Local Dependence

Can we have a CLT with dependent random variables?

Eg. Let $\underline{X}^1, \underline{X}^2, \ldots, \underline{X}^n, \ldots \in \mathbb{R}^d$ be L^2 i.i.d. random vectors with $\mathbb{E}[\underline{X}^n] = 0$.

Then consider the sequence

$$\underbrace{X^1_1, X^1_2, \ldots, X^1_d}_{\text{dependent}}, \underbrace{X^2_1, X^2_2, \ldots X^2_d}_{\text{dependent}}, \ldots, \underbrace{X^n_1, X^n_2, \ldots, X^n_d}_{\text{dependent}}, \ldots$$

independent "neighborhoods"

In this case, $S_{nd} = X^1_1 + X^1_2 + \cdots + X^n_d = Y_1 + \cdots + Y_n$

where $Y_j := X^j_1 + \cdots + X^j_d$

$\{Y_j\}_{j=1}^\infty$ are independent, id.

$\sigma^2 = \mathbb{E}[Y_1^2] = \mathbb{E}[\|\underline{X}^1\|^2]$.

$0 \le k < d$

$$\frac{S_{nd+k}}{\sqrt{nd+k}} - \frac{S_{nd}}{\sqrt{nd}} \to_p 0.$$

$$\frac{S_{nd}}{\sqrt{nd}} = \frac{1}{\sqrt{d}} \frac{Y_1 + \cdots + Y_n}{\sqrt{n}} \to_w \mathcal{N}\left(0, \frac{\sigma^2}{d}\right)$$

Eg. Let $\{X_n\}_{n=0}^\infty$ be i.i.d. Let $Y_n = X_n X_{n-1}$ for $n \ge 1$.

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

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.

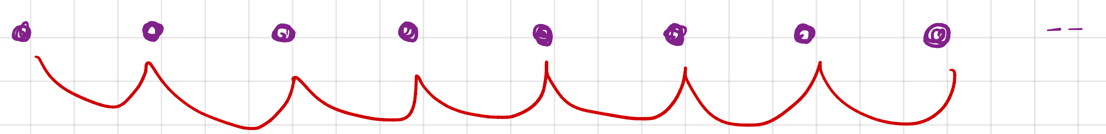
Eg. If $\{X_n\}_{n\in\mathbb{N}}$ are independent, the dependency graph is trivial $E = \phi$.

Eg. In the example $X^1_1, X^1_2, \dots, X^1_d; X^2_1, X^2_2, \dots, X^2_d, \dots$ disjoint union of d-vertex graphs.

$\deg G \le d-1 \rightarrow$

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

$\deg G = 2 \rightarrow$



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

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

Pf. $Var[S] = \mathbb{E}\Big[\Big(\sum_{j\in V}\mathring{X}_j\Big)^2\Big] = \sum_{i,j\in V}\mathbb{E}[\mathring{X}_i\mathring{X}_j] = \sum_{i\in V}\mathbb{E}[\mathring{X}_i^2] + \sum_{i\in V}\sum_{j\neq i}\mathbb{E}[\mathring{X}_i\mathring{X}_j]$

if $\{i,j\}\notin E$ $\mathring{X}_i, \mathring{X}_j$ are indep.

$\mathbb{E}[\mathring{X}_i\mathring{X}_j] = \mathbb{E}[\mathring{X}_i]\mathbb{E}[\mathring{X}_j] = 0.$

$$= \sum_{i\in V} Var[X_i] + \sum_{i\in V}\sum_{j\sim i}\mathbb{E}[\mathring{X}_i\mathring{X}_j] \qquad j\sim i \Leftrightarrow \{i,j\}\in E.$$

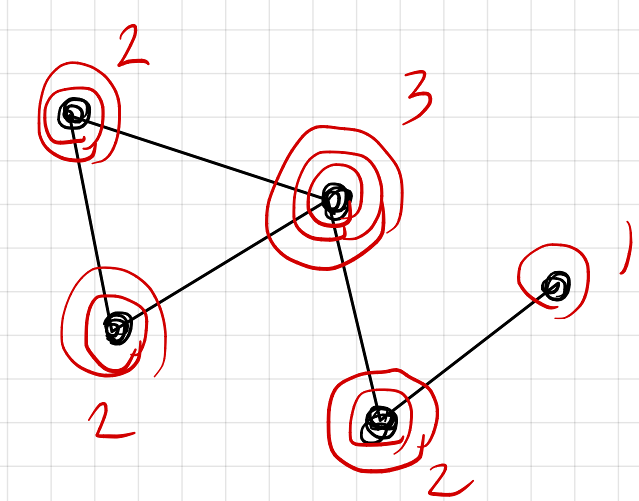
$ab \leq \frac{1}{2}(a^2+b^2)$

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

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

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

$$= \frac{1}{2}\sum_{j\in V}\underset{\deg(j)}{\big(\ \uparrow\ \big)}Var[X_j]$$

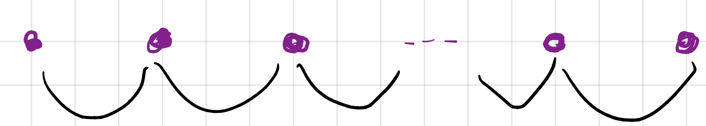


///

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 $\sigma^2 = \mathrm{Var}\left[\sum_{j\in V} X_j\right] \le (1+\deg G)\sum_{j\in V}\mathrm{Var}[X_j]$.
If $W = \frac{1}{\sigma}\sum_{j\in V} X_j$, and $Z \overset{d}{=} \mathcal{N}(0,1)$, then

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

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 $j\ge 1$. $\mathbb{E}[Y_j^k] = \mathbb{E}[X_{j-1}^k X_j^k] = \mathbb{E}[X_{j-1}^k]\,\mathbb{E}[X_j^k] = \mathbb{E}[X_1^k]^2$ $(k\le 4)$
Dependency graph of $\{Y_j\}_{j=1}^{n}$ is  $\deg = 2$

Note: $(X_{j-1},X_j) \overset{d}{=} (X_j,X_{j+1})$ $\therefore$ $f(X_{j-1},X_j) \overset{d}{=} f(X_j,X_{j+1})$ $\forall$ measurable f
$= \mu_{X_1}\otimes\mu_{X_1}$ $\uparrow$ $f(x,y)=x\cdot y \Rightarrow \{Y_j\}_{j=1}^{\infty}$ are identically distributed

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

$i < j$

$\Downarrow$

$i-1 < i \leq j-1 < j$

$$= \mathbb{E}[X_{i-1} X_i X_{j-1} X_j]$$

$$= \mathbb{E}[\underbrace{\cancel{X_{i-1}}}_{0}]\,\mathbb{E}[X_i X_{j-1}]\,\mathbb{E}[\underbrace{\cancel{X_j}}_{0}] = 0.$$

$$\therefore \mathrm{Var}[Y_1 + \cdots + Y_n] = \sum_{j=1}^{n} \mathrm{Var}[Y_j] = n.$$

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

$$d_{W_1}(W, Z) \leq \frac{6}{\sqrt{\pi}\, n} \sqrt{(3)^3 \underbrace{\sum_{j=1}^{n} \mathbb{E}[Y_j^4]}_{\substack{n\,\mathbb{E}[Y_1^4] \\ = n\,\mathbb{E}[X_1^4]^2}}} + \frac{3^2}{n^{3/2}} \underbrace{\sum_{j=1}^{n} \mathbb{E}[|Y_j|^3]}_{\substack{n\,\mathbb{E}[|Y_1|^3] \\ = n\,\mathbb{E}[|X_1|^3]^2}} = \frac{\frac{18\sqrt{3}}{\sqrt{\pi}}\mathbb{E}[X_1^4] + 9\,\mathbb{E}[|X_1|^3]^2}{\sqrt{n}}.$$

New for a less contrived example...

Erdős-Renyi Random Graphs

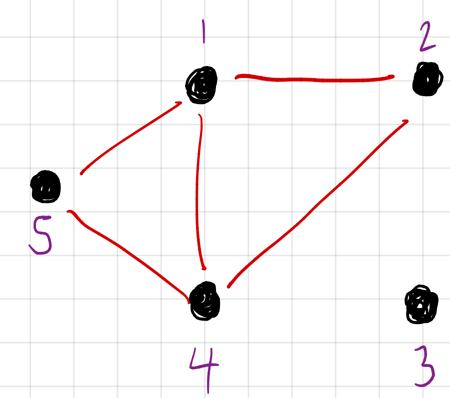
$V = \{1, 2, \ldots, n\}$

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

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

E.g.

$$\begin{array}{c|ccccc} & 1 & 2 & 3 & 4 & 5 \\ \hline 1 & & & & & \\ 2 & 1 & & & & \\ 3 & 0 & 0 & & & \\ 4 & 1 & 1 & 0 & & \\ 5 & 1 & 0 & 0 & 1 & \end{array}$$



Some statistics about these random graphs are easy to compute:

E.g. $\#\text{Edges} = \sum_{i<j} X_{ij} \overset{d}{=} \mathrm{Binom}\left(\binom{n}{2}, p\right)$ $\quad \mathbb{E}[\#\text{Edges}] = p\binom{n}{2}$

E.g. $\deg(i) = \sum_{j \sim i} X_{ij} \overset{d}{=} \mathrm{Binom}(n-1, p)$ $\quad \mathbb{E}[\deg(i)] = p(n-1)$

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

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

Harder Problem:

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

$\downarrow$

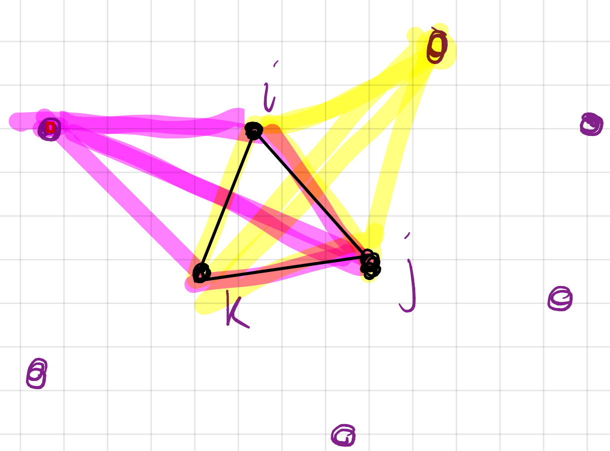
A triangle is a triple $\{i,j,k\}$ wlog $i<j<k$
s.t. $\{i,j\}$, $\{j,k\}$, $\{i,k\}$ are all edges in $G(n,p)$

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

Eg. $E[T_{n,p}] = \sum_{i<j<k} E[X_{ij} X_{jk} X_{ik}] = p^3 \, \#\{(i,j,k) : 1\le i<j<k\le n\} = p^3 \binom{n}{3}$.

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

$X_{i_1 j_1} X_{j_1 k_1} X_{i_1 k_1}$ 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, $\deg(i,j,k) = 3(n-3)$.

A fun calculation shows that

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

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

with $W_{n,p} = \dfrac{T_{n,p} - \mathbb{E}[T_{n,p}]}{\sqrt{\mathrm{Var}[T_{n,p}]}}$ and $Z \overset{d}{=} \mathcal{N}(0,1)$,

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

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 $W_{n,p_n} \to_w \mathcal{N}(0,1)$ iff $np_n \to \infty$. See [Ruciński, PTRF 1988].