# Couplings for Normal Approximations with Stein's Method

Gesine Reinert

Abstract. Coupling constructions for Poisson approximation using the Chen-Stein method is now a standard technique; a systematic study of related couplings for normal approximation using Stein's method has begun only a few years ago. This small survey of coupling methods for normal approximations includes size-bias couplings, which are natural for nonnegative random variables such as counts, and zero-bias couplings, which may be applied to mean zero variables and are especially useful for random variables with vanishing third moments.

## 1. Introduction

In 1972, Stein [**47**] published a very elegant method to prove normal approximations. It is based on the fact that a random variable $Z$ is standard normal if and only if for all smooth, real-valued functions $f$,

$$E\{Zf'(Z) - f''(Z)\} = 0.$$

(This is easily seen using dominated convergence and integration by parts.) Stein [**47**] then showed that for any smooth, real-valued function $h$ there is a function $f$ solving the now-called "Stein equation"

$$xf'(x) - f''(x) = h(x) - \Phi h, \tag{1.1}$$

$\Phi h$ denoting the expectation of $h$ with respect to the standard normal density. Moreover, there is a solution $f$ of the Stein equation (1.1) satisfying

$$\|f'\| \le \sqrt{\frac{\pi}{2}}\|h - \Phi h\|; \quad \|f''\| \le (\sup h - \inf h); \quad \|f^{(3)}\| \le 2\|h'\|; \tag{1.2}$$

where $\|\cdot\|$ denotes the supremum norm (see Stein [**48**], p.25 and Baldi *et al.* [**5**]). Now, for any random variable $W$, taking expectations in (1.1) gives

$$\mathbb{E}h(W) - \Phi h = \mathbb{E}Wf'(W) - \mathbb{E}f''(W). \tag{1.3}$$

Thus the distance of $W$ from the normal, in terms of a test function $h$, can be bounded by bounding the right-hand side of (1.3); the immediate bound on the distance is one of the key advantages of Stein's method compared to moment-generating functions or characteristic functions. Typical classes of test functions $h$ are $C^4(\mathbb{R})$ (for weak convergence), or the class of indicator functions of half-lines (giving the Kolmogorov distance).

---

1991 *Mathematics Subject Classification.* 60F05, 60E10, 62D05.
The author was supported in part by NSF Grant DMS-9505075.

Nearly twenty years later, Barbour **[7]** and Götze **[32]** proved similar results for more general Gaussian approximations; Barbour **[7]** considered diffusion approximations, Götze **[32]** multivariate normal approximations, with the bound

$$\|f^{(j)}\| \leq j^{-1}\|h^{(j)}\|, j = 1, 2, \ldots \tag{1.4}$$

A key tool for solving the higher-dimensional case is the generator method developed by Barbour **[6]**, **[7]**, **[8]**; the left-hand side of (1.1) can be written as $Af(x)$, where $A$ is the generator of an Ornstein-Uhlenbeck process. Thus semigroup theory can be applied to solve the generator equation. Note that the target distribution, here the standard normal, is the stationary distribution of this Markov process.

Stein's method has been generalized to many other distributions, foremost the Poisson distribution (see Chen **[21]**, Arratia *et al.* **[1]**, Barbour *et al.* **[15]**, Aldous **[2]**, to cite but a few). Other distributions include the uniform distribution (Diaconis **[26]**), the binomial distribution (Ehm **[28]**), the compound Poisson distribution (Barbour *et al.* **[9]**, Barbour and Utev **[17]**, Roos **[46]**), the multinomial distribution (Loh **[36]**), the gamma distribution (Luk **[38]**; for the $\chi^2$ distribution see also Mann **[39]**), the geometric distribution (Peköz **[40]**) and, more generally, Pearson curves (Diaconis and Zabell **[27]**, Loh **[37]**).

The most obvious advantage of Stein's method is that it yields immediate bounds on the distance. Moreover in many situations where dependence comes into play the application is straightforward; many examples are of combinatorial nature. An early success of Stein's method is the work by Bolthausen **[18]** for a combinatorial central limit theorem; he was the first to obtain the correct order for this approximation. In examples from random graph theory, where the method of moments used to be the most popular technique, Stein's method allowed not only to provide rates of convergence for the first time, but also to weaken conditions; see, for instance, Barbour *et al.* **[16]**. Another advantage of Stein's method is that it can also be used to derive lower bounds for the approximations; Hall and Barbour **[34]** applied it to give lower bounds for the rate of convergence in the central limit theorem for independent random variables.

Unfortunately such a straightforward application of Stein's method may not yield the correct order for the rate of convergence; additional work may be needed to sharpen the bounds. An example are dependency graphs; first, Baldi and Rinott **[3]** proved a normal approximation using the method of moments, without any result on the rate of convergence. Next, Baldi and Rinott **[4]** employed Stein's method to obtain a bound of the order $n^{-1/4}$; Baldi *et al.* **[5]** derived related results for the number of local maxima in a graph whose vertices are randomly ranked. About five years later, Rinott **[43]** improved the bounds considerably to the correct order $n^{-1/2}$, for bounded random variables. Shortly after, Dembo and Rinott **[25]** simplified this bound, whereas Goldstein and Rinott **[31]** and Rinott and Rotar **[44]** provide multivariate extensions. Another, earlier example that illustrates the effort in obtaining optimal bounds is Bolthausen's **[18]** proof of the Berry-Esséen theorem, the first to yield the correct order when applying Stein's method.

Evidently, Stein's method is in place when the right-hand side of (1.3) is easier to bound than the left-hand side of (1.3). A typical approach for evaluating the right-hand side of (1.3) is to employ couplings, and often the success of the method is connected with finding an effective coupling for the right-hand side of (1.3).

In what follows, we will only consider random variables $W$ that are the sum of $n$ random variables $X_1, \ldots X_n$; that is, $W = \sum_{i=1}^n X_i$. For convenience we will assume throughout that $\mathrm{Var}(W) = 1$. Moreover we will only discuss smooth test functions, because the treatment of nonsmooth test functions is slightly more technical, and the purpose of this paper is to lay out the basic methods. References for nonsmooth test functions will be given. As our main two examples, we will firstly consider that $X_1, \dots X_n$ are independent, and secondly that $X_1, \dots X_n$ is a simple random sample from a finite population. Moreover, for $\|f''\|$ in (1.2), Stein [**48**] proved the bound $2\|h - \Phi h\|$; Baldi *et al.* [**5**] showed the improved inequality in (1.2). In what follows, we will use the improved inequality for theorems we cite, even if the theorems used the first inequality.

In Section 2 we describe the perhaps most common coupling, the "local" approach. It is very effective if each $X_i$ depends only on a small number of the other $X_j, j \neq i$. Typical examples are $m$-dependent sequences.

In case of global but weak dependence between the $X_1, \dots X_n$, exchangeable pair couplings are usually more natural. This approach is discussed in Section 3.

Section 4 gives a coupling that is particularly adapted to describe counts; the size bias coupling. In Barbour *et al.* [**15**] it has been developed as a major tool for proving Poisson approximations; its importance for normal approximations has been described in Baldi *et al.* [**5**], Goldstein and Rinott [**31**], and Stein [**49**]. A drawback in the context of normal approximations is that it requires $W$ to be nonnegative, with positive mean.

For mean zero $W$, and in particular for symmetric $W$, the zero bias coupling discussed in Section 5 might give better results, especially when the test functions are smooth. The zero-bias coupling is a "second-order" refinement of exchangeable pair ideas .

Finally, Section 6 collects other couplings that work well in special cases. It illustrates that specific problems may benefit from constructing couplings that do not fit into the above classes, and thus illustrates the dynamical structure of the field.

## 2. The local approach

Suppose $X_1, \dots X_n$ are independent, mean zero, and let $\sigma_i^2 = \mathrm{Var}(X_i)$. Put $W = \sum_{i=1}^n X_i$, and let $\mathrm{Var}(W) = \sum_{i=1}^n \sigma_i^2 = 1$. For each $i = 1, \dots, n$ put

$$W_i = W - X_i = \sum_{j \neq i} X_j. \tag{2.1}$$

Thus $W$ and $W_i$ are defined on the same probability space. For any smooth function $f$ we have, by Taylor expansion

$$\begin{aligned}
\mathbb{E} W f'(W) &= \sum_{i=1}^n \mathbb{E} X_i f'(W) \\
&= \sum_{i=1}^n \mathbb{E} X_i f'(W_i) + \sum_{i=1}^n \mathbb{E} X_i^2 f''(W_i) + R,
\end{aligned}$$

where

$$|R| \leq \frac{1}{2} \|f^{(3)}\| \sum_{i=1}^n \mathbb{E}|X_i^3|.$$

Using the independence we obtain for the right-hand side of (1.3), with $\mu = 0$, that

$$\begin{aligned} \mathbb{E}Wf'(W) - \mathbb{E}f''(W) &= \sum_{i=1}^n \mathbb{E}X_i^2 \mathbb{E}f''(W_i) - \mathbb{E}f''(W) + R \\ &= \sum_{i=1}^n \sigma_i^2 \mathbb{E}\{f''(W_i) - f''(W)\} + R. \end{aligned}$$

Taylor expansion and the bounds (1.2) now give the following theorem.

THEOREM 2.1. *Suppose $X_1, \dots X_n$ are independent, mean zero, and let $\sigma_i^2 = \mathrm{Var}(X_i)$. Put $W = \sum_{i=1}^n X_i$, and let $\mathrm{Var}(W) = 1$. For any continuous, bounded function $h$ with piecewise continuous, bounded first derivative, we have*

$$|\mathbb{E}h(W) - \Phi h| \leq ||h'|| \left( 2 \sum_{i=1}^n \sigma_i^3 + \sum_{i=1}^n \mathbb{E}|X_i^3| \right).$$

This approach, first used by Stein [**47**], was employed by Bolthausen [**18**] to obtain sharp rates in the Berry-Esséen Theorem, by Barbour and Hall [**13**] for smoother metrics, by Barbour and Hall [**14**] for the non-identically distributed case, and by Hall and Barbour [**34**] for reversing the Berry-Esséen Theorem. For more general Gaussian approximations of sums of independent random elements, it was employed by Barbour [**7**] for a functional CLT, by Götze [**32**] for a multivariate CLT, and by Reinert [**41**] for a Gaussian approximation of empirical measures.

The local approach might well be the most widely applied coupling approach for Stein's method, being both effective and easy to construct in situations of local dependence; the amount of literature where it is used is large, and listing it all would be beyond the scope of this paper. As briefly mentioned in the introduction, the correct bound on the rate of convergence for bounded variates has been obtained by Rinott [**43**]; Goldstein and Rinott [**31**] and Rinott and Rotar [**44**] give multivariate extensions. Important applications include $m$-dependent sequences, where the generalization using Taylor expansion is straightforward, dependency graphs (Rinott [**43**]), and sums of dissociated random variables (Chen [**22**], Barbour and Eagleson [**10**]). Refinements with different types of neighborhoods are derived by Chen [**23**], and by Barbour *et al.* [**16**] for decomposable random variables; typical applications are graph-related statistics, see also Goldstein and Rinott [**31**].

Moreover the local approach can be seen as a special case of a conditional expectation coupling used by Stein [**47**], [**48**]. Let $(\tilde{\Omega}, \tilde{B}, \tilde{\mathbb{P}})$ be a probability space, let $B$ and $C$ be sub-$\sigma$-algebras of $\tilde{B}$, let $G$ be a $\tilde{B}$-measurable random variable such that $\mathbb{E}|G| < \infty$, and let $W^*$ be $C$-measurable. Put

$$W = \mathbb{E}(G|B).$$

Assume $\mathbb{E}W = 0$ and $\mathrm{Var}(W) = 1$. Then (see Stein [**48**], p.106, Theorem 1)

THEOREM 2.2. *Let $W = \mathbb{E}(G|B)$ be constructed as above. For any continuous, bounded function $h : \mathbb{R} \to \mathbb{R}$ with bounded, piecewise continuous derivatives, we have*

$$\begin{aligned} |\mathbb{E}h(W) - \Phi h| \leq\ & (\sup h - \inf h) \sqrt{\mathbb{E}\{1 - \mathbb{E}(G(W - W^*)|B)\}^2} \\ & + \sqrt{\frac{\pi}{2}} ||h - \Phi h|| \mathbb{E}|\mathbb{E}(G|C)| + ||h'|| \mathbb{E}|G|(W - W^*)^2. \end{aligned}$$

In the independent example, an index $I$ would be chosen uniformly from $\{1, \dots, n\}$; we would put $\tilde{B} = \sigma\{X_1, \dots, X_n, I\}$, $B = \sigma\{X_1, \dots, X_n\}$, and $C = \sigma\{X_j, j \neq I; I\}$. Furthermore put $G = nX_I$ and $W^* = W - X_I$. Then we have $W = \mathbb{E}(G|B) = \sum_{i=1}^n X_i$. The conditional expectation formulation also extends to mixing sequences (Stein [**47**], Chen [**23**]), and can be applied whenever the dependence between any $X_i$ and $X_j, j \neq i$ is strong only for a few indices $j$, and very weak for the other indices.

However, this approach does not work well for global weak dependence structures, such as given by the example of simple random sampling.

## 3. Exchangeable pair couplings

Note that the local coupling can be described in a more abstract setting as follows. Let $I$ be chosen randomly, uniformly from $\{1, \dots, n\}$ and put $W' = W_I$; recall (2.1). Assume as usual that $\mathrm{Var}(W) = 1$. Then we have for all smooth functions $f$ that

$$\mathbb{E}Wf'(W) = n\mathbb{E}(W - W')(f'(W) - f'(W')). \tag{3.1}$$

Taylor expansion gives

$$\mathbb{E}Wf'(W) - \mathbb{E}f''(W) = \mathbb{E}\{1 - n(W - W')^2\}f''(W) + R,$$

where $R$ is a remainder term that can be bounded by

$$|R| \leq \frac{n}{2}\|f^{(3)}\|\mathbb{E}|W - W'|^3.$$

Using the Cauchy-Schwarz inequality gives

$$\begin{aligned}
&|\mathbb{E}Wf'(W) - \mathbb{E}f''(W)| \\
&\leq \|f''\|\sqrt{\mathbb{E}(1 - n\mathbb{E}((W - W')^2|W)} + \frac{n}{2}\|f^{(3)}\|\mathbb{E}|W - W'|^3 \\
&\leq (\sup h - \inf h)\sqrt{\mathbb{E}(1 - n\mathbb{E}((W - W')^2|W)} + \frac{n}{2}\|h'\|\mathbb{E}|W - W'|^3.
\end{aligned}$$

Indeed, all that is required for this derivation is an equation of the type (3.1). Another method to achieve this type of equation is the method of exchangeable pairs. A pair $(W, W')$ of random variables defined on the same probability space is called *exchangeable* if for all measurable sets $B$ and $B'$,

$$\mathbb{P}(W \in B, W' \in B') = \mathbb{P}(W \in B', W' \in B).$$

Following Stein [**48**] we assume that $W$ is mean zero, variance 1, and that there is a $0 < \lambda < 1$ such that

$$\mathbb{E}(W'|W) = (1 - \lambda)W. \tag{3.2}$$

This assumption can be related to regression; if $(W, W')$ is bivariate normal with correlation $\rho$, then $1 - \lambda = \rho$. Under (3.2) it is easy to see that

$$\mathbb{E}Wf'(W) = \frac{1}{2\lambda}\mathbb{E}(W - W')(f'(W) - f'(W')),$$

so that equation (3.1) is satisfied with $n$ replaced by $\frac{1}{2\lambda}$. The same reasoning as above gives the following result (see Stein [**48**]).

THEOREM 3.1. *Let $(W, W')$ be an exchangeable pair satisfying (3.2) and assume* $\mathbb{E}W = 0, \mathrm{Var}(W) = 1$. *For any continuous, bounded function* $h : \mathbb{R} \to \mathbb{R}$ *with bounded, piecewise continuous derivatives, we have*

$$
\begin{aligned}
&|\mathbb{E}h(W) - \Phi h| \\
&\leq \quad (\sup h - \inf h)\sqrt{\mathbb{E}\left(1 - \frac{1}{2\lambda}\mathbb{E}((W - W')^2|W)\right)^2} + \frac{1}{4\lambda}\|h'\|\mathbb{E}|W - W'|^3.
\end{aligned}
$$

EXAMPLE 3.2. Let us assume that $X_1, \dots, X_n$ are independent, mean zero; let $\sigma_i^2 = \mathrm{Var}(X_i)$ and $\sum_{i=1}^n \sigma_i^2 = 1$. As usual, we consider $W = \sum_{i=1}^n X_i$. To construct $W'$ such that $(W, W')$ is an exchangeable pair, pick an index $I$ uniformly from $\{1, \dots, n\}$. If $I = i$, we replace $X_i$ by an independent copy $X_i^*$, and we put

$$W' = W - X_I + X_I^*. \tag{3.3}$$

Then $(W, W')$ is exchangeable, and

$$\mathbb{E}(W'|W) = W - \frac{1}{n}W + \mathbb{E}X_I^* = \left(1 - \frac{1}{n}\right)W,$$

so (3.2) is satisfied with $\lambda = \frac{1}{n}$. Theorem 3.1 thus gives

$$
\begin{aligned}
&|\mathbb{E}h(W) - \Phi h| \\
&\leq \quad (\sup h - \inf h)\sqrt{\mathbb{E}\left(1 - \frac{1}{2\lambda}\mathbb{E}((W - W')^2|W)\right)} + \frac{1}{4\lambda}\|h'\|\mathbb{E}|W - W'|^3.
\end{aligned}
$$

Bounding the expectations gives (see [**48**], p.37)

$$|\mathbb{E}h(W) - \Phi h| \quad \leq \quad (\sup h - \inf h)\sqrt{\sum_{i=1}^n \mathbb{E}X_i^4 - \sigma_i^4} + \frac{1}{2}\|h'\|\sum_{i=1}^n \left(\mathbb{E}|X_i|^3 + 3\sigma_i^3\right).$$

The construction used in Example 3.2 is typical for the exchangeable pair approach.

CONSTRUCTION 3.3. Let $X_1, \dots, X_n$ be possibly dependent, mean zero variates with existing variances, and let $W = \sum_{i=1}^n X_i$. Pick an index $I$ uniformly from $\{1, \dots, n\}$. If $I = i$, replace $X_i$ by an independent copy $X_i^*$. If $X_i^* = x$, construct $\hat{X}_j, j \neq i$ such that

$$\mathcal{L}(\hat{X}_j, j \neq i) = \mathcal{L}(X_j, j \neq i | X_i = x). \tag{3.4}$$

Then put

$$W' = \sum_{j \neq I} \hat{X}_j + X_I^*. \tag{3.5}$$

The pair $(W, W')$ is exchangeable. If, in addition, (3.2) holds, then Theorem 3.1 can be applied. However, not always will the exchangeable pair $(W, W')$ constructed above satisfy (3.2); see, for example, Rinott and Rotar [**45**].

Compared to Theorem 2.1 the bound obtained in Example 3.2 is worse. However, a similar construction leads to an exchangeable pair for the simple random sampling example, which caused problems in the local approach; see Stein [**48**].

EXAMPLE 3.4. Let $X_1, \dots, X_n$ be a simple random sample of size $n$ from a finite population $A$. Assume for convenience that all elements of $A$ are distinct. We follow Construction 3.3. To construct $\hat{X}_j, j \neq i$ satisfying (3.4), we choose $\hat{X}_j, j \neq i$ as a simple random sample of size $n-1$ from $A \setminus \{X_I^*\}$. In particular, if $X_I^* \notin \{X_j, j \neq i\}$, then we may choose $\hat{X}_j = X_j, j \neq i$. If $X_I^* \in \{X_j, j \neq i\}$, so that $X_I^* = X_J$, say, then let $\hat{X}_j = X_j, j \neq i, J$, and choose $\hat{X}_j$ uniformly from $A \setminus \{X_I^*, \hat{X}_j, j \neq i, J\}$. Then (3.2) is satisfied with $\lambda = \frac{2}{n-1}$. Note that $W$ and $W'$ differ for at most two summands, so that the coupling is efficient.

Other examples where Construction 3.3 works well include random permutations (Stein [**48**], Fulman [**29**], for example), random allocations (Stein [**48**]) and combinatorial central limit theorems (Bolthausen [**18**], Bolthausen and Götze [**19**]). Note that Construction 3.3, if repeated, yields a Markov chain - this relates to the generator approach developed by Barbour [**6**], [**7**], [**8**]. Indeed, this construction can be used to derive a generator associated with a target distribution; see, e.g., Reinert [**41**].

Conversely, an exchangeable pair can be constructed from a reversible Markov chain; see Rinott and Rotar [**45**]. Let $X_1, \dots, X_n$ be random variables, and suppose that $\mathcal{L}(X_1, \dots, X_n)$ is the stationary distribution of a reversible, ergodic Markov chain $(X_1(t), \dots, X_n(t))_{t=0,1,\dots}$. Let $W = W(X_1, \dots, X_n)$ be the quantity of interest. Put

$$\begin{aligned} W &= W(X_1(t), \dots, X_n(t)) \\ W' &= W(X_1(t+1), \dots, X_n(t+1)), \end{aligned}$$

then $(W, W')$ is an exchangeable pair.

Moreover, the approach is not restricted to requiring that Condition (3.2) is satisfied. Following Rinott and Rotar [**45**], assume that $(W, W')$ is an exchangeable pair such that $\mathbb{E}W = 0, \mathbb{E}W^2 = 1$, and let $R = R(W)$ be such that

$$\mathbb{E}(W'|W) = (1-\lambda)W + R \tag{3.6}$$

for some $0 < \lambda < 1$. Similarly as for Theorem 3.1, we can show that, under the above setting, for $W$ real-valued,

THEOREM 3.5. *Let $(W, W')$ be an exchangeable pair such that Condition (3.6) is satisfied, and assume $\mathbb{E}W = 0, \mathrm{Var}(W) = 1$. For any continuous, bounded function $h$ with piecewise continuous, bounded first derivative, we have*

$$\begin{aligned} |\mathbb{E}h(W) - \Phi h| &\leq (\sup h - \inf h)\sqrt{\mathbb{E}(1 - \frac{1}{2\lambda}\mathbb{E}((W-W')^2|W)} \\ &\quad + \frac{1}{4\lambda}\|h'\|\mathbb{E}|W-W'|^3 + \sqrt{\frac{\pi}{2}}\frac{1}{\lambda}\|h - \Phi h\|\mathbb{E}|R|. \end{aligned}$$

Thus, if $R$ is small, the normal approximation will be good. Examples include the anti-voter model and weighted $U$-statistics; see Rinott and Rotar [**45**].

Finally it must be remarked that the method of exchangeable pairs has a much wider range than normal approximations; in particular it can be used to estimate ratios of probabilities, see Stein [**49**]. Moreover Diaconis [**26**] employed it for for the uniform distribution, and more generally, for the convergence of a Markov chain to its stationary distribution. Recently, Mann [**39**] has applied it to yield a $\chi^2$ approximation for statistics based on the multinomial distribution. This illustrated the vast potential of the method of exchangeable pairs.

In contrast to the local approach, where we used a coupling that reduced the variability (by leaving out a summand, or by taking conditional expectations), the exchangeable pair coupling introduces additional randomness. This is a first example of what, following Stein [**49**], might be called the method of auxiliary randomization. Section 4 and Section 5 provide more examples for this method.

## 4. Size-bias couplings

There are situations where couplings other than exchangeable pairs seem more natural. A wide class of examples is provided in the context of Poisson approximations, see Barbour *et al.* [**15**], where counts are considered. Then size bias couplings seem to be more adapted. In the context of Stein's method for normal approximations, they have been explored by Baldi *et al.* [**5**], Stein [**49**], Goldstein and Rinott [**31**], Dembo and Rinott [**25**], and Reinert [**42**]. In Baldi *et al.* [**5**], and Stein [**49**], sums of $0-1$ random variables are considered. Goldstein and Rinott [**31**] generalize it to multivariate normal approximations of any sums, Reinert [**42**] uses it for empirical processes, and Dembo and Rinott [**25**] prove approximations for nonsmooth functions.

Let $W$ be a nonnegative random variable, $\mathrm{Var}(W)=1$, and $\mathbb{E}W=\mu>0$. $W^*$ is said to have the $W$-size biased distribution if, for all functions $g$ for which the expectation exists,

$$\mathbb{E}Wg(W)=\mu\mathbb{E}g(W^*).$$

Thus, if $w$ is discrete, say, then, for all $w$ we have $\mathbb{P}(W^*=w)=\frac{w}{\mu}\mathbb{P}(W=w)$. This illustrates that size biasing corresponds to sampling proportional to size; the larger a subpopulation, the more likely it is to be in the sample.

If $(W,W^*)$ are defined on the same probability space, with $W^*$ having the $W$-size biased distribution, then the right-hand side of equation (1.3) becomes

$$\begin{aligned}\mathbb{E}(W-\mu)f('W-\mu) &= \mu\mathbb{E}(f'(W^*-\mu)-f'(W-\mu))\\ &= \mu\mathbb{E}f''(W-\mu)(W^*-W)+R,\end{aligned}$$

using Taylor expansion, where

$$|R|\le\mu\|f^{(3)}\|\mathbb{E}(W^*-W)^2.$$

Note that $\mu\mathbb{E}(W^*-W)=\mathbb{E}W^2-\mu^2=\mathrm{Var}(W)=1$. Thus

$$\begin{aligned}\mu\mathbb{E}f''(W-\mu)(W^*-W)-\mathbb{E}f''(W-\mu) &= \mathbb{E}f''(W-\mu)(\mu\mathbb{E}(W^*-W|W)-1)\\ &\le \|f''\|\sqrt{\mathbb{E}(\mu\mathbb{E}(W^*-W|W)-1)^2}\\ &= \|f''\|\mu\sqrt{\mathrm{Var}\mathbb{E}(W^*-W|W)}.\end{aligned}$$

This gives (see Goldstein and Rinott [**31**])

THEOREM 4.1. *Let $W$ be nonnegative, $\mathbb{E}W=\mu>0$, and* $\mathrm{Var}(W)=1$. *Let $W^*$ have the $W$-size biased distribution. For any continuous, bounded function $h$ with piecewise continuous, bounded first derivative, we have*

$$\begin{aligned}&|\mathbb{E}h(W-\mu)-\Phi h|\\ &\le (\sup h-\inf h)\mu\sqrt{\mathrm{Var}\mathbb{E}(W^*-W|W)}+\|h'\|\mu\mathbb{E}(W^*-W)^2.\end{aligned}$$

EXAMPLE 4.2. let $X_1, \dots X_n$ be independent, nonnegative, $\mathbb{E}X_i = \mu_i > 0$, $W = \sum_{i=1}^n X_i$, $\mathbb{E}W = \mu$, $\mathrm{Var}(W) = 1$. Choose an index $I$ from $\{1, \dots, n\}$ according to

$$\mathbb{P}(I = i) = \frac{\mu_i}{\mu}, \tag{4.1}$$

that is, choose index $i$ proportionally to its expectation. If $I = i$, replace $X_i$ by $X_i^*$ having the $X_i$-size bias distribution, and put

$$W^* = W - X_I + X_I^*. \tag{4.2}$$

Then $W^*$ has the $W$-size biased distribution. (Note the similarity to (3.3)). Using Theorem 4.1 gives

$$|\mathbb{E}h(W - \mu) - \Phi h| \quad \leq \quad (\sup h - \inf h)\frac{\mu}{n} + \|h'\| \sum_{i=1}^n \mathbb{E}|X_i|^3.$$

In the usual scaling , $X_i \asymp \frac{1}{\sqrt{n}}$, in which case $\mu \asymp \sqrt{n}$ and the bound is of order $\frac{1}{\sqrt{n}}$. Depending on $\sigma_i, i = 1, \dots, n$, this bound may be better than the one obtained in Theorem 2.1 by the local approach.

Goldstein and Rinott [**31**] also give a general construction.

CONSTRUCTION 4.3. Let $X_1, \dots X_n$ be nonnegative, $\mathbb{E}X_i = \mu_i > 0$, $W = \sum_{i=1}^n X_i$, $\mathbb{E}W = \mu$, $\mathrm{Var}(W) = 1$. Choose an index $I$ from $\{1, \dots, n\}$ according to (4.1). If $I = i$, replace $X_i$ by a variate $X_i^*$ having the $X_i$-size bias distribution. If $X_i^* = x$, construct $\hat{X}_j, j \neq i$ such that

$$\mathcal{L}(\hat{X}_j, j \neq i) = \mathcal{L}(X_j, j \neq i | X_i = x).$$

Then

$$W^* = \sum_{j \neq I} \hat{X}_j + X_I^*$$

has the $W$-size biased distribution. (The difference from Construction 3.3 are the choice of $I$ and the distribution of $X_i^*$.)

The above construction can also be applied to the example of simple random sampling.

EXAMPLE 4.4. As in Example 3.4, let $X_1, \dots, X_n$ be a simple random sample of size $n$ from a finite population $A$, where all elements of $A$ are distinct. We follow Construction 4.3. Once $X_I^*$ is constructed, we may continue as in Example 3.4; if $X_I^* \notin \{X_j, j \neq I\}$, then we may choose $\hat{X}_j = X_j, j \neq I$, whereas if $X_I^* = X_J$ for some $J \in \{j, j \neq I\}$, then let $\hat{X}_j = X_j, j \neq I, J$, and choose $\hat{X}_J$ uniformly from $A \setminus \{X_I^*, \hat{X}_j, j \neq I, J\}$. This procedure is known as Midzuno's procedure, see Luk [**38**], and is used to obtain unbiased ratio estimators. Note that, as again $W$ and $W^*$ differ for at most two summands, the coupling is efficient.

Size bias couplings for functions of variables are described in Dembo and Rinott [**25**]. Typical examples are counts, as occurring in random graphs (the number of vertices of a fixed degree, for example) and in random allocations.

However, a necessary ingredient is that $W$ is nonnegative. For bounded variables $W$ one might shift the distribution to the positive axis, but this is not very natural, in particular when the distribution of $W$ is symmetric around zero, because $W$ should be closer to normal than any shifted version of it. This drawback motivated the introduction of the zero bias coupling described in the next section.

## 5. Zero-bias couplings

Let $W$ be mean zero, variance one. We say that a random variable $W^*$ has the $W$-zero biased distribution if, for all $g$ for which the expectation exists,

$$\mathbb{E}Wg(W) = \mathbb{E}g'(W^*). \tag{5.1}$$

This notion was introduced in Goldstein and Reinert [**30**]. Related analytical ideas, without coupling constructions, appear in Ho and Chen [**35**], Chen [**24**], and in Cacoullos *et al.* [**20**]. Bolthausen [**18**] uses a related coupling, without formalizing it. As the standard normal distribution is the unique fixed point of (5.1), it seems to be another natural approach for normal approximations.

Using (5.1), the right-hand side of Equation (1.3) can be written as

$$\begin{aligned}\mathbb{E}Wf'(W) - \mathbb{E}f'(W) &= \mathbb{E}(f''(W^*) - f''(W)) \\ &\le R,\end{aligned}$$

where

$$|R| \le \|f^{(3)}\| \mathbb{E}|W^* - W|.$$

Thus we save a step in the Taylor expansion. Expanding one step further, though, gives the following, sharper result, see Goldstein and Reinert [**30**], where we use the bounds (1.4), as derivatives higher than second order occur.

THEOREM 5.1. *Let $W$ be mean zero, variance 1, and let $W^*$ have the $W$-zero biased distribution. For any bounded, continuous function $h$ with bounded derivatives up to order 4, we have*

$$|\mathbb{E}h(W) - \Phi h| \le \frac{1}{3}\|h^{(3)}\|\sqrt{\mathbb{E}(\mathbb{E}(W^* - W|W))^2} + \frac{1}{8}\|h^{(4)}\|\mathbb{E}(W - W^*)^2.$$

EXAMPLE 5.2. Let $X_1, \ldots X_n$ be independent, mean zero, $\mathrm{Var}(X_i) = \sigma_i^2$, $W = \sum_{i=1}^n X_i$, and assume $\mathrm{Var}(W) = 1$. Choose an index $I$ from $\{1, \ldots, n\}$ according to

$$\mathbb{P}(I = i) = \sigma_i^2,$$

that is, choose index $i$ proportionally to its variance. (This resembles (4.1), where $i$ is drawn proportionally to its expectation.) If $I = i$, replace $X_i$ by $X_i^*$ having the $X_i$-zero biased distribution, and put

$$W^* = W - X_I + X_I^*. \tag{5.2}$$

Then $W^*$ has the $W$-zero biased distribution; note the similarity to (3.3) and (4.2). Using Theorem 5.1 gives

$$|\mathbb{E}h(W - \mu) - \Phi h| \le \frac{1}{3}\|h^{(3)}\|\sqrt{\mathbb{E}\left(\mathbb{E}X_I^3 - \frac{W}{n}\right)^2} + \frac{1}{8}\|h^{(4)}\|\left(\frac{\mathbb{E}X_I^4}{3} + \mathbb{E}X_I^2\right).$$

In the special case that $\mathbb{E}X_i^3 = 0$ we obtain

$$|\mathbb{E}h(W - \mu) - \Phi h| \le \frac{1}{3}\|h^{(3)}\|\frac{1}{n} + \frac{1}{8}\|h^{(4)}\|\left(\frac{\mathbb{E}X_I^4}{3} + \mathbb{E}X_I^2\right). \tag{5.3}$$

With the scaling $X_i \asymp \frac{1}{\sqrt{n}}$ we thus obtain a $\frac{1}{n}$ - bound on the rate of convergence for smooth test functions.

The feature of a $\frac{1}{n}$ - bound on the rate of convergence for smooth test functions under vanishing third moment conditions is a main advantage of the zero bias coupling. (It could also be derived using Edgeworth expansions, but those assume some smoothness of the density, see for instance Hall [**33**], Chapter 2.8). For nonsmooth test functions it is easy to see that the rate of $\frac{1}{\sqrt{n}}$ is unimprovable - consider the sum of $n$ independent centered binomials and, as test function, the indicator of the negative half axis.

Despite the similarity of the construction for the independent case in Example 5.2 to Construction 4.3 and Construction 3.3, a general construction is much more involved, unfortunately; see Goldstein and Reinert [**30**]. This more complicated construction, Construction 5.3 below, makes it seem advisable to mainly consider applications with vanishing third moments, as then there is hope for the better rate for smooth test functions.

CONSTRUCTION 5.3. Let $X_1, \ldots X_n$ be mean zero, $W = \sum_{i=1}^n X_i$, and assume $\mathrm{Var}(W) = 1$. Denote the distribution of $X_1, \ldots X_n$ by $dF_n$. Suppose that for each $i = 1, \ldots, n$ there exists a distribution $dF_{n,i}(x_1, \ldots, x_{i-1}, x_i', x_i'', x_{i+1}, \ldots, x_n)$ on $n+1$ random variables $X_1, \ldots, X_{i-1}, X_i', X_i'', X_{i+1}, \ldots, X_n$ such that

$$(X_1, \ldots, X_{i-1}, X_i', X_i'', X_{i+1}, \ldots, X_n) \quad \stackrel{d}{=} \quad (X_1, \ldots, X_{i-1}, X_i'', X_i', X_{i+1}, \ldots, X_n),$$

and

$$(X_1, \ldots, X_{i-1}, X_i, X_{i+1}, \ldots, X_n) \quad \stackrel{d}{=} \quad (X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n).$$

Suppose that there is a $\rho$ such that for all $f$ for which $\mathbb{E}Wf(W)$ exists,

$$\sum_{i=1}^n \mathbb{E}X_i' f(W_i + X_i'') \quad = \quad \rho \mathbb{E}Wf(W), \tag{5.4}$$

where $W_i = W - X_i$ as in (2.1). Let

$$\sum_{i=1}^n v_i^2 > 0 \quad \text{where} \quad v_i^2 = \mathbb{E}(X_i' - X_i'')^2,$$

and let $I$ be a random index independent of the $X's$ such that

$$\mathbb{P}(I = i) = v_i^2 / \sum_{j=1}^n v_j^2.$$

Further, for $i$ such that $v_i > 0$, let $\hat{X}_1, \ldots, \hat{X}_{i-1}, \hat{X}_i', \hat{X}_i'', \hat{X}_{i+1}, \ldots, \hat{X}_n$ be chosen according to the distribution

$$\begin{aligned} & d\hat{F}_{n,i}(\hat{x}_1, \ldots, \hat{x}_{i-1}, \hat{x}_i', \hat{x}_i'', \hat{x}_{i+1}, \ldots, \hat{x}_n) \\ = \quad & \frac{(\hat{x}_i' - \hat{x}_i'')^2}{v_i^2} dF_{n,i}(\hat{x}_1, \ldots, \hat{x}_{i-1}, \hat{x}_i', \hat{x}_i'', \hat{x}_{i+1}, \ldots, \hat{x}_n). \end{aligned}$$

Then, with $U$ a uniform $U[0,1]$ variate which is independent of the $X$'s and the index $I$,

$$W^* = U\hat{X}_I' + (1-U)\hat{X}_I'' + \sum_{j \neq I} \hat{X}_j$$

has the $W$-zero biased distribution.

Although this construction is rather involved, it is not so difficult to apply it to the simple random sampling example.

EXAMPLE 5.4. As in Example 3.4, let $X_1, \dots, X_n$ be a simple random sample of size $n$ from a finite population $A$, where all elements of $A$ are distinct. Assume that $\sum_{a \in A} a^3 = 0$. In Construction 5.3, because of exchangeability we may choose $I = 1$. Independently of $X_1, \dots, X_n$, pick a pair $(\hat{X}_1', \hat{X}_1'')$ from the distribution

$$q(u, v) = \frac{(u-v)^2}{2N} \mathbf{1}(\{u, v\} \subset A).$$

Firstly then, independently of the chosen sample $\mathbf{X}$, pick $(\hat{X}_1', \hat{X}_1'')$ from the distribution $q(u, v)$. The random variables $(\hat{X}_1', \hat{X}_1'')$ are now placed as the first two components in the vector $\hat{\mathbf{X}}$. The remaining $n-1$ random variables $\hat{\mathbf{X}}$ are sampled by rejection. If the two sets $\{X_2, \dots, X_n\}$ and $\{\hat{X}_1', \hat{X}_1''\}$ do not intersect, fill in the remaining $n-1$ components of $\hat{\mathbf{X}}$ with $(X_2, \dots, X_n)$. If the sets have an intersection, remove from the vector $(X_2, \dots, X_n)$ the two random variables (or single random variable) that intersect and replace them (or it) with values obtained by a simple random sample of size two (one) from $A \setminus \{\hat{X}_1', \hat{X}_1'', X_2, \dots, X_n\}$. This new vector now fills in the remaining $n-1$ positions in $\hat{\mathbf{X}}$. In Goldstein and Reinert [**30**] it is shown that this construction satisfies Condition (5.4) with $\rho = -n/(N-n)$, and that it yields a bound of order $\frac{1}{n}$ for the normal approximation of $W$, provided the elements of $A$ are scaled to be $\asymp \frac{1}{\sqrt{n}}$.

The zero-bias coupling also displays an interesting connection with the method of exchangeable pairs. Let $(W, W')$ be an exchangeable pair with distribution function $dF$ such that Condition 3.2 is satisfied. Pick a pair $(\hat{W}, \hat{W}')$ from the distribution

$$\frac{(\hat{w} - \hat{w}')^2}{\mathbb{E}(W - W')^2} dF(\hat{w}, \hat{w}').$$

Let $U$ be an independent $U(0,1)$ variable. Then

$$W^* = U\hat{W} + (1-U)\hat{W}'$$

has the $W$-zero biased distribution (see Goldstein and Reinert [**30**]).

Other examples where the zero bias coupling might be useful include the anti-voter model, $U$-statistics, and permutation statistics.

## 6. Other couplings

There are many other couplings that work well in specific situations. One example is symmetric arrays as treated by Barbour and Eagleson [**10**]. Assume $a_1, \dots, a_n$ and $b_1, \dots, b_n$ are real numbers with $\sum_{i=1}^n a_i = \sum_{j=1}^n b_j = 0$. Let $\pi$ be a random permutation of $\{1, \dots, n\}$, chosen uniformly, and put

$$X_i = a_i b_{\pi(i)}.$$

Let, as usual, $W = \sum_{i=1}^n X_i$. Assume that $\mathrm{Var}(W) = (n-1)^{-1} \sum_{i=1}^n a_i^2 \sum_{j=1}^n b_j^2 = 1$. Now, if $\pi(i)$ is known, we automatically know $X_i$. This can be put to use as follows.

$$\begin{aligned} \mathbb{E}Wf'(W) &= \frac{1}{n}\sum_{i,j=1}^{n} a_i b_j \mathbb{E}(f'(W)|\pi(i)=j) \\ &= \frac{1}{n}\sum_{i,j=1}^{n} a_i b_j \mathbb{E}f'(W+D_{i,j}), \end{aligned}$$

where

$$D_{i,j} = (a_i - a_{\pi^{-1}(j)})(b_j - b_{\pi(i)}).$$

Now

$$\begin{aligned} &\frac{1}{n}\sum_{i,j=1}^{n} a_i b_j \mathbb{E}f'(W+D_{i,j}) \\ &\approx \frac{1}{n}\sum_{i,j=1}^{n} a_i b_j \mathbb{E}D_{i,j} f'(W) \\ &= \frac{1}{n}\sum_{i,j=1}^{n} a_i b_j \frac{1}{n(n-1)} \sum_{k\neq j}\sum_{l\neq i} (a_i - a_l)(b_j - b_k)\mathbb{E}f''(W + \tilde{D}_{i,k;l,j}), \end{aligned}$$

where $\tilde{D}_{i,k;l,j} = D_{i,k} + D_{l,j}$ whenever $\{\pi(i), \pi(j)\} \cap \{k, j\}$ is empty, and when the intersection is nonempty, the expressions are slightly modified. After some work this yields a normal approximation. This approach can be generalized to obtain a Wald-Wolfowitz Theorem for processes, see Barbour [**7**], [**8**] and Barbour and Eagleson [**11**].

Furthermore, sometimes couplings for normal approximations are used in a different sense. A variate $T$ is first coupled to a variate $T_0$ using the structure inherent of the problem, and then a normal approximation is shown or known to hold for $T_0$. This is applied by Bolthausen and Götze [**19**] to multivariate sampling statistics, and by Barbour *et al.* [**12**] to iterations of expanding maps.

Finally it should be emphasized that the above is a collection of techniques. Depending on the problem that is to be solved, they might provide useful tools. Yet there is always the possibility that a coupling of a different nature might yield better results. Moreover, a concentration inequality approach has been proved to be another powerful tool when using Stein's method for normal approximations; see Chen [**24**] for an overview.

**Acknowledgements.** I would like to thank the organizers for the opportunity to present the paper, and the referee for many useful suggestions. Larry Goldstein has provided many useful discussions over the last years that certainly influenced my perspective. Finally, many thanks to Andrew Barbour and Charles Stein for teaching me on this subject.

## References

[1] Arratia, R., Goldstein, L., and Gordon, L. (1989). Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.* **17** 9–25.

[2] Aldous, D. (1989). Stein's method in a two-dimensional coverage problem. *Statist. Probab. Lett.* **8**, 307 - 314.

[3] Baldi, P., and Rinott, Y. (1989). Asymptotic normality of some graph-related statistics. *J. Appl. Prob.* **26**, 171–175.

[4] Baldi, P., and Rinott, Y. (1989). On normal approximations of distributions in terms of dependency graphs. *Ann. Probab.* **17**, 1646–1650.

[5] Baldi, P., Rinott, Y., and Stein, C. (1989). A normal approximation for the number of local maxima of a random function on a graph. In *Probability, Statistics and Mathematics, Papers in Honor of Samuel Karlin.* T. W. Anderson, K.B. Athreya and D. L. Iglehart, eds., Academic Press, 59–81.

[6] Barbour, A.D. (1988). Stein's method and Poisson process convergence. *J. Appl. Probab.* **25 (A)**, 175–184.

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

[8] Barbour, A.D. (1997). Stein's method. *Encyclopaedia of Statistical Science,* Update, Vol.1, Wiley, New York, 513–521.

[9] Barbour, A.D., Chen, L.H.Y., and Loh, W.-L. (1992) Compound Poisson approximation for nonnegative 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. Prob.* **17**, 147–162.

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

[12] Barbour, A.D., Gerrard, R., and Reinert, G. (1997). Iterates of expanding maps. Preprint.

[13] Barbour, A.D., and Hall, P. (1983). On bounds to the rate of convergence in the central limit theorem. *Bull. London Math. Soc.* **17**, 151–156.

[14] Barbour, A.D., and Hall, P. (1984). Stein's method and the Berry-Esséen theorem. *Austral. J. Statist.* **26 (1)**, 8–15.

[15] Barbour, A.D., Holst, L., and Janson, S. (1992). *Poisson Approximation.* Oxford Science Publications.

[16] Barbour, A.D., Karoński, M., and Ruciński, A. (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Comb. Theory*, Ser. B, **47**, 125–145.

[17] Barbour, A.D., and Utev, S. (1997). Compound Poisson approximation in total variation. Preprint.

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

[19] Bolthausen, E., and Götze, F. (1993). The rate of convergence for multivariate sampling statistics. *Ann. Statist.* **21**, 1692–1710.

[20] Cacoullos, T., Papathanasiou, V., and Utev, S. (1994). Variational inequalities with examples and an application to the central limit theorem. *Ann. Probab.* **22**, 1607–1618.

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

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

[23] 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.

[24] Chen, L.H.Y. (1996). Stein's method: some perspectives with applications. Preprint.

[25] Dembo, A., and Rinott, Y. (1996). Some examples of normal approximations by Stein's method. In *Random Discrete Structures*, IMA volume 76, Aldous, D. and Pemantle, R., eds., Springer-Verlag, 25–44.

[26] Diaconis, P. (1989). An example for Stein's method. Stanford Stat. Dept. Tech. Rep.

[27] Diaconis, P., and Zabell, S. (1991). Closed form summation for classical distributions: variations on a theme of de Moivre. *Statistical Science* **6**, 284-302.

[28] Ehm, W. (1991). Binomial approximation to the Poisson binomial distribution. *Statist. Probab. Lett.* **11**, 7–16.

[29] Fulman, J. (1997). A Stein's method proof for the asymptotic normality of descents and inversions in the symmetric group. Preprint.

[30] Goldstein, L., and Reinert, G. (1997) Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.,* in print.

[31] Goldstein, L., and Rinott, Y. (1996). On multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Prob.* **33**, 1–17.

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

[33] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag.
[34] Hall, P., and Barbour, A.D. (1984). Reversing the Berry-Esséen theorem. *Proc. AMS* **90**, 107–110.
[35] Ho, S.-T., and Chen, L.H.Y. (1978). An $L_p$ bound for the remainder in a combinatorial central limit theorem. *Ann. Probab.* **6**, 231–249.
[36] Loh, W.-L. (1992). Stein's method and multinomial approximation. *Ann. Appl. Probab.* **2**, 536–554.
[37] Loh, W.-L. (1997). Stein's method and Pearson Type 4 approximations. Research report.
[38] Luk, H.M. (1994). Stein's method for the gamma distribution and related statistical applications. Ph.D. thesis. University of Southern California, Los Angeles, USA.
[39] Mann, B. (1997). Stein's method for $\chi^2$ of a multinomial. Preprint.
[40] Peköz, E. (1995). Stein's method for geometric approximation. Tech. Report #225, Stat. Dept. U. California. Berkeley.
[41] Reinert, G. (1994). A weak law of large numbers for empirical measures via Stein's method, and applications. Ph.D. thesis. University of Zurich, Switzerland.
[42] Reinert, G. (1997). A Gaussian approximation for random measures via Stein's method. Preprint.
[43] Rinott, Y. (1994). On normal approximation rates for certain sums of dependent random variables. *J. Comp. Appl. Math.* **55**, 135–143.
[44] 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 Analysis.* **56**, 333–350.
[45] Rinott, Y., and Rotar, V. (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the anti-voter model and weighted $U$-statistics. Preprint.
[46] Roos, M. (1994). Stein's method for compound Poisson approximation: the local approach. *Ann. Appl. Probab.* **4**, 1177–1187.
[47] Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2**, 583–602. Univ. California Press, Berkeley.
[48] Stein, C. (1986). *Approximate Computation of Expectations.* IMS, Hayward, California.
[49] Stein, C. (1992). A way of using auxiliary randomization. In *Probability Theory,* L. H. Y. Chen, K. P. Choi, K. Hu and J.-H. Lou, eds., Walter de Gruyter, Berlin - New York, 159–180.

Department of Mathematics, University of California, Los Angeles CA 90095-1555
*E-mail address*: `reinert@stat.ucla.edu`