Weighted dependency graphs

Valentin Féray

Institut für Mathematik, Universität Zürich

Final conference of the MADACA project
Domaine de Chalès, June 20th – June 24th 2016

Central limit theorems

Theorem

If $Y_1, Y_2, \dots$ are independent identically distributed variables with finite variance, and $X_n = \sum_{i=1}^n Y_i$, then

$$\frac{X_n - \mathbb{E}(X_n)}{\sqrt{\operatorname{Var} X_n}} \xrightarrow{d} \mathcal{N}(0,1). \tag{CLT}$$

Central limit theorems

Theorem

If $Y_1, Y_2, \dots$ are independent identically distributed variables with finite variance, and $X_n = \sum_{i=1}^n Y_i$, then

$$\frac{X_n - \mathbb{E}(X_n)}{\sqrt{\operatorname{Var} X_n}} \xrightarrow{d} \mathcal{N}(0,1). \tag{CLT}$$

Relax identical distribution hypothesis $\longrightarrow$ Lindeberg condition.

Central limit theorems

Theorem

If $Y_1, Y_2, \dots$ are independent identically distributed variables with finite variance, and $X_n = \sum_{i=1}^n Y_i$, then

$$\frac{X_n - \mathbb{E}(X_n)}{\sqrt{\operatorname{Var} X_n}} \xrightarrow{d} \mathcal{N}(0,1). \tag{CLT}$$

Relax identical distribution hypothesis $\longrightarrow$ Lindeberg condition.

Relax independence hypothesis: leads to CLT for Markov chains, martingales, mixing sequences, exchangeable pairs, determinantal point processes, dependency graphs, ...

Central limit theorems

Theorem

If $Y_1, Y_2, \dots$ are independent identically distributed variables with finite variance, and $X_n = \sum_{i=1}^n Y_i$, then

$$\frac{X_n - \mathbb{E}(X_n)}{\sqrt{\operatorname{Var} X_n}} \xrightarrow{d} \mathcal{N}(0,1). \tag{CLT}$$

Relax identical distribution hypothesis $\longrightarrow$ Lindeberg condition.

Relax independence hypothesis: leads to CLT for Markov chains, martingales, mixing sequences, exchangeable pairs, determinantal point processes, dependency graphs, ...

Goal of the talk: give an extension of dependency graphs that has a wide range of application.

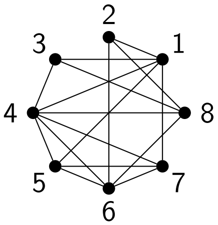
Dependency graphs

(Petrovskaya/Leontovich, Janson, Baldi/Rinott, Mikhailov, 80's)

A problem in random graphs

Erdős-Rényi model of random graphs $G(n, p)$:

- G has n vertices labelled $1, \ldots, n$;
- each edge $\{i, j\}$ is taken independently with probability p;

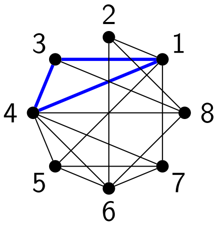


Example : $n = 8, p = 1/2$

A problem in random graphs

Erdős-Rényi model of random graphs $G(n, p)$:

- G has n vertices labelled $1, \dots, n$;
- each edge $\{i, j\}$ is taken independently with probability p;



Example : $n = 8, p = 1/2$

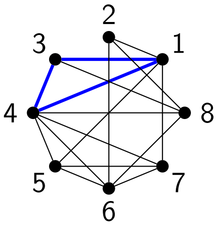
Question

Fix $p \in (0; 1)$. Does the number of triangles T_n satisfy a CLT?

A problem in random graphs

Erdős-Rényi model of random graphs $G(n,p)$:

- G has n vertices labelled $1, \dots, n$;
- each edge $\{i,j\}$ is taken independently with probability p;



Example : $n = 8, p = 1/2$

Question

Fix $p \in (0;1)$. Does the number of triangles T_n satisfy a CLT?

$$T_n = \sum_{\Delta = \{i,j,k\} \subset [n]} Y_\Delta, \text{ where } Y_\Delta(G) = \begin{cases} 1 & \text{if } G \text{ contains the triangle } \Delta; \\ 0 & \text{otherwise.} \end{cases}$$

T_n is a sum of mostly independent variables.

Dependency graphs

Definition (Petrovskaya and Leontovich, 1982, Janson, 1988)

A graph L with vertex set A is a dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if

- if A_1 and A_2 are disconnected subsets in L, then $\{Y_\alpha, \alpha \in A_1\}$ and $\{Y_\alpha, \alpha \in A_2\}$ are independent.

Roughly: there is an edge between pairs of dependent random variables.

Dependency graphs

Definition (Petrovskaya and Leontovich, 1982, Janson, 1988)

A graph L with vertex set A is a dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if

- if A_1 and A_2 are disconnected subsets in L, then $\{Y_\alpha, \alpha \in A_1\}$ and $\{Y_\alpha, \alpha \in A_2\}$ are independent.

Roughly: there is an edge between pairs of dependent random variables.

Example

Consider $G = G(n,p)$. Let $A = \{\Delta \in \binom{[n]}{3}\}$ (set of potential triangles) and

$$\{\Delta_1,\ \Delta_2\} \in E_L \text{ iff } \Delta_1 \text{ and } \Delta_2 \text{ share an edge in } G.$$

Then L is a dependency graph for the family $\{Y_\Delta, \Delta \in \binom{[n]}{3}\}$.

Dependency graphs

Definition (Petrovskaya and Leontovich, 1982, Janson, 1988)

A graph L with vertex set A is a dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if

- if A_1 and A_2 are disconnected subsets in L, then $\{Y_\alpha, \alpha \in A_1\}$ and $\{Y_\alpha, \alpha \in A_2\}$ are independent.

Roughly: there is an edge between pairs of dependent random variables.

Example

Note: L has degree $O(n)$

Consider $G = G(n,p)$. Let $A = \{\Delta \in \binom{[n]}{3}\}$ (set of potential triangles) and

$$\{\Delta_1,\, \Delta_2\} \in E_L \text{ iff } \Delta_1 \text{ and } \Delta_2 \text{ share an edge in } G.$$

Then L is a dependency graph for the family $\{Y_\Delta, \Delta \in \binom{[n]}{3}\}$.

Janson's normality criterion

Setting: for each n,

- $\{Y_{n,i}, 1 \leq i \leq N_n\}$ is a family of bounded random variables; $|Y_{n,i}| < M$ a.s.
- we have a dependency graph L_n with maximal degree $\Delta_n - 1$.
- we set $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.

Janson's normality criterion

Setting: for each n,

- $\{Y_{n,i}, 1 \leq i \leq N_n\}$ is a family of bounded random variables; $|Y_{n,i}| < M$ a.s.
- we have a dependency graph L_n with maximal degree $\Delta_n - 1$.
- we set $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.

Theorem (Janson, 1988)

Assume that $\left(\frac{N_n}{\Delta_n}\right)^{1/s} \frac{\Delta_n}{\sigma_n} \to 0$ *for some integer* s. *Then* X_n *satisfies a CLT.*

Janson's normality criterion

Setting: for each n,

- $\{Y_{n,i}, 1 \leq i \leq N_n\}$ is a family of bounded random variables; $|Y_{n,i}| < M$ a.s.
- we have a dependency graph L_n with maximal degree $\Delta_n - 1$.
- we set $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.

Theorem (Janson, 1988)

Assume that $\left(\frac{N_n}{\Delta_n}\right)^{1/s} \frac{\Delta_n}{\sigma_n} \to 0$ *for some integer* s*. Then* X_n *satisfies a CLT.*

For triangles, $N_n = \binom{n}{3}$, $\Delta_n = O(n)$, while $\sigma_n \asymp n^2$. (for fixed p)

Corollary

Fix p *in* $(0,1)$*. Then* T_n *satisfies a CLT.*

(also true for $p_n \to 0$ with $np_n \to \infty$; originally proved by Rucinski, 1988).

Applications of dependency graphs to CLT results

- mathematical modelization of cell populations (Petrovskaya, Leontovich, 82);
- subgraph counts in random graphs (Janson, Baldi, Rinott, Penrose, 88, 89, 95, 03);
- Geometric probability (Avram, Bertsimas, Penrose, Yukich, Bárány, Vu, 93, 05 , 07);
- pattern occurrences in random permutations (Bóna, Janson, Hitchenko, Nakamura, Zeilberger, 07, 09, 14).
- m-dependence (Hoeffding, Robbins, 53, ...; now widely used in statistics) is a special case.

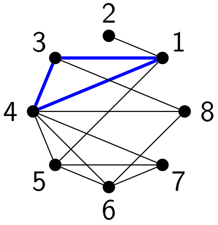
(Some of these applications use variants of Janson's normality criterion, which are more technical to state and omitted here...)

Not an application of dependency graphs

Random graph $G(n, M)$:

- G has n vertices labelled $1, \ldots, n$;
- The edge-set of G is taken uniformly among all possible edge-sets of cardinality M.

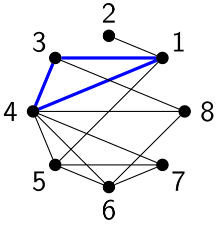
Example with $n = 8$ and $M = 14$



Not an application of dependency graphs

Random graph $G(n, M)$:

- G has n vertices labelled $1, \dots, n$;
- The edge-set of G is taken uniformly among all possible edge-sets of cardinality M.



Example with $n = 8$ and $M = 14$

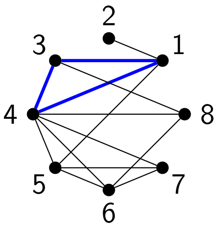
If $p = M/\binom{n}{2}$, each edge appears with probability p, but no independence any more!

Not an application of dependency graphs

Random graph $G(n, M)$:

- G has n vertices labelled $1, \dots, n$;
- The edge-set of G is taken uniformly among all possible edge-sets of cardinality M.

Example with $n = 8$ and $M = 14$



If $p = M / \binom{n}{2}$, each edge appears with probability p, but no independence any more!

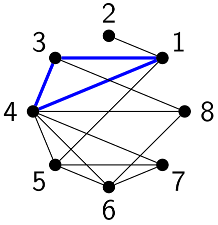
Question

Fix $p \in (0; 1)$ and $M = p\binom{n}{2}$. Does the number of triangles T_n in $G(n, M_n)$ satisfy a CLT?

Not an application of dependency graphs

Random graph $G(n, M)$:

- G has n vertices labelled $1, \dots, n$;
- The edge-set of G is taken uniformly among all possible edge-sets of cardinality M.



Example with $n = 8$ and $M = 14$

If $p = M / \binom{n}{2}$, each edge appears with probability p, but no independence any more!

Question

Fix $p \in (0;1)$ and $M = p\binom{n}{2}$. Does the number of triangles T_n in $G(n, M_n)$ satisfy a CLT?

T_n still writes as a sum of Y_Δ, but the Y_Δ are pairwise dependent!

Solution: edge-weighted dependency graphs

weighted graphs = graphs with weights in $[0,1]$ on edges.

Definition

A weighted graph $\widetilde{L}$ with vertex set A is a weighted dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if

(*skipped for the moment*).

Roughly: the smaller the weight on the edge $\{Y_\alpha, Y_\beta\}$ is, the closer to independence Y_α and Y_β should be.

Solution: edge-weighted dependency graphs

weighted graphs = graphs with weights in $[0,1]$ on edges.

Definition

A weighted graph $\widetilde{L}$ with vertex set A is a weighted dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if

(*skipped for the moment*).

Roughly: the smaller the weight on the edge $\{Y_\alpha, Y_\beta\}$ is, the closer to independence Y_α and Y_β should be.

Example

Consider $G = G(n, M)$, where $M = p\binom{n}{2}$. Let $A = \{\Delta \in \binom{[n]}{3}\}$ and

$$\mathrm{wt}_{\widetilde{L}}(\{\Delta_1, \Delta_2\}) = \begin{cases} 1 & \text{if } \Delta_1 \text{ and } \Delta_2 \text{ share an edge in } G. \\ 1/n^2 & \text{otherwise.} \end{cases}$$

Then $\widetilde{L}$ is a weighted dependency graph for the family $\{Y_\Delta, \Delta \in \binom{[n]}{3}\}$.

Solution: edge-weighted dependency graphs

weighted graphs = graphs with weights in $[0,1]$ on edges.

Definition

A weighted graph $\widetilde{L}$ with vertex set A is a weighted dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if

(*skipped for the moment*).

Roughly: the smaller the weight on the edge $\{Y_\alpha, Y_\beta\}$ is, the closer to independence Y_α and Y_β should be.

Note: $\widetilde{L}$ has degree $O(n^3)$, but **weighted** degree $O(n)$.

Example

Consider $G = G(n, M)$, where $M = p\binom{n}{2}$. Let $A = \{\Delta \in \binom{[n]}{3}\}$ and

$$\mathrm{wt}_{\widetilde{L}}(\{\Delta_1, \Delta_2\}) = \begin{cases} 1 & \text{if } \Delta_1 \text{ and } \Delta_2 \text{ share an edge in } G. \\ 1/n^2 & \text{otherwise.} \end{cases}$$

Then $\widetilde{L}$ is a weighted dependency graph for the family $\{Y_\Delta, \Delta \in \binom{[n]}{3}\}$.

A normality criterion for weighted dependency graphs

Setting: for each n,

- $\{Y_{n,i}, 1 \leq i \leq N_n\}$ is a family of bounded random variables; $|Y_{n,i}| < M$ a.s.
- we have a weighted dependency graph $\widetilde{L}_n$ with weighted maximal degree $\Delta_n - 1$.
- we set $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.

A normality criterion for weighted dependency graphs

Setting: for each n,

- $\{Y_{n,i}, 1 \le i \le N_n\}$ is a family of bounded random variables; $|Y_{n,i}| < M$ a.s.
- we have a weighted dependency graph $\widetilde{L}_n$ with weighted maximal degree $\Delta_n - 1$.
- we set $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.

Theorem (F., 2016)

Assume that $\left(\frac{N_n}{\Delta_n}\right)^{1/s} \frac{\Delta_n}{\sigma_n} \to 0$ *for some integer* s. *Then* X_n *satisfies a CLT.*

A normality criterion for weighted dependency graphs

Setting: for each n,

- $\{Y_{n,i}, 1 \le i \le N_n\}$ is a family of bounded random variables; $|Y_{n,i}| < M$ a.s.
- we have a weighted dependency graph $\widetilde{L}_n$ with weighted maximal degree $\Delta_n - 1$.
- we set $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.

Theorem (F., 2016)

Assume that $\left(\frac{N_n}{\Delta_n}\right)^{1/s} \frac{\Delta_n}{\sigma_n} \to 0$ *for some integer* s*. Then* X_n *satisfies a CLT.*

For triangles in $G(n, M_n)$, $N_n = \binom{n}{3}$, $\Delta_n = O(n)$, while $\sigma_n \asymp n^{3/2}$.

Corollary

Fix p *in* $(0,1)$ *and set* $M_n = p\binom{n}{2}$*. Then* T_n *satisfies a CLT.*

(also true for $n \ll M_n \ll n^2$; originally proved by Janson, 1994).

Applications of weighted dependency graphs

- crossings in random pair-partitions;
- subgraph counts in $G(n, M)$;
- random permutations;
- particles in symmetric simple exclusion process;
- subword counts in Markov chains;
- patterns in multiset permutations*, in set-partitions*;
- spins in Ising model*;
- determinantal point process**.

*in progress with Jehanne Dousse and Marko Thiel. **project
(Some of these applications use a variant of the above normality criterion, which is more technical to state...)

Cumulants

What are (mixed) cumulants?

- The r-th mixed cumulant κ_r of r random variables is a specific r-linear symmetric polynomial in joint moments. Examples:

$$\kappa_1(X) := \mathbb{E}(X), \quad \kappa_2(X,Y) := \mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

$$\kappa_3(X,Y,Z) := \mathbb{E}(XYZ) - \mathbb{E}(XY)\mathbb{E}(Z) - \mathbb{E}(XZ)\mathbb{E}(Y) \\ - \mathbb{E}(YZ)\mathbb{E}(X) + 2\mathbb{E}(X)\mathbb{E}(Y)\mathbb{E}(Z).$$

Notation: $\kappa_\ell(X) := \kappa_\ell(X, \dots, X)$.

What are (mixed) cumulants?

- The r-th mixed cumulant κ_r of r random variables is a specific r-linear symmetric polynomial in joint moments. Examples:

$$\kappa_1(X) := \mathbb{E}(X), \quad \kappa_2(X,Y) := \mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$
$$\kappa_3(X,Y,Z) := \mathbb{E}(XYZ) - \mathbb{E}(XY)\mathbb{E}(Z) - \mathbb{E}(XZ)\mathbb{E}(Y) - \mathbb{E}(YZ)\mathbb{E}(X) + 2\mathbb{E}(X)\mathbb{E}(Y)\mathbb{E}(Z).$$

  Notation: $\kappa_\ell(X) := \kappa_\ell(X,\dots,X)$.

- If a set of variables can be split in two mutually independent sets, then its mixed cumulant vanishes.

What are (mixed) cumulants?

- The r-th mixed cumulant κ_r of r random variables is a specific r-linear symmetric polynomial in joint moments. Examples:

$$\kappa_1(X) := \mathbb{E}(X), \quad \kappa_2(X,Y) := \mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

$$\kappa_3(X,Y,Z) := \mathbb{E}(XYZ) - \mathbb{E}(XY)\mathbb{E}(Z) - \mathbb{E}(XZ)\mathbb{E}(Y) - \mathbb{E}(YZ)\mathbb{E}(X) + 2\mathbb{E}(X)\mathbb{E}(Y)\mathbb{E}(Z).$$

  Notation: $\kappa_\ell(X) := \kappa_\ell(X,\dots,X)$.

- If a set of variables can be split in two mutually independent sets, then its mixed cumulant vanishes.

- If, for each r big enough, we have $\kappa_r(X_n) = o(\mathrm{Var}(X_n)^{r/2})$, then X_n satisfies a CLT. (Janson, 1988)

Sketch of proof of Janson's normality criterion

Setting: for each n,

- $\{Y_{n,i}, 1 \leq i \leq N_n\}$ is a family of bounded random variables; $|Y_{n,i}| < M$ a.s.
- we have a dependency graph L_n with maximal degree $\Delta_n - 1$.
- we set $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.

Sketch of proof of Janson's normality criterion

Setting: for each n,

- $\{Y_{n,i}, 1 \leq i \leq N_n\}$ is a family of bounded random variables; $|Y_{n,i}| < M$ a.s.
- we have a dependency graph L_n with maximal degree $\Delta_n - 1$.
- we set $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.

Fix $r \geq 1$. Then

$$\kappa_r(X_n) = \sum_{i_1,\dots,i_r} \kappa(Y_{n,i_1}, \cdots, Y_{n,i_r}).$$

Sketch of proof of Janson's normality criterion

Setting: for each n,

- $\{Y_{n,i}, 1 \leq i \leq N_n\}$ is a family of bounded random variables; $|Y_{n,i}| < M$ a.s.
- we have a dependency graph L_n with maximal degree $\Delta_n - 1$.
- we set $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.

Fix $r \geq 1$. Then

$$\kappa_r(X_n) = \sum_{i_1,\dots,i_r} \kappa(Y_{n,i_1}, \cdots, Y_{n,i_r}).$$

Each summand is 0, unless, up to reordering, each i_j is a neighbour in L_n of either $i_1,\dots,i_{j-1}$.

Sketch of proof of Janson's normality criterion

Setting: for each n,

- $\{Y_{n,i}, 1 \leq i \leq N_n\}$ is a family of bounded random variables; $|Y_{n,i}| < M$ a.s.
- we have a dependency graph L_n with maximal degree $\Delta_n - 1$.
- we set $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.

Fix $r \geq 1$. Then

$$\kappa_r(X_n) = \sum_{i_1,\dots,i_r} \kappa(Y_{n,i_1}, \cdots, Y_{n,i_r}).$$

Each summand is 0, unless, up to reordering, each i_j is a neighbour in L_n of either $i_1, \dots, i_{j-1}$.
$\rightarrow$ only $(r!)^2\, N_n\, \Delta_n^{r-1}$ non-zero terms, each of which is easily bounded by a constant.

Sketch of proof of Janson's normality criterion

Setting: for each n,

- $\{Y_{n,i}, 1 \leq i \leq N_n\}$ is a family of bounded random variables; $|Y_{n,i}| < M$ a.s.
- we have a dependency graph L_n with maximal degree $\Delta_n - 1$.
- we set $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.

Fix $r \geq 1$. Then

$$\kappa_r(X_n) = \sum_{i_1,\dots,i_r} \kappa(Y_{n,i_1}, \cdots, Y_{n,i_r}).$$

Each summand is 0, unless, up to reordering, each i_j is a neighbour in L_n of either $i_1, \dots, i_{j-1}$.
$\to$ only $(r!)^2\, N_n\, \Delta_n^{r-1}$ non-zero terms, each of which is easily bounded by a constant.

$$|\kappa_r(X_n)| \leq C_r N_n\, \Delta_n^{r-1}. \quad \square$$

Back to weighted dependency graphs

$\rightarrow$ what we should require in the definition of weighted dependency graphs is a bound on mixed cumulants.

Back to weighted dependency graphs

$\rightarrow$ what we should require in the definition of weighted dependency graphs is a bound on mixed cumulants.

Definition (F., 2016)

A weighted graph $\widetilde{L}$ with vertex set A is a weighted dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if, for any $\alpha_1, \dots, \alpha_r$ in A,

$$\left|\kappa(Y_{\alpha_1}, \cdots, Y_{\alpha_r})\right| \leq C_r\, \mathcal{M}\left(\widetilde{L}[\alpha_1, \cdots, \alpha_r]\right).$$

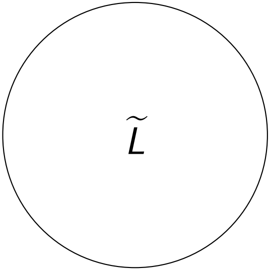
Back to weighted dependency graphs

$\rightarrow$ what we should require in the definition of weighted dependency graphs is a bound on mixed cumulants.

Definition (F., 2016)

A weighted graph $\widetilde{L}$ with vertex set A is a weighted dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if, for any $\alpha_1, \dots, \alpha_r$ in A,

$$\left|\kappa(Y_{\alpha_1}, \cdots, Y_{\alpha_r})\right| \leq C_r \, \mathcal{M}\left(\widetilde{L}[\alpha_1, \cdots, \alpha_r]\right).$$



Back to weighted dependency graphs

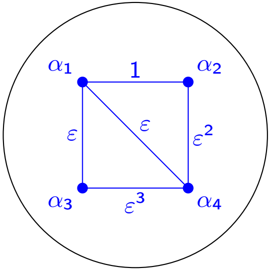
$\rightarrow$ what we should require in the definition of weighted dependency graphs is a bound on mixed cumulants.

Definition (F., 2016)

A weighted graph $\widetilde{L}$ with vertex set A is a weighted dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if, for any $\alpha_1, \dots, \alpha_r$ in A,

$$\left|\kappa(Y_{\alpha_1}, \cdots, Y_{\alpha_r})\right| \leq C_r\, \mathcal{M}\left(\widetilde{L}[\alpha_1, \cdots, \alpha_r]\right).$$

$\widetilde{L}[\alpha_1, \cdots, \alpha_r]$: graph induced by $\widetilde{L}$ on vertices $\alpha_1, \cdots, \alpha_r$.



Back to weighted dependency graphs

$\rightarrow$ what we should require in the definition of weighted dependency graphs is a bound on mixed cumulants.

Definition (F., 2016)

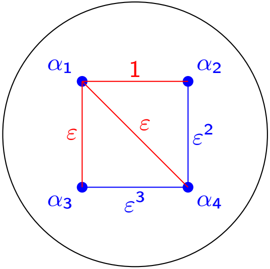
A weighted graph $\widetilde{L}$ with vertex set A is a weighted dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if, for any $\alpha_1, \dots, \alpha_r$ in A,

$$\left|\kappa(Y_{\alpha_1}, \cdots, Y_{\alpha_r})\right| \leq C_r \, \mathcal{M}\left(\widetilde{L}[\alpha_1, \cdots, \alpha_r]\right).$$

$\widetilde{L}[\alpha_1, \cdots, \alpha_r]$: graph induced by $\widetilde{L}$ on vertices $\alpha_1, \cdots, \alpha_r$.

$\mathcal{M}(K)$: Maximum weight of a spanning tree of K.

In the example, $\mathcal{M}\left(\widetilde{L}[\alpha_1, \cdots, \alpha_4]\right) = \varepsilon^2$.



Back to weighted dependency graphs

$\rightarrow$ what we should require in the definition of weighted dependency graphs is a bound on mixed cumulants.

Definition (F., 2016)

A weighted graph $\widetilde{L}$ with vertex set A is a weighted dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if, for any $\alpha_1, \dots, \alpha_r$ in A,

$$\left|\kappa(Y_{\alpha_1}, \cdots, Y_{\alpha_r})\right| \leq C_r \, \mathcal{M}\big(\widetilde{L}[\alpha_1, \cdots, \alpha_r]\big).$$

⚠ This is a simplified version of the definition; some of the applications need a more general but more technical version.

On the normality criterion for weighted dependency graphs

Proof is rather easy, similar as Janson's.
(The reordering is given by Prim's minimum spanning tree algorithm.)

On the normality criterion for weighted dependency graphs

Proof is rather easy, similar as Janson's.
(The reordering is given by Prim's minimum spanning tree algorithm.)

Question

How to prove that something is a weighted dependency graph for a family $\{Y_\alpha, \alpha \in A\}$?

i.e. prove a bound on every cumulant $\kappa(Y_{\alpha_1}, \cdots, Y_{\alpha_r})$.
A priori not easy!

On the normality criterion for weighted dependency graphs

Proof is rather easy, similar as Janson's.
(The reordering is given by Prim's minimum spanning tree algorithm.)

Question

How to prove that something is a weighted dependency graph for a family $\{Y_\alpha, \alpha \in A\}$?

i.e. prove a bound on every cumulant $\kappa(Y_{\alpha_1}, \cdots, Y_{\alpha_r})$.
A priori not easy!

$\rightarrow$ in the next section, we give 3 general tools for that.

Finding weighted dependency graphs

Stability by powers

Setting:

- Let $\{Y_\alpha,\, \alpha \in A\}$ be r.v. with weighted dependency graph $\widetilde{L}$;
- fix an integer $m \geq 2$;
- for a multiset $B = \{\alpha_1, \cdots, \alpha_m\}$ of elements of A, denote

$$\boldsymbol{Y}_B := Y_{\alpha_1} \cdots Y_{\alpha_m}.$$

Stability by powers

Setting:

- Let $\{Y_\alpha, \alpha \in A\}$ be r.v. with weighted dependency graph $\widetilde{L}$;
- fix an integer $m \geq 2$;
- for a multiset $B = \{\alpha_1, \cdots, \alpha_m\}$ of elements of A, denote

$$\boldsymbol{Y}_B := Y_{\alpha_1} \cdots Y_{\alpha_m}.$$

Tool 1

The set of r.v. $\{\boldsymbol{Y}_B\}$ has a weighted dependency graph $\widetilde{L}^m$, where

$$\mathrm{wt}_{\widetilde{L}^m}(\boldsymbol{Y}_B, \boldsymbol{Y}_{B'}) = \max_{\alpha \in B, \alpha' \in B'} \mathrm{wt}_{\widetilde{L}}(Y_\alpha, Y_{\alpha'}).$$

In short: if we have a dependency graph for some variables Y_α, we have also one for monomials in the Y_α.

Back to triangles (1/2)

It is enough to find a weighted dependency graph for edge indicators:

$$Y_e = \mathbf{1}_{e \in G} \quad (e \in \binom{[n]}{2}).$$

(Indeed, triangle indicators are product of edge indicators.)

Back to triangles (1/2)

It is enough to find a weighted dependency graph for edge indicators:

$$Y_e = \mathbf{1}_{e \in G} \quad (e \in \binom{[n]}{2}).$$

(Indeed, triangle indicators are product of edge indicators.)

We need to bound cumulants of the shape

$$\kappa\big(Y_{e_1}, \cdots, Y_{e_r}\big). \tag{1}$$

A priori, there can be repetitions in the sequence $e_1, \cdots, e_r$.

Back to triangles (1/2)

It is enough to find a weighted dependency graph for edge indicators:

$$Y_e = \mathbf{1}_{e \in G} \quad (e \in \binom{[n]}{2}).$$

(Indeed, triangle indicators are product of edge indicators.)

We need to bound cumulants of the shape

$$\kappa\big(Y_{e_1}, \cdots, Y_{e_r}\big). \tag{1}$$

A priori, there can be repetitions in the sequence $e_1, \cdots, e_r$.

Tool 2 (informal version)

In (1), we can replace repeated variables or variables linked by edges of weight 1 by their products

$\rightarrow$ with Bernoulli variables, we can forget repetitions.

Back to triangles (2/2)

We only need to prove

$$\kappa\big(Y_{e_1}, \cdots, Y_{e_r}\big) = O(n^{-2(r-1)}),$$

for distinct edges.

Back to triangles (2/2)

We only need to prove

$$\kappa\big(Y_{e_1},\cdots,Y_{e_r}\big)=O(n^{-2(r-1)}),$$

for distinct edges.

Joint moments are explicit: let $E_n=\binom{n}{2}$,

$$J_\ell := \mathbb{E}(Y_{e_1}\dots Y_{e_\ell})=\frac{\binom{E_n-\ell}{M_n-\ell}}{\binom{E_n}{M_n}}=\frac{(E_n-\ell)!M_n!}{E_n!(M_n-\ell)!}.$$

Example: for $r=3$, we need to prove

$$J_3-3\,J_2\,J_1+2J_1^3=O(n^{-4}).$$

For fixed r, easy to check with a computer algebra system, but not easy to prove for general r.

The multiplicative criterion

Tool 3 (for edges in $G(n, M)$)

The bounds

$$\kappa\big(Y_{e_1}, \cdots, Y_{e_r}\big) = O(n^{-2(r-1)}) \quad \text{(for all } r \geq 1\text{)}$$

are equivalent to

$$\prod_{i=1}^{r} J_i^{-\binom{r}{i}} = 1 + O(n^{-2(r-1)})$$

The multiplicative criterion

Tool 3 (for edges in $G(n, M)$)

The bounds

$$\kappa\left(Y_{e_1}, \cdots, Y_{e_r}\right) = O(n^{-2(r-1)}) \quad \text{(for all } r \geq 1)$$

are equivalent to

$$\prod_{i=1}^{r} J_i^{-\binom{r}{i}} = 1 + O(n^{-2(r-1)})$$

Second statement is much easier to handle:

- it is "multiplicative" in J_i: can be done separately for each factorial factor.
- lots of cancellations in LHS.

$\rightarrow$ form here, quite easy to prove that we have a weighted dependency graph...

Other dependency graphs

Uniform random permutations

Let σ be a uniform random permutation of size n. Set

$$Y_{(i,s)} = \mathbf{1}_{\sigma(i)=s}.$$

Joint moment for distinct $i_1, \cdots, i_r$, $s_1, \cdots, s_r$:

$$\mathbb{E}\big(Y_{i_1,s_1} \cdots Y_{i_r,s_r}\big) = \frac{(n-r)!}{n!}.$$

Uniform random permutations

Let σ be a uniform random permutation of size n. Set

$$Y_{(i,s)} = \mathbf{1}_{\sigma(i)=s}.$$

Joint moment for distinct $i_1, \cdots, i_r$, $s_1, \cdots, s_r$:

$$\mathbb{E}\big(Y_{i_1,s_1} \cdots Y_{i_r,s_r}\big) = \frac{(n-r)!}{n!}.$$

- It is a quotient of factorial factor, so it satisfies the multiplicity criterion.
- Thus we have a weighted dependency graphs for the $Y_{i,s}$;
- and, therefore, also for monomials in $Y_{i,s}$.

☺ No need to do any computation here.

Uniform random permutations

Let σ be a uniform random permutation of size n. Set

$$Y_{(i,s)} = \mathbf{1}_{\sigma(i)=s}.$$

Joint moment for distinct $i_1, \cdots, i_r$, $s_1, \cdots, s_r$:

$$\mathbb{E}\big(Y_{i_1,s_1} \cdots Y_{i_r,s_r}\big) = \frac{(n-r)!}{n!}.$$

- It is a quotient of factorial factor, so it satisfies the multiplicity criterion.
- Thus we have a weighted dependency graphs for the $Y_{i,s}$;
- and, therefore, also for monomials in $Y_{i,s}$.

☺ No need to do any computation here.

$\rightarrow$ gives a bivariate extension of a functional CLT of Janson and Barbour.

Markov chains

Setting:

- Let $(M_i)_{i \geq 0}$ be an irreducible aperiodic Markov chain on a finite space state S;
- Assume M_0 is distributed with the stationary distribution π;
- Set $Y_{i,s} = \mathbf{1}_{M_i = s}$.

Markov chains

Setting:

- Let $(M_i)_{i \geq 0}$ be an irreducible aperiodic Markov chain on a finite space state S;
- Assume M_0 is distributed with the stationary distribution π;
- Set $Y_{i,s} = \mathbf{1}_{M_i = s}$.

Proposition

We have a weighted dependency graph $\widetilde{L}$ *with* $\mathrm{wt}_{\widetilde{L}}\left(\{Y_{i,s}, Y_{j,t}\}\right) = |\lambda_2|^{j-i}$, *where* λ_2 *is the second eigenvalue of the transition matrix.*

$\rightarrow$ CLT for linear statistics $\sum_{i=1}^{N} f(M_i) = \sum_{i,s} f(s) Y_{i,s}$.
Already known (huge literature on the subject).

Markov chains

Setting:

- Let $(M_i)_{i \geq 0}$ be an irreducible aperiodic Markov chain on a finite space state S;
- Assume M_0 is distributed with the stationary distribution π;
- Set $Y_{i,s} = \mathbf{1}_{M_i = s}$.

Proposition

We have a weighted dependency graph $\widetilde{L}$ with $\mathrm{wt}_{\widetilde{L}}\left(\{Y_{i,s}, Y_{j,t}\}\right) = |\lambda_2|^{j-i}$, where λ_2 is the second eigenvalue of the transition matrix.

Corollary (using the stability by product)

We have a weighted dependency graph $\widetilde{L}^m$ for monomials $Y_{i_1,s_1}, \dots, Y_{i_m,s_m}$.

$\rightarrow$ gives a CLT for the number of copies of a given word in $(M_i)_{0 \leq i \leq N}$.
(Answers a question of Bourdon and Vallée.)

Discrete determinantal point processes

Setting: S discrete state space; X random subset of S.

Definition

X is a discrete determinantal point process (DPP) with kernel K if for any distinct $s_1, \dots, s_r$ in S,

$$\mathbb{P}(\{s_1, \dots, s_r\} \subseteq X) = \mathbb{E}\left(\prod_{i=1}^{r} \mathbf{1}_{s_i \in X}\right) = \det\Big(K(s_i, s_j)\Big)_{1 \le i,j \le r}.$$

Discrete determinantal point processes

Setting: S discrete state space; X random subset of S.

Definition

X is a discrete determinantal point process (DPP) with kernel K if for any distinct $s_1, \dots, s_r$ in S,

$$\mathbb{P}(\{s_1, \dots, s_r\} \subseteq X) = \mathbb{E}\left(\prod_{i=1}^{r} \mathbf{1}_{s_i \in X}\right) = \det\Big(K(s_i, s_j)\Big)_{1 \le i,j \le r}.$$

Strange definition (not even clear *a priori* if such a process exists at all), but there are lots of example:

- random Young diagrams, taken with Poissonized Plancherel measure;
- mid-time positions of non-intersecting random walks conditioned to come back to their starting positions.
- eigenvalues of random matrices (continuous DPP);

Discrete determinantal point processes

Setting: S discrete state space; X random subset of S.

Definition

X is a discrete determinantal point process (DPP) with kernel K if for any distinct $s_1, \dots, s_r$ in S,

$$\mathbb{P}(\{s_1, \dots, s_r\} \subseteq X) = \mathbb{E}\left(\prod_{i=1}^{r} \mathbf{1}_{s_i \in X}\right) = \det \Big(K(s_i, s_j)\Big)_{1 \le i,j \le r}.$$

Lemma (Soshnikov, 2000)

If X is a discrete determinantal point process with kernel K, then, for any distinct $s_1, \dots, s_r$ in S,

$$\kappa(\mathbf{1}_{s_1 \in X}, \dots, \mathbf{1}_{s_r \in X}) = \sum_\sigma \varepsilon(\sigma) \prod_i K(s_i, s_{\sigma(i)}),$$

where the sum runs over cyclic permutation in S_r.

Discrete determinantal point processes

Setting: S discrete state space; X random subset of S.

Definition

X is a discrete determinantal point process (DPP) with kernel K if for any distinct $s_1, \dots, s_r$ in S,

$$\mathbb{P}(\{s_1, \dots, s_r\} \subseteq X) = \mathbb{E}\left(\prod_{i=1}^{r} \mathbf{1}_{s_i \in X}\right) = \det\Big(K(s_i, s_j)\Big)_{1 \le i,j \le r}.$$

Soshnikov cumulant formula $\Rightarrow$ for each DPP, we have a weighted dependency graph with weights $K(s_i, s_j)$.

again, CLT for linear statistic is known;
Project: investigate CLT for "multilinear" statistics.

Conclusion

- We provide a general tool to prove CLT for sums of weakly dependent random variables.
- Other examples (d-regular graphs)?
- Can we prove other type of results: speed of convergence, large deviations, ...? (with P.-L. Méliot and A. Nikeghbali, we have such results for usual dependency graphs.)
- A theory of continuous weighted dependency graphs (to handle continuous determinantal point processes)?