Lecture 27: Subgraph counts continued

We are in the middle of trying to understand when subgraph counts are approximately Poisson. Last time we showed Theorem 26.3, which gave a bound on the total variation distance between W equal to the number of copies of a subgraph H and an appropriate Poisson distribution, in terms of the mean and variance of W. The following result rewrites the bound of Theorem 26.3 by expressing the mean and variance of W in terms of certain properties of H. Let Γ denote the set of all copies of H in K_n, the complete graph on n vertices.

Corollary 27.1. *Fix n and p, let H be a graph with no isolated vertices and e_H edges, let W be the number of copies H in $G \sim \mathrm{ER}(n,p)$, and write $\mu = \mathbb{E}[W]$.*

1. *For each $\alpha \in \Gamma$ and $H' \subseteq H$, let $\Gamma_\alpha^{H'} \subseteq \Gamma$ be the set of subgraphs of K_n isomorphic to H whose intersection with α is H'. Then for any $\alpha \in \Gamma$,*

$$d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{Po}(\mu)\big) \le \min\{1,\mu\}\bigg(p^{e_H} + \sum_{\substack{H' \subsetneq H \\ e_{H'}>0}} |\Gamma_\alpha^{H'}| p^{e_H - e_{H'}}\bigg).$$

2. *For each $\alpha \in \Gamma$ and $t = 1, \dots, e_H$, let $\Gamma_\alpha^t \subseteq \Gamma$ be the set of subgraphs of K_n isomorphic to H with exactly t edges not in α. Then for any $\alpha \in \Gamma$,*

$$d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{Po}(\mu)\big) \le \min\{1,\mu\}\left(p^{e_H} + \sum_{t=1}^{e_H-1} |\Gamma_\alpha^t| \left(p^t - p^{e_H}\right)\right).$$

Proof. As usual, write $W = \sum_{\alpha\in\Gamma} X_\alpha$, where X_α is the indicator that there is a copy of H in G at α. The corollary follows from Theorem 26.3 after deriving the mean and variance of W. For Part 1, the terms $|\Gamma_\alpha^{H'}|$ account for the number of covariance terms for different types of pairs of indicators. In detail, noting the key property that $\mathrm{Cov}(X_\alpha, X_\beta) = 0$ if $\alpha \cap \beta = \emptyset$, we have

$$\begin{aligned}
\mathrm{Var}(W) &= \sum_{\alpha\in\Gamma} \mathrm{Var}(X_\alpha) + \sum_{\alpha\in\Gamma}\sum_{\beta\neq\alpha} \mathrm{Cov}(X_\alpha, X_\beta) \\
&= \mu(1-p^{e_H}) + \sum_{\alpha\in\Gamma} \sum_{\substack{H'\subsetneq H \\ e_{H'}>0}} \sum_{\beta\in\Gamma_\alpha^{H'}} \mathrm{Cov}(X_\alpha, X_\beta) \\
&= \mu(1-p^{e_H}) + \sum_{\alpha\in\Gamma} p^{e_H} \sum_{\substack{H'\subsetneq H \\ e_{H'}>0}} \sum_{\beta\in\Gamma_\alpha^{H'}} \left(\mathbb{E}[X_\beta | X_\alpha = 1] - p^{e_H}\right) \\
&= \mu\bigg(1 - p^{e_H} + \sum_{\substack{H'\subsetneq H \\ e_{H'}>0}} |\Gamma_\alpha^{H'}| \big(p^{e_H - e_{H'}} - p^{e_H}\big)\bigg),
\end{aligned}$$

since $\mu = \sum_{\alpha\in\Gamma} p^{e_H}$, and for $\beta \in \Gamma_\alpha^{H'}$, $\mathbb{E}[X_\beta | X_\alpha = 1] = p^{e_H - e_{H'}}$. Similarly, for Part 2,

$$\begin{aligned}
\mathrm{Var}(W) &= \mu(1-p^{e_H}) + \sum_{\alpha\in\Gamma} \sum_{t=1}^{e_H} \sum_{\beta\in\Gamma_\alpha^t} \mathrm{Cov}(X_\alpha, X_\beta) \\
&= \mu(1-p^{e_H}) + \sum_{\alpha\in\Gamma} p^{e_H} \sum_{t=1}^{e_H} \sum_{\beta\in\Gamma_\alpha^t} \left(\mathbb{E}[X_\beta | X_\alpha = 1] - p^{e_H}\right) \\
&= \mu\left(1 - p^{e_H} + \sum_{t=1}^{e_H-1} |\Gamma_\alpha^t| \left(p^t - p^{e_H}\right)\right),
\end{aligned}$$

since $\mu = \sum_{\alpha\in\Gamma} p^{e_H}$, and for $\beta \in \Gamma_\alpha^t$, $\mathbb{E}[X_\beta | X_\alpha = 1] = p^t$. □

We can use the first part of the corollary to shed light on the issues brought up at the start of the subgraph counts discussion at the end of Lecture 26. Recall that discussion, and that for H a fixed subgraph with no isolated vertices, we define

$$r_H := \min\Big\{ \frac{v_{H'}}{e_{H'}} : H' \text{ is a subgraph of } H \Big\}.$$

We also let a_H be the size of the automorphism group of H, which is the number of ways to permute the vertices of H with their edges so that the resulting graph is H. For example, if H is a complete graph on k vertices, then $a_H = k!$, if H is a k-cycle, then $a_H = 2k$, coming from the k rotations, each combined with a reflection. We saw this concept in Lecture 3 when working out the threshold for the number of copies of H in an Erdős-Rényi random graph.

Theorem 27.2. *Let H be a graph with no isolated vertices such that for all proper subgraphs $H' \subsetneq H$,*

$$\frac{v_H}{e_H} < \frac{v_{H'}}{e_{H'}} \tag{27.1}$$

(such H are said to be strictly balanced*). Then $r_H = v_H/e_H$, and setting $p_n = (\nu\, a_H)^{1/e_H} n^{-r_H}$ for fixed $\nu > 0$, let W_n be the number of copies H in $G_n \sim \mathrm{ER}(n, p_n)$. Then $\mu_n := \mathbb{E}[W_n] \to \nu$, and*

$$d_{\mathrm{TV}}\big(\mathscr{L}(W_n), \mathrm{Po}(\nu)\big) \xrightarrow[n\to\infty]{} 0.$$

A bound on the total variation distance can be read from (27.5) *and* (27.6) *below.*

Proof. That $r_H = v_H/e_H$ is by the definition of r_H and (27.1). To prove the convergence of the mean, write W_n as a sum of indicators as above, and then we have

$$\begin{aligned}
\mathbb{E}[W_n] &= \sum_{\alpha\in\Gamma} p^{e_H} \\
&= |\Gamma| p^{e_H} \\
&= \binom{n}{v_H} \frac{v_H!}{a_H} p^{e_H} \\
&= \nu\, n(n-1)\cdots(n-v_H+1)\, n^{-v_H} \qquad (27.2)\\
&= \nu + \mathrm{O}(n^{-1}). \qquad (27.3)
\end{aligned}$$

For the bound, we first use the triangle inequality to find

$$d_{\mathrm{TV}}\big(\mathscr{L}(W_n), \mathrm{Po}(\nu)\big) \le d_{\mathrm{TV}}\big(\mathscr{L}(W_n), \mathrm{Po}(\mu_n)\big) + d_{\mathrm{TV}}\big(\mathrm{Po}(\mu_n), \mathrm{Po}(\nu)\big). \tag{27.4}$$

For the second term, we couple $\mathrm{Po}(\mu_n)$ and $\mathrm{Po}(\nu)$ as follows. By (27.2), $\mu_n \le \nu$, and so we set $Z_n \sim \mathrm{Po}(\mu_n)$ and $Z = Z_n + Y_n$, where $Y_n \sim \mathrm{Po}(\nu - \mu_n)$ is independent of Z_n. Then (Z_n, Z) is the desired coupling, and

$$d_{\mathrm{TV}}\big(\mathrm{Po}(\mu_n), \mathrm{Po}(\nu)\big) \le \mathbb{P}(Z_n \ne Z) = \mathbb{P}(Y_n \ne 0) = 1 - e^{-(\nu-\mu_n)} \le \nu - \mu_n = \mathrm{O}(n^{-1}), \tag{27.5}$$

where the last equality is because of (27.3). For the first term of (27.4), we use the first part of Corollary 27.1 to find that

$$d_{\mathrm{TV}}\big(\mathscr{L}(W_n), \mathrm{Po}(\mu_n)\big) \le p^{e_H} + \sum_{\substack{H'\subsetneq H \\ e_{H'}>0}} |\Gamma_\alpha^{H'}| p^{e_H - e_{H'}}.$$

Now, we claim $|\Gamma_\alpha^{H'}| = \mathrm{O}(n^{v_H - v_{H'}})$, which holds because each $\beta \in \Gamma_\alpha^{H'}$ overlaps with α in $v_{H'}$ vertices, and so we are free to choose the remaining $v_H - v_{H'}$ vertices among the order n not

in α (there is also a combinatorial factor that does not depend on n). Therefore, there are constants c' and c'' such that

$$\begin{aligned} d_{\mathrm{TV}}\big(\mathscr{L}(W_n), \mathrm{Po}(\mu_n)\big) &\le p^{e_H} + c' \sum_{\substack{H' \subsetneq H \\ e_{H'}>0}} n^{v_H - v_{H'}} p^{e_H - e_{H'}} \\ &= p^{e_H} + c' n^{v_H} p^{e_H} \sum_{\substack{H' \subsetneq H \\ e_{H'}>0}} \left(n^{v_{H'}} p^{e_{H'}}\right)^{-1} \\ &\le p^{e_H} + c'' n^{v_H} p^{e_H} \sum_{\substack{H' \subsetneq H \\ e_{H'}>0}} \left(n^{v_{H'}/e_{H'} - v_H/e_H}\right)^{-e_{H'}} \qquad (27.6) \\ &\xrightarrow[n\to\infty]{} 0, \end{aligned}$$

where the convergence to zero is because $n^{v_H} p^{e_H}$ is of constant order, and by (27.1), for all proper subgraphs $H' \subsetneq H$,

$$v_{H'}/e_{H'} - v_H/e_H > 0. \qquad \square$$

The second part of the Corollary 27.1 is useful in explicit small examples, as we now show.

Example 27.3 (Triangles)**.** Let H be a triangle. In this case, $e_H = 3$, $|\Gamma^2_\alpha| = 3(n-3)$, and $|\Gamma^1_\alpha| = 0$ since triangles can only share zero, one, or all three edges. Thus Corollary 27.1 implies that for W the number of triangles in $G \sim \mathrm{ER}(n,p)$ and $\mu = \binom{n}{3}p^3$, we have

$$d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{Po}(\mu)\big) \le \min\{1,\mu\}\left(p^3 + 3(n-3)p^2(1-p)\right). \qquad (27.7)$$

Since $\mu = \binom{n}{3}p^3$, we can view (27.7) as an asymptotic result with p of order $1/n$. In this case, (27.7) is of order $1/n$. However, note that the bound is small as long as $p \ll n^{-1/2}$, where there is no convergence (for $n^{-1} \ll p \ll n^{-1/2}$, the mean and variance go to infinity and so W is close to a Poisson with large mean, which is also close to a normal distribution).

Example 27.4 (k-cycles)**.** More generally, we can let H be a k-cycle (a triangle is 3-cycle). Now note that for some constants c_t and C_k,

$$|\Gamma^t_\alpha| \le \binom{k}{k-t} c_t n^{t-1} \le C_k n^{t-1},$$

since we choose the $k-t$ edges shared in the k-cycle α, and then we have order n^{t-1} sequences of vertices to create a cycle with t edges outside of the $k-t$ edges shared with α. The second equality follows by maximizing $\binom{k}{k-t}c_t$ over the possible values of t. We can now find for W the number of k-cycles in $G \sim \mathrm{ER}(n,p)$ and $\mu = \mathbb{E}[W]$,

$$d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{Po}(\mu)\big) \le \min\{1,\mu\}\left(p^k + C_k p \sum_{t=1}^{k-1} (np)^{t-1}\right). \qquad (27.8)$$

To interpret this bound asymptotically, we note that $\mu = |\Gamma| p^k$ and

$$|\Gamma| = \binom{n}{k}\frac{(k-1)!}{2},$$

since the number of non-isomorphic k-cycles on K_k is $k!/(2k)$ (since $k!$ is the number of permutations of the vertices, which over counts by a factor of $2k$ due to reflections and rotations). Thus μ is of order $(np)^k$ for fixed k, so if we take p to be of order $1/n$, then there is Poisson convergence and (27.8) is of order $1/n$. The bound is small as long as $p \ll n^{-(k-2)/(k-1)}$.