Lecture 9. A guest lecture by Nathan Ross

3.5. **Stein's method via local dependency.** In the last two lectures, we learned the basics of Stein's method, which is a powerful toolbox of establishing central limit theorems with rate of convergence. Let F be a real-valued random variable of interest, which is expected to be approximately normal with matched mean and matched variance. Say, $\mathbb{E}[F] = 0$ and $\mathrm{Var}(F) = 1$. The fundamental Stein's bounds (see, e.g., Proposition 3.3) tell us that in order to bound the distributional distance between F and a standard normal $Z \sim \mathcal{N}(0, 1)$, it is enough for us to control the expectation $\mathbb{E}[f(F)F] - \mathbb{E}[f'(F)]$ with f in a suitable family of test functions. To achieve this goal, one often requires to utilize the particular structure of the random variable F, for example, independence and symmetry inherited from the construction of the statistic F (see Example 1), and the exchangeable pairs structure as in Theorem 3.4. As we have seen in Proposition 3.5, the construction of an exchangeable pair approach requires one to introduce the **auxiliary randomness**, which is quite standard within the Stein's method.

In what follows, we will introduce a notion of local dependency, with which we can combine the Stein's bounds as well. This notion goes back to Charles Stein and Louis H.Y. Chen [St72, Cha75] in the 70s, and it was extensively studied in the paper [CS04] by Chen and Shao under various dependence settings. See also [BR89, CR10] for more details. Let us start with a simple definition.

Definition 3.8. Given a collection of random variables $\mathbb{X} = \{X_i\}_{i\in\mathcal{V}}$ and a graph $\mathcal{G}$ with vertex set $\mathcal{V}$, we say $\mathcal{G}$ is a **dependency graph** for $\mathbb{X}$ if the following property holds:

$$\text{whenever } I, J \subset \mathcal{V} \text{ are disjoint with no edge between them,}$$
$$\{X_i\}_{i\in I} \text{ and } \{X_j\}_{j\in J} \text{ are independent.}$$

In particular, if there is no edge between $i, j \in \mathcal{V}$, then X_i is independent from X_j; in other words, the neighbors of the vertex i, together with i itself, form a **dependency neighborhood**.

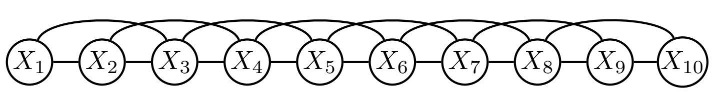
Here are a few examples of dependency graphs.

Example 2. (i) If $\mathbb{X} = \{X_1, ..., X_n\}$ is a family of independent random variables, any graph on $[n]$ is a dependency graph for $\mathbb{X}$. For example, the empty graph (i.e., empty edge set) and the complete graph (i.e., every two vertices are connected by an edge) are dependency graphs for $\mathbb{X}$. (*Which one is better?*)

(ii) [Sliding window statistics] Fix an integer $k \geq 1$. Let $Y_1, ..., Y_{n+k-1}$ be independent random variables with values in some measure space $(E, \mathcal{E}, \mu)$, and $\varphi_i : E^k \to \mathbb{R}$ be measurable. Define

$$X_i = \varphi_i(Y_i, Y_{i+1}, ..., Y_{i+k-1}),\ i = 1, ..., n.$$

When $k = 1$, we are back to the independent case as in (i). When $(k, n) = (3, 10)$, we have an example of dependency graph:



(edges connect variables with $|i - j| < 3$)

One can see that the maximum degree is 4 in the above picture.[21] The above is the smallest possible dependency graph.

[21]We write $i \sim j$ if the two (distinct) vertices $i, j \in [n]$ are directly connected by an edge. The degree of a vertex i is defined by $\deg(i) = \mathrm{card}\{j : j \sim i\}$, the number of vertices that are directly connected to i via an edge. The maximum degree is $\max\{\deg(i) : i \in [n]\}$.

Lemma 3.9. *Suppose $\{Z_v\}_{v\in\mathcal{V}}$ is a family of square-integrable random variables admitting a dependency graph with maximum degree D. Then,*

$$\operatorname{Var}\Big(\sum_{v\in\mathcal{V}} Z_v\Big) \le (D+1)\sum_{v\in\mathcal{V}} \mathbb{E}[Z_v^2].$$

Proof. We begin with

$$\operatorname{Var}\Big(\sum_{v\in\mathcal{V}} Z_v\Big) = \sum_{u,v\in\mathcal{V}} \operatorname{Cov}(Z_u, Z_v),$$

and note that for each u, there are at most $(D+1)$ vertices v such that $\operatorname{Cov}(Z_u,Z_v)\neq 0$. Using $|\operatorname{Cov}(Z_u,Z_v)| \le \|Z_u\|_2\|Z_v\|_2 \le \frac{1}{2}\big(\mathbb{E}[Z_v^2]+\mathbb{E}[Z_u^2]\big)$, we get

$$\begin{aligned}\operatorname{Var}\Big(\sum_{v\in\mathcal{V}} Z_v\Big) &\le \frac{1}{2}\sum_{u,v\in\mathcal{V}} \big(\mathbb{E}[Z_v^2]+\mathbb{E}[Z_u^2]\big)\mathbf{1}_{\{\operatorname{Cov}(Z_u,Z_v)\neq 0\}} \\ &= \sum_{u,v\in\mathcal{V}} \mathbb{E}[Z_v^2]\mathbf{1}_{\{\operatorname{Cov}(Z_u,Z_v)\neq 0\}} \quad \text{by symmetry} \\ &\le (D+1)\sum_{v\in\mathcal{V}}\mathbb{E}[Z_v^2].\end{aligned}$$

Hence, the proof is completed. □

In the following, we present a central limit theorem in case of the local dependency.

Theorem 3.10. *Let $\{X_i\}_{1\le i\le n}$ be a family of centered random variables with $\mathbb{E}[X_i^4]<\infty$ for each i. Assume that the associated dependency graph has maximum degree $\mathfrak{m}$, and*

$$\sigma := \sqrt{\operatorname{Var}\Big(\sum_{i=1}^n X_i\Big)} > 0.$$

Then, with $F := \frac{1}{\sigma}\sum_{i=1}^n X_i$, we have

$$d_{\mathrm{W}}(F,\mathcal{N}(0,1)) \le (\mathfrak{m}+1)^2\underbrace{\Big(\sigma^{-3}\sum_{i=1}^n \mathbb{E}[|X_i|^3]\Big)}_{\text{1st term}} + \frac{2\sqrt{2/\pi}}{\sigma^2}(\mathfrak{m}+1)^{3/2}\sqrt{\sum_{i=1}^n \mathbb{E}[|X_i|^4]}. \tag{3.11}$$

The first term in (3.11) appears in the usual CLT, and the second term is often of comparable order. It is easy to see that when $\{X_i\}$ are i.i.d., then the above bound (3.11) is consistent with the one in Proposition 3.5.

Proof of Theorem 3.10. In view of Proposition 3.3, it suffices to bound the following expectation

$$\mathbb{E}[\varphi'(F) - F\varphi(F)]$$

for $\varphi\in C^2(\mathbb{R})$ with $\|\varphi\|_\infty\le 2$, $\|\varphi'\|_\infty \le \sqrt{2/\pi}$, and $\|\varphi''\|_\infty\le 2$.

Let $\{\mathcal{N}_i : i\in[n]\}$ be the dependency neighborhood, i.e., $\mathcal{N}_i = \{j : j \leftrightsquigarrow i\}$, where $j \leftrightsquigarrow i$ means $j=i$ or j,i are distinct and connected by an edge (denoted by $i\sim j$). Then, the cardinality $\operatorname{card}(\mathcal{N}_i)$ of $\mathcal{N}_i$ is bounded by $\mathfrak{m}+1$, and we also have the following equivalence:

$$j\in\mathcal{N}_i \iff i \leftrightsquigarrow j \iff i\in\mathcal{N}_j. \tag{3.12}$$

Put

$$F_i = F - \frac{1}{\sigma} \sum_{j \in \mathcal{N}_i} X_j$$

for $i \in [n]$. Then, F_i is independent from X_i. In particular, we have

$$\mathbb{E}[\varphi(F_i) X_i] = 0. \tag{3.13}$$

Now let us deal with the expectation $\mathbb{E}[F\varphi(F)]$ to derive some expression like $\mathbb{E}[\varphi'(F)\eta]$:

$$\begin{aligned}
\mathbb{E}[F\varphi(F)] &= \frac{1}{\sigma} \sum_{i=1}^{n} \mathbb{E}[X_i \varphi(F)] \\
&= \frac{1}{\sigma} \sum_{i=1}^{n} \mathbb{E}\big[X_i\big(\varphi(F) - \varphi(F_i)\big)\big] \quad \text{by (3.13)} \\
&= \Big(\frac{1}{\sigma} \sum_{i=1}^{n} \mathbb{E}\big[X_i \varphi'(F)(F - F_i)\big]\Big) - \frac{1}{2\sigma} \sum_{i=1}^{n} \mathbb{E}\big[X_i \varphi''(\widehat{F}_i)(F - F_i)^2\big],
\end{aligned} \tag{3.14}$$

where in the last step, we did Taylor expansion for φ around F with $\widehat{F}_i$ some random variable with values between F and F_i. Let us first do some algebra and deduce from (3.14), with $\|\varphi'\|_\infty \le \sqrt{2/\pi}$, and $\|\varphi''\|_\infty \le 2$, that

$$\begin{aligned}
\big|\mathbb{E}[F\varphi(F)] - \mathbb{E}[\varphi'(F)]\big| &\le \Big|\mathbb{E}\Big[\varphi'(F)\Big(-1 + \frac{1}{\sigma^2} \sum_{i=1}^{n} \sum_{j \in \mathcal{N}_i} X_j X_i\Big)\Big]\Big| + \frac{\|\varphi''\|_\infty}{2\sigma^3} \sum_{i=1}^{n} \mathbb{E}\Big|X_i\big(\sum_{j \in \mathcal{N}_i} X_j\big)^2\Big| \\
&\le \frac{\sqrt{2/\pi}}{\sigma^2} \mathbb{E}\Big| -\sigma^2 + \sum_{i=1}^{n} \sum_{j \in \mathcal{N}_i} X_j X_i\Big| + \frac{1}{\sigma^3} \sum_{i=1}^{n} \mathbb{E}\Big|X_i\big(\sum_{j \in \mathcal{N}_i} X_j\big)^2\Big| \\
&=: \frac{\sqrt{2/\pi}}{\sigma^2} \mathbf{T}_1 + \frac{1}{\sigma^3} \mathbf{T}_2.
\end{aligned} \tag{3.15}$$

• **bounding $\mathbf{T}_1$.** First we can deduce from (3.14) with $\varphi(x) = x$ that

$$\mathbb{E}\Big(\sum_{i=1}^{n} \sum_{j \in \mathcal{N}_i} X_j X_i\Big) = \sigma^2.$$

For simplicity, we write $\sum_{i=1}^{n} \sum_{j \in \mathcal{N}_i} = \sum_{i \leftrightsquigarrow j}$ in view of the equivalence (3.12). Next, by Cauchy-Schwarz, we get

$$\big(\mathbf{T}_1\big)^2 \le \mathrm{Var}\Big(\sum_{i \leftrightsquigarrow j} X_i X_j\Big) = \sum_{i \leftrightsquigarrow j} \sum_{k \leftrightsquigarrow \ell} \mathrm{Cov}(X_i X_j, X_k X_\ell). \tag{3.16}$$

The above covariance is nonzero only when $k \in \mathcal{N}_i \cup \mathcal{N}_j$ or $\ell \in \mathcal{N}_i \cup \mathcal{N}_j$ for any given (i,j) with $i \leftrightsquigarrow j$. The number of pairs (k,ℓ) with the above conditions and subject to $k \leftrightsquigarrow \ell$ can be bounded as follows:

(1) given (i,j), the cardinality of $\mathcal{N}_i \cup \mathcal{N}_j$ is bounded by $2(\mathfrak{m}+1)$, then there are at most $2(\mathfrak{m}+1)$ vertices k in $\mathcal{N}_i \cup \mathcal{N}_j$;
(2) given (i,j) and fixing $k \in \mathcal{N}_i \cup \mathcal{N}_j$, we have at most $(\mathfrak{m}+1)$ vertices $\ell \leftrightsquigarrow k$
(3) given (i,j), if $k \notin \mathcal{N}_i \cup \mathcal{N}_j$, then we must have $\ell \in \mathcal{N}_i \cup \mathcal{N}_j$, there are at most $2(\mathfrak{m}+1)$ such vertices ℓ, and we have at most $(\mathfrak{m}+1)$ vertices $k \leftrightsquigarrow \ell$;
(4) thus, given (i,j) with $i \leftrightsquigarrow j$, there are at most $4(\mathfrak{m}+1)^2$ nonzero terms in (3.16).

Therefore, if we view $\{X_iX_j\}_{i\in[n],j\in\mathcal{N}_i}$ as a new family of random variables with a dependency graph, then the graph has maximum degree $\le 4(\mathfrak{m}+1)^2-1$.[22] Thus, we deduce from Lemma 3.9 with (3.16), $ab \le \frac{1}{2}(a^2+b^2)$, and (3.12) that

$$\begin{aligned}\big(\mathbf{T}_1\big)^2 &\le 4(\mathfrak{m}+1)^2 \sum_{i=1}^n \sum_{j\in\mathcal{N}_i} \mathbb{E}[X_i^2X_j^2] \\ &\le 4(\mathfrak{m}+1)^2 \sum_{i=1}^n \sum_{j\in\mathcal{N}_i} \frac{1}{2}\mathbb{E}[X_i^4+X_j^4] \\ &\le 4(\mathfrak{m}+1)^3 \sum_{i=1}^n \mathbb{E}[X_i^4].\end{aligned} \tag{3.17}$$

• **bounding $\mathbf{T}_2$.** Using the inequality of arithmetic and geometric means, we have

$$\mathbb{E}|X_iX_jX_k| \le \frac{1}{3}\mathbb{E}\big(|X_i|^3+|X_j|^3+|X_k|^3\big).$$

Thus,

$$\begin{aligned}\mathbf{T}_2 &\le \sum_{i=1}^n \sum_{k,j\in\mathcal{N}_i} \mathbb{E}|X_iX_jX_k| \le \frac{1}{3}\sum_{i=1}^n \sum_{k,j\in\mathcal{N}_i} \mathbb{E}\big(|X_i|^3+|X_j|^3+|X_k|^3\big) \\ &= \left(\frac{1}{3}\sum_{i=1}^n \mathbb{E}\big(|X_i|^3\big) \sum_{k,j\in\mathcal{N}_i} 1\right) + \frac{2}{3}\sum_{i=1}^n \sum_{k,j\in\mathcal{N}_i} \mathbb{E}\big(|X_k|^3\big) \\ &\le \left(\frac{(\mathfrak{m}+1)^2}{3}\sum_{i=1}^n \mathbb{E}\big(|X_i|^3\big)\right) + \frac{2(\mathfrak{m}+1)}{3}\sum_{i=1}^n \sum_{k\in\mathcal{N}_i} \mathbb{E}\big(|X_k|^3\big) \\ &\le (\mathfrak{m}+1)^2 \sum_{i=1}^n \mathbb{E}\big(|X_i|^3\big).\end{aligned}$$

Together with (3.17) and (3.15), it leads to the desired conclusion (3.11). □

• **Application I: Headruns.** Fix an integer $k \ge 1$. Let $Y_1, ..., Y_{n+k-1}$ be i.i.d. Bernoulli random variables with success probability $p \in (0,1)$. Then

$$\prod_{j=i}^{i+k-1} Y_j \tag{3.18}$$

is equal to one if there is a *run* of ones of length k starting at the position i. The k-run statistic is a classical concept in probability and combinatorial statistics that counts the number of consecutive sequences (runs) of length k in a binary or categorical sequence. It has powerful applications in DNA sequence analysis, where nucleotides (A, C, G, T) form a symbolic sequence that can be modeled as a random string. Here is a short example of a DNA sequence – a real-like but synthetic one (not from any specific organism):

ATGCGTACCGTTAGCTGACCTGAAATTTGCGCGGCTAA,

where each letter represents one nucleotide base: A = Adenine, T = Thymine, C = Cytosine, and G = Guanine. In the above sequence (length =38), there are only two 3-runs. The statistics in

[22] -1 comes from the exclusion of a vertex itself when counting its degree.

(3.18) can model, for instance, the similarity between two DNA sequences. Here is an example:

sequence 1: ATG**CG**TACCGT**T**A**GC**T**G**ACC**T**G**A**AAT**T**TGCGC**G**GCTAA
sequence 2: GAT**CG**GTAACC**T**T**GC**A**G**GAT**T**A**A**GCG**T**CCGAT**G**CAAGC
matched? : $\underline{\checkmark\checkmark}$ $\checkmark$ $\underline{\checkmark\checkmark}$ $\checkmark$ $\checkmark$ $\checkmark$ $\checkmark$ $\checkmark$

we see that there are only two 2-runs of consecutive matches.

It is a natural question to ask about how many k-runs there are, and what is its asymptotic behavior as the number of observations tends to infinity. That is, letting

$$W_n = \sum_{i=1}^{n} \prod_{j=i}^{i+k-1} Y_j, \tag{3.19}$$

we are interested at the fluctuation of properly scaled W_n as $n \to +\infty$.

Proposition 3.11. *Let W_n be given as in (3.19). Then,*

$$d_{\mathrm{W}}\big(\tfrac{1}{\sigma_n}(W_n - \mathbb{E}[W_n]), \mathcal{N}(0,1)\big) \lesssim \frac{k^2}{\sqrt{n}}, \tag{3.20}$$

where the above implicit constant depends on (p,k) but not on n.

Proof. Let $X_i = \prod_{j=i}^{i+k-1} Y_j$. Then, we are in the setting of Example 2-(ii). It is not difficult to see that the corresponding dependency graph for $\{X_1, ..., X_n\}$ has maximum degree $\mathfrak{m} \le 2(k-1)$. Then, it is enough to apply Theorem 3.10. Let us first do some moment estimates.

• **Basic moments.** Since $X_i := \prod_{j=i}^{i+k-1} Y_j = \mathbf{1}_{\{Y_i = \cdots = Y_{i+k-1} = 1\}}$, we have

$$\mathbb{E}[X_i] = p^k \quad \text{and} \quad \mathrm{Var}(X_i) = p^k(1-p^k).$$

It is also easy to verify that for $j \ge 1$, $\mathbb{E}[X_i X_{i+j}] = p^{k+j}\mathbf{1}_{\{1 \le j \le k-1\}} + p^{2k}\mathbf{1}_{\{j \ge k\}}$. Therefore,

$$\begin{aligned} \sigma_n^2 := \mathrm{Var}(W_n) &= \sum_{i,j=1}^{n} \big(p^{k+|j-i|} - p^{2k}\big)\mathbf{1}_{\{|j-i| \le k\}} \\ &= np^k(1-p^k) + 2\sum_{1 \le i < j \le n} \big(p^{k+|j-i|} - p^{2k}\big)\mathbf{1}_{\{|j-i| \le k\}} \\ &\ge np^k(1-p^k), \end{aligned} \tag{3.21}$$

where the second term in (3.21) is clearly strictly positive and of order n as $n \to \infty$ (for fixed p, k). That is, we have $\sigma_n \asymp \sqrt{n}$, although the above lower bound suffices for our purpose.

• **Dependency graph bound.** Set $\xi_i = X_i - \mathbb{E}X_i$. Then, $\{\xi_i\}$ satisfy the dependency graph condition with $\mathfrak{m} \le 2(k-1)$. Then, it follows from Theorem 3.10 that

$$d_{\mathrm{W}}\Big(\frac{1}{\sigma_n}\sum_{i=1}^{n} \xi_i, \mathcal{N}(0,1)\Big) \le (\mathfrak{m}+1)^2\Big(\sigma_n^{-3}\sum_{i=1}^{n} \mathbb{E}[|\xi_i|^3]\Big) + \frac{2\sqrt{2/\pi}}{\sigma_n^2}(\mathfrak{m}+1)^{3/2}\sqrt{\sum_{i=1}^{n} \mathbb{E}[|\xi_i|^4]}.$$

Since $|\xi_i| \le 1$, we have the trivial bounds: $\mathbb{E}|\xi_i|^3 \le 1$ and $\mathbb{E}|\xi_i|^4 \le 1$. Thus, the desired bound (3.20) follows immediately from (3.21). $\square$

Remark 3.12. The bound in (3.20) holds for fixed (p,k), while as one can see from our proof, a finer estimate of the variance σ_n^2 could lead to an asymptotically negligible rate of convergence when k, p, n vary together in a certain way.

• **Application II: Triangles in Erdős-Rényi random graphs.** Erdős-Rényi random graphs mark the birth of modern random graph theory (1959-1961); see, e.g., the book [vdH17] for more details. An Erdős-Rényi random graph can be constructed as follows:

(i) Let K_n be a complete graph on the vertex set $[n]$, i.e., every two distinct vertices are connected by an edge;

(ii) let E_n be the edge set of K_n, and let $\{Y_e\}_{e\in E_n}$ be a sequence of i.i.d. Bernoulli random variables with $\mathbb{P}(Y_e = 1) = 1 - \mathbb{P}(Y_e = 0) = p \in (0,1)$;

(iii) when $Y_e = 1$, we say the edge e is open; the subgraph of K_n with the vertex set $[n]$ and the open edges is called an Erdős-Rényi random graph, denoted by $\mathcal{G}(n,p)$.

Let T be the number of triangles formed in the Erdős-Rényi random graph $\mathcal{G}(n,p)$. The following result suggests that T is asymptotically normal in some regime where $p \to 0$ along $n \to \infty$; see, e.g., [BKR89, Rol22] for more results.

Proposition 3.13. *Let the above notations prevail. It holds for any $n \geq 3$ and $p \in (0,1)$ that*

$$\sigma^2 := \mathrm{Var}(T) = \binom{n}{3} p^3 [1 - p^3 + 3(n-3)p^2(1-p)]$$

and

$$\begin{aligned} d_{\mathrm{W}}\Big(\frac{1}{\sigma}\big(T - \mathbb{E}[T]\big),\ \mathcal{N}(0,1)\Big) \leq{}& \frac{(3n-8)^2}{\sigma^3}\binom{n}{3} p^3(1-p^3)[(1-p^3)^2 + p^6] \\ &+ \frac{\sqrt{26}(3n-8)^{\frac{3}{2}}}{\sigma^2\sqrt{\pi}} \sqrt{\binom{n}{3} p^3(1-p^3)[(1-p^3)^3 + p^9]}. \end{aligned} \tag{3.22}$$

In particular, if $p \sim n^{-\alpha}$ with $\alpha \in [0, \frac{2}{9})$, the bound in (3.22) tends to zero.

Proof. to be added later (not very soon). Or you can try to fill in the details. □

$\propto\propto\propto\propto\propto\propto$ **End of Lecture 9 (Sept. 30, 2025)** $\infty\infty\infty\infty\infty\infty\infty$