CLT FOR DEPENDENT R.V'S

CLT Let $X_1,\dots,X_n$ be mean zero random variables such that each X_i may depend on at most d r.v.s. Let $W=X_1+\dots+X_n$ have $\mathrm{Var}(W)=1$. Then

$$W_1(W,Z) \le Cd^2 \sum_{i=1}^{n} \mathbb{E}|X_i|^3$$

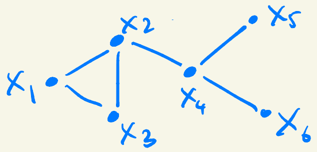
where $Z\sim N(0,1)$.

Formally: consider a graph $G=(V,E)$ with $V=\{1,\dots,n\}$. ("Dependency graph")

Assume that:

(a) The degrees of all vertices are $\le d-1$

(b) $\{X_i : i\in V_1\} \perp\!\!\!\perp \{X_j : j\in V_2\}$ whenever $\nexists$ edges from $V_1\subset V$ to $V_2 \subset V$.



Example: for this graph, $d=3$ and $\{X_1,X_2,X_3\} \perp\!\!\!\perp \{X_5,X_6\}$

Proof By Stein's lemma (p.133), it is enough to prove that

$$\mathbb{E}Wf(W) = \mathbb{E}f'(W) + O\left(\sum_{i=1}^{n}\mathbb{E}|X_i|^3\right) \quad \text{whenever } \|f''\|_\infty \le 1.$$

- $\mathbb{E}Wf(W) = \sum_i \mathbb{E}X_i f(W)$. X_i, W may be dependent. So let's remove the "dependence neighborhood" $N_i := \{i\}\cup\{j\in V : (i,j)\in E\}$:

$W_i := \sum_{j\notin N_i} X_j$ Taylor approximation, using $\|f''\|_\infty \le 1$:

$$\begin{tikzcd} W_i \arrow[r, "\sum_{j\in N_i} X_j"] & W \end{tikzcd}$$

$$f(W) = f(W_i) + (W-W_i)f'(W_i) + O\left((W-W_i)^2\right).$$

Multiply both sides by X_i, sum over i, take $\mathbb{E}$ $\Rightarrow$

$$\mathbb{E}Wf(W) = \sum_i \underbrace{\mathbb{E}X_i f(W_i)}_{\substack{\text{independence}\,\| \\ 0}} + \underbrace{\sum_i \mathbb{E}X_i \overbrace{(W-W_i)}^{\substack{\sum_{j\in N_i} X_j \\ \|}} f'(W_i)}_{\substack{\| \\ \text{(I)}}} + O\Big(\underbrace{\sum_i \mathbb{E}X_i \overbrace{(W-W_i)}^{\substack{\sum_{j\in N_i} X_j \\ \|}}{}^2}_{\substack{\| \\ \text{(II)}}}\Big) \qquad (*)$$

$$\textcircled{I} = \sum_i \sum_{j \in N_i} \mathbb{E}[X_i X_j f'(W_i)]$$

X_j, W_j may still be dependent.

So let's remove the "dependence neighborhood" of j as well:

$$\boxed{W_{ij} := \sum_{k \notin N_i \cup N_j} X_k}$$

Taylor approximation, using $\|f''\|_\infty \le 1$:

this sum contains only the terms X_k for which $k \in N_j$

$$f'(W_i) = f'(W_{ij}) + O(W_i - W_{ij}) = f'(W_{ij}) + O\Big(\sum_{k \in N_j} |X_k|\Big).$$

$$\Rightarrow \textcircled{I} = \underbrace{\sum_i \sum_{j \in N_i} \mathbb{E}[\underbrace{X_i X_j}\, \underbrace{f'(W_{ij})}_{\text{independent}}]}_{I_a} + O\Big(\underbrace{\sum_i \sum_{j \in N_i} \sum_{k \in N_j} \mathbb{E}|X_i X_j X_k|}_{I_b}\Big)$$

- $\mathbb{E}[X_i X_j f'(W_{ij})] = \mathbb{E}[X_i X_j]\, \mathbb{E}[f'(W_{ij})]$.

Let's replace $W_{ij} \mapsto W$. Taylor approximation, using $\|f''\|_\infty \le 1$:

$$f'(W_{ij}) = \mathbb{E} f'(W) + O(W - W_{ij}) = f'(W) + O\Big(\sum_{k \in N_i \cup N_j} X_k\Big).$$

$$\Rightarrow I_a = \underbrace{\sum_i \sum_{j \in N_i} \mathbb{E}[X_i X_j]}_{\overset{\text{indep}}{=} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[X_i X_j] = \mathbb{E}\left(\sum_{i=1}^n X_i\right)^2 = \mathbb{E}W^2 = 1} f'(W) + O\Big(\sum_i \sum_{j \in N_i} \sum_{k \in N_i \cup N_j} \underbrace{\mathbb{E}|X_i X_j| \cdot \mathbb{E}|X_k|}_{\wedge\, CS;\ L^1 \le L^2}\Big)$$

$$\|X_i\|_{L^2} \|X_j\|_{L^2} \|X_k\|_{L^2} \le \|X_i\|_{L^3} \|X_j\|_{L^3} \|X_k\|_{L^3} \overset{\text{am-gm}}{\le} \tfrac{1}{3}\left(\mathbb{E}|X_i|^3 + \mathbb{E}|X_j|^3 + \mathbb{E}|X_k|^3\right)$$

$$\Rightarrow I = I_a + I_b = \mathbb{E} f'(W) + O\Big(\underbrace{\sum_i \sum_{j \in N_i} \sum_{k \in N_i \cup N_j}}\left(\mathbb{E}|X_i|^3 + \mathbb{E}|X_j|^3 + \mathbb{E}|X_k|^3\right)\Big)$$

∀ triple i, j, k in this sum,
all i, j, k are within graph distance 2 from each other

⇒
- ∀ i, there are $\le d^2$ pairs (j, k) in the sum,
- ∀ j, there are $\le d^2$ pairs (i, k) in the sum,
- ∀ k, there are $\le d^2$ pairs (i, j) in the sum.

$$\Rightarrow I = \mathbb{E} f'(W) + O\Big(d^2 \sum_i \mathbb{E}|X_i|^3\Big).$$

$$\textcircled{\text{II}} \leq \sum_i \mathbb{E}\Big|X_i \underbrace{\Big(\sum_{j\in N_i} X_j\Big)^2}_{\text{expand}}\Big| \underset{\Delta}{\leq} \sum_i \sum_{j,k\in N_i} \underbrace{\mathbb{E}|X_i X_j X_k|}_{\substack{\wedge\!\mid\ \text{Hölder} \\ \|X_i\|_{L^3}\|X_j\|_{L^3}\|X_k\|_{L^3}}} \overset{\text{as in } \textcircled{\text{I}}}{\leq} d^2 \sum_i \mathbb{E}|X_i|^3$$

$$\overset{(*)}{\Rightarrow} \quad \mathbb{E}Wf(W) = \text{I} + O(\text{II}) = \mathbb{E}f'(W) + d^2 \sum_i \mathbb{E}|X_i|^3 . \qquad \square$$

- **Example**: $W = X_1X_2 + X_2X_3 + \cdots + X_{n-1}X_n$ where all X_i are mean 0 indep.

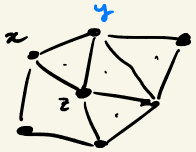
  $d=3 \Rightarrow W$ satisfies CLT. More generally, time series.

HW: $W = \sum_{i,j=1}^{n} X_iX_j$ is NOT always approx. normal.

Application: Triangles in Random Graphs

- Consider an Erdös-Rényi random graph

  $G \sim G(n,p)$ with $p = \text{const}$, $n \to \infty$.



$\Delta = 4$ here

$\Delta := \#(\text{triangles in } G)$ satisfies CLT?

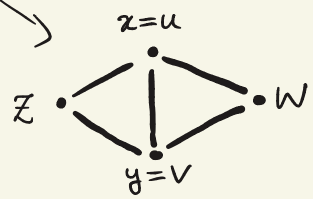
$$= \underbrace{\sum_{xyz}}_{\binom{n}{3} \text{ triples}} \mathbb{1}_{\Delta_{xyz}} \quad \text{where} \quad \Delta_{xyz} = \{\text{triple } xyz \text{ is a triangle}\}$$

- Dependency graph $G = (V, E)$: $V := \{\text{unordered triples } xyz\}$.

  $\Delta_{xyz} \perp\!\!\!\perp \Delta_{uvw}$ unless the triples share an edge $\Rightarrow$ E: connect $xyz \leftrightarrow uvw$ if they share an edge. For a given xyz, there are $\sim n$ such triples uvw

  $\Downarrow$

  $\boxed{d \lesssim n}$



- $\mathbb{P}(\Delta_{xyz}) = p^3 \quad \Rightarrow \quad \boxed{\mathbb{E}\Delta = \binom{n}{3} p^3 \asymp n^3}$

- $\mathrm{Var}(\Delta) = \sum_{xyz,\, uvw} \underbrace{\mathrm{Cov}\left(\mathbb{1}_{\Delta_{xyz}}, \mathbb{1}_{\Delta_{uvw}}\right)}_{\overset{\shortparallel}{?}}$

(a) All covariances are ≥ 0

[xyz being a Δ may only increase the prob. of uvw being a Δ.]

(b) At least $\gtrsim n^4$ covariances are $\gtrsim 1$

[For such two triples,

$$\mathrm{cov}\left(\mathbb{1}_{\Delta_{xyz}}, \mathbb{1}_{\Delta_{uvw}}\right) = \mathbb{P}(\Delta_{xyz} \,\&\, \Delta_{uvw}) - \mathbb{P}(\Delta_{xyz})\mathbb{P}(\Delta_{uvw}) = p^5 - p^6 \gtrsim 1$$

and there are $\binom{n}{4} \asymp n^4$ such pairs of triples.]

(a) & (b) $\Rightarrow$ $\boxed{\underbrace{\mathrm{Var}(\Delta)}_{\overset{\shortparallel}{\sigma^2}} \gtrsim n^4}$

- Use CLT (p.137) for $W := \dfrac{\Delta - \mathbb{E}\Delta}{\sqrt{\mathrm{Var}\Delta}} = \sum_{xyz} \dfrac{\overbrace{\mathbb{1}_{\Delta_{xyz}}}^{\mathrm{Ber}(p^3)} - p^3}{\sigma} \Rightarrow$

$$W_1(W, Z) \lesssim \frac{d^2}{\sigma^3} \underbrace{\sum_{xyz}}_{\lesssim n^3 \text{ terms}} \underbrace{\mathbb{E}\left|\mathbb{1}_{\Delta_{xyz}} - p^3\right|^3}_{\overset{\wedge\!\shortparallel}{1}} \leq \frac{n^2}{n^6} \cdot n^3 \lesssim \frac{1}{n}.$$

$\Rightarrow$ **THM** The # of triangles Δ in Erdös-Rényi graph $G(n,p)$ is approximately normal:

$$d\left(\frac{\Delta - \mathbb{E}\Delta}{\sqrt{\mathrm{Var}\Delta}}, Z\right) \leq \frac{C(p)}{n} \quad \text{where } Z \sim N(0,1)$$

Remark A finer analysis $\Rightarrow$ CLT holds iff $np^3 \to \infty$.