Mod-ϕ convergence II: dependency graphs

Valentin Féray
(joint work with Pierre-Loïc Méliot and Ashkan Nikeghbali)

Institut für Mathematik, Universität Zürich

Workshop on "Cumulants, concentration and superconcentration"
Dec. 6th-8th, 2016

Reminder

Definition (uniform control on cumulants)

A sequence (S_n) admits a uniform control on cumulants with DNA (D_n, N_n, A) and limits σ^2 and L if $D_n = o(N_n)$, $N_n \to +\infty$ and

$$\forall r \geq 2, \ \ |\kappa^{(r)}(S_n)| \leq N_n\,(2D_n)^{r-1}\,r^{r-2}\,A^r;$$

$$\frac{\kappa^{(2)}(S_n)}{N_n\,D_n} = (\sigma_n)^2 \to_{n\to\infty} \sigma^2; \qquad \frac{\kappa^{(3)}(S_n)}{N_n\,(D_n)^2} = L_n \to_{n\to\infty} L.$$

This yields mod-Gaussian convergence of a suited renormalization of S_n, hence deviation probability estimate and a bound on the speed on convergence in a CLT.

Today's talk

See a setting where the inequality above holds: dependency graphs and a weighted variant.

Many applications: subgraph counts in random graphs, patterns in random permutations, magnetization in Ising models, linear statistics of Markov chains.

We will also discuss a weaker framework, which is useful to prove CLT, but where we cannot prove mod-Gaussian convergence.

Dependency graphs

Definition (Malyshev, '80, Petrovskaya/Leontovich, '82, Janson, '88)

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 (Malyshev, '80, Petrovskaya/Leontovich, '82, Janson, '88)

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 (Malyshev, '80, Petrovskaya/Leontovich, '82, Janson, '88)

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; $|Y_{n,i}| < A$ a.s.
- we have a dependency graph L_n with maximal degree $D_n - 1$.
- we set $S_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(S_n)$.

Janson's normality criterion

Setting: for each n,

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

Theorem (Janson, 1988)

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

Janson's normality criterion

Setting: for each n,

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

Theorem (Janson, 1988)

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

For triangles, $N_n = \binom{n}{3}$, $D_n = O(n)$, 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...)

A bound for cumulants

Setting: for each n,

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

A bound for cumulants

Setting: for each n,

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

Lemma (Janson, 1988)

$$\kappa_r(S_n) \leq C_r\, N_n\, D_n^{r-1}\, A^r,$$

for some universal constant C_r.

A bound for cumulants

Setting: for each n,

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

Lemma (Janson, 1988)

$$\kappa_r(S_n) \leq C_r \, N_n \, D_n^{r-1} \, A^r,$$

for some universal constant C_r.

Döring and Eichelsbacher, 2012: one can take $C_r = (2e)^r (r!)^3$.

A bound for cumulants

Setting: for each n,

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

Lemma (Janson, 1988)

$$\kappa_r(S_n) \leq C_r\, N_n\, D_n^{r-1}\, A^r,$$

for some universal constant C_r.

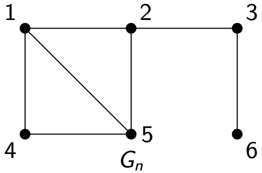
Döring and Eichelsbacher, 2012: one can take $C_r = (2e)^r (r!)^3$.

FMN, 2013-2017: one can take $C_r = 2^{r-1} r^{r-2}$,
i.e. we have the uniform bound on cumulants.

A bound for mixed cumulants

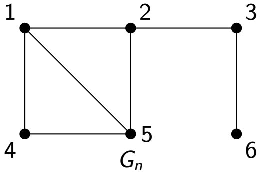
How to bound $\kappa_r(X_n)$?

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



$\kappa(Y_{n,i_1}, \cdots, Y_{n,i_r})$ is the mixed cumulants; multilinear version of cumulants.

A bound for mixed cumulants



How to bound $\kappa_r(X_n)$?

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

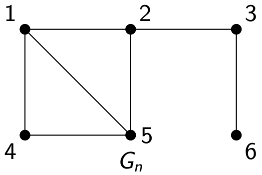
Most terms are zero: $\kappa(Y_{n,i_1}, \cdots, Y_{n,i_r}) = 0$ unless the induced graph $G_n[\{i_1, \dots, i_r\}]$ is connected.

e.g. $\kappa(Y_{n,1}, Y_{n,3}, Y_{n,4}, Y_{n,5}) = 0$

A bound for mixed cumulants

How to bound $\kappa_r(X_n)$?

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



Most terms are zero: $\kappa(Y_{n,i_1}, \cdots, Y_{n,i_r}) = 0$ unless the induced graph $G_n[\{i_1, \dots, i_r\}]$ is connected.

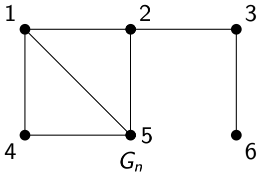
Usual strategy: bound each term $\kappa(Y_{n,i_1}, \cdots, Y_{n,i_r})$ and the number of non-zero terms.

We prove a bound depending on the induced graph $G_n[\{i_1, \dots, i_r\}]$.

A bound for mixed cumulants

How to bound $\kappa_r(X_n)$?

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



Most terms are zero: $\kappa(Y_{n,i_1}, \cdots, Y_{n,i_r}) = 0$ unless the induced graph $G_n[\{i_1, \dots, i_r\}]$ is connected.

Proposition (FMN, 2013-2017)

$$|\kappa(Y_{n,i_1}, \cdots, Y_{n,i_r})| \leq 2^{\ell-1} A^r \, \mathrm{ST}\left(G_n[\{i_1, \dots, i_r\}]\right),$$

where ST(G) *denotes the number of spanning tree of a graph* G.

e.g. $|\kappa(Y_{n,1}, Y_{n,2}, Y_{n,4}, Y_{n,5})| \leq 2^3 A^4 \, 8;$

$|\kappa(Y_{n,1}, Y_{n,2}, Y_{n,3}, Y_{n,4})| \leq 2^3 A^4 \, 1.$

From mixed cumulants to cumulants

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

Thus

$$|\kappa_r(S_n)| \leq \sum_{i_1,\dots,i_r} 2^{r-1} A^r \,\mathrm{ST}\left(G_n[\{i_1,\dots,i_r\}]\right).$$

From mixed cumulants to cumulants

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

Thus

$$|\kappa_r(S_n)| \leq 2^{r-1} A^r \sum_{T \text{ Cayley tree}} \left| \left\{ (i_1, \dots, i_r \text{ s.t. } T \subset G_n[\{i_1, \dots, i_r\}] \right\} \right| .$$

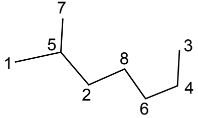
From mixed cumulants to cumulants

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

Thus

$$|\kappa_r(S_n)| \leq 2^{r-1} A^r \sum_{T \text{ Cayley tree}} \left| \left\{ (i_1, \dots, i_r \text{ s.t. } T \subset G_n[\{i_1, \dots, i_r\}] \right\} \right|.$$

Fix a Cayley tree T. For how many lists $i_1, \dots, i_r$ is it contained in $G_n[\{i_1, \dots, i_r\}]$?

$T =$ 

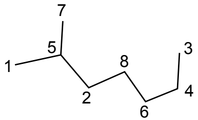
From mixed cumulants to cumulants

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

Thus

$$|\kappa_r(S_n)| \leq 2^{r-1} A^r \sum_{T \text{ Cayley tree}} \left| \{ (i_1, \dots, i_r \text{ s.t. } T \subset G_n[\{i_1, \dots, i_r\}] \} \right| .$$

Fix a Cayley tree T. For how many lists $i_1, \dots, i_r$ is it contained in $G_n[\{i_1, \dots, i_r\}]$?

$T =$ 

- Choose any i_1: N_n choices ;

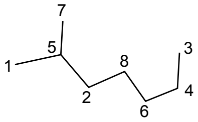
From mixed cumulants to cumulants

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

Thus

$$|\kappa_r(S_n)| \leq 2^{r-1} A^r \sum_{T \text{ Cayley tree}} \left| \left\{ (i_1, \dots, i_r \text{ s.t. } T \subset G_n[\{i_1, \dots, i_r\}] \right\} \right| .$$

Fix a Cayley tree T. For how many lists $i_1, \dots, i_r$ is it contained in $G_n[\{i_1, \dots, i_r\}]$?

$T =$ 

- Choose any i_1: N_n choices ;
- i_5 should be a neighbour of i_1 in G_n (or i_1 itself): D_n choices ;

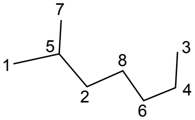
From mixed cumulants to cumulants

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

Thus

$$|\kappa_r(S_n)| \leq 2^{r-1} A^r \sum_{T \text{ Cayley tree}} \left| \left\{ (i_1, \dots, i_r \text{ s.t. } T \subset G_n[\{i_1, \dots, i_r\}] \right\} \right|.$$

Fix a Cayley tree T. For how many lists $i_1, \dots, i_r$ is it contained in $G_n[\{i_1, \dots, i_r\}]$?

$T =$ 

- Choose any i_1: N_n choices ;
- i_5 should be a neighbour of i_1 in G_n (or i_1 itself): D_n choices ;
- i_2 should be a neighbour of i_5 in G_n (or i_5 itself). Also D_n choices.
- ...

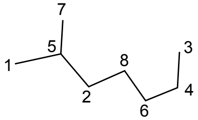
From mixed cumulants to cumulants

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

Thus

$$|\kappa_r(S_n)| \leq 2^{r-1} A^r \sum_{T \text{ Cayley tree}} \left| \left\{ (i_1, \dots, i_r \text{ s.t. } T \subset G_n[\{i_1, \dots, i_r\}] \right\} \right|.$$

Fix a Cayley tree T. For how many lists $i_1, \dots, i_r$ is it contained in $G_n[\{i_1, \dots, i_r\}]$? $N_n D_n^{r-1}$

$T =$ 

- Choose any i_1: N_n choices ;
- i_5 should be a neighbour of i_1 in G_n (or i_1 itself): D_n choices ;
- i_2 should be a neighbour of i_5 in G_n (or i_5 itself). Also D_n choices.
- ...

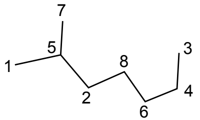
From mixed cumulants to cumulants

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

Thus

$$|\kappa_r(S_n)| \le 2^{r-1} A^r \sum_{T \text{ Cayley tree}} \left| \{ (i_1, \dots, i_r \text{ s.t. } T \subset G_n[\{i_1, \dots, i_r\}] \} \right| .$$

Fix a Cayley tree T. For how many lists $i_1, \dots, i_r$ is it contained in $G_n[\{i_1, \dots, i_r\}]$? $N_n D_n^{r-1}$

$T =$ 

- Choose any i_1: N_n choices ;
- i_5 should be a neighbour of i_1 in G_n (or i_1 itself): D_n choices ;
- i_2 should be a neighbour of i_5 in G_n (or i_5 itself). Also D_n choices.
- ... $|\kappa_r(T_n)| \le 2^{r-1} r^{r-2} N_n D_n^{r-1} A^r$

Setting: for each n,

- $\{Y_{n,i}, 1 \leq i \leq N_n\}$ is a family of bounded random variables; $|Y_{n,i}| < A$ a.s.
- we have a dependency graph L_n with maximal degree $D_n - 1$.
- we set $S_n = \sum_{i=1}^{N_n} Y_{n,i}$, $\sigma_n^2 = \mathrm{Var}(S_n)$ and $L_n^3 = \kappa^3(S_n)$.

Setting: for each n,

- $\{Y_{n,i}, 1 \leq i \leq N_n\}$ is a family of bounded random variables; $|Y_{n,i}| < A$ a.s.
- we have a dependency graph L_n with maximal degree $D_n - 1$.
- we set $S_n = \sum_{i=1}^{N_n} Y_{n,i}$, $\sigma_n^2 = \mathrm{Var}(S_n)$ and $L_n^3 = \kappa^3(S_n)$.

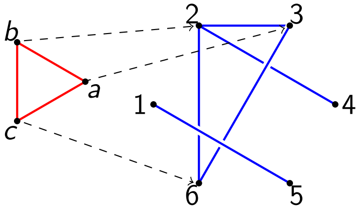
Theorem (FMN, 2017)

Assume that $\frac{\sigma_n^2}{N_n D_n} \to \sigma^2 > 0$. *Then the error term in the CLT for* $\frac{1}{\sigma_n}(S_n - \mathbb{E}(S_n))$ *is* $O(\sqrt{D_n/N_n})$.
If furthermore $\frac{\sigma_n^2}{N_n D_n} \to L^3 \neq 0$, *the normality zone is* $o\big((D_n/N_n)^{1/6}\big)$.

Proof: uniform bound of cumulants + general results from yesterday.

The speed of convergence was already known from Rinott (1994), through Stein's method.

Copies of F in a random graph G

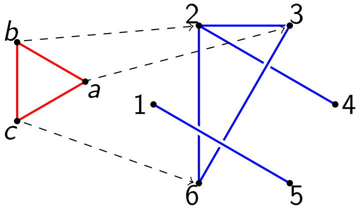


Proposition

The number of copies of a fixed F in $G(n,p)$ (p fixed) admits a uniform control on cumulants with DNA $(n^{|V_G|-2}, n^{|V_G|}, 1)$ and $\sigma^2 > 0$.

Application 1: subgraph counts

Copies of F in a random graph G



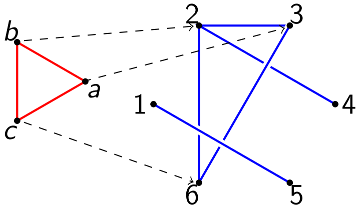
Proposition

The number of copies of a fixed F in $G(n,p)$ (p fixed) admits a uniform control on cumulants with DNA $(n^{|V_G|-2}, n^{|V_G|}, 1)$ and $\sigma^2 > 0$.

Proof: we have a dependency graph with $N_n = n^{|V_G|}$ and $D_n = n^{|V_G|-2}$, which gives us the bound for higher cumulants.
Estimates for second and third cumulants by computation.

Application 1: subgraph counts

Copies of F in a random graph G



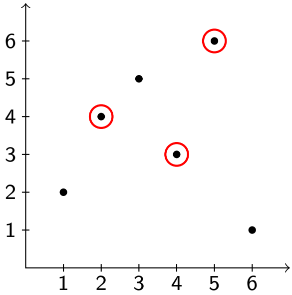
Proposition

The number of copies of a fixed F in $G(n,p)$ (p fixed) admits a uniform control on cumulants with DNA $(n^{|V_G|-2}, n^{|V_G|}, 1)$ and $\sigma^2 > 0$.

☹ When $p = p_n \to 0$, we do not have a good uniform control on cumulants.

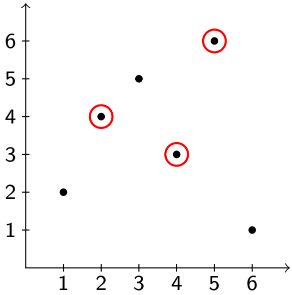
Application 2: patterns in permutation

If τ and π are permutations of size n and k ($k \leq n$), an occurrence of π in τ is an embedding of the diagram of π in the diagram of τ.
Ex: an occurrence of 213 in 245361.



Application 2: patterns in permutation

If τ and π are permutations of size n and k ($k \leq n$), an occurrence of π in τ is an embedding of the diagram of π in the diagram of τ.
Ex: an occurrence of 213 in 245361.



Proposition

The number of occurrences of a fixed π in a uniform random permutation τ of size n admits a uniform control on cumulants with DNA $(n^{k-1}, n^k, 1)$ and $\sigma^2 > 0$.

Proof: again, dependency graph + careful estimate of second/third cumulants.

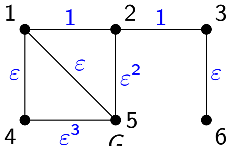
Transition

We will now discuss a weighted analogue of dependency graphs.

Goal: consider sum of pairwise dependent random variables, where the dependencies are asymptotically small.

Uniform weighted dependency graph

Let $S = \sum_{i=1}^{N} Y_i$ as above and let G be an edge weighted graph with vertex set $[N]$ (weights are in $[0,1]$).



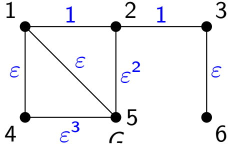
Proposition

Assume that, for any $i_1, \dots, i_r$ in $[N]$, we have

$$\left|\kappa(Y_{i_1}, \dots, Y_{i_r})\right| \le C^r \sum_{T \text{ Cayley tree}} \prod_{\{j,k\} \in E_T} w(\{i_j, i_k\}), \qquad \text{(UWDG)}$$

Uniform weighted dependency graph

Let $S = \sum_{i=1}^{N} Y_i$ as above and let G be an edge weighted graph with vertex set $[N]$ (weights are in $[0,1]$).



Proposition

Assume that, for any $i_1, \dots, i_r$ in $[N]$, we have

$$\left|\kappa(Y_{i_1}, \dots, Y_{i_r})\right| \leq C^r \sum_{T \text{ Cayley tree}} \prod_{\{j,k\} \in E_T} w(\{i_j, i_k\}), \qquad \text{(UWDG)}$$

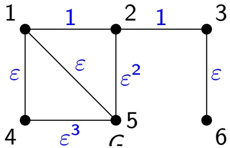
e.g. $\kappa(Y_{n,1}, Y_{n,3}, Y_{n,4}, Y_{n,5}) = 0;$

$|\kappa(Y_{n,1}, Y_{n,2}, Y_{n,3}, Y_{n,4})| \leq C^4\, \varepsilon;$

$|\kappa(Y_{n,1}, Y_{n,2}, Y_{n,4}, Y_{n,5})| \leq C^4\, (\varepsilon^2 + \varepsilon^3 + 3\varepsilon^4 + \varepsilon^5 + 2\varepsilon^6);$

Uniform weighted dependency graph

Let $S = \sum_{i=1}^{N} Y_i$ as above and let G be an edge weighted graph with vertex set $[N]$ (weights are in $[0,1]$).



Proposition

Assume that, for any $i_1, \dots, i_r$ in $[N]$, we have

$$\left|\kappa(Y_{i_1}, \dots, Y_{i_r})\right| \le C^r \sum_{T \text{ Cayley tree}} \prod_{\{j,k\} \in E_T} w(\{i_j, i_k\}), \qquad \text{(UWDG)}$$

then $\left|\kappa_r(S)\right| \le C^r\, N\, D^{r-1}$, where $D-1$ is the maximal weighted degree of the graph.

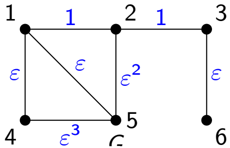
e.g. $\kappa(Y_{n,1}, Y_{n,3}, Y_{n,4}, Y_{n,5}) = 0;$

$|\kappa(Y_{n,1}, Y_{n,2}, Y_{n,3}, Y_{n,4})| \le C^4\, \varepsilon;$

$|\kappa(Y_{n,1}, Y_{n,2}, Y_{n,4}, Y_{n,5})| \le C^4\, (\varepsilon^2 + \varepsilon^3 + 3\varepsilon^4 + \varepsilon^5 + 2\varepsilon^6);$

Uniform weighted dependency graph

Let $S = \sum_{i=1}^{N} Y_i$ as above and let G be an edge weighted graph with vertex set $[N]$ (weights are in $[0,1]$).



Proposition

Assume that, for any $i_1, \dots, i_r$ in $[N]$, we have

$$\left|\kappa(Y_{i_1}, \dots, Y_{i_r})\right| \le C^r \sum_{T \text{ Cayley tree}} \prod_{\{j,k\} \in E_T} w(\{i_j, i_k\}), \qquad \text{(UWDG)}$$

then $\left|\kappa_r(S)\right| \le C^r \, N \, D^{r-1}$, where $D-1$ is the maximal weighted degree of the graph.

Proof: simple adaptation of the case of dependency graphs (which corresponds to edges of G having weights 0 or 1).
When (UWDG) holds, we say that G is a C-uniform weighted dependency graph for $\{Y_i, 1 \le i \le N\}$.

Linear statistics in 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 $Y_i = f_i(M_i)$ for some functions f_i uniformly bounded by B.

Linear statistics in 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 $Y_i = f_i(M_i)$ for some functions f_i uniformly bounded by B.

Proposition (FMN, 2017, based on Saulis, Statelivičius, 1991)

There exists $\varepsilon > 0$ depending on the transition matrix such that the complete graph on $\mathbb{N}_{\geq 0}$ with weights $w(\{s,t\}) = \varepsilon^{|s-t|}$ is a $4B$-uniform weighted dependency graph for the Y_i's.

The maximal weighted degree of the restriction to $[n]$ is constant!

Linear statistics in 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 $Y_i = f_i(M_i)$ for some functions f_i uniformly bounded by B.

Proposition (FMN, 2017, based on Saulis, Statelivičius, 1991)

There exists $\varepsilon > 0$ depending on the transition matrix such that the complete graph on $\mathbb{N}_{\geq 0}$ with weights $w(\{s,t\}) = \varepsilon^{|s-t|}$ is a $4B$-uniform weighted dependency graph for the Y_i's.

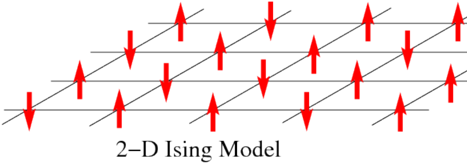
The maximal weighted degree of the restriction to $[n]$ is constant!

$\rightarrow$ deviation and speed of convergence estimates for linear statistics of Markov chains (for speed of convergence, see Bolthausen, 1980).

Ising model

Statistical mechanics model to modelize ferromagnetism.

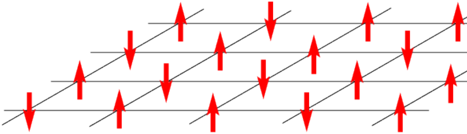
$$P(\omega) \propto \exp\left(\beta \sum_{i \sim j} \omega_i \omega_j + h \sum_i \omega_i\right)$$



Ising model

Statistical mechanics model to modelize ferromagnetism.

$$P(\omega) \propto \exp\left(\beta \sum_{i\sim j} \omega_i\omega_j + h\sum_i \omega_i\right)$$

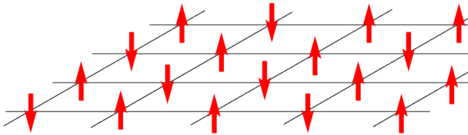

2–D Ising Model

Defined a priori for finite subsets $\Lambda \Subset \mathbb{Z}^d$, but we can take the "thermodynamic limit" $\Lambda \uparrow \mathbb{Z}^d$ (with $+$ boundary conditions).

Ising model

Statistical mechanics model to modelize ferromagnetism.

$$P(\omega) \propto \exp\left(\beta \sum_{i\sim j} \omega_i \omega_j + h \sum_i \omega_i\right)$$



2–D Ising Model

Defined a priori for finite subsets $\Lambda \Subset \mathbb{Z}^d$, but we can take the "thermodynamic limit" $\Lambda \uparrow \mathbb{Z}^d$ (with $+$ boundary conditions).

Proposition (Duneau, Iagolnitzer, Souillard, 1974)

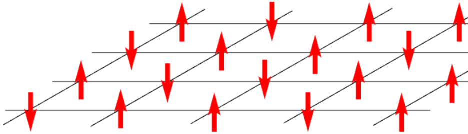
In the termodynamic limit, for $h \neq 0$ or $h = 0$ and sufficiently small β, there exists $\varepsilon(d), C(d) > 0$ such that the complete graph on $\mathbb{N}_{\geq 0}$ with weights $w(\{s,t\}) = \varepsilon(d)^{\|s-t\|}$ is a $C(d)$-uniform weighted dependency graph for the σ_i's.

The same cannot be true for large β ($\beta > \beta_c(d)$).

Ising model

Statistical mechanics model to modelize ferromagnetism.

$$P(\omega) \propto \exp\left(\beta \sum_{i\sim j} \omega_i \omega_j + h \sum_i \omega_i\right)$$


2–D Ising Model

Defined a priori for finite subsets $\Lambda \Subset \mathbb{Z}^d$, but we can take the "thermodynamic limit" $\Lambda \uparrow \mathbb{Z}^d$ (with $+$ boundary conditions).

Proposition (Duneau, Iagolnitzer, Souillard, 1974)

In the termodynamic limit, for $h \neq 0$ or $h = 0$ and sufficiently small β, there exists $\varepsilon(d), C(d) > 0$ such that the complete graph on $\mathbb{N}_{\geq 0}$ with weights $w(\{s,t\}) = \varepsilon(d)^{\|s-t\|}$ is a $C(d)$-uniform weighted dependency graph for the σ_i's.

$\rightarrow$ deviation and speed of convergence estimates for the $\sum \sigma_i$.

(Non-uniform) weighted dependency graphs

Definition (Reminder)

Let C be a positive constant.
We say that G is a C-uniform weighted dependency graph for $\{Y_i, 1 \leq i \leq N\}$ if, for any $i_1, \dots, i_r$ in $[N]$, we have

$$\left|\kappa(Y_{i_1}, \dots, Y_{i_r})\right| \leq C^r \sum_{T \text{ Cayley tree}} \prod_{\{j,k\} \in E_T} w(\{i_j, i_k\}), \qquad \text{(UWDG)}$$

(Non-uniform) weighted dependency graphs

Definition (F. 2016)

Let $\boldsymbol{C} = (C_r)_{r \geq 2}$ be a sequence of positive numbers.
We say that G is a $\boldsymbol{C}$-weighted dependency graph for $\{Y_i, 1 \leq i \leq N\}$ if, for any $i_1, \dots, i_r$ in $[N]$, we have

$$\left|\kappa(Y_{i_1}, \dots, Y_{i_r})\right| \leq C_r \sum_{T \text{ Cayley tree}} \prod_{\{j,k\} \in E_T} w(\{i_j, i_k\}), \qquad \text{(UWDG)}$$

The bounds are now non-uniform in r.

(Non-uniform) weighted dependency graphs

Definition (F. 2016)

Let $\boldsymbol{C} = (C_r)_{r \geq 2}$ be a sequence of positive numbers.
We say that G is a $\boldsymbol{C}$-weighted dependency graph for $\{Y_i, 1 \leq i \leq N\}$ if, for any $i_1, \ldots, i_r$ in $[N]$, we have

$$\left|\kappa(Y_{i_1}, \ldots, Y_{i_r})\right| \leq C_r \max_{T \text{ Cayley tree}} \prod_{\{j,k\} \in E_T} w(\{i_j, i_k\}), \tag{UWDG}$$

We can replace the sum by a max. Spanning trees of maximal weight can be found efficiently (e.g. Prim's algorithm).

(Non-uniform) weighted dependency graphs

Definition (F. 2016)

Let $\boldsymbol{C} = (C_r)_{r \geq 2}$ be a sequence of positive numbers.
We say that G is a $\boldsymbol{C}$-weighted dependency graph for $\{Y_i, 1 \leq i \leq N\}$ if, for any $i_1, \dots, i_r$ in $[N]$, we have

$$\left|\kappa(Y_{i_1}, \dots, Y_{i_r})\right| \leq C_r \max_{T \text{ Cayley tree}} \prod_{\{j,k\} \in E_T} w(\{i_j, i_k\}), \qquad \text{(UWDG)}$$

Lemma

$$\left|\kappa_r\left(\sum Y_i\right)\right| \leq C_r \, N \, D^{r-1}$$

Good to prove CLT, but not for mod-Gaussian convergence.

Applications of non-uniform weighted dependency graphs

- crossings in random pair-partitions;
- subgraph counts in $G(n, M)$;
- random permutations;
- particles in symmetric simple exclusion process;
- linear functional of Markov chain;
- spins in Ising model (with Jehanne Dousse);
- (*) patterns in multiset permutations, in set-partitions (with Marko Thiel);

*in progress. In blue: the ones which are also uniform WDG.

(Some of these applications use a variant of the above definition/lemma, which is more technical to state. . .)

Stability by powers of weighted dependency graphs

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 of weighted dependency graphs

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

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

Proposition (F., 2016)

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.

☹ No analogue for uniform weighted dependency graphs.

Multilinear statistics in 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 $Y_{i,s} = \mathbf{1}_{M_i = s}$.

Multilinear statistics in 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 $\mathrm{wt}_{\widetilde{L}}\left(\{Y_{i,s}, Y_{j,t}\}\right) = |\varepsilon|^{j-i}$.

We already know that: in fact this is a uniform weighted dependency graph.

Multilinear statistics in 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* $\mathrm{wt}_{\widetilde{L}}\left(\{Y_{i,s}, Y_{j,t}\}\right) = |\varepsilon|^{j-i}$.

Corollary (using the stability by product)

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

Multilinear statistics in 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 $\mathrm{wt}_{\widetilde{L}}\left(\{Y_{i,s}, Y_{j,t}\}\right) = |\varepsilon|^{j-i}$.

Corollary (using the stability by product)

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

$\rightarrow$ 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.)

(But no uniform control on cumulants here ☹)

Remark on weighted dependency graphs for Ising models

- (F., Dousse, 2016) Using the stability by products, we can have CLTs for number of occurrences of a given spin pattern (like the number of $+$'s that are surrounded by $-$'s); but no uniform control on cumulants.

Remark on weighted dependency graphs for Ising models

- (F., Dousse, 2016) Using the stability by products, we can have CLTs for number of occurrences of a given spin pattern (like the number of $+$'s that are surrounded by $-$'s); but no uniform control on cumulants.
- For $h = 0$ and sufficiently large β, there is also a weighted dependency graph for spins (Malyshev, Minlos, 1991, F., Dousse, 2016).
  But, recall that in this case, we can prove that there is no uniform control on cumulants and hence no uniform weighted dependency graph.

Conclusion

- Dependency graphs provide uniform bounds on cumulants (lots of examples, including subgraph counts in $G(n,p)$ for fixed p, pattern occurrences in random permutations);

Conclusion

- Dependency graphs provide uniform bounds on cumulants (lots of examples, including subgraph counts in $G(n,p)$ for fixed p, pattern occurrences in random permutations);
- There is a weighted variant which include linear statistics of Markov chain, magnetization in Ising model (except at low temperature);

Conclusion

- Dependency graphs provide uniform bounds on cumulants (lots of examples, including subgraph counts in $G(n, p)$ for fixed p, pattern occurrences in random permutations);
- There is a weighted variant which include linear statistics of Markov chain, magnetization in Ising model (except at low temperature);
- If you relax assumptions to only get a CLT and not a uniform control on cumulants, the weighted version applies to even more examples;

Conclusion

- Dependency graphs provide uniform bounds on cumulants (lots of examples, including subgraph counts in $G(n,p)$ for fixed p, pattern occurrences in random permutations);
- There is a weighted variant which include linear statistics of Markov chain, magnetization in Ising model (except at low temperature);
- If you relax assumptions to only get a CLT and not a uniform control on cumulants, the weighted version applies to even more examples;
- We would really like to get uniform bounds on cumulants for these extra examples. . .