Lecture XIII. AN APPLICATION TO THE THEORY OF RANDOM GRAPHS

Consider a random graph $G(n)$ on n vertices in which each possible edge is present with probability p, independently of all others. Let $W_{n,k}$ (also abbreviated W_n) be the number of isolated trees of order k in $G(n)$. Conditions are given for W_n to have approximately a Poisson distribution. This lecture is based on a paper of Barbour (1982), who also gave conditions for a normal approximation to be valid.

I shall use essentially the same notation as Barbour. Denoting the set of vertices by $\{1,\ldots,n\}$, I shall think of the random graph $G(n)$ as a random subset of the set of all two-element subsets $\{i,j\}$ of $\{1,\ldots,n\}$. If $\{i,j\} \in G(n)$ I shall say that $\{i,j\}$ is an edge of the random graph $G(n)$, which will be constructed by having the events $\{\{i,j\} \in G(n)\}$ occur independently with common probability p. Let D_n be the set of all k-tuples $i = (i_1,i_2,\ldots,i_k)$ of natural numbers with $1 \leq i_1 < i_2 <\ldots< i_k \leq n$. For each $i \in D_n$ let $X_i = 1$ if there is in $G(n)$ an isolated tree spanning the vertices $i_1,\ldots,i_k$, and otherwise let $X_i = 0$. A tree is, by definition, a connected graph containing no cycles, and it is isolated if $G(n)$ has no edge with one vertex in the tree and one not in the tree. Then W_n, the number of isolated trees of order k in $G(n)$ is given by

$$W_n = \sum_{i \in D_n} X_i. \tag{1}$$

The expectation λ of W_n is given by

(2)
$$\lambda = EW_n = \binom{n}{k} P\{X_i = 1\}$$
$$= \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)+\binom{k}{2}-k+1} .$$

The argument for this is as follows. By a theorem of Cayley (see, for example, Graver and Watkins (1977), p. 322) there are k^{k-2} different trees on k specified vertices. In order that a given isolated tree on these k vertices be realized by the process indicated it is necessary and sufficient that the k-1 connections of the specified tree be made, but none of the $\binom{k}{2}-k+1$ other connections among these k vertices, and that none of the k(n-k) possible connections of these k vertices to vertices outside this set be made. Let us also compute the variance of W_n. If i and i' are disjoint elements of D_n, then, by essentially the same argument as in (2),

(3)
$$EX_i X_{i'} = k^{2(k-2)} p^{2(k-1)} (1-p)^{2k(n-2k)+\binom{2k}{2}-2(k-1)} ,$$

but if i and i' are neither identical nor disjoint, $EX_i X_{i'} = 0$. It follows that

(4)
$$\text{Var } W_n - EW_n = EW_n^2 - EW_n - (EW_n)^2$$
$$= \binom{n}{k}\binom{n-k}{k} k^{2(k-2)} p^{2(k-1)} (1-p)^{2kn-2k^2-3k+2}$$
$$- \binom{n}{k}^2 k^{2(k-2)} p^{2(k-1)} (1-p)^{2kn-k^2-3k+2}$$
$$= \left\{ \left[\prod_{i=0}^{k-1} \left(1 - \frac{k}{n-i}\right) \right] (1-p)^{-k^2} - 1 \right\} \lambda^2 .$$

Later we shall have to make a careful study of the dependence of the mean and variance of W_n on n, p, and k.

Now let us look at the Poisson approximation for the distribution of W_n. For arbitrary $f: Z^+ \to R$ and $i \in D_n$ we have

(5) $$EX_i f(W_n) = P\{X_i=1\}Ef(W^*_{n-k}+1)$$

where W^*_{n-k} is the number of isolated trees of order k in the graph G* obtained from G(n) by dropping the vertices $i_1,\ldots,i_k$ and all edges containing any of these vertices. Summing (5) over i, using the fact that W^*_{n-k} has the same distribution as W_{n-k}, we obtain

(6) $$EW_n f(W_n) = \lambda Ef(W_{n-k}+1).$$

Consequently

(7) $$E[\lambda f(W_n+1) - W_n f(W_n)] = \lambda E[f(W_n+1) - f(W_{n-k}+1)].$$

Substituting for f the function $U_\lambda h$, defined by (VIII.18), we obtain, for arbitrary $h\colon Z^+ \to R$,

(8) $$Eh(W_n) - \mathcal{P}_\lambda h = E[\lambda U_\lambda h(W_n+1) - W_n U_\lambda h(W_n)]$$
$$= \lambda E[U_\lambda h(W_n+1) - U_\lambda h(W_{n-k}+1)].$$

In particular, for $h = h_A$ defined by

(9) $$h_A(w) = \begin{cases} 1 \text{ if } w \in A \\ 0 \text{ if } w \notin A, \end{cases}$$

we obtain

(10) $$P\{W_n \in A\} - e^{-\lambda} \sum_{w\in A} \frac{\lambda^w}{w!}$$
$$= \lambda E[U_\lambda h_A(W_n+1) - U_\lambda h_A(W_{n-k}+1)].$$

But we have seen in (VIII.42) that, for all λ, w, and A,

(11) $$|U_\lambda h_A(w+1) - U_\lambda h_A(w)| \leq 1\wedge\lambda^{-1}.$$

It follows from (10) and (11) that

(12) $$|P\{W_n \in A\} - e^{-\lambda} \sum_{w \in A} \frac{\lambda^w}{w!}| \le (1 \wedge \lambda) E|W_n - W_{n-k}|.$$

In order to bound $E|W_n-W_{n-k}|$ we first observe that

(13) $$(W_n-W_{n-k})_+ \le \sum_{j=n-k+1}^{n} Y_j$$

where Y_j equals one if j belongs to an isolated tree of order k in G(n), but otherwise zero. Consequently

(14) $$E(W_n-W_{n-k})_+ \le k\binom{n-1}{k-1} k^{k-2}\, p^{k-1}(1-p)^{k(n-k)+\binom{k-1}{2}}$$

$$= \frac{k\binom{n-1}{k-1}}{\binom{n}{k}} \lambda = \frac{k^2}{n} \lambda.$$

Of course the argument for the inequality (14) is that there are k terms on the right-hand side of (13), that each point j can form a tree of order k with any of the $\binom{n-1}{k-1}$ (k-1)-element subsets of the remaining points and that there are k^{k-2} trees on these k points. The remaining factor is the probability that a particular such tree will be realized. Furthermore, an upper bound for $W_{n-k}-W_n$ is the number of isolated trees of order k in G(n-k) that are destroyed by being connected to vertices in $\{n-k+1,\ldots,n\}$. Consequently, writing $\lambda(n)$ and $\lambda(n-k)$ for EW_n and EW_{n-k}, we have

(15) $$E(W_{n-k}-W_n)_+ \le [1 - (1-p)^{k^2}]\lambda(n-k).$$

Finally, (12), (14), and (15) yield

(16) $$|P\{W_n \in A\} - e^{-\lambda} \sum_{w \in A} \frac{\lambda^w}{w!}|$$

$$\le \left\{\frac{k^2}{n} + [1 - (1-p)^{k^2}] \frac{\lambda(n-k)}{\lambda(n)}\right\} [\lambda(n) \wedge \lambda^2(n)].$$

Now we must study the behavior of $\lambda(n,k,p)$ (which was abbreviated as λ

or $\lambda(n)$ in the above) as a function of n, k, and p, in part as an aid in the study of the bound given in (16) for the error in the Poisson approximation to the distribution of W_n. It will be convenient to write

$$\rho = -\log(1-p) \tag{17}$$

and

$$c = n\rho . \tag{18}$$

Then, by (2)

$$\lambda(n,k,p) = \alpha(k)\beta(k,p)\gamma(n,k,p), \tag{19}$$

where

$$\alpha(k) = \frac{k^{k+\frac{1}{2}}e^{-k}}{k!} \tag{20}$$

$$\beta(k,p) = \frac{k^{k-2}\, p^{k-1}(1-p)^{-k^2+\binom{k-1}{2}}}{k^{k+\frac{1}{2}}e^{-k}} \tag{21}$$

$$= k^{-5/2}e^{k}\, p^{k-1}(1-p)^{-\frac{k^2+3k}{2}+1}$$

and

$$\gamma(n,k,p) = n_{(k)}(1-p)^{nk} = [\prod_{j=1}^{k-1}(1-\frac{j}{n})]n^k\, e^{-kc} \tag{22}$$

$$= \exp[-\frac{k(k-1)}{2n} - \frac{\theta k^3}{3n^2}]n^k\, e^{-kc}$$

for $k < \frac{n}{2}$, with $0 < \theta < 1$. It follows that

$$\frac{\lambda(n,k,p)}{\alpha(k)} \tag{23}$$

$$= n\, c^{k-1}\, e^{k(1-c)}\exp[(\frac{k^2+3k}{2}-1)\rho - \frac{k(k-1)}{2c}\rho - \frac{\theta k^3}{3n^2}]k^{-5/2}(\frac{p}{\rho})^{k-1}.$$

We shall also need to evaluate the second term in braces in (16). We have

(24) $$\frac{\lambda(n-k)}{\lambda(n)} = \prod_{j=0}^{k-1} \left(1 - \frac{k}{n-j}\right)(1-p)^{-k^2}$$

$$< \exp[k^2(\rho - \frac{1}{n})].$$

Thus (16) yields

(25) $$\left|P\{W_{n,k} \in A\} - e^{-\lambda(n)} \sum_{w\in A} \frac{(\lambda(n))^w}{w!}\right|$$

$$\leq \left[\frac{k^2}{n} + (e^{k^2\rho}-1)e^{-\frac{k^2}{n}}\right]\lambda(n).$$

Let us first try to get some idea of the behavior of λ and then return to the bound in (25). If n, k, p are varied in such a way that $k^2/n \to 0$ and $k^2p \to 0$ then, by (23),

(26) $$\lambda(n,k,p) \sim nk^{-5/2}(ce^{1-c})^{k-1}\, e^{1-c}\, \alpha(k)$$

where $c = n\rho \sim np$, and $\alpha(k)$ is bounded away from 0 and ∞ by Stirling's formula. Since ce^{1-c} attains a maximum value of 1 at c = 1, (26) shows that the expected number of isolated k-trees with k much larger than log n is small unless np is close to 1. When np is sufficiently close to 1 the expected number of isolated k-trees approaches 0 only for k appreciably larger than $n^{2/5}$. Of course all of these remarks are subject to the condition imposed earlier that $k^2/n \to 0$ and $k^2p \to 0$.

Now let us return to the evaluation of the bound in (25) subject only to the condition that k^2p remain bounded. Then (26) still gives the correct order of magnitude of λ so that, for some constant B, (25) yields

(27) $$\left|P\{W_{n,k} \in A\} - e^{-\lambda(n)} \sum_{w\in A} \frac{(\lambda(n))^w}{w!}\right|$$

$$\leq Bk^{-\frac{1}{2}}(1+c)e^{1-c}(ce^{1-c})^{k-1}.$$

The bound on the r.h.s. of (27) approaches 0 if $k \to \infty$ or $c \to \infty$ or $k \geq 2$ and $c \to 0$.

Barbour went on to show that for fixed k, the error in the approximation to the distribution of $W_{n,k}$ by a normal distribution with mean $\lambda(n,k,p)$ and variance

(28) $$\sigma^2(n,k,p) = \lambda[1 + \lambda\{\exp(k^2(\rho - \frac{1}{n}) - \theta k^3/n^2) - 1\}]$$

is of the order of $\sigma^{-1}(n,k,p)$, uniformly in n and p. This suggests that the bound (27) is not sharp in order of magnitude in the neighborhood of c = 1.