(Weighted) dependency graphs and asymptotic normality

Valentin Féray

CNRS, Institut Élie Cartan de Lorraine (IECL)

L^2 workshop in Probability and Statistics
Luxembourg, September 15th, 2022

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.

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

Available tools:

- computation of characteristic functions;
- Stein's method;
- moment (or cumulant) methods

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.

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

Available tools:

- computation of characteristic functions;
- Stein's method;
- 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

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)

(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

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

where $C_1 > 0$ 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

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

where $C_1 > 0$ 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

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

where $C_1 > 0$ 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 $Y_I = \mathbf{1}[u \text{ occurs at position } I \text{ in } \boldsymbol{w}]$.
Then $X_n = \sum_{I \in \binom{[n]}{\ell}} Y_I$.

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

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 $A_1, A_2 \subset A$:

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

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 $A_1, A_2 \subset A$:

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

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,

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

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

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

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

Note: $\kappa_\ell(X) := \kappa_\ell(X,\dots,X)$ is the usual cumulant of a single r.v.

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:

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

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

  Note: $\kappa_\ell(X) := \kappa_\ell(X,\dots,X)$ is the usual cumulant of a single r.v.

- 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:

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

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

  Note: $\kappa_\ell(X) := \kappa_\ell(X,\dots,X)$ is the usual cumulant of a single r.v.

- 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,

- $\{Y_{n,i}, 1 \leq i \leq N_n\}$ 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 $s \geq 3$.

Fix $r \geq 1$. Then

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

Each summand is 0, unless 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, ...

→ 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,

- $\{Y_{n,i}, 1 \leq i \leq N_n\}$ 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 $s \geq 3$.

Fix $r \geq 1$. Then

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

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

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

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 of random points, ... (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

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

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 uniform random graphs with fixed number of edges ($G(n, M)$) or fixed vertex degrees;

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 uniform random graphs with fixed number of edges ($G(n,M)$) or fixed vertex degrees;
- "local statistics" $\sum_{i_1,\dots,i_r} F(i_1,\dots,i_r,\pi(i_1),\dots,\pi(i_r))$ in a uniform random permutation π;

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 uniform random graphs with fixed number of edges ($G(n,M)$) or fixed vertex degrees;
- "local statistics" $\sum_{i_1,\dots,i_r} F(i_1,\dots,i_r,\pi(i_1),\dots,\pi(i_r))$ in a uniform random permutation π;
- statistical physics: exclusion process (SSEP), 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 uniform random graphs with fixed number of edges ($G(n,M)$) or fixed vertex degrees;
- "local statistics" $\sum_{i_1,\dots,i_r} F(i_1,\dots,i_r,\pi(i_1),\dots,\pi(i_r))$ in a uniform random permutation π;
- statistical physics: exclusion process (SSEP), 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,

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

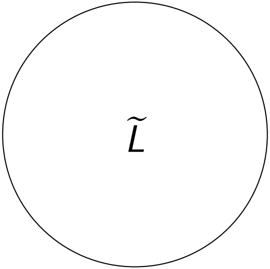
Weighted dependency graphs

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,

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



Weighted dependency graphs

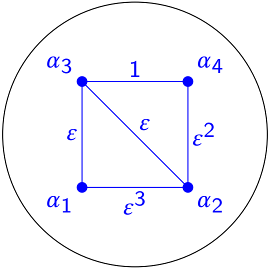
We use weighted graphs, i.e. graphs with a weight in $[0,1]$ on each edge (weight $0 \equiv$ no edge).

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,

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

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



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)

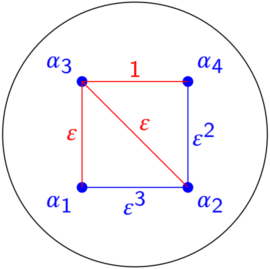
Fix $\boldsymbol{C} = (C_r)_{r\geq 1}$. A weighted graph $\widetilde{L}$ with vertex set A is a **$\boldsymbol{C}$-weighted dependency graph** for the family $\{Y_\alpha, \alpha \in A\}$ if, for any $\alpha_1, \dots, \alpha_r$ in A,

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

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

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

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



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,

- $\{Y_{n,i}, 1 \le i \le N_n\}$ 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.

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.

This helps in proving correctness of weighted dependency graph!

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

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

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

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

$$\sum_J \lambda_2^{\mathrm{md}(I,J)} \leq \sum_{t,s} \left(\sum_J \lambda_2^{|i_t - j_s|} \right) \leq \sum_{t,s} \binom{n}{\ell - 1} \sum_j \lambda_2^{|i_t - j|} = \mathcal{O}(n^{\ell-1}).$$

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

$$\sum_J \lambda_2^{\mathrm{md}(I,J)} \le \sum_{t,s} \left(\sum_J \lambda_2^{|i_t - j_s|} \right) \le \sum_{t,s} \binom{n}{\ell - 1} \sum_j \lambda_2^{|i_t - j|} = \mathcal{O}(n^{\ell-1}).$$

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).

→ 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).

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!