> *Since there are some problems about mathjax, I also upload the pdf version [/Network / 6 Normal Approximation of Local Structure].*

Normal Approximation

In this part, we want to approximate the number of triangle with normal distribution in the sense of weak convergence. This approach is called Stein 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 Normal

$W \sim \mathcal{N}(0,1)$ if and only if for all continuous and piecewise continuously differentiable functions $f : \mathbb{R} \to \mathbb{R}$ with $\mathbb{E}\,|f'(W)| < \infty$

$$\mathbb{E}[Wf(W)] = \mathbb{E}f'(W).$$

Stein Equation

Stein Equation for normal distribution (for h) is

$$f'(w) - wf(w) = h(w) - Nh.$$

where $Nh = \mathbb{E}h(Z)$

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

$$\begin{aligned} f_h(w) &= e^{\frac{w^2}{2}} \int_{-\infty}^{w} (h(x) - Nh)e^{-x^2/2}dx \\ &= -e^{\frac{w^2}{2}} \int_{w}^{\infty} (h(x) - Nh)e^{-x^2/2}dx. \end{aligned}$$

and it satisfies

$$\|f_h\| \leq 2\|h'\|; \quad \|f_h'\| \leq \sqrt{\frac{2}{\pi}}\|h'\|; \quad \|f_h''\| \leq 2\|h'\|.$$

Stein Summary

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

$$\sup_{h\in\mathcal{H}} |\mathbb{E}h(W) - Nh| = \sup_{f_h} |\mathbb{E}f_h'(W) - \mathbb{E}[Wf_h(W)]|$$

Application

Sum of iid Binomial distribution

Let $X_1, X_2, \ldots$ be independent with mean zero and same variance $\text{Var}(X_i) = \frac{1}{n}$; denote $W = \sum_{i=1}^n X_i$. For any continuous and piecewise continuously differentiable function $h : \mathbb{R} \to \mathbb{R}$ we have

$$|\mathbb{E}h(W) - Nh| \leqslant \|h'\| \left(\frac{2}{\sqrt{n}} + \sum_{i=1}^n \mathbb{E}\left|X_i^3\right|\right)$$

Local Dependence

dissociated decomposition

Denote $X_i, i \in I$ where I be a finite index set. Denote $W = \sum_{i\in I} X_i$ and suppose $\forall i \in I$ there is a set $K_i \subset I$ with

$$W = W_i + Z_i \text{ with } Z_i = \sum_{k\in K_i} X_k, \quad i \in I$$

s.t. $W_i \perp X_i$. Further assume that

$$W_i = W_{ik} + V_{ik} \text{ with } V_{ik} = \sum_{j\in K_k\setminus K_i} X_j$$

for $i \in I, k \in K_i$ so that W_{ik} is independent of the pair (X_i, X_k).

Bound for dissociated decomposition

With the notation above, assume that

- $X_i, i \in I$ are mean zero and that $\text{Var}(W) = 1$;
- $k \in K_i \iff i \in K_k$.

Then, for any continuous and piecewise continuously differentiable function h,

$$
\begin{aligned}
&|\mathbb{E}h(W)-Nh| \\
&\leqslant \|4h'\| \sum_{i\in I}\sum_{j,k\in K_i}\left(\mathbb{E}\left(|X_iX_jX_k|+\mathbb{E}\left(|X_iX_j|\right)\mathbb{E}\left(|X_k|\right)\right)\right)
\end{aligned}
$$

Normal Approximation for Triangles in ER model

Denote W denote as standardised number of triangles in ER model with $p \leq \frac{1}{2}$ and $n \geq 3$. Then, for any continuous and piecewise continuously differentiable h,

$$
|\mathbb{E}h(W)-Nh| \leqslant \frac{4}{3} \times \frac{29n^5p^3}{\sigma^3}\|h'\|
$$

When p does not depend on n then this expression is of order $O\left(n^{-1}\right)$.

Note: Standardised version of $Y_\alpha = a_{u,v}a_{v,w}a_{u,w}$ (Triangle with vertexes (u,v,w)) is

$$
X_\alpha = \frac{Y_\alpha - p^3}{\sqrt{\operatorname{Var}(T)}}
$$