Asymptotic normality via (weighted) dependency graphs

Valentin Féray

CNRS, Université de Lorraine
Institut Élie Cartan de Lorraine, Nancy

Rencontres de probabilités,
Rouen, October 20th, 2021

What is this talk about ?

General problem: A sequence of random variables X_n is asymptotically normal, i.e.

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

How to prove that a sequence is asymptotically normal?

What is this talk about ?

General problem: A sequence of random variables X_n is asymptotically normal, i.e.

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

How to prove that a sequence is asymptotically normal?

A powerful tool: analytic methods, based on the (bivariate/probability) generating functions of the sequence.

Problem: we do not always know how to compute generating functions.

What is this talk about ?

General problem: A sequence of random variables X_n is asymptotically normal, i.e.

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

How to prove that a sequence is asymptotically normal?

Other standard tool: moment (or cumulant) methods.

Today: (weighted) dependency graphs, based on cumulants and independence (or weak dependencies) between variables.

What is this talk about ?

General problem: A sequence of random variables X_n is asymptotically normal, i.e.

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

How to prove that a sequence is asymptotically normal?

Other standard tool: moment (or cumulant) methods.

Today: (weighted) dependency graphs, based on cumulants and independence (or weak dependencies) between variables.

Various examples of applications: occurrences of patterns in combinatorial objects or statistical physics models, length of nearest neighbour graphs of Poisson point processes, ...

Outline of the talk

1. Dependency graphs
   - A motivating example: substrings in random words
   - An asymptotic normality criterion
   - Substructure counts in graphs and permutations
   - Lengths of nearest neighbour graphs
2. Weighted dependency graphs
   - Definition and an extended normality criterion
   - Back to subwords and subgraphs: Markovian texts and $G(n,M)$
   - Patterns in set-partitions
   - Applications in statistical physics

Transition

1. Dependency graphs
   - A motivating example: substrings in random words
   - An asymptotic normality criterion
   - Substructure counts in graphs and permutations
   - Lengths of nearest neighbour graphs
2. Weighted dependency graphs
   - Definition and an extended normality criterion
   - Back to subwords and subgraphs: Markovian texts and $G(n, M)$
   - Patterns in set-partitions
   - Applications in statistical physics

Substrings in random words (1/2)

(following Flajolet, Guivarc'h, Szpankowski, and Vallée, '01)

Let $\boldsymbol{w}$ be a random word of size n with independent (identically distributed) letters taken in a finite alphabet $\mathscr{A}$.

Fix a word u, called "pattern" of length ℓ.

An occurrence of u in w is a ℓ-tuple $i_1 < \cdots < i_\ell$ s.t. $w_{i_1} = u_1, \dots, w_{i_\ell} = u_\ell$.

Example: two occurrences of aab in $w = a\underline{a}bb\underline{a}baa\underline{b}$ (one in blue, one underlined)

(Variants: consecutive occurrences, allowing gaps of given lengths).

Substrings in random words (1/2)

(following Flajolet, Guivarc'h, Szpankowski, and Vallée, '01)

Let $\boldsymbol{w}$ be a random word of size n with independent (identically distributed) letters taken in a finite alphabet $\mathscr{A}$.

Fix a word u, called "pattern" of length ℓ.

An occurrence of u in w is a ℓ-tuple $i_1 < \cdots < i_\ell$ s.t. $w_{i_1} = u_1, \ldots, w_{i_\ell} = u_\ell$.

Example: two occurrences of aab in $w = a\underline{a}bb\underline{a}baa\underline{b}$ (one in blue, one underlined)

Question

Asymptotic behaviour of the number X_n of occurrences of u in $\boldsymbol{w}$?

Motivations: intrusion detection in computer science, discovering meaningful strings of DNA, ...

Substrings in random words (2/2)

Theorem (FGSV, '01)

We have

$$\mathbb{E}[X_n] \sim C_1 n^{\ell}, \qquad \mathrm{Var}[X_n] = C_2 n^{2\ell-1} + O(n^{2\ell-2}),$$

where $C_1 > 0$ and $C_2 \geq 0$ are computable constants.
Moreover, if $C_2 > 0$, then X_n is asymptotically normal.

Substrings in random words (2/2)

Theorem (FGSV, '01)

We have

$$\mathbb{E}[X_n] \sim C_1 n^{\ell}, \qquad \mathrm{Var}[X_n] = C_2 n^{2\ell-1} + O(n^{2\ell-2}),$$

where $C_1 > 0$ and $C_2 \geq 0$ are computable constants.
Moreover, if $C_2 > 0$, then X_n is asymptotically normal.

The proof of asymptotic normality uses the method of moments.

I will sketch it using cumulants and dependency graphs (essentially the same proof, but presented differently, and in a general context).

Substrings in random words (2/2)

Theorem (FGSV, '01)

We have

$$\mathbb{E}[X_n] \sim C_1 n^{\ell}, \qquad \mathrm{Var}[X_n] = C_2 n^{2\ell-1} + O(n^{2\ell-2}),$$

where $C_1 > 0$ and $C_2 \geq 0$ are computable constants.
Moreover, if $C_2 > 0$, then X_n is asymptotically normal.

The proof of asymptotic normality uses the method of moments.

I will sketch it using cumulants and dependency graphs (essentially the same proof, but presented differently, and in a general context).

Notation: for $I \subseteq [n]$, $|I| = \ell$, set $Y_I = \mathbf{1}[u \text{ occurs at position } I \text{ in } \boldsymbol{w}]$.
Then $X_n = \sum_{I \in \binom{[n]}{\ell}} Y_I$.

Transition

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 the following holds for any $A_1, A_2 \subset A$:

$$\begin{array}{c}\text{there is no edge}\\ \text{between } A_1 \text{ and } A_2\end{array} \Longrightarrow \begin{array}{c}\{Y_\alpha, \alpha \in A_1\} \text{ and } \{Y_\alpha, \alpha \in A_2\}\\ \text{are independent}\end{array}$$

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 the following holds for any $A_1, A_2 \subset A$:

$$\begin{array}{c}\text{there is no edge}\\ \text{between } A_1 \text{ and } A_2\end{array} \Longrightarrow \begin{array}{c}\{Y_\alpha, \alpha \in A_1\} \text{ and } \{Y_\alpha, \alpha \in A_2\}\\ \text{are independent}\end{array}$$

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

Example

Consider our random word problem. Let $A = \binom{[n]}{\ell}$ and

$$\{I_1, I_2\} \in E_L \text{ iff } I_1 \cap I_2 \neq \emptyset.$$

Then L is a dependency graph for the family $\{Y_I, I \in \binom{[n]}{\ell}\}$.

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 the following holds for any $A_1, A_2 \subset A$:

$$\begin{array}{c}\text{there is no edge}\\ \text{between } A_1 \text{ and } A_2\end{array} \implies \begin{array}{c}\{Y_\alpha, \alpha \in A_1\} \text{ and } \{Y_\alpha, \alpha \in A_2\}\\ \text{are independent}\end{array}$$

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

Example

Note: L is regular of degree $\mathcal{O}(n^{\ell-1})$

Consider our random word problem. Let $A = \binom{[n]}{\ell}$ and

$$\{I_1, I_2\} \in E_L \text{ iff } I_1 \cap I_2 \neq \emptyset.$$

Then L is a dependency graph for the family $\{Y_I, I \in \binom{[n]}{\ell}\}$.

Janson's normality criterion

Setting: for each n,

- $\{Y_{n,i}, 1 \le i \le N_n\}$ is a family of bounded random variables; $|Y_{n,i}| < M_n$ a.s.
- we have a dependency graph L_n with maximal degree $D_n - 1$.
- 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,

- $\{Y_{n,i}, 1 \le i \le N_n\}$ is a family of bounded random variables; $|Y_{n,i}| < M_n$ a.s.
- we have a dependency graph L_n with maximal degree $D_n - 1$.
- 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 $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} M_n \to 0$ *for some integer* s.
Then X_n *is asymptotically normal.*

Janson's normality criterion

Setting: for each n,

- $\{Y_{n,i}, 1 \le i \le N_n\}$ is a family of bounded random variables; $|Y_{n,i}| < M_n$ a.s.
- we have a dependency graph L_n with maximal degree $D_n - 1$.
- 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 $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} M_n \to 0$ *for some integer* s.
Then X_n *is asymptotically normal.*

Example: For occurrences of u in $\boldsymbol{w}$, we have

$$M_n = 1,\ N_n = \Theta(n^{\ell}),\ D_n = \Theta(n^{\ell-1}) \text{ and } \sigma_n = \Theta(n^{\ell-1/2}),$$

implying asymptotic normality (assuming the variance estimates!).

Janson's normality criterion

Setting: for each n,

- $\{Y_{n,i}, 1 \le i \le N_n\}$ is a family of bounded random variables; $|Y_{n,i}| < M_n$ a.s.
- we have a dependency graph L_n with maximal degree $D_n - 1$.
- 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 $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} M_n \to 0$ *for some integer* s.
Then X_n *is asymptotically normal.*

In roughly the same setting (when $s = 3$), we also have bounds on the speed of convergence and deviation estimates: (see Baldi, Rinott, '89, Rinott, '94 and F., Méliot, Nikeghbali, '16, '17).

Main tool in the proof: (mixed) cumulants

- Definition: mixed cumulants are multilinear functionals defined by

$$\kappa_r(X_1,\dots,X_r) = [t_1 \cdots t_r] \log\left(\mathbb{E}\left[\exp\left(\sum_{j=1}^r t_j X_j\right)\right]\right).$$

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

Main tool in the proof: (mixed) cumulants

- Definition: mixed cumulants are multilinear functionals defined by

$$\kappa_r(X_1,\dots,X_r) = [t_1 \cdots t_r] \log\left(\mathbb{E}\left[\exp\left(\sum_{j=1}^r t_j X_j\right)\right]\right).$$

  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.

Main tool in the proof: (mixed) cumulants

- Definition: mixed cumulants are multilinear functionals defined by

$$\kappa_r(X_1,\dots,X_r) = [t_1\cdots t_r]\log\left(\mathbb{E}\left[\exp\left(\sum_{j=1}^r t_j X_j\right)\right]\right).$$

  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.

- Let $\sigma_n = \sqrt{\mathrm{Var}(X_n)}$. If, for some $s \geq 3$ and any $r \geq s$, we have $\kappa_r(X_n) = o(\sigma_n^r)$, then X_n is asymptotically normal. (Janson, 1988)

Sketch of proof of Janson's normality criterion

Setting: for each n,

- $\{Y_{n,i}, 1 \le i \le N_n\}$ is a family of bounded r.v.; $|Y_{n,i}| < M_n$ a.s.
- we have a dependency graph L_n with maximal degree $D_n - 1$.
- we set $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.
- we assume $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} M_n \to 0$ for some $s \ge 3$.

Sketch of proof of Janson's normality criterion

Setting: for each n,

- $\{Y_{n,i}, 1 \le i \le N_n\}$ is a family of bounded r.v.; $|Y_{n,i}| < M_n$ a.s.
- we have a dependency graph L_n with maximal degree $D_n - 1$.
- we set $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.
- we assume $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} M_n \to 0$ for some $s \ge 3$.

Fix $r \ge 1$. 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,

- $\{Y_{n,i}, 1 \le i \le N_n\}$ is a family of bounded r.v.; $|Y_{n,i}| < M_n$ a.s.
- we have a dependency graph L_n with maximal degree $D_n - 1$.
- we set $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.
- we assume $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} M_n \to 0$ for some $s \ge 3$.

Fix $r \ge 1$. 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 the induced graph $L_n[i_1, \cdots, i_r]$ is connected.

Sketch of proof of Janson's normality criterion

Setting: for each n,

- $\{Y_{n,i}, 1 \le i \le N_n\}$ is a family of bounded r.v.; $|Y_{n,i}| < M_n$ a.s.
- we have a dependency graph L_n with maximal degree $D_n - 1$.
- we set $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.
- we assume $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} M_n \to 0$ for some $s \ge 3$.

Fix $r \ge 1$. 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 of either i_1, . . . , or i_{j-1}.

Sketch of proof of Janson's normality criterion

Setting: for each n,

- $\{Y_{n,i}, 1 \le i \le N_n\}$ is a family of bounded r.v.; $|Y_{n,i}| < M_n$ a.s.
- we have a dependency graph L_n with maximal degree $D_n - 1$.
- we set $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.
- we assume $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} M_n \to 0$ for some $s \ge 3$.

Fix $r \ge 1$. 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 of either i_1, ..., or i_{j-1}. We have $r!$ choices for the reordering, N_n choices for i_1, D_n choices for i_2, $2D_n$ choices for i_3, ...

$\to$ at most $(r!)^2 N_n D_n^{r-1}$ non-zero terms, each of which is bounded by $C_r M_n^r$.

Sketch of proof of Janson's normality criterion

Setting: for each n,

- $\{Y_{n,i}, 1 \le i \le N_n\}$ is a family of bounded r.v.; $|Y_{n,i}| < M_n$ a.s.
- we have a dependency graph L_n with maximal degree $D_n - 1$.
- we set $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.
- we assume $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} M_n \to 0$ for some $s \ge 3$.

Fix $r \ge 1$. Then

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

$\to$ at most $(r!)^2 N_n D_n^{r-1}$ non-zero terms, each of which is bounded by $C_r M_n^r$.

$$\begin{aligned} |\kappa_r(X_n)| &\le C_r (r!)^2 N_n D_n^{r-1} M_n^r \\ &= o(\sigma_n^r) \qquad \text{(for } r \ge s\text{, using the assumption)} \quad \square \end{aligned}$$

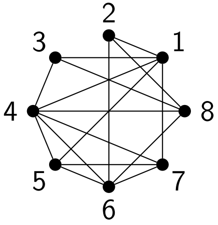
Transition

1. Dependency graphs
   - A motivating example: substrings in random words
   - An asymptotic normality criterion
   - Substructure counts in graphs and permutations
   - Lengths of nearest neighbour graphs
2. Weighted dependency graphs
   - Definition and an extended normality criterion
   - Back to subwords and subgraphs: Markovian texts and $G(n, M)$
   - Patterns in set-partitions
   - Applications in statistical physics

Triangle counts in Erdős-Rényi random graphs (1/2)

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;

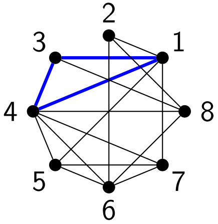


Example : $n=8, p=1/2$

Triangle counts in Erdős-Rényi random graphs (1/2)

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

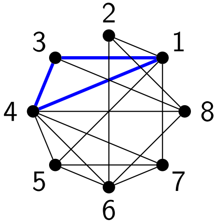
Question

Fix $p \in (0;1)$. Is the number of triangles T_n asymptotically normal?

Triangle counts in Erdős-Rényi random graphs (1/2)

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 $p \in (0;1)$. Is the number of triangles T_n asymptotically normal?

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

Triangle counts in Erdős-Rényi random graphs (2/2)

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 $\{Y_\Delta, \Delta \in \binom{[n]}{3}\}$.

Triangle counts in Erdős-Rényi random graphs (2/2)

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 $\{Y_\Delta, \Delta \in \binom{[n]}{3}\}$.

We have (for fixed p)

$$M_n = 1,\, N_n = \binom{n}{3},\, D_n = \mathcal{O}(n), \text{ while } \sigma_n = \Theta(n^2).$$

(The variance estimates is easily obtained by expanding $\mathrm{Var}(\sum Y_\Delta)$.)

Triangle counts in Erdős-Rényi random graphs (2/2)

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 $\{Y_\Delta, \Delta \in \binom{[n]}{3}\}$.

We have (for fixed p)

$$M_n = 1, N_n = \binom{n}{3}, D_n = \mathcal{O}(n), \text{ while } \sigma_n = \Theta(n^2).$$

(The variance estimates is easily obtained by expanding $\mathrm{Var}(\sum Y_\Delta)$.)

Janson's assumption is fulfilled for $s = 3$.
$\Rightarrow$ T_n is asymptotically normal.
(known at least since Rucinsky, 1988)

Triangle counts in Erdős-Rényi random graphs (2/2)

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 $\{Y_\Delta, \Delta \in \binom{[n]}{3}\}$.

We have (for fixed p)

$$M_n = 1,\, N_n = \binom{n}{3},\, D_n = \mathcal{O}(n), \text{ while } \sigma_n = \Theta(n^2).$$

(The variance estimates is easily obtained by expanding $\mathrm{Var}(\sum Y_\Delta)$.)

Janson's assumption is fulfilled for $s = 3$.
$\Rightarrow$ T_n is asymptotically normal.
(known at least since Rucinsky, 1988)

Note: this generalizes to $p = p_n \gg n^{-1}$ and other subgraph counts, using a more involved normality criterion.

Pattern occurrences in uniform random permutations (1/3)

Definition

An occurrence of a pattern τ in σ is a subsequence $\sigma_{i_1} \dots \sigma_{i_k}$ that is order-isomorphic to τ, *i.e.* $\sigma_{i_s} < \sigma_{i_t} \Leftrightarrow \tau_s < \tau_t$.

Examples of occurrences of 213:

245361

82346175

Question

Fix a pattern π. What is the asymptotic behaviour of the number X_n^π of occurrences of π in a uniform random permutation $\boldsymbol{\sigma}$ of size n?

Again we write $X_n^\pi = \sum_{I \in \binom{[n]}{\ell}} Y_I$,

where $Y_I = \mathbf{1}[\pi \text{ occurs at the set of position } I \text{ in } \boldsymbol{\sigma}]$.

Pattern occurrences in uniform random permutations (2/3)

- Recall that a uniform random permutation $\boldsymbol{\sigma}$ can be obtained by standardizing a sequence of i.i.d. continuous random variables $U_1, \dots, U_n$: i.e. $\boldsymbol{\sigma}_i$ is the rank of U_i in the set $\{U_1, \dots, U_n\}$.

Pattern occurrences in uniform random permutations (2/3)

- Recall that a uniform random permutation $\boldsymbol{\sigma}$ can be obtained by standardizing a sequence of i.i.d. continuous random variables $U_1, \dots, U_n$: i.e. $\boldsymbol{\sigma}_i$ is the rank of U_i in the set $\{U_1, \dots, U_n\}$.

- With this construction, Y_I depends only on $(U_i, i \in I)$: e.g. for $\pi = 132$,
$$Y_I = \mathbf{1}\left[\boldsymbol{\sigma}_{i_1} < \boldsymbol{\sigma}_{i_3} < \boldsymbol{\sigma}_{i_2}\right] = \mathbf{1}\left[U_{i_1} < U_{i_3} < U_{i_2}\right].$$

Pattern occurrences in uniform random permutations (2/3)

- Recall that a uniform random permutation $\boldsymbol{\sigma}$ can be obtained by standardizing a sequence of i.i.d. continuous random variables $U_1, \dots, U_n$: i.e. $\boldsymbol{\sigma}_i$ is the rank of U_i in the set $\{U_1, \dots, U_n\}$.
- With this construction, Y_I depends only on $(U_i, i \in I)$: e.g. for $\pi = 132$,
$$Y_I = \mathbf{1}[\boldsymbol{\sigma}_{i_1} < \boldsymbol{\sigma}_{i_3} < \boldsymbol{\sigma}_{i_2}] = \mathbf{1}[U_{i_1} < U_{i_3} < U_{i_2}].$$
- Therefore the graph L with vertex set $\binom{[n]}{\ell}$ and
$$I_1 \sim_L I_2 \; \Leftrightarrow I_1 \cap I_2 \neq \emptyset$$
is a dependency graph for the family $\{Y_I, I \in \binom{[n]}{\ell}\}$.

Pattern occurrences in uniform random permutations (3/3)

Can we apply Janson's criterion?

$$M_n = 1,\, N_n = \Theta(n^{\ell}),\, D_n = \mathcal{O}(n^{\ell-1}),\, \sigma_n = \Theta(n^{\ell-1/2}).$$

Janson's criterion is fulfilled for $s = 3$:

$\longrightarrow X_n^{\pi} = \sum_{I \in \binom{[n]}{\ell}} Y_I$ is asymptotically normal
(Janson–Nakamura–Zeilberger '15).

Pattern occurrences in uniform random permutations (3/3)

Can we apply Janson's criterion?

$$M_n = 1,\, N_n = \Theta(n^{\ell}),\, D_n = \mathcal{O}(n^{\ell-1}),\, \sigma_n = \Theta(n^{\ell-1/2}).$$

Janson's criterion is fulfilled for $s = 3$:

$\longrightarrow X_n^{\pi} = \sum_{I \in \binom{[n]}{\ell}} Y_I$ is asymptotically normal
(Janson–Nakamura–Zeilberger '15).

(the variance estimates is not trivial;
Bóna '10, Dimitrov–Khare '21: direct proof for the monotone/general case,
Janson–Nakamura–Zeilberger '15: proof using U-statistics for all patterns,
Hofer '18/F. '19: alternative proof using the law of total variance and extending to vincular patterns/patterns in multiset permutations,
Janson '21: U-statistics approach to the vincular pattern case).

Transition

1. Dependency graphs
   - A motivating example: substrings in random words
   - An asymptotic normality criterion
   - Substructure counts in graphs and permutations
   - Lengths of nearest neighbour graphs

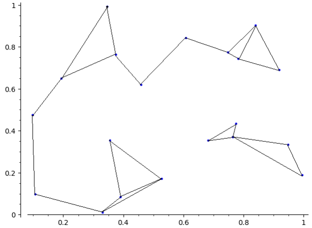
2. Weighted dependency graphs
   - Definition and an extended normality criterion
   - Back to subwords and subgraphs: Markovian texts and $G(n, M)$
   - Patterns in set-partitions
   - Applications in statistical physics

k-nearest neighbour graphs: the problem and its history

Consider a Poisson point process of points in the unit square $[0,1]^2$ of intensity n.

Fix $k \geq 1$. Let $\boldsymbol{G}_n^{(k)}$ be its k-nearest neighbour graph: each point is connected to the k nearest points.

Example with 20 points and $k = 2$:



Question

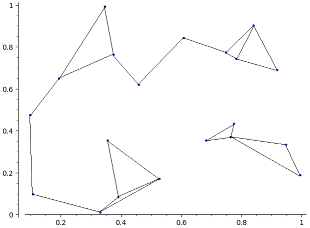
Asymptotics behaviour of the total length X_n of $\boldsymbol{G}_n^{(k)}$?

k-nearest neighbour graphs: the problem and its history

Consider a Poisson point process of points in the unit square $[0,1]^2$ of intensity n.

Fix $k \geq 1$. Let $\boldsymbol{G}_n^{(k)}$ be its k-nearest neighbour graph: each point is connected to the k nearest points.

Example with 20 points and $k = 2$:



Question

Asymptotics behaviour of the total length X_n of $\boldsymbol{G}_n^{(k)}$?

Miles, '70: $\mathbb{E}[X_n] \sim C_k n^{1/2}$, for some explicit C_k.

Bickel, Breiman, '83: for $k = 1$, X_n is asymptotically normal.

Avram, Bertsimas, '93: for any $k \geq 1$, X_n is asymptotically normal (and analogue results for the length of Voronoi diagram and of Delaunay triangulation).

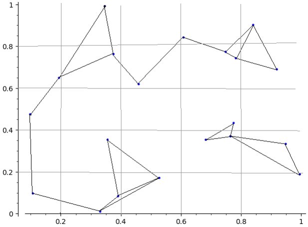
k-nearest neighbours: proof of asymptotic normality (1/2)

(following Avram & Bertsimas, '93)

Set $m = \sqrt{\frac{n}{\log(n)}}$ and divide the square $[0,1]^2$ into m^2 *boxes*. Write

$$X_n = \sum_{1 \le i,j \le m} Y_{i,j},$$

where $Y_{i,j}$ is the length of the graph in box (i,j).



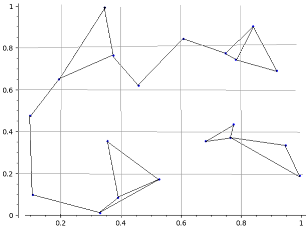
k-nearest neighbours: proof of asymptotic normality (1/2)

(following Avram & Bertsimas, '93)

Set $m = \sqrt{\frac{n}{\log(n)}}$ and divide the square $[0,1]^2$ into m^2 *boxes*. Write

$$X_n = \sum_{1 \le i,j \le m} Y_{i,j},$$

where $Y_{i,j}$ is the length of the graph in box (i,j).



The number of points in each cube is Poisson(λ), where $\lambda := n/m^2 \sim \log(n)$.

Lemma

With probability tending to 1, each box contains at least one point and at most $e\lambda$ points.

(We call A_n this event.)

k-nearest neighbours: proof of asymptotic normality (2/2)

Conditionally on A_n,

- there is no edge in $\boldsymbol{G}_n^{(k)}$ spanning over more than $\sqrt{k}+1$ boxes;
- thus $Y_{i,j}$ and $Y_{i'j'}$ are independent unless $\|(i,j)-(i',j')\|_1 \le 2\sqrt{k}+2$;
- we have a dependency graph of bounded degree for the family $\{Y_{i,j}, 1 \le i,j \le m\}$.

k-nearest neighbours: proof of asymptotic normality (2/2)

Conditionally on A_n,

- there is no edge in $\boldsymbol{G}_n^{(k)}$ spanning over more than $\sqrt{k}+1$ boxes;
- thus $Y_{i,j}$ and $Y_{i'j'}$ are independent unless $\|(i,j)-(i',j')\|_1 \le 2\sqrt{k}+2$;
- we have a dependency graph of bounded degree for the family $\{Y_{i,j}, 1 \le i,j \le m\}$.

Can we apply Janson's criterion? $N_n = m^2 = \widetilde{\mathcal{O}}(n)$, $D_n = \mathcal{O}(1)$,

- $|Y_{i,j}| \le M_n$ with $M_n = \mathcal{O}(\lambda m^{-1}) = \widetilde{\mathcal{O}}(n^{-1/2})$
  (since there are at most $e\lambda$ points in each box, there are at most $\mathcal{O}(\lambda)$ edges, each of length at most $\mathcal{O}(m^{-1/2})$);
- $\sigma_n \ge \Theta(1)$ (tricky argument).

Notation: $\widetilde{\mathcal{O}}$ is $\mathcal{O}$ up to logarithmic factors.

k-nearest neighbours: proof of asymptotic normality (2/2)

Conditionally on A_n,

- there is no edge in $\boldsymbol{G}_n^{(k)}$ spanning over more than $\sqrt{k}+1$ boxes;
- thus $Y_{i,j}$ and $Y_{i'j'}$ are independent unless $\|(i,j)-(i',j')\|_1 \le 2\sqrt{k}+2$;
- we have a dependency graph of bounded degree for the family $\{Y_{i,j}, 1 \le i,j \le m\}$.

Can we apply Janson's criterion? $N_n = m^2 = \widetilde{\mathcal{O}}(n)$, $D_n = \mathcal{O}(1)$,

- $|Y_{i,j}| \le M_n$ with $M_n = \mathcal{O}(\lambda m^{-1}) = \widetilde{\mathcal{O}}(n^{-1/2})$ (since there are at most $e\lambda$ points in each box, there are at most $\mathcal{O}(\lambda)$ edges, each of length at most $\mathcal{O}(m^{-1/2})$);
- $\sigma_n \ge \Theta(1)$ (tricky argument).

Janson's assumption is fulfilled for $s=3$. Thus X_n is asymptotically normal, conditionally on A_n. Since $\mathbb{P}[A_n] \to 1$, X_n is asymptotically normal, unconditionally.

Transition

Motivation: models with "weak dependencies"

In many models, we do not have independence, but only *weak dependencies*:

- subword occurrences in a text generated by a Markovian source;
- subgraph counts in Erdős-Rényi random graphs $G(n,M)$ ($G(n,M)$: fixed number M of edges);
- number of exceedances (i s.t. $\sigma(i) \geq i$) in a uniform random permutation;
- patterns in other combinatorial objects, such as multiset permutations, set partitions, ...;
- some statistical physics models, stationary distribution of symmetric simple exclusion process and Ising model.

Goal: extend Janson's normality criterion, to cover the above frameworks.

Weighted dependency graphs

We use weighted graphs, i.e. graphs with a weight in $[0,1]$ on each edge (weight $0 \equiv$ no edge).

Definition (F., '18)

Fix $\boldsymbol{C} = (C_r)_{r \geq 1}$. A weighted graph $\widetilde{L}$ with vertex set A is a $\boldsymbol{C}$-weighted dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if, for any $\alpha_1, \dots, \alpha_r$ in A,

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

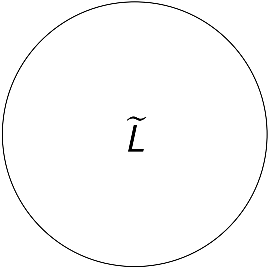
Weighted dependency graphs

We use weighted graphs, i.e. graphs with a weight in $[0,1]$ on each edge (weight $0 \equiv$ no edge).

Definition (F., '18)

Fix $\boldsymbol{C} = (C_r)_{r \geq 1}$. A weighted graph $\widetilde{L}$ with vertex set A is a $\boldsymbol{C}$-weighted dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if, for any $\alpha_1, \dots, \alpha_r$ in A,

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



Weighted dependency graphs

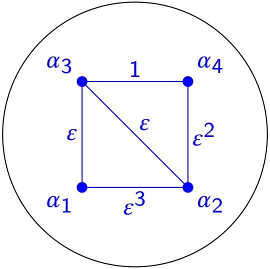
We use weighted graphs, i.e. graphs with a weight in $[0,1]$ on each edge (weight $0 \equiv$ no edge).

Definition (F., '18)

Fix $\boldsymbol{C} = (C_r)_{r \geq 1}$. A weighted graph $\widetilde{L}$ with vertex set A is a $\boldsymbol{C}$-weighted dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if, for any $\alpha_1, \dots, \alpha_r$ in A,

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

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



Weighted dependency graphs

We use weighted graphs, i.e. graphs with a weight in $[0,1]$ on each edge (weight $0 \equiv$ no edge).

Definition (F., '18)

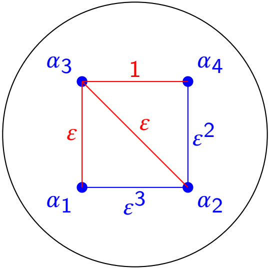
Fix $\boldsymbol{C} = (C_r)_{r \geq 1}$. A weighted graph $\widetilde{L}$ with vertex set A is a $\boldsymbol{C}$-weighted dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if, for any $\alpha_1, \dots, \alpha_r$ in A,

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

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

$\mathscr{M}(K)$: Maximum weight of a spanning tree of K (= product of the edge weights).

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



Weighted dependency graphs

We use weighted graphs, i.e. graphs with a weight in $[0,1]$ on each edge (weight $0 \equiv$ no edge).

Definition (F., '18)

Fix $\boldsymbol{C} = (C_r)_{r \geq 1}$. A weighted graph $\widetilde{L}$ with vertex set A is a $\boldsymbol{C}$-weighted dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if, for any $\alpha_1, \dots, \alpha_r$ in A,

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

Intuition: the smaller the edge weights are, the smaller the cumulant should be. The edge weights quantify the dependencies between variables.

Weighted dependency graphs

We use weighted graphs, i.e. graphs with a weight in $[0,1]$ on each edge (weight $0 \equiv$ no edge).

Definition (F., '18)

Fix $\boldsymbol{C} = (C_r)_{r \geq 1}$. A weighted graph $\widetilde{L}$ with vertex set A is a $\boldsymbol{C}$-weighted dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if, for any $\alpha_1, \dots, \alpha_r$ in A,

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

Intuition: the smaller the edge weights are, the smaller the cumulant should be. The edge weights quantify the dependencies between variables.

⚠ Unlike for usual dependency graphs, proving that something is a weighted dependency graph needs work!

Weighted dependency graphs

We use weighted graphs, i.e. graphs with a weight in $[0,1]$ on each edge (weight $0 \equiv$ no edge).

Definition (F., '18)

Fix $\boldsymbol{C} = (C_r)_{r\geq 1}$. A weighted graph $\widetilde{L}$ with vertex set A is a $\boldsymbol{C}$-weighted dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if, for any $\alpha_1, \dots, \alpha_r$ in A,

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

Intuition: the smaller the edge weights are, the smaller the cumulant should be. The edge weights quantify the dependencies between variables.

⚠ Unlike for usual dependency graphs, proving that something is a weighted dependency graph needs work!

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

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 $\boldsymbol{C}$-weighted dependency graph $\widetilde{L}_n$ with weighted maximal degree $D_n - 1$ (with a sequence $\boldsymbol{C} = (C_r)_{r \geq 1}$ independent of n).
- 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 $\boldsymbol{C}$-weighted dependency graph $\widetilde{L}_n$ with weighted maximal degree $D_n - 1$ (with a sequence $\boldsymbol{C} = (C_r)_{r \ge 1}$ independent of n).
- we set $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.

Theorem (F., '18)

Assume that $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} \to 0$ for some integer s. Then X_n is asymptotically normal.

Sketch of proof of the normality criterion

$$\left|\kappa_r(X_n)\right| \le \sum_{i_1,\dots,i_r} \left|\kappa\left(Y_{n,i_1},\cdots,Y_{n,i_r}\right)\right| \le C_r \sum_{i_1,\dots,i_r} \mathscr{M}\left(\widetilde{L}[i_1,\cdots,i_r]\right)$$

$$\le C_r \sum_{i_1,\dots,i_r} \left[\sum_{\substack{T \text{ spanning tree} \\ \text{of } \widetilde{L}[i_1,\cdots,i_r]}} \prod_{(j,k)\in E_T} w_{i_j,i_k} \right]$$

Sketch of proof of the normality criterion

$$\left|\kappa_r(X_n)\right| \le \sum_{i_1,\dots,i_r} \left|\kappa\left(Y_{n,i_1},\cdots,Y_{n,i_r}\right)\right| \le C_r \sum_{i_1,\dots,i_r} \mathscr{M}\left(\widetilde{L}[i_1,\cdots,i_r]\right)$$

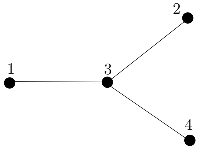
$$\le C_r \sum_{i_1,\dots,i_r} \left[\sum_{\substack{T \text{ spanning tree} \\ \text{of } \widetilde{L}[i_1,\cdots,i_r]}} \prod_{(j,k)\in E_T} w_{i_j,i_k} \right] \le C_r \sum_{\substack{T \text{ spanning} \\ \text{tree of } K_r}} \left[\sum_{i_1,\dots,i_r} \prod_{(j,k)\in E_T} w_{i_j,i_k} \right].$$

Sketch of proof of the normality criterion

$$|\kappa_r(X_n)| \leq \sum_{i_1,\dots,i_r} |\kappa(Y_{n,i_1},\cdots,Y_{n,i_r})| \leq C_r \sum_{i_1,\dots,i_r} \mathscr{M}\left(\widetilde{L}[i_1,\cdots,i_r]\right)$$

$$\leq C_r \sum_{i_1,\dots,i_r} \left[\sum_{\substack{T \text{ spanning tree} \\ \text{of } \widetilde{L}[i_1,\cdots,i_r]}} \prod_{(j,k)\in E_T} w_{i_j,i_k} \right] \leq C_r \sum_{\substack{T \text{ spanning} \\ \text{tree of } K_r}} \left[\sum_{i_1,\dots,i_r} \prod_{(j,k)\in E_T} w_{i_j,i_k} \right].$$

Fix a spanning tree T of K_r (=Cayley tree). We want to bound $\Sigma_T := \sum_{i_1,\dots,i_r} \prod_{(j,k)\in E_T} w_{i_j,i_k}$. On an example:

$T =$ 

;

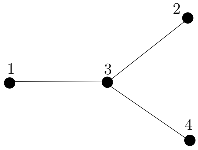
$$\Sigma_T = \underbrace{\sum_{i_1}}\left(\underbrace{\sum_{i_3} w_{i_1,i_3}}\left(\underbrace{\sum_{i_2} w_{i_2,i_3}}\left(\underbrace{\sum_{i_4} w_{i_3,i_4}}\right)\right)\right).$$

Sketch of proof of the normality criterion

$$\left|\kappa_r(X_n)\right| \le \sum_{i_1,\dots,i_r} \left|\kappa(Y_{n,i_1},\cdots,Y_{n,i_r})\right| \le C_r \sum_{i_1,\dots,i_r} \mathscr{M}\left(\widetilde{L}[i_1,\cdots,i_r]\right)$$

$$\le C_r \sum_{i_1,\dots,i_r} \left[\sum_{\substack{T \text{ spanning tree} \\ \text{of } \widetilde{L}[i_1,\cdots,i_r]}} \prod_{(j,k)\in E_T} w_{i_j,i_k} \right] \le C_r \sum_{\substack{T \text{ spanning} \\ \text{tree of } K_r}} \left[\sum_{i_1,\dots,i_r} \prod_{(j,k)\in E_T} w_{i_j,i_k} \right].$$

Fix a spanning tree T of K_r (=Cayley tree). We want to bound $\Sigma_T := \sum_{i_1,\dots,i_r} \prod_{(j,k)\in E_T} w_{i_j,i_k}$. On an example:

$T =$ 

;

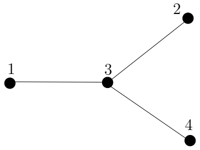
$$\Sigma_T = \underbrace{\sum_{i_1}}\left(\underbrace{\sum_{i_3} w_{i_1,i_3}}\left(\underbrace{\sum_{i_2} w_{i_2,i_3}}\left(\underbrace{\sum_{i_4} w_{i_3,i_4}}_{\le D_n}\right)\right)\right).$$

Sketch of proof of the normality criterion

$$\left|\kappa_r(X_n)\right| \leq \sum_{i_1,\dots,i_r} \left|\kappa(Y_{n,i_1},\cdots,Y_{n,i_r})\right| \leq C_r \sum_{i_1,\dots,i_r} \mathscr{M}\left(\widetilde{L}[i_1,\cdots,i_r]\right)$$

$$\leq C_r \sum_{i_1,\dots,i_r} \left[\sum_{\substack{T \text{ spanning tree} \\ \text{of } \widetilde{L}[i_1,\cdots,i_r]}} \prod_{(j,k)\in E_T} w_{i_j,i_k} \right] \leq C_r \sum_{\substack{T \text{ spanning} \\ \text{tree of } K_r}} \left[\sum_{i_1,\dots,i_r} \prod_{(j,k)\in E_T} w_{i_j,i_k} \right].$$

Fix a spanning tree T of K_r (=Cayley tree). We want to bound $\Sigma_T := \sum_{i_1,\dots,i_r} \prod_{(j,k)\in E_T} w_{i_j,i_k}$. On an example:

$T =$ 

;

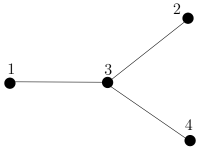
$$\Sigma_T = \underbrace{\sum_{i_1}}\left(\underbrace{\sum_{i_3} w_{i_1,i_3}} \left(\underbrace{\sum_{i_2} w_{i_2,i_3}}_{\leq D_n} \left(\underbrace{\sum_{i_4} w_{i_3,i_4}}_{\leq D_n} \right)\right)\right).$$

Sketch of proof of the normality criterion

$$\left|\kappa_r(X_n)\right| \le \sum_{i_1,\dots,i_r} \left|\kappa(Y_{n,i_1},\cdots,Y_{n,i_r})\right| \le C_r \sum_{i_1,\dots,i_r} \mathscr{M}\left(\widetilde{L}[i_1,\cdots,i_r]\right)$$

$$\le C_r \sum_{i_1,\dots,i_r} \left[\sum_{\substack{T \text{ spanning tree} \\ \text{of } \widetilde{L}[i_1,\cdots,i_r]}} \prod_{(j,k)\in E_T} w_{i_j,i_k} \right] \le C_r \sum_{\substack{T \text{ spanning} \\ \text{tree of } K_r}} \left[\sum_{i_1,\dots,i_r} \prod_{(j,k)\in E_T} w_{i_j,i_k} \right].$$

Fix a spanning tree T of K_r (=Cayley tree). We want to bound $\Sigma_T := \sum_{i_1,\dots,i_r} \prod_{(j,k)\in E_T} w_{i_j,i_k}$. On an example:

$T =$ 

;

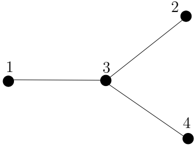
$$\Sigma_T = \underbrace{\sum_{i_1}} \left(\underbrace{\sum_{i_3} w_{i_1,i_3}}_{\le D_n} \left(\underbrace{\sum_{i_2} w_{i_2,i_3}}_{\le D_n} \left(\underbrace{\sum_{i_4} w_{i_3,i_4}}_{\le D_n} \right)\right)\right).$$

Sketch of proof of the normality criterion

$$\left|\kappa_r(X_n)\right| \leq \sum_{i_1,\dots,i_r} \left|\kappa(Y_{n,i_1},\cdots,Y_{n,i_r})\right| \leq C_r \sum_{i_1,\dots,i_r} \mathscr{M}\left(\widetilde{L}[i_1,\cdots,i_r]\right)$$

$$\leq C_r \sum_{i_1,\dots,i_r} \left[\sum_{\substack{T \text{ spanning tree} \\ \text{of } \widetilde{L}[i_1,\cdots,i_r]}} \prod_{(j,k)\in E_T} w_{i_j,i_k} \right] \leq C_r \sum_{\substack{T \text{ spanning} \\ \text{tree of } K_r}} \left[\sum_{i_1,\dots,i_r} \prod_{(j,k)\in E_T} w_{i_j,i_k} \right].$$

Fix a spanning tree T of K_r (=Cayley tree). We want to bound $\Sigma_T := \sum_{i_1,\dots,i_r} \prod_{(j,k)\in E_T} w_{i_j,i_k}$. On an example:

$T =$ 

;

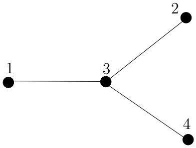
$$\Sigma_T = \underbrace{\sum_{i_1}}_{\leq N_n} \left(\underbrace{\sum_{i_3} w_{i_1,i_3}}_{\leq D_n} \left(\underbrace{\sum_{i_2} w_{i_2,i_3}}_{\leq D_n} \left(\underbrace{\sum_{i_4} w_{i_3,i_4}}_{\leq D_n} \right)\right)\right).$$

Sketch of proof of the normality criterion

$$\left|\kappa_r(X_n)\right| \leq \sum_{i_1,\dots,i_r} \left|\kappa(Y_{n,i_1},\cdots,Y_{n,i_r})\right| \leq C_r \sum_{i_1,\dots,i_r} \mathscr{M}\left(\widetilde{L}[i_1,\cdots,i_r]\right)$$

$$\leq C_r \sum_{i_1,\dots,i_r} \left[\sum_{\substack{T \text{ spanning tree} \\ \text{of } \widetilde{L}[i_1,\cdots,i_r]}} \prod_{(j,k)\in E_T} w_{i_j,i_k} \right] \leq C_r \sum_{\substack{T \text{ spanning} \\ \text{tree of } K_r}} \left[\sum_{i_1,\dots,i_r} \prod_{(j,k)\in E_T} w_{i_j,i_k} \right].$$

Fix a spanning tree T of K_r (=Cayley tree). We want to bound $\Sigma_T := \sum_{i_1,\dots,i_r} \prod_{(j,k)\in E_T} w_{i_j,i_k}$. On an example:

$T =$ 

;

$$\Sigma_T = \underbrace{\sum_{i_1}}_{\leq N_n} \left(\underbrace{\sum_{i_3} w_{i_1,i_3}}_{\leq D_n} \left(\underbrace{\sum_{i_2} w_{i_2,i_3}}_{\leq D_n} \left(\underbrace{\sum_{i_4} w_{i_3,i_4}}_{\leq D_n} \right)\right)\right).$$

In general, $\Sigma_T \leq N_n D_n^{r-1}$ and

$$\left|\kappa_r(X_n)\right| \leq C_r\, r^{r-2}\, N_n\, D_n^{r-1}. \quad \square$$

Stability by powers

Setting:

- Let $\{Y_\alpha, \alpha \in A\}$ be r.v. with $\boldsymbol{C}$-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 $\boldsymbol{C}$-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

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

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

where $\boldsymbol{C}^{(m)}$ depends only on $\boldsymbol{C}$ and m.

Convention: $\mathrm{wt}_{\widetilde{L}}(Y_\alpha, Y_\alpha) = 1$.

Stability by powers

Setting:

- Let $\{Y_\alpha, \alpha \in A\}$ be r.v. with $\boldsymbol{C}$-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

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

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

where $\boldsymbol{C}^{(m)}$ depends only on $\boldsymbol{C}$ and m.

In short: if we have a wieght dependency graph for $\{Y_\alpha\}$, we have also one for monomials in the Y_α.
(And potentially asymptotic normality of polynomials in the Y_α).

Transition

A weighted dependency graph for Markov chain

Setting:

- Let $(w_i)_{i \geq 1}$ be an irreducible aperiodic Markov chain on a finite space state $\mathscr{A}$;
- Assume w_1 is distributed with the stationary distribution π;
- Set $Z_{i,s} = \mathbf{1}_{w_i = s}$.

A weighted dependency graph for Markov chain

Setting:

- Let $(w_i)_{i\geq 1}$ be an irreducible aperiodic Markov chain on a finite space state $\mathscr{A}$;
- Assume w_1 is distributed with the stationary distribution π;
- Set $Z_{i,s} = \mathbf{1}_{w_i=s}$.

Proposition

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

Concretely, this means that, for $i_1 < \cdots < i_r$,

$$\left|\kappa(Z_{i_1,s_1}, \dots, Z_{i_r,s_r})\right| \leq C_r \, |\lambda_2|^{i_r - i_1}.$$

This was proved by Saulis and Statulevičius ('90).

A weighted dependency graph for Markov chain

Setting:

- Let $(w_i)_{i \geq 1}$ be an irreducible aperiodic Markov chain on a finite space state $\mathscr{A}$;
- Assume w_1 is distributed with the stationary distribution π;
- Set $Z_{i,s} = \mathbf{1}_{w_i = s}$.

Proposition

We have a weighted dependency graph $\widetilde{L}$ with $\mathrm{wt}_{\widetilde{L}}\big(\{Z_{i,s}, Z_{j,t}\}\big) = |\lambda_2|^{j-i}$ (for $i < j$), 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 $Z_{I;S} := Z_{i_1,s_1} \cdots Z_{i_m,s_m}$, with $\mathrm{wt}_{\widetilde{L}^m}(Z_{I;S}, Z_{I,T}) = |\lambda_2|^{\mathrm{md}(I,J)}$, where $\mathrm{md}(I,J)$ is the minimal distance between I and J.

Subword occurrences in Markovian text (1/2)

Let $(w_i)_{i\geq 1}$ be a Markov chain as before and fix a pattern (= a word) u of length ℓ on $\mathscr{A}$.

For $I = \{i_1, \cdots, i_\ell\} \subset \mathbb{N}$ $(i_1 < \cdots < i_\ell)$, we set

$$\begin{aligned} Y_I &= \mathbf{1}\big[u \text{ occurs at position } I \text{ in } \boldsymbol{w}\big]; \\ &= Z_{i_1,u_1} \cdots Z_{i_s,u_s}. \end{aligned}$$

Subword occurrences in Markovian text (1/2)

Let $(w_i)_{i\geq 1}$ be a Markov chain as before and fix a pattern (= a word) u of length ℓ on $\mathscr{A}$.

For $I = \{i_1, \cdots, i_\ell\} \subset \mathbb{N}$ $(i_1 < \cdots < i_\ell)$, we set

$$\begin{aligned} Y_I &= \mathbf{1}\big[u \text{ occurs at position } I \text{ in } \boldsymbol{w}\big]; \\ &= Z_{i_1,u_1} \cdots Z_{i_s,u_s}. \end{aligned}$$

We have a weighted dependency graph for $\left(Y_I, I \in \binom{[n]}{\ell}\right)$, which is a restriction of the one for the $Z_{I,S}$.

Subword occurrences in Markovian text (1/2)

Let $(w_i)_{i\geq 1}$ be a Markov chain as before and fix a pattern (= a word) u of length ℓ on $\mathscr{A}$.

For $I = \{i_1, \cdots, i_\ell\} \subset \mathbb{N}$ $(i_1 < \cdots < i_\ell)$, we set

$$Y_I = \mathbf{1}\left[u \text{ occurs at position } I \text{ in } \boldsymbol{w}\right]; \\ = Z_{i_1,u_1} \cdots Z_{i_s,u_s}.$$

We have a weighted dependency graph for $\left(Y_I, I \in \binom{[n]}{\ell}\right)$, which is a restriction of the one for the $Z_{I,S}$.

What is its maximal weighted degree D_n? Fix $I = \{i_1, \cdots, i_\ell\}$, we have

$$\sum_J |\lambda_2|^{\mathrm{md}(I,J)} \leq \sum_J \sum_{s,t\leq\ell} |\lambda_2|^{|i_s - j_t|} \leq \ell^2 \sum_J |\lambda_2|^{|i_1 - j_1|} \\ \leq \ell^2 \binom{n-1}{\ell-1} \sum_{j_1} |\lambda_2|^{|i_1 - j_1|} = \mathscr{O}(n^{\ell-1}).$$

Subword occurrences in Markovian text (2/2)

Let $X_n = \sum_I Y_I$ be the number of occurrences of u in a Markovian text $\boldsymbol{w}$.
Recall that $\left(Y_I, I \in \binom{[n]}{\ell}\right)$ admits a weighted dependency graph.

Can we apply the normality criterion?

Subword occurrences in Markovian text (2/2)

Let $X_n = \sum_I Y_I$ be the number of occurrences of u in a Markovian text $\boldsymbol{w}$. Recall that $\left(Y_I, I \in \binom{[n]}{\ell}\right)$ admits a weighted dependency graph.

Can we apply the normality criterion? $M = 1$, $N_n = \binom{n}{\ell}$, $D_n = \mathcal{O}(n^{\ell-1})$ and $\sigma_n = \sqrt{\mathrm{Var}(X_n)} = (C + o(1))n^{\ell-1/2}$, for a computable constant $C \geq 0$ (Bourdon, Vallée, '01).

Subword occurrences in Markovian text (2/2)

Let $X_n = \sum_I Y_I$ be the number of occurrences of u in a Markovian text $\boldsymbol{w}$. Recall that $\left(Y_I, I \in \binom{[n]}{\ell}\right)$ admits a weighted dependency graph.

Can we apply the normality criterion? $M = 1$, $N_n = \binom{n}{\ell}$, $D_n = \mathcal{O}(n^{\ell-1})$ and $\sigma_n = \sqrt{\mathrm{Var}(X_n)} = (C + o(1))n^{\ell-1/2}$, for a computable constant $C \geq 0$ (Bourdon, Vallée, '01).

$\rightarrow$ when $C > 0$, the normality criterion satisfied for $s = 3$.

Conclusion: when $C > 0$, the number X_n of occurrences of u in a Markovian text $\boldsymbol{w}$ is asymptotically normal.

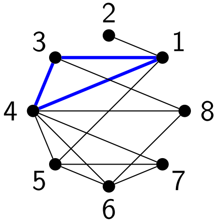
(Answers partially a question of Bourdon–Vallée, '01).

Erdős-Rényi graph model $G(n, M)$

Subgraph count $G(n, M)$

- 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!

Weighted dependency graphs for $G(n, M)$

Consider $\boldsymbol{G} \sim G(n, M)$ with $M = p\binom{n}{2}$, p fixed in $(0,1)$.
Let $A_2 = \binom{[n]}{2}$ and for $e \in A_2$, let $\mathbf{1}[e]$ be the edge indicator variable.

Proposition

The complete graph on A_2 with weights $1/n^2$ is a $\boldsymbol{C}$-weighted dependency graph for $\{\mathbf{1}[e], e \in A_2\}$, for some fixed sequence $\boldsymbol{C} = (C_r)_{r \geq 1}$.

Weighted dependency graphs for $G(n,M)$

Consider $\boldsymbol{G} \sim G(n,M)$ with $M = p\binom{n}{2}$, p fixed in $(0,1)$.
Let $A_2 = \binom{[n]}{2}$ and for $e \in A_2$, let $\mathbf{1}[e]$ be the edge indicator variable.

Proposition

The complete graph on A_2 with weights $1/n^2$ is a $\boldsymbol{C}$-weighted dependency graph for $\{\mathbf{1}[e], e \in A_2\}$, for some fixed sequence $\boldsymbol{C} = (C_r)_{r \geq 1}$.

Concretely, this means

$$\left|\kappa(\mathbf{1}[e_1], \dots, \mathbf{1}[e_r])\right| \leq C_r n^{-2d+2},$$

where d is the number of distinct edges in $\{e_1, \dots, e_r\}$.

Weighted dependency graphs for $G(n,M)$

Consider $\boldsymbol{G} \sim G(n,M)$ with $M = p\binom{n}{2}$, p fixed in $(0,1)$.
Let $A_2 = \binom{[n]}{2}$ and for $e \in A_2$, let $\mathbf{1}[e]$ be the edge indicator variable.

Proposition

The complete graph on A_2 with weights $1/n^2$ is a $\boldsymbol{C}$-weighted dependency graph for $\{\mathbf{1}[e], e \in A_2\}$, for some fixed sequence $\boldsymbol{C} = (C_r)_{r\geq 1}$.

Concretely, this means

$$\left|\kappa(\mathbf{1}[e_1],\dots,\mathbf{1}[e_r])\right| \leq C_r n^{-2d+2},$$

where d is the number of distinct edges in $\{e_1,\dots,e_r\}$.

General fact: for Bernoulli variables, it is enough to establish the bounds on cumulants of distinct variables.

Weighted dependency graphs for $G(n,M)$

Consider $\boldsymbol{G} \sim G(n,M)$ with $M = p\binom{n}{2}$, p fixed in $(0,1)$.
Let $A_2 = \binom{[n]}{2}$ and for $e \in A_2$, let $\mathbf{1}[e]$ be the edge indicator variable.

Proposition

The complete graph on A_2 with weights $1/n^2$ is a $\boldsymbol{C}$-weighted dependency graph for $\{\mathbf{1}[e], e \in A_2\}$, for some fixed sequence $\boldsymbol{C} = (C_r)_{r \geq 1}$.

What needs to be proved (set $N = \binom{n}{2}$):

$$\sum_{\pi \text{ set-partition of } [r]} (-1)^{|\pi|-1}(|\pi|-1)! \left(\prod_{B \in \pi} \binom{N-|B|}{M-|B|} \Big/ \binom{N}{M} \right) = \mathcal{O}(n^{-2r+2}).$$

(all terms on the LHS have degree 0 in N and M; showing that the sum has degree at most -1 is easy, that it has degree $-r+1$ not so much.)

Weighted dependency graphs for $G(n,M)$

Consider $\boldsymbol{G} \sim G(n,M)$ with $M = p\binom{n}{2}$, p fixed in $(0,1)$.
Let $A_2 = \binom{[n]}{2}$ and for $e \in A_2$, let $\mathbf{1}[e]$ be the edge indicator variable.

Proposition

The complete graph on A_2 with weights $1/n^2$ is a $\boldsymbol{C}$-weighted dependency graph for $\{\mathbf{1}[e], e \in A_2\}$, for some fixed sequence $\boldsymbol{C} = (C_r)_{r \geq 1}$.

Corollary

The complete graph on $A_3 = \{\Delta \in \binom{[n]}{3}\}$ with weights

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

is a weighted dependency graph for the triangle indicator variables.

Asymptotic normality of the number of triangles in $G(n,M)$

Corollary (copied from previous slide)

The complete graph on $A_3 = \{\Delta \in \binom{[n]}{3}\}$ *with weights*

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

is a weighted dependency graph for the triangle indicator variables.

Can we apply the normality criterion?

Asymptotic normality of the number of triangles in $G(n,M)$

Corollary (copied from previous slide)

The complete graph on $A_3 = \{\Delta \in \binom{[n]}{3}\}$ with weights

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

is a weighted dependency graph for the triangle indicator variables.

Can we apply the normality criterion? $N_n = \binom{n}{3}$, $D_n = n$.

One can estimate the variance as $\Theta(n^3)$ (smaller than for $G(n,p)$!).

The criterion is fulfilled for $s = \mathbf{5}$, thus T_n is asymptotically normal.

Asymptotic normality of the number of triangles in $G(n,M)$

Corollary (copied from previous slide)

The complete graph on $A_3 = \{\Delta \in \binom{[n]}{3}\}$ *with weights*

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

is a weighted dependency graph for the triangle indicator variables.

Can we apply the normality criterion? $N_n = \binom{n}{3}$, $D_n = n$.

One can estimate the variance as $\Theta(n^3)$ (smaller than for $G(n,p)$!).

The criterion is fulfilled for $s = \mathbf{5}$, thus T_n is asymptotically normal.

- This can be generalized to $p = p_n \gg 1/n$ and to other subgraph counts (recovers a result of Janson, '94).
- similar bounds on cumulants can be found in $G(n,d)$ (random regular graph), see Janson '20.

Transition

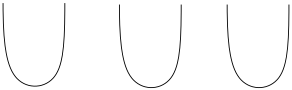
Stam's algortihm for generating set-partitions

How to generate a uniform random a set-partition of $[n]$?

Stam's algortihm for generating set-partitions

How to generate a uniform random a set-partition of $[n]$?

- Take M at random with distribution: $\mathbb{P}(M=m) = \frac{1}{eB_n} \frac{m^n}{m!}$ (B_n: Bell number) and consider M urns.
  Note: M concentrates around $n/\log n$.

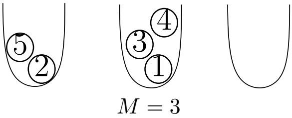


$M = 3$

Stam's algortihm for generating set-partitions

How to generate a uniform random a set-partition of $[n]$?

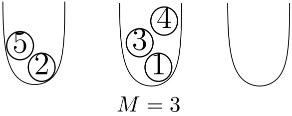
- Take M at random with distribution: $\mathbb{P}(M = m) = \frac{1}{eB_n} \frac{m^n}{m!}$ (B_n: Bell number) and consider M urns.
  Note: M concentrates around $n/\log n$.
- Drop numbers from 1 to n independently uniformly in the urns.



Stam's algortihm for generating set-partitions

How to generate a uniform random a set-partition of $[n]$?

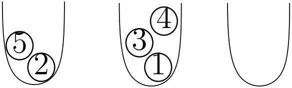
- Take M at random with distribution: $\mathbb{P}(M=m) = \frac{1}{eB_n} \frac{m^n}{m!}$ (B_n: Bell number) and consider M urns.
  Note: M concentrates around $n/\log n$.
- Drop numbers from 1 to n independently uniformly in the urns.
- Forget empty urns and the order on the urns, you get a set-partition: in the example, $\{1,3,4\},\{2,5\}$.



Stam's algortihm for generating set-partitions

How to generate a uniform random a set-partition of $[n]$?

- Take M at random with distribution: $\mathbb{P}(M=m) = \frac{1}{eB_n}\frac{m^n}{m!}$ (B_n: Bell number) and consider M urns.
  Note: M concentrates around $n/\log n$.
- Drop numbers from 1 to n independently uniformly in the urns.
- Forget empty urns and the order on the urns, you get a set-partition: in the example, $\{1,3,4\},\{2,5\}$.



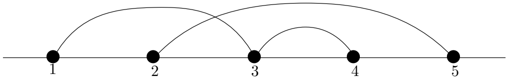
$M = 3$

Proposition (Stam, '83)

The resulting set partition π of $[n]$ is uniformly distributed. Moreover, the number of empty urns is Poisson(1)-distributed and independent from π.

Patterns in set-partitions

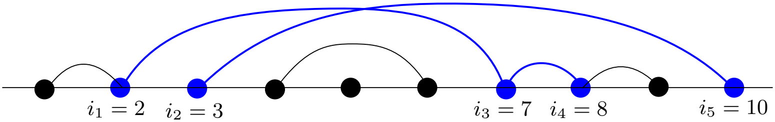
We think at partitions as arch systems, e.g. $\{1,3,4\},\{2,5\}$ is



Definition

An occurrence of a set-partition $\mathscr{A}$ of size ℓ in another set-partition π is a list $(i_1,\dots,i_\ell)$ s.t. (i_j,i_k) is an arch of π whenever (j,k) is an arch of $\mathscr{A}$.

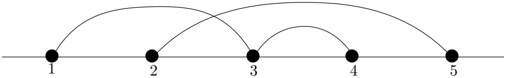
Example: an occurrence of $\{1,3,4\},\{2,5\}$



⚠ encapsulates contraints on the i_j's, but also on intermediate points (in the example, i_1 and i_3 should be in the same part, but none of the points inbetween).

Patterns in set-partitions

We think at partitions as arch systems, e.g. $\{1,3,4\},\{2,5\}$ is



Definition

An occurrence of a set-partition $\mathscr{A}$ of size ℓ in another set-partition π is a list $(i_1,\dots,i_\ell)$ s.t. (i_j,i_k) is an arch of π whenever (j,k) is an arch of $\mathscr{A}$.

Background:

- standard well-studied examples: crossings, nestings, k-crossings, k-nestings;
- the general notion was defined (in even more generality) by Chern, Diaconis, Kane, Rhodes, '14;
- the same authors proved the asymptotic normality of the number of crossings ('15).

A weighted dependency graph for set-partitions

Let $\boldsymbol{\pi}$ be a uniform random set-partition of size n and $\mathbf{1}[\widehat{ij}]$ be the indicator variable of the arc $\{i,j\}$ $(1 \le i < j \le n)$.

Proposition

The complete graph with weights

$$w\left(\mathbf{1}[\widehat{ij}], \mathbf{1}[\widehat{i'j'}]\right) = \begin{cases} 1 & \text{if } i = i' \text{ or } j = j'; \\ 1/n & \text{otherwise.} \end{cases}$$

is a $(\boldsymbol{C}, \boldsymbol{\Psi})$-weighted dependency graph for the family $\{\mathbf{1}[\widehat{ij}], i < j\}$, for some $\boldsymbol{C} = (C_r)_{r \ge 1}$ **depending on** *n with $C_r = \widetilde{\mathcal{O}}(1)$ and some $\boldsymbol{\Psi}$.*

Here, we need the general definition of weighted dependency graph, which involves some function $\boldsymbol{\Psi}$ as parameter.

A weighted dependency graph for set-partitions

Let π be a uniform random set-partition of size n and $\mathbf{1}[\widehat{ij}]$ be the indicator variable of the arc $\{i,j\}$ $(1 \le i < j \le n)$.

Proposition

The complete graph with weights

$$w(\mathbf{1}[\widehat{ij}], \mathbf{1}[\widehat{i'j'}]) = \begin{cases} 1 & \text{if } i = i' \text{ or } j = j'; \\ 1/n & \text{otherwise.} \end{cases}$$

is a $(\boldsymbol{C}, \boldsymbol{\Psi})$-weighted dependency graph for the family $\{\mathbf{1}[\widehat{ij}], i < j\}$, for some $\boldsymbol{C} = (C_r)_{r \ge 1}$ **depending on** *n with $C_r = \widetilde{\mathcal{O}}(1)$ and some $\boldsymbol{\Psi}$.*

It is enough to prove that for distinct $i_1, \dots, i_r$ and distinct $j_1, \dots, j_r$

$$\kappa(\mathbf{1}[\widehat{i_1 j_1}], \dots, \mathbf{1}[\widehat{i_r j_r}]) = \widetilde{\mathcal{O}}(n^{-2r+1})$$

Elements of proof: use Stam's urn model, first control cumulants conditionally on M, and then use the law of total cumulance.

Asymptotic normality of patterns in set partition

Using the stability by product of weighted-dependency graphs, we get:

Proposition (F., '19)

Fix a pattern $\mathscr{A}$. Let $\mathbf{1}[\pi_I = \mathscr{A}]$ be the indicator of having the pattern $\mathscr{A}$ at position I. This family has a $(\boldsymbol{C}, \boldsymbol{\Psi})$-weighted dependency graph with weights $w(\mathbf{1}[\pi_I = \mathscr{A}], \mathbf{1}[\pi_{I'} = \mathscr{A}]) = \begin{cases} 1 & \text{if } I \cap I' \neq \emptyset; \\ 1/n & \text{otherwise.} \end{cases}$

Asymptotic normality of patterns in set partition

Using the stability by product of weighted-dependency graphs, we get:

Proposition (F., '19)

Fix a pattern $\mathscr{A}$. Let $\mathbf{1}[\pi_I = \mathscr{A}]$ be the indicator of having the pattern $\mathscr{A}$ at position I. This family has a $(\boldsymbol{C}, \boldsymbol{\Psi})$-weighted dependency graph with weights

$$w(\mathbf{1}[\pi_I = \mathscr{A}], \mathbf{1}[\pi_{I'} = \mathscr{A}]) = \begin{cases} 1 & \text{if } I \cap I' \neq \emptyset; \\ 1/n & \text{otherwise.} \end{cases}$$

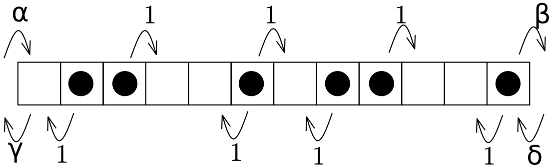
Using a generalization of the above normality criterion, we get

Corollary (F., '19)

For any pattern $\mathscr{A}$, the number $X_n^{\mathscr{A}}$ of occurrences of $\mathscr{A}$ in a uniform random set-partition $\boldsymbol{\pi}$ of $[n]$ is asymptotically normal.

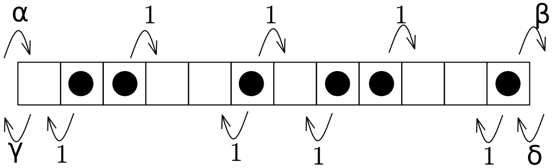
Transition

Symmetric simple exclusion process (SSEP)



$\tau = (\tau_1, \cdots, \tau_N)$ particle configuration with stationary distribution.

Symmetric simple exclusion process (SSEP)



$\tau = (\tau_1, \cdots, \tau_N)$ particle configuration with stationary distribution.

Theorem

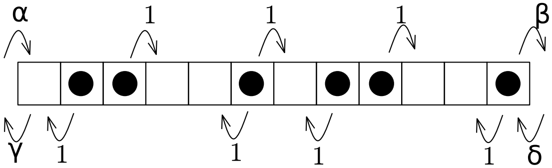
The complete graph on $[N]$ with weight $1/N$ on each edge is a weighted dependency graph for the family $\{\tau_i, 1 \le i \le N\}$.

Concretely, for $i_1, \cdots, i_r$,

$$\kappa(\tau_{i_1}, \ldots, \tau_{i_r}) = \mathcal{O}_r(N^{-d+1}),$$

where $d = |\{i_1, \ldots, i_r\}|$.

Symmetric simple exclusion process (SSEP)



$\tau = (\tau_1, \cdots, \tau_N)$ particle configuration with stationary distribution.

Theorem

The complete graph on $[N]$ with weight $1/N$ on each edge is a weighted dependency graph for the family $\{\tau_i, 1 \le i \le N\}$.

Ingredients of the proof

- enough to prove the bound for distinct $i_1, \dots, i_r$;
- joint moments of the τ_i given by matrix ansatz;
- this gives an induction formula for cumulants (Derrida, Lebowitz, Speer, 2006), from which we deduce easily the upper bound.

A functional central limit theorem

Set $X_N(t) = \sum_{i=1}^{Nt} \tau_i$ be the particle distribution function.

Theorem (F., '18)

There exists a continuous Gaussian process Z on $[0,1]$ with explicit covariance function such that, in the space $\mathscr{D}([0,1])$,

$$\widetilde{X_N}(t) := \frac{X_N(t) - \mathbb{E}X_N(t)}{\sqrt{N}} \xrightarrow{d} Z$$

Essentially similar to a result of Derrida–Enaud–Landim–Olla '05 on the fluctuations of the density of particles.

A functional central limit theorem

Set $X_N(t) = \sum_{i=1}^{Nt} \tau_i$ be the particle distribution function.

Theorem (F., '18)

There exists a continuous Gaussian process Z on $[0,1]$ with explicit covariance function such that, in the space $\mathscr{D}([0,1])$,

$$\widetilde{X_N}(t) := \frac{X_N(t) - \mathbb{E}X_N(t)}{\sqrt{N}} \xrightarrow{d} Z$$

Essentially similar to a result of Derrida–Enaud–Landim–Olla '05 on the fluctuations of the density of particles.

Any interest in asymptotic normality of higher order polynomials in the τ_i?

A functional central limit theorem

Set $X_N(t) = \sum_{i=1}^{Nt} \tau_i$ be the particle distribution function.

Theorem (F., '18)

There exists a continuous Gaussian process Z on $[0,1]$ with explicit covariance function such that, in the space $\mathscr{D}([0,1])$,

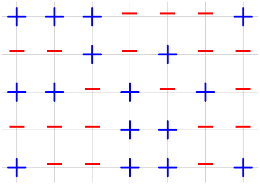
$$\widetilde{X_N}(t) := \frac{X_N(t) - \mathbb{E}X_N(t)}{\sqrt{N}} \xrightarrow{d} Z$$

Derrida et al.'s result holds more generally for ASEP (A=asymmetric, i.e. particles jump backwards at rate q instead of 1).

Question

Is the same weighted graph also a weighted dependency graphs for particles in ASEP? Or should we use weights $1/|i-j|$?

Ising model

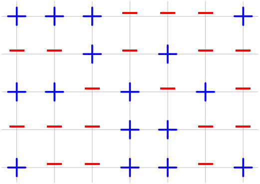


$$\mathbb{P}(\omega) \propto \exp\left[-H(\omega)\right];$$
$$H(\omega) = -\beta \sum_{x\sim y} \omega_x \omega_y - h \sum_x \omega_x.$$

Theorem

In presence of a magnetic field or at very low or very large temperature, there exists $\varepsilon = \varepsilon(d,h,\beta) > 0$ *such that the complete graph on* $\mathbb{Z}^d$ *with weight* $\varepsilon^{\|x-y\|_1}$ *on the edge* $\{x,y\}$ *is a weighted dependency graph for* $\{\sigma_x, x \in \mathbb{Z}^d\}$

Ising model



$$\mathbb{P}(\omega) \propto \exp\left[-H(\omega)\right];$$
$$H(\omega) = -\beta \sum_{x\sim y} \omega_x \omega_y - h \sum_x \omega_x.$$

Theorem

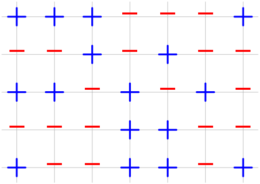
In presence of a magnetic field or at very low or very large temperature, there exists $\varepsilon = \varepsilon(d,h,\beta) > 0$ such that the complete graph on $\mathbb{Z}^d$ with weight $\varepsilon^{\|x-y\|_1}$ on the edge $\{x,y\}$ is a weighted dependency graph for $\{\sigma_x, x \in \mathbb{Z}^d\}$

Concretely, this means that

$$\kappa(\sigma_{x_1},\dots,\sigma_{x_r}) = \mathcal{O}_r(\varepsilon^{\ell_T(x_1,\dots,x_r)}),$$

where $\ell_T(x_1,\dots,x_r)$ is the smallest length of a tree connecting $x_1,\dots,x_r$.

Ising model



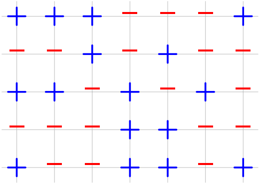
$$\mathbb{P}(\omega) \propto \exp\left[-H(\omega)\right];$$
$$H(\omega) = -\beta \sum_{x\sim y} \omega_x \omega_y - h \sum_x \omega_x.$$

Theorem

In presence of a magnetic field or at very low or very large temperature, there exists $\varepsilon = \varepsilon(d,h,\beta) > 0$ such that the complete graph on $\mathbb{Z}^d$ with weight $\varepsilon^{\|x-y\|_1}$ on the edge $\{x,y\}$ is a weighted dependency graph for $\{\sigma_x, x \in \mathbb{Z}^d\}$

This was proved by Duneau, Iagolnitzer and Souillard ('74) (with magnetic field or in very high temperature) and Malyshev and Minlos ('91) in very low temperature.
Proofs based on cluster expansion...

Ising model



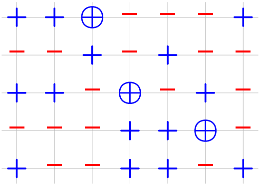
$$\mathbb{P}(\omega) \propto \exp\left[-H(\omega)\right];$$
$$H(\omega) = -\beta \sum_{x \sim y} \omega_x \omega_y - h \sum_x \omega_x.$$

Theorem

In presence of a magnetic field or at very low or very large temperature, there exists $\varepsilon = \varepsilon(d,h,\beta) > 0$ such that the complete graph on $\mathbb{Z}^d$ with weight $\varepsilon^{\|x-y\|_1}$ on the edge $\{x,y\}$ is a weighted dependency graph for $\{\sigma_x, x \in \mathbb{Z}^d\}$

Question: does it hold near the critical point?
(At the critical point, the answer is NO, since already covariances do not decay exponentially)

Ising model: asymptotic normality of global patterns



Circled spins:
South-East chain of +

$S_n :=$ number of south-East chains of $\oplus$ within $\Lambda_n = [-n, n]^2$.

Theorem (Dousse, F., '19)

S_n *is asymptotically normal.*

(generalizes to more "pattern" counts and any dimension.)

Conclusion

- Dependency graphs are a powerful simple tool to prove asymptotic normality, particularly for substructure counts in models exhibiting some independence;
- We proposed an extension to handle models without independence, but with weak dependencies.
- Plenty of applications (both for the initial framework and for the extended one)!

Conclusion

- Dependency graphs are a powerful simple tool to prove asymptotic normality, particularly for substructure counts in models exhibiting some independence;
- We proposed an extension to handle models without independence, but with weak dependencies.
- Plenty of applications (both for the initial framework and for the extended one)!

Thank you for your attention!