Proceedings of the Seventh Young Statisticians Meeting
Andrej Mrvar (Editor)
Metodološki zvezki, 21, Ljubljana: FDV, 2003

Normal Approximation by Stein's Method

Martin Raič[1]

Abstract

The aim of this paper is to give an overview of Stein's method, which has turned out to be a powerful tool for estimating the error in normal, Poisson and other approximations, especially for sums of dependent random variables. We focus on the normal approximation of random variables posessing decompositions of Barbour, Karoński, and Ruciński (1989), which are particularly useful in combinatorial structures, where there is no natural ordering of the summands. We highlight two applications: Nash equilibria and linear rank statistics.

1 Introduction

In 1970, Stein introduced a powerful new technique for estimating the rate of convergence of sums of weakly dependent random variables to the standard normal distribution. His approach was subsequently extended by Chen (1975) to the Poisson approximation. The general approach is presented in Stein's (1986) monograph, along with the specializations relevant for normal and Poisson approximation. An overview of normal approximation by Stein's method is given in Rinott and Rotar (2000). In the present paper, we give a somewhat different presentation: we pay greater attention on the derivation of the method and highlight a different concept of weak dependence.

A comprehensive presentation of the Poisson approximation is given in Barbour, Holst, and Janson (1992). Further extensions have been made, among others, to approximation by binomial and multinomial (see Loh, 1992), uniform (see Diaconis, 1989) and gamma distribution (see Luk, 1994).

Stein's method can be readily extended to multivariate and functional settings. Thus, the Poisson approximation can be extended to approximation by a Poisson process (see Barbour, Holst, and Janson, 1992; and Barbour and Xia, 2000). Extensions to multivariate normal and Brownian motion approximation are presented in Götze (1991) and Barbour (1990). Further extensions to measure valued processes have also been undertaken (see Reinert, 1995).

[1] Department of Mathematics, Faculty of Mathematics and Physics, University of Ljubljana, Slovenia.

Stein's method can also be refined to approximations admitting higher rates of convergence than approximations by the 'classical' limit distributions such as normal and Poisson. Barbour (1986) considers Edgeworth expansions for sums of independent random variables. Schneller (1989) derives Edgeworth expansions of the first and the second order for linear rank statistics. In the case of the Poisson approximation, refinements involving compound Poisson approximation (see Barbour, Chen, and Loh, 1992), compound Poisson process approximation (see Barbour and Månsson, 2002) and approximation by the Poisson–Charlier signed measure (see Barbour and Čekanavičius, 2002) have been considered.

A remarkable feature of Stein's method is that it can be applied in many circumstances where dependence plays a part. Stein (1986) discusses applications to Latin rectangles, random allocations, the binary expansion of a random integer and isolated trees in random graphs. More applications to random graphs are given in Barbour, Karoński, and Ruciński (1989). As already mentioned, Stein's method is applicable to linear rank statistics (see Bolthausen, 1984; Schneller, 1989; Barbour, Holst and Janson, 1992; Bolthausen and Götze, 1993; and Goldstein and Reinert, 1997). Other areas of application include the analysis of DNA sequences (see Arratia, Gordon, and Waterman, 1990), additive functionals of correlated Gaussian random variables or components of a multinomial random vector (see Dembo and Rinott, 1996), matrix correlation statistics (see Barbour and Eagleson, 1986), extreme value theory (see Smith, 1988), dissociated statistics (see Barbour and Eagleson, 1985), patterns and runs (see Chryssaphinou and Papastavridis, 1988), reliability theory (see Godbole, 1993), random fields (see Takahata, 1983), spacings and the scan statistics (see Glaz, Naus, Róos, and Wallenstein, 1994; and Dembo and Rinott, 1996), antivoter model, U-statistics (see Rinott and Rotar, 1997) and many others.

The paper is organized as follows. In Section 2, we give an outline of the method. In Section 3, we consider sums of independent random variables and derive Stein's method for normal approximation. The results of Section 3 are extended in Section 4 to dependent random variables posessing a certain kind of dependence structure.

All the estimates in Sections 3 and 4 are given for sufficiently smooth or Lipschitz test functions. In particular, one can derive estimates of the form:

$$\left|\mathbb{E}\, f(W) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(z)\, e^{-\frac{1}{2}z^2}\, dz\right| \le \varepsilon\, M_1(f) \tag{1.1}$$

where $M_1(f)$ is the Lipschitz constant of the function f and where ε is small. In other words, the error in normal approximation is estimated in the Wasserstein L^1 metric, naturally arising from Stein's method. However, in a statistical context, it is more natural to consider the Kolmogorov distance:

$$\delta := \sup_{w \in \mathbb{R}} \left|\mathbb{P}(W \le w) - \Phi(w)\right| \tag{1.2}$$

In general, we only have $\delta = O(\sqrt{\varepsilon})$ (Cf. Section 5). However, one can often show that $\delta = O(\varepsilon)$, but only at the cost of much greater effort. In Section 5, we derive bounds in the Kolmogorov distance comparable to the bounds derived in Section 4 for two cases: for sums of independent random variables (i. e., we prove the classical

Berry–Esséen theorem) and for bounded random variables posessing the dependence structure from Section 4. Finally, in Section 6, we illustrate the whole approach with two examples: Nash equilibria and linear rank statistics.

2 A general approach

Let W be a random variable whose distribution we want to approximate. In other words, for test functions f from a suitable linear space $\mathcal{F}$, we would like to approximate:

$$\mathbb{E}\, f(W) \approx \mathcal{N} f \tag{2.1}$$

where $\mathcal{N} \colon \mathcal{F} \to \mathbb{R}$ is a linear functional which is easier to compute as the expectations on the left-hand side. In order to estimate the error in (2.1), the main idea of Stein's method is to find an easily computable linear operator $\mathcal{A}$ from a linear space $\mathcal{G}$ to $\mathcal{F}$, such that $\mathbb{E}\, \mathcal{A}g(W)$ is small in some sense for all $g \in \mathcal{G}$.

In order to find $\mathcal{A}$, Stein (1986) suggests to first find an exchangeable pair (W, W'). Then the operator:

$$\mathcal{A}_0 g(w) := \mathbb{E}\big[g(W') - g(W) \bigm| W = w\big] \tag{2.2}$$

obviously satisfies $\mathbb{E}\, \mathcal{A}_0 g(W) = 0$, so that $\mathcal{A}$ can be sought as an approximation to $\mathcal{A}_0$. More generally, one can take an exchangeable pair (X, X') and a linear operator $\mathcal{T}_0$ mapping $\mathcal{G}$ into the space of antisymmetric functions; then define:

$$\mathcal{A}_0 g(w) := \mathbb{E}\big[\mathcal{T}_0 g(X, X') \bigm| W = w\big] \tag{2.3}$$

Notice that (2.3) reduces to (2.2) when taking $X = W$, $X' = W'$ and $\mathcal{T}_0 g(w, w') = g(w') - g(w)$.

Once we have found $\mathcal{A}$, one way of finding $\mathcal{N}$ and estimating the error in (2.1) is to find a projector $\mathcal{P}$ (in the space $\mathcal{F}$) with $\ker \mathcal{P} \subseteq \operatorname{im} \mathcal{A}$ and to solve the *Stein equation*:

$$\mathcal{A}g = f - \mathcal{P}f \tag{2.4}$$

leading to:

$$\mathbb{E}\, f(W) = \mathbb{E}\, \mathcal{P}f(W) + \mathbb{E}\, \mathcal{A}g(W) \tag{2.5}$$

Consequently, $\mathcal{N}f := \mathbb{E}\, \mathcal{P}f(W)$ is a good approximation to $\mathbb{E}\, f(W)$ provided that $\mathbb{E}\, \mathcal{A}g(W)$ is small. Clearly, this only makes sense if $\mathcal{P}f$ is such a function that $\mathbb{E}\, \mathcal{P}f(W)$ is easier to compute as $\mathbb{E}\, f(W)$. In fact, $\mathcal{P}f$ can usually be chosen to be a constant, so that we can identify $\mathcal{N}f \equiv \mathcal{P}f$. This is because one can choose $\mathcal{A}_0$ and $\mathcal{A}$ with images of codimension one.

Another way to obtain an approximation to $\mathbb{E}\, f(W)$ together with the error estimate is to observe that $\mathcal{A}$ may be an infinitesimal generator of a Markov process (or at least an operator semigroup). Indeed, in view of (2.2), the exchangeable pair (W, W') can be extended to a stationary Markov chain with generator $\mathcal{A}_0$. Now suppose that $\mathcal{A}$ generates an operator semigroup $\mathcal{P}_t$, $t \geq 0$ with $\mathcal{P} = \lim_{t\to\infty} \mathcal{P}_t$. Then it follows from the Dynkin formula (see Ethier and Kurtz, 1986) that:

$$\mathcal{P}f - f = \int_0^\infty \mathcal{A}\mathcal{P}_t f \, dt \tag{2.6}$$

Similarly as before, $\mathbb{E}\,\mathcal{P}f(W)$ is a good approximation to $\mathbb{E}\,f(W)$ if the expectations $\mathbb{E}\,\mathcal{A}\mathcal{P}_t f(W)$ are sufficiently small and tend to zero sufficiently fast when $t \to \infty$. Moreover, if $\mathcal{P}_t f$ converges to $\mathcal{P}f$ sufficiently fast, observe that:

$$g = -\int_0^\infty (\mathcal{P}_t f - \mathcal{P}f)\,dt \tag{2.7}$$

solves (2.4) (note that $\mathcal{A}\mathcal{P} = 0$).

Remark. If $\mathcal{A}$ generates an ergodic Markov process with stationary distribution ν, we have $\mathcal{N}f = \mathcal{P}f = \int f\,d\nu$, so that ν approximates the distribution of W.

Remark. Using (2.6), we need not find an explicit solution g to the Stein equation (2.4); in view of (2.6), $\mathbb{E}\,\mathcal{P}_t f(W)$, $t \geq 0$, is a path from $\mathbb{E}\,f(W)$ to $\mathcal{N}f$, and the error in the approximation is estimated using bounds on the derivative along the path. This is known as *continuation method* in numerical analysis (see Allgower and Georg, 1990).

3 Stein's method for the normal approximation

Let $X_1, X_2, \ldots X_n$ be independent random variables with sum W. Without loss of generality, we can assume that $\mathbb{E}\,X_i = 0$ for all i and that $\mathrm{var}(W) = 1$. Moreover, let $X'_1, \ldots X'_n$ be independent copies of $X_1, \ldots X_n$. Taking $W_i := W - X_i$ and $W'_i := W_i + X'_i$, the pair (W, W'_i) is clearly exchangeable. In way of (2.2), define:

$$S_{1i} := g(W'_i) - g(W) \tag{3.1}$$

Excheangability implies that $\mathbb{E}\,S_{1i} = 0$. According to Section 2, we shall seek the Stein operator $\mathcal{A}g(W)$ as an approximation to $\mathbb{E}(S_{1i} \mid W)$. Denote $X \sim Y$ for random variables with $\mathbb{E}(X \mid W) = \mathbb{E}(Y \mid W)$. Assuming that g is twice continuously differentiable, Taylor's expansions of W'_i centered at W_i and of W_i centered at W with remainders in integral form yield:

$$\begin{aligned} S_{1i} &= \big(g(W'_i) - g(W_i)\big) + \big(g(W_i) - g(W)\big) \sim \\ &\sim g'(W_i)X'_i - g'(W)X_i + U_{1i} \end{aligned} \tag{3.2}$$

where:

$$U_{1i} := (1-\theta)g''(W_i + \theta X'_i)X'^2_i + \theta g''(W_i + \theta X_i)X_i^2 \tag{3.3}$$

and where θ is uniformly distributed over $[0,1]$ and independent of all other variates. By independence, $g'(W_i)X'_i \sim 0$. Adding over i, we obtain:

$$\sum_{i=1}^n S_{1i} \sim -g'(W)W + \sum_{i=1}^n U_{1i} \tag{3.4}$$

Now write:

$$U_{1i} = U_{2i} + R_i \tag{3.5}$$

where:

$$U_{2i} := (1-\theta)g''(W)X'^2_i + \theta g''(W'_i)X_i^2 \sim \frac{1}{2}\Big[g''(W)X'^2_i + g''(W'_i)X_i^2\Big] \tag{3.6}$$

The remainder R_i can be estimated in the following way: for every sufficiently smooth f, define:

$$M_r(f) := \begin{cases} \sup\limits_{\substack{x,y\in\mathbb{R}\\ x\neq y}} \left| \dfrac{f^{(r-1)}(x) - f^{(r-1)}(y)}{x-y} \right| & ;\ f \in \mathcal{C}^{(r-1)}(\mathbb{R}) \\ \infty & ;\ \text{otherwise} \end{cases} \tag{3.7}$$

Using the inequality $xy^2 \le \frac{1}{3}x^3 + \frac{2}{3}y^3$ holding for all $x, y \ge 0$, we find that:

$$\begin{aligned} |R_i| &\le M_3(g)(1-\theta)|X_i||X_i'|^2 + \theta(1-\theta)|X_i'|^3 + \theta|X_i'||X_i|^2 + \theta^2|X_i|^3 \le \\ &\le M_3(g)\Big[\big(\theta^2 + \tfrac{1}{3}\theta + \tfrac{1}{3}\big)|X_i|^3 + \big(-\theta^2 + \tfrac{2}{3}\theta + \tfrac{2}{3}\big)|X_i'|^3\Big] \sim \\ &\sim M_3(g)\big(\tfrac{5}{6}|X_i|^3 + \tfrac{2}{3}|X_i'|^3\big) \end{aligned} \tag{3.8}$$

Now we turn to U_{2i}. For the first term, we have by independence:

$$g''(W)X_i'^2 \sim \sigma_i^2 g''(W) \tag{3.9}$$

where $\sigma_i^2 = \mathbb{E}\,X_i'^2 = \mathbb{E}\,X_i^2 = \operatorname{var}(X_i)$. To handle the second term, we shall apply a minor correction to S_{1i}. In view of (2.3), define:

$$S_{2i} := \frac{1}{2}\Big[g''(W_i')X_i^2 - g''(W)X_i'^2\Big] \tag{3.10}$$

By exchangeability, we have $\mathbb{E}\,S_{2i} = 0$. Combining (3.9) and (3.10), we obtain:

$$U_{2i} \sim \sigma_i^2 g''(W) + S_{2i} \tag{3.11}$$

Collecting (3.2), (3.4), (3.5) and (3.11), we finally find:

$$\sum_{i=1}^n (S_{1i} - S_{2i}) \sim \mathcal{A}g(W) + \sum_{i=1}^n R_i \tag{3.12}$$

where:

$$\mathcal{A}g(w) := g''(w) - g'(w)w \tag{3.13}$$

Now since the expectations of S_{1i} and S_{2i} vanish, the absolute values of expectations of the two terms on the r. h. s. of (3.12) must be equal. Consequently,

$$|\mathbb{E}\,\mathcal{A}g(W)| \le M_3(g)\sum_{i=1}^n \left(\frac{5}{6}\,\mathbb{E}\,|X_i|^3 + \frac{2}{3}\,\mathbb{E}\,|X_i'|^3\right) = \frac{3}{2}\sum_{i=1}^n \mathbb{E}\,|X_i|^3 \tag{3.14}$$

Remark. The expression on the right-hand side is typically small. In particular, if all random variables X_i are identically distributed, i. e. $X_i \stackrel{\mathrm{d}}{=} n^{-1/2}X$, where X can be any random variable with $\mathbb{E}\,X = 0$ and $\operatorname{var}(X) = 1$, we have $\sum_{i=1}^n \mathbb{E}\,|X_i|^3 = n^{-1/2}\,\mathbb{E}\,|X|^3$.

Having found $\mathcal{A}$ and derived (3.14), there are two possibilities to approximate the distribution of W as described in Section 2. The first one is to solve the Stein equation (2.4), which in our case reduces to an ordinary differential equation. Writing $h := g'$, we have in fact to find a solution to the equation:

$$h'(x) - h(x)x = f(x) - \mathcal{P}f \tag{3.15}$$

with $M_2(h) = M_3(g) < \infty$. First observe that every function f with $M_r(f) < \infty$ for some r is of *polynomial growth* in the following sense:

Definition. A function $f\colon \mathbb{R} \to \mathbb{R}$ is of *polynomial growth* if its norm defined by:

$$\|f\|_r := \sup_{x\in\mathbb{R}} \frac{|f(x)|}{1+|x|^r} \tag{3.16}$$

is bounded for some $r \geq 0$.

Now suppose that a function h of polynomial growth solves:

$$h'(x) - h(x)x = f(x) \tag{3.17}$$

in weak sense, i. e. h is absolutely continuous on every finite interval and (3.17) holds for almost all $x \in \mathbb{R}$. Clearly, f must be Lebesgue integrable on every finite interval and:

$$h(x) = e^{\frac{1}{2}x^2}\left(h(0) + \int_0^x f(z)e^{-\frac{1}{2}z^2}\,dz\right) \tag{3.18}$$

Since h is of polynomial growth, we have:

$$\lim_{x\to\infty} e^{-\frac{1}{2}x^2}h(x) = \lim_{x\to-\infty} e^{-\frac{1}{2}x^2}h(x) = 0 \tag{3.19}$$

Therefore,

$$\int_{-\infty}^{\infty} f(z)\, e^{-\frac{1}{2}z^2}\,dz = 0 \tag{3.20}$$

where the integral is taken to be the limit of the Lebesgue integrals over finite intervals, and h satisfies:

$$h(x) = e^{\frac{1}{2}x^2}\int_{-\infty}^{x} f(z)\, e^{-\frac{1}{2}z^2}\,dz = -e^{\frac{1}{2}x^2}\int_{x}^{\infty} f(z)\, e^{-\frac{1}{2}z^2}\,dz \tag{3.21}$$

Having established necessary conditions for a solution to (3.17) to be of polynomial growth, we now turn to sufficient conditions. Define:

$$\mathcal{N}f := \mathcal{P}f := \int_{-\infty}^{\infty} f(z)\,\phi(z)\,dz \tag{3.22}$$

where ϕ denotes the standard normal density:

$$\phi(z) := \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2} \tag{3.23}$$

Lemma 3.1 *Let $f\colon \mathbb{R} \to \mathbb{R}$ be a measurable function of polynomial growth. Then the function:*

$$h(x) = e^{\frac{1}{2}x^2} \int_{-\infty}^{x} (f(z) - \mathcal{N}f)\, e^{-\frac{1}{2}z^2}\, dz = -e^{\frac{1}{2}x^2} \int_{x}^{\infty} (f(z) - \mathcal{N}f)\, e^{-\frac{1}{2}z^2}\, dz \tag{3.24}$$

is the unique solution of polynomial growth to (3.15)*; in fact, for every $r \geq 1$, there is a constant C_r, such that:*

$$\|h\|_{r-1} \leq C_r \|f\|_r \tag{3.25}$$

Moreover,

$$M_2(h) \leq 2M_1(f) \tag{3.26}$$

Proof. For the first part, it only remains to prove (3.25). That estimate can be derived by the observation that for every $x \geq 0$ and $r \geq 1$,

$$\begin{aligned} \int_x^{\infty} z^r\, e^{-\frac{1}{2}z^2}\, dz &= \int_0^{\infty} (u+x)^r\, e^{-\frac{1}{2}(u+x)^2}\, du \leq \\ &\leq 2^{r-1} \int_0^{\infty} (u^r + x^r) e^{-\frac{1}{2}(u+x)^2}\, du \leq \\ &\leq 2^{r-1} e^{-\frac{1}{2}x^2} \int_0^{\infty} \left(u^r e^{-\frac{1}{2}u^2} + x^r e^{-ux}\right) du \leq \\ &\leq e^{-\frac{1}{2}x^2} \left[2^{\frac{3(r-1)}{2}} \Gamma\left(\frac{r+1}{2}\right) + 2^{r-1} x^{r-1} \right] \end{aligned} \tag{3.27}$$

For the second part, see Stein (1986). □

The operator semigroup approach can also be used to our advantage. Observe that the operator $\mathcal{A}$ generates the *Ornstein–Uhlenbeck semigroup* (see Rogers and Williams, 1994):

$$\mathcal{P}_t f(x) := \int_{-\infty}^{\infty} f\left(e^{-t}x + \sqrt{1 - e^{-2t}}\, z\right) \phi(z)\, dz \tag{3.28}$$

Lemma 3.2 *For a measurable function $f\colon \mathbb{R} \to \mathbb{R}$ of polynomial growth, the integral:*

$$g(x) = -\int_0^{\infty} \left(\mathcal{P}_t f(x) - \mathcal{N}f\right) dt \tag{3.29}$$

exists for all $x \in \mathbb{R}$ and solves:

$$g''(x) - g'(x)x = f(x) - \mathcal{N}f \tag{3.30}$$

Moreover, g' is of polynomial growth, coincides with h defined in (3.24) *and:*

$$M_3(g) = M_2(h) \leq \min\left\{ \frac{\sqrt{2\pi}}{4}\, M_2(f), \frac{1}{3}\, M_3(f) \right\} \tag{3.31}$$

Proof. In view of (2.7), write:

$$\begin{aligned} \mathcal{P}_t f(x) - \mathcal{N}f = \mathcal{P}_t f(x) - \mathcal{N}\mathcal{P}_t f &= \int_{-\infty}^{\infty} \Big[\mathcal{P}_t f(x) - \mathcal{P}_t f(y)\Big]\phi(y)\,dy = \\ &= \int_{-\infty}^{\infty}\int_y^x \frac{d}{ds}\mathcal{P}_t f(s)\,ds\,\phi(y)\,dy \end{aligned} \tag{3.32}$$

In order to evaluate $d/dx\,\mathcal{P}_t f(x)$, write:

$$\begin{aligned} \mathcal{P}_t f(x) &= \int_{-\infty}^{\infty} f\big(e^{-t}x + \sqrt{1-e^{-2t}}\,z\big)\,\phi(z)\,dz = \\ &= \int_{-\infty}^{\infty} f\big(\sqrt{1-e^{-2t}}u\big)\,\phi\left(u - \frac{1}{\sqrt{e^{2t}-1}}\,x\right)\,du \end{aligned} \tag{3.33}$$

Differentiation yields:

$$\begin{aligned} \frac{d}{dx}\mathcal{P}_t f(x) &= -\frac{1}{\sqrt{e^{2t}-1}}\int_{-\infty}^{\infty} f\big(\sqrt{1-e^{-2t}}u\big)\,\phi'\left(u - \frac{1}{\sqrt{e^{2t}-1}}\,x\right)\,du = \\ &= -\frac{1}{\sqrt{e^{2t}-1}}\int_{-\infty}^{\infty} f\big(e^{-t}x + \sqrt{1-e^{-2t}}\,z\big)\,\phi'(z)\,dz \end{aligned} \tag{3.34}$$

Since f is of polynomial growth, $\|f\|_r$ is bounded for some $r \ge 0$. An easy calculation shows that $\|d/dx\,\mathcal{P}_t f(x)\|_r$ is also bounded. Using (3.32), one can then check that:

$$\int_0^{\infty} \|\mathcal{P}_t f - \mathcal{N}f\|_{r+1}\,dt < \infty \tag{3.35}$$

Hence the integral on the r. h. s. of (3.29) exists for all $x \in \mathbb{R}$. Moreover, since $\mathcal{P}_t$, $t \ge 0$, is a continuous operator semigroup on the space of all functions ψ with $\|\psi\|_{r+1} < \infty$, it follows from Dynkin formula that $\mathcal{A}g = f - \mathcal{N}f$, where $\mathcal{A}$ is the infinitesimal generator of the semigroup. Notice that $\mathcal{A}g$ coincides with the l. h. s. of (3.30) provided that $g \in \mathcal{C}^{(2)}(\mathbb{R})$ with g'' being of polynomial growth.

Now suppose that $f \in \mathcal{C}^{(1)}(\mathbb{R})$ and that f' is of polynomial growth. Differentiation of (3.29) together with (3.34) yields:

$$g'(x) = -\int_0^{\infty} \frac{d}{dx}\mathcal{P}_t f(x)\,dt = \int_0^{\infty}\frac{1}{\sqrt{e^{2t}-1}}\,f\big(e^{-t}x + \sqrt{1-e^{-2t}}\,z\big)\,\phi'(z)\,dz\,dt \tag{3.36}$$

$$g''(x) = \int_0^{\infty} \frac{e^{-t}}{\sqrt{e^{2t}-1}}\,f'\big(e^{-t}x + \sqrt{1-e^{-2t}}\,z\big)\,\phi'(z)\,dz\,dt \tag{3.37}$$

implying the desired properties of g. Hence g solves (3.30). Moreover, it follows from (3.36) that g' is of polynomial growth; in fact, one can easily check that $\|g'\|_r \le C_r\|f\|_r$ for some C_r independent of f. Since h is the unique solution to (3.15) being of polynomial growth, g' must coincide with h.

Now choose some $r \ge 1$ and consider the space of all functions f with $\|f\|_r < \infty$. Since the operators mapping f to g' and to h are bounded with respect to the norm

$\|\cdot\|_r$ and coincide on the dense subspace of all functions f, such that f' is continuous and of polynomial growth, g' coincides with h and therefore g solves (3.30) for every function f of polynomial growth.

Finally, in order to derive (3.31), observe that by differentiating (3.29) and (3.37) and noting that every Lipschitz function is differentiable almost everywhere and satisfies the fundamental theorem of calculus (see Rudin, 1987):

$$\begin{aligned} g'''(x) &= -\int_0^\infty e^{-3t} \int_{-\infty}^\infty f'''\big(e^{-t}x + \sqrt{1-e^{-2t}}\, z\big)\, \phi(z)\, dz\, dt = \\ &= \int_0^\infty \frac{e^{-2t}}{\sqrt{e^{2t}-1}}\, f''\big(e^{-t}x + \sqrt{1-e^{-2t}}\, z\big)\, \phi'(z)\, dz\, dt \end{aligned} \tag{3.38}$$

The estimate (3.31) obviously follows. □

The operator semigroup is important because it can be readily extended to multivariate and functional settings. In the d-variate setting, we simply replace ϕ by the d-variate standard normal density and $\mathcal{A}$ by:

$$\mathcal{A}g(w) := \Delta g(w) - \langle \nabla g(w), w\rangle \tag{3.39}$$

For extensions to the functional setting, see Barbour (1990).

In the multivariate setting, the estimate (3.31) remains unchanged and the proof can more or less be obtained by rewriting the proof of (3.31), where g is defined analogously to (3.29) and where $M_r(f)$ are suitable generalizations of the r. h. s. of (3.7). On the other hand, $M_3(g)$ cannot be estimated in terms of $M_1(f)$. In fact, it turns out by straightforward calculation that for the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by:

$$f(x,y) := \max\{\min\{x,y\}, 0\} \tag{3.40}$$

the corresponding function g is twice continuously differentiable, but $\partial^2 g/\partial x\, \partial y$ is not Lipschitz.

However, we can go one step further by observing that we only need ∇g, not g. Hence, instead of seeking the solution g to the equation $\mathcal{A}g = f - \mathcal{N}f$, it suffices to find a vector field h satisfying:

$$\operatorname{div} h(w) - \langle h(w), w\rangle = f(w) - \int_{\mathbb{R}^d} f(x)\, \phi(x)\, dx \tag{3.41}$$

Open problem. Let $m \geq 2$. Is there a constant C_m, such that for every Lipschitz function $f\colon \mathbb{R}^m \to \mathbb{R}$, there is a vector field h satisfying (3.41) and such that the Lipschitz constant of its second partial derivatives is bounded by C_m? If yes, how rapidly does C_m grow with m?

To summarize the results of the whole section, we state the following results implied by Lemmas 3.1 and 3.2, describing the behavior of the solutions to the Stein equation:

Corollary 3.3 *For every function $f\colon \mathbb{R} \to \mathbb{R}$ with finite $M_1(f)$, $M_2(f)$ or $M_3(f)$, there is a solution h to* (3.15) *satisfying:*

$$M_2(h) \le M(f) := \min\left\{2M_1(f), \frac{\sqrt{2\pi}}{4} M_2(f), \frac{1}{3} M_3(f)\right\} \tag{3.42}$$

provided that the quantity on the r. h. s. is finite.

Having examined solutions to the Stein equation, we can now bound the error in the normal approximation: in order to approximate the distribution of *any* random variable W (not necessarily a sum of independent random variables), we only need to estimate $\left|\mathbb{E}[h'(W) - h(W)W]\right|$ in terms of $M_2(h)$; then we apply Corollary 3.3, which has been proved once and for all.

Notice that bounds in terms of $M(f)$ imply weak convergence: any sequence of probability measures μ_n satisfying $\left|\int f\, d\mu_n - \mathcal{N}f\right| \le \varepsilon_n M(f)$, where $\varepsilon_n \to 0$, converges weakly to the standard normal distribution.

For sums of independent random variables, (3.14) together with Corollary 3.3 yields:

Theorem 3.4 *For independent random variables $X_1, \dots X_n$ with sum W and with $\mathbb{E}\, X_i = 0$ and $\operatorname{var}(W) = 1$, we have for every sufficiently smooth f,*

$$\left|\mathbb{E}\, f(W) - \mathcal{N}f\right| \le \frac{3}{2} M(f) \sum_{i=1}^{n} \mathbb{E}\, |X_i|^3 \tag{3.43}$$

The estimate in (3.43) is analogous to the one in the Berry–Esséen theorem, except that it is based on smooth or Lipschitz test functions. A proof of the Berry–Esséen theorem by means of Stein's method will be given in Section 5. However, the main strength of Stein's method is that it is readily applicable to sums of *dependent* random variables. This will be done in the next section.

4 A CLT for decomposable random variables

As we have seen in the previous section, the distribution of a random variable W will be close to the standard normal if the Stein expectation:

$$\mathbb{E}\big[h'(W) - h(W)W\big] \tag{4.1}$$

is small with respect to $M_2(h)$. In order to estimate (4.1) for sums of dependent random variables, several techniques have been developed. All of them are based on constructing auxiliary random variables W^* on a joint probability space with W, having certain properties and being close to W. This is called *"auxiliary randomization"* by Stein; the idea appears under the name of *"coupling"* in other contexts (see e.g., Lindvall, 1992 and references therein).

The properties of W^* that can be used to our advantage can be quite different. Stein (1986) chooses W^*, such that its conditional expectation with respect to W is a linear function of W (inspired by the fact that this is true if the joint distribution of (W, W^*) is bivariate normal). This idea was subsequently extended by Rinott and Rotar (1997).

Another idea is to construct W^* with a certain distribution depending on the distribution of W. Baldi, Rinott, and Stein (1989) (see also Dembo and Rinott, 1996; and Goldstein and Rinott, 1996) propose the *size-biassed coupling*: for non-negative W, W^* is chosen so that $\mathbb{P}(W^* \in dw) = w\,\mathbb{P}(W \in dw)/\,\mathbb{E}\,W$, or, in other words, $\mathbb{E}\,f(W^*) = \mathbb{E}\,f(W)W/\,\mathbb{E}\,W$. Another approach is the so called *zero biassed coupling* proposed by Goldstein and Reinert (1997), where for W with $\mathbb{E}\,W = 0$, $\mathbb{E}\,f'(W^*) = \mathbb{E}\,f(W)W$ (although the same construction had in fact already been used by Chen and Ho (1978) and Stein (1986) to prove the classical Berry–Esséen theorem).

In most applications, the random variable to be approximated can be regarded as a sum of weakly dependent random variables. For $W = \sum_{i\in I} X_i$, the Stein expectation takes the form:

$$\mathbb{E}\,h'(W) - \sum_{i\in I} \mathbb{E}\,h(W)X_i \tag{4.2}$$

One way of evaluating $\mathbb{E}\,h(W)X_i$ is to examine the conditional distribution of W given X_i. In particular, it is often useful to construct auxiliary random variables being close to W and having the conditional distribution of W given $X_i = x$. This allows us, for instance, to construct good size biassed couplings.

Observe that the conditional distribution of W given X_i will be close to its unconditional distribution if W can be written as a sum of two random variables, the first one independent of X_i and the second one being small. This elegant and powerful approach was already introduced by Stein (1972), but considerably simplified by Barbour, Karoński, and Ruciński (1989). In fact, it can be used in most of the applications of normal approximation by Stein's method that have been studied. In particular, it extends the concept of *local dependence* (see Chen, 1978; Chen, 1986; Stein, 1986; Arratia, Goldstein, and Gordon, 1990; Rinott, 1994; and Goldstein and Rinott, 1996). We shall take a closer look at local dependence at the end of this section.

Following Barbour, Karoński, and Ruciński (1989), suppose that W is a random variable decomposed in the following way:

$$W = \sum_{i\in I} X_i \tag{4.3}$$

$$\mathbb{E}\,X_i = 0,\ i \in I; \quad \operatorname{var}(W) = 1 \tag{4.4}$$

$$W = W_i + Z_i,\ i \in I, \quad \text{where } W_i \text{ is independent of } X_i \tag{4.5}$$

$$Z_i = \sum_{k\in K_i} Z_{ik}, \quad i \in I \tag{4.6}$$

$$\begin{gathered} W_i = W_{ik} + V_{ik}, \quad i \in I,\ k \in K_i \\ \text{where } W_{ik} \text{ is independent of the pair } (X_i, Z_{ik}), \end{gathered} \tag{4.7}$$

and where I and K_i are index sets. We also assume that:

$$\sum_{i\in I}\bigl(\mathbb{E}\,|X_i|^2\bigr)^{1/2} < \infty, \quad \sum_{i\in I}\sum_{k\in K_i}\mathbb{E}\,|X_i|\,|Z_{ik}| < \infty \tag{4.8}$$

The following theorem is a slight modification of the result derived by Barbour, Karoński, and Ruciński (1989):

Theorem 4.1 *For every random variable W decomposed as in* (4.3)–(4.7) *and every function $f\colon\ \mathbb{R}\to\mathbb{R}$, we have:*

$$\begin{aligned}\bigl|\mathbb{E}\,f(W)-\mathcal{N}f\bigr| \le M(f)\biggl[\frac{1}{2}\sum_{i\in I}\mathbb{E}\,|X_i|Z_i^2 + \\ +\sum_{i\in I}\sum_{k\in K_i}\bigl(\mathbb{E}\,|X_iZ_{ik}V_{ik}| + \mathbb{E}\,|X_iZ_{ik}|\,\mathbb{E}\,|Z_i+V_{ik}|\bigr)\biggr]\end{aligned} \tag{4.9}$$

where $M(f)$ is as in (3.42) *and where $\mathcal{N}f$ is as in* (3.22)*.*

Remark. If the X_i's are independent, we can set $Z_i := X_i$, $K_i := \{0\}$ and $V_{i0} := 0$, and Theorem 4.1 yields (3.43).

Proof of Theorem 4.1. By Theorem 3.1, it suffices to estimate:

$$\mathbb{E}\Bigl[h'(W)-h(W)W\Bigr] = \sum_{i\in I}\mathbb{E}\Bigl[h'(W)\,\mathbb{E}\,X_iW - h(W)X_i\Bigr] \tag{4.10}$$

Taylor's expansion of W centered at W_i yields:

$$\mathbb{E}\,h(W)X_i = \mathbb{E}\,h(W_i)X_i + \mathbb{E}\,h'(W_i)X_iZ_i + \rho_{1i} \tag{4.11}$$

where:

$$|\rho_{1i}| \le \frac{1}{2}\,M_2(h)\,\mathbb{E}\,|X_i|Z_i^2 \tag{4.12}$$

By independence, the first term on the r. h. s. of (4.11) vanishes, while the second term can be rewritten in the following way, using Taylor's expansion and independence:

$$\begin{aligned}\mathbb{E}\,h'(W_i)X_iZ_i &= \sum_{k\in K_i}\mathbb{E}\,h'(W_i)X_iZ_{ik} = \\ &= \sum_{k\in K_i}\mathbb{E}\,h'(W_{ik})X_iZ_{ik} + \rho_{2i} = \\ &= \sum_{k\in K_i}\mathbb{E}\,h'(W_{ik})\,\mathbb{E}\,X_iZ_{ik} + \rho_{2i}\end{aligned} \tag{4.13}$$

where:

$$|\rho_{2i}| \le M_2(h)\sum_{k\in K_i}\mathbb{E}\,|X_iZ_{ik}V_{ik}| \tag{4.14}$$

Finally, for the first term of (4.10), we have:

$$\begin{aligned}\mathbb{E}\,h'(W)\,\mathbb{E}\,X_iW &= \mathbb{E}\,h'(W)\,\mathbb{E}\,X_iZ_i = \sum_{k\in K_i}\mathbb{E}\,h'(W)\,\mathbb{E}\,X_iZ_{ik}\\ &= \sum_{k\in K_i}\mathbb{E}\,h'(W_{ik})\,\mathbb{E}\,X_iZ_{ik} + \rho_{3i}\end{aligned} \tag{4.15}$$

where:

$$|\rho_{3i}| \le M_2(h)\sum_{k\in K_i}\mathbb{E}\,|X_iZ_{ik}|\,\mathbb{E}\,|Z_i+V_{ik}| \tag{4.16}$$

Combining (4.11), (4.13), (4.15) and adding over i, only the remainders survive, so that:

$$\mathbb{E}\big[h'(W) - h(W)W\big] = \sum_{i\in I}\big(\rho_{3i} - \rho_{1i} - \rho_{2i}\big) \tag{4.17}$$

which, together with Theorem 3.1, proves the desired result. □

Now we turn to locally dependent random variables according to Rinott's (1994) definition:

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 for any two disjoint subsets $K, L \subset I$ which are not connected by an edge of Γ, the collections $\{X_k : k \in K\}$ and $\{X_l : l \in L\}$ are independent.

The following theorem is useful for *finite dependence*, i. e. the case where the maximum degree of the dependence graph is finite.

Theorem 4.2 *Let Γ be a dependence graph for a collection $X_i, i \in I$ with $\mathbb{E}\,X_i = 0$ and $\operatorname{var}(W) = 1$, where $W = \sum_{i\in I} X_i$. Then we have:*

$$\big|\mathbb{E}\,f(W) - \mathcal{N}f\big| \le \frac{7}{2}M(f)(D+1)^2\sum_{i\in I}\mathbb{E}\,|X_i|^3 \tag{4.18}$$

where D denotes the maximum degree of Γ.

Proof. In order to satisfy (4.5)–(4.7), define K_i to be the set of all vertices adjacent to i with respect to Γ (including i itself) and put:

$$Z_{ik} := X_k, \quad V_{ik} := \sum_{l\in K_k\setminus K_i} X_l \tag{4.19}$$

Theorem 4.1 yields:

$$\begin{aligned}\big|\mathbb{E}\,f(W) - \mathcal{N}f\big| \le M(f)\Bigg[&\frac{1}{2}\sum_{i\in I}\sum_{k\in K_i}\sum_{l\in K_i}\mathbb{E}\,|X_iX_kX_l| +\\ &+ \sum_{i\in I}\sum_{k\in K_i}\sum_{l\in K_k\setminus K_i}\mathbb{E}\,|X_iX_kX_l| + \sum_{i\in I}\sum_{k\in K_i}\sum_{l\in K_i\cup K_k}\mathbb{E}\,|X_iX_k|\,\mathbb{E}\,|X_l|\Bigg]\end{aligned} \tag{4.20}$$

The proof is now completed by applying the inequality $xyz \le \frac{1}{3}(x^3 + y^3 + z^3)$ $(x, y, z \ge 0)$ and straightforward calculation. □

5 Berry–Esséen theorem

In this section, we shall estimate the error in normal approximation in the Kolmogorov metric, i. e. for a given random variable W, we shall estimate:

$$\sup_{w\in\mathbb{R}} \bigl|\mathbb{P}(W \le w) - \Phi(w)\bigr| \tag{5.1}$$

where:

$$\Phi(w) := \int_{-\infty}^{w} \phi(z)\, dz \tag{5.2}$$

and where ϕ is as in (3.23). In other words, we take the following test functions:

$$f_w(x) := \begin{cases} 1 & ;\ x \le w \\ 0 & ;\ x > w \end{cases} \tag{5.3}$$

Unfortunately, the results of the previous section cannot be applied directly because the first derivative of the solution h_w to the Stein equation:

$$h_w'(x) - h_w(x)x = f_w(x) - \Phi(w) \tag{5.4}$$

given by (3.24) is not continuous, let alone Lipschitz or smooth. One way to overcome this obstacle is to use the zero bias transformation mentioned in Section 4: see Chapter IX in Stein (1986). However, this approach requires a deconvolution procedure, which may cause problems and has remained limited to sums of i. i. d. random variables.

Another useful idea is to approximate the functions f_w by Lipschitz functions in the following way:

$$f_{w,\varepsilon}(x) := \begin{cases} 1 & ;\ x \le w - \varepsilon \\ 0 & ;\ x \ge w + \varepsilon \\ \text{linear} & ;\ w - \varepsilon \le x \le w + \varepsilon \end{cases} \tag{5.5}$$

A straightforward calculation shows that:

$$\begin{aligned} \mathbb{P}(W \le w) - \Phi(w) &\le \mathbb{E}\, f_{w+\varepsilon,\varepsilon}(W) - \mathcal{N} f_{w+\varepsilon,\varepsilon} + \frac{\varepsilon}{\sqrt{2\pi}} \\ \mathbb{P}(W \le w) - \Phi(w) &\ge \mathbb{E}\, f_{w-\varepsilon,\varepsilon}(W) - \mathcal{N} f_{w-\varepsilon,\varepsilon} - \frac{\varepsilon}{\sqrt{2\pi}} \end{aligned} \tag{5.6}$$

where $\mathcal{N}$ is as in (3.22). Furthermore, defining $h_{w,\varepsilon}$ analogously to h_w, Lemma 3.1 yields:

$$M_2(h_{w,\varepsilon}) \le 2M_1(f_{w,\varepsilon}) \le \frac{1}{\varepsilon} \tag{5.7}$$

Now suppose that:

$$\bigl|\mathbb{E}\bigl[h'(W) - h(W)W\bigr]\bigr| \le \delta\, M_2(h) \tag{5.8}$$

Combining (5.6) and (5.7), we obtain:

$$\bigl|\mathbb{P}(W \le w) - \Phi(w)\bigr| \le \frac{\delta}{\varepsilon} + \frac{\varepsilon}{\sqrt{2\pi}} \tag{5.9}$$

and optimization over ε unfortunately only yields:

$$\left|\mathbb{P}(W \le w) - \Phi(w)\right| \le \frac{2}{\sqrt[4]{2\pi}} \sqrt{\delta} \tag{5.10}$$

which is in most cases a crude bound. Typically, it can be shown that the l. h. s. of (5.10) is of order $O(\delta)$, as for Lipschitz and smooth test functions. However, the derivation of the latter requires a more delicate argument with the key observation that $h'_{w,\varepsilon}$ has only one jump at w, but is Lipschitz on $(-\infty, w)$ as well as on (w, ∞). Alternatively, we can use the fact that $h''_{w,\varepsilon}$ is in fact only large on a small interval. Hence, in view of (4.11), (4.13) and (4.15), the estimate of the l. h. s. of (5.10) depends on proving a conditional concentration inequality for a perturbation of W. More precisely, one has to prove that for a certain class of random variables X, $\mathbb{P}(a \le X \le b) \le \varepsilon + c(b-a)$ for all $a < b$, where ε and c are universal constants and where ε is small. This can be done separately – by the zero bias transformation (see Chen and Ho, 1978) or by a bootstrapping argument (see Chen, 1986; and Chen and Shao, 2001). The other possibility is to refer back to normal approximation, requiring an inductive (see Bolthausen, 1984; Götze, 1991; and Bolthausen and Götze, 1993) or a bootstrapping argument (see Rinott, 1994; and Rinott and Rotar, 1996).

Rinott (1994) proves a central limit theorem for sums of bounded locally dependent random variables. In the present paper, we first extend that result to random variables with bounded decompositions in the sense of Section 4:

Theorem 5.1 *Let W be a random variable decomposed as in* (4.3)–(4.7)*. Suppose that:*

$$|X_i| \le A_i,\ |Z_{ik}| \le B_{ik},\ |V_{ik}| \le C_{ik},\ |Z_i + V_{ik}| \le C'_{ik} \tag{5.11}$$

for some constants A_i, B_{ik}, C_{ik} and C'_{ik}. Then:

$$\sup_{w\in\mathbb{R}} \left|\mathbb{P}(W \le w) - \Phi(w)\right| \le 13.7 \sum_{i\in I} A_i B_i^2 + \sum_{i\in I} \sum_{k\in K_i} A_i B_{ik} (6.8\, C_{ik} + 9.3\, C'_{ik}) \tag{5.12}$$

where $B_i := \sum_{k\in K_i} B_{ik}$.

Proof. Similarly as in (4.17), we have:

$$\left|\mathbb{E}\, f_{w,\varepsilon}(W) - \mathcal{N} f_{w,\varepsilon}\right| = \sum_{i\in I} \bigl(R_{3i} h'_{w,\varepsilon} - R_{1i} h'_{w,\varepsilon} - R_{2i} h'_{w,\varepsilon}\bigr) \tag{5.13}$$

where $f_{w,\varepsilon}$ and $h_{w,\varepsilon}$ are as before, where:

$$\begin{aligned} R_{1i} g &= \mathbb{E}\bigl(g(W_i + \theta_1 Z_i) - g(W_i)\bigr) X_i Z_i \\ R_{2i} g &= \sum_{k\in K_i} \mathbb{E}\bigl(g(W_i) - g(W_{ik})\bigr) X_i Z_{ik} \\ R_{3i} g &= \sum_{k\in K_i} \mathbb{E}\bigl(g(W) - g(W_{ik})\bigr)\, \mathbb{E}\, X_i Z_{ik} \end{aligned} \tag{5.14}$$

and where θ_1 is uniformly distributed over $[0,1]$ and independent of all other variates. Notice that $h'_{w,\varepsilon} = g_{w,\varepsilon} + f_{w,\varepsilon}$, where:

$$g_{w,\varepsilon}(x) = h_{w,\varepsilon}(x)x \tag{5.15}$$

Now write:

$$\begin{aligned} g_{w,\varepsilon}(x) - g_{w,\varepsilon}(y) = h_{w,\varepsilon}(x)(x-z) - h_{w,\varepsilon}(y)(y-z) + \\ + \big(h_{w,\varepsilon}(x) - h_{w,\varepsilon}(y)\big)z \end{aligned} \tag{5.16}$$

and notice that $\sup_{x\in\mathbb{R}} |h_{w,\varepsilon}(x)| \le \sqrt{2\pi}/4$ and that $M_1(h_{w,\varepsilon}) \le 1$ (see Lemma 2 in Chapter II of Stein, 1986). Choosing $z = W$, a straightforward calculation then shows that:

$$|R_{1i} g_{w,\varepsilon}| \le \left(\frac{3}{8}\sqrt{2\pi} + \frac{1}{2}\,\mathbb{E}\,|W|\right) A_i B_i^2 \le \left(\frac{3}{8}\sqrt{2\pi} + \frac{1}{2}\right) A_i B_i^2 \tag{5.17}$$

noting that $\mathbb{E}\,|W| \le (\mathbb{E}\,W^2)^{1/2} = 1$. Similar estimates can be derived for $R_{2i} g_{w,\varepsilon}$ and $R_{3i} g_{w,\varepsilon}$. Summing up and adding over i, we obtain as a result:

$$\begin{aligned} \sum_{i\in I}\sum_{s=1}^{3} |R_{si} g_{w,\varepsilon}| \le \beta_1 := \frac{4+5\sqrt{2\pi}}{8} \sum_{i\in I} A_i B_i^2 + \\ + \sum_{i\in I}\sum_{k\in K_i} A_i B_{ik}\left(C_{ik} + \frac{2+\sqrt{2\pi}}{2}\,C'_{ik}\right) \end{aligned} \tag{5.18}$$

Now we shall estimate the most delicate part, i. e. $R_{si} f_{w,\varepsilon}$. First observe that because of the monotonicity of $f_{w,\varepsilon}$,

$$\begin{aligned} |R_{1i} f_{w,\varepsilon}| &\le \mathbb{E}\big(f_{w,\varepsilon}(W-B_i) - f_{w,\varepsilon}(W+B_i)\big) A_i B_i \\ |R_{2i} f_{w,\varepsilon}| &\le \sum_{k\in K_i} \mathbb{E}\big(f_{w,\varepsilon}(W-B_i-C_{ik}) - f_{w,\varepsilon}(W+B_i+C_{ik})\big) A_i B_{ik} \\ |R_{3i} f_{w,\varepsilon}| &\le \sum_{k\in K_i} \mathbb{E}\big(f_{w,\varepsilon}(W-C'_{ik}) - f_{w,\varepsilon}(W+C'_{ik})\big) A_i B_{ik} \end{aligned} \tag{5.19}$$

Now write:

$$\begin{aligned} f_{w,\varepsilon}(x) - f_{w,\varepsilon}(y) &= \int_0^1 f'_{w,\varepsilon}\big((1-\theta)x + \theta y\big)(x-y)\,d\theta = \\ &= \frac{y-x}{2\varepsilon}\int_0^1 \mathbf{1}\big(w-\varepsilon \le (1-\theta)y + \theta x \le w+\varepsilon\big)\,d\theta \end{aligned} \tag{5.20}$$

Hence,

$$|R_{1i} f_{w,\varepsilon}| \le \frac{1}{\varepsilon}\int_0^1 \mathbb{P}\big(w-\varepsilon \le W + (2\theta-1)B_i \le w+\varepsilon\big)\,d\theta\, A_i B_i^2 \tag{5.21}$$

Denoting:

$$\delta := \sup_{w\in\mathbb{R}} \big|\mathbb{P}(W \le w) - \Phi(w)\big| \tag{5.22}$$

the random variable W satisfies the following concentration inequality:

$$\mathbb{P}(a \leq W \leq b) \leq 2\delta + \frac{b-a}{\sqrt{2\pi}} \tag{5.23}$$

Therefore,

$$|R_{1i} f_{w,\varepsilon}| \leq 2\left(\frac{1}{\sqrt{2\pi}} + \frac{\delta}{\varepsilon}\right) A_i B_i^2 \tag{5.24}$$

Analogously, we can estimate $|R_{2i} f_{w,\varepsilon}|$ and $|R_{3i} f_{w,\varepsilon}|$. Collecting all three estimates and adding over i, we obtain:

$$\sum_{i\in I} \sum_{s=1}^{3} |R_{si} f_{w,\varepsilon}| \leq 2\left(\frac{1}{\sqrt{2\pi}} + \frac{\delta}{\varepsilon}\right) \beta_2 \tag{5.25}$$

where:

$$\beta_2 := 2\sum_{i\in I} A_i B_i^2 + \sum_{i\in I} \sum_{k\in K_i} A_i B_{ik}(C_{ik} + C'_{ik}) \tag{5.26}$$

Combining (5.6), (5.13), (5.18) and (5.25), we obtain:

$$\delta \leq \beta_1 + \frac{2}{\sqrt{2\pi}}\beta_2 + \frac{2\delta}{\varepsilon}\beta_2 + \frac{\varepsilon}{\sqrt{2\pi}} \tag{5.27}$$

Choosing $\varepsilon := 4\beta_2$, we obtain:

$$\delta \leq \beta_1 + \frac{6}{\sqrt{2\pi}}\beta_2 + \frac{\delta}{2} \tag{5.28}$$

Therefore,

$$\delta \leq 2\beta_1 + \frac{12}{\sqrt{2\pi}}\beta_2 \tag{5.29}$$

which, together with some numerical estimates, yields the desired result. □

For locally dependent and uniformly bounded random variables, we obtain the following result:

Corollary 5.2 *Let $W := X_1 + \ldots + X_n$ be a sum of random variables with a dependence graph of maximum degree D. As before, suppose that $\mathbb{E}\, X_i = 0$ and* $\mathrm{var}(W) = 1$. *Furthermore, suppose that for all i, $|X_i| \leq B$ for some $B > 0$. Then for all $w \in \mathbb{R}$,*

$$\big|\mathbb{P}(W \leq w) - \Phi(w)\big| \leq 40n(D+1)^2 B^3 \tag{5.30}$$

This result is due to Rinott (1994). In fact, an estimate with the constant 27 instead of 40 follows easily from Theorem 2.2 in Rinott (1994), noting that $1 = \mathrm{var}(W) \leq n(D+1)B^2$. Further extensions of Theorem (5.2) are given by Rinott and Rotar (1996), where the results are formulated in the multivariate context. A similar argument can also be applied to the antivoter model and degenerate U-statistics (see Rinott and Rotar, 1997).

The constants obtained in (5.12) are relatively large and could be improved to some extent. In particular, instead of the simple, crude estimates applied in (5.19), one could consider the absolute value of the conditional expectations given X_i and Z_{ik} rather than the unconditional expectation of the absolute value. This variant of the argument requires an additional assumption that the random variables W_{ik} have the same distribution as W. This causes no loss of generality: if W_{ik} is independent of (X_i, Z_{ik}), it is always possible to construct a random variable W'_{ik} (on an extended probability space) which is also independent of (X_i, Z_{ik}), has the same distribution as W and is sufficiently close to W_{ik}.

Apart from better constants, the approach mentioned above is also more flexible. It can be, for instance, applied to certain random graph counts, such as the number of isolated trees (see Barbour, 1982; Cf. Barbour, Karoński, and Ruciński, 1989), where Theorem 5.1 does not work. Details may appear somewhere else.

It would also be interesting to consider the cases where the random variables such as X_i and Z_{ik} are not bounded and to derive a Berry–Esséen type estimate of the error in the CLT in terms of their third moments. In this case, the 'bootstrapping' argument fails. Instead, one has to consider conditional distributions of certain perturbations of W separately. This leads to a more complicated inductive argument. Bolthausen (1984) uses this type of argument to prove the Berry–Esséen theorem and to derive a combinatorial central limit theorem for linear rank statistics (see Theorem 6.2). The same technique is used by Götze (1991) and Bolthausen and Götze (1993) to derive a multivariate central limit theorem for independent random vectors and approximately linear rank statistics, respectively.

In the general setting, the inductive approach turns out to be rather complicated. For the sake of simplicity, we shall only demonstrate the method to prove the *Berry–Esséen theorem*:

Theorem 5.3 *Let $X_1, \dots X_n$ be independent random variables with sum W. Suppose that $\mathbb{E}\, X_i = 0$ and $\operatorname{var}(W) = 1$. Then for all $w \in \mathbb{R}$,*

$$\left|\mathbb{P}(W \le w) - \Phi(w)\right| \le K \sum_{i=1}^{n} \mathbb{E}\,|X_i|^3 \tag{5.31}$$

where K is a universal constant.

Regarding possible extensions to dependent random variables, the main feature of the inductive approach is that it is likely to give rise to results of the same quality as Theorem 5.3. In particular, for local dependence, the following result seems likely to be true:

Conjecture 5.4 *Let $W := \sum_{i \in I} X_i$ be a sum of random variables with a dependence graph of maximum degree D. As usual, suppose that $\mathbb{E}\, X_i = 0$ and $\operatorname{var}(W) = 1$. Then for all $w \in \mathbb{R}$,*

$$\left|\mathbb{P}(W \le w) - \Phi(w)\right| \le K(D+1)^2 \sum_{i \in I} \mathbb{E}\,|X_i|^3 \tag{5.32}$$

where K is a universal constant.

Proof of Theorem 5.3. Given $n \in \mathbb{N}$, let K_n be the greatest number K, such that (5.31) holds for any sum of n independent random variables centered and scaled as in Theorem 5.3. By Jensen's inequality,

$$\sup_{w\in\mathbb{R}} \bigl|P(W \le w) - \Phi(w)\bigr| \le 1 = \left(\sum_{i=1}^{n} \mathbb{E}\,|X_i|^2\right)^{3/2} \le \sqrt{n}\sum_{i=1}^{n} \mathbb{E}\,|X_i|^3 \tag{5.33}$$

so that, trivially, $K_n \le \sqrt{n}$ and is therefore finite. However, it has to be shown that the constants K_n are uniformly bounded.

Similarly as in (4.11)–(4.15), we have:

$$\begin{aligned}\mathbb{E}\, f_{w,\varepsilon}(W) - \mathcal{N}f &= \mathbb{E}\Bigl[h'_{w,\varepsilon}(W) - h_{w,\varepsilon}(W)W\Bigr] = \\ &= \sum_{i=1}^{n} \mathbb{E}\Bigl[h'_{w,\varepsilon}(W)\sigma_i^2 - h_{w,\varepsilon}(W)X_i\Bigr] = \\ &= \sum_{i=1}^{n} \mathbb{E}\Bigl[h'_{w,\varepsilon}(W)\sigma_i^2 - h_{w,\varepsilon}(W_i)X_i - h'_{w,\varepsilon}(W_i+\theta_1 X_i)X_i^2\Bigr]\end{aligned} \tag{5.34}$$

where $\sigma_i^2 := \mathrm{var}(X_i)$, $W_i = W - X_i$ and where θ_1 is uniformly distributed over $[0,1]$ and independent of all other variates. Because of independence, the second term vanishes, so that:

$$\mathbb{E}\, f_{w,\varepsilon}(W) - \mathcal{N}f_{w,\varepsilon} = \sum_{i=1}^{n} R_i h'_{w,\varepsilon} \tag{5.35}$$

where:

$$R_i g := \mathbb{E}\Bigl[g(W)\sigma_i^2 - g(W_i+\theta_1 X_i)X_i^2\Bigr] \tag{5.36}$$

Again by independence, one can write:

$$R_i g = R_{1i} g - R_{2i} g \tag{5.37}$$

where:

$$\begin{aligned}R_{1i} g &:= \mathbb{E}\Bigl[g(W) - g(W_i)\Bigr]\sigma_i^2 \\ R_{2i} g &:= \mathbb{E}\Bigl[g(W_i+\theta_1 X_i) - g(W_i)\Bigr]X_i^2\end{aligned} \tag{5.38}$$

As in the proof of Theorem 5.1, observe that $h'_{w,\varepsilon} = f_{w,\varepsilon} + g_{w,\varepsilon}$, where $g_{w,\varepsilon}$ is as in (5.15). The quantities $R_{is} g_{w,\varepsilon}$ will be estimated similarly as in (5.16)–(5.17): choosing $z = W_i$, we obtain:

$$\begin{aligned}|R_{1i} g_{w,\varepsilon}| &\le \left(\frac{\sqrt{2\pi}}{4} + \mathbb{E}\,|W_i|\right)\sigma_i^2\, \mathbb{E}\,|X_i| \le \frac{4+\sqrt{2\pi}}{4}\,\mathbb{E}\,|X_i|^3 \\ |R_{2i} g_{w,\varepsilon}| &\le \left(\frac{\sqrt{2\pi}}{8} + \frac{1}{2}\,\mathbb{E}\,|W_i|\right)\mathbb{E}\,|X_i|^3 \le \frac{4+\sqrt{2\pi}}{8}\,\mathbb{E}\,|X_i|^3\end{aligned} \tag{5.39}$$

noting that $\mathbb{E}\,|W_i| \le (\mathrm{var}(W_i))^{1/2} \le (\mathrm{var}(W))^{1/2} = 1$.

To estimate $R_i f_{w,\varepsilon}$, first write:

$$R_i f_{w,\varepsilon} = \mathbb{E}\big(f_{w,\varepsilon}(W) - \tfrac{1}{2}\big)\sigma_i^2 - \mathbb{E}\big(f_{w,\varepsilon}(W_i + \theta_1 X_i) - \tfrac{1}{2}\big)X_i^2 \tag{5.40}$$

Since $0 \le f_{w,\varepsilon} \le 1$, we have:

$$|R_i f_{w,\varepsilon}| \le \sigma_i^2 \tag{5.41}$$

On the other hand, $R_i f_{w,\varepsilon}$ can be estimated by means of (5.20). Here we deviate from the argument of (5.19). Instead of using boundedness (which is no longer assumed), we consider, as already mentioned, conditional expectations, leading to:

$$\begin{aligned}
|R_{1i} f_{w,\varepsilon}| &\le \frac{1}{2\varepsilon}\,\mathbb{E}\Big[\mathbb{P}\big(w-\varepsilon \le W_i + \theta_2 X_i \le w+\varepsilon \;\big|\; X_i, \theta_2\big)\sigma_i^2 |X_i|\Big] \\
|R_{2i} f_{w,\varepsilon}| &\le \frac{1}{2\varepsilon}\,\mathbb{E}\Big[\mathbb{P}\big(w-\varepsilon \le W_i + \theta_1\theta_2 X_i \le w+\varepsilon \;\big|\; X_i, \theta_1, \theta_2\big)|X_i|^3\Big]
\end{aligned} \tag{5.42}$$

where θ_2 is another random variable which is uniformly distributed over $[0,1]$ and independent of everything else.

Now we shall apply the inductive hypothesis. Noting that $\operatorname{var}(W_i)^{-1/2} W_i$ is also a sum of $n-1$ (and hence n) centered independent random variables and has variance 1 and using the definition of K_n, we have, similarly as in (5.23):

$$\begin{aligned}
\mathbb{P}(a \le W_i \le b) &\le \frac{b-a}{\sqrt{2\pi}} + 2\sup_{x\in\mathbb{R}} \big|\mathbb{P}(W_i \le x) - \Phi(x)\big| = \\
&= \frac{b-a}{\sqrt{2\pi}} + 2\sup_{x\in\mathbb{R}} \big|\mathbb{P}\big(\operatorname{var}(W_i)^{1/2} W_i \le x\big) - \Phi(x)\big| \le \\
&\le \frac{b-a}{\sqrt{2\pi}} + \frac{2}{(1-\sigma_i^2)^{3/2}}\,K_n\beta
\end{aligned} \tag{5.43}$$

where:

$$\beta = \sum_{i=1}^{n} \mathbb{E}\,|X_i|^3 \tag{5.44}$$

Combining (5.42) and (5.43), and using independence, we obtain:

$$|R_i f_{w,\varepsilon}| \le \frac{3}{2}\left(\frac{1}{\sqrt{2\pi}} + \frac{\beta K_n}{\varepsilon(1-\sigma_i^2)^{3/2}}\right)\mathbb{E}\,|X_i|^3 \tag{5.45}$$

which together with (5.41) implies:

$$|R_i f_{w,\varepsilon}| \le \frac{3}{2\sqrt{2\pi}}\,\mathbb{E}\,|X_i|^3 + \min\left\{\sigma_i^2, \frac{3\beta K_n}{2\varepsilon}\,\mathbb{E}\,|X_i|^3(1-\sigma_i^2)^{-3/2}\right\} \tag{5.46}$$

It can be shown that $\min\big\{a, x(1-\sigma^2)^{-3/2}\big\} \le x + \sqrt{3/2}\,a\sigma$, so that:

$$|R_i f_{w,\varepsilon}| \le \frac{3}{2}\left(\frac{1}{\sqrt{2\pi}} + \frac{\beta K_n}{\varepsilon}\right)\mathbb{E}\,|X_i|^3 + \sqrt{\frac{3}{2}}\,\sigma_i^3 \tag{5.47}$$

Collecting (5.39) and (5.47) and taking some numerical estimates, we obtain:

$$|R_i h'_{w,\varepsilon}| \left(4.27 + 1.5\,\frac{\beta K_n}{\varepsilon}\right)\mathbb{E}\,|X_i|^3 \tag{5.48}$$

Combining this with (5.6) and (5.35), we obtain:

$$\sup_{w\in\mathbb{R}} \left|\mathbb{P}(W \le w) - \Phi(w)\right| \le \left(4.27 + 1.5\,\frac{\beta K_n}{\varepsilon}\right)\beta + \frac{\varepsilon}{\sqrt{2\pi}} \tag{5.49}$$

Taking the minimum over ε, dividing by β and taking the supremum over all sums W of n centered independent centered random variables with $\operatorname{var}(W) = 1$, we find that:

$$K_n \le 4.27 + 1.55\sqrt{K_n} \tag{5.50}$$

Since K_n is finite, (5.50) implies $K_n \le 8.9$, As a result, we have proved that Theorem 5.3 holds with $K = 8.9$. □

Remark. The constant $K = 8.9$ is far from being optimal. It has been shown by Van Beek (1972) that it can be reduced to 0.7975; the optimal constant is, to the best of the authors's knowledge, not yet known (but bounded from below by $(\sqrt{10}+3)/(6\sqrt{2\pi}) \approx 0.409$, which is the best asymptotic constant for the i. i. d. case: see Bhattacharya and Ranga Rao, 1986). Although the calculations given above can be improved to some extent, it seems unlikely that Van Beek's bound can be attained this way; the best constant obtained up to now by Stein's method is 4.1 and was obtained by Chen and Shao (2001). On the other hand, Van Beek's result was proved by an entirely different argument based on characteristic functions. Those methods, however, do not seem amenable for the case of dependent random variables in such a generality as, for instance, Theorem 4.1.

6 Applications

6.1 Nash equilibria

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 $V_{\mathbf{i}}^{(k)} \ge V_{\mathbf{j}}^{(k)}$ 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 N the number of Nash equilibria.

We shall consider games with random payoffs, such that the payoff vectors $V_{\mathbf{i}} = \left(V_{\mathbf{i}}^{(1)}, \dots V_{\mathbf{i}}^{(p)}\right)$, $\mathbf{i} \in \{1, \dots s\}^p =: I$, are independent and identically distributed. Rinott and Scarsini (2000) investigate asymptotic distribution of N under several conditions on the distribution of $V_{\mathbf{i}}$. In this paper, we focus on sufficient conditions for convergence to the normal distribution.

Writing:

$$N = \sum_{\mathbf{i}\in I} Y_{\mathbf{i}} \tag{6.1}$$

where:

$$Y_{\mathbf{i}} := \begin{cases} 1 & \text{if } \mathbf{i} \text{ is a Nash equilibrium point} \\ 0 & \text{otherwise} \end{cases} \tag{6.2}$$

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. Clearly, the same holds for $X_{\mathbf{i}} := \operatorname{var}(N)^{-1/2}(Y_{\mathbf{i}} - \mathbb{E}\, Y_{\mathbf{i}})$. Noting that Γ is a regular graph with vertices of degree $(s-1)p + \binom{p}{2}(s-1)^2 \le (sp)^2 - 1$, Theorems 4.2 and 5.2 yield:

$$\begin{aligned} \bigl|\mathbb{E}\, f(W) - \mathcal{N} f\bigr| &\le \frac{7}{2} M(f)(sp)^4 s^p \operatorname{var}(N)^{-3/2} Q \\ \bigl|\mathbb{P}(W \le w) - \Phi(w)\bigr| &\le 40\, (sp)^4 s^p \operatorname{var}(N)^{-3/2} Q \end{aligned} \tag{6.3}$$

where:

$$W := \operatorname{var}(N)^{-1/2}(N - \mathbb{E}\, N) = \sum_{\mathbf{i} \in I} X_{\mathbf{i}} \tag{6.4}$$

and 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:

$$\rho := \frac{\operatorname{cov}\bigl(V_{\mathbf{i}}^{(k)}, V_{\mathbf{i}}^{(l)}\bigr)}{\operatorname{var}\bigl(V_{\mathbf{i}}^{(k)}\bigr)^{1/2} \operatorname{var}\bigl(V_{\mathbf{i}}^{(l)}\bigr)^{1/2}} > 0 \tag{6.5}$$

Rinott and Scarsini (2000) show that $\operatorname{var}(N) \ge c(\rho) s^p Q$ for some $c(\rho) > 0$, provided that sp is large. Hence,

$$\begin{aligned} \bigl|\mathbb{E}\, f(W) - \mathcal{N} f\bigr| &\le C_1(\rho)\, M(f)\, \frac{(sp)^4}{s^{p/2} Q^{1/2}} \\ \bigl|\mathbb{P}(W \le w) - \Phi(w)\bigr| &\le C_2(\rho)\, \frac{(sp)^4}{s^{p/2} Q^{1/2}} \end{aligned} \tag{6.6}$$

for some $C_1(\rho), C_2(\rho) > 0$. The above result appears in Rinott and Scarsini (2000), together with the study of asymptotic behavior of Q.

6.2 Linear rank statistics

Consider a random variable W of the following form:

$$W := \sum_{i=1}^{n} a(i, \pi(i)) \tag{6.7}$$

where $a(i,k) \in \mathbb{R}$, $i \in \mathbb{N}_n$, $k \in \mathbb{N}_N$, $N \ge n$ and where π is a random mapping drawn from the uniform distribution over all injections from $\mathbb{N}_n$ to $\mathbb{N}_N$; for $r \in \mathbb{N}$, we denote $\mathbb{N}_r := \{1, 2, \dots r\}$. Without loss of generality, we can (and will) assume that for all $i \in \mathbb{N}_n$,

$$\sum_{j=1}^{n} a(i,j) = 0 \tag{6.8}$$

In addition, we shall assume that:

$$\operatorname{var}(W) = \frac{1}{N-1} \sum_{i=1}^{n} \sum_{j=1}^{N} a(i,j)^2 - \frac{1}{N(N-1)} \left(\sum_{i=1}^{n} a(i,j) \right)^2 = 1 \tag{6.9}$$

Observe that the random variables $\pi(i)$ are exchangeable and therefore cannot posess the structure of local dependence. Nevertheless, decompositions (4.5)–(4.7) can still be constructed in the following way. Firstly, for any two finite sets $I, J \subset \mathbb{N}$ of the same cardinality, let $\tau_{I,J}\colon\ \mathbb{N} \to \mathbb{N}$ be the map which maps the r-th element of $I \setminus J$ to the r-th element of $J \setminus I$ and vice versa; the other elements are left unchanged. Thus $\tau_{I,J}$ maps the set I bijectively onto J and $\mathbb{N} \setminus I$ onto $\mathbb{N} \setminus J$. The following lemma is straightforward and is therefore left without proof.

Lemma 6.1 *Let $A, B \subset \mathbb{N}$ with $|A| \le |B| < \infty$, where $|\cdot|$ denotes cardinality, and let π be a uniformly distributed random injection $A \to B$. Then for any subset $I \subset A$ and any random set $J \subset B$ drawn from the uniform distribution over $\{J \subset B : |J| = |I|\}$ and independent of π, the random mapping:*

$$\tau_{\pi(I),J} \circ \pi\colon\ A \setminus I \to B \tag{6.10}$$

is independent of the family $\{\pi(i) : i \in I\}$.

Now let J_1 and J_2 be random sets drawn from the uniform distribution over all subsets of $\mathbb{N}_N$ with one, respectively two elements, independent of each other and jointly independent of π. Using Lemma 6.1, observe that (4.5)–(4.7) can be satisfied by putting:

$$\begin{aligned}
&X_i := a(i, \pi(i)), \quad && W_i := \sum_{k \in I \setminus \{i\}} a\big(k, \tau_{\{\pi(i)\},J_1}(\pi(k))\big), && K_i := \mathbb{N}_n \\
&Z_{ii} := a(i, \pi(i)), \quad && Z_{ik} := a(k, \pi(k)) - a\big(k, \tau_{\{\pi(i)\},J_1}(\pi(k))\big) && ;\ k \ne i \\
&W_{ii} := W_i, \quad && W_{ik} := \sum_{l \in I \setminus \{i,k\}} a\big(l, \tau_{\{\pi(i),\pi(k)\},J_2}(\pi(l))\big) && ;\ k \ne i
\end{aligned} \tag{6.11}$$

Theorem 6.2 *For a random variable W satisfying* (6.7)–(6.9), *we have:*

$$\big|\mathbb{E}\, f(W) - \mathcal{N} f\big| \le M(f) \left[\frac{3}{2} + 20\frac{n}{N} + 24\left(\frac{n}{N}\right)^2\right] \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{N} |a(i,j)|^3 \tag{6.12}$$

$$\sup_{w \in \mathbb{R}} \big|\mathbb{P}(W \le w) - \Phi(w)\big| \le K\, \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{N} |a(i,j)|^3 \tag{6.13}$$

for a universal constant K.

Remark. When $N \to \infty$ and n remains fixed, the random variables $a(i, \pi(i))$ are almost independent; notice that the bound in (6.12) approaches the bound in (3.43).

Remark. Extensions of Theorem 6.2 to to the multivariate and functional setting have also been undertaken (see Bolthausen and Götze, 1993, and Barbour, 1990).

Proof of Theorem 6.2. We shall only prove (6.12); for the proof of (6.13), see Bolthausen (1984). To prove (6.12), observe that for any distinct $i, j, l \in \mathbb{N}_n$,

$$\begin{aligned}
|Z_{ik}| &\le \mathbf{1}(\pi(k) \in J_1)\big(|a(k, \pi(i))| + |a(k, \pi(k))|\big) \\
Z_{ik} Z_{il} &= V_{ii} = 0, \\
|Z_i + V_{ik}| &\le |a(i, \pi(i))| + |a(k, \pi(k))| + \\
&\qquad + \sum_{l=1}^{n} \mathbf{1}(\pi(l) \in J_2)\big(|a(l, \pi(i))| + |a(l, \pi(k))| + |a(l, \pi(l))|\big)
\end{aligned} \tag{6.14}$$

Theorem 4.1 yields, after a straightforward calculation:

$$
\begin{aligned}
\bigl|\mathbb{E}\, f(W) - \mathcal{N} f\bigr| \le M(f) \Biggl[& \sum_{i \in \mathbb{N}_n} \Bigl(\mathbb{E}\,|X_i|^3 + 2\,\mathbb{E}\,|X_i|^2\,\mathbb{E}\,|X_i| \Bigr) + \\
& + \sum_{i \in \mathbb{N}_n} \sum_{k \in \mathbb{N}_n \setminus \{i\}} \Bigl(4\,\mathbb{E}\,|X_i^2 Z_{ik}| + 3\,\mathbb{E}\,|X_i Z_{ik}^2| + 2\,\mathbb{E}\,|X_i|^2\,\mathbb{E}\,|Z_{ik}| + \\
& + 2\,\mathbb{E}\,|X_i Z_{ik}(Z_i + V_{ik})| + 2\,\mathbb{E}\,|X_i Z_{ik}|\,\mathbb{E}\,|Z_i + V_{ik}| \Bigr) \Biggr]
\end{aligned}
\tag{6.15}
$$

The proof is now completed by noting that:

$$
\mathbb{P}(\pi(k) \in J_1 \mid \pi) = \frac{1}{N}, \quad \mathbb{P}(\pi(l) \in J_2 \mid \pi, J_1) = \frac{2}{N} \tag{6.16}
$$

and applying the inequality $xyz \le \frac{1}{3}(x^3 + y^3 + z^3)$ holding for any $x, y, z \ge 0$. □

References

[1] Allgower, E.L. and Georg, K. (1990): Numerical Continuation Methods. An Introduction. Berlin: Springer-Verlag.

[2] Arratia, R., Goldstein, L., and Gordon, L. (1990): Poisson approximation and the Chen–Stein method. *Stat. Science*, **5**, 403–434.

[3] Arratia, R., Gordon, L., and Waterman, L. S. (1990): The Erdös–Renyi law in distribution, for coin tossing and sequence matching. *Stat. Science*, **5**, 403–434.

[4] Baldi, P., Rinott, Y., and Stein, C. (1989): On normal approximations for the number of local maxima of a random function on a graph. *Probability, Statistics, and Mathematics*, 58–81. Papers in honor of Samuel Karlin, Boston: Academic Press.

[5] Barbour, A.D. (1986): Asymptotic expansions based on smooth test functions in the central limit theorem. *Probab. Theory Related Fields*, **72**, 289–303

[6] Barbour, A.D. (1982): Poisson convergence and random graphs.*Math. Proc. Cambridge Philos. Soc*, **92**, 349–359.

[7] Barbour, A.D. (1990): Stein's method for diffusion approximations, *Probab. Theory Related Fields*, **84**, 297–322.

[8] Barbour, A.D. and Čekanavičius, V. (2002): Total variation asymptotics for sums of independent integer random variables. *Ann. Probab.*, **30**, 509–545.

[9] Barbour, A.D., Chen, L.H.Y., and Loh, W.–L. (1992): Compound Poisson approximation for non-negative random variables via Stein's method. *Ann. Probab.*, **20**, 1843–1866.

[10] Barbour, A.D. and Eagleson, G.K. (1985): Multiple comparisons and sums of dissociated random variables. *Adv. Appl. Probab.*, **17**, 147–162.

[11] Barbour, A.D. and Eagleson, G.K. (1986): Random association of symmetric arrays. *Stoch. Analysis Appl.*, **4**, 239–281.

[12] Barbour, A.D., Holst, L., and Janson, S. (1992): *Poisson Approximation.* Oxford: Clarendon Press.

[13] 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.

[14] Barbour, A.D., and Månsson, M. (2002): Compound Poisson process approximation. *Ann. Probab.*, **30**, 1492–1537.

[15] Barbour, A.D., and Xia, A. (2000): Estimating Stein's constants for compound Poisson approximation. *Bernoulli*, **6**, 581–590.

[16] Beek, P. Van (1972): An application of the Fourier method to the sharpening of the Berry–Esséen inequality. *Z. Wahrsch. Verw. Gebiete*, **23**, 187–197.

[17] Bhattacharya, R.N., and Ranga Rao, R. (1986): *Normal Approximation and Asymptotic Expansion.* Melbourne: Krieger.

[18] Bolthausen, E. (1984): An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete*, **66**, 379–386.

[19] Bolthausen, E., and Götze, F. (1993): The rate of convergence for multivariate sampling statistics. *Ann. Statist.*, **2**, 1692–1710.

[20] Chen, L.H.Y. (1975): Poisson approximation for dependent trials. *Ann. Probab.*, **3**, 534–545.

[21] Chen, L.H.Y. (1986): The rate of convergence in a central limit theorem for dependent random variables with arbitrary index set. *IMA Preprint Series #243, Univ. Minnesota.*

[22] Chen, L.H.Y. (1978): Two central limit problems for dependent random variables. *Z. Wahrsch. Verw. Gebiete*, **43**, 223–243.

[23] Chen, L.H.Y. and Ho, S.–T (1978): An L_p bound for the remainder in a combinatorial central limit theorem. *Ann. Probab.*, **6**, 231–249.

[24] Chen, L.H.Y. and Shao, Q.–M. (2001): A non-uniform Berry–Esséen bound via Stein's method. *Probab. Theory Related Fields*, **120**, 236–254

[25] Chryssaphinou, O. and Papastavridis, S. (1988): A limit theorem for the number of non-overlapping occurences of a pattern in a sequence of independent trials. *J. Appl. Probab.*, **25**, 428–431.

[26] Dembo, A., and Rinott, Y. (1996): Examples of normal approximations by Stein's method. Random Discrete Structures, 25–44, *IMA Vol. Math. Appl.*. **76**, New York: Springer-Verlag.

[27] Diaconis, P. (1989): *An Example for Stein's Method*. Stanford Stats Dept Techn. Report.

[28] Durrett, R. (1996): *Probability: Theory and Examples.* Second edition. Belmont: Duxbury Press.

[29] Ethier, S.N. and Kurtz, T.G. (1986): *Markov Processes: Characterization and Convergence*. New York: John Wiley & Sons.

[30] Glaz, J., Naus, J., Róos, M., and Wallenstein, S. (1994): Poisson approximations for distribution and moments of ordered m-spacings. *J. Appl. Probab.*, **31**, 271–281.

[31] Godbole, A.P. (1993): Approximate reliabilities of m-consecutive-k-out-of-n: Failure systems. *Statist. Sinica*, **3**, 321–328.

[32] Goldstein, L. and Reinert, G. (1997): Stein's method and the zero bias transformation with applications to simple random sampling. *Ann. Appl. Probab.*, **7**, 935–952.

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

[34] Götze, F. (1991): On the rate of convergence in the multivariate CLT, *Ann. Probab.*, **19**, 724–739.

[35] Lindvall, T. (1992): Lectures on the coupling method. New York: John Wiley & Sons.

[36] Loh, W.–L. (1992): Stein's method and multinomial approximation. *Ann. Appl. Probab.*, **2**, 536–554.

[37] Luk, M. (1994): Stein's method for the gamma distribution and related statistical applications. *Ph. D. Thesis*, Univ. Southern California.

[38] Reinert, G. (1995): A weak law of large numbers for empirical measures via Stein's method. *Ann. Probab.*, **23**, 334–354.

[39] Rinott, Y. (1994): On normal approximation rates for certain sums of dependent random variables. *J. Appl. Comp. Math.*, **55**, 135–143.

[40] 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.

[41] Rinott, Y., and Rotar, V. (2000): Normal approximations by Stein's method. *Decis. Econ. Finance*, **23**, 15–29.

[42] 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.

[43] Rinott, Y., and Scarsini, M. (2000): On the number of pure strategy Nash equilibria in random games. *Games Econom. Behav.*, **33**, 274–293.

[44] Rogers, L.C.G. and Williams, D. (1994): *Diffusions, Markov Processes, and Martingales. Volume 1: Foundations.* Chichester: John Wiley & Sons.

[45] Rudin, W. (1987): *Real and Complex Analysis.* New York: McGraw-Hill.

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

[47] Smith, R.L. (1989): Extreme value theory for dependent sequences via the Stein–Chen method for Poisson approximation. *Stoch. Proc. Appl.*, **30**, 317–327.

[48] Stein, C. (1972): A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proceedings of the VIth Berkley Symposium on Mathematical Statistics and Probability, Vol. II: Probability Theory*, Berkeley: Univ. California Press, 583–602.

[49] Stein, C. (1986): *Approximate Computation of Expectations.* Hayward: IMS.

[50] Takahata, H. (1993): On the rates in the central limit theorem for weakly dependent random fields. *Z. Wahrsch. Verw. Gebiete*, **64**, 445–456.