_{ik}^*) - h'(W_i + \theta Z_i)] X_i Z_{ik} = \\
&= - \sum_{i \in I} \sum_{k \in K_i} \mathbf{E} \int_{W_{ik}^*}^{W_i + \theta Z_i} h''(s) ds X_i Z_{ik}
\end{aligned} \tag{4.12}$$

The proof is now completed by making use of the conditional version of Lemma 4.2 given \mathcal{H}_{ik} and θ , with W_{ik}^* in place of W , W_{ik} in place of W_0 and with $W_i + \theta Z_i = (1 - \theta)W_i + \theta W$ in place of W' . \square

5 Proofs of main results

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

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

where $\mathbf{N}(\mu, \sigma)\{f\}$ denotes the expectation of f with respect to the appropriate normal density. We introduce the *Mills ratio*, which will be defined in the following way:

$$\psi(x) := \sqrt{2\pi} e^{\frac{1}{2}x^2} \Phi(x) \tag{5.2}$$

Notice that the Mills ratio satisfies:

$$\psi'(x) - \psi(x)x = 1 \tag{5.3}$$

Lemma 5.1. *The Mills ratio as given in (5.2) and all its derivatives are strictly positive functions. Moreover,*

$$\Phi(-x)\psi'(x) + \Phi(x)\psi'(-x) = 1 \tag{5.4}$$

$$\int_x^\infty \Phi(-t) dt \psi''(x) + \int_{-\infty}^x \Phi(t) dt \psi''(-x) = 1 \tag{5.5}$$

Proof. Write:

$$\psi(x) = e^{\frac{1}{2}x^2} \int_{-\infty}^x e^{-\frac{1}{2}z^2} dz = \int_0^\infty e^{tx - \frac{1}{2}t^2} dt \tag{5.6}$$

Differentiation under the integral sign yields:

$$\psi^{(r)}(x) = \int_0^\infty t^r e^{tx - \frac{1}{2}t^2} dt > 0 \quad (5.7)$$

The equalities (5.4) can be proved by making use of (5.3) and (5.2) in turn and by straightforward calculation. \square

Now we turn to the study the solutions of the Stein equation (5.1) together with their derivatives. For a measurable function $f: \mathbb{R} \rightarrow \mathbb{R}$, denote by $\|f\|_\infty$ its supremum norm, by $\|f\|_1$ its L^1 -norm, by $V(f)$ its total variation and:

$$\|f\|_0 := \sup_{x \in \mathbb{R}} f(x) - \inf_{x \in \mathbb{R}} f(x) \quad (5.8)$$

We shall use the Lebesgue theory of differentiation (see Rudin [33]). For an absolutely continuous function f , the notation f' will be used for any function equal to the derivative of f at almost every point where f is differentiable. We shall therefore speak about *versions* of f' .

Lemma 5.2. *For every bounded measurable function f , there exist a unique absolutely continuous bounded function h solving the Stein equation (5.1) (for some version of h'). Denoting by h' the version that satisfies (5.1), h satisfies:*

$$\|h\|_\infty \leq \|f\|_1 \quad (5.9)$$

$$\|h'\|_\infty \leq \|f\|_0 \quad (5.10)$$

$$V(h') \leq 2V(f) \quad (5.11)$$

Moreover, if f is absolutely continuous, h' is also absolutely continuous and for every version of f' , there exists a version of h'' satisfying:

$$\|h''\|_\infty \leq \|f'\|_0 \quad (5.12)$$

$$V(h'') \leq 2V(f') \quad (5.13)$$

Proof of Lemma 5.2. Since any two solutions of (5.1) differ by a multiply of ψ , which is unbounded, the bounded solution, if any, is unique. Define:

$$h(w) = \frac{1}{\sqrt{2\pi}} \psi(-w) \int_{-\infty}^w f(x) e^{-\frac{1}{2}x^2} dx - \frac{1}{\sqrt{2\pi}} \psi(w) \int_w^\infty f(x) e^{-\frac{1}{2}x^2} dx \quad (5.14)$$

Since ψ is increasing, we can estimate:

$$|h(w)| \leq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^w |f(x)| \psi(-x) e^{-\frac{1}{2}x^2} dx + \frac{1}{\sqrt{2\pi}} \int_w^{\infty} |f(x)| \psi(x) e^{-\frac{1}{2}x^2} dx \quad (5.15)$$

Now since $\psi(-x)e^{-\frac{1}{2}x^2}/\sqrt{2\pi} = \Phi(-x) \leq 1$ and $\psi(x)e^{-\frac{1}{2}x^2}/\sqrt{2\pi} = \Phi(x) \leq 1$, (5.9) obviously follows.

Differentiating (5.14), we find that h is absolutely continuous and that there is a version of h' satisfying:

$$\begin{aligned} h'(w) = f(w) - \frac{1}{\sqrt{2\pi}} \psi'(-w) \int_{-\infty}^w f(x) e^{-\frac{1}{2}x^2} dx - \\ - \frac{1}{\sqrt{2\pi}} \psi'(w) \int_w^{\infty} f(x) e^{-\frac{1}{2}x^2} dx \end{aligned} \quad (5.16)$$

Combining (5.3), (5.14) and (5.16), we conclude that h together with h' satisfies (5.1). Moreover, (5.16) together with (5.4) implies (5.10).

To prove (5.11), we first integrate (5.16) by parts and apply (5.4), leading to:

$$h'(w) = \psi'(-w) \int_{-\infty}^w \Phi(x) df(x) - \psi'(w) \int_w^{\infty} \Phi(-x) df(x) \quad (5.17)$$

Thus, we can write $h'(w)$ as the limit of the Riemann–Stieltjes sums:

$$\sum_{i=1}^n G(w, \xi_i) (f(x_i) - f(x_{i-1})) \quad (5.18)$$

where:

$$G(w, x) = \begin{cases} -\psi'(w) \Phi(-x) & ; w \leq x \\ \psi'(-w) \Phi(x) & ; w > x \end{cases} \quad (5.19)$$

and where $x_0 < \xi_1 < x_1 < \xi_2 < \dots < \xi_n < x_n$ and $w \in \{x_0, \dots, x_n\}$ (notice that $h'(w) = \int_{-\infty}^{\infty} G(w, x) df(x)$ at any w where f is continuous). Using (5.4) and the fact that ψ' is increasing, we find that the total variation of the function $w \mapsto G(w, x)$ equals 2 for all x . The estimate (5.11) now follows.

Now suppose that f is absolutely continuous. In this case, we can write (5.17) in terms of the usual Riemann integral:

$$h'(w) = \psi'(w) \int_{-\infty}^w f'(x) \Phi(x) dx - \psi'(w) \int_w^{\infty} f'(x) \Phi(-x) dx \quad (5.20)$$

Similarly as before, we find that h' is also absolutely continuous. Making use of (5.4), we find that in every Lebesgue point of f' ,

$$h''(w) = f'(w) - \psi''(-w) \int_{-\infty}^w f'(x) \Phi(x) dx - \psi''(w) \int_w^{\infty} f'(x) \Phi(-x) dx \quad (5.21)$$

Using (5.4), we obtain (5.12).

Finally, to prove (5.13), assume without loss of generality that $V(f') < \infty$. Integrating by parts and making use of (5.5), we obtain:

$$h''(w) = \psi''(-w) \int_{-\infty}^w \int_{-\infty}^x \Phi(t) dt df'(x) - \psi''(w) \int_w^{\infty} \int_x^{\infty} \Phi(-t) dt df'(x) \quad (5.22)$$

Again, we can write $h''(w)$ as the limit of the Riemann–Stieltjes sums:

$$\sum_{i=1}^n H(w, \xi_i) (f(x_i) - f(x_{i-1})) \quad (5.23)$$

where:

$$H(w, x) = \begin{cases} -\psi''(w) \int_x^{\infty} \Phi(-t) dt & ; w \leq x \\ \psi''(-w) \int_{-\infty}^x \Phi(t) dt & ; w > x \end{cases} \quad (5.24)$$

and where $x_0 < \xi_1 < x_1 < \xi_2 < \dots < \xi_n < x_n$ and $w \in \{x_0, \dots, x_n\}$. Using (5.5) and the monotonicity of ψ'' , we find again that the total variation of the function $w \mapsto H(w, x)$ equals 2 for all x and (5.13) is proved. \square

Lemma 5.2 allows us to take control over h'' provided that we have control over f' . Since the usual test functions (e. g. indicators of half-lines) are not AC, we shall need to approximate them by suitable smooth functions. For a non-decreasing function f and $\varepsilon > 0$, define:

$$M_{\varepsilon}^{+}(f)(w) := \frac{1}{\varepsilon} \int_w^{w+\varepsilon} f(x) dx \quad M_{\varepsilon}^{-}(f)(w) := \frac{1}{\varepsilon} \int_{w-\varepsilon}^w f(x) dx \quad (5.25)$$

For f with $V(f) < \infty$, we can write $f = f_{\uparrow} - f_{\downarrow}$, where f_{\uparrow} and f_{\downarrow} are non-decreasing functions with $V(f) = V(f_{\uparrow}) + V(f_{\downarrow})$. Now define:

$$M_{\varepsilon}^{+}(f) := M_{\varepsilon}^{+}(f_{\uparrow}) - M_{\varepsilon}^{-}(f_{\downarrow}) \quad M_{\varepsilon}^{-}(f) := M_{\varepsilon}^{-}(f_{\uparrow}) - M_{\varepsilon}^{+}(f_{\downarrow}) \quad (5.26)$$

and notice that the definition is independent of the decomposition.

Lemma 5.3. For every function $f: \mathbb{R} \rightarrow \mathbb{R}$ with $V(f) < \infty$ and every $\varepsilon > 0$, the functions $M_\varepsilon^+(f)$ and $M_\varepsilon^-(f)$ satisfy:

$$M_\varepsilon^-(f) \leq f \leq M_\varepsilon^+(f) \quad (5.27)$$

and:

$$V(M_\varepsilon^+(f)) \leq V(f), \quad V(M_\varepsilon^-(f)) \leq V(f) \quad (5.28)$$

$$\|M_\varepsilon^+(f) - f\|_1 \leq \frac{\varepsilon}{2} V(f), \quad \|M_\varepsilon^-(f) - f\|_1 \leq \frac{\varepsilon}{2} V(f) \quad (5.29)$$

Moreover, $M_\varepsilon^+(f)$ and $M_\varepsilon^-(f)$ are absolutely continuous and there are versions of $(M_\varepsilon^+(f))'$ and $(M_\varepsilon^-(f))'$ satisfying:

$$V((M_\varepsilon^+(f))') \leq \frac{2V(f)}{\varepsilon}, \quad V((M_\varepsilon^-(f))') \leq \frac{2V(f)}{\varepsilon} \quad (5.30)$$

Proof. Observe that it suffices to prove (5.27)–(5.30) for non-decreasing functions. The estimate (5.27) is evident. Taking $w \mapsto -f(-w)$ in place of f , observe that it suffices to prove (5.28), (5.29) and (5.30) for $M_\varepsilon^+(f)$. To prove (5.28), observe that if f is non-decreasing, we have:

$$\begin{aligned} V(M_\varepsilon^+(f)) &= \lim_{w \rightarrow \infty} M_\varepsilon^+(f)(w) - \lim_{w \rightarrow -\infty} M_\varepsilon^+(f)(w) = \\ &= \lim_{w \rightarrow \infty} f(w) - \lim_{w \rightarrow -\infty} f(w) = \\ &= V(f) \end{aligned} \quad (5.31)$$

To prove (5.29), observe that by Fubini's theorem:

$$\begin{aligned} \|M_\varepsilon^+(f) - f\|_1 &= \int_{-\infty}^{\infty} M_\varepsilon^+(f)(w) - f(w) \, dw = \\ &= \frac{1}{\varepsilon} \int_{-\infty}^{\infty} \int_w^{w+\varepsilon} (f(x) - f(w)) \, dx \, dw = \\ &= \frac{1}{\varepsilon} \int_0^\varepsilon \int_{-\infty}^{\infty} (f(w+t) - f(w)) \, dw \, dt = \\ &= \frac{1}{\varepsilon} \int_0^\varepsilon \int_{-\infty}^{\infty} \int_w^{w+t} df(x) \, dw \, dt = \\ &= \frac{1}{\varepsilon} \int_0^\varepsilon \int_{-\infty}^{\infty} \int_{x-t}^x dw \, df(x) \, dt = \\ &= \frac{\varepsilon}{2} V(f) \end{aligned} \quad (5.32)$$

Finally, observe that $M_\varepsilon^+(f)$ is a primitive function of the function $w \mapsto (f(w + \varepsilon) - f(w))/\varepsilon$ and is therefore absolutely continuous. The estimate (5.30) also obviously follows. \square

The following assertion allows us to express the solutions of (1.9) 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 easy and is therefore left without proof.

Lemma 5.4. *Suppose that h solves:*

$$h'(w) - h(w)w = f(w + \lambda) - \mathbf{N}(\lambda, 1)\{f\} \quad (5.33)$$

Then the function:

$$\tilde{h}(w) := e^{\lambda w} h(w - \lambda) \quad (5.34)$$

solves:

$$\tilde{h}'(w) - \tilde{h}(w)w = e^{\lambda w} \left(f(w) - \mathbf{N}(\lambda, 1)\{f\} \right) \quad (5.35)$$

Now we turn to the proofs of Theorems 2.1–2.4. We shall derive them from a more general result, estimating the expectations of suitable test functions. Fix a random variable W decomposed as in Section 2. For every bounded measurable function $f: \mathbb{R} \rightarrow \mathbb{R}$ and each $\lambda \in \mathbb{R}$ with $\mathbf{E} e^{\lambda W} < \infty$, define:

$$\delta_\lambda(f) := \left| \mathbf{E} e^{\lambda W} \left[f(W) - \mathbf{N}(\lambda, 1)\{f\} \right] \right| \quad (5.36)$$

The following result is the key to the proofs of Theorems 2.1–2.4.

Lemma 5.5. *For each f with finite $\|f\|_1$ and $V(f)$ and each $\lambda \in \mathbb{R}$ with $\mathbf{E} e^{\lambda W} < \infty$, we have:*

$$\delta_\lambda(f) \leq \left[(4.66 + 4.73|\lambda|)V(f) + \lambda^2\|f\|_1 \right] \mathbf{E} e^{\lambda W} \beta(|\lambda|) \quad (5.37)$$

where β is as in (2.14).

Before proving Lemma 5.5, we prove some additional auxiliary results and introduce some new quantities. For a function f with $0 < V(f) < \infty$ and $\|f\|_1 < \infty$, and constants $\lambda \in \mathbb{R}$, $a > 0$ and $b > 0$, such that $\mathbf{E} e^{\lambda W} < \infty$, define:

$$\delta_{\lambda,a,b}(f) := \frac{\delta_\lambda(f)}{\left[(1 + a|\lambda|)V(f) + b\lambda^2\|f\|_1 \right] \mathbf{E} e^{\lambda W}} \quad (5.38)$$

Next, define:

$$\delta^* := \sup_{0 < V(f) < \infty} \frac{\delta_0(f)}{V(f)}, \quad \delta_{\lambda,a,b}^* := \sup_{\substack{0 < V(f) < \infty \\ \|f\|_1 < \infty}} \delta_{\lambda,a,b}(f) \quad (5.39)$$

Notice that $\delta^* = \delta_{0,a,b}^*$ and that $\delta_{\lambda,a,b}^*$ is finite, because for every bounded f , we trivially have $\delta_\lambda(f) \leq \|f\|_0 \mathbf{E} e^{\lambda W}$; if, in addition, $V(f) < \infty$ and $\|f\|_1 < \infty$, we have $\lim_{x \rightarrow -\infty} f(x) = \lim_{x \rightarrow \infty} f(x) = 0$, so that we can estimate:

$$\delta_\lambda(f) \leq \frac{1}{2} V(f) \mathbf{E} e^{\lambda W} \quad (5.40)$$

and consequently,

$$\delta_{\lambda,a,b}^* \leq \frac{1}{2(1+a|\lambda|)} \quad (5.41)$$

Lemma 5.6. *Suppose that there is a function $f = f_1 + f_2: \mathbb{R} \rightarrow \mathbb{R}$, such that $V(f_1)$ and $V(f_2)$ are finite and f_1 and f_2 are absolutely continuous. Then for arbitrary versions of f'_1 and f'_2 , we have:*

$$\delta_0(f) \leq \beta(0) \left[\|f'_1\|_0 + \frac{2V(f_2)}{\sqrt{2\pi}} + 2\delta^* V(f'_2) \right] \quad (5.42)$$

Moreover, for $\lambda \in \mathbb{R}$, $a, b > 0$ and $\|f\|_1 < \infty$, we have:

$$\begin{aligned} \delta_\lambda(f) \leq \mathbf{E} e^{\lambda W} \beta(|\lambda|) & \left[\lambda^2 \|f\|_1 + |\lambda| V(f) + \|f'_1\|_0 + \frac{2V(f_2)}{\sqrt{2\pi}} \right] + \\ & + 2 \mathbf{E} e^{\lambda W} \delta_{\lambda,a,b}^* \beta(|\lambda|) \left[(1+a|\lambda|) V(f'_2) + b\lambda^2 V(f_2) \right] \end{aligned} \quad (5.43)$$

Proof. As (5.42) is essentially a special case of (5.43), we shall only derive the latter estimate. Let h be the unique solution of (5.33) and let \tilde{h} be as in (5.34). By Lemma 5.4, there is a version of \tilde{h}' , such that (5.35) holds, so that:

$$\mathbf{E} e^{\lambda W} [f(W) - \mathbf{N}(\lambda, 1)\{f\}] = \mathbf{E} [\tilde{h}'(W) - \tilde{h}(W)W] \quad (5.44)$$

The key tool for estimating the r. h. s. will be Lemma 4.1, so that it suffices to estimate $\mathbf{E} |\tilde{h}''(W+t)|$ for all $t \in \mathbb{R}$. Using (5.34), (5.9) and (5.10) in turn and making use of the fact that $\|f\|_0 \leq \frac{1}{2} V(f)$ if $\|f\|_1 < \infty$, we obtain:

$$\begin{aligned} \mathbf{E} |\tilde{h}''(W+t)| & \leq \mathbf{E} e^{\lambda(W+t)} \left[\lambda^2 |h(W-\lambda+t)| + 2|\lambda| |h'(W-\lambda+t)| + \right. \\ & \quad \left. + |h''(W-\lambda+t)| \right] \leq \\ & \leq e^{|\lambda t|} \left[\lambda^2 \mathbf{E} e^{\lambda W} \|f\|_1 + |\lambda| \mathbf{E} e^{\lambda W} V(f) + \mathbf{E} e^{\lambda W} |h''(W-\lambda+t)| \right] \end{aligned} \quad (5.45)$$

Now it only remains to estimate the last term. Let h_1 and h_2 be the corresponding solutions of the Stein equation (5.33) for f_1 and f_2 . Clearly, $h = h_1 + h_2$. Notice that by Lemma 5.2, h'_1 and h'_2 are absolutely continuous and there exist versions of h''_1 and h''_2 satisfying (5.12) and (5.13). For h''_1 , we use (5.12) to estimate:

$$\mathbf{E} e^{\lambda W} |h''_1(W + t)| \leq \mathbf{E} e^{\lambda W} \|f'_1\|_0 \quad (5.46)$$

For h''_2 , we shall apply the ‘bootstrapping’ argument, i. e. we shall refer back to the approximate normality of W . The expectation with respect to the standard normal distribution can be estimated as follows, using (5.11):

$$\mathbf{N}(t, 1)\{|h''_2|\} \leq \frac{1}{\sqrt{2\pi}} \|h''_2\|_1 = \frac{1}{\sqrt{2\pi}} V(h'_2) \leq \frac{2}{\sqrt{2\pi}} V(f_2) \quad (5.47)$$

To estimate the difference, we refer to (5.39) and use (5.11) and (5.13), leading to:

$$\begin{aligned} \mathbf{E} e^{\lambda W} \left(|h''_2(W - \lambda + t)| - \mathbf{N}(t, 1)\{|h''_2|\} \right) &\leq \\ &\leq \mathbf{E} e^{\lambda W} \delta_{\lambda, a, b}^* \left[(1 + a|\lambda|)V(|h''_2|) + b\lambda^2 \|h''_2\|_1 \right] \leq \\ &\leq \mathbf{E} e^{\lambda W} \delta_{\lambda, a, b}^* \left[(1 + a|\lambda|)V(h''_2) + b\lambda^2 V(h'_2) \right] \leq \\ &\leq 2 \mathbf{E} e^{\lambda W} \delta_{\lambda, a, b}^* \left[(1 + a|\lambda|)V(f'_2) + b\lambda^2 V(f_2) \right] \end{aligned} \quad (5.48)$$

Collecting (5.45), (5.46), (5.47) and (5.48), we obtain:

$$\begin{aligned} |\tilde{h}''_\lambda(W + t)| &\leq e^{|\lambda t|} \mathbf{E} e^{\lambda W} \left[\lambda^2 \|f\|_1 + |\lambda|V(f) + \|f'_1\|_0 + \frac{2V(f_2)}{\sqrt{2\pi}} \right] + \\ &\quad + 2e^{|\lambda t|} \mathbf{E} e^{\lambda W} \delta_{\lambda, a, b}^* \left[(1 + a|\lambda|)V(f'_2) + b\lambda^2 V(f_2) \right] \end{aligned} \quad (5.49)$$

Now use Lemma 4.1 and the result follows. \square

Lemma 5.7. *For bounded real functions $f^- \leq f \leq f^+$, we have:*

$$\begin{aligned} \delta_\lambda(f) &\leq \max \left\{ \delta_\lambda(f^+) + \frac{1}{\sqrt{2\pi}} \|f^+ - f\|_1 \mathbf{E} e^{\lambda W}, \right. \\ &\quad \left. \delta_\lambda(f^-) + \frac{1}{\sqrt{2\pi}} \|f^- - f\|_1 \mathbf{E} e^{\lambda W} \right\} \end{aligned} \quad (5.50)$$

Proof. Observe that:

$$\begin{aligned}
\mathbf{E} e^{\lambda W} (f(W) - \mathbf{N}(\lambda, 1)\{f\}) &\leq \mathbf{E} e^{\lambda W} (f^+(W) - \mathbf{N}(\lambda, 1)\{f\}) \leq \\
&\leq \mathbf{E} e^{\lambda W} (f^+(W) - \mathbf{N}(\lambda, 1)\{f^+\}) + \mathbf{N}(\lambda, 1)\{f^+ - f\|_1\} \mathbf{E} e^{\lambda W} \leq \\
&\leq \delta_\lambda(f^+) + \frac{1}{\sqrt{2\pi}} \|f^+ - f\|_1 \mathbf{E} e^{\lambda W}
\end{aligned} \tag{5.51}$$

Similarly, one can derive:

$$\mathbf{E} e^{\lambda W} (f(W) - \mathbf{N}(\lambda, 1)\{f\}) \geq -\delta_\lambda(f^-) - \frac{1}{\sqrt{2\pi}} \|f^- - f\|_1 \mathbf{E} e^{\lambda W} \tag{5.52}$$

This completes the proof. \square

Proof of Lemma 5.5. For $\beta(|\lambda|) = 0$, the result follows immediately from Lemma 5.6. Suppose that $\beta(|\lambda|) > 0$ and take arbitrary $a, b, \varepsilon > 0$. Approximating f by $M_\varepsilon^+(f)$ and $M_\varepsilon^-(f)$ and combining Lemmas 5.3, 5.6 (where taking $f_1 = 0$) and 5.7, we find after some calculation that:

$$\begin{aligned}
\delta_\lambda(f) &\leq V(f) \mathbf{E} e^{\lambda W} \beta(|\lambda|) \left[\frac{\varepsilon}{2\sqrt{2\pi} \beta(|\lambda|)} + \frac{2}{\sqrt{2\pi}} + \frac{4\delta_{\lambda,a,b}^*}{\varepsilon} \right] + \\
&\quad + |\lambda| V(f) \mathbf{E} e^{\lambda W} \beta(|\lambda|) \left[1 + \frac{4a\delta_{\lambda,a,b}^*}{\varepsilon} + \frac{\varepsilon|\lambda|}{2} + 2|\lambda|b\delta_{\lambda,a,b}^* \right] + \\
&\quad + \lambda^2 \|f\|_1 \mathbf{E} e^{\lambda W} \beta(|\lambda|)
\end{aligned} \tag{5.53}$$

Now we apply (5.41) to estimate $2|\lambda|b\delta_{\lambda,a,b}^* \leq b/a$. Next, recalling (5.40) and applying the inequality $\min\{x, y + z\} \leq \sqrt{xy} + z$ for $x = \frac{1}{2}V(f) \mathbf{E} e^{\lambda W}$ and $y = \frac{1}{2}\lambda^2 V(f) \mathbf{E} e^{\lambda W} \beta(|\lambda|)\varepsilon$, we find that:

$$\begin{aligned}
\delta_\lambda(f) &\leq V(f) \mathbf{E} e^{\lambda W} \beta(|\lambda|) \left[\frac{\varepsilon}{2\sqrt{2\pi} \beta(|\lambda|)} + \frac{2}{\sqrt{2\pi}} + \frac{4\delta_{\lambda,a,b}^*}{\varepsilon} \right] + \\
&\quad + |\lambda| V(f) \mathbf{E} e^{\lambda W} \beta(|\lambda|) \left[1 + \frac{4a\delta_{\lambda,a,b}^*}{\varepsilon} + \frac{1}{2} \sqrt{\frac{\varepsilon}{\beta(|\lambda|)}} + \frac{b}{a} \right] + \\
&\quad + \lambda^2 \|f\|_1 \mathbf{E} e^{\lambda W} \beta(|\lambda|)
\end{aligned} \tag{5.54}$$

Without loss of generality, we can assume that $V(f) > 0$. Dividing by $[(1 + a|\lambda|)V(f) + b\lambda^2] \mathbf{E} e^{\lambda W}$, applying the inequality:

$$\frac{\alpha_1 V(f) + \beta_1 a |\lambda| V(f) + \gamma_1 b \lambda^2 \|f\|_1}{\alpha_2 V(f) + \beta_2 a |\lambda| V(f) + \gamma_2 b \lambda^2 \|f\|_1} \leq \max \left\{ \frac{\alpha_1}{\alpha_2}, \frac{\beta_1}{\beta_2}, \frac{\gamma_1}{\gamma_2} \right\} \quad (5.55)$$

and taking supremum over f , we obtain:

$$\delta_{\lambda,a,b}^* \leq \beta(|\lambda|) \max \left\{ \frac{\varepsilon}{2\sqrt{2\pi} \beta(|\lambda|)} + \frac{2}{\sqrt{2\pi}} + \frac{4\delta_{\lambda,a,b}^*}{\varepsilon}, \right. \\ \left. \frac{1}{a} + \frac{4\delta_{\lambda,a,b}^*}{\varepsilon} + \frac{1}{2a} \sqrt{\frac{\varepsilon}{\beta(|\lambda|)}} + \frac{b}{a^2}, \frac{1}{b} \right\} \quad (5.56)$$

or, taking $K_{\lambda,a,b}^* := \delta_{\lambda,a,b}^*/\beta(|\lambda|)$ and $c := \varepsilon/\beta(|\lambda|)$:

$$K_{\lambda,a,b}^* \leq \max \left\{ \frac{c}{2\sqrt{2\pi}} + \frac{2}{\sqrt{2\pi}} + \frac{4K_{\lambda,a,b}^*}{c}, \frac{1}{a} + \frac{4K_{\lambda,a,b}^*}{c} + \frac{\sqrt{c}}{2a} + \frac{b}{a^2}, \frac{1}{b} \right\} \quad (5.57)$$

Now choose $c > 4$ so that the unique root k of the equation:

$$k = \frac{c}{2\sqrt{2\pi}} + \frac{2}{\sqrt{2\pi}} + \frac{4k}{c}, \quad (5.58)$$

is minimal, that is:

$$c = 4(1 + \sqrt{2}), \quad k = \frac{2(3 + 2\sqrt{2})}{\sqrt{2\pi}} \quad (5.59)$$

Furthermore, choose a and b so that:

$$k = \frac{1}{a} + \frac{4k}{c} + \frac{\sqrt{c}}{2a} + \frac{b}{a^2} = \frac{1}{b} \quad (5.60)$$

Combining (5.57)–(5.60), we now find that $K_{\lambda,a,b}^* \leq k$, or equivalently:

$$\delta_{\lambda,a,b}^* \leq k\beta(|\lambda|) \quad (5.61)$$

As one can easily check that $k < 4.66$, $ka < 4.73$ and $kb = 1$, the proof is complete. \square

Now let $\lambda > 0$ and $x \in \mathbb{R}$. Taking:

$$f_{x,\lambda}(w) := e^{-\lambda w} \mathbf{1}[w \geq x] \quad (5.62)$$

noting that:

$$\mathbf{N}(\lambda, 1)\{f_{x,\lambda}\} = e^{-\frac{1}{2}\lambda^2} \Phi(-x), \quad V(f_{x,\lambda}) = 2e^{-\lambda x}, \quad \|f_{x,\lambda}\|_1 = \frac{e^{-\lambda x}}{\lambda} \quad (5.63)$$

and applying Lemma 5.5, we find that:

$$\left| \mathbf{P}(W \geq x) - e^{-\frac{1}{2}\lambda^2} \mathbf{E} e^{\lambda W} \Phi(-x) \right| \leq (9.32 + 10.46\lambda) e^{-\lambda x} \mathbf{E} e^{\lambda W} \beta(\lambda) \quad (5.64)$$

In the sequel, the constants will be substantially improved by taking different smoothing, which smooths only the non-smooth part of $f_{x,\lambda}$.

Lemma 5.8. *For all $x \in \mathbb{R}$ and $\lambda \geq 0$, such that $\mathbf{E} e^{\lambda W} < \infty$, we have:*

$$\left| \mathbf{P}(W \geq x) - e^{-\frac{1}{2}\lambda^2} \mathbf{E} e^{\lambda W} \Phi(-x) \right| \leq (4.66 + 8.58\lambda) e^{-\lambda x} \mathbf{E} e^{\lambda W} \beta(\lambda) \quad (5.65)$$

Proof. For every $\varepsilon > 0$, define:

$$f_{x,\lambda,\varepsilon}^+(w) := \begin{cases} 0 & ; w \leq x - \varepsilon \\ \frac{e^{-\lambda x}(w - x + \varepsilon)}{\varepsilon} & ; x - \varepsilon \leq w \leq x \\ e^{-\lambda w} & ; w \geq x \end{cases} \quad (5.66)$$

and:

$$f_{x,\lambda,\varepsilon}^-(w) := \begin{cases} 0 & ; w \leq x \\ \frac{e^{-\lambda(x+\varepsilon)}(w - x)}{\varepsilon} & ; x \leq w \leq x + \varepsilon \\ e^{-\lambda w} & ; w \geq x + \varepsilon \end{cases} \quad (5.67)$$

An easy calculation shows that $f_{x,\lambda,\varepsilon}^- \leq f_{x,\lambda} \leq f_{x,\lambda,\varepsilon}^+$ and:

$$\|f_{x,\lambda,\varepsilon}^+ - f_{x,\lambda}\|_1 = \frac{\varepsilon}{2} e^{-\lambda x}, \quad \|f_{x,\lambda,\varepsilon}^- - f_{x,\lambda}\|_1 \leq \frac{\varepsilon}{2} e^{-\lambda x} \quad (5.68)$$

$$V(f_{x,\lambda,\varepsilon}^+) = 2e^{-\lambda x}, \quad V(f_{x,\lambda,\varepsilon}^-) \leq 2e^{-\lambda x} \quad (5.69)$$

Furthermore, write $f_{x,\lambda,\varepsilon}^+ = f_{x,\lambda,\varepsilon,1}^+ + f_{x,\lambda,\varepsilon,2}^+$, where:

$$f_{x,\lambda,\varepsilon,1}^+(w) = (e^{-\lambda w} - e^{-\lambda x}) \mathbf{1}[w \geq x] \quad (5.70)$$

Noting that some version of $(f_{x,\lambda,\varepsilon,1}^+)'$ satisfies:

$$\|(f_{x,\lambda,\varepsilon,1}^+)'\|_0 = \lambda e^{-\lambda x} \quad (5.71)$$

and that:

$$V(f_{x,\lambda,\varepsilon,2}^+) = e^{-\lambda x} \quad \text{and} \quad V((f_{x,\lambda,\varepsilon,2}^+)') = \frac{2e^{-\lambda x}}{\varepsilon} \quad (5.72)$$

for some version of $(f_{x,\lambda,\varepsilon,2}^+)'$, and applying Lemma 5.6 together with (5.63), we find that for all $a, b > 0$:

$$\delta_\lambda(f_{x,\lambda,\varepsilon}^+) \leq e^{-\lambda x} \beta(\lambda) \mathbf{E} e^{\lambda W} \left[\frac{2}{\sqrt{2\pi}} + 4\lambda + \frac{\varepsilon}{2} \lambda^2 + \delta_{\lambda,a,b}^* \left(\frac{4(1+a\lambda)}{\varepsilon} + 2b\lambda^2 \right) \right] \quad (5.73)$$

Noting that (5.61) holds provided that a and b are as in (5.60), we have:

$$\begin{aligned} \delta_\lambda(f_{x,\lambda,\varepsilon}^+) &\leq e^{-\lambda x} \mathbf{E} e^{\lambda W} \left[\left(\frac{2}{\sqrt{2\pi}} + 4\lambda + \frac{\varepsilon}{2} \lambda^2 \right) \beta(\lambda) + \right. \\ &\quad \left. + \left(\frac{18.61(1+1.016\lambda)}{\varepsilon} + 2\lambda^2 \right) \beta(\lambda)^2 \right] \end{aligned} \quad (5.74)$$

Similarly, one can show that the same estimate also holds for $f_{x,\lambda,\varepsilon}^-$. By Lemma 5.7 and (5.68), we then have:

$$\begin{aligned} \delta_\lambda(f_{x,\lambda}) &\leq e^{-\lambda x} \mathbf{E} e^{\lambda W} \left[\left(\frac{2}{\sqrt{2\pi}} + 4\lambda + \frac{\varepsilon}{2} \lambda^2 \right) \beta(\lambda) + \right. \\ &\quad \left. + \left(\frac{18.64(1+1.016\lambda)}{\varepsilon} + 2\lambda^2 \right) \beta(\lambda)^2 + \frac{\varepsilon}{2\sqrt{2\pi}} \right] \end{aligned} \quad (5.75)$$

Optimizing in ε , we find that:

$$\begin{aligned} \delta_\lambda(f_{x,\lambda}) &\leq e^{-\lambda x} \mathbf{E} e^{\lambda W} \beta(\lambda) \left[\frac{2}{\sqrt{2\pi}} + 4\lambda + 2\lambda^2 \beta(\lambda) + \right. \\ &\quad \left. + 3.854 \sqrt{(1+1.016\lambda)(1+\sqrt{2\pi}\lambda^2\beta(\lambda))} \right] \end{aligned} \quad (5.76)$$

Making use of the inequality between the arithmetic and the geometric mean, we obtain:

$$\delta_\lambda(f_{x,\lambda}) \leq e^{-\lambda x} \mathbf{E} e^{\lambda W} \left[4.66\beta(\lambda) + 5.96\lambda\beta(\lambda) + 6.86\lambda^2\beta(\lambda)^2 \right] \quad (5.77)$$

The proof is now completed by recalling (5.40) and applying the inequality $\min\{1, x+y\} \leq \sqrt{x} + y$. \square

Lemma 5.9. *For every $\lambda \in \mathbb{R}$, we have:*

$$e^{\frac{1}{2}\lambda^2 - \beta^*(|\lambda|)} \leq \mathbf{E} e^{\lambda W} \leq e^{\frac{1}{2}\lambda^2 + \beta^*(|\lambda|)} \quad (5.78)$$

Proof. Define:

$$F(\lambda) = \ln \mathbf{E} e^{\lambda W} \quad (5.79)$$

and notice that:

$$F'(\lambda) = \frac{\mathbf{E} e^{\lambda W} W}{\mathbf{E} e^{\lambda W}} \quad (5.80)$$

so that:

$$F'(\lambda) - \lambda = -\frac{1}{\mathbf{E} e^{\lambda W}} \mathbf{E} \left[\chi'_\lambda(W) - \chi_\lambda(W) W \right] \quad (5.81)$$

where:

$$\chi_\lambda(x) := e^{\lambda x} \quad (5.82)$$

Again, we shall apply Lemma 4.1. Noting that:

$$\mathbf{E} \chi''_\lambda(W + t) = \lambda^2 \mathbf{E} e^{\lambda(W+t)} \leq \lambda^2 e^{|\lambda t|} \mathbf{E} e^{\lambda W} \quad (5.83)$$

it follows that:

$$|F'(\lambda) - \lambda| \leq \lambda^2 \beta(|\lambda|) \quad (5.84)$$

Now write:

$$F(\lambda) = \int_0^\lambda F'(t) dt = \frac{\lambda^2}{2} + \int_0^\lambda (F'(t) - t) dt \quad (5.85)$$

which together with (5.84) yields $|F(\lambda) - \frac{1}{2}\lambda^2| \leq \beta^*(|\lambda|)$. This completes the proof. \square

Proof of Theorems 2.1 and 2.2. First, observe that it suffices to approximate $\mathbf{P}(W \geq x)$; the approximations to $\mathbf{P}(W \leq -x)$ follow from the fact that $-W$ can be decomposed in the same way as W , in particular with the same $\beta(\lambda)$. The estimate (2.15) follows from Lemmas 5.8 and 5.9. To derive (2.17), use Lemma 5.9 and apply the Markov inequality. Thus Theorem 2.2 is proved. Theorem 2.1 follows immediately by setting $\lambda = 0$. \square

Proof of Theorem 2.3. The result follows from Theorem 2.2: to prove (2.18) and (2.19), choose $\lambda := x$; to prove (2.20), choose $\lambda := M$. \square

Proof of Theorem 2.4. Similarly as before, we shall use Theorem 2.2. Choosing $\lambda := x$ and noting that:

$$\beta^*(\lambda) \leq \frac{1}{3}\beta_1 e^{\beta_2 \lambda} \lambda^3 \quad \text{and} \quad \beta_1 e^{\beta_2 \lambda} \leq \frac{1}{2M} \quad (5.86)$$

proves (2.23) and (2.24). To prove (2.25), choose:

$$\lambda := M \left(1 + \frac{1}{3} \ln \frac{x}{M} \right) \quad (5.87)$$

and Theorem 2.2 yields:

$$\max\{\mathbf{P}(W \geq x), \mathbf{P}(W \leq -x)\} \leq \exp \left\{ M^2 F \left(\frac{x}{M} \right) \right\} \quad (5.88)$$

where:

$$F(y) = \frac{1}{6}y^{1/3} \left(1 + \frac{1}{3} \ln y \right)^3 + \frac{1}{2} \left(1 + \frac{1}{3} \ln y \right)^2 - y \left(1 + \frac{1}{3} \ln y \right) \quad (5.89)$$

In view of (2.25), it suffices to show that:

$$F(y) \leq -\frac{1}{3} - \frac{1}{3}y \ln y \quad (5.90)$$

for all for $y \geq 1$. Differentiating and noting that $F(1) = -\frac{1}{3}$, we find that it suffices to show that:

$$\frac{\left(1 + \frac{1}{3} \ln y \right)^3}{18y^{2/3}} + \frac{\left(1 + \frac{1}{3} \ln y \right)^2}{6y^{2/3}} + \frac{1 + \frac{1}{3} \ln y}{3y} \leq 1 \quad (5.91)$$

A straightforward calculation shows that:

$$\frac{\left(1 + \frac{1}{3} \ln y \right)^3}{18y^{2/3}} \leq \frac{3}{16e} < 0.07, \quad \frac{\left(1 + \frac{1}{3} \ln y \right)^2}{6y^{2/3}} \leq \frac{1}{6}, \quad \frac{1 + \frac{1}{3} \ln y}{3y} \leq \frac{1}{3} \quad (5.92)$$

for all $y \geq 1$. This completes the proof. \square

Acknowledgements

Part of this work was done while the author was visiting the Institute for Mathematical Sciences, National University of Singapore, in 2003. The author would like to thank the Institute for kind hospitality and financial support, the Applied Mathematics Department of the University of Zürich for travel grant and Andrew D Barbour for his help with visit arranging. The author also wishes to thank Mihael Perman for reading the text.

References

- [1] ALESHKYAVICHENE, A. K. (1990). Probabilities of large deviations for U -statistics and von Mises functionals. *Teor. Veroyatnost. i Primenen.* **35**, 3–14.
- [2] BARBOUR, A. D. (1982). Poisson convergence and random graphs. *Math. Proc. Cambridge Philos. Soc.* **92**, 349–359.
- [3] BARBOUR, A. D., HOLST, L. AND JANSON, S. (1992). *Poisson approximation* vol. 2 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York.
- [4] 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.
- [5] BOLTHAUSEN, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete* **66**, 379–386.
- [6] BOLTHAUSEN, E. AND GÖTZE, F. (1993). The rate of convergence for multivariate sampling statistics. *Ann. Statist.* **21**, 1692–1710.
- [7] BOROVSKIKH, Y. V. AND WEBER, N. C. (2001). Large deviation results for a U -statistical sum with product kernel. *Bull. Austral. Math. Soc.* **63**, 151–165.
- [8] CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probability* **3**, 534–545.
- [9] 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.
- [10] CHEN, L. H. Y. AND SHAO, Q.-M. (2003). Uniform and non-uniform bounds in normal approximation for nonlinear statistics. Preprint.
- [11] CHEN, L. H. Y. AND SHAO, Q.-M. (2004). Normal approximation under local dependence. *Ann. Probab.* **32**, 1985–2028.

- [12] 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.
- [13] DEMBO, A. AND RINOTT, Y. (1996). Some examples of normal approximations by Stein’s method. In *Random discrete structures (Minneapolis, MN, 1993)*. vol. 76 of *IMA Vol. Math. Appl.* Springer, New York pp. 25–44.
- [14] DUFOUR, J.-M. AND HALLIN, M. (1992). Simple exact bounds for distributions of linear signed rank statistics. *J. Statist. Plann. Inference* **31**, 311–333.
- [15] 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.
- [16] GOLDSTEIN, L. AND RINOTT, Y. (1996). Multivariate normal approximations by Stein’s method and size bias couplings. *J. Appl. Probab.* **33**, 1–17.
- [17] GÖTZE, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19**, 724–739.
- [18] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics* **19**, 293–325.
- [19] INGLOT, T. AND LEDWINA, T. (1993). Cramér type large deviations for some U -statistics. *Theory Probab. Appl.* **38**, 651–659.
- [20] INOGAMOV, R. T. (1986). On large deviations in a combinatorial central limit theorem. *Izv. Akad. Nauk UzSSR Ser. Fiz.-Mat. Nauk* 27–30, 82.
- [21] KALLENBERG, W. C. M. (1982). Cramér type large deviations for simple linear rank statistics. *Z. Wahrsch. Verw. Gebiete* **60**, 403–409.
- [22] KEENER, R. W., ROBINSON, J. AND WEBER, N. C. (1998). Tail probability approximations for U -statistics. *Statist. Probab. Lett.* **37**, 59–65.

- [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] MALEVIČ, T. L. AND ABDALIMOV, B. (1979). Probabilities of large deviations for U -statistics. *Teor. Veroyatnost. i Primenen.* **24**, 215–220.
- [25] PETROV, V. V. (1966). A generalization of Cramér’s limit theorem. In *Selected Translations of Math. Statistics and Probability*. vol. 6. Amer. Math. Soc., Providence, R. I. pp. 1–8.
- [26] PURI, M. L. AND SEOH, M. (1988). On the rate of convergence in normal approximation and large deviation probabilities for a class of statistics. *Theory Probab. Appl.* **33**, 682–697. (1989).
- [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. Probab.* **7**, 1080–1105.
- [30] RINOTT, Y. AND ROTAR, V. (2000). Normal approximations by Stein’s method. *Decis. Econ. Finance* **23**, 15–29.
- [31] RINOTT, Y. AND SCARSINI, M. (2000). On the number of pure strategy Nash equilibria in random games. *Games Econom. Behav.* **33**, 274–293.
- [32] RUBIN, H. AND SETHURAMAN, J. (1965). Probabilities of moderate deviations. *Sankhyā Ser. A* **27**, 325–346.
- [33] RUDIN, W. (1987). *Real and complex analysis*. McGraw-Hill Book Co., New York.
- [34] SAKHANENKO, A. I. (1991). Estimates of Berry-Esseen type for the probabilities of large deviations. *Siberian Math. J.* **32**, 647–656. (1992).

- [35] SCHNELLER, W. (1989). Edgeworth expansions for linear rank statistics. *Ann. Statist.* **17**, 1103–1123.
- [36] SEOH, M., RALESCU, S. S. AND PURI, M. L. (1985). Cramér type large deviations for generalized rank statistics. *Ann. Probab.* **13**, 115–125.
- [37] SLASTNIKOV, A. D. (1978). Limit theorems for probabilities of moderate deviations. *Teor. Veroyatnost. i Primenen.* **23**, 340–357.
- [38] 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*. Univ. California Press, Berkeley, Calif. pp. 583–602.
- [39] STEIN, C. (1986). *Approximate computation of expectations*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA.
- [40] VANDEMAELE, M. (1982). On large deviation probabilities for U -statistics. *Teor. Veroyatnost. i Primenen.* **27**, 573–574.