> *Since there are some problems about mathjax, I also upload the pdf version [/Network / 5 Poisson Approximation of Local Structure].*

Poisson Approximation

In this part, we want to approximate the number of subgraphs like edge, triangle with Poisson distribution. This approach is called Stein-Chen method.

Total Variation Distance

Total variation distance measures the similarity between two distributions. It is defined as

$$d_{TV}(P,Q) = \sup_{A \subset \mathbb{Z}^+} |P(A) - Q(A)|.$$

We use $\mathbb{Z}^+$ because Poisson distribution is discrete. For continuous distribution, we use $A \subset \mathbb{R}$.

A Stein characterisation

An equivalent expression of Poisson

$Z \sim Po(\lambda)$ if and only if for any bounded $g : \mathbb{N}_0 \to \mathbb{R}$

$$\mathbb{E}\{\lambda g(Z+1) - Zg(Z)\} = 0$$

Stein Equation

Stein Equation is

$$\lambda g(j+1) - jg(j) = I(j \in A) - \mathrm{Po}\{\lambda\}(A).$$

Note that the right side becomes total variation after taking expectation and maximum. The solution of g is

$$\begin{aligned} g(j+1) &= \frac{j!}{\lambda^{j+1}} e^{\lambda} \sum_{k=0}^{j} e^{-\lambda} \frac{\lambda^k}{k!} [I(k \in A) - \mathrm{Po}\{\lambda\}(A)] \\ &= -\frac{j!}{\lambda^{j+1}} e^{\lambda} \sum_{k=j+1}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} [I(k \in A) - \mathrm{Po}\{\lambda\}(A)] \end{aligned}$$

and it satisfies $\sup_j |g(j)| \leq \min\left(1, \lambda^{-\frac{1}{2}}\right)$ and
$\Delta g := \sup_j |g(j+1) - g(j)| \leq \min\left(1, \lambda^{-1}\right)$.

Stein-Chen Equation

Attributing to the equivalent expression of Poisson, we could measure the distribution of W with Poisson distribution with

$$d_{TV}(\mathcal{L}(W), \mathrm{Po}\{\lambda\}) = \sup_g |\mathbb{E}\{\lambda g(W+1) - Wg(W)\}|.$$

Application

Sum of iid Binomial distribution

Let $X_1, X_2, \ldots$ be independent $\mathrm{Be}(p_i)$, and $W = \sum_{i=1}^n X_i$. With
$\lambda = \sum_{i=1}^n p_i = \mathbb{E}(W)$,

$$\begin{aligned} d_{TV}(\mathcal{L}(W), \mathrm{Po}\{\lambda\}) &= \sup_g |\mathbb{E}\{\lambda g(W+1) - Wg(W)\}| \\ &\leq \sup_g \Delta g \sum_{i=1}^n p_i^2 \\ &\leq \min\left(1, \lambda^{-1}\right) \sum_{i=1}^n p_i^2 \end{aligned}$$

The approximation of the edge distribution in ER model is a trivial extension of this type.

Local Approach (With Local Dependency)

When there is dependency, we generalize to local approach. Let I be an index set and let $X_\alpha \sim Be\left(p_\alpha\right), \alpha \in I$, and $W = \sum_{\alpha \in I} X_\alpha$. Let $\lambda = \sum_{\alpha \in I} p_\alpha$. Suppose that for each $\alpha \in I$ there exists a set $A_\alpha \subset I$ s.t. $X_\alpha \perp \sum_{\beta \notin A_\alpha} X_\beta$. Define

$$\eta_\alpha = \sum_{\beta \in A_\alpha} X_\beta,$$

then with $\lambda = \mathbb{E}(W)$,

$$d_{TV}(\mathcal{L}(W), Po(\lambda)) \leq \sum_{\alpha \in I} \left[\left(p_\alpha \mathbb{E}\left(\eta_\alpha\right) + \mathbb{E}\left(X_\alpha\left(\eta_\alpha - X_\alpha\right)\right)\right] \min\left(1, \lambda^{-1}\right).$$

$$\begin{aligned} & d_{TV}(\mathcal{L}(W), Po(\lambda)) \\ & \quad \leqslant \textstyle\sum_{\alpha \in I} \left[\left(p_\alpha \mathbb{E}\left(\eta_\alpha\right) + \mathbb{E}\left(X_\alpha\left(\eta_\alpha - X_\alpha\right)\right)\right] \min\left(1, \lambda^{-1}\right) \end{aligned}$$

Triangles in ER model

The existence of triangles in ER model are dependent when there are common edge between them, so we use the local approach to bound the total variation distance. Consider ER model with n vertexes and the probability of edge existence is p, then with $\lambda = \binom{n}{3} p^3$,

$$d_{TV}(\mathcal{L}(W), Po(\lambda)) \leqslant \binom{n}{3} p^3 \left(3np^3 + 3np^2\right) \min\left(1, \lambda^{-1}\right)$$

Note that λ is exactly $\mathbb{E}(W)$.