

Local Dependence

Can we have a CLT with dependent random variables?

E.g. Let $\underline{X}^1, \underline{X}^2, \dots, \underline{X}^n, \dots \in \mathbb{R}^d$ be L^2 iid. random vectors

Then consider the sequence with $E[\underline{X}^n] = 0$.

$\underline{X}_1^1, \underline{X}_2^1, \dots, \underline{X}_d^1, \underline{X}_1^2, \underline{X}_2^2, \dots, \underline{X}_d^2, \dots, \underline{X}_1^n, \underline{X}_2^n, \dots, \underline{X}_d^n, \dots$

In this case, $S_{nd} = \underline{X}_1^1 + \underline{X}_2^1 + \dots + \underline{X}_d^n$

E.g. Let $\{X_n\}_{n=0}^\infty$ be iid. Let $Y_n = X_n X_{n-1}$ for $n \geq 1$.

The Y_n, Y_m are independent if $|n-m| \geq 2$... but doesn't fit \rightarrow this mold.

Def. Let $\{X_j\}_{j \in V}$ be a collection of random variables.

The dependency graph for $\{X_j\}_{j \in V}$ is the (loop free) graph (V, E) with vertex set V , where $\{i, j\} \in E$ iff X_i, X_j are not independent.

Eg. If $\{X_n\}_{n \in \mathbb{N}}$ are independent, the dependency graph is trivial $E = \emptyset$.

Eg. In the example $X_1^1, X_2^1, \dots, X_d^1, X_1^2, X_2^2, \dots, X_d^2, \dots$

Eg. If $\{X_n\}_{n=0}^{\infty}$ are iid and $Y_n = X_n X_{n-1}$ for $n \geq 1$,



The degree of a vertex $i \in V$ is the number of edges in E adjacent to i . The degree $\deg G$ of a (finite) graph is the max degree of any vertex.

Lemma: Let $\{X_j\}_{j \in V}$ be independent L^2 random variables.

Let (V, E) be the dependency graph of $\{X_j\}_{j \in V}$.

Let $S = \sum_{j \in V} X_j$. Then

$$\text{Var}[S] \leq \sum_{i \in V} (1 + \text{deg}(i)) \text{Var}[X_i]$$

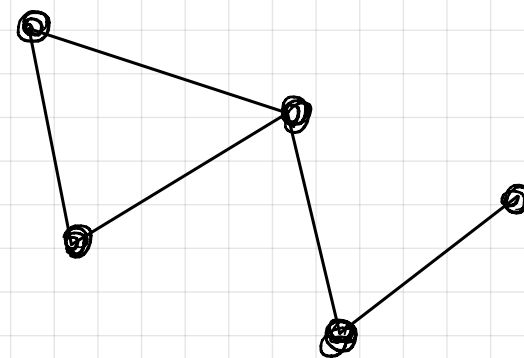
Pf. $\text{Var}[S] = \mathbb{E}\left[\left(\sum_{j \in V} X_j\right)^2\right] = \sum_{i, j \in V} \mathbb{E}[X_i X_j] = \sum_{i \in V} \mathbb{E}[X_i^2] + \sum_{i \in V} \sum_{j \neq i} \mathbb{E}[X_i X_j]$

$$= \sum_{i \in V} \text{Var}[X_i] + \sum_{i \in V} \sum_{j \sim i} \mathbb{E}[X_i X_j]$$

$$\leq \sum_{i \in V} \text{Var}[X_i] + \frac{1}{2} \sum_{i \in V} \sum_{j \sim i} (\text{Var}[X_i] + \text{Var}[X_j])$$

$$\frac{1}{2} \sum_{i \in V} \sum_{j \sim i} (\text{Var}[X_i] + \text{Var}[X_j])$$

$$= \frac{1}{2} \sum_{i \in V} \text{deg}(i) \text{Var}[X_i] + \frac{1}{2} \sum_{i \in V} \sum_{j \sim i} \text{Var}[X_j]$$



Theorem: Let $\{X_j\}_{j \in V}$ be L^4 random variables with $\mathbb{E}[X_j] = 0$. Let G be the dependency graph. Set $\sigma^2 = \text{Var}[\sum_{j \in V} X_j] \leq (1 + \text{deg}G) \sum_{j \in V} \text{Var}[X_j]$.


If $W = \frac{1}{\sigma} \sum_{j \in V} X_j$, and $Z \stackrel{d}{=} \mathcal{N}(0, 1)$, then

$$d_{W_1}(W, Z) \leq \frac{6}{\sqrt{\pi} \sigma^2} \sqrt{(1 + \text{deg}G)^3 \sum_{j \in V} \mathbb{E}[X_j^4]} + \frac{(1 + \text{deg}G)^2}{\sigma^3} \sum_{j \in V} \mathbb{E}[|X_j|^3]$$

The proof is similar (but a bit more fiddly) to our proof of the (Wasserstein) Berry-Esseen theorem, using Stein's method. (See [Ross, §3.2] for details.)

Eg. $\{X_j\}_{j=0}^{\infty}$ iid L^4 random variables, with $\mathbb{E}[X_j] = 0$, $\mathbb{E}[X_j^2] = 1$.

Let $Y_j = X_j X_{j-1}$ for $j \geq 1$.

Dependency graph of $\{Y_j\}_{j=1}^n$ is 

Note: $(X_{j-1}, X_j) \stackrel{d}{=} (X_j, X_{j+1}) \quad \therefore f(X_{j-1}, X_j) \stackrel{d}{=} f(X_j, X_{j+1}) \quad \forall \text{ measurable } f$
 \uparrow
 $f(x, y) = x \cdot y \Rightarrow \{Y_j\}_{j=1}^{\infty}$ are identically distributed

Note: if $i \neq j$, $\text{Cov}(Y_i, Y_j) = \mathbb{E}[Y_i Y_j]$

\therefore with $W = \frac{1}{\sqrt{n}}(Y_1 + \dots + Y_n)$,

$$d_{W_1}(W, Z) \leq \frac{6}{\sqrt{\pi}n} \sqrt{(3)^3 \sum_{j=1}^n \mathbb{E}[Y_j^4]} + \frac{3^2}{n^{3/2}} \sum_{j=1}^n \mathbb{E}[|Y_j|^3]$$

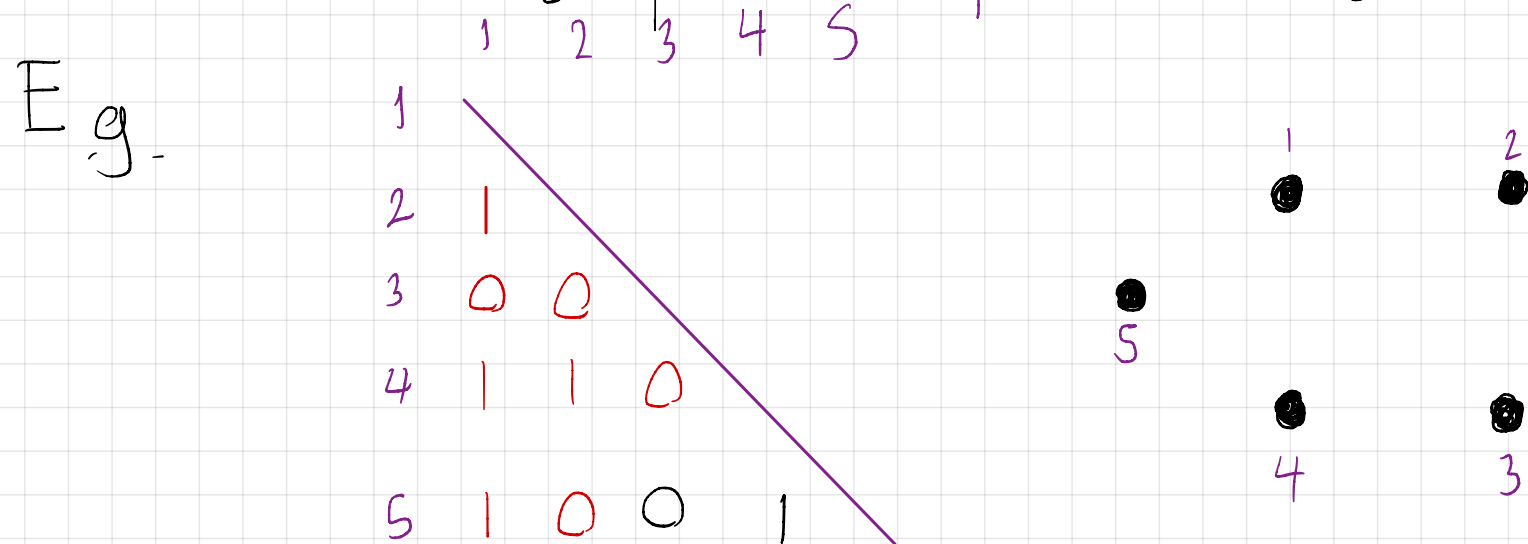
Now for a less contrived example...

Erdős-Renyi Random Graphs

$$V = \{1, 2, \dots, n\}$$

Let $p \in [0, 1]$. Let $\{X_{ij}\}_{1 \leq i < j \leq n}$ be iid $\text{Bern}(p)$.

The random graph $G(n, p)$ has edge $\{i, j\}$ iff $X_{ij} = 1$.



Some statistics about these random graphs are easy to compute:

E.g. # Edges

E.g. $\deg(i)$

We can allow $p = p_n$ to change with n , and study $G(n, p_n)$ in different scaling limits.

E.g. If $p_n \sim \frac{\lambda}{n}$, $\deg(i)$

Harder Problem:

Let $T_{n,p} = \# \text{triangles in } G(n,p)$.

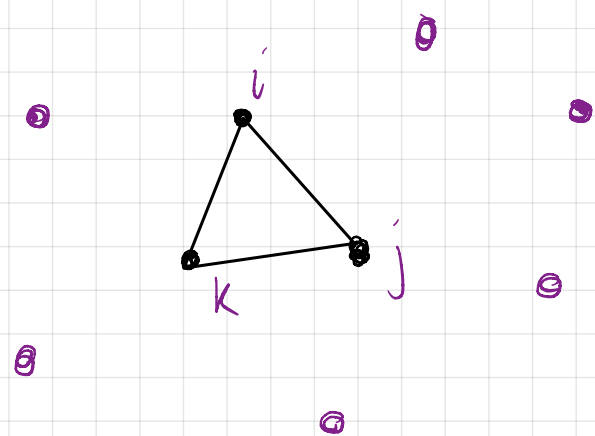
↓
A triangle is a triple $\{i,j,k\}$ wlog $i < j < k$
s.t. $\{i,j\}, \{j,k\}, \{i,k\}$ are all edges in $G(n,p)$

$$\text{Thus, } T_{n,p} = \sum_{1 \leq i < j < k \leq n} X_{ij} X_{jk} X_{ik}$$

$$\text{E.g. } \mathbb{E}[T_{n,p}] = \sum_{i < j < k} \mathbb{E}[X_{ij} X_{jk} X_{ik}]$$

Dependency graph: $V = \{(i,j,k) : 1 \leq i < j < k \leq n\}$

$X_{i_1 j_1} X_{j_1 k_1} X_{i_1 k_1}$ and $X_{i_2 j_2} X_{j_2 k_2} X_{i_2 k_2}$ are independent iff $\{(i_1, j_1), (j_1, k_1), (i_1, k_1)\} \cap \{(i_2, j_2), (j_2, k_2), (i_2, k_2)\} = \emptyset$



\therefore in the dependency graph, $\text{deg}(i,j,k) = 0$

A fun calculation shows that

$$\text{Var}[T_{n,p}] = \binom{n}{3} p^3 (1-p^3 + 3(n-3)p^2(1-p))$$

Putting this all together, Stein's method shows that:

with $W_{n,p} = \frac{T_{n,p} - \mathbb{E}[T_{n,p}]}{\sqrt{\text{Var}[T_{n,p}]}}$ and $Z \stackrel{d}{=} N(0,1)$,

$$d_{W_1}(W_{n,p}, Z) \leq \frac{\text{Const.}}{n p^{9/2}}$$

Thus, if $p_n \sim n^{-\alpha}$ for $\alpha < \frac{2}{9}$, then the # of triangles in $G(n, p_n)$ is asymptotically normal.

See [Ross, §3.2] for details.

This is not the optimal result. A more refined asymptotic analysis shows $W_{n,p_n} \rightarrow_w N(0,1)$ iff $np_n \rightarrow \infty$. See [Ruciński, PTRF 1988].