Problem sheet 2

Problem 2.1. Let $\mathcal{G}_n$ be an Erdős-Rényi graph on n vertices, where each edge is present independently with probability c/n, $c > 0$. Let $\mathcal{X}_n$ be the number of triangles in $\mathcal{G}_n$, where we call $\{u, v, w\} \subset \mathcal{V}_n$ a triangle if $u \sim v, v \sim w$ and $w \sim u$. Find the limiting distribution of $\mathcal{X}_n$, as $n \to \infty$.

Proof. Let $u, v, w \in [n]$ be three distinct vertices. By assumption the form a triangle with probability

$$p_n := \mathbb{P}(u \sim v, v \sim w, w \sim u) = \tfrac{c^3}{n^3}.$$

Let $X_{\{u,v,w\}} = \mathbb{1}_{\{u\sim v, v\sim w, w\sim u\}}$ be the indicator of the event that u, v, w form a triangle, which is Bernoulli(p_n) distributed. Then, by independence of all edges, the random variables $(X_{\{x,y,z\}})$ form a collection of Bernoulli(p_n) random variables, which are independent, whenever the triangles in question share no common edge, as

$$\mathbb{P}(X_{\{u,v,w\}} = 1, X_{\{u,v,x\}} = 1) = \mathbb{P}(u \sim v, v \sim w, v \sim x, w \sim u, x \sim u) = \tfrac{c^5}{n^5} \neq p_n^2. \quad (1)$$

Still, by definition

$$\mathcal{X}_n = \sum_{u=1}^{n} \sum_{\substack{v=1 \\ v\neq u}}^{n} \sum_{\substack{w=1 \\ w\notin\{u,v\}}}^{n} X_{\{u,v,w\}}.$$

If all these random variables were independent, then $\mathcal{X}_n$ would be binomially distributed with parameters $\binom{n}{3}$ and p_n, and as

$$\binom{n}{3} p_n = \frac{n^3 - 3n^2 + 2n}{6} \frac{c^3}{n^3} \longrightarrow \frac{c^3}{6},$$

the limiting distribution would be Poisson($c^3/6$). We aim to showing that this is still the case. First of all, recall that a discrete random variable is uniquely defined via its probability generating function

$$g_n(t) = \mathbb{E}t^{\mathcal{X}_n} = \sum_{j}^{\binom{n}{3}} t^j \underbrace{\mathbb{P}(\mathcal{X}_n = j)}_{=:\mathbf{p}(n)}.$$

The important property, giving the probability generating function its name, is $\mathbf{p}(r) = g_n^{(r)}(0)/r!$. However, by definition, $g_n^{(r)}(0) = \mathbb{E}[\mathcal{X}_n(\mathcal{X}_n - 1)\dots(\mathcal{X}_n - r + 1)]$, the r-th factorial moment. Hence, we have to show that factorial moments of $\mathcal{X}_n$ converge towards the factorial moments of a Poisson random variable. Using that the PGF of a Poisson(λ) random variable is $e^{\lambda(t-1)}$, it is straightforward that the factorial moments are simply of the form λ^r. Hence, we have to show, that for each $r \in \mathbb{N}$

$$\mathbb{E}[(\mathcal{X}_n)_r] := \mathbb{E}[\mathcal{X}_n(\mathcal{X}_n - 1)\dots(\mathcal{X}_n - r + 1)] \to= \left(\tfrac{c^3}{6}\right)^r =: \lambda^r.$$

To easy notation, let $\tau_1, \dots, \tau_{\binom{n}{3}}$ denote all potential triangles, i.e., each τ represents a subset of size 3 of $[n]$. Consequently,

$$\mathcal{X}_n = \sum_{i=1}^{\binom{n}{3}} X_{\tau_i}$$

We start with $r = 1$, for which we immediately infer $\mathbb{E}(\mathcal{X}_n)_1 \to \lambda$ by our observations above. For the second factorial moment, we obtain

$$\mathbb{E}[(\mathcal{X}_n)_2] = \mathbb{E}[\mathcal{X}_n^2] - \mathbb{E}\mathcal{X}_n = \sum_{i\neq j} \mathbb{E}[X_{\tau_i}X_{\tau_j}] + \sum_i \mathbb{E}[X_{\tau_i}^2] - \sum_i \mathbb{E}X_{\tau_i} = \sum_{i\neq j} \mathbb{E}[X_{\tau_i}X_{\tau_j}],$$

as $X_\tau^2 = X_\tau$. Similarly, for the third factorial moment we first have

$$\mathbb{E}(\mathcal{X}_n)_3 = \mathbb{E}\mathcal{X}_n^3 - 3\mathbb{E}\mathcal{X}_n^2 + 2\mathbb{E}\mathcal{X}_n = \mathbb{E}\mathcal{X}_n^3 - \mathbb{E}\mathcal{X}_n^2 - 2\underbrace{(\mathbb{E}\mathcal{X}_n^2 - \mathbb{E}\mathcal{X}_n)}_{=\sum_{i\neq j}\mathbb{E}[X_{\tau_i}X_{\tau_j}]}.$$

For the third moment, we infer

$$\mathbb{E}\mathcal{X}_n^3 = \sum_{\substack{i,j,k\\ i\neq j, i\neq k, k\neq j}} \mathbb{E}[X_{\tau_i}X_{\tau_j}X_{\tau_k}] + 3\sum_{i\neq j} \mathbb{E}[X_{\tau_i}X_{\tau_j}] + \sum_i \mathbb{E}X_{\tau_i}$$

and therefore

$$\begin{aligned}\mathbb{E}(\mathcal{X}_n)_3 &= \sum_{\substack{i,j,k\\ i\neq j, i\neq k, k\neq j}} \mathbb{E}[X_{\tau_i}X_{\tau_j}X_{\tau_k}] + \sum_{i\neq j} \mathbb{E}[X_{\tau_i}X_{\tau_j}] + \mathbb{E}\mathcal{X}_n - \mathbb{E}\mathcal{X}_n^2\\ &= \sum_{\substack{i,j,k\\ i\neq j, i\neq k, k\neq j}} \mathbb{E}[X_{\tau_i}X_{\tau_j}X_{\tau_k}].\end{aligned}$$

Proceeding as such, we inductively infer

$$\mathbb{E}(\mathcal{X}_n)_r = \sum_{i_1,\dots,i_r} \mathbb{E}\Big[\prod_{j=1}^r X_{\tau_j}\Big]$$

and we are left to show that

$$\sum_{i_1,\dots,i_r} \mathbb{E}\Big[\prod_{j=1}^r X_{\tau_j}\Big] \longrightarrow \lambda^r$$

for each $r \in \mathbb{N}$. For $r = 1$ this has already been done above. For $r = 2$, we have to distinguish two cases: Either $\tau_i \cap \tau_j = \emptyset$, in which case the occurrence of the two triangles is completely independent, or τ_i and τ_j share an edge, in which case the expectation is given by (1). Note that if the triangles share two edges, they must be the same, implying $i = j$, which is no part of the sum. The number of disjoint triangles is $\binom{n}{3}\binom{n-3}{3}$ and therefore

$$\sum_{\substack{i\neq j\\ \tau_i\cap\tau_j=\emptyset}} \mathbb{E}[X_{\tau_i}X_{\tau_j}] = \binom{n}{3}\binom{n-3}{3}\Big(\frac{c^3}{n^3}\Big)^2 \longrightarrow \Big(\frac{c^3}{6}\Big)^2 = \lambda^2.$$

As there are $\binom{n}{3}(n-3)$ many triangles that share an edge, we infer

$$\sum_{\substack{i\neq j\\ \tau_i\cap\tau_j\neq\emptyset}} \mathbb{E}[X_{\tau_i}X_{\tau_j}] = \binom{n}{3}(n-3)\frac{c^5}{n^5} \longrightarrow 0.$$

Together,

$$\sum_{i\neq j}\mathbb{E}[X_{\tau_i}X_{\tau_j}]\longrightarrow\lambda^2.$$

Here (and in the following), we used that $\binom{n-k}{3}=(n^3/6)+o(n^3)$ as $n\to\infty$ and k fixed. Finally, consider the case of general r. Again, we have to distinguish the two cases of disjoint triangles, in which case the expectation again factorises and the case, in which some of the triangles share an edge. Assuming n to be large enough, we have for the first case

$$\begin{aligned}\sum_{\substack{i_1,\dots,i_r\\ \text{disjoint}}}\mathbb{E}\Big[\prod_{j=1}^{r}X_{\tau_j}\Big]&=\binom{n}{3}\binom{n-3}{3}\dots\binom{n-3(r-1)}{3}\Big(\frac{c^3}{n^3}\Big)^r\\ &=\Big(\frac{c^3}{6}\Big)^r(n^{3r}+o(n^{3r}))n^{-3r}\longrightarrow\lambda^r.\end{aligned}$$

It therefore remains to show that all the summands left out tend to zero. Consider $\tau_{i_1},\dots,\tau_{i_r}$ that are not all pairwise disjoint, so that at least two of the involved triangles share an edge and let $m=|\tau_{i_1}\cup\dots\cup\tau_{i_r}|$ be the total number of distinct vertices among all the triangles. Let $t=t_{i-1,\dots,i_r}$ be the number of all distinct *edges* needed to draw all these triangles. Then always $t>m$. This is due to the fact, that no triangle is allowed to appear twice. Therefore, whenever two triangles are not disjoint they either share a vertex or an edge. In the first case, 6 distinct edges are needed to draw the two triangles (in fact they are independent). In the second case, two additional edges are needed to create the second triangle but only a single additional vertex. Let $\mathcal{V}_m=\{\tau=(\tau_{i_1},\dots,\tau_r)\colon|\tau|=m\}$ the set of all r many triangles on m vertices. Let $\underline{t}_m=\min\{t_\tau\colon\tau\in\mathcal{V}_m\}$ the minimal number of edges among all the triangle-collections in $\mathcal{V}_m$. By the above, $\underline{t}_m\geq m+1$. Additionally, $|\mathcal{V}_m|=C\binom{n}{m}$, where the constant C counts the number of possibilities to sort m vertices into triangles, which does not depend on n. Therefore, $|\mathcal{V}_m|=O(n^m)$, and thus

$$\sum_{\tau\in\mathcal{V}_m}p_n^{t_\tau}\leq O(n^m)O(n^{-m-1})=O(n^{-1}).$$

In order to be not completely distinct, the different triangles must share at least one vertex, in which case $m=3r-1$. Further, τ can clearly not contain less than 4 different vertices. Therefore,

$$\sum_{\substack{i_1,\dots,i_r\\ \text{not disjoint}}}\mathbb{E}\Big[\prod_{j=1}^{r}X_{\tau_j}\Big]\leq\sum_{m=4}^{3r-1}O(n^{-1})\leq 3rO(n^{-1})\to 0,$$

as r is fixed. This finishes the proof. □

Problem 2.2. Show that the space $\mathscr{G}$ of rooted graphs is *not* compact.

Proof. Recall that, in a metric space, compactness is equivalent to sequential compactness, meaning that each sequence has a convergent subsequent. For $n\in\mathbb{N}$, let

$\mathcal{S}_n \in \mathscr{G}$, the rooted star of size n. That is, $\mathcal{S}_n$ consists of a root that has n children. Consider the sequence $(\mathcal{S}_n : n \in \mathbb{N})$. As two stars, $\mathcal{S}_n$ and $\mathcal{S}_m$ differ in the number of children of the root, they differ at graph distance 1 from the root and only agree on the root. Therefore, $d(\mathcal{S}_n, \mathcal{S}_m) = 1$ for all $n \neq m$. Therefore, $(\mathcal{S}_n)$ cannot contain a convergent subsequence, as each convergent sequence is also a Cauchy sequence, which cannot be the case here. □

Problem 2.3 (Poisson Thinning)**.** Let N be a Poisson distribution with parameter $\lambda > 0$. Let X be a multinomial distribution with N trials and probabilities $p_1, \dots, p_m$, i.e., we perform N independent experiments with m potential outcomes and outcome j has probability p_j. Let N_j denote the number of outcomes j. Show that, $N_1, \dots, N_m$ are independently Poisson distributed with respective parameters $\lambda p_1, \dots, \lambda p_m$.

Proof. See Lemma 3.11 in the book [1]. □

Problem 2.4 (Uniform model and Erdős-Rényi)**.** Fix $m \in \mathbb{N}$ and consider the space $\mathscr{G}_{n,m} := \{\mathcal{G} : |\mathcal{V}(G)| = n, |\mathcal{E}(\mathcal{G})| = m\}$, the space of all graphs on n vertices and m edges. Consider the uniform measure on $\mathscr{G}_{n,m}$, i.e.,

$$\mathbb{P}_{n,m}(\mathcal{G}) := \frac{1}{\binom{\binom{n}{2}}{m}}, \qquad \text{for any } \mathcal{G} \in \mathscr{G}_{n,m}. \tag{2}$$

The corresponding random graph drawn with this distribution is denoted by $\mathcal{G}(n,m)$. Let $\mathcal{G}(n,p)$ the Erdős-Rényi random graph on n vertices, where each edge is present independently with probability $p \in (0,1)$. Show that $\mathcal{G}(n,p)$, conditioned on having exactly m edges, is distributed as $\mathcal{G}(n,m)$.

Proof.

$$\mathbb{P}(\mathcal{G}(n,p) | \mathcal{G}(n,p) \in \mathscr{G}_{n,m}) = \frac{\mathbb{P}(\mathcal{G}(n,p), |\mathcal{E}(\mathcal{G}(n,p))| = m)}{\mathbb{P}(\mathcal{G}(n,p) \in \mathscr{G}_{n,m})} \tag{3}$$

Note that $\mathbb{P}(\mathcal{G}(n,p) \in \mathscr{G}_{n,m}) = \sum_{\mathcal{G} \in \mathscr{G}_{n,m}} \mathbb{P}(\mathcal{G}(n,p) = \mathcal{G}) = |\mathscr{G}_{n,m}| p^m (1-p)^{\binom{n}{2} - m}$, therefore, from (3), we get,

$$\mathbb{P}(\mathcal{G}(n,p) | \mathcal{G}(n,p) \in \mathscr{G}_{n,m}) = \frac{|p^m (1-p)^{\binom{n}{2} - m}}{|\mathscr{G}_{n,m}| p^m (1-p)^{\binom{n}{2} - m}} = \frac{1}{|\mathscr{G}_{n,m}|} = \frac{1}{\binom{\binom{n}{2}}{m}},$$

which is exactly the law in (2). □

References

[1] Remco van der Hofstad. *Random Graphs and Complex Networks. Vol. 2*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2024.