The Annals of Applied Probability
2013, Vol. 23, No. 2, 617–636
DOI: 10.1214/12-AAP848

A BERRY–ESSEEN BOUND WITH APPLICATIONS TO VERTEX DEGREE COUNTS IN THE ERDŐS–RÉNYI RANDOM GRAPH

BY LARRY GOLDSTEIN[1]

University of Southern California

Applying Stein's method, an inductive technique and size bias coupling yields a Berry–Esseen theorem for normal approximation without the usual restriction that the coupling be bounded. The theorem is applied to counting the number of vertices in the Erdős–Rényi random graph of a given degree.

1. Introduction. We present a new Berry–Esseen theorem for sums Y of dependent variables by combining Stein's method, size bias couplings and the inductive technique of Bolthausen (1984) originally developed for the combinatorial central limit theorem. We apply the theorem to asses the accuracy of the normal approximation to the distribution of the number of vertices of degree d in the classical Erdős–Rényi (1959) random graph G_n having n vertices connected by independent edges with common success probability depending on n and a parameter θ. Over the range of parameters considered, the theorem yields a bound that is the same up to constants as the one obtained earlier by Barbour, Karoński and Ruciński (1989) for the weaker smooth function metric (19).

Stein's method [Stein (1972, 1986)] often proceeds by coupling a random variable Y of interest to a related variable Y', using, for example, the method of exchangeable pairs, size bias couplings or zero bias couplings; for an overview see Chen, Goldstein and Shao (2010). The chief innovation here is the removal of an inconvenient restriction present in a number of results that provide Kolmogorov distance bounds using Stein's method, that the difference $|Y - Y'|$ between Y and the coupled Y' be bounded almost surely by a constant. Through the use of an unbounded coupling, in Theorem 2.1 we are able to extend the previous work by Kordecki (1990) on the number of isolated, or degree zero, vertices of G_n to all positive degrees.

To describe Theorem 1.1, our general result, recall that for a nonnegative random variable Y with finite, nonzero mean μ, we say that Y^s has the Y-size bias distribution if

$$E[Yf(Y)] = \mu E[f(Y^s)] \tag{1}$$

for all functions f for which these expectations exist.

Received May 2010; revised December 2011.
[1]Supported in part by NSA Grant H98230-11-1-0162.
MSC2010 subject classifications. 60F05, 60C05, 05C80.
Key words and phrases. Stein's method, size bias coupling, inductive method, random graphs.

In employing the size bias version of Stein's method [see Baldi, Rinott and Stein (1989), Goldstein and Rinott (1996) and Chen, Goldstein and Shao (2010)], the goal is to construct, on the same space as Y, a variable Y^s with the Y-size bias distribution such that Y and Y^s are close is some sense. Previous applications of the size bias coupling technique for obtaining Berry–Esseen bounds by Stein's method, requiring that $|Y^s - Y|$ be bounded, include Goldstein (2005), Goldstein and Penrose (2010) and Goldstein and Zhang (2011).

Let $\mathbb{N} = \{1, 2, \ldots\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Our abstract framework consists of random elements indexed by $n \geq n_0$ for some $n_0 \in \mathbb{N}_0$ whose distributions $\mathcal{L}_\theta(\cdot)$ depend on n, left implicit when clear from context, and a parameter θ in a topological space Θ_n. We also assume that Θ_n is endowed with a σ-algebra, taken to be the one generated by the collection of open sets unless specified otherwise.

In our application the parameter θ lies in a subset Θ_n of the real numbers $\mathbb{R}$ and interest centers on the distributions of the nonnegative random variables Y_n counting the number of degree $d \in \mathbb{N}$ vertices of the Erdős–Rényi random graph G_n. For sums of exchangeable indicator variables such as Y_n, Lemma 1.1 below says, essentially, that to construct a variable Y_n^s with the Y_n-size bias distribution, one chooses an indicator uniformly and sets it to one if it was not so already, and then "adjusts" the remaining indicators, if necessary, to have their original distribution given that the selected indicator is one. Applying Lemma 1.1 when Y_n counts the number of vertices in G_n having degree d results in the construction of Barbour, Holst and Janson (1992), where nothing is changed if a uniformly chosen vertex already has degree d, and otherwise edges to the chosen vertex are added if the vertex has degree less than d, or removed if it has degree in excess of d. As it is possible that the chosen vertex has, say, $n - 1$ edges, the resulting coupling fails to be bounded in n. Nevertheless, when there is only a small probability that a very large number of edges will need to be added or removed, the coupling can be controlled using moments on bounds K_n that satisfy $|Y_n^s - Y_n| \leq K_n$.

After coupling, the second ingredient in our method has an inductive flavor. We construct a variable V_n such that its distribution, conditional on a collection J_n of random elements, is that of Y_n reduced in size by some "small" amount L_n, with parameter $\psi_{n,\theta}$ "close" to the original θ. Formally, we require that

$$\mathcal{L}_\theta(V_n | J_n) = \mathcal{L}_{\psi_{n,\theta}}(Y_{n-L_n}) \tag{2}$$

hold on an event where the size of L_n is controlled, and that a bound B_n on the absolute difference $|Y_n - V_n|$ not be "too large." As bounds to the normal for Y_n can be expressed in terms of quantities that include bounds to the normal for reduced versions of the same problem, a recursive inequality for the sought after bound can be produced.

In the graph degree problem, V_n counts the number of degree d vertices in the graph obtained by removing a uniformly chosen vertex from G_n, along with all its incident edges, and the set J_n consists of the identity of the chosen vertex, and its degree. Conditionally on J_n, the graph that remains is an Erdős–Rényi graph on the reduced vertex set, with the same connectivity as before. As with the bound K_n, it is not required that B_n be almost surely bounded by a constant; though $|Y_n - V_n|$ may be large in the graph degree problem, it is unlikely that it will be.

Tension exists in choosing the set J_n that appears in the conditioning equality (2). In order to reduce the larger problem to a smaller one so that induction may be applied, working conditionally we must be able to treat the bounds K_n and B_n, and the parameters of the reduced problem, L_n and $\psi_{n,\theta}$, as constants. Hence we require that these variables be measurable with respect to $\mathcal{F}_n$, the σ-algebra generated by the conditioning collection J_n. Though this restriction necessitates that $\mathcal{F}_n$ be large enough to contain, say, information on $Y_n^s - Y_n$, it must also be small enough so that L_n and B_n are not too large, and that the conditioning "leaves enough randomness" to yield a useful recursion for the ultimate bound.

At the heart of our main result, and Stein's method for normal approximation, is the characterization that Z is a standard normal random variable if and only if

$$E[Zf(Z)] = E[f'(Z)]$$

for all absolutely continuous functions f for which the above expectations exist. This characterization leads to the Stein equation, when, given a test function h on which to evaluate the difference $Eh(W) - Eh(Z)$ between the expectation of the random variable W of interest and the standard normal Z, one solves

$$f'(w) - wf(w) = h(w) - Eh(Z)$$

for f. Using f, one evaluates this difference by substituting W for w, and takes expectation on the left-hand side, rather than the right. Though we focus on manipulation of the Stein equation using the size bias coupling, many variations are possible; see Chen, Goldstein and Shao (2010) for an overview.

Throughout, for $n_0 \in \mathbb{N}$ and all $n \ge n_0$ and $\theta \in \Theta_n$, we let $\mu_{n,\theta} = E_\theta Y_n$ and $\sigma_{n,\theta}^2 = \mathrm{Var}_\theta(Y_n)$ indicate the mean and variance of Y_n under $\mathcal{L}_\theta$. The value $r_{n,\theta}$ appearing in Theorem 1.1 is a function that determines the quality of the bound to the normal, while the sequence $s_{n,\theta}$ is used to control L_n, and hence the size of the smaller subproblem V_n related to Y_n. Without further mention, $\mu_{n,\theta}, \sigma_{n,\theta}^2$ and $r_{n,\theta}$ are assumed to be measurable in $\theta \in \Theta_n$, a condition satisfied for all natural examples, including the one considered here. To avoid repetition, the distribution of random variables indicated after $\theta \in \Theta_n$ has been fixed is with respect to $\mathcal{L}_\theta$. The random variable Z will always denote the standard normal.

To familiarize the reader with the conditions of Theorem 1.1, toward the end of this section we present its application in the simple case where a bounded size bias coupling of Y_n^s to Y_n exists.

THEOREM 1.1. *For some $n_0 \in \mathbb{N}_0$ and all $n \ge n_0$, let Y_n be a nonnegative random variable with mean $\mu_{n,\theta} = E_\theta Y_n$ and positive variance $\sigma_{n,\theta}^2 = \mathrm{Var}_\theta(Y_n)$ for all $\theta \in \Theta_n$, and set*

$$W_{n,\theta} = \frac{Y_n - \mu_{n,\theta}}{\sigma_{n,\theta}}, \tag{3}$$

the standardized value of Y_n. *Let* $r_{n,\theta}$ *be positive for all* $n \geq n_0$ *and all* $\theta \in \Theta_n$, *and for all* $r \geq 0$ *let*

$$\Theta_{n,r} = \{\theta \in \Theta_n : r_{n,\theta} \geq r\}.$$

Assume there exists $r_1 > 0$ *and* $n_1 \geq n_0$ *such that*

$$\max_{n_0 \leq n < n_1} \sup_{\theta \in \Theta_{n,r_1}} r_{n,\theta} < \infty. \tag{4}$$

Further, suppose that for all $n \geq n_1$ *and* $\theta \in \Theta_{n,r_1}$, *there exist random variables* $Y_n^s, K_n, L_n, \psi_{n,\theta}, V_n$ *and* B_n *on the same space as* Y_n, *and a* σ*-algebra* $\mathcal{F}_n$, *generated by a collection of random elements* J_n, *such that the following conditions hold*:

1. *The random variable* Y_n^s *has the* Y_n*-size bias distribution, and*

$$\begin{gathered} \Psi_{n,\theta} = \sqrt{\operatorname{Var}_\theta\bigl(E_\theta(Y_n^s - Y_n | Y_n)\bigr)} \qquad \textit{satisfies} \\ \sup_{n \geq n_1, \theta \in \Theta_{n,r_1}} \frac{r_{n,\theta}\mu_{n,\theta}\Psi_{n,\theta}}{\sigma_{n,\theta}^2} < \infty. \end{gathered} \tag{5}$$

2. *The random variable* K_n *is* $\mathcal{F}_n$*-measurable*, $|Y_n^s - Y_n| \leq K_n$ *and*

$$\sup_{n \geq n_1, \theta \in \Theta_{n,r_1}} \frac{r_{n,\theta}\mu_{n,\theta} E_\theta[(1 + |W_{n,\theta}|)K_n^2]}{\sigma_{n,\theta}^3} < \infty \tag{6}$$

with $W_{n,\theta}$ *as given in* (3).

3. *The random variable* L_n *takes values in* $\{0, 1, \ldots, n\}$, *there exists a positive integer valued sequence* $\{s_{n,\theta}\}_{n \geq n_1}$ *satisfying* $n - s_{n,\theta} \geq n_0$, *the variables* L_n *and* $\psi_{n,\theta}$ *are* $\mathcal{F}_n$*-measurable, for some* $F_{n,\theta} \in \mathcal{F}_n$ *satisfying* $F_{n,\theta} \subset \{L_n \leq s_{n,\theta}\}$,

$$\psi_{n,\theta} \in \Theta_{n-L_n} \quad \textit{and} \quad \mathcal{L}_\theta(V_n | J_n) = \mathcal{L}_{\psi_{n,\theta}}(Y_{n-L_n}) \qquad \textit{on } F_{n,\theta} \tag{7}$$

and

$$\sup_{n \geq n_1, \theta \in \Theta_{n,r_1}} \frac{r_{n,\theta}^2 \mu_{n,\theta}}{\sigma_{n,\theta}^3} E_\theta\bigl[K_n^2\bigl(1 - \mathbf{1}(F_{n,\theta})\bigr)\bigr] < \infty. \tag{8}$$

4. *There exists* $\{c_1, c_2\} \subset (0, \infty)$ *such that*

$$\sigma_{n,\theta}^2 \leq c_1 \sigma_{n-L_n,\psi_{n,\theta}}^2 \quad \textit{and} \quad r_{n,\theta} \leq c_2 r_{n-L_n,\psi_{n,\theta}} \qquad \textit{on } F_{n,\theta}.$$

5. *The random variable* B_n *is* $\mathcal{F}_n$*-measurable*, $|Y_n - V_n| \leq B_n$ *and*

$$\sup_{n \geq n_1, \theta \in \Theta_{n,r_1}} \frac{r_{n,\theta}^2 \mu_{n,\theta} E_\theta[K_n^2 B_n]}{\sigma_{n,\theta}^4} < \infty. \tag{9}$$

6. *Either*:

(a) *there exists* $l_{n,0} \in \{0, \dots, n\}$ *such that* $P_\theta(L_n = l_{n,0}) = 1$ *for all* $\theta \in \Theta_{n,r_1}$, *or*

(b) *the set* Θ_{n,r_1} *is a compact subset of* Θ_n, *and the functions of* θ

$$t_{n,\theta,l} = E_\theta\left(\frac{K_n^2}{E_\theta K_n^2}\mathbf{1}(L_n = l)\right), \qquad l \in \{0, 1, \dots, n\} \tag{10}$$

are continuous on Θ_{n,r_1} *for* $l \in \{0, 1, \dots, s_n\}$ *where* $s_n = \sup_{\theta \in \Theta_{n,r_1}} s_{n,\theta}$.

Then there exists a constant C *such that for all* $n \ge n_0$ *and* $\theta \in \Theta_n$,

$$\sup_{z \in \mathbb{R}} |P_\theta(W_{n,\theta} \le z) - P(Z \le z)| \le C/r_{n,\theta}. \tag{11}$$

When higher moments exist a number of the conditions of the theorem may be verified using standard inequalities. In particular, by the Cauchy–Schwarz inequality a sufficient condition for (6) is

$$\sup_{n \ge n_1, \theta \in \Theta_{n,r_1}} \frac{r_{n,\theta}\mu_{n,\theta}k_{n,\theta,4}^{1/2}}{\sigma_{n,\theta}^3} < \infty \qquad \text{where } k_{n,\theta,m} = E_\theta K_n^m, \tag{12}$$

and, when $F_{n,\theta} = \{L_n \le s_{n,\theta}\}$, a sufficient condition for (8) is

$$\sup_{n \ge n_1, \theta \in \Theta_{n,r_1}} \frac{r_{n,\theta}^2\mu_{n,\theta}k_{n,\theta,4}^{1/2}l_{n,\theta,2}^{1/2}}{\sigma_{n,\theta}^3 s_{n,\theta}} < \infty \qquad \text{where } l_{n,\theta,m} = E_\theta L_n^m,$$

since, additionally using the Markov inequality yields

$$\begin{aligned} E_\theta[K_n^2\mathbf{1}(L_n > s_{n,\theta})] &\le k_{n,\theta,4}^{1/2}P_\theta(L_n > s_{n,\theta})^{1/2} = k_{n,\theta,4}^{1/2}P_\theta(L_n^2 > s_{n,\theta}^2)^{1/2} \\ &\le \frac{k_{n,\theta,4}^{1/2}l_{n,\theta,2}^{1/2}}{s_{n,\theta}}. \end{aligned}$$

Similarly, a sufficient condition for (9) is

$$\sup_{n \ge n_1, \theta \in \Theta_{n,r_1}} \frac{r_{n,\theta}^2\mu_{n,\theta}k_{n,\theta,4}^{1/2}b_{n,\theta,2}^{1/2}}{\sigma_{n,\theta}^4} < \infty \qquad \text{where } b_{n,\theta,m} = E_\theta B_n^m. \tag{13}$$

Regarding (7) we remark that by $\mathcal{L}_\theta(Y_{n-L_n})$ we mean the mixture distribution $\sum_{m=n_0}^n \mathcal{L}_\theta(Y_m)P(L_n = n - m)$, which can be defined without requiring that $Y_{n_0}, \dots, Y_n$ and L_n all be defined on the same space. A general prescription for size biasing a sum of nonnegative variables is given in Goldstein and Rinott (1996); specializing to exchangeable indicators yields the following result.

LEMMA 1.1. *Let* $Y = \sum_{\alpha \in \mathcal{I}} X_\alpha$ *be a finite sum of nontrivial exchangeable Bernoulli variables* $\{X_\alpha, \alpha \in \mathcal{I}\}$, *and suppose that for* $\alpha \in \mathcal{I}$ *the variables* $\{X_\beta^\alpha, \beta \in \mathcal{I}\}$ *have joint distribution*

$$\mathcal{L}(X_\beta^\alpha, \beta \in \mathcal{I}) = \mathcal{L}(X_\beta, \beta \in \mathcal{I} | X_\alpha = 1).$$

Then

$$Y^\alpha = \sum_{\beta \in \mathcal{I}} X_\beta^\alpha$$

has the Y-size biased distribution Y^s, as does the mixture Y^I when I is a random index with values in $\mathcal{I}$, independent of all other variables.

PROOF. First, fixing $\alpha \in \mathcal{I}$, we show that Y^α satisfies (1). For given f,

$$E[Yf(Y)] = \sum_{\beta \in \mathcal{I}} E[X_\beta f(Y)] = \sum_{\beta \in \mathcal{I}} P[X_\beta = 1] E[f(Y)|X_\beta = 1].$$

As exchangeability implies that $E[f(Y)|X_\beta = 1]$ does not depend on β, we have

$$E[Yf(Y)] = \left(\sum_{\beta \in \mathcal{I}} P[X_\beta = 1]\right) E[f(Y)|X_\alpha = 1] = E[Y]E[f(Y^\alpha)],$$

demonstrating the first result. The second follows easily using that Y^I is a mixture of random variables all of which have distribution Y^s. □

Employing size bias couplings and Stein's method, Chen and Röllin (2010) prove a general result to compute bounds to the normal in the Waserstein metric. In particular, Corollary 2.2 and Construction 3A of Chen and Röllin (2010) yield

$$d_W(\mathcal{L}_\theta(W_{n,\theta}), \mathcal{L}(Z)) \le 0.8 \frac{\mu_{n,\theta}\Psi_{n,\theta}}{\sigma_{n,\theta}^2} + \frac{\mu_{n,\theta} k_{n,\theta,2}}{\sigma_{n,\theta}^3}. \tag{14}$$

To compare (14) with one conclusion of Theorem 1.1, as well as to familiarize the reader with the roles of some of the variables appearing in its formulation, we now consider its application in the simple case where a bounded size bias coupling exists, that is, when the bound K_n on $|Y_n^s - Y_n|$ can be taken to be a constant, say k_n, almost surely. In such cases we set J_n to be the empty set, and note that any constant is measurable with respect to the trivial σ-algebra that J_n generates. Conditions 3 through 6 are easily satisfied in this case for any candidate $r_{n,\theta}$. In particular, taking $L_n = 0, s_{n,\theta} = 1$ and $F_{n,\theta} = \{L_n \le s_{n,\theta}\}$, with $J_n = \varnothing$, (7) of Condition 3 holds with $\psi_{n,\theta} = \theta$ and $V_n = Y_n$, and (8) holds as $1 - \mathbf{1}(F_{n,\theta}) = 0$ a.s. As $(n - L_n, \psi_{n,\theta}) = (n, \theta)$, Condition 4 holds with $c_1 = c_2 = 1$. As $V_n = Y_n$ we may take $B_n = 0$ in Condition 5, and as $L_n = 0$ Condition 6a is satisfied. Hence, only Conditions 1 and 2 are in force, and Theorem 1.1 obtains with

$$r_{n,\theta}^{-1} = \frac{\mu_{n,\theta}\Psi_{n,\theta}}{\sigma_{n,\theta}^2} + \frac{\mu_{n,\theta} k_n^2}{\sigma_{n,\theta}^3},$$

yielding a Kolmogorov bound that, up to constants, agrees with the Wasserstein bound (14) in this particular case.

Bounded size bias couplings exist when Y_n is the sum of independent, bounded nonnegative random variables, or a sum of bounded, nonnegative locally dependent variables with bounded dependence neighborhood sizes, as studied, for instance, in Goldstein (2005). In addition, bounded size bias couplings can also be constructed in cases of global dependence; see Goldstein and Zhang (2011) or Goldstein and Penrose (2010).

We next apply Theorem 1.1 to vertex degree counts in the Erdős–Rényi random graph. The proof of Theorem 1.1 is given in Section 3.

2. Vertex degree in the Erdős–Rényi random graph. We apply Theorem 1.1 to bound the error in the normal approximation to the distribution of the number of vertices of a given degree in the Erdős–Rényi (1959) random graph G_n; see also Bollobás (1985). With $n \in \mathbb{N}$ we take the vertex set of G_n to be $\mathcal{I}_n = \{1, \dots, n\}$, and the indicators $\xi_{u,v}$ of the presence of edges between distinct vertices u and v to be independent Bernoulli variables with a common success probability. No vertex is connected to itself, and we set $\xi_{u,u} = 0$ for all $u \in \mathcal{I}_n$.

The number Y_n of vertices of degree d of G_n has been the object of much study. For a sequence of graphs with connectivity probability p depending on $n \in \mathbb{N}$, Karoński and Ruciński (1987) proved the asymptotic normality of Y_n when $n^{(d+1)/d} p \to \infty$ and $np \to 0$, or $np \to \infty$ and $np - \log n - d \log\log n \to -\infty$; see also Palka (1984) and Bollobás (1985). Asymptotic normality of Y_n when $np \to c > 0$ was obtained by Barbour, Karoński and Ruciński (1989), and Kordecki (1990) for nonsmooth functions of Y_n in the case $d = 0$. Neammanee and Suntadkarn (2009) obtain a Kolmogorov distance bound between Y_n and the normal with rate $n^{-1/2+\varepsilon}$ for all $\varepsilon > 0$ when $\operatorname{Var}(Y_n)$ is of order n. Other univariate results on asymptotic normality of counts on random graphs are given in Janson and Nowicki (1991), and references therein. Goldstein and Rinott (1996) obtain smooth function bounds for the vector whose k components count the number of vertices of fixed degrees $d_1, d_2, \dots, d_k$ when $p = \theta/(n-1) \in (0,1)$ for fixed θ, implying asymptotic multivariate joint normality.

We focus on the counts of vertices of some fixed degree $d \in \mathbb{N}$, the case $d = 0$ of isolated vertices having already been handled by Kordecki (1990). Set

$$\Theta_n = (0, n-1) \cap (0, b] \qquad \text{for all } n \ge d+1 \tag{15}$$

with b some arbitrarily large constant, and let the connectivity probability between the vertices of G_n be given by $\theta/(n-1)$ for $n \ge d+1, \theta \in \Theta_n$. For $v \in \mathcal{I}_n$ let

$$D_n(v) = \sum_{w \in \mathcal{I}_n} \xi_{v,w}, \qquad X_{n,v} = \mathbf{1}(D_n(v) = d) \quad \text{and} \quad Y_n = \sum_{v \in \mathcal{I}_n} X_{n,v},$$

the degree of vertex v, the indicator that vertex v has degree d, and the number of vertices of degree d of G_n, respectively.

From Goldstein and Rinott (1996), for all $n \ge d+1$ and $\theta \in \Theta_n$, the mean $\mu_{n,\theta}$ and variance $\sigma^2_{n,\theta}$ of Y_n are given explicitly by

$$\mu_{n,\theta} = n\tau_{n,\theta} \quad \text{and} \quad \sigma^2_{n,\theta} = n\tau^2_{n,\theta}\left[\frac{(d-\theta)^2}{\theta(1-\theta/(n-1))} - 1\right] + n\tau_{n,\theta}, \tag{16}$$

where

$$\tau_{n,\theta} = \binom{n-1}{d}\left(\frac{\theta}{n-1}\right)^d \left(1 - \frac{\theta}{n-1}\right)^{n-1-d}. \tag{17}$$

THEOREM 2.1. *For any $d \in \mathbb{N}$ and $b > 0$ there exists a constant C such that for all $n \ge d+1$ and all $\theta \in \Theta_n$ given in* (15), *the normalized count $W_{n,\theta}$ in* (3) *of the number Y_n of vertices with degree d in the Erdős–Rényi random graph G_n on n vertices, with edges connecting each distinct pair independently with probability $\theta/(n-1)$, satisfies*

$$\sup_{z\in\mathbb{R}} |P_\theta(W_{n,\theta} \le z) - P(Z \le z)| \le C/r_{n,\theta} \qquad \text{for all } n \ge d+1,$$

where Z is a standard normal variable and

$$r_{n,\theta} = \sqrt{n\tau_\theta} \qquad \text{with } \tau_\theta = e^{-\theta}\theta^d/d!. \tag{18}$$

By applying Stein's method, Barbour, Karoński and Ruciński (1989) obtain a bound of order $1/\sqrt{n\tau_{n,\theta}}$ in the metric d_L defined as the supremum over Lipschitz functions

$$d_L(\mathcal{L}(X), \mathcal{L}(Y)) = \sup_h \frac{|Eh(X) - Eh(Y)|}{\|h\| + \|h'\|}. \tag{19}$$

As Lemma 2.1 shows that $\tau_{n,\theta}/\tau_\theta$ converges uniformly to 1 over Θ_n, the Kolmogorov bound of order $1/\sqrt{n\tau_\theta}$ provided by Theorem 2.1 is of the same order as the d_L bound. As remarked in Barbour, Karoński and Ruciński (1989), a bound of size ε_n in the d_L metric yields a bound in the Kolmogorov metric of order $O(\varepsilon_n^{1/2})$, which can at times be improved to $O(\varepsilon_n)$ "at the cost of much greater effort."

Though we do not cover the case $d = 0$ of isolated vertices, handled in Kordecki (1990), our proof can be extended to apply there by appending additional arguments that are separate, but similar to, those for the case $d \in \mathbb{N}$. Note, for example, the difference in the behavior of the function τ_θ at zero for these two ranges of d.

Following Lemma 1.1 for the case of vertex degrees yields a coupling where for each $n \ge d+1$ and vertex $v \in \mathcal{I}_n$ one constructs a graph G_n^v from G_n having the distribution of G_n conditioned on $X_{n,v} = 1$, or equivalently, on $D_n(v) = d$; this coupling has previously been applied by Barbour, Holst and Janson (1992) and Goldstein and Rinott (1996). The graph G_n^v is obtained from G_n by adding or removing edges of v as needed. Mixing over v as indicated by Lemma 1.1 yields a variable Y_n^s having the Y_n-size bias distribution.

In the course of constructing G_n^v one also obtains a set $\mathcal{R}_n^v$ holding the collection of vertices other than v that are affected by the size bias operation. In particular, if $D_n(v) = d$, then $G_n^v = G_n$ and $\mathcal{R}_n^v = \varnothing$. If $D_n(v) > d$, then G_n^v is formed by removing from G_n the edges between v and the vertices in the subset $\mathcal{R}_n^v$ of neighbors $\{u : \xi_{u,v} = 1\}$ of v, chosen with uniform conditional distribution given G_n over all subsets of the neighbors of v of size $D_n(v) - d$. Similarly, if $D_n(v) < d$, then G_n^v is formed by adding edges to G_n between v and vertices in $\mathcal{R}_n^v$, chosen with uniform conditional distribution given G_n over all subsets of the nonneighbors $\{u : u \neq v, \xi_{u,v} = 0\}$ of v of size $d - D_n(v)$.

Now let $X_{n,w}^v$ be the indicator that vertex w has degree d in G_n^v and

$$Y_n^v = \sum_{w \in \mathcal{I}_n} X_{n,w}^v,$$

the number of degree d vertices in G_n^v. When I_n is chosen uniformly over $\mathcal{I}_n$, independent of all other variables, Lemma 1.1 yields that $Y_n^s = Y_n^{I_n}$ has the Y_n-size biased distribution. Similarly setting $\mathcal{R}_n^s = \mathcal{R}_n^{I_n}$, all vertices not in $\{I_n\} \cup \mathcal{R}_n^s$ have the same degree in both G_n and G_n^s, and as $I_n \notin \mathcal{R}_n^s$, letting

$$\mathcal{A}_n = \{I_n\} \cup \mathcal{R}_n^s \qquad \text{we have } |\mathcal{A}_n| = 1 + |d - D_n(I_n)|. \tag{20}$$

We prove Theorem 2.1 by verifying the hypotheses of Theorem 1.1 for the size bias construction just given. With $\tau_{n,\theta}$ as in (17), and recalling (16), let

$$\delta_{n,\theta} = \tau_{n,\theta}\left[\frac{(d-\theta)^2}{\theta(1-\theta/(n-1))} - 1\right] + 1 \qquad \text{so that } \sigma_{n,\theta}^2 = n\tau_{n,\theta}\delta_{n,\theta}, \tag{21}$$

and correspondingly, with τ_θ as in (18), let

$$\delta_\theta = \tau_\theta\left[\frac{(d-\theta)^2}{\theta} - 1\right] + 1. \tag{22}$$

With the help of a technical lemma placed at the end of this section, we present the proof of Theorem 2.1. Throughout we let C_j denote a constant not depending on n or θ, and not necessarily the same at each occurrence.

PROOF OF THEOREM 2.1. Let $n_0 = d + 1$. For $n \geq n_0$ and $\theta \in \Theta_n$ the binomial and Poisson probabilities $\tau_{n,\theta}$ and τ_θ in (17) and (18), respectively, lie in $(0, 1)$, and hence $\sigma_{n,\theta}^2$ of (16) and $r_{n,\theta}$ are positive for all such n and θ. Let $r_1 > 0$ be arbitrary. In place of naming n_1 explicitly, we show the remaining conditions of Theorem 1.1 are satisfied for all n sufficiently large. Since $r_{n,\theta} \leq \sqrt{n}$ inequality (4) holds for any $n_1 \geq n_0$.

From Chen, Goldstein and Shao [(2010), equation (12.17)], following Goldstein and Rinott (1996), for Y_n^s having the Y_n-size biased distribution as constructed above, we obtain

$$\Psi_{n,\theta}^2 \leq C_1 n^{-1}(24\theta + 48\theta^2 + 144\theta^3 + 48d^2 + 144\theta d^2 + 12)$$

and hence

$$\sup_{\theta\in\Theta_n} \Psi_{n,\theta} \le \frac{C_2}{\sqrt{n}}.$$

To complete the verification of Condition 1, Lemma 2.1 gives that over Θ_n the ratio $\delta_\theta/\delta_{n,\theta} = \delta_\theta\mu_{n,\theta}/\sigma_{n,\theta}^2$ converges uniformly to 1, and δ_θ in (22) is bounded away from zero. Hence for all n sufficiently large and all $\theta\in\Theta_n$, we have

$$\frac{\mu_{n,\theta}}{\sigma_{n,\theta}^2} \le \frac{2}{\delta_\theta} \le C_3 \quad \text{and so} \quad \frac{r_{n,\theta}\mu_{n,\theta}\Psi_{n,\theta}}{\sigma_{n,\theta}^2} \le C_4\sqrt{\tau_\theta} \le C_4 \tag{23}$$

as $\tau_\theta \le 1$ for all $\theta\in\Theta_n$.

Turning to Condition 2, let

$$J_n = (I_n, D_n(I_n)) \quad \text{and} \quad \mathcal{F}_n = \sigma\{J_n\};$$

that is, $\mathcal{F}_n$ is the σ-algebra generated by the chosen vertex and its degree. Further, let

$$K_n = 1 + d + D_n(I_n).$$

Clearly K_n is $\mathcal{F}_n$-measurable, and recalling that vertices not in $\mathcal{A}_n$ of (20) have the same degree in both G_n and G_n^s, taking the difference between Y_n^s and Y_n yields

$$Y_n^s - Y_n = \sum_{w\in\mathcal{A}_n} (X_{n,w}^{I_n} - X_{n,w}),$$

and (20) yields

$$|Y_n^s - Y_n| = 1 + |d - D(I_n)| \le K_n.$$

Next, for all $m\in\mathbb{N}$ we have

$$K_n^m \le 2^{m-1}\big((1+d)^m + D_n(I_n)^m\big). \tag{24}$$

To bound the moments of K_n, using Riordan (1937) for the first equality below, with $S_{j,m}$ the Stirling numbers of the second kind, $(n)_j$ the falling factorial, $C_{5,m} = m\max_{1\le j\le m} S_{j,m}$ and $D\sim \mathrm{Bin}(n-1,p)$, we obtain

$$\begin{aligned} ED^m &= \sum_{j=1}^m S_{j,m}(n-1)_j p^j \le \sum_{j=1}^m S_{j,m}(n-1)^j p^j \\ &\le C_{5,m}\big((n-1)p + (n-1)^m p^m\big). \end{aligned}$$

In particular $E_\theta D_n(v)^m \le C_{5,m}(b+b^m)$, and as $D_n(I_n)$ is the mixture of the identical distributions $D_n(v)$ over $v\in\mathcal{I}_n$, it obeys the same upper bound. Taking expectation in (24), we find that there exists constants $C_{6,m}, m\in\mathbb{N}$ such that

$$k_{n,\theta,m} \le C_{6,m} \qquad \text{for all } n\in\mathbb{N}, \theta\in\Theta_n \text{ and } m\in\mathbb{N}. \tag{25}$$

Now, using (25) for the first inequality in (26), the first inequality in (23) for the second inequality, the second equality of (21) for the first equality, and Lemma 2.1 both to obtain the third inequality, and the boundedness of δ_θ away from zero for the fourth, we obtain that for all n sufficiently large and $\theta \in \Theta_n$,

$$(26) \qquad \frac{r_{n,\theta}\mu_{n,\theta}k_{n,\theta,4}^{1/2}}{\sigma_{n,\theta}^3} \le \frac{C_{6,4}{}^{1/2} r_{n,\theta}\mu_{n,\theta}}{\sigma_{n,\theta}^3} \le \frac{C_7 r_{n,\theta}}{\sigma_{n,\theta}} = \frac{C_7\sqrt{\tau_\theta}}{\sqrt{\tau_{n,\theta}\delta_{n,\theta}}} \le \frac{C_8}{\sqrt{\delta_\theta}} \le C_9.$$

Hence inequality (12), sufficient for (6), is satisfied, and Condition 2 holds.

Turning to Condition 3, for $n \ge d+2$, let

$$(27) \quad L_n = 1, \qquad s_{n,\theta} = 1, \qquad \psi_{n,\theta} = \left(\frac{n-2}{n-1}\right)\theta \quad \text{and} \quad F_{n,\theta} = \{L_n \le s_{n,\theta}\},$$

and note therefore that conditions holding on $F_{n,\theta}$ must hold on the entire probability space. Clearly L_n takes values in $\{0, 1, \dots, n\}$ as required and $n - s_{n,\theta} \ge n_0$ for any $n \ge d+2$. Being constants, L_n and $\psi_{n,\theta}$ are $\mathcal{F}_n$ measurable, hence $F_{n,\theta} \in \mathcal{F}_n$. By (27) and $\theta \in \Theta_n$ we have that $\psi_{n,\theta} \in (0, b] \cap (0, n-2) = \Theta_{n-1} = \Theta_{n-L_n}$, verifying the first part of (7).

Regarding the second part of (7), let H_n be the graph G_n with the vertex I_n and its incident edges removed, relabeling the remaining vertices $\{1, \dots, n-1\}$ by preserving their relative order. Let V_n be the number of degree d vertices of H_n. By counting the number of degree d vertices, the distributional equality in (7) is a consequence of

$$(28) \qquad \mathcal{L}_\theta(H_n | I_n, D_n(I_n)) = \mathcal{L}_{\psi_{n,\theta}}(G_{n-1}).$$

The graph H_n is determined by $\{\xi_{u,v} : \{u, v\} \subset \mathcal{I}_n \setminus \{I_n\}\}$, which is independent of the σ-algebra generated by $\{I_n, \xi_{I_n,v}, v \in \mathcal{I}_n\}$, with respect to which I_n and $D_n(I_n)$ are measurable. Hence H_n is independent of the conditioning event in (28), and therefore its conditional and unconditional distribution agree. In particular, conditional on $\{I_n, D_n(I_n)\}$, the edge indicators of H_n are independent with common success probability

$$\frac{\theta}{n-1} = \frac{\psi_{n,\theta}}{n-2},$$

so (28) holds. Inequality (8) holds trivially, as $P(L_n > 1) = 0$. Hence Condition 3 holds.

By Lemma 2.1, Condition 4 holds with $c_1 = c_2 = 2$.

Regarding Condition 5, as only the degrees of vertex I_n and its neighbors are different in the graphs G_n and H_n, we have

$$|Y_n - V_n| \le 1 + D(I_n) \le K_n,$$

and we set $B_n = K_n$, so $\mathcal{F}_n$-measurable. We now finish the verification of Condition 5 by showing (13), sufficient for (9), is satisfied. By (25), that $\mu_{n,\theta} = n\tau_{n,\theta}$

and the second equality in (21), for all n sufficiently large and all $\theta \in \Theta_n$, we have

$$\frac{r_{n,\theta}^2 \mu_{n,\theta} k_{n,\theta,4}^{1/2} b_{n,\theta,2}^{1/2}}{\sigma_{n,\theta}^4} \le \frac{\tau_\theta (C_{6,4} C_{6,2})^{1/2}}{\tau_{n,\theta} \delta_{n,\theta}^2} \le C_{10},$$

where the final inequality follows from Lemma 2.1, yielding that $\tau_{n,\theta}/\tau_\theta$ and $\delta_{n,\theta}/\delta_\theta$ converge uniformly to 1 on Θ_n, and that δ_θ is bounded away from zero on $(0, b]$.

Lastly, Condition 6a holds with $l_{n,0} = 1$ for all $n \ge d + 2$, completing the verification of all conditions of Theorem 1.1. □

The proof of Lemma 2.1 is straightforward, and is therefore omitted.

LEMMA 2.1. *With $\tau_{n,\theta}$, τ_θ, $\delta_{n,\theta}$ and δ_θ given by* (17), (18), (21) *and* (22), *respectively, for all $d \in \mathbb{N}$ and all $b > 0$ the function δ_θ is bounded away from zero and infinity over* $(0, b]$, *and the ratios*

$$\frac{\tau_{n,\theta}}{\tau_\theta}, \qquad \frac{\delta_{n,\theta}}{\delta_\theta}, \qquad \frac{r_{n,\theta}}{r_{n-1,\psi_{n,\theta}}} \quad \text{and} \quad \frac{\sigma_{n,\theta}^2}{\sigma_{n-1,\psi_{n,\theta}}^2}$$

and their reciprocals converge uniformly to 1 *on* $(0, b]$ *as n tends to infinity.*

3. Proof of Theorem 1.1.

We begin the proof of Theorem 1.1 with the following lemma.

LEMMA 3.1. *Suppose that for some $n_1 \in \mathbb{N}_0$ the nonnegative numbers f, $\{p_{n,l}\}_{n \ge n_1, 0 \le l \le n}$ and $\{a_n\}_{n \ge 0}$ satisfy*

$$a_n \le \sum_{l=0}^{n} a_{n-l} p_{n,l} + f \qquad \text{for all } n \ge n_1 \quad \text{and}$$

$$\tau \in (0,1) \qquad \text{where } \tau = \sup_{n \ge n_1} \sum_{l=0}^{n} p_{n,l}. \tag{29}$$

Then $\sup_{n \ge 0} a_n < \infty$.

PROOF. As for all $n \ge n_1$ we have $p_{n,0} \le \tau < 1$, letting

$$q_{n,l} = \frac{p_{n,l}}{1 - p_{n,0}} \qquad \text{for } 1 \le l \le n \quad \text{and} \quad a = \frac{f}{1-\tau},$$

(29) implies

$$a_n \le \sum_{l=1}^{n} a_{n-l} q_{n,l} + a \qquad \text{with } 0 \le \sum_{l=1}^{n} q_{n,l} \le \frac{\tau - p_{n,0}}{1 - p_{n,0}} \le \tau \text{ for all } n \ge n_1.$$

Letting $\alpha = \max_{0 \le n \le n_1} a_n$ and $c = \max\{a, \alpha(1-\tau)\}$, the sequence $\{b_n\}_{n \ge 0}$ defined by

$$b_n = \alpha \qquad \text{for } 0 \le n \le n_1 \quad \text{and} \quad b_{n+1} = \tau b_n + c \qquad \text{for } n \ge n_1$$

has, for $n \ge n_1$, the explicit form

$$b_n = \gamma \tau^{n-n_1} + \frac{c}{1-\tau} \qquad \text{where } \gamma = \alpha - \frac{c}{1-\tau}.$$

Since $\gamma \le 0$ and $\tau \in (0,1)$, the sequence $\{b_n\}_{n \ge 0}$ is nondecreasing with limit $c/(1-\tau)$, and hence is bounded. We complete the proof by showing that for all $n \in \mathbb{N}_0$ we have $a_m \le b_m$ for all $0 \le m \le n$. Clearly the statement holds for $0 \le n \le n_1$. Assuming it true for some $n \ge n_1$, using the induction hypotheses, the definition of c and that b_n is nondecreasing,

$$\begin{aligned} a_{n+1} &\le \sum_{l=1}^{n+1} a_{n+1-l} q_{n+1,l} + a \le \sum_{l=1}^{n+1} b_{n+1-l} q_{n+1,l} + c \le b_n \sum_{l=1}^{n+1} q_{n+1,l} + c \\ &\le \tau b_n + c = b_{n+1}. \end{aligned}$$ □

The following proof is based on the inductive argument of Bolthausen (1984).

PROOF OF THEOREM 1.1. With $r \ge 0$, recall that $\Theta_{n,r} = \{\theta \in \Theta_n : r_{n,\theta} \ge r\}$, and let

$$\delta(n,r) = \sup_{z \in \mathbb{R}, \theta \in \Theta_{n,r}} |P_\theta(W_{n,\theta} \le z) - P(Z \le z)| \qquad \text{for } n \ge n_0. \tag{30}$$

First note that (11) of Theorem 1.1 can be made to hold whenever $r_{n,\theta} < r_1$ by taking $C \ge r_1$. By (4) the cases $n_0 \le n < n_1$ and $r_{n,\theta} \ge r_1$ can be handled in this same manner. Hence it suffices to show that there exists some C such that

$$\delta(n,r) \le C/r \qquad \text{for } n \ge n_1 \text{ and } r \ge r_1. \tag{31}$$

For $z \in \mathbb{R}$ and $\lambda > 0$ let $h_{z,\lambda}$ be the smoothed indicator

$$h_{z,\lambda}(x) = \begin{cases} 1, & x \le z, \\ 1 + (z-x)/\lambda, & z < x \le z + \lambda, \\ 0, & x > z + \lambda \end{cases}$$

and let $Nh_{z,\lambda} = Eh_{z,\lambda}(Z)$ with Z a standard normal variable. Let $f(x)$ be the unique bounded solution to the Stein equation for $h_{z,\lambda}(x)$ [see, e.g., Chen, Goldstein and Shao (2010)]

$$h_{z,\lambda}(x) - Nh_{z,\lambda} = f'(x) - xf(x). \tag{32}$$

Let $n \ge n_1$, $\theta \in \Theta_{n,r}$ for some $r \ge r_1$, $z \in \mathbb{R}$ and $\lambda > 0$. Recalling $W_{n,\theta} = (Y_n - \mu_{n,\theta})/\sigma_{n,\theta}$, with a slight abuse of notation, set

$$W_{n,\theta}^s = \frac{Y_n^s - \mu_{n,\theta}}{\sigma_{n,\theta}}.$$

Substituting $W_{n,\theta}$ for x in (32) and taking expectation, and dropping the subscript θ when not essential below, we obtain

$$E_\theta h_{z,\lambda}(W_n) - Nh_{z,\lambda} = E_\theta[f'(W_n) - W_n f(W_n)]. \tag{33}$$

Beginning with the second term on the right-hand side of (33), from the definition of $W_{n,\theta}$ and the size bias relation (1), we have

$$E_\theta[W_n f(W_n)] = \frac{1}{\sigma_n} E_\theta[(Y_n - \mu_n) f(W_n)] = \frac{\mu_n}{\sigma_n} E_\theta\bigl(f(W_n^s) - f(W_n)\bigr).$$

Taking absolute value and applying the triangle inequality, we obtain

$$\begin{aligned}
&|E_\theta h_{z,\lambda}(W_n) - Nh_{z,\lambda}| \\
&\quad = |E_\theta[f'(W_n) - W_n f(W_n)]| \\
&\quad = \left| E_\theta\left[f'(W_n) - \frac{\mu_n}{\sigma_n}\bigl(f(W_n^s) - f(W_n)\bigr)\right]\right| \\
&\quad = \frac{\mu_n}{\sigma_n}\left| E_\theta\left[\frac{\sigma_n}{\mu_n} f'(W_n) - \bigl(f(W_n^s) - f(W_n)\bigr)\right]\right| \\
&\quad = \frac{\mu_n}{\sigma_n}\left| E_\theta\left[\left(\frac{\sigma_n}{\mu_n} - (W_n^s - W_n)\right) f'(W_n) + (W_n^s - W_n) f'(W_n) \right.\right. \\
&\qquad\qquad\qquad\qquad \left.\left. - \bigl(f(W_n^s) - f(W_n)\bigr)\right]\right| \\
&\quad \le \frac{\mu_n}{\sigma_n}\left| E_\theta\left[\left(\frac{\sigma_n}{\mu_n} - (W_n^s - W_n)\right) f'(W_n)\right]\right| \\
&\qquad + \frac{\mu_n}{\sigma_n}\left| E_\theta\left[\int_0^{W_n^s - W_n} [f'(W_n) - f'(W_n + t)]\,dt\right]\right|.
\end{aligned} \tag{34}$$

From the size bias relation (1) with $f(x) = x$, we obtain $\mu_n E_\theta[Y_n^s] = E_\theta[Y_n^2]$, and therefore

$$E_\theta[W_n^s - W_n] = E_\theta\left[\frac{Y_n^s - Y_n}{\sigma_n}\right] = \frac{1}{\sigma_n}\left[\frac{E_\theta Y_n^2}{\mu_n} - \mu_n\right] = \frac{1}{\sigma_n \mu_n}\sigma_n^2 = \frac{\sigma_n}{\mu_n}. \tag{35}$$

Now applying (35) and $|f'(x)| \le 1$ from Chen and Shao [(2004), equation (4.6)] [see also Chen, Goldstein and Shao (2010), Lemma 2.5], by conditioning on W_n the first term of (34) may be bounded by

$$\begin{aligned}
&\frac{\mu_n}{\sigma_n}\left| E_\theta\left[E_\theta\left(\frac{\sigma_n}{\mu_n} - (W_n^s - W_n) \Big| W_n\right) f'(W_n)\right]\right| \\
&\qquad \le \frac{\mu_n}{\sigma_n}\sqrt{\operatorname{Var} E_\theta(W_n^s - W_n | W_n)} = \frac{\mu_n}{\sigma_n^2}\Psi_n,
\end{aligned} \tag{36}$$

recalling the definition of Ψ_n in (5).

Moving now to the second term of (34), Bolthausen [(1984), equation (2.4)] gives

$$|f(x)| \le 1 \quad \text{and} \quad |xf(x)| \le 1,$$

and combining these inequalities with $|f'(x)| \le 1$ and (32) as in Bolthausen [(1984), equation (2.5)] yields

$$|f'(x) - f'(x+t)| \le |t|\biggl(1 + |x| + \frac{1}{\lambda}\int_0^1 1_{[z,z+\lambda]}(x+ut)\,du\biggr).$$

Hence, applying the bound $|Y_n^s - Y_n| \le K_n$, the second term in (34) may be bounded by

$$\frac{\mu_n}{\sigma_n} E_\theta \int_{-K_n/\sigma_n}^{K_n/\sigma_n} |t|\biggl(1 + |W_n| + \frac{1}{\lambda}\int_0^1 1_{[z,z+\lambda]}(W_n+ut)\,du\biggr)\,dt, \tag{37}$$

yielding three terms.

For the first two terms in (37) we obtain

$$\frac{2\mu_n}{\sigma_n} E_\theta\biggl((1+|W_n|)\int_0^{K_n/\sigma_n} t\,dt\biggr) = \frac{\mu_n}{\sigma_n^3} E_\theta[(1+|W_n|)K_n^2]. \tag{38}$$

Next, as $|t| \le K_n/\sigma_n$ in the region of integration, we may bound the expectation of the remaining term in (37) by

$$\frac{\mu_n}{\lambda\sigma_n^2} E_\theta\biggl(K_n \int_{-K_n/\sigma_n}^{K_n/\sigma_n}\int_0^1 \mathbf{1}_{[z,z+\lambda]}(W_n+ut)\,du\,dt\biggr). \tag{39}$$

Clearly,

$$\mathbf{1}_{[z,z+\lambda]}(W_n+ut) \le (1-\mathbf{1}_{F_{n,\theta}}) + \mathbf{1}_{[z,z+\lambda]}(W_n+ut)\mathbf{1}_{F_{n,\theta}}. \tag{40}$$

Substituting (40) into (39), the first term in (40) gives rise to the expression

$$\frac{\mu_n}{\lambda\sigma_n^2} E_\theta\biggl(K_n \int_{-K_n/\sigma_n}^{K_n/\sigma_n}\int_0^1 (1-\mathbf{1}_{F_{n,\theta}})\,du\,dt\biggr) = \frac{2\mu_n}{\lambda\sigma_n^3} E_\theta[K_n^2(1-\mathbf{1}_{F_{n,\theta}})]. \tag{41}$$

Substituting the second term in (40) into (39), conditioning on $\mathcal{F}_n$ and invoking the $\mathcal{F}_n$ measurability of K_n and $F_{n,\theta}$ provided by Conditions 2 and 3, respectively, yields

$$\begin{aligned} &\frac{\mu_n}{\lambda\sigma_n^2} E_\theta\biggl(K_n \int_{-K_n/\sigma_n}^{K_n/\sigma_n}\int_0^1 \mathbf{1}(z \le W_n+ut \le z+\lambda)\mathbf{1}_{F_{n,\theta}}\,du\,dt\biggr) \\ &\quad = \frac{\mu_n}{\lambda\sigma_n^2} E_\theta\biggl(K_n \int_{-K_n/\sigma_n}^{K_n/\sigma_n}\int_0^1 P_\theta^{\mathcal{F}_n}(z \le W_n+ut \le z+\lambda)\mathbf{1}_{F_{n,\theta}}\,du\,dt\biggr), \end{aligned} \tag{42}$$

where $P_\theta^{\mathcal{F}_n}$ denotes conditional probability with respect to $\mathcal{F}_n$. To handle the indicator in (42), note that Condition 3 implies that $n - L_n \ge n_0$ on $F_{n,\theta}$. Hence on $F_{n,\theta}$ we may define

$$\underline{W_{n,\theta}} = \frac{V_n - \mu_{n-L_n,\psi_{n,\theta}}}{\sigma_{n-L_n,\psi_{n,\theta}}}$$

and write

$$W_n = \left(\frac{\sigma_{n-L_n,\psi_n}}{\sigma_n}\right)\underline{W_n} + \left(\frac{Y_n - V_n}{\sigma_n}\right) - \left(\frac{\mu_n - \mu_{n-L_n,\psi_n}}{\sigma_n}\right) \tag{43}$$
$$:= \rho_n \underline{W_n} + T_{n,1} - T_{n,2}.$$

By Conditions 5 and 3 we have $|T_{n,1}| \le B_n/\sigma_n$ and that ρ_n, B_n and $T_{n,2}$ are $\mathcal{F}_n$-measurable. Using (43) we may write

$$\begin{aligned}
&P_\theta^{\mathcal{F}_n}(z \le W_n + ut \le z + \lambda)\mathbf{1}_{F_{n,\theta}} \\
&\quad = P_\theta^{\mathcal{F}_n}\big(\rho_n^{-1}(z - T_{n,1} + T_{n,2} - ut) \le \underline{W_n} \\
&\qquad\qquad \le \rho_n^{-1}(z - T_{n,1} + T_{n,2} - ut + \lambda)\big)\mathbf{1}_{F_{n,\theta}} \\
&\quad \le P_\theta^{\mathcal{F}_n}\big(\rho_n^{-1}(z + T_{n,2} - ut) - B_n/\sigma_{n-L_n,\psi_n} \le \underline{W_n} \\
&\qquad\qquad \le \rho_n^{-1}(z + T_{n,2} - ut) + B_n/\sigma_{n-L_n,\psi_n} + \rho_n^{-1}\lambda\big)\mathbf{1}_{F_{n,\theta}} \\
&\quad = P_\theta^{\mathcal{F}_n}(Q_n - B_n/\sigma_{n-L_n,\psi_n} \le \underline{W_n} \le Q_n + B_n/\sigma_{n-L_n,\psi_n} + \rho_n^{-1}\lambda)\mathbf{1}_{F_{n,\theta}},
\end{aligned} \tag{44}$$

where we have set

$$Q_n = \rho_n^{-1}(z + T_{n,2} - ut).$$

Recalling (30), we have

$$\begin{aligned}
&P_\theta(z \le W_{n,\theta} \le z + \lambda) \\
&\quad \le |P_\theta(z \le W_{n,\theta} \le z + \lambda) - P(z \le Z \le z + \lambda)| + P(z \le Z \le z + \lambda) \\
&\quad \le 2\delta(n, r_{n,\theta}) + \lambda/\sqrt{2\pi}.
\end{aligned} \tag{45}$$

Since the endpoints of the interval bounding $\underline{W_n}$ in (44) are $\mathcal{F}_n$-measurable, using Condition 3 and (45) with the appropriate substitutions, expression (44) is bounded by

$$\begin{aligned}
&\big(2\delta(n - L_n, r_{n-L_n,\psi_{n,\theta}}) + (2B_n/\sigma_{n-L_n,\psi_{n,\theta}} + \rho_n^{-1}\lambda)/\sqrt{2\pi}\big)\mathbf{1}_{F_{n,\theta}} \\
&\quad \le \big(2\delta(n - L_n, r_{n,\theta}/c_2) + \big(2\sqrt{c_1}B_n/\sigma_{n,\theta} + \sqrt{c_1}\lambda\big)/\sqrt{2\pi}\big)\mathbf{1}_{F_{n,\theta}},
\end{aligned}$$

where we have applied Condition 4, and that $\delta(n, r)$ is nonincreasing in r. As this last quantity does not depend on u or t, substitution into (42) yields the bound

$$\frac{2\mu_{n,\theta}}{\lambda\sigma_{n,\theta}^3} E_\theta\big[K_n^2\big(2\delta(n - L_n, r_{n,\theta}/c_2) + \big(2\sqrt{c_1}B_n/\sigma_{n,\theta} + \sqrt{c_1}\lambda\big)/\sqrt{2\pi}\big)\big]\mathbf{1}_{F_{n,\theta}}. \tag{46}$$

Expression (46) leads to three terms. By (6), that $F_{n,\theta} \subset \{L_n \le s_{n,\theta}\}$, and since $n - s_{n,\theta} \ge n_0$ for all $\theta \in \Theta_{n,r_1}$ implies $s_n = \sup_{\theta\in\Theta_{n,r_1}} s_{n,\theta} \le n - n_0$, there exists a positive constant C_1 such that the first term satisfies

$$
\begin{aligned}
&\frac{4\mu_{n,\theta}}{\lambda\sigma_{n,\theta}^3} E_\theta[K_n^2 \delta(n-L_n, r_{n,\theta}/c_2)\mathbf{1}_{F_{n,\theta}}] \\
&\quad = \frac{4\mu_{n,\theta}k_{n,\theta,2}}{\lambda\sigma_{n,\theta}^3} E_\theta\biggl[\frac{K_n^2}{E_\theta K_n^2}\delta(n-L_n, r_{n,\theta}/c_2)\mathbf{1}_{F_{n,\theta}}\biggr] \\
&\quad \le \frac{C_1}{\lambda r_{n,\theta}} E_\theta\biggl[\frac{K_n^2}{E_\theta K_n^2}\delta(n-L_n, r_{n,\theta}/c_2)\mathbf{1}_{F_{n,\theta}}\biggr] \\
&\quad \le \frac{C_1}{\lambda r_{n,\theta}} \sum_{l=0}^{s_n} \delta(n-l, r_{n,\theta}/c_2) t_{n,\theta,l},
\end{aligned}
\tag{47}
$$

where $t_{n,\theta,l}$, given in (10), satisfy

$$
\sum_{l=0}^{n} t_{n,\theta,l} = 1 \qquad \text{for all } \theta \in \Theta_{n,r}. \tag{48}
$$

Dropping the indicator $\mathbf{1}_{F_{n,\theta}}$, the sum of the second and third terms of (46) are bounded by

$$
\frac{4\sqrt{c_1}\mu_n E_\theta[K_n^2 B_n]}{\sqrt{2\pi}\lambda\sigma_n^4} + \frac{\sqrt{2c_1}\mu_n}{\sqrt{\pi}\sigma_n^3} E_\theta K_n^2. \tag{49}
$$

Collecting terms (36), (38), (41), (48) and (49), and letting

$$
c_{n,\theta,1} = \frac{\mu_{n,\theta}}{\sigma_{n,\theta}^2}\Psi_{n,\theta} + \frac{\mu_{n,\theta}}{\sigma_{n,\theta}^3} E_\theta\biggl[\biggl(\biggl(1 + \frac{\sqrt{2c_1}}{\sqrt{\pi}}\biggr) + |W_{n,\theta}|\biggr)K_n^2\biggr] \quad \text{and}
$$

$$
c_{n,\theta,2} = \frac{2\mu_{n,\theta}}{\sigma_{n,\theta}^3} E_\theta[K_n^2(1-\mathbf{1}_{F_{n,\theta}})] + \frac{4\sqrt{c_1}\mu_{n,\theta}E_\theta[K_n^2 B_n]}{\sqrt{2\pi}\sigma_{n,\theta}^4},
$$

for all $z \in \mathbb{R}$ we have

$$
\begin{aligned}
&|E_\theta h_{z,\lambda}(W_{n,\theta}) - Nh_{z,\lambda}| \\
&\quad \le \frac{C_1}{\lambda r_{n,\theta}} \sum_{l=0}^{s_n} \delta(n-l, r_{n,\theta}/c_2) t_{n,\theta,l} + c_{n,\theta,1} + \frac{1}{\lambda} c_{n,\theta,2}.
\end{aligned}
\tag{50}
$$

Note that Conditions 1 and 2, and 3 and 5, respectively, yield the existence of positive constants C_2 and C_3 that

$$
c_{n,\theta,1} \le C_2/r_{n,\theta} \quad \text{and} \quad c_{n,\theta,2} \le C_3/r_{n,\theta}^2. \tag{51}
$$

As $\mathbf{1}(w \le z) \le h_{z,\lambda}(w) \le \mathbf{1}(w \le z+\lambda)$ we obtain

$$
\begin{aligned}
&P_\theta(W_{n,\theta} \le z) - P(Z \le z) \\
&\quad \le |E_\theta h_{z,\lambda}(W_{n,\theta}) - Eh_{z,\lambda}(Z)| + Eh_{z,\lambda}(Z) - P(Z \le z)
\end{aligned}
$$

with $Eh_{z,\lambda}(Z) - P(Z \le z) \le P(z \le Z \le z+\lambda) \le \lambda/\sqrt{2\pi}$. Along with a similar lower bound obtained by considering $h_{z-\lambda,\lambda}(w)$, in view of (50) and (51) we have that for every $z \in \mathbb{R}$

$$\begin{aligned}&|P_\theta(W_{n,\theta} \le z) - P(Z \le z)|\\&\quad\le \frac{C_1}{\lambda r_{n,\theta}} \sum_{l=0}^{s_n} \delta(n-l, r_{n,\theta}/c_2) t_{n,\theta,l} + \frac{C_2}{r_{n,\theta}} + \frac{C_3}{\lambda r_{n,\theta}^2} + \frac{\lambda}{\sqrt{2\pi}}.\end{aligned}$$

Letting $\lambda = 2c_2C_1/r_{n,\theta}$, and, noting that the right-hand side does not depend on z, taking supremum over $z \in \mathbb{R}$ yields

$$\begin{aligned}&\sup_{z\in\mathbb{R}} |P_\theta(W_{n,\theta} \le z) - P(Z \le z)|\\&\quad\le \sum_{l=0}^{s_n} \delta(n-l, r_{n,\theta}/c_2) t_{n,\theta,l}/2c_2 + C_4/r_{n,\theta} \qquad (52)\\&\quad\le \sum_{l=0}^{s_n} \delta(n-l, r/c_2) t_{n,\theta,l}/2c_2 + C_4/r\end{aligned}$$

for $C_4 = C_2 + C_3/2c_2C_1 + 2c_2C_1/\sqrt{2\pi}$, where for the last inequality we have used that $\theta \in \Theta_{n,r}$, and that $\delta(n,r)$ and $1/r$ are nonincreasing functions of r. Taking supremum over Θ_{n,r_1} on the right-hand side of (52), then over $\Theta_{n,r} \subset \Theta_{n,r_1}$ on the left yields

$$\delta(n,r) \le \sup_{\theta\in\Theta_{n,r_1}} \sum_{l=0}^{s_n} \delta(n-l, r/c_2) t_{n,\theta,l}/2c_2 + C_4/r. \qquad (53)$$

Suppose first that Condition 6a is satisfied, so that $L_n = l_{n,0}$ almost surely for some $l_{0,n} \in \{0,\dots,n\}$ for all $\theta \in \Theta_{n,r_1}$. If $l_{0,n} > s_n$ then (10) and (53) yield $\delta(n,r) \le C_4/r$, proving (31). Otherwise $t_{n,\theta,l} = \mathbf{1}(l = l_{n,0})$ for $0 \le l_{n,0} \le s_n$, and inequality (53) specializes to

$$\delta(n,r) \le \delta(n - l_{n,0}, r/c_2)/2c_2 + C_4/r. \qquad (54)$$

When Condition 6b is satisfied, the sum in (53) is a continuous function of θ on the compact set Θ_{n,r_1}, and hence achieves its supremum at some $\theta_n^* \in \Theta_{n,r_1}$. Letting $p_{n,l} = t_{n,\theta_n^*,l}/2$, from (53) and (48) we have

$$\delta(n,r) \le \sum_{l=0}^{s_n} \delta(n-l, r/c_2) p_{n,l}/c_2 + C_4/r \qquad \text{with } \sum_{l=0}^{n} p_{n,l} = 1/2. \qquad (55)$$

As (54) is the special case of (55) when $p_{n,l} = \mathbf{1}(l = l_{n,0})/2$, it suffices to handle the latter.

Let $a_n = 0$ for $0 \le n < n_0$, and $a_n = \sup_{r \ge r_1} r\delta(n,r)$ for $n \ge n_0$. For all $r \ge r_1$ and $n \ge n_0$ we have

$$\begin{aligned}(r/c_2)\delta(n, r/c_2) &\le \sup_{s\,:\,s\ge r_1} (s/c_2)\delta(n, s/c_2)\\ &= \sup_{s\,:\,s\ge r_1/c_2} s\delta(n,s)\\ &\le \Big[\sup_{s\,:\,r_1/c_2\le s<r_1} s\delta(n,s)\Big]\mathbf{1}(c_2>1) + \sup_{s\,:\,s\ge r_1} s\delta(n,s)\\ &\le r_1 + a_n.\end{aligned}$$

Using that $n \ge n_1$ implies $n - s_n \ge n_0$, multiplication by r in (55) yields, with $f = r_1/2 + C_4$, that for all $n \ge n_1$

$$\begin{aligned}r\delta(n,r) &\le \sum_{l=0}^{s_n} (r/c_2)\delta(n-l, r/c_2)p_{n,l} + C_4 \le \sum_{l=0}^{s_n}(r_1 + a_{n-l})p_{n,l} + C_4\\ &\le \sum_{l=0}^{n} a_{n-l}p_{n,l} + f.\end{aligned}$$

Taking supremum on the left-hand side over $r \ge r_1$ and recalling (55) now yields

$$a_n \le \sum_{l=0}^{n} a_{n-l}p_{n,l} + f \qquad \text{with } \sum_{l=0}^{n} p_{n,l} = 1/2 \text{ for all } n \ge n_1.$$

Lemma 3.1 now implies $\sup_{n\ge n_1} a_n < \infty$. Hence, there exists a constant C such that $\delta(n,r) \le C/r$ for all $n \ge n_1$ and all $r \ge r_1$; that is, (31) holds. □

REFERENCES

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* (T. W. Anderson, K. B. Athreya and D. L. Iglehart, eds.) 59–81. Academic Press, Boston, MA. MR1031278

BARBOUR, A. D., HOLST, L. and JANSON, S. (1992). *Poisson Approximation. Oxford Studies in Probability* **2**. The Clarendon Press Oxford Univ. Press, New York. MR1163825

BARBOUR, A. D., KAROŃSKI, M. and RUCIŃSKI, A. (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B* **47** 125–145. MR1047781

BOLLOBÁS, B. (1985). *Random Graphs*. Academic Press, London. MR0809996

BOLTHAUSEN, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete* **66** 379–386. MR0751577

CHEN, L. H. Y., GOLDSTEIN, L. and SHAO, Q. M. (2010). *Normal Approximation by Stein's Method*. Springer, Berlin.

CHEN, L. H. Y. and RÖLLIN, A. (2010). Stein couplings for normal approximation. Preprint.

CHEN, L. H. Y. and SHAO, Q.-M. (2004). Normal approximation under local dependence. *Ann. Probab.* **32** 1985–2028. MR2073183

ERDŐS, P. and RÉNYI, A. (1959). On random graphs. I. *Publ. Math. Debrecen* **6** 290–297. MR0120167

GOLDSTEIN, L. (2005). Berry–Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. *J. Appl. Probab.* **42** 661–683. MR2157512

GOLDSTEIN, L. and PENROSE, M. D. (2010). Normal approximation for coverage models over binomial point processes. *Ann. Appl. Probab.* **20** 696–721. MR2650046

GOLDSTEIN, L. and RINOTT, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33** 1–17. MR1371949

GOLDSTEIN, L. and ZHANG, H. (2011). A Berry–Esseen bound for the lightbulb process. *Adv. in Appl. Probab.* **43** 875–898. MR2858224

JANSON, S. and NOWICKI, K. (1991). The asymptotic distributions of generalized U-statistics with applications to random graphs. *Probab. Theory Related Fields* **90** 341–375. MR1133371

KAROŃSKI, M. and RUCIŃSKI, A. (1987). Poisson convergence and semi-induced properties of random graphs. *Math. Proc. Cambridge Philos. Soc.* **101** 291–300. MR0870602

KORDECKI, W. (1990). Normal approximation and isolated vertices in random graphs. In *Random Graphs '87* (*Poznań*, 1987) 131–139. Wiley, Chichester. MR1094128

NEAMMANEE, K. and SUNTADKARN, A. (2009). On the normal approximation of the number of vertices in a random graph. *ScienceAsia* **25** 203–210.

PALKA, Z. (1984). On the number of vertices of given degree in a random graph. *J. Graph Theory* **8** 167–170. MR0732030

RIORDAN, J. (1937). Moment recurrence relations for binomial, Poisson and hypergeometric frequency distributions. *Ann. Math. Statist.* **8** 103–111.

STEIN, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California*, *Berkeley*, *Calif.*, 1970/1971), *Vol. II*: *Probability Theory* 583–602. Univ. California Press, Berkeley, CA. MR0402873

STEIN, C. (1986). *Approximate Computation of Expectations*. IMS, Hayward, CA. MR0882007

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
KAPRIELIAN HALL, ROOM 108
3620 VERMONT AVENUE
LOS ANGELES, CALIFORNIA 90089-2532
USA
E-MAIL: larry@math.usc.edu