POISSON LIMIT OF THE NUMBER OF MONOCHROMATIC CLIQUES

Somabha Mukherjee[1] and Bhaswar B. Bhattacharya[2]

University of Pennsylvania

July 5, 2017

[1]Department of Statistics , University of Pennsylvania
[2]Department of Statistics, University of Pennsylvania

Outline

Our Setup

- Each of the vertices $\{1, ..., n\}$ of a simple, undirected non-random graph G_n is colored independently of the others, using one of c_n $(\to \infty)$ different colors, chosen uniformly at random.
- We shall denote the color of vertex i by X_i.
- $T(K_m, G_n)$ $(m \text{ fixed})$: Number of monochromatic m-cliques in G_n .
- Assumptions:
  1. $\lim_{n\to\infty} \mathbb{E}T(K_m, G_n) = \lambda \in (0, \infty)$,
  2. $\lim_{n\to\infty} \mathbb{V}T(K_m, G_n) = \lambda$.
- We proved:

$$T(K_m, G_n) \xrightarrow{d} \text{Pois}(\lambda) \text{ as } n \to \infty.$$

- We could actually prove the same Poisson convergence result for $T(H, G_n)$: the number of copies of a fixed connected graph H on m vertices in G_n, under the same conditions on $\mathbb{E}T(H, G_n)$ and $\mathbb{V}T(H, G_n)$.

Existing Literature

- Bhattacharya, Diaconis and Mukherjee [1] showed that under the uniform coloring scheme, the number of monochromatic edges $T(K_2, G_n)$ of a random graph G_n (under any arbitrary probability distribution) converges weakly to Pois(λ), if we only assume that $\mathbb{E}(T(K_2, G_n)|G_n) \xrightarrow{d} \lambda$.
- Their result is more general, in the sense that they actually proved that if $\mathbb{E}(T(K_2, G_n)|G_n)$ converges weakly to a random variable Z, then $T(K_2, G_n) \xrightarrow{d} W$, where W is a Z-mixture of Poisson random variables, i.e.

$$\mathbb{P}(W = k) = \frac{1}{k!}\mathbb{E}(e^{-Z}Z^k) \ .$$

- Bhattacharya and Mukherjee [2] established the Poisson convergence result for the number of monochromatic triangles and 2-stars, under the same setup, but with the first two moment assumptions same as ours.
- Our work provides a complete answer to their first open problem, where they ask whether the same phenomenon extends to other connected monochromatic subgraphs.

Outline

What do the Two Assumptions Imply?

- $T(K_m, G_n) = \sum_{H \in \Lambda_m(G_n)} 1(H \text{ is monochromatic})$, where $\Lambda_m(G_n)$ is the set of all m-cliques in G_n .
- Let $N(K_m, G_n) =$ number of m-cliques in $G_n = |\Lambda_m(G_n)|$.
- $\mathbb{E}T(K_m, G_n) = \frac{N(K_m, G_n)}{c_n^{m-1}}$,

$$\mathbb{V}T(K_m, G_n) = \frac{1}{c_n^{m-1}}\left(1 - \frac{1}{c_n^{m-1}}\right) N(K_m, G_n) + \sum_{k=2}^{m-1} \frac{1}{c_n^{2m-k-1}}\left(1 - \frac{1}{c_n^{k-1}}\right) J_k ,$$

where

$$J_k = \Big|\{(F, H) : F, H \in \Lambda_m(G_n) \text{ and } |V(F) \cap V(H)| = k\}\Big|.$$

- Hence, $N(K_m, G_n) = \Theta(c_n^{m-1})$ and $J_k = o(c_n^{2m-k-1})$ $(2 \leq k \leq m-1)$.

A Decomposition of the Number of Monochromatic m-cliques

- For evey $2 \leq k \leq m-1$ and every tuple $(i_1, ..., i_k)$ of distinct vertices of G_n, let $\gamma_k(i_1, ..., i_k)$ be the number of m-cliques having $i_1, ..., i_k$ as vertices.
- For each $\epsilon > 0$, define:
  $A_{n,\epsilon} = \{H \in \Lambda_m(G_n) : \gamma_k(H_{i_1}, ..., H_{i_k}) \leq \epsilon c_n^{m-k} \ \forall\ 2 \leq k \leq m-1,\ \forall\ 1 \leq i_1 < ... < i_k \leq m\}$,
  where for a graph H with m vertices, the vertices are ordered as $H_1 < ... < H_m$.
- Let $T_{1,\epsilon}(K_m, G_n) = \sum_{H \in A_{n,\epsilon}} 1(H \text{ is monochromatic})$ and

$$T_{2,\epsilon}(K_m, G_n) = T(K_m, G_n) - T_{1,\epsilon}(K_m, G_n) \ .$$

- We will show that $T_{1,\epsilon}(K_m, G_n) \xrightarrow{d} \text{Pois}(\lambda)$ and $T_{2,\epsilon}(K_m, G_n) \xrightarrow{P} 0$ as $n \to \infty$ followed by $\epsilon \to 0$.

$T_{2,\epsilon}(K_m, G_n)$ converges to 0 in L^1

- To begin with, note that

$$T_{2,\epsilon}(K_m, G_n) \leq \sum_{i_1,\dots,i_m} \prod_{(k,l)\in\langle m\rangle} a_{i_k i_l}(G_n)\mathbf{1}\left(X_{i_1} = \dots = X_{i_m}\right) \sum_{q=2}^{m-1} \mathbf{1}\left(\gamma_q(i_1,\dots,i_q) > \epsilon c_n^{m-q}\right).$$

- Let us introduce the notation $\langle m\rangle = \{(a,b) \in \{1,2,\dots,m\}^2 : a < b\}$.
- Hence, for sufficiently large n (for example, when $\epsilon c_n > 2$), we have:

$$\begin{aligned}
&\mathbb{E}T_{2,\epsilon}(K_m, G_n) \\
&\leq \frac{1}{c_n^{m-1}} \sum_{i_1,\dots,i_m} \prod_{(k,l)\in\langle m\rangle} a_{i_k i_l}(G_n) \sum_{q=2}^{m-1} \mathbf{1}\left(\gamma_q(i_1,\dots,i_q) > \epsilon c_n^{m-q}\right) \\
&\leq \frac{1}{c_n^{m-1}} \sum_{q=2}^{m-1} \sum_{i_1,\dots,i_q} \prod_{(k,l)\in\langle q\rangle} a_{i_k i_l}(G_n) \frac{\gamma_q(i_1,\dots,i_q)}{\epsilon c_n^{m-q}} \mathbf{1}\left(\gamma_q(i_1,\dots,i_q) > \epsilon c_n^{m-q}\right) \\
&\quad \sum_{i_{q+1},\dots,i_m} \prod_{(k,l)\in\langle m\rangle\setminus\langle q\rangle} a_{i_k i_l}(G_n)
\end{aligned}$$

$$
\begin{aligned}
&\lesssim \sum_{q=2}^{m-1} \sum_{i_1,\ldots,i_q} \prod_{(k,l)\in\langle q\rangle} a_{i_k i_l}(G_n) \frac{[\gamma_q(i_1,\ldots,i_q)]^2}{\epsilon c_n^{2m-q-1}} \mathbf{1}\left(\gamma_q(i_1,\ldots,i_q) > \epsilon c_n^{m-q}\right) \\
&\lesssim \sum_{q=2}^{m-1} \left[\frac{1}{\epsilon c_n^{2m-q-1}} \sum_{i_1,\ldots,i_q} \prod_{(k,l)\in\langle q\rangle} a_{i_k i_l} \binom{\gamma_q(i_1,\ldots,i_q)}{2} \right] \\
&\lesssim \sum_{q=2}^{m-1} \left[\frac{1}{\epsilon c_n^{2m-q-1}} \sum_{t=q}^{m-1} J_t \right] \\
&= \sum_{q=2}^{m-1} \frac{o\left(c_n^{2m-q-1}\right)}{\epsilon c_n^{2m-q-1}} \\
&= o(1).
\end{aligned}
$$

Binomial Approximation of $T_{1,\epsilon}(K_m, G_n)$

- Let $\{Z_{i_1,\ldots,i_m} : i_1 < \ldots < i_m\}$ be a collection of i.i.d. Ber(c_n^{1-m}) random variables, and define

$$W_{m,\epsilon}(G_n) = \sum_{i_1<i_2<\ldots<i_m} \prod_{(k,l)\in\langle m\rangle} a_{i_k i_l}(G_n) Z_{i_1,\ldots,i_m}$$

$$\mathbf{1}\left(\gamma_q(i_1, \ldots, i_q) \le \epsilon c_n^{m-q}\ \forall\ 2 \le q \le m-1,\ \forall\ 1 \le i_1 < \ldots < i_q \le m\right)$$

- For any other permutation $(j_1, \ldots, j_m)$ of $(i_1, \ldots, i_m)$, set $Z_{j_1,\ldots,j_m} = Z_{i_1,\ldots,i_m}$.
- Our next target is to show that for every natural number r, the r^{th} moments of $T_{1,\epsilon}(K_m, G_n)$ and $W_{m,\epsilon}(G_n)$ are asymptotically close as $n \to \infty$ followed by $\epsilon \to 0$. But HOW DOES THIS HELP?
- Well, it is a simple method of moments argument! Observe that $W_{m,\epsilon}(G_n) \sim \text{Bin}\left(|A_{n,\epsilon}|, c_n^{1-m}\right)$. We know that $N(K_m, G_n)c_n^{1-m} \to \lambda$.
- Now, $\frac{N(K_m,G_n)-|A_{n,\epsilon}|}{c_n^{m-1}} = \mathbb{E}T_{2,\epsilon}(K_m, G_n) \to 0$. So, $|A_{n,\epsilon}|c_n^{1-m} \to \lambda$.
- Hence, $\mathbb{E}W_{m,\epsilon}(G_n)^r \to \mathbb{E}\text{Pois}(\lambda)^r$ for every natural number r.

Expressions for r^{th} Moments

- $\mathbb{E}T_{1,\epsilon}(K_m, G_n)^r = \sum_{H^{(1)} \in A_{n,\epsilon}} \dots \sum_{H^{(r)} \in A_{n,\epsilon}} \mathbb{E}\left(\prod_{i=1}^{r} \mathbf{1}(H^{(i)} \text{ is monochromatic}) \right).$
- $\mathbb{E}W_{m,\epsilon}(G_n)^r = \sum_{H^{(1)} \in A_{n,\epsilon}} \dots \sum_{H^{(r)} \in A_{n,\epsilon}} \mathbb{E}\left(\prod_{i=1}^{r} Z_{H_1^{(i)},\dots,H_m^{(i)}} \right).$
- Let $\Gamma_r = \{(H^{(1)}, \dots, H^{(r)}) : H^{(i)} \in A_{n,\epsilon} \ \forall\ 1 \le i \le r\}$.
- For each $A = (H^{(1)}, \dots, H^{(r)}) \in \Gamma_r$, let $H(A) = \bigcup_{i=1}^{r} H^{(i)}$ in the graph union sense, and let $a(A) = \left|\{H^{(1)}, \dots, H^{(r)}\}\right|$, i.e. the number of distinct graphs among $H^{(1)}, \dots, H^{(r)}$.
- Let $\nu(H)$ denotes the number of connected components of a graph H. In these notations, $\mathbb{E}T_{1,\epsilon}(K_m, G_n)^r = \sum_{A \in \Gamma_r} \left(\frac{1}{c_n}\right)^{\left|V(H(A))\right| - \nu(H(A))}$ and $\mathbb{E}W_{m,\epsilon}(G_n)^r = \sum_{A \in \Gamma_r} \left(\frac{1}{c_n}\right)^{(m-1)a(A)}.$

Closeness of the r^{th} Moments

- From the previous slide, we get that:

$$\left|\mathbb{E}T_{1,\epsilon}(K_m, G_n)^r - \mathbb{E}W_{m,\epsilon}(G_n)^r\right| \leq \sum_{A\in\Gamma_r}\left|\left(\frac{1}{c_n}\right)^{(m-1)a(A)} - \left(\frac{1}{c_n}\right)^{\left|V(H(A))\right|-\nu(H(A))}\right| .$$

- For any $A \in \Gamma_r$, it can be shown that:

$$(m-1)a(A) \geq \left|V\left(H(A)\right)\right| - \nu(H(A)) .$$

- We showed that the number of tuples $A \in \Gamma_r$ with $|V(H(A))| = v$ and $\nu(H(A)) = \nu$, for which the above inequality is strict, is $\epsilon O(c_n^{v-\nu})$.
- Since v and ν can take only a bounded number of values, the r^{th} moment difference is seen to be $\epsilon|O_n(1) - o_n(1)| = \epsilon O_n(1)$.
- This shows that $\mathbb{E}T_{1,\epsilon}(K_m, G_n)^r - \mathbb{E}W_{m,\epsilon}(G_n)^r$ converge to 0, as $n \to \infty$, followed by $\epsilon \to 0$, thereby completing the proof.
- The details are sketched in the next few slides.

Further Details

- For each $1 \le a \le r, m \le v \le mr$ and $1 \le \nu \le r$, define

$$\Gamma^r_{a,v,\nu} = \Big\{ A \in \Gamma_r : a(A) = a, \big|V(H(A))\big| = v \text{ and } \nu(H(A)) = \nu \Big\}. \quad (1)$$

- Then, we have:

$$\begin{aligned} &\Big| \mathbb{E} T_{1,\epsilon}(K_m, G_n)^r - \mathbb{E} W_{m,\epsilon}(G_n)^r \Big| \\ \le\ & \sum_{a=1}^{r} \sum_{v=m}^{mr} \sum_{\nu=1}^{r} \sum_{A \in \Gamma^r_{a,v,\nu}} \left| \left(\frac{1}{c_n}\right)^{(m-1)a} - \left(\frac{1}{c_n}\right)^{v-\nu} \right| \\ =\ & \sum_{a=1}^{r} \sum_{v=m}^{mr} \sum_{\nu=1}^{r} \left| \left(\frac{1}{c_n}\right)^{(m-1)a} - \left(\frac{1}{c_n}\right)^{v-\nu} \right| \left|\Gamma^r_{a,v,\nu}\right| \end{aligned}$$

- It thus suffices to show that for every fixed $1 \le a \le r, m \le v \le mr$ and $1 \le \nu \le r$,

$$\left| \left(\frac{1}{c_n}\right)^{(m-1)a} - \left(\frac{1}{c_n}\right)^{v-\nu} \right| \left|\Gamma^r_{a,v,\nu}\right| \to 0$$

as $n \to \infty$ followed by $\epsilon \to 0$.

A Useful Lemma

- **Lemma:** Let $A = (H^{(1)}, ..., H^{(s)}) \in \Gamma_s$ for some natural number s, and suppose that $H(A)$ is connected. Then, one of the following two always holds: (1) There exists an ordering $(G^{(1)}, ..., G^{(s)})$ of $(H^{(1)}, ..., H^{(s)})$ such that for each $2 \leq t \leq s$, either

$$\left| V(G^{(t)}) \cap \cup_{u=1}^{t-1} V(G^{(u)}) \right| = 1$$

or $G^{(t)}$ equals one of $G^{(1)}, ..., G^{(t-1)}$.
(2) There exists an ordering $(G^{(1)}, ..., G^{(s)})$ of $(H^{(1)}, ..., H^{(s)})$ such that

$$\left| V(G^{(t)}) \cap \cup_{u=1}^{t-1} V(G^{(u)}) \right| \geq 1 \ \forall \ 2 \leq t \leq s \text{ and}$$

$$2 \leq \left| V(G^{(t)}) \cap \cup_{u=1}^{t-1} V(G^{(u)}) \right| \leq m - 1 \text{ for some } 2 \leq t \leq s.$$

- Let $A = (G^{(1)}, ..., G^{(s)}) \in \Gamma_s$ & $\left|V(G^{(t)}) \cap \cup_{u=1}^{t-1} V(G^{(u)})\right| \geq 1 \; \forall \; 2 \leq t \leq s$. For each $k = 1, 2, ..., m-1$, define

$$s_k = \left|\left\{2 \leq t \leq s : \left|V(G^{(t)}) \cap \cup_{u=1}^{t-1} V(G^{(u)})\right| = m - k\right\}\right|.$$

Also, define

$$s_0 = \left|\left\{2 \leq t \leq s : \left|V(G^{(t)}) \cap \cup_{u=1}^{t-1} V(G^{(u)})\right| = m, G^{(t)} \notin \{G^{(1)}, ..., G^{(t-1)}\}\right\}\right|.$$

Now, we have:

$$\left|V(H(G))\right| = m + \sum_{k=1}^{m-1} k s_k \text{ and}$$

$$a(G) = 1 + \sum_{k=0}^{m-1} s_k.$$

Hence, we have:

$$\left|V(H(G))\right| \leq m + (m-1)\sum_{k=1}^{m-1} s_k \leq (m-1)a(G) + 1,$$

with equality holding if and only if $s_0 = s_1 = ... = s_{m-2} = 0$.

- If $\Gamma^r_{a,v,\nu}$ is empty, or contains an A with the property that each of the connected components $(H(A))_1, \ldots, (H(A))_\nu$ of $H(A)$, expressed as tuples $A_1,\ldots,A_\nu$, satisfies case (1) of the lemma, then

$$\left| \left(\frac{1}{c_n}\right)^{(m-1)a} - \left(\frac{1}{c_n}\right)^{v-\nu} \right| \left|\Gamma^r_{a,v,\nu}\right| = 0.$$

- So, suppose that for every element A of $\Gamma^r_{a,v,\nu}$, there exists $1 \leq i \leq \nu$, such that A_i satisfies case (2) of the lemma.
- For each A_i $(1 \leq i \leq \nu)$, denote the quantities $s_0, \ldots, s_{m-1}$ for A_i as $s^i_0, \ldots, s^i_{m-1}$, respectively.
- So, for a fixed array of quantities $(s^i_j)_{0 \leq j \leq m-1, 1 \leq i \leq \nu}$, the number of elements of $\Gamma^r_{a,v,\nu}$ corresponding to these array, is $\leq$ (upto constant multiples):

$$\begin{aligned}
&\prod_{k=1}^{\nu} N(K_m, G_n)^{1+s^k_{m-1}} \prod_{u=1}^{m-2} (\epsilon c_n^u)^{s^k_u} \\
\lesssim \quad & \epsilon^{\sum_{k=1}^{\nu} \sum_{u=1}^{m-2} s^k_u} c_n^{(m-1)(\nu+\sum_{k=1}^{\nu} s^k_{m-1})} c_n^{\sum_{k=1}^{\nu} \sum_{u=1}^{m-2} u s^k_u} \\
\leq \quad & \epsilon c_n^{\sum_{k=1}^{\nu} \sum_{u=1}^{m-1} u s^k_u + (m-1)\nu} \\
= \quad & \epsilon c_n^{v-\nu} \ .
\end{aligned}$$

Completing the Calculation for $T_{1,\epsilon}$

- Since the array $(s_j^i)_{0\le j\le m-1, 1\le i\le \nu}$ is constrained within the **finite** set $\{0, ..., r\}^{m\times\nu}$, we conclude that

$$\left|\Gamma^r_{a,v,\nu}\right| \le rm\nu\epsilon c_n^{v-\nu} \lesssim \epsilon c_n^{v-\nu} \ .$$

- Hence, $\left|\left(\frac{1}{c_n}\right)^{(m-1)a} - \left(\frac{1}{c_n}\right)^{v-\nu}\right| \left|\Gamma^r_{a,v,\nu}\right| \lesssim \epsilon \left|1 - c_n^{v-\nu-(m-1)a}\right| .$
- Clearly, the right hand side of the last inequality goes to 0 as $n \to \infty$ followed by $\epsilon \to 0$.
- This completes the entire proof!

Outline

Concluding Remarks

- The truncation we performed on the number of m-cliques supported by every tuple of distinct vertices of an m-clique in G_n , is crucial for the closeness of the moments of the main term $T_{1,\epsilon}(K_m, G_n)$ and the corresponding Binomial variable $W_{m,\epsilon}(G_n)$, and at the same time, ensures that the remainder term $T_{2,\epsilon}(K_m, G_n)$ is $o_{\mathbb{P}}(1)$.
- The proof for the Poisson limit of the number of monochromatic copies of an arbitrary fixed, connected graph is almost similar to the proof for cliques, barring a few technicalities.
- For example, in the general case, $W_{m,\epsilon}(G_n) = \sum_{F \in A_{n,\epsilon}} Z_{F_1,\ldots,F_n}$ may not have a Binomial distribution, because of the possibility of the existence of more than one copy of the graph H with the same vertex set. This hampers independence of the summands.
- We dealt with this issue by splitting the above sum into a main term consisting of those copies of H whose vertex sets do not support any other copy, and a remainder term consisting of those copies of H whose vertex sets support at least another copy. We then showed that the remainer term converges to 0 in L^r for every natural number r.

Our Most General Result

- The following is the most general result proved by us:
  General Result: Let H_0 be a fixed, connected graph on m vertices and for each $1 \leq k \leq 2^{\binom{m}{2}}$, define:

  $$R_k = \left|\{S \subseteq V(G_n) : |S| = m \text{ and } G_n[S] \text{ contains exactly } k \text{ copies of } H_0\}\right| .$$

  Also, for $2 \leq k \leq m-1$, define J_k as the number of all ordered pairs of copies of H_0 in G_n, that have k vertices in common. Assume that the following two conditions hold:

  1. $\frac{R_k}{c_n^{m-1}} \to \lambda_k \ (\geq 0)$ as $n \to \infty, \ \forall \ 1 \leq k \leq 2^{\binom{m}{2}}$,
  2. $J_k = o(c_n^{2m-k-1}) \ \forall \ 2 \leq k \leq m-1$.

  Then,

  $$T(H_0, G_n) \xrightarrow{d} \sum_{k=1}^{2^{\binom{m}{2}}} k \ Pois(\lambda_k) \ ,$$

  where the Poisson random variables in the limit are all independent.

Outline

References

Bhattacharya, B. B., Diaconis, P., Mukherjee, S., *Universal Poisson and Normal limit theorems in graph coloring problems with connections to extremal combinatorics,* Annals of Applied Probability, Vol. 27 (1), 337-394, 2017.

Bhattacharya, B. B., Mukherjee, S., *Limit theorems for Monochromatic 2-stars and triangles,* arXiv version, 2017.