Graphes de dépendence (pondérés) et normalité asymptotique

Valentin Féray

CNRS, Institut Élie Cartan de Lorraine (IECL)

Séminaire *Probabilités et Statistiques* de l'IECL
Nancy, 12 novembre 2020

What is this talk about ?

Consider some sequence of r.v. X_n (e.g., number of substructures of a given type in some probabilistic model).

Goal: prove that some X_n satisfies is asymptotically normal, i.e.

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

What is this talk about ?

Consider some sequence of r.v. X_n (e.g., number of substructures of a given type in some probabilistic model).

Goal: prove that some X_n satisfies is asymptotically normal, i.e.

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

A powerful tool: analytic methods, in particular bivariate generating functions and Hwang's quasi-power theorem.

Problem: the bivariate generating function might be intractable.

What is this talk about ?

Consider some sequence of r.v. X_n (e.g., number of substructures of a given type in some probabilistic model).

Goal: prove that some X_n satisfies is asymptotically normal, i.e.

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

Other standard tool: moment (or cumulant) methods.

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

Outline of the talk

1. Dependency graphs
   - A motivating example: substrings in random words
   - An asymptotic normality criterion

2. Weighted dependency graphs
   - Definition and an extended normality criterion
   - Back to subwords: Markovian texts
   - Applications in statistical physics

Transition

1. Dependency graphs
   - A motivating example: substrings in random words
   - An asymptotic normality criterion

2. Weighted dependency graphs
   - Definition and an extended normality criterion
   - Back to subwords: Markovian texts
   - 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 $i_1 < \dots < i_\ell$ 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)

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

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

where C_1 and C_2 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

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

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

The proof of the 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

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

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

The proof of the 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 $Y_I = \mathbf{1}\big[u$ occurs at position I in $\boldsymbol{w}\big]$.
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 $A_1, A_2 \subset A$:

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

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

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 $A_1, A_2 \subset A$:

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

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 $A_1, A_2 \subset A$:

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

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

Example

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; $|Y_{n,i}| < M$ 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; $|Y_{n,i}| < M$ 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 $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} \to 0$ *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; $|Y_{n,i}| < M$ 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 $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} \to 0$ *for some integer* s.
Then X_n *is asymptotically normal.*

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

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

so that asymptotic normality follows (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; $|Y_{n,i}| < M$ 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 $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} \to 0$ *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

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

  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.; $|Y_{n,i}| < M$ 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 $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} \to 0$ 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.; $|Y_{n,i}| < M$ 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 $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} \to 0$ 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.; $|Y_{n,i}| < M$ 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 $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} \to 0$ 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.; $|Y_{n,i}| < M$ 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 $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} \to 0$ 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.; $|Y_{n,i}| < M$ 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 $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} \to 0$ 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 $C_r M^r$.

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.; $|Y_{n,i}| < M$ 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 $\left(\frac{N_n}{D_n}\right)^{1/s} \frac{D_n}{\sigma_n} \to 0$ for some $s \ge 3$.

Fix $r \ge 1$. Then

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

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

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

Applications of dependency graphs to asymptotic normality results

- mathematical modelization of cell populations (Petrovskaya, Leontovich, 82);
- subgraph counts in random graphs (Janson, Baldi, Rinott, Penrose, 88, 89, 95, 03);
- Geometric probability: length of k neighbour graphs (Avram, Bertsimas, Penrose, Yukich, Bárány, Vu, 93, 05 , 07);
- pattern occurrences in random permutations (Bóna, Janson, Hitchenko, Nakamura, Zeilberger, Hofer, 07, 09, 14, 18).

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;

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);

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;

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, ...;

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, ...;
- spins or patterns of spins in Ising model.

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, ...;
- spins or patterns of spins in 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,

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

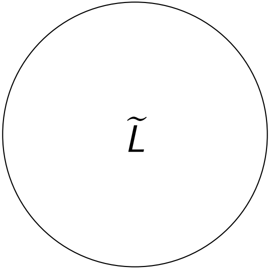
Weighted dependency graphs

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

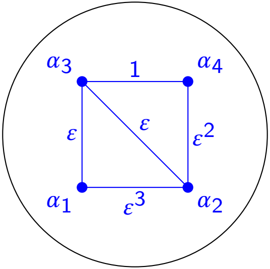
We use weighted graphs, i.e. graphs with a weight in $[0,1]$ on each edge (weight $0 \equiv$ no edge).

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,

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

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



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)

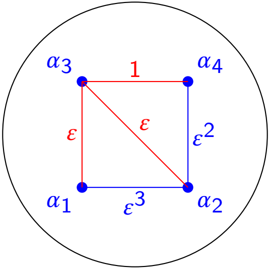
Fix $\boldsymbol{C} = (C_r)_{r \geq 1}$. A weighted graph $\widetilde{L}$ with vertex set A is a $\boldsymbol{C}$-weighted dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ if, for any $\alpha_1, \ldots, \alpha_r$ in A,

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

$\widetilde{L}[\alpha_1, \cdots, \alpha_r]$: 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,

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

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!

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!

⚠ 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,

- $\{Y_{n,i}, 1 \le i \le N_n\}$ 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 $\sigma_n^2 = \mathrm{Var}(X_n)$.

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.*

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.*

Note: if $s = 3$ and $C_r \le K^r (r!)^{\gamma}$, we also have bounds on the speed of convergence and deviation estimates.

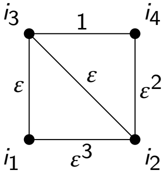
Sketch of proof of the normality criterion (1/2)

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

Sketch of proof of the normality criterion (1/2)

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

Prim's algorithm. We can construct the spanning tree T of $\widetilde{L}[i_1,\cdots,i_r]$ of maximal weight as follows:

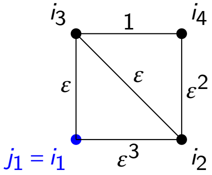


Sketch of proof of the normality criterion (1/2)

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

Prim's algorithm. We can construct the spanning tree T of $\widetilde{L}[i_1,\cdots,i_r]$ of maximal weight as follows:

- Start with any vertex j_1, e.g. $j_1 = i_1$;

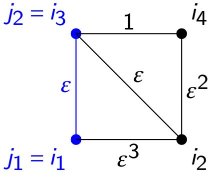


Sketch of proof of the normality criterion (1/2)

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

Prim's algorithm. We can construct the spanning tree T of $\widetilde{L}[i_1,\cdots,i_r]$ of maximal weight as follows:

- Start with any vertex j_1, e.g. $j_1 = i_1$;
- take j_2 which maximizes the weight of $\{j_1, j_2\}$ and add $\{j_1, j_2\}$ to T;

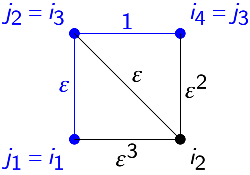


Sketch of proof of the normality criterion (1/2)

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

Prim's algorithm. We can construct the spanning tree T of $\widetilde{L}[i_1,\cdots,i_r]$ of maximal weight as follows:

- Start with any vertex j_1, e.g. $j_1 = i_1$;
- take j_2 which maximizes the weight of $\{j_1,j_2\}$ and add $\{j_1,j_2\}$ to T;
- take j_3 which maximizes either the weight of $\{j_1,j_3\}$ or $\{j_2,j_3\}$ and add the corresponding edge to T;

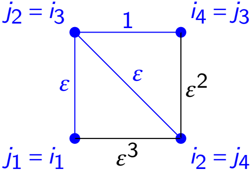


Sketch of proof of the normality criterion (1/2)

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

Prim's algorithm. We can construct the spanning tree T of $\widetilde{L}[i_1,\cdots,i_r]$ of maximal weight as follows:

- Start with any vertex j_1, e.g. $j_1 = i_1$;
- take j_2 which maximizes the weight of $\{j_1, j_2\}$ and add $\{j_1, j_2\}$ to T;
- take j_3 which maximizes either the weight of $\{j_1, j_3\}$ or $\{j_2, j_3\}$ and add the corresponding edge to T; and so on...

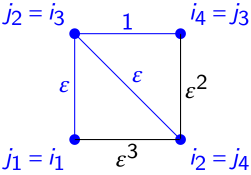


Sketch of proof of the normality criterion (1/2)

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

Prim's algorithm. We can construct the spanning tree T of $\widetilde{L}[i_1,\cdots,i_r]$ of maximal weight as follows:

- Start with any vertex j_1, e.g. $j_1 = i_1$;
- take j_2 which maximizes the weight of $\{j_1,j_2\}$ and add $\{j_1,j_2\}$ to T;
- take j_3 which maximizes either the weight of $\{j_1,j_3\}$ or $\{j_2,j_3\}$ and add the corresponding edge to T; and so on...



$\Rightarrow$ there is a reordering $(j_1,\dots,j_r)$ of $(i_1,\dots,i_r)$ such that

$$\mathscr{M}\left(\widetilde{L}[i_1,\cdots,i_r]\right) = \prod_{t=1}^{r} \max\left(w(\{j_1,j_t\}),\dots,w(\{j_{t-1},j_t\})\right).$$

Sketch of proof of the normality criterion (2/2)

$$\begin{aligned}\left|\kappa_r(X_n)\right| &\le C_r \sum_{i_1,\dots,i_r} \mathscr{M}\big(\widetilde{L}[i_1,\cdots,i_r]\big)\\ &\le r!\,C_r \sum_{j_1,\dots,j_r} \left(\prod_{t=1}^{r} \max\big(w(\{j_1,j_t\}),\dots,w(\{j_{t-1},j_t\})\big)\right)\end{aligned}$$

(reordering argument from the previous slide)

Sketch of proof of the normality criterion (2/2)

$$\begin{aligned}
\left|\kappa_r(X_n)\right| &\leq C_r \sum_{i_1,\dots,i_r} \mathscr{M}\big(\widetilde{L}[i_1,\cdots,i_r]\big) \\
&\leq r!\, C_r \sum_{j_1,\dots,j_r} \left(\prod_{t=1}^{r} \max\big(w(\{j_1,j_t\}),\dots,w(\{j_{t-1},j_t\})\big) \right) \\
&\leq r!\, C_r \sum_{j_1,\dots,j_{r-1}} \left(\prod_{t=1}^{r-1} \max\big(w(\{j_1,j_t\}),\dots,w(\{j_{t-1},j_t\})\big) \right) \cdot S_{j_1,\dots,j_{r-1}},
\end{aligned}$$

where

$$S_{j_1,\dots,j_{r-1}} = \sum_{j_r} \max\big(w(\{j_1,j_r\}),\dots,w(\{j_{r-1},j_r\})\big)$$

.

Sketch of proof of the normality criterion (2/2)

$$
\begin{aligned}
\left|\kappa_r(X_n)\right| &\le C_r \sum_{i_1,\dots,i_r} \mathscr{M}\big(\widetilde{L}[i_1,\cdots,i_r]\big) \\
&\le r!\, C_r \sum_{j_1,\dots,j_r} \left(\prod_{t=1}^{r} \max\big(w(\{j_1,j_t\}),\dots,w(\{j_{t-1},j_t\})\big) \right) \\
&\le r!\, C_r \sum_{j_1,\dots,j_{r-1}} \left(\prod_{t=1}^{r-1} \max\big(w(\{j_1,j_t\}),\dots,w(\{j_{t-1},j_t\})\big) \right) \cdot S_{j_1,\dots,j_{r-1}},
\end{aligned}
$$

where

$$
\begin{aligned}
S_{j_1,\dots,j_{r-1}} &= \sum_{j_r} \max\big(w(\{j_1,j_r\}),\dots,w(\{j_{r-1},j_r\})\big) \\
&\le \sum_{j_r} w(\{j_1,j_r\}) + \cdots + w(\{j_{r-1},j_r\}) = \widetilde{\deg}(j_1) + \cdots + \widetilde{\deg}(j_{r-1}) \le (r-1) D_n.
\end{aligned}
$$

Sketch of proof of the normality criterion (2/2)

$$\begin{aligned}
\left|\kappa_r(X_n)\right| &\le C_r \sum_{i_1,\dots,i_r} \mathscr{M}\left(\widetilde{L}[i_1,\cdots,i_r]\right) \\
&\le r!\,C_r \sum_{j_1,\dots,j_r} \left(\prod_{t=1}^{r} \max\left(w(\{j_1,j_t\}),\dots,w(\{j_{t-1},j_t\})\right)\right) \\
&\le r!\,C_r \sum_{j_1,\dots,j_{r-1}} \left(\prod_{t=1}^{r-1} \max\left(w(\{j_1,j_t\}),\dots,w(\{j_{t-1},j_t\})\right)\right) \cdot S_{j_1,\dots,j_{r-1}},
\end{aligned}$$

where

$$\begin{aligned}
S_{j_1,\dots,j_{r-1}} &= \sum_{j_r} \max\left(w(\{j_1,j_r\}),\dots,w(\{j_{r-1},j_r\})\right) \\
&\le \sum_{j_r} w(\{j_1,j_r\}) + \cdots + w(\{j_{r-1},j_r\}) = \widetilde{\deg}(j_1) + \cdots + \widetilde{\deg}(j_{r-1}) \le (r-1) D_n.
\end{aligned}$$

Iterating, we get $\left|\kappa_r(X_n)\right| \le r!\,C_r\,N_n\,(r-1)!\,D_n^{r-1}$. We conclude as in the usual case. $\square$

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 dependency graph for some variables Y_α, we have also one for monomials in the Y_α.
(And potentially asymptotic normality for polynomials in the Y_α).

Transition

1. Dependency graphs
   - A motivating example: substrings in random words
   - An asymptotic normality criterion

2. Weighted dependency graphs
   - Definition and an extended normality criterion
   - Back to subwords: Markovian texts
   - Applications in statistical physics

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 $\mathrm{wt}_{\widetilde{L}}(\{Z_{i,s}, Z_{j,t}\}) = |\lambda_2|^{j-i}$ (for $i<j$), where λ_2 is the second eigenvalue of the transition matrix.

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

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

It turns out that 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 $\mathrm{wt}_{\widetilde{L}}\left(\{Z_{i,s}, Z_{j,t}\}\right) = |\lambda_2|^{j-i}$ (for $i < j$), where λ_2 is the second eigenvalue of the transition matrix.

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 $\mathrm{wt}_{\widetilde{L}^m}(Z_{I;S}, Z_{J,T}) = |\lambda_2|^{\mathrm{md}(I,J)}$, 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 (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? $M = 1$, $N_n = \binom{n}{\ell}$, and...

degree: Fix $I = \{i_1, \cdots, i_\ell\}$, we have

$$\sum_J \lambda_2^{\mathrm{md}(I,J)} \leq \sum_J \lambda_2^{|i_1 - j_1|} \leq \binom{n}{\ell - 1} \sum_{j_1} \lambda_2^{|i_1 - j_1|} = \mathcal{O}(n^{\ell-1}).$$

The maximal weighted degree D_n is $\mathcal{O}(n^{\ell-1})$.

variance: $\sigma_n = \sqrt{\mathrm{Var}(X_n)} = (C + o(1)) n^{\ell - 1/2}$, for a computable constant C (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? $M = 1$, $N_n = \binom{n}{\ell}$, and...

degree Fix $I = \{i_1, \cdots, i_\ell\}$, we have

$$\sum_J \lambda_2^{\mathrm{md}(I,J)} \leq \sum_J \lambda_2^{|i_1 - j_1|} \leq \binom{n}{\ell-1} \sum_{j_1} \lambda_2^{|i_1-j_1|} = \mathcal{O}(n^{\ell-1}).$$

The maximal weighted degree D_n is $\mathcal{O}(n^{\ell-1})$.

variance $\sigma_n = \sqrt{\mathrm{Var}(X_n)} = (C + o(1))n^{\ell-1/2}$, for a computable constant C (Bourdon, Vallée, '01).

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

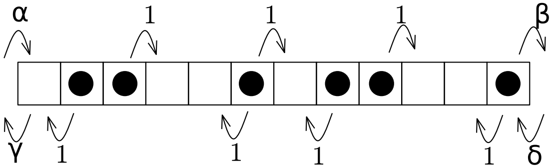
Conclusion: when $C > 0$, the number X_n of occurrences of u in a Markovian text $\boldsymbol{w}$ is asymptotically normal.

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

Transition

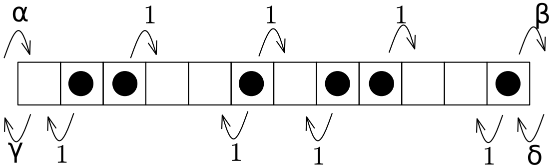
1. Dependency graphs
   - A motivating example: substrings in random words
   - An asymptotic normality criterion
2. Weighted dependency graphs
   - Definition and an extended normality criterion
   - Back to subwords: Markovian texts
   - Applications in statistical physics

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

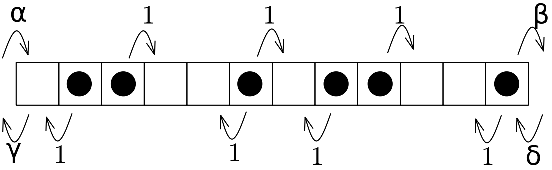
The complete graph on $[N]$ with weight $1/N$ on each edge is a weighted dependency graph for the family $\{\tau_i, 1 \le i \le N\}$.

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 $i_1, \ldots, i_r$;
- 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.

An invariance principle

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 $\mathscr{C}([0,1])$,

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

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

An invariance principle

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 $\mathscr{C}([0,1])$,

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

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

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

An invariance principle

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 $\mathscr{C}([0,1])$,

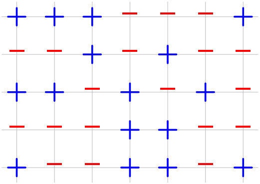
$$\widetilde{X_N}(t) := \frac{X_N(t) - \mathbb{E}X_N(t)}{\sqrt{N}} \xrightarrow{d} Z$$

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

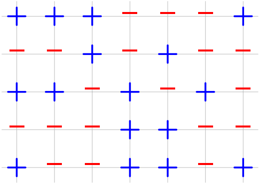


$$\begin{aligned} \mathbb{P}(\omega) &\propto \exp\left[-H(\omega)\right]; \\ H(\omega) &= -\beta \textstyle\sum_{x\sim y} \omega_x \omega_y - h \sum_x \omega_x. \end{aligned}$$

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



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

Theorem

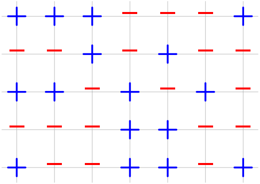
In presence of a magnetic field or at very low or very large temperature, there exists $\varepsilon = \varepsilon(d,h,\beta) > 0$ such that the complete graph on $\mathbb{Z}^d$ with weight $\varepsilon^{\|x-y\|_1}$ on the edge $\{x,y\}$ is a weighted dependency graph for $\{\sigma_x, x \in \mathbb{Z}^d\}$

Concretely, this means that

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

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

Ising model



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

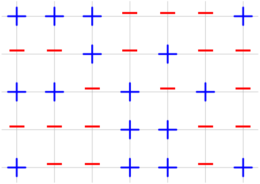
Theorem

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



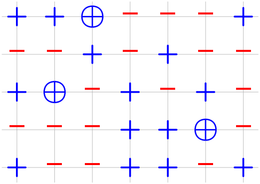
$$\mathbb{P}(\omega) \propto \exp\big[-H(\omega)\big];$$
$$H(\omega) = -\beta \textstyle\sum_{x\sim y} \omega_x \omega_y - h \sum_x \omega_x.$$

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)

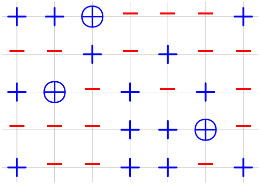
Ising model: asymptotic normality for global patterns



Circled spins:
occurrence of the $+$ pattern 231

(notion inspired from patterns in permutations.)

Ising model: asymptotic normality for global patterns



Circled spins:
occurrence of the $+$ pattern 231

$S_n^{\mathscr{P}} :=$ number of occurrences of $\mathscr{P}$ within $\Lambda_n = [-n, n]^d$.

Theorem (Dousse, F., '19)

Assume $\mathrm{Var}(S_n^{\mathscr{P}}) \geq cst|\Lambda_n|^{2|\mathscr{P}|-2+\eta}$ *for* $\eta > 0$*. Then we have* $S_n^{\mathscr{P}}$ *is asymptotically normal. Moreover, the lower bound of the variance is fulfilled for patterns of only positive spins (as in the example).*

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!