

Multivariate normal approximations by Stein's method and size bias couplings

Larry Goldstein* and Yosef Rinott†

arXiv:math.PR/0510586 v1 27 Oct 2005

Abstract

Stein's method is used to obtain two theorems on multivariate normal approximation. Our main theorem, Theorem 1.2, provides a bound on the distance to normality for any nonnegative random vector. Theorem 1.2 requires multivariate size bias coupling, which we discuss in studying the approximation of distributions of sums of dependent random vectors. In the univariate case, we briefly illustrate this approach for certain sums of nonlinear functions of multivariate normal variables. As a second illustration, we show that the multivariate distribution counting the number of vertices with given degrees in certain random graphs is asymptotically multivariate normal and obtain a bound on the rate of convergence. Both examples demonstrate that this approach may be suitable for situations involving non-local dependence. We also present Theorem 1.4 for sums of vectors having a local type of dependence. We apply this theorem to obtain a multivariate normal approximation for the distribution of the random p -vector which counts the number of edges in a fixed graph both of whose vertices have the same given color when each vertex is colored by one of p colors independently. All normal approximation results presented here do not require an ordering of the summands related to the dependence structure. This is in contrast to hypotheses of classical central limit theorems and examples, which involve e.g., martingale, Markov chain, or various mixing assumptions.

Keywords and phrases: Stein's method, coupling, size bias, random graphs, multivariate central limit theorems.

MSC 1991 Classification: 60F05,60B12,05C80

*Department of Mathematics DRB-155, USC, Los Angeles, CA 90089-1113. This work was supported in part by NSF grant DMS 90-05833

†Department of Mathematics, UCSD, La Jolla, CA 92093. This work was supported in part by NSF grant DMS 92-05759.

1 Introduction

Stein's method has been successful in assessing the quality of normal and Poisson approximations under various dependence structures. See Stein (1972), Stein (1986), Barbour, Holst and Janson (1992), and references therein. Significant multivariate (or functional) versions of Stein's method appear for example in Barbour (1990), and Götze (1991). An important part of Stein's method is the construction of auxiliary random variables (coupling), which are used in the computation of the bounds on the distance between a given random variable and its normal or Poisson approximant.

The coupling variables constructed in the application of Stein's method may not appear explicitly in the final bounds. Applications of results where the auxiliary random variables do appear in the bounds require their explicit construction. Although this feature may make such results more difficult to use, couplings yielding useful bounds may often be found where other methods seem to fail. For the Poisson, theorems of this nature may be found in Barbour, Holst and Janson (1992) and references within. In the main result of this paper, Theorem 1.2, we obtain bounds for normal approximations in terms of such couplings, and provide general guidelines and methods for their construction so that these methods may be applied. The couplings studied here, an instance of a construction of a joint distribution with given marginals, is of independent interest. It is known that constructions of multivariate couplings may be problematic, see, e.g., Dall'Aglio, Kotz and Salinetti (1991). Nevertheless we are able to provide methods for the required multidimensional coupling constructions, which we illustrate in two applications involving nonlocal dependence, Theorems 4.1 and 4.2.

By the same techniques used to prove Theorem 1.2, Stein's method and the analysis of the properties of the solution to the partial differential equation (13), we obtain Theorem 1.4, a result complementary to Theorem 1.2. Theorem 1.4 provides a multivariate normal approximation under conditions of local dependence. Unlike Theorem 1.2, coupling variables do not appear explicitly in Theorem 1.4.

In order to introduce the couplings needed for the proof and applications of Theorem 1.2, we require the following definition.

Definition 1.1 *Given a nonnegative random variable W with distribution $dF(w)$ and mean λ , W^* is said to have the W -size biased distribution if it has distribution $w dF(w)/\lambda$.*

Note that the distribution of W^* may be characterized by the relation

$$EWG(W) = \lambda EG(W^*) \tag{1}$$

for all functions G for which the expectations exist. Size biased distributions are well known in sampling theory and renewal theory, for example. The following one dimensional version of our main result illustrates the relevance of size biased coupling to normal approximations.

Theorem 1.1 *Let W be a nonnegative random variable with mean $EW = \lambda$, variance $\sigma^2 = \text{Var}(W)$, and let W^* be jointly distributed with W , having the W -size biased distribution. Then for any piecewise continuously differentiable h ,*

$$\begin{aligned} & \left| Eh\left(\frac{W - \lambda}{\sigma}\right) - \Phi h \right| \\ & \leq 2 \|h\| \frac{\lambda}{\sigma^2} \sqrt{\text{Var} E(W^* - W|W)} + \|h'\| \frac{\lambda}{\sigma^3} E(W^* - W)^2, \end{aligned} \tag{2}$$

where $\|\cdot\|$ denotes the supremum norm, and $\Phi h = Eh(Z)$ with Z a standard normal variate.

Theorem 1.1 is an extension of a result of Baldi, Rinott and Stein (1989). The theorem requires the construction of W^* on a joint space with W ; hence, obtaining good bounds in any particular application depends on the construction of a W^* which will be close to W in an appropriate sense.

However, since the resulting bound is valid for any construction for which the marginal distribution of W^* coincides with the W -size biased distribution, one has the flexibility to choose constructions which result in good computable bounds.

Here is a brief description of a method which leads to a construction of a W -size biased variate W^* when $W = X_1 + \dots + X_n$ is a sum of random variables. To begin, if X_1, \dots, X_n are iid nonnegative random variables with finite mean, then W^* can be constructed by replacing any single summand, say, X_1 by an independent variable X_1^* with the X_1 -size biased distribution, i.e. $W^* = X_1^* + X_2 + \dots + X_n$. More generally, if W is a sum of non-iid variates, then a like construction of W^* may be given by replacing X_I by X_I^* , where the random index is chosen independently with $P(I = i) = EX_i / \sum EX_j$, and adjusting the remaining variables to their conditional distribution given the new value of X_I .

A special case of this idea is Midzuno's procedure (e.g. Cochran (1977)), where a size biased variable is used to obtain unbiased ratio estimators in finite population sampling. To describe Midzuno's procedure, let nonnegative "sizes" X_1, \dots, X_n be obtained by sampling from a finite population without replacement, and W be their sum. Then, W^* is realized by sampling the first variate in proportion to its size, removing it from the population, and sampling the other variables without replacement from the population that remains, that is, sampling from the resulting conditional distribution.

Further flexibility is obtained by realizing that for any representation of W in the form

$$W = \psi_1(U_1) + \dots + \psi_n(U_n), \quad (3)$$

the above construction of W^* may be accomplished by choosing a random index I such that $P(I = i) = E\psi_i(U_i) / \sum_{j=1}^n E\psi_j(U_j)$, and if $I = i$, replacing U_i by an independent variable with distribution $\psi_i(u)P(U_i \in du) / E\psi_i(U_i)$, and adjusting the remaining U variables. Therefore, the theorem may be applied whenever one can find a transformation such that the variables U_1, \dots, U_n have a dependence structure that allows the computation of the conditional distribution required in the bounds. Further details and examples of these size bias coupling constructions and their applications will be provided in Sections 2 and 4.

For the multivariate case, we need a more general notion of size biasing, and we replace the $*$ notation by a superscript β in order to identify in which "coordinate" or variable the variates are size biased.

Definition 1.2 *Let \mathcal{I} an arbitrary index set and let $\mathbf{X} = \{X_\alpha : \alpha \in \mathcal{I}\}$ be a collection of nonnegative random variables with joint distribution $dF(\mathbf{x})$ and means $EX_\alpha = \lambda_\alpha$. For $\beta \in \mathcal{I}$, we say that $\mathbf{X}^\beta = \{X_\alpha^\beta : \alpha \in \mathcal{I}\}$ has the \mathbf{X} -size biased distribution in the β^{th} coordinate if \mathbf{X}^β has the joint distribution $x_\beta dF(\mathbf{x}) / \lambda_\beta$.*

The distribution of \mathbf{X}^β is characterized by the relations

$$EX_\beta G(\mathbf{X}) = \lambda_\beta EG(\mathbf{X}^\beta) \quad (4)$$

for all functions G for which the above expectations exist. When the function G depends on \mathbf{X} only through X_β , equation (4) yields $EX_\beta G(X_\beta) = \lambda_\beta EG(X_\beta^\beta)$, hence, comparing to (1), we see that the β^{th} coordinate of \mathbf{X}^β , that is, the variate X_β^β , has the X_β -size biased distribution in the sense of definition 1.1.

By considering the case where the collection \mathbf{X} consists of only a single random variable, we see that equation (4) reduces to (1); hence definition 1.1 is a special case of definition 1.2.

We will apply Definition 1.2 to a vector $\mathbf{W} = (W_1, \dots, W_p) \in R^p$ by identifying it with the collection $\{W_j : j \in \mathcal{I}\}$ with $\mathcal{I} = \{1, \dots, p\}$. Letting $EW_i = \lambda_i$, we see that the vector $\mathbf{W}^i = (W_1^i, \dots, W_p^i)$ is characterized by

$$EW_i G(\mathbf{W}) = \lambda_i EG(\mathbf{W}^i). \quad (5)$$

Relation (5) leads to a multivariate normal approximation theorem, for which we introduce the following notation (see e.g., Horn and Johnson (1985)). Given a vector \mathbf{a} in R^p , let $\|\mathbf{a}\| = \max_{1 \leq i \leq p} |a_i|$. Given a $p \times p$ matrix $A = (a_{ij})$ we set $\|A\| = \max_{1 \leq i, j \leq p} |a_{ij}|$, and more generally for any array, $\|\cdot\|$ will denote its maximal absolute value. For an array of functions, say $A(\mathbf{w}) = \{a_i(\mathbf{w})\}$, where i could stand for a multiple index, $\|A\| = \sup_{\mathbf{w}} \max_i |a_i(\mathbf{w})|$. For a smooth function $h : R^p \rightarrow R$ we let ∇h or Dh denote the vector of first partial derivatives of h , D^2h the usual Hessian matrix of second order partial derivatives and $D^k h$ the k^{th} derivative of h in general.

Our main multivariate result is the following theorem:

Theorem 1.2 *Let \mathbf{W} be a random vector in R^p with nonnegative components. Set $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p) = E\mathbf{W}$, and assume $\text{Var}\mathbf{W} = \Sigma = (\sigma_{ij})$ is invertible. For each $i = 1, \dots, p$ let $(\mathbf{W}, \mathbf{W}^i)$ be random vectors defined on a joint probability space with \mathbf{W}^i having the \mathbf{W} -size biased distribution in the i^{th} coordinate as in (5). Let $h : R^p \rightarrow R$ be a function having bounded mixed partial derivatives up to order 3. Let $\Phi h = Eh(\mathbf{Z})$, where \mathbf{Z} denotes a standard (mean zero, covariance I) normal vector in R^p . Then*

$$\begin{aligned} |Eh(\Sigma^{-1/2}(\mathbf{W} - \boldsymbol{\lambda})) - \Phi h| \leq & \\ & \frac{p^2}{2} \|\Sigma^{-1/2}\|^2 \|D^2h\| \sum_{i=1}^p \sum_{j=1}^p \lambda_i \sqrt{\text{Var}E[W_j^i - W_j | \mathbf{W}]} \\ & + \frac{1}{2} \frac{p^3}{3} \|\Sigma^{-1/2}\|^3 \|D^3h\| \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p \lambda_i E|(W_j^i - W_j)(W_k^i - W_k)|. \end{aligned} \quad (6)$$

Note that the theorem does not require the joint construction of $(\mathbf{W}^1, \dots, \mathbf{W}^p)$. Although Theorems 1.1 and 1.2 are stated for nonnegative variates, they may be applied to general variates by translation and truncation.

In Section 2, we discuss the construction of the vectors \mathbf{W}^i required for Theorem 1.2 when the components of \mathbf{W} are sums of dependent random variables. Specifically, when $\mathbf{X} = \{X_\alpha, \alpha \in \mathcal{I}\}$ is any collection of nonnegative random variables, and A_1, \dots, A_p are any subsets of \mathcal{I} , we may apply Theorem 1.2 to the vector $\mathbf{W} = (W_1, \dots, W_p)$ where $W_j = \sum_{\alpha \in A_j} X_\alpha$. In particular, we obtain a result for a sum $\mathbf{W} = (W_1, \dots, W_p) = \sum_{u=1}^n \mathbf{X}_u$ of nonnegative dependent random vectors, $\mathbf{X}_u = (X_{u1}, \dots, X_{up})$, $u = 1, \dots, n$, by letting \mathcal{I} be a set of double indices and $A_j = \{1, \dots, n\} \times j$.

We briefly indicate how size biased variables arise in one dimensional normal approximations. Given a random variable W and a test function h , one can compute $E[h(\frac{W-\lambda}{\sigma}) - \Phi h]$ by computing $E[f'(W) - \frac{(W-\lambda)}{\sigma^2} f(W)]$ where f is the bounded solution of the Stein equation

$$f'(w) - \frac{(w-\lambda)}{\sigma^2} f(w) = h(\frac{w-\lambda}{\sigma}) - \Phi h. \quad (7)$$

If W^* has the W -size biased distribution, and therefore satisfies $EWf(W) = EWEf(W^*)$, we obtain

$$Eh(\frac{W-\lambda}{\sigma}) - \Phi h = E[f'(W) - \frac{(W-\lambda)}{\sigma^2} f(W)] = E[f'(W) - \frac{\lambda}{\sigma^2} (f(W^*) - f(W))].$$

Taylor expansion of $E[f(W^*) - f(W)]$ is then the first step in obtaining the bound in Theorem 1.1. Note that the one dimensional versions of the theorems are not exactly special cases of their multivariate counterparts. In the multivariate case, equation (7) will be replaced by the partial differential equation (13), resulting in different orders of the derivatives of h appearing in the one and multidimensional theorems.

The size biased coupling approach handles cases where there is global dependence among the summand variables. In contrast the following univariate and multivariate results not based on size biased couplings are very useful in cases of local dependence. The following theorem is due to Stein (1986); our Theorem 1.4 is a multivariate version.

Theorem 1.3 Let $X_v, v = 1, \dots, n$, be random variables with $EX_v = 0$. Let $S_v, v = 1, \dots, n$ be subsets of $\{1, \dots, n\}$ and set $W = \sum_{v=1}^n X_v$ and denote

$$\sum_{v=1}^n \sum_{u \in S_v} EX_v X_u = \sigma^2,$$

assuming $\sigma^2 > 0$. Then for any $h : R \rightarrow R$ which is continuous and piecewise continuously differentiable,

$$\begin{aligned} |Eh(W/\sigma) - \Phi h| &\leq \frac{2\|h\|}{\sigma^2} \sqrt{E \left\{ \sum_{v=1}^n \sum_{u \in S_v} (X_v X_u - EX_v X_u) \right\}^2} \\ &+ \sqrt{\frac{\pi}{2}} \frac{\|h\|}{\sigma} \sum_{v=1}^n E|E[X_v | X_u : u \notin S_v]| \\ &+ \frac{1}{\sigma^3} \|h'\| \sum_{v=1}^n E|X_v \sum_{u \in S_v} X_u \sum_{t \in S_v} X_t|. \end{aligned} \quad (8)$$

In typical applications of Theorem 1.3, we have X_v independent of $\{X_u : u \notin S_v\}$ and the second term of the bound in (8) vanishes. In this case we may view S_v as a *dependency neighborhood* of X_v . Generally, the bound in (8) is small if these neighborhoods are small, so this theorem is useful when the dependence is local.

The following result, which is particularly useful for normal approximations of sums of locally dependent random vectors, extends Theorem 1.3 to the multivariate case.

Theorem 1.4 Let $\{X_\alpha, \alpha \in \mathcal{I}\}$ be random variables with $EX_\alpha = 0$. Let A_1, \dots, A_p be subsets of \mathcal{I} , and set $\mathbf{W} = (W_1, \dots, W_p)$ where $W_j = \sum_{\alpha \in A_j} X_\alpha$. For each $\alpha \in \mathcal{I}$ let $S_\alpha \subseteq \mathcal{I}$, and assume that $\Sigma = (\sigma_{ij})$ is symmetric positive definite, where

$$\sigma_{ij} = \sum_{\alpha \in A_i} \sum_{\beta \in A_j \cap S_\alpha} EX_\alpha X_\beta, \quad i, j = 1, \dots, p.$$

Let $h : R^p \rightarrow R$ be a function having bounded mixed partial derivatives up to order 3, and $\Phi h = Eh(\mathbf{Z})$ where \mathbf{Z} denotes a standard normal vector in R^p . Then

$$\begin{aligned} |Eh(\Sigma^{-1/2} \mathbf{W}) - \Phi h| &\leq \frac{p^2}{2} \|\Sigma^{-1/2}\|^2 \|D^2 h\| \sum_{i=1}^p \sum_{j=1}^p \sqrt{E \left\{ \sum_{\alpha \in A_i} \sum_{\beta \in A_j \cap S_\alpha} (X_\alpha X_\beta - EX_\alpha X_\beta) \right\}^2} \\ &+ p \|\Sigma^{-1/2}\| \|Dh\| \sum_{i=1}^p \sum_{\alpha \in A_i} E|E[X_\alpha | X_\beta : \beta \notin S_\alpha]| \\ &+ \frac{1}{2} \frac{p^3}{3} \|\Sigma^{-1/2}\|^3 \|D^3 h\| \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p \sum_{\alpha \in A_i} E|X_\alpha \sum_{\beta \in A_j \cap S_\alpha} X_\beta \sum_{\gamma \in A_k \cap S_\alpha} X_\gamma|. \end{aligned} \quad (9)$$

Note that σ^2 of Theorem 1.3 and Σ of Theorem 1.4 are not necessarily equal to the covariance of W and the covariance matrix of \mathbf{W} , respectively. In Theorem 1.4, the symmetry of Σ is guaranteed if the sets $S_\alpha \subseteq \mathcal{I}$ are symmetric in the sense that $\beta \in S_\alpha$ if and only if $\alpha \in S_\beta$. In particular, in applying Theorem 1.4 to a sum of mean zero random vectors, $\mathbf{W} = \sum_{u=1}^n \mathbf{X}_u$, where $\mathbf{X}_u = (X_{u1}, \dots, X_{up})$ for $u = 1, \dots, n$, it is natural to take neighborhoods of the form $S_\alpha = S_{(u,i)} = T_u \times \{1, \dots, p\}$, where T_u are symmetric subsets of $\{1, \dots, n\}$.

In our applications, the sets S_α contain $\{\beta \in \mathcal{I} : \text{Cov}(X_\beta, X_\alpha) \neq 0\}$; in any such case, $\Sigma = \text{Cov}(\mathbf{W})$. In particular, $\Sigma = \text{Cov}(\mathbf{W})$ if X_β is independent of X_α for every $\beta \notin S_\alpha$. In the general case, the above (somewhat unusual) choice of Σ simplifies the form of the bound. With the more natural choice $\Sigma = \text{Cov}(\mathbf{W})$, the present technique applies to yield a version of the above theorem, but an additional term in the bound may result.

Applications of the above theorems are given in Section 4. As an illustration involving non-local dependence, we apply Theorem 1.1 and its multivariate extension, Theorem 1.2, to show that the multivariate distribution counting the number of vertices with given degrees in certain random graphs is asymptotically multivariate normal and obtain a bound on the rate of convergence. To illustrate a case of local dependence, we apply Theorem 1.3, and its multivariate extension, Theorem 1.4, to obtain a multivariate normal approximation for the distribution of the random p -vector which counts the number of edges in a fixed graph both of whose vertices have the same given color when each vertex is colored by one of p colors independently. Applications related to representations of W as in (3) will be given where the U variables are normal and multinomial. The ideas and results presented here have been applied in work of Luk (1994) in finite population sampling, and Reinert (1994) in the study of empirical measures.

The proofs of Theorems 1.2 and 1.4 are given in Section 3, Theorems 1.1 and 1.3 may be proved similarly.

The theorems presented here supply approximations in terms of expectations of smooth test functions h , allowing our main theorems to be presented in a form where they can be readily applied under unrestrictive, simple conditions. In the context of Stein's method, Stein (1986), Baldi, Rinott and Stein (1989), Götze (1991), and Rinott (1994) among others, consider also non smooth functions, usually at the expense of added technical detail or some loss of information in the bounds. It is possible to obtain certain multivariate version of our results for nonsmooth functions h using the methodology developed in Götze (1991), see Rinott and Rotar (1994). In the present paper, our main focus is in the coupling structure. The issue of smooth versus non-smooth function approximation is discussed in Barbour, Karoński and Ruciński (1989).

2 Construction of size biased couplings

The construction of size biased variables required for the application of Theorems 1.1 and 1.2 is the focus of this section. While the details depend on the case at hand, this section will provide general guidelines that extend and unify ideas which appeared in Baldi, Rinott, Stein (1989), and Stein (1992), where only univariate sums of zero-one variables were studied.

The following lemma is the key in the construction of coupled variables satisfying equations (1) and (5) required in Theorems 1.1 and 1.2 respectively. Readers interested only in the univariate case may read the lemma below with $\mathcal{I} = A = B$.

Lemma 2.1 *Let \mathcal{I} be an arbitrary index set, and let $\mathbf{X} = \{X_\alpha : \alpha \in \mathcal{I}\}$ be a collection of nonnegative random variables. For any subset $B \subseteq \mathcal{I}$, set $X_B = \sum_{\beta \in B} X_\beta$, and $\lambda_B = EX_B$. Suppose $B \subseteq \mathcal{I}$ with $\lambda_B < \infty$, and for $\beta \in B$ let \mathbf{X}^β have the \mathbf{X} -size biased distribution in coordinate β as in Definition 1.2. Let \mathbf{X}^B be a random variable distributed as the mixture of the distributions \mathbf{X}^β , $\beta \in B$ with weights λ_β/λ_B . Then*

$$EX_B G(\mathbf{X}) = \lambda_B EG(\mathbf{X}^B). \quad (10)$$

Hence, for any $A \subseteq \mathcal{I}$, if G is a function of X_A only, then

$$EX_B G(X_A) = \lambda_B EG(X_A^B), \quad (11)$$

where

$$X_A^B = \sum_{\alpha \in A} X_\alpha^B.$$

In particular, by taking $A = B$ in (11) we have $EX_A G(X_A) = \lambda_A G(X_A^A)$, and hence X_A^A has the X_A -size biased distribution in the sense of definition 1.1.

Proof: For a function G on \mathbf{X} , we have $EG(\mathbf{X}^\beta) = EX_\beta G(\mathbf{X})/\lambda_\beta$ by equation (4). Multiplying by λ_β/λ_B and summing over $\beta \in B$ yields (10). The remainder of the lemma now follows. \square

Construction of \mathbf{X}^B : Since \mathbf{X}^B is a mixture of the distributions \mathbf{X}^β for $\beta \in B$ with weights λ_β/λ_B , given the collection $\mathbf{X} = \{X_\alpha : \alpha \in \mathcal{I}\}$ first choose an independent index $I \in B$ according to the distribution $P(I = \beta) = \lambda_\beta/\lambda_B$. If $I = \beta$, construct X_β^β to have the X_β -size biased distribution $x_\beta P(X_\beta \in dx_\beta)/\lambda_\beta$. If $X_\beta^\beta = x$ then the remaining variates X_α^β , $\alpha \neq \beta$ are constructed so that $P(\mathbf{X}^\beta \in d\mathbf{x}) = P(\mathbf{X} \in d\mathbf{x} | X_\beta = x)$. This construction yields

$$P(\mathbf{X}^\beta \in d\mathbf{x}) = P(\mathbf{X} \in d\mathbf{x} | X_\beta = x_\beta) x_\beta P(X_\beta \in dx_\beta)/\lambda_\beta,$$

that is, $\mathbf{X}^\beta \sim x_\beta dF(\mathbf{x})/\lambda_\beta$, and indeed \mathbf{X}^β has the \mathbf{X} -size biased distribution in the β^{th} coordinate as given in Definition 1.2.

In the univariate case, with $W = \sum_{\alpha \in \mathcal{I}} X_\alpha = X_{\mathcal{I}}$, and $A = B = \mathcal{I}$, equation (11) in Lemma 2.1 shows that a construction of W^* satisfying (1) may be obtained by setting $W^* = X_{\mathcal{I}}^{\mathcal{I}}$. Hence W^* may be constructed as follows: a summand X_β of W , chosen with probability EX_β/EW , is replaced by a new value from its size biased distribution, and the remaining summands are adjusted to have the conditional distribution of \mathbf{X} conditioned on the event that for the chosen β , X_β takes the new value.

If the variates $\{X_\alpha : \alpha \in \mathcal{I}\}$ are independent the last step is not needed since by independence the conditioning is irrelevant. In this case, the construction of a W -size biased variable W^* reduces to size biasing a single randomly chosen summand X_I . In the case that $\{X_\alpha : \alpha \in \mathcal{I}\}$ are all zero-one variates we simply have $X_\beta^\beta \equiv 1$, so in the case that W is a sum of independent zero-one random variables, the coupling is accomplished by choosing an index $I = \beta$ with probabilities proportional to $P(X_\beta = 1)$, setting $X_\beta = 1$, and leaving the remaining variates unchanged.

In the multivariate case, the connection between Lemma 2.1 and Theorem 1.2 for approximating sums of random variables is as follows. Given $\{X_\alpha : \alpha \in \mathcal{I}\}$, let A_1, \dots, A_p be subsets of \mathcal{I} , and set

$$\mathbf{W} = (W_1, \dots, W_p), \quad \text{where } W_j = X_{A_j} \quad \text{and} \quad \mathbf{W}^i = (W_1^i, \dots, W_p^i) \quad \text{where } W_j^i = X_{A_j}^{A_i}.$$

When $B = A_i$ and $G(\mathbf{X})$ is a function depending on \mathbf{X} only through \mathbf{W} , equation (10) yields equation (5). Therefore, one may obtain the vector \mathbf{W}^i satisfying (5) by constructing \mathbf{X}^{A_i} using a random index $I \in A_i$ with $P(I = \beta) = \lambda_\beta/\lambda_{A_i}$ as described above.

In particular, the sum of random vectors $\mathbf{W} = \sum_{u=1}^n \mathbf{X}_u$ where $\mathbf{X}_u = (X_{u1}, \dots, X_{up})$, corresponds to the choice $\mathcal{I} = \{1, \dots, n\} \times \{1, \dots, p\}$ and $A_j = \{(u, j) : u = 1, \dots, n\}$. In Section 4, we apply this multivariate construction in the setup where X_{uj} is the indicator of the event that the degree of the vertex u in the random graph $K = K_{n,\pi}$ equals a prescribed number d_j . Hence, W_j is the number of vertices of K of degree d_j . The coupling in this case is accomplished as follows. Since EX_{ui} are equal for indices in A_i , to construct \mathbf{X}^{A_i} it is required to choose an index, say v , uniformly over $\{1, \dots, n\}$ and size bias X_{vi} for this v . As X_{vi} is an indicator, size biasing is accomplished by replacing X_{vi} by the constant 1. The above construction now requires that the remaining variables have their original distribution conditioned on $X_{vi} = 1$. If X_{vi} was initially 1, that is, if the degree of vertex v was d_i , no change is required. Otherwise, by adding or removing randomly chosen edges as appropriate, the degree of v is made to be d_i , thereby size biasing the indicator X_{vi} . This procedure results in a new graph K^i , in which the other variables now have the proper conditional distribution.

The following comments pertain to the random choice of index that appears in the above constructions. In certain cases the size biased distribution can be constructed with a deterministic

index, however, such constructions may lead to larger bounds than those obtained using randomization.

We specialize to $p = 1$ and the expression $\text{Var}E[W^* - W | W]$ which appears in the first term in the bound in (6). If X_1, \dots, X_n are exchangeable then it is easy to see that W may be size biased by size biasing X_1 ; that is, in the above description one can set $I \equiv 1$ deterministically and the rest is done as above. This is true for the example where W counts the number of vertices of degree d in the random graph K . Often in order to make calculations tractable it is necessary to condition on a larger σ -field $\mathcal{F} \supset \sigma\{W\}$ and replace $\text{Var}E[W^* - W | W]$ by the larger quantity $\text{Var}E[W^* - W | \mathcal{F}]$. However, $\text{Var}E[W^* - W | \mathcal{F}]$ may not give rise to a useful bound unless I is randomized. This is the case in the random graph example when $\text{Var}E[W^* - W | W]$ is replaced by $\text{Var}E[W^* - W | K]$.

This difficulty can be seen even in the case of independent, identically distributed zero-one random variables, where conditioning on \mathbf{X} is the analog to conditioning on K . There, size biasing with a random I leads to $\text{Var}E[W^* - W | \mathbf{X}] = \text{Var}E[1 - X_I | \mathbf{X}] = \text{Var}E[X_I | \mathbf{X}] = \text{Var}(W/n)$ of order $1/n$, but setting $I \equiv 1$, the quantity $\text{Var}E[W^* - W | \mathbf{X}] = \text{Var}(X_1)$, a constant.

The following lemma of Dembo and Rinott (1994) shows how to size bias a sum of the form $W = \sum_{j=1}^n \psi_j(U_j)$ by working with the argument distribution U_1, \dots, U_n .

Lemma 2.2 *Let $\mathbf{U} = (U_1, \dots, U_n)$ be a random n vector, and let ψ_i be nonnegative functions such that $E\psi_i(U_i) < \infty$, $i = 1, \dots, n$. Let $\mathbf{Y}^{(i)} = (Y_1^{(i)}, \dots, Y_n^{(i)})$ satisfy $P(\mathbf{Y}^{(i)} \in d\mathbf{y}) = P(\mathbf{U} \in d\mathbf{y})\psi_i(y_i)/E\psi_i(U_i)$. Let I be a random variable taking values in $\{1, \dots, n\}$, distributed independently of all the above variables, with $P(I = i) = E\psi_i(U_i)/\sum_{j=1}^n E\psi_j(U_j)$.*

Let $W = \sum_{j=1}^n \psi_j(U_j)$ have the distribution F . Then $W^ = \sum_{j=1}^n \psi_j(Y_j^{(I)})$ has the distribution $w dF(w)/\lambda$, where $\lambda = EW$.*

Note that with F_j denoting the marginal distribution function of U_j , the distribution of $\mathbf{Y}^{(i)}$ is obtained by letting $Y_i^{(i)}$ have the marginal distribution $\psi_i(\cdot)dF_i(\cdot)/E\psi_i(U_i)$, and if $Y_i^{(i)} = u$, letting $(Y_1^{(i)}, \dots, Y_{i-1}^{(i)}, Y_{i+1}^{(i)}, \dots, Y_n^{(i)})$ have the distribution of $(U_1, \dots, U_{i-1}, U_{i+1}, \dots, U_n)$ conditioned on $U_i = u$.

To summarize, given $W = \sum_{j=1}^n \psi_j(U_j)$, this suggests the following:

Construction of W^* . Choose a random index I as in the lemma. If $I = i$, let $Y_i \sim \psi_i(u)dF_i(u)/E\psi_i(U_i)$. If Y_i is assigned the value u , let $(Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n)$ have the conditional distribution of $(U_1, \dots, U_{i-1}, U_{i+1}, \dots, U_n)$ given $U_i = u$. Now set $W^* = \sum_{j=1}^n \psi_j(Y_j)$.

If (U_1, \dots, U_n) are Gaussian or multinomial, then an explicit construction of such variables $(Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n)$ having the required conditional distribution, jointly with (U_1, \dots, U_n) is possible. More details on applications of such constructions to sums of nonlinear functions of Gaussian and multinomial variables are given in Section 4.

3 Proofs

Before proving Theorems 1.2 and 1.4, we need the following lemma, the proof of which can be found in Barbour (1990), or Götze (1991).

Let \mathbf{Z} be a standard p -variate normal vector and for $u \geq 0$, define

$$(T_u h)(\mathbf{w}) = E\{h(\mathbf{w}e^{-u} + \sqrt{1 - e^{-2u}}\mathbf{Z})\}.$$

Lemma 3.1 *Let $h : R^p \rightarrow R$ have three bounded derivatives. Then*

$$g(\mathbf{w}) = - \int_0^\infty [T_u h(\mathbf{w}) - \Phi h] du$$

solves

$$\text{tr}D^2g(\mathbf{w}) - \mathbf{w} \cdot \nabla g(\mathbf{w}) = h(\mathbf{w}) - \Phi h,$$

and for any k^{th} partial derivative we have the bound

$$\left| \frac{\partial^k}{\prod_{j=1}^k \partial w_{i_j}} g(\mathbf{w}) \right| \leq \frac{1}{k} \sup_{\mathbf{w}} \left| \frac{\partial^k}{\prod_{j=1}^k \partial w_{i_j}} h(\mathbf{w}) \right| \leq \frac{1}{k} \|D^k h\|.$$

Further, for any $\boldsymbol{\lambda} \in R^p$ and positive definite $p \times p$ matrix Σ , f defined by the change of variable

$$f(\mathbf{w}) = g(\Sigma^{-1/2}(\mathbf{w} - \boldsymbol{\lambda})) \quad (12)$$

solves

$$\text{tr}\Sigma D^2 f(\mathbf{w}) - (\mathbf{w} - \boldsymbol{\lambda}) \cdot \nabla f(\mathbf{w}) = h(\Sigma^{-1/2}(\mathbf{w} - \boldsymbol{\lambda})) - \Phi h, \quad (13)$$

and hence

$$\left| \frac{\partial^k}{\prod_{j=1}^k \partial w_{i_j}} f(\mathbf{w}) \right| \leq \frac{p^k}{k} \|\Sigma^{-1/2}\|^k \|D^k h\|. \quad (14)$$

Proof: One can follow Barbour (1990) to show that g is a solution, and that under the assumptions above, by dominated convergence,

$$D^k g(\mathbf{w}) = - \int_0^\infty e^{-ku} E\{D^k h(\mathbf{w}e^{-u} + \sqrt{1 - e^{-2u}}\mathbf{Z})\} du.$$

The Lemma now follows by straightforward calculations. \square

Proof of Theorem 1.2 Given h , let f be the solution of (13) given by (12). Writing out the expressions in (13) we have

$$E\{h(\Sigma^{-1/2}(\mathbf{W} - \boldsymbol{\lambda})) - \Phi h\} = E\left\{ \sum_{i=1}^p \sum_{j=1}^p \sigma_{ij} \frac{\partial^2}{\partial w_i \partial w_j} f(\mathbf{W}) - \sum_{i=1}^p (W_i - \lambda_i) \frac{\partial}{\partial w_i} f(\mathbf{W}) \right\}. \quad (15)$$

Recall that \mathbf{W}^i can be characterized by (5):

$$EW_i G(\mathbf{W}) = \lambda_i EG(\mathbf{W}^i),$$

holding for all functions $G : R^p \rightarrow R$ for which the expectations exist. Identity (5) is equivalent to

$$E(W_i - \lambda_i)G(\mathbf{W}) = \lambda_i E[G(\mathbf{W}^i) - G(\mathbf{W})]. \quad (16)$$

For the coordinate function $G(\mathbf{w}) = w_j$ we obtain

$$\sigma_{ij} = \text{Cov}(W_i, W_j) = EW_i W_j - \lambda_i \lambda_j = E\lambda_i (W_j^i - W_j); \quad (17)$$

when $i = j$ this recovers the one dimensional relation given in Baldi, Rinott, Stein (1989), $E\lambda(W^* - W) = \sigma^2$, where W^* has the W -size biased distribution. Equation (15), and (16) with $G = \frac{\partial}{\partial w_i} f$, yield

$$E\{h(\Sigma^{-1/2}(\mathbf{W} - \boldsymbol{\lambda})) - \Phi h\} = E\left\{ \sum_{i=1}^p \sum_{j=1}^p \sigma_{ij} \frac{\partial^2}{\partial w_i \partial w_j} f(\mathbf{W}) - \sum_{i=1}^p \lambda_i \left[\frac{\partial}{\partial w_i} f(\mathbf{W}^i) - \frac{\partial}{\partial w_i} f(\mathbf{W}) \right] \right\}. \quad (18)$$

Taylor expansion of $\frac{\partial}{\partial w_i} f(\mathbf{W}^i)$ centered at \mathbf{W} , with remainder in integral form, and simple calculations show that (18) equals

$$\begin{aligned} & -E\left\{ \sum_{i=1}^p \sum_{j=1}^p [\lambda_i (W_j^i - W_j) - \sigma_{ij}] \frac{\partial^2}{\partial w_i \partial w_j} f(\mathbf{W}) \right\} \\ & -E\left\{ \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p \lambda_i \int_0^1 (1-t) \frac{\partial^3}{\partial w_i \partial w_j \partial w_k} f(\mathbf{W} + t(\mathbf{W}^i - \mathbf{W})) (W_j^i - W_j)(W_k^i - W_k) dt \right\}. \end{aligned} \quad (19)$$

In the first term, we condition on \mathbf{W} , apply the Cauchy-Schwarz inequality and use (17), and then apply the bound (14) with $k = 2$ to obtain the first term in (6). The second term in (19) gives the second term in (6) by applying (14) with $k = 3$. \square

Proof of Theorem 1.4 Our proof extends and simplifies the proof of Stein (1986) in the univariate case.

With $\mathbf{W} = (W_1, \dots, W_p)$ where $W_j = \sum_{\alpha \in A_j} X_\alpha$, let $W_j^{(\alpha)} = \sum_{\beta \in A_j \cap S_\alpha^c} X_\beta$, and $\mathbf{W}^{(\alpha)} = (W_1^{(\alpha)}, \dots, W_p^{(\alpha)})$.

Let f be the solution of (13) given by (12 for a test function h . Writing (13) while noting that $\boldsymbol{\lambda} = \mathbf{0}$, and subtracting and adding a term at the end of the expression, we obtain

$$\begin{aligned} E\{h(\Sigma^{-1/2}\mathbf{W}) - \Phi h\} &= E\left\{\sum_{i=1}^p \sum_{j=1}^p \sigma_{ij} \frac{\partial^2}{\partial w_i \partial w_j} f(\mathbf{W}) - \sum_{i=1}^p W_i \frac{\partial}{\partial w_i} f(\mathbf{W})\right\} \\ &= E\left\{\sum_{i=1}^p \sum_{j=1}^p \sigma_{ij} \frac{\partial^2}{\partial w_i \partial w_j} f(\mathbf{W}) \right. \\ &\quad \left. - \sum_{i=1}^p \left[\sum_{\alpha \in A_i} X_\alpha \left[\frac{\partial}{\partial w_i} f(\mathbf{W}) - \frac{\partial}{\partial w_i} f(\mathbf{W}^{(\alpha)}) \right] + \sum_{\alpha \in A_i} X_\alpha \frac{\partial}{\partial w_i} f(\mathbf{W}^{(\alpha)}) \right] \right\}. \end{aligned} \quad (20)$$

Taylor expansion of $\frac{\partial}{\partial w_i} f(\mathbf{W}^{(\alpha)})$ centered at \mathbf{W} and some rearrangement shows that the i^{th} summand in the above expression equals

$$\begin{aligned} E\left\{\sum_{j=1}^p \frac{\partial^2}{\partial w_i \partial w_j} f(\mathbf{W}) [\sigma_{ij} - \sum_{\alpha \in A_i} X_\alpha (W_j - W_j^{(\alpha)})]\right\} &+ E\left\{\sum_{\alpha \in A_i} X_\alpha \frac{\partial}{\partial w_i} f(\mathbf{W}^{(\alpha)})\right\} \\ &- E\left\{\sum_{j=1}^p \sum_{k=1}^p \sum_{\alpha \in A_i} X_\alpha \int_0^1 (1-t) \frac{\partial^3}{\partial w_i \partial w_j \partial w_k} f(\mathbf{W} + t(\mathbf{W}^{(\alpha)} - \mathbf{W})) (W_j - W_j^{(\alpha)}) (W_k - W_k^{(\alpha)}) dt\right\}. \end{aligned} \quad (21)$$

Using (14), applying the Cauchy-Schwarz inequality to the first expectation in (21), and elementary calculations on the remaining two terms yield the three terms of (9) respectively. \square

4 Examples

4.1 Sums of nonlinear functions

Various detailed applications of Theorem 1.1 in the setting of (3) and Lemma 2.2, and related references, are given in Dembo and Rinott (1994). We highlight two problems:

Theorem 4.1 *Let $\mathbf{U} = (U_1, \dots, U_n)$ have the multivariate normal distribution $N(\mathbf{0}, \Sigma)$, where $\Sigma = \{\rho_{ij}\}$ satisfies $\rho_{ii} = 1$ for all i , and $\max_{i \neq j} |\rho_{ij}| \leq r < 1/3$. Let $W = \sum_{i=1}^n \psi(U_i)$ where $0 \leq \psi(u) \leq K e^{K|u|^q}$ for some $K > 0$ and $q < 2$, and ψ scaled such that $N\psi = 1$ (hence $EW = n$). Denote $\sigma^2 = \text{Var}W$. Suppose $\max_i \sum_{j=1}^n |\rho_{ij}| \leq B < \infty$. Define $D = \max\{\|h\|, \|h'\|\}$. Then, for some $C = C(r, K, q) < \infty$,*

$$\left| E h\left(\frac{W - n}{\sigma}\right) - \Phi h \right| \leq \frac{4DB(nBC)^{1/2}}{\sigma^2} + \frac{DCB^2n}{\sigma^3}. \quad (22)$$

The construction of W^* utilizes the well known structure of conditional distributions in the Gaussian case.

An analogous Normal approximation holds for $W = \sum_{i=1}^n \psi(U_i)$, when \mathbf{U} is a vector of multinomial variables with equal (or commensurate) cell probabilities and $\sum_{i=1}^n U_i = kn$ for some integer

$k > 0$. In view of Lemma 2.2 the construction of the coupling can be done as follows: thinking about (U_1, \dots, U_n) as counting the distribution of kn balls in n cells, choose a cell at random, and if cell i is chosen reset the number of balls in it according to the distribution $\psi(u)P(U_i \in u)/E\psi(U_i)$. If doing this requires the addition of balls into cell i , these balls are chosen with equal probability per ball from the other cells. If the resetting requires a reduction in the number of balls in cell i , then a suitable number of balls is redistributed at random in the remaining cells. Now W^* is the sum of the function ψ applied to these adjusted cell counts. This defines W^* on a joint space with W , allowing the calculation of the bound in Theorem 1.1. This construction generalizes to any situation where (U_1, \dots, U_n) have the same distribution as that of some n iid variables conditioned on their sum.

4.2 Graph degree counts

Let $K = K_{n,\pi}$ be a random graph on the vertex set $\{1, \dots, n\}$, where each pair of vertices has probability π of making up an edge, independently of all other such pairs. For distinct, fixed d_i , $i = 1, \dots, p$, let W_i be the number of vertices of degree d_i . Set $\mathbf{W} = (W_1, \dots, W_p)$, $E\mathbf{W} = \boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)$, $\text{Var}\mathbf{W} = \Sigma = (\sigma_{ij})$. For explicit expressions of $\boldsymbol{\lambda}$ and Σ , see (25) below. Set

$$\beta(i) = \binom{n-1}{d_i} \pi^{d_i} (1-\pi)^{n-1-d_i}, \quad \text{and} \quad B = \left[\frac{1}{\min_i \beta(i) (1 - \sum_{i=1}^p \beta(i))} \right]. \quad (23)$$

The theorem below can be extended to the case $n\pi_n \rightarrow c > 0$ as $n \rightarrow \infty$; for simplicity we assume $0 < \pi = c/(n-1) < 1$.

Theorem 4.2 *If $\pi = \pi_n = c/(n-1)$, then for any $h : R^p \rightarrow R$, having bounded mixed partial derivatives up to order 3,*

$$|Eh(\Sigma^{-1/2}(\mathbf{W}-\boldsymbol{\lambda})) - \Phi h| \leq n^{-1/2} M \left\{ p^3 B \|D^2 h\| (1+c^{5/2}) ((n-1)/(n-1-c))^{1/2} + p^5 B^{3/2} \|D^3 h\| (1+c^2) \right\}, \quad (24)$$

where

$$\lambda_i = n\beta(i), \quad \sigma_{ij} = \text{Cov}(W_i, W_j) = n\beta(i)\beta(j) \left[\frac{(d_i - c)(d_j - c)}{c(1 - c/(n-1))} - 1 \right] + \delta_{i,j} n\beta(i), \quad (25)$$

$\delta_{i,j}$ is 1 if $i = j$, 0 otherwise, and M is a universal constant. Asymptotic joint normality obviously follows.

For the case $p = 1$, Karoński and Ruciński (1987) proved asymptotic normality when $\pi n^{(d+1)/d} \rightarrow \infty$ and $\pi n \rightarrow 0$, or $\pi n \rightarrow \infty$ and $\pi n - \log n - d \log \log n \rightarrow -\infty$. See also Palka (1984) and Bollobás (1985). Asymptotic normality when $\pi n \rightarrow c > 0$, was obtained by Barbour, Karoński and Ruciński (1989). See also Kordecki (1990) for the case of the one dimensional distribution of the number of vertices of degree zero, for nonsmooth h . Numerous univariate results on asymptotic normality of counts on random graphs, including counts of the type discussed in Theorems 4.2 and 4.3, are given in Janson and Nowicki (1991) and references therein.

We remark that the calculation of a bound on the conditional variance in Theorem 1.2, as well as other terms, is usually involved in nontrivial cases. The technical details omitted in the following sketch of the proof of Theorem 4.2 are available in Goldstein and Rinott (1994).

Sketch of Proof: Let $D(v)$ denote the degree of vertex v in K and set $X_{vi} = 1$ if $D(v) = d_i$ and 0 otherwise. We have $W_i = \sum_{v=1}^n X_{vi}$, $i = 1, \dots, p$. Note that $D(v) \sim \text{Binomial}(n-1, \pi)$ which approaches $\text{Poisson}(c)$ as $n \rightarrow \infty$. Note that $\beta(i)$ in (23) equals $EX_{vi} = P(D(v) = d_i)$, and so the expression for λ_i in (25) is immediate. Also, by conditioning on the existence of an edge between vertices v and u and then unconditioning, we can compute $EX_{vi}X_{uj}$, and a straightforward

calculation leads to the expression for σ_{ij} in (25). Using the fact that the maximum absolute value of $\|\Sigma^{-1/2}\|$ is bounded by the largest eigenvalue of $\Sigma^{-1/2}$, and invoking the Rayleigh-Ritz characterization of eigenvalues, we obtain with some calculation, $\|\Sigma^{-1/2}\| \leq N^{-1/2}B^{1/2}$.

The construction required for the application of Theorem 1.2 is straightforward. Fix i and let V be uniformly distributed on the vertex set $\{1, \dots, n\}$, independent of K . (Note that here V is uniform because $\gamma_{vi} = EX_{vi}$, $v = 1, \dots, n$ are all equal for a fixed i .) Now $D(V)$ denotes the degree of the randomly chosen vertex V . If $D(V) > d_i$ define K^i to differ from K only in that $D(V) - d_i$ edges selected uniformly from the $D(V)$ edges at V are removed from the edge-set. If $D(V) < d_i$ define K^i to be the graph obtained from K by adding $d_i - D(V)$ edges of the form (V, u) , where the vertices u are selected uniformly from the $n - 1 - D(V)$ vertices not connected to V . If $D(V) = d_i$, $K^i = K$. Clearly, If $V = v$, then in the new graph K^i the degree of v is d_i so that the indicator X_{vi} is size biased to 1, and the distribution of K^i is the same as the conditional distribution of K given $X_{vi} = 1$.

Define \mathbf{W}^i to be related to K^i as \mathbf{W} is related to K , that is, set $X_{vj}^i = 1$ if in the graph K^i , $D(v) = d_j$, otherwise 0, and $W_j^i = \sum_{u=1}^n X_{uj}^i$, $j = 1, \dots, p$. From the discussion in Section 2 it follows that this procedure defines \mathbf{W}^i as in Definition 1.2 and Theorem 1.2.

In order to obtain a tractable bound to the first term on the right hand side of (6), we condition on a larger σ -field, as discussed in Section 2. Specifically, we use the relation $\text{Var}E[W_j^i - W_j | \mathbf{W}] \leq \text{Var}E[W_j^i - W_j | K]$, and show that

$$\text{Var}E[W_j^i - W_j | K] = O((1 + c^3)(1 + \frac{d_i^2}{1 - c/(n-1)})/n). \quad (26)$$

Let \mathcal{E} denotes the edge set of K and $|\cdot|$ cardinality. Conditioning on $V = v$ and then taking expectation, recalling $P(V = v) = 1/n$, we obtain

$$\begin{aligned} E[W_j^i - W_j | K] &= \frac{1}{n} \sum_{v: D(v) > d_i} \left[\left| \{u : (u, v) \in \mathcal{E}, D(u) = d_j + 1\} \right| - \left| \{u : (u, v) \in \mathcal{E}, D(u) = d_j\} \right| \right] \frac{D(v) - d_i}{D(v)} \\ &+ \frac{1}{n} \sum_{v: D(v) < d_i} \left[\left| \{u : (u, v) \notin \mathcal{E}, D(u) = d_j - 1\} \right| - \left| \{u : (u, v) \notin \mathcal{E}, D(u) = d_j\} \right| \right] \frac{d_i - D(v)}{n - 1 - D(v)} \\ &+ \frac{1}{n} \left| \{v : D(v) \neq d_i\} \right| \delta_{i,j} - \frac{1}{n} \left| \{v : D(v) = d_j\} \right| (1 - \delta_{i,j}). \end{aligned} \quad (27)$$

To understand the first term, for example, note that that if $V = v$ and $D(v) > d_i$, then $X_{vj}^i - X_{uj} = 1$ if $(u, v) \in \mathcal{E}$, $D(u) = d_j + 1$, and (u, v) is one of the $d_i - D(v)$ edges removed at v at random, chosen with probability $(D(v) - d_i)/D(v)$.

The calculation of a bound on the variance of the expression in (27) can be done by computing the covariances between the terms. They involve conditioning on events to induce independence of terms appearing as products in the covariances, the use of simple coupling arguments and various moment inequalities.

The bound for the second term on the right hand side of (6) is obtained by noting that $|W_j^i - W_j| \leq |D(V) - d_i| + 1$ and applying simple calculations related to the Binomial distribution of $D(V)$. \square

4.3 Graph Vertex Color Matching

Let $G = G_n$ be a fixed regular graph on a vertex set \mathcal{V} of size n , with each vertex $v \in \mathcal{V}$ of degree d . The regularity of G implies the set \mathcal{E} of edges of G has size $N = nd/2$. Let $C = \{1, \dots, p\}$ be a set

of p colors, and suppose that each vertex $v \in \mathcal{V}$ is independently assigned color i with probability π_i . Let $B = \left[\frac{1}{\min_i \pi_i^2(1-\pi_i)} \right]$.

Theorem 4.3 *Let $\mathbf{W} = (W_1, \dots, W_p)$ where $W_i, i = 1, \dots, p$ is the number of edges of G that have both vertices of color i . Then $E\mathbf{W} = \boldsymbol{\lambda} = N(\pi_1^2, \dots, \pi_p^2)$ and $\mathbf{Var}(\mathbf{W}) = \Sigma = (\sigma_{ij})$ as given in equations (29) and (30), and for any $h : R^p \rightarrow R$, having bounded mixed partial derivatives up to order 3,*

$$|Eh(\Sigma^{-1/2}(\mathbf{W} - \boldsymbol{\lambda})) - \Phi h| \leq N^{-1/2} M \left\{ p^4 B \|D^2 h\| d^{3/2} + p^6 B^{3/2} \|D^3 h\| d^2 \right\}, \quad (28)$$

where M is a universal constant. Asymptotic joint normality obviously follows.

Proof: First we will obtain $\Sigma = \mathbf{Var}\mathbf{W}$ in order to bound $\|\Sigma^{-1/2}\|$. Let $X_{ei} = 1$ if edge e has color i on both vertices and 0 otherwise, so $W_i = \sum_{e \in \mathcal{E}} X_{ei}$ counts the number of edges with both vertices of color i . We have $EX_{ei} = \pi_i^2$, and $\mathbf{Var}X_{ei} = \pi_i^2(1 - \pi_i^2)$. Given an edge e , let S_e denote the set of $2d - 1$ edges that share a vertex with e , including e itself. For the $2(d - 1)$ edges $f \in S_e, f \neq e$, $\mathbf{Cov}(X_{ei}, X_{fi}) = \pi_i^3 - \pi_i^4$. For $f \notin S_e$, this covariance is 0 by independence. Thus,

$$\sigma_{ii} = \mathbf{Var}(W_i) = N\pi_i^2(1 - \pi_i^2) + 2N(d - 1)(\pi_i^3 - \pi_i^4). \quad (29)$$

For different colors $i \neq j$ for $f \in S_e$, we have $\mathbf{Cov}(X_{ei}, X_{fj}) = -\pi_i^2\pi_j^2$; again, for $f \notin S_e$ this covariance is 0. Hence,

$$\text{for } i \neq j, \quad \sigma_{ij} = \mathbf{Cov}(W_i, W_j) = -N(2d - 1)\pi_i^2\pi_j^2. \quad (30)$$

Let A and H be the diagonal matrices with i^{th} diagonal entry $\pi_i^3, \pi_i^2 - \pi_i^3$ respectively, and let b be a column vector with i^{th} component π_i^2 . Then $\Sigma = N(2d - 1)[A - bb^t] + NH$. In order to show that $\Sigma \succeq NH$, let D be the diagonal matrix with diagonal entries $\pi_i^{3/2}$, and g the column vector with entries $\pi_i^{1/2}$. Then $A - bb^t = D(I - gg^t)D$. Since $\Sigma\pi_i = 1$, it is easy to see that the smallest eigenvalue of $I - gg^t$ is 0. Hence, $A - bb^t$ is nonnegative definite and $\Sigma \succeq NH$ is established. It follows that $\|\Sigma^{-1/2}\| \leq N^{-1/2}B^{1/2}$.

We now apply Theorem 1.4 to the mean zero variables $X_{ei} - \pi_i^2$. When the square in the first term in the bound (9) is expanded and expectation is taken, most terms vanish by independence, and because $|S_e| = 2d - 1$, the number of summands which do not vanish under the root sign is of the order Nd^3 . The second term in (9) vanishes, and in the third term each expectation is of order d^2 . \square

References

- Baldi, P. Rinott, Y. (1989). *On normal approximations of distributions in terms of dependency graphs*, Annals of Probability **17**, 1646-1650.
- Baldi, P. Rinott, Y. and Stein, C. (1989). *A normal approximations 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.
- Barbour, A.D. (1990) *Stein's method for diffusion approximations*, Probab. Th. Rel. Fields **84** 297-322.
- Barbour, A. D., Karoński, M. and Ruciński, A. (1989). *A central limit theorem for decomposable random variables with applications to random graphs*, J. Combinatorial Theory B **47**, 125-145.

- Bollobás, B. (1985) *Random Graphs*. Academic Press, 1985.
- Brewer, K. and Hanif, M. (1983) *Sampling with unequal probabilities*, Lecture Notes in Statistics, vol. 15. Springer-Verlag, New York.
- Cochran, W. (1977) *Sampling Techniques* John Wiley & Sons, New York.
- Dall'Aglio, S, Kotz, S., and Salinetti, G. (Eds.) (1991) *Advances in Probability Distributions With Given Marginals*. Kluwer Academic Publishers, Dordrecht, Boston, London.
- Dembo, A, and Rinott, Y. (1994). *Some examples of Normal approximations by Stein's method*. To appear in IMA Conference Proceedings, Aldous, D. and Pemantle, R. Eds.
- Goldstein, L., and Rinott, Y. (1994) *On multivariate normal approximations by Stein's method and size bias couplings: Technical Report*.
- Götze, F. (1991) *On the rate of convergence in the multivariate CLT*. Annals of Probability, **19**, 724-739.
- Horn, R. A., and Johnson, C. A. (1985) *Matrix Analysis*. Cambridge University Press 1985.
- Janson, S. and Nowicki, K. (1991) *The asymptotic distributions of generalized U-statistics with applications to random graphs*. Probability Theory and Related Fields **90**, 341-375.
- Karoński, M. and Ruciński A. (1987), *Poisson convergence of semi-induced properties of random graphs*. Math. Proc. Comb. Phil. Soc. **101** 291-300.
- Kordecki, W. (1990) *Normal approximation and isolated vertices in random graphs*. In Random Graphs 1987, M. Karoński, J. Jaworski and A. Ruciński eds. John Wiley & Sons, New York.
- Luk, H. M. (1994) *Stein's method for the Gamma distribution and related statistical applications*. PhD dissertation, USC.
- Palka, Z. (1984) *On the number of vertices of a given degree in a random graph*. J. Graph Theory **8**, 167-170.
- Reinert, G. (1994) *A weak law of large numbers for empirical measures via Stein's method, and applications*. PhD dissertation.
- Rinott, Y. (1994) *On normal approximation rates for certain sums of dependent random variables*. To appear in Journal of applied and Comp. Math.
- Rinott, Y. and Rotar, V. (1994) *A multivariate CLT for local dependence with $n^{-1/2} \log n$ rate*.
- 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.
- Stein, C. (1986) *Approximate Computation of Expectations*. IMS, Hayward, Calif., 1986.
- Stein, C. (1992) *A way of using auxiliary randomization* Probability Theory, pp. 159-180. Walter de Gruyter & Co., Berlin - New York.