Math. Proc. Camb. Phil. Soc. (1982), **92**, 349–359
Printed in Great Britain

Poisson convergence and random graphs

BY A. D. BARBOUR
Gonville and Caius College, Cambridge

(*Received* 16 *August* 1981; *Revised* 29 *March* 1982)

1. *Introduction*

Approximation by the Poisson distribution arises naturally in the theory of random graphs, as in many other fields, when counting the number of occurrences of individually rare and unrelated events within a large ensemble. For example, one may be concerned with the number of times that a particular small configuration is repeated in a large graph, such questions being considered, amongst others, in the fundamental paper of Erdős and Rényi(4). The technique normally used to obtain such approximations in random graph theory is based on showing that the factorial moments of the quantity concerned converge to those of a Poisson distribution as the size of the graph tends to infinity. Since the rth factorial moment is just the expected number of ordered r-tuples of events occurring, it is particularly well suited to evaluation by combinatorial methods. Unfortunately, such a technique becomes very difficult to manage if the mean of the approximating Poisson distribution is itself increasing with the size of the graph, and this limits the scope of the results obtainable.

The purpose of this paper is to show how an alternative approach, based on that of Stein(6) for the Normal approximation, and developed in the Poisson context by Chen(3) and Barbour and Eagleson(1), can be adapted to circumvent this particular difficulty. The new approach is doubly attractive in that, at the same time, it yields an upper bound on the total variation distance between the approximating and the true distributions. The technique is described in Section 2, in the context of some classical problems raised by Erdős and Rényi. In Section 3, Stein's method is used to establish convergence in distribution to the Normal, together with the correct rate of convergence, in a case where a Poisson approximation would not be appropriate.

2. *Poisson convergence*

The method rests on the following simple observation: that, for each $\lambda > 0$ and each $A \subset \mathbb{Z}^+$, the function $x = x_{\lambda, A}: \mathbb{Z}^+ \to R$ defined by

$$x(0) = 0; \quad x(m+1) = \lambda^{-m-1} e^{\lambda} m! [\mathscr{P}_\lambda(A \cap C_m) - \mathscr{P}_\lambda(A)\mathscr{P}_\lambda(C_m)], \quad m \geqslant 0, \quad (2.1)$$

where $C_m = \{0, 1, \ldots, m\}$, has the properties

(i) $\|x\| \equiv \sup_{m \in \mathbb{Z}^+} |x(m)| \leqslant 1 \wedge 1{\cdot}4\lambda^{-\frac{1}{2}}$,

(ii) $\Delta x \equiv \sup_{m \in \mathbb{Z}^+} |x(m+1) - x(m)| \leqslant 2(1 \wedge \lambda^{-1})$,

(iii) for any non-negative-integer-valued random variable R,

$$E\{\lambda x(R+1) - Rx(R)\} = P[R \in A] - \mathscr{P}_\lambda(A),$$

https://doi.org/10.1017/S0305004100059995 Published online by Cambridge University Press

where $\mathscr{P}_\lambda(.)$ denotes the probability that a Poisson random variable with mean λ falls in A.

Properties (i) and (ii) are proved in Barbour and Eagleson (1), and property (iii) is straightforward. The advantage of such a representation for the departure from the Poisson of the distribution of R is not immediately apparent, but the following examples show its use when R is composed of a sum of individually rather unimportant components.

As a first example, consider the problem of isolated trees in $G(n,p)$, the random graph on n vertices in which each edge is present, independently of all others, with probability p. For each k-tuple $\mathbf{i} \in D_n \equiv \{(i_1, \ldots, i_k), 1 \leqslant i_1 < \ldots < i_k \leqslant n\}$, let $X_{\mathbf{i}} = 1$ if there is an isolated tree of order k spanning the vertices $i_1, \ldots, i_k$, and set $X_{\mathbf{i}} = 0$ otherwise: write $W_n = \Sigma_{\mathbf{i} \in D_n} X_{\mathbf{i}}$. Then, for any $x: \mathbb{Z}^+ \to R$,

$$E\{X_{\mathbf{i}} x(W_n)\} = P[X_{\mathbf{i}} = 1]\, E\{x(W_{n-k}+1)\}, \tag{2·2}$$

and so, writing

$$\lambda = \sum_{\mathbf{i} \in D_n} P[X_{\mathbf{i}} = 1] = \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)+\binom{k}{2}-k+1}, \tag{2·3}$$

it follows that

$$E\{\lambda x(W_n+1) - W_n x(W_n)\} = \lambda E\{x(W_n+1) - x(W_{n-k}+1)\}. \tag{2·4}$$

To exploit this expression, couple W_n and W_{n-k} in the following way. Construct $G(n,p)$ by starting with a $G(n-k,p)$ and then adding the remaining vertices and randomly selected edges: let W_{n-k} denote the number of isolated trees of order k in the initial realization of $G(n-k,p)$, and W_n that in the resulting $G(n,p)$. Then, with W_{n-k} and W_n coupled in this way, the obvious inequality

$$|\lambda E\{x(W_n+1) - x(W_{n-k}+1)\}| \leqslant \lambda \Delta x E|W_n - W_{n-k}|$$

allows one to make a useful estimate of the right-hand side of (2·4). To see this, note first that

$$(W_n - W_{n-k})^+ \leqslant \sum_{j=n-k+1}^{n} Y_j, \tag{2·5}$$

where $Y_j = 1$ if j belongs to an isolated tree of order k in $G(n,p)$, and $Y_j = 0$ otherwise: hence

$$\begin{aligned} E\{(W_n - W_{n-k})^+\} &\leqslant k \binom{n-1}{k-1} k^{k-2} p^{k-1} (1-p)^{\binom{k}{2}-k+1+k(n-k)} \\ &\leqslant \{k^{k-1}(npe^{-np})^{k-1} e^{-np}/(k-1)!\}\,(1-p)^{-k^2} \equiv c_1(n,k,p). \end{aligned} \tag{2·6}$$

On the other hand, an upper bound for $(W_{n-k} - W_n)^+$ is the number of isolated trees of order k in $G(n-k,p)$ which then become connected to vertices among $\{n-k+1, \ldots, n\}$, so that

$$\begin{aligned} E\{(W_{n-k} - W_n)^+\} &\leqslant \{1-(1-p)^{k^2}\}\, EW_{n-k} \\ &\leqslant k^2 p \{k^{k-2}(p(n-k)\, e^{-(n-k)p})^k p^{-1}/k!\}\,(1-p)^{-k(k+3)/2} \\ &\equiv c_2(n,k,p). \end{aligned} \tag{2·7}$$

Hence, taking $x = x_{\lambda,A}$ with λ defined in (2·3) and any choice of A, it follows from properties (iii) and (ii) that

$$|P[W_n \in A] - \mathscr{P}_\lambda(A)| \leqslant 2(\lambda \wedge 1)(c_1 + c_2).$$

https://doi.org/10.1017/S0305004100059995 Published online by Cambridge University Press

Noting also that $k! \geqslant k^{k+1/2}e^{-k}$ and that $0 \leqslant xe^{1-x} \leqslant 1$ for all $x \geqslant 0$, one can deduce the following result.

THEOREM 1. *If $p = p_n$ and $k = k_n$ are chosen in such a way that $k_n^2 p_n = O(1)$ as $n \to \infty$, then*

$$d(\mathscr{L}(W_n), \mathscr{P}_{\lambda_n}) \leqslant 2\{c_1(n, k_n, p_n) + c_2(n, k_n, p_n)\} \to 0$$

as $n \to \infty$, if any of the following three conditions is satisfied:

$$\text{(i) } k_n \to \infty, \quad \text{(ii) } np_n \to \infty, \quad \text{(iii) } k_n \geqslant 2 \text{ and } np_n \to 0.$$

Here, $\mathscr{L}(W_n)$ denotes the distribution of W_n and d is the total variation metric: for random variables U and V on the non-negative integers,

$$d(\mathscr{L}(U), \mathscr{L}(V)) \equiv \sup_{A \subset \mathbb{Z}^+} |P[U \in A] - P[V \in A]|.$$

A similar argument could also be used to discuss, for instance, the total number of trees exceeding a certain size. Note, too, that no useful information is obtained in the interesting case $k_n = k$, fixed, and $p_n \sim c/n$.

The same problem could have been considered for $G_{n,M}$, the random graph uniformly distributed over all graphs with n vertices and M edges, which was the setting adopted by Erdős and Rényi (4). The detail becomes rather messier, because the graph obtained by deleting k vertices from a $G_{n,\binom{n}{2}p_n}$ is not a typical representative of a

$$G_{n-k,\binom{n}{2}p_n-k+1},$$

and a random number of further edges have to be added or subtracted in order to couple a $G_{n,\binom{n}{2}p_n}$ and a $G_{n-k,\binom{n}{2}p_n-k+1}$ on the same probability space. However, the argument is not greatly different, and the order of magnitude of the error in the approximation remains the same as that given in (2·6) and (2·7).

Instead of looking at total variation distance, Erdős and Rényi, in their theorem 2*b*, considered the convergence in distribution as $n \to \infty$ of $\lambda_n^{-\frac{1}{2}}(W_n - \lambda_n)$ to the standard Normal distribution, for k fixed and $\lambda_n \to \infty$, including the case $p_n \sim c/n$. Such results are rather weaker than the convergence in total variation established in Theorem 1, since they imply only that

$$P[\lambda_n^{-\frac{1}{2}}(W_n - \lambda_n) \in A'] - P[\mathscr{N}(0,1) \in A'] \to 0$$

for Borel sets $A' \subset R$ such that the Lebesgue measure of the boundary of A' is zero, whereas Theorem 1 allows any A' in the corresponding relation

$$P[\lambda_n^{-\frac{1}{2}}(W_n - \lambda_n) \in A'] - P(\lambda_n^{-\frac{1}{2}}(\mathscr{P}_{\lambda_n} - \lambda_n) \in A'] \to 0.$$

Moreover, Theorem 1 does not require the set A' in such expressions to be fixed as n varies, since it states that

$$\sup_{A \subset \mathbb{Z}^+} |P[W_n \in A] - \mathscr{P}_{\lambda_n}(A)| \to 0.$$

Thus one could also, for instance, deduce from Theorem 1 by taking $A = 2\mathbb{Z}^+$ that, in those cases where $\lambda_n \to \infty$, the probability of finding an even number of trees of size k_n approaches $\frac{1}{2}$. As it happens, the proof given by Erdős and Rényi of their theorem 2*b* is not correct, and their conclusion, that $\lambda_n^{-\frac{1}{2}}(W_n - \lambda_n) \xrightarrow{D} \mathscr{N}(0,1)$, is not true for $p_n \sim c/n$ (except when $c = 1$), though it follows in the other cases from

https://doi.org/10.1017/S0305004100059995 Published online by Cambridge University Press

Theorem 1 above. The case $k_n = k, p_n \sim c/n$ is considered, as a problem of convergence to the normal distribution, in Section 3.

As a further example, let G be a given prescribed graph on n vertices, and let H be derived by colouring the edges of G independently, and not necessarily with equal probabilities. Let $\mathscr{A}$ be a collection of coloured labelled subgraphs of G, and, for $\alpha \in \mathscr{A}$, set $X_\alpha = 1$ if $\alpha \subset H$, $X_\alpha = 0$ otherwise. Let $p_\alpha = P[X_\alpha = 1]$, and define $W = \Sigma_{\alpha \in \mathscr{A}} X_\alpha$, $\mu = \Sigma_{\alpha \in \mathscr{A}} p_\alpha$. Then we can write

$$\mu x(W+1) - Wx(W) = \sum_{\alpha \in \mathscr{A}} (p_\alpha - X_\alpha) x(W_\alpha + 1) + \sum_{\alpha \in \mathscr{A}} p_\alpha \{x(W+1) - x(W_\alpha + 1)\} + \sum_{\alpha \in \mathscr{A}} X_\alpha \{x(W_\alpha + 1) - x(W)\}, \quad (2{\cdot}8)$$

where $W_\alpha = \Sigma_{\beta \in \mathscr{A}, \beta \cap \alpha = \varnothing} X_\alpha$, and $\beta \cap \alpha$ is defined to be the set of edges (not necessarily of the same colour) common to α and β. Now.

$$|x(W+1) - x(W_\alpha + 1)| \leqslant (2\|x\| \wedge \Delta x) \sum_{\beta \cap \alpha \neq \varnothing} X_\beta,$$

and

$$|X_\alpha \{x(W_\alpha + 1) - x(W)\}| \leqslant (2\|x\| \wedge \Delta x) X_\alpha \sum_{\substack{\beta \cap \alpha \neq \varnothing \\ \beta \neq \alpha}} X_\beta:$$

thus, taking expectations in (2·8) and using the independence of X_α and W_α,

$$d(\mathscr{L}(W), \mathscr{P}_\mu) \leqslant 2(1 \wedge \mu^{-1}) \{ \sum_{\substack{\beta, \alpha \in \mathscr{A} \\ \beta \cap \alpha \neq \varnothing}} p_\alpha p_\beta + \sum_{\substack{\beta, \alpha \in \mathscr{A} \\ \beta \cap \alpha \neq \varnothing \\ \beta \neq \alpha}} E(X_\alpha X_\beta) \}. \quad (2{\cdot}9)$$

As an example of the use of (2·9), let $G = K_n$, and colour each edge black, independently of the others, with probability p, so that H is equivalent to a realization of $G(n, p)$. Let $\mathscr{A}$ be the set of all copies of a black *strictly balanced* graph F (see Bollobás(2)), with k vertices, l edges and automorphism group of order a. Then $p_\alpha = p^l$ for all $\alpha \in \mathscr{A}$, and $E(X_\alpha X_\beta) = p^{2l-j}$, where $j = |\beta \cap \alpha|$. In order to evaluate the right-hand side of (2·9), argue as in Bollobás(2), p. 199, to obtain the estimate

$$\sum_{\substack{\beta, \alpha \in \mathscr{A} \\ \beta \cap \alpha \neq \varnothing}} \{p_\alpha p_\beta + E(X_\alpha X_\beta)\} \leqslant 2 \sum_{k < s \leqslant 2k-2} \binom{n}{s} \left\{ \binom{s}{k} \frac{k!}{a} \right\}^2 p^{(sl+1)/k},$$

where the range of s covers the possible numbers of vertices in the union of two F-graphs which share at least one edge. Hence it follows that

$$d(\mathscr{L}(W), \mathscr{P}_\mu) \leqslant 2p^l + 4 \sum_{k < s \leqslant 2k-2} \frac{s^{2k}}{s!\, a} (np^{l/k})^{s-k} p^{l/k}.$$

Thus if, for example, F, k, l, a and p all depend upon n in such a way that $p^l |\mathscr{A}| \sim \mu$, fixed, $d(\mathscr{L}(W_n), \mathscr{P}_\mu) \to 0$ if $k^{2k} = o(n^{1/l})$, which is a condition very similar to that of theorem 2 of Bollobás(2). However, there is now no need to demand that $p^l |\mathscr{A}|$ should be asymptotically constant. So, for example, if F is fixed, $d(\mathscr{L}(W_n), \mathscr{P}_{\mu_n}) \to 0$ if

$$p = o(n^{-k/[l+1/(k-2)]}):$$

this, in the case where F is the complete graph on k vertices, improves upon theorem 2 of Schürger(5). Note also that

$$\operatorname{var} W = \sum_{\alpha \cap \beta \neq \varnothing} \{E(X_\alpha X_\beta) - p_\alpha p_\beta\},$$

https://doi.org/10.1017/S0305004100059995 Published online by Cambridge University Press

so that, in this application, if the estimate (2·9) does not converge to zero as $N \to \infty$, then μ_n^{-1} var $W_n \not\to 1$, and a Poisson approximation is unlikely to be a good one.

3. *Isolated trees of order k in $G(n, c/n)$*

In this section, we return to the problem, raised when discussing the number of isolated trees of order k, of the failure to prove Poisson convergence in $G(n, p_n)$ when $p_n \sim c/n$. Let W_n denote the number of isolated trees of order k in a realization of $G(n, p_n)$. Then, using an adaptation of Stein's original technique, the following result may be proved.

THEOREM 2. *If k is fixed and $np_n \to c$ as $n \to \infty$, then an n-independent constant C can be found such that*

$$\sup_x |P[\sigma_n^{-1}(W_n - \lambda_n) \leqslant x] - \Phi(x)| \leqslant Cn^{-\frac{1}{2}},$$

where

$$\lambda_n = EW_n \sim nk^{k-2}c^{k-1}e^{-kc}/k!$$

and

$$\sigma_n^2 = \operatorname{var} W_n \sim \lambda_n\{1 + (c-1)(kc)^{k-1}e^{-kc}/k!\}.$$

Remarks. 1. The argument is complicated by the desire to obtain the natural convergence rate of $O(n^{-\frac{1}{2}})$, and a very much simpler proof would suffice to establish a convergence rate of $O(n^{-\frac{1}{4}})$.

2. It can be seen that Poisson convergence, in the sense that $d(\mathscr{L}(W_n), \mathscr{P}_{\lambda_n}) \to 0$, would be impossible for $c \neq 1$.

Rather than prove Theorem 2 as it stands, we prove the following more general version, which overlaps the domain of validity of Theorem 1. Although the form of convergence established here is weaker than that of Theorem 1, the error estimates themselves are in some circumstances better.

THEOREM 3. *For each fixed $k \geqslant 2$, there exists a constant $C(k)$ such that*

$$\sup_x |P[\sigma_n^{-1}(W_n - \lambda_n) \leqslant x] - \Phi(x)| \leqslant C\sigma_n^{-1} \tag{3·1}$$

uniformly for all n and p_n, where $\lambda_n = EW_n$ and $\sigma_n^2 = \operatorname{var} W_n$.

Proof. The factorial moments of W_n can be expressed in the form

$$E(W_n)_r = \{n^k k^{k-2} p_n^{k-1}(1-p_n)^{nk}/k!\}^r \times \{1 - rk(rk-1)/2n + \tfrac{1}{2}rp_n(rk^2 + 3k - 2) + O_{k,r}(n^{-2}, p_n^2)\}. \tag{3·2}$$

From this, it follows first that there exists a constant K such that, for all n and p_n,

$$\lambda_n p_n \leqslant K \quad \text{and} \quad \lambda_n/n \leqslant K; \tag{3·3}$$

and then that

$$\begin{aligned}\sigma_n^2 = E(W_n)_2 + \lambda_n - \lambda_n^2 &= \lambda_n\{1 + \lambda_n k^2(p_n - n^{-1})\} + O(\lambda_n^2 n^{-2}, \lambda_n^2 p_n^2) \\ &= \lambda_n\{1 + \lambda_n k^2(p_n - n^{-1})\} + O(1).\end{aligned} \tag{3·4}$$

Here, and subsequently, K denotes a generic constant, not necessarily the same at each appearance, which depends on k alone: the same is true, unless otherwise stated,

https://doi.org/10.1017/S0305004100059995 Published online by Cambridge University Press

of the constants implied in any O or o expressions. Now straightforward calculation shows that

$$\begin{aligned}(n^{-1}-p_n)\,k^2\,.n^k k^{k-2}p_n^{k-1}(1-p_n)^{nk}/k! &\leqslant k^{-\frac{1}{2}}\,e^k\{(np_n)^{k-1}-(np_n)^k\}\,e^{-np_nk}\\ &\leqslant k^{-\frac{1}{2}}\,e^k(1-k^{-\frac{1}{2}})^{k-1}\,k^{-\frac{1}{2}}\,e^{-k+k^{\frac{1}{2}}}\\ &\leqslant \begin{cases} k^{-1}\,e^{k^{-\frac{1}{2}}}, & k\geqslant 3,\\ \frac{1}{2}e^{\sqrt{2}}(1-1/\sqrt{2}), & k=2,\end{cases}\\ &< 0{\cdot}65.\end{aligned}$$

Thus, from (3·4), there exist $n_0 = n_0(k)$ and $p_0 = p_0(k)$ such that, for all $n \geqslant n_0$ and $p_n \leqslant p_0$,

$$\sigma_n^2 \geqslant 0{\cdot}2\lambda_n + O(1).$$

Hence there exists $\lambda_0 = \lambda_0(k)$ such that, for all n and p_n with $n \geqslant n_0$ and $p_n \leqslant p_0$ for which $\lambda_n \geqslant \lambda_0$,

$$\sigma_n^2 \geqslant 0{\cdot}1\lambda_n. \tag{3·5}$$

That there exists a constant $K_1 = K_1(k)$ such that

$$\sigma_n^2 \leqslant K_1\lambda_n \tag{3·6}$$

in the same range of values of n and p_n is much simpler to establish, from (3·3) and (3·4).

What of values of n and p_n outside these ranges? Well, if λ_n is bounded, equation (3·4) implies a corresponding bound on σ_n^2. Now λ_n is uniformly bounded both in $p_n > p_0$ and in $n < n_0$, for any n_0 and $p_0 > 0$. Hence, over those values of n and p_n for which (3·5) and (3·6) do not hold, σ_n^2 is bounded above by some constant $K_2 = K_2(k)$. So, suppose that (3·1) can be established for all those values of n and p_n such that (3·5) and (3·6) hold and which also entail

$$\sigma_n^2 \geqslant 1: \tag{3·7}$$

then (3·1) will have been shown true for *all* choices of n and p_n, provided that C is chosen also to exceed $\max(K_2, 1)$. Of course, the conclusion (3·1) has no interest when $C\sigma_n^{-1} > 1$: but, because of this observation, (3·5)–(3·7) may always be taken to hold.

The main body of the argument now consists of estimating $E\{h'(U_n) - U_n h(U_n)\}$, where $U_n = \sigma_n^{-1}(W_n - \lambda_n)$ and

$$h(x) = e^{x^2/2}\int_{-\infty}^{x} e^{-u^2/2}f(u)\,du = -e^{x^2/2}\int_{x}^{\infty} e^{-u^2/2}f(u)\,du, \tag{3·8}$$

where, for any choice of $y \in R$ and $\epsilon > 0$,

$$f(u) = f_{y,\epsilon}(u) = \left\{\begin{array}{ll} -c, & u \leqslant y,\\ -c + (u-y)/\epsilon, & y \leqslant u \leqslant y+\epsilon,\\ 1-c, & y+\epsilon \leqslant u,\end{array}\right\} \tag{3·9}$$

and $c = c_{y,\epsilon}$ is chosen so that $\int_{-\infty}^{\infty} e^{-u^2/2}f(u)\,du = 0$. Clearly,

$$h'(U_n) - U_n h(U_n) = f(U_n), \tag{3.10}$$

and

$$P[U_n \geqslant y+\epsilon] \leqslant Ef(U_n) + c \leqslant P[U_n \geqslant y],$$

https://doi.org/10.1017/S0305004100059995 Published online by Cambridge University Press

allowing the eventual approximation

$$Ef_{y,\epsilon}(U_n)+c_{y,\epsilon} \leqslant P[U_n \geqslant y] \leqslant Ef_{y-\epsilon,\epsilon}(U_n)+c_{y-\epsilon,\epsilon}: \tag{3·11}$$

this, with a suitable choice of $\epsilon = \epsilon_n$, is used to derive the required result. Now $|h(x)|$ and $|h'(x)|$ are both bounded uniformly in x, y and ϵ, by H_0 and H_1, say, respectively, but $|h''(x)| \leqslant H_2/\epsilon$ is the best general estimate on the second derivative. Since, from (3·11), it is the small values of ϵ that make for good distributional approximation, this ϵ^{-1} term causes difficulties. Any uniform estimate of h'', when making Taylor approximations, leads to convergence rates of order $\sigma_n^{-\frac{1}{2}}$, and improving this rate to σ_n^{-1} is what causes most of the complication of the proof.

As a first step, write $U_n' = \sigma_n^{-1}(W_n'-\lambda_n)$ and $V_n = W_n'-W_n$, where W_n' denotes the number of isolated trees of order k in that realization of $G(n+k,p_n)$ obtained by adding k vertices and an appropriate independent selection of edges to the realization of $G(n,p_n)$ which yielded W_n. Then, trivially,

$$E\{h'(U_n)-U_nh(U_n)\} = E\{h'(U_n)-U_n'h(U_n')\}+E_1,$$

where

$$E_1 = E\{U_n'h(U_n')-U_nh(U_n)\}$$

is to be estimated later. For the first term, arguing as for (2·2) and using $\lambda_n = EW_n$,

$$E\{h'(U_n)-U_n'h(U_n')\} = E\{h'(U_n)+(\lambda_n/\sigma_n)[h(U_n')-h(U_n+\sigma_n^{-1})]\}.$$

So, writing

$$E_2 = -(\lambda_n/\sigma_n)E\{h(U_n+\sigma_n^{-1})-h(U_n)-\sigma_n^{-1}h'(U_n)\}$$

and

$$E_3 = (\lambda_n/\sigma_n)E\{h(U_n')-h(U_n)-\sigma_n^{-1}V_nh'(U_n)\},$$

both to be estimated later, and collecting terms, it follows that

$$E\{h'(U_n)-U_nh(U_n)\} = E\{h'(U_n)[1+(\lambda_n/\sigma_n^2)(V_n-1)]\}+\sum_{j=1}^{3}E_j. \tag{3.12}$$

The next step is to note that

$$E\{h'(U_n)V_n\} = E\{h'(U_n)E\{V_n|G(n,p_n)\}\},$$

and to estimate the variability of $X_n \equiv E\{V_n|G(n,p_n)\}$. Now X_n can be written as Z_n-Y_n, where

$Z_n \equiv E\{$number of isolated trees of order k in $G(n+k,p_n)$ which contain at least one vertex from $\{n+1,\ldots,n+k\}|G(n,p_n)\}$

and

$Y_n \equiv E\{$number of isolated trees of order k in $G(n,p_n)$ that become connected to $\{n+1,\ldots,n+k\}$ in $G(n+k,p_n)|G(n,p_n)\}$.

Let

$Z_n^* = kE\{$number of isolated trees of order k in $G(n+k,p_n)$ which contain the vertex $n+1$ and no others from $\{n+1,\ldots,n+k\}|G(n,p_n)\}$:

clearly, $Z_n^* \leqslant Z_n$, and also

$$E\{Z_n-Z_n^*\} = \sum_{j=2}^{k}\binom{k}{j}\binom{n}{k-j}k^{k-2}p_n^{k-1}(1-p_n)^{nk+\binom{k}{2}-k+1}$$
$$\leqslant Kn^{-1}(np_n)^{k-1}e^{-nkp_n} = O(n^{-1}); \tag{3.13}$$

K, as always, depends only on k, and not on n, p_n, y or ϵ.

https://doi.org/10.1017/S0305004100059995 Published online by Cambridge University Press

Furthermore, $Y_n = W_n\{1-(1-p_n)^{k^2}\}$ and

$$Z_n^* = k(1-p_n)^{k(k-1)} \sum_{j=1}^{k-1} \left\{ \sum_{\substack{1\leqslant s_1<\ldots<s_j \\ 1\leqslant r_i,\ 1\leqslant i\leqslant j \\ \sum_{i=1}^{j} r_i s_i = k-1}} \prod_{i=1}^{j} \left[(s_i p_n/(1-p_n))^{r_i} \binom{W_{n,s_i}}{r_i} \right] \right\}, \tag{3·14}$$

where $W_{n,s}$ is the number of isolated trees of order s in $G(n,p_n)$: so that each of Y_n and Z_n^* is a finite sum of non-negative terms of the form

$$a_n \prod_{i=1}^{k} (n^{-1}W_{n,i})^{m_i} (np_n)^{m_i},$$

where $a_n = O(1)$ as $n\to\infty$, and where $\Sigma_{i=1}^k m_i \leqslant k-1$. The random variables $n^{-1}W_{n,i}$ are all bounded between 0 and 1, and a simple calculation based on (3·2) shows that

$$\operatorname{var}(n^{-1}W_{n,i}) \leqslant K(i)\, n^{-1}(np_n)^{i-1} e^{-np_n i}.$$

Hence, by a standard argument, it follows that

$$E\left| \prod_{i=1}^{k} (n^{-1}W_{n,i})^{m_i} - \prod_{i=1}^{k} (E\{n^{-1}W_{n,i}\})^{m_i} \right| (np_n)^{\Sigma m_i}$$
$$\leqslant (np_n)^{\Sigma m_i} \sum_{i=1}^{k} m_i E|n^{-1}W_{n,i} - E(n^{-1}W_{n,i})| \leqslant Kn^{-\frac{1}{2}}, \tag{3·15}$$

and, a fortiori, that

$$(np_n)^{\Sigma m_i}\left[E\left\{ \prod_{i=1}^{k} (n^{-1}W_{n,i})^{m_i} \right\} - \prod_{i=1}^{k} (E\{n^{-1}W_{n,i}\})^{m_i} \right] = O(n^{-\frac{1}{2}}) \tag{3·16}$$

also. Collecting (3·13)–(3·16), it has been established that

$$|E\{h'(U_n)V_n\} - EV_n . Eh'(U_n)| \leqslant H_1\{E|Z_n - Z_n^*| + E|Z_n^* - Y_n - E(Z_n^* - Y_n)|$$
$$+ |E(Z_n^* - Z_n)|\} = O(n^{-\frac{1}{2}}).$$

Thus

$$|E\{h'(U_n)[1+(\lambda_n/\sigma_n^2)(V_n-1)]\}| \leqslant Kn^{-\frac{1}{2}} + H_1|1+(\lambda_n/\sigma_n^2)(EV_n-1)|;$$

and since, from (3·2),

$$EV_n = \lambda_n\{k^2(n^{-1}-p_n) + O(n^{-2}, p_n^2)\},$$

it follows that

$$\lambda_n(1-EV_n) = \sigma_n^2 + O(1),$$

and hence that

$$|E\{h'(U_n)[1+(\lambda_n/\sigma_n^2)(V_n-1)]\}| \leqslant K(n^{-\frac{1}{2}} + \sigma_n^{-2}). \tag{3.17}$$

There remain the three errors

$$E_1 = E\{(U_n+v_n)h(U_n+v_n) - U_n h(U_n)\},$$
$$E_2 = -(\lambda_n/\sigma_n)E\{h(U_n+\sigma_n^{-1}) - h(U_n) - \sigma_n^{-1}h'(U_n)\}$$

and

$$E_3 = (\lambda_n/\sigma_n)E\{h(U_n+v_n) - h(U_n) - v_n h'(U_n)\},$$

where $v_n = \sigma_n^{-1}V_n$. We start by establishing some inequalities circumscribing the distributions of U_n and v_n.

The first such is that $EU_n^4 = O(1)$ as $n\to\infty$. This follows because

$$E\{(W_n-\lambda_n)^4\} = O(\sigma_n^{-4}),$$

https://doi.org/10.1017/S0305004100059995 Published online by Cambridge University Press

as may be checked by using the expressions (3·2) for the first four factorial moments of W_n. Using a fourth moment Chebyshev inequality, we then have

$$P[|U_n| > \sigma_n^{\frac{1}{2}}] \leqslant K\sigma_n^{-2}. \tag{3·18}$$

Next, consider v_n. Clearly, as for (2·5), $v_n^+ \leqslant k\sigma_n^{-1}$. Furthermore, as in the argument preceding (2·7), conditional on $G(n, p_n)$, $-\sigma_n v_n^-$ is stochastically smaller than a random variable with the binomial distribution $B(W_n, k^2 p_n)$, and hence

$$\begin{aligned} P[-v_n^- \geqslant x | G(n,p_n)] &\leqslant e^{-x\sigma_n}\{1 + k^2 p_n(e-1)\}^{W_n} \\ &\leqslant \exp\{-x\sigma_n + 2\lambda_n k^2 p_n(e-1)\} \\ &\leqslant K e^{-x\sigma_n}, \end{aligned} \tag{3·19}$$

uniformly for all realizations of $G(n, p_n)$ such that $W_n \leqslant 2\lambda_n$. This, in particular, includes all realizations for which $|U_n| \leqslant \sigma_n^{\frac{1}{2}}$, because of (3·6) and (3·7).

Also,

$$\begin{aligned} E|v_n|^4 &\leqslant \sigma_n^{-4}(k^4 + E\{B(W_n, k^2 p_n)\}^4) \\ &\leqslant K\sigma_n^{-4}(1 + E\{(W_n p_n)^4 + W_n p_n\}) \\ &\leqslant K\sigma_n^{-4}(1 + (\lambda_n p_n)^4 + \lambda_n p_n) \leqslant K\sigma_n^{-4}, \end{aligned}$$

this last from calculations based on (3·2), and so, for $1 \leqslant r \leqslant 4$,

$$E|v_n|^r \leqslant K\sigma_n^{-r}; \tag{3·20}$$

hence, also,

$$P[|v_n| > \sigma_n^{-\frac{1}{2}}] \leqslant K\sigma_n^{-2}. \tag{3·21}$$

Now take the estimation of E_1. For each choice of y and ϵ, $\sup_{x \in R} |xh(x)| \leqslant 1$, and so, from (3·18) and (3·21),

$$E\{|(U_n + v_n)h(U_n + v_n) - U_n h(U_n)| \, I[\{|U_n| > \sigma_n^{\frac{1}{2}}\} \cup \{|v_n| > \sigma_n^{-\frac{1}{2}}\}]\} \leqslant K\sigma_n^{-2}: \tag{3·22}$$

here, and subsequently, for any random event A, $I[A]$ denotes the (indicator) random variable taking the value 1 when A occurs and zero otherwise. On the other hand, arguing for $u \geqslant 0$ – a similar argument goes through for $u \leqslant 0$ – we have, from (3·8),

$$\begin{aligned} (u+v)h(u+v) - uh(u) = &-\{(u+v)e^{uv+v^2/2} - u\}\int_0^\infty e^{-uw-w^2/2} f(u+w)\,dw \\ &+ \int_0^v (u+v)\exp\{uv + \tfrac{1}{2}v^2 - uw - \tfrac{1}{2}w^2\} f(u+w)\,dw. \end{aligned}$$

In $|u| \leqslant \sigma_n^{\frac{1}{2}}$ and $|v| \leqslant \sigma_n^{-\frac{1}{2}}$, $\exp(uv + \frac{1}{2}v^2) \leqslant e^{\frac{3}{2}}$ since, by (3·7), $\sigma_n \geqslant 1$; hence the first term is in modulus no larger than

$$e^{\frac{3}{2}}\{|v| + u^2|v| + \tfrac{1}{2}uv^2\}(u^{-1} \wedge \sqrt{(\tfrac{1}{2}\pi)}),$$

since $|f(x)| \leqslant 1$, and the second no larger than

$$e^{\frac{3}{2}}|v|\,(u + |v|).$$

Now

$$E\{|U_n v_n| \, I[\{|U_n| \leqslant \sigma_n^{\frac{1}{2}}\} \cap \{|v_n| \leqslant \sigma_n^{-\frac{1}{2}}\}]\} \leqslant (EU_n^2)^{\frac{1}{2}}(Ev_n^2)^{\frac{1}{2}} \leqslant K\sigma_n^{-1},$$

from (3.20) and because $EU_n^4 = O(1)$ as $n \to \infty$; and also

$E\{|v_n|^r I|U_n| \leqslant \sigma_n^{\frac{1}{2}}]\} = O(\sigma_n^{-1})$ for $r \geqslant 1$, from (3·20). Hence $|E_1| \leqslant K\sigma_n^{-1}$, as required.

https://doi.org/10.1017/S0305004100059995 Published online by Cambridge University Press

There remain E_2 and E_3, of which we only estimate E_3 explicitly, E_2 being similar but easier. Since, for all u and v,

$$|h(u+v)-h(u)-vh'(u)| \leqslant 2H_0+H_1|v|,$$

it follows from (3·18), (3·20) and (3·21) that

$$\begin{aligned}(\lambda_n/\sigma_n)\,&|E\{(h(U_n+v_n)-h(U_n)-v_nh'(U_n))\,I[\{|U_n|>\sigma_n^{\frac{1}{2}}\}\cup\{|v_n|>\sigma_n^{-\frac{1}{2}}\}]\}|\\ &\leqslant(\lambda_n/\sigma_n)\{2H_0P[\{|U_n|>\sigma_n^{\frac{1}{2}}\}\cup\{|v_n|>\sigma_n^{-\frac{1}{2}}\}]\\ &\quad+H_1E(|v_n|\,I[\{|v_n|\leqslant\sigma_n^{-\frac{1}{2}}\}\cap\{|U_n|>\sigma_n^{\frac{1}{2}}\}])+H_1E(|v_n|\,I[|v_n|>\sigma_n^{-\frac{1}{2}}])\}\\ &\leqslant K(\lambda_n/\sigma_n)\,(\sigma_n^{-2}+\sigma_n^{-\frac{5}{2}}+H_1\sigma_n^{\frac{3}{2}}E|v_n|^4)\\ &\leqslant K\sigma_n^{-1}, \end{aligned}\tag{3·23}$$

this last by (3·5).

For the rest, if $u\geqslant 0$, use (3·8) and (3·10) to write

$$\begin{aligned}h(u+v)-h(u)-vh'(u) = &-(e^{uv+v^2/2}-1-uv)\int_0^\infty e^{-uw-w^2/2}f(u+w)\,dw\\ &+e^{uv+v^2/2}\int_0^v e^{-uw-w^2/2}f(u+w)\,dw-vf(u).\end{aligned}$$

In $|u|\leqslant\sigma_n^{\frac{1}{2}}$ and $|v|\leqslant\sigma_n^{-\frac{1}{2}}$, the first of these terms is bounded in modulus by

$$\tfrac{1}{2}\{v^2+e^{\frac{3}{2}}(uv+\tfrac{1}{2}v^2)^2\}(u^{-1}\wedge\sqrt{(\tfrac{1}{2}\pi)}),$$

whereas

$$\begin{aligned}&\left|\int_0^v[\exp(uv+\tfrac{1}{2}v^2-uw-\tfrac{1}{2}w^2)-1]f(u+w)\,dw\right|\\ &\qquad\leqslant\int_0^{|v|}e^{\frac{3}{2}}\{w|u|+\tfrac{1}{2}(v^2-w^2)\}\,dw\leqslant K(v^2|u|+|v|^3):\end{aligned}$$

thus, on $|u|\leqslant\sigma_n^{\frac{1}{2}}$ and $|v|\leqslant\sigma_n^{-\frac{1}{2}}$, since again a similar estimate holds for $u\leqslant 0$, it follows that

$$\left|h(u+v)-h(u)-vh'(u)-\int_0^v[f(u+w)-f(u)]\,dw\right|\leqslant Kv^2(1+|u|).$$

Hence

$$\begin{aligned}(\lambda_n/\sigma_n)\Big|&E\{[h(U_n+v_n)-h(U_n)-v_nh'(U_n)]\,I[\{|U_n|\leqslant\sigma_n^{\frac{1}{2}}\}\cap\{|v_n|\leqslant\sigma_n^{-\frac{1}{2}}\}]\}\\ &-E\left\{\int_0^{v_n}[f(U_n+w)-f(U_n)]\,dw\,I[\{|U_n|\leqslant\sigma_n^{\frac{1}{2}}\}\cap\{|v_n|\leqslant\sigma_n^{-\frac{1}{2}}\}]\right\}\Big|\\ &\qquad\leqslant K(\lambda_n/\sigma_n)\{Ev_n^2+(EU_n^2)^{\frac{1}{2}}(Ev_n^4)^{\frac{1}{2}}\}\leqslant K\sigma_n^{-1}\end{aligned}\tag{3·24}$$

by (3·5) and (3·20), and it remains to consider $\int_0^v[f(u+w)-f(u)]\,dw$.

Recalling the definition (3·9) of f, it is easy to make the following crude estimate:

$$\left|\int_0^v[f(u+w)-f(u)]\,dw\right|\leqslant\{(|v|-g(u))\wedge(|v|-g(u))^2/2\epsilon\}\,I[|v|\geqslant g(u)],\tag{3·25}$$

where $g=g_{v,\epsilon}$ is defined by

$$g(u)=\begin{cases}(y-u)^+, & u\leqslant y+\epsilon/2,\\ (u-y-\epsilon)^+, & u\geqslant y+\epsilon/2.\end{cases}\tag{3·26}$$

Hence, from (3·19), uniformly over realizations of $G(n,p_n)$ such that $|U_n|\leqslant\sigma_n^{\frac{1}{2}}$,

$$E\left\{\left|\int_0^{v_n}[f(U_n+w)-f(U_n)]\,dw\right|\,|G(n,p_n)\right\}\leqslant K\sigma_n^{-1}e^{-g(U_n)\sigma_n}(1\wedge 1/\epsilon\sigma_n).\tag{3·27}$$

https://doi.org/10.1017/S0305004100059995 Published online by Cambridge University Press

Let Δ_n denote $\sup_{x \in R} |P[U_n \leqslant x] - \Phi(x)|$. Since $g_{y,\epsilon}(u)$ is an increasing function of $|y-u+\epsilon/2|$ and since the standard normal density is bounded by $1/\sqrt{(2\pi)}$, we have, from (3·27), the estimate

$$\begin{aligned}(\lambda_n/\sigma_n)\left| E\left\{\int_0^{v_n} [f(U_n+w)-f(U_n)]\,dw\, I[\{|U_n| \leqslant \sigma_n^{\frac{1}{2}}\} \cap \{|v_n| \leqslant \sigma_n^{-\frac{1}{2}}\}]\right\}\right| \\ \leqslant K(1 \wedge 1/\epsilon\sigma_n)\left\{2\Delta_n + (2\pi)^{-\frac{1}{2}}[\epsilon + 2\int_0^\infty e^{-t\sigma_n}\,dt\right\} \\ \leqslant K(\epsilon^{-1}\Delta_n + 1)\,\sigma_n^{-1}, \qquad (3.28)\end{aligned}$$

uniformly in y. Collecting (3·23), (3·24) and (3·28), it follows that

$$E_3 \leqslant K\sigma_n^{-1}(1+\epsilon^{-1}\Delta_n). \qquad (3·29)$$

Since a similar argument yields the same estimate for E_2, we have arrived finally at the estimate

$$|E\{h'(U_n) - U_n h(U_n)\}| \leqslant K\sigma_n^{-1}(1+\epsilon^{-1}\Delta_n), \qquad (3.30)$$

uniformly in y.

Returning to (3·11), we note that, from (3·9),

$$(2\pi)^{\frac{1}{2}} c_{y,\epsilon} = \int_y^{y+\epsilon} \epsilon^{-1}(u-y)\,e^{-u^2/2}\,du + \int_{y+\epsilon}^\infty e^{-u^2/2}\,du,$$

so that

$$|c_{y,\epsilon} - (1-\Phi(y))| \leqslant \epsilon/\sqrt{(2\pi)}$$

and

$$|c_{y-\epsilon,\epsilon} - (1-\Phi(y))| \leqslant \epsilon/\sqrt{(2\pi)}.$$

Hence, from (3·10), (3·11) and (3·30), we deduce that

$$\Delta_n = \sup_{x \in R} |P[U_n \geqslant x] - (1-\Phi(x))| \leqslant K\sigma_n^{-1}(1+\epsilon^{-1}\Delta_n) + \epsilon/\sqrt{(2\pi)}.$$

Choosing $\epsilon = \alpha\sigma_n^{-1}$ for any $\alpha > 0$, this yelds the inequality

$$\Delta_n(1-K/\alpha) \leqslant (K + \alpha/\sqrt{(2\pi)})\,\sigma_n^{-1},$$

or, choosing $\alpha = 2K$, say, $\Delta_n \leqslant 4K\sigma_n^{-1}$. The theorem is proved.

REFERENCES

(1) Barbour, A. D. and Eagleson, G. K. Poisson approximation for some statistics based on exchangeable trials. (1982).

(2) Bollobás, B. Threshold functions for small subgraphs. *Math. Proc. Cambridge Phil. Soc.* **90** (1981), 197–206.

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

(4) Erdős, P. and Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* V, (1960), 17–61.

(5) Schürger, K. Limit theorems for complete subgraphs of random graphs. *Period. Mat. Hung.* **10** (1979), 47–53.

(6) Stein, C. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. VIth Berk. Symp. Math. Stat. Prob.* **2** (1970), 583–602.

https://doi.org/10.1017/S0305004100059995 Published online by Cambridge University Press