Graphons de probabilités, limites de graphes pondérés aléatoires et chaînes de Markov branchantes cachées

Julien Weibel

sous la direction de Romain Abraham et Jean-François Delmas

Institut Denis Poisson - Université d'Orléans CERMICS - École nationale des ponts et chausées

8 novembre 2024, Orléans







Graphons de probabilités et limites de graphes pondérés aléatoires

1) Graphons de probabilités et limites de graphes pondérés aléatoires

2) Chaînes de Markov branchantes cachées

Les graphes et les réseaux



Figure: Jpatokal. (2009) Carte du traffic aérien autour du monde, circa juin 2009. Wikimedia commons.

Définition d'un graphe



Figure: Un exemple d'un graphe étiqueté non-orienté (à gauche) et d'un graphe étiqueté orienté (à droite).

Un graphe (simple orienté) G = (V, E) est composé :

- d'un ensemble de points V appelés les sommets
- d'un ensemble d'*arêtes* $E \subset V \times V$ tel que :
 - *E* n'a pas de boucle, i.e. $(x, x) \notin E$.

De plus, le graphe G est dit non-orienté si :

• E est symétrique, i.e. $(x, y) \in E$ si et seulment si $(y, x) \in E$.

Graphe, matrice d'adjacence et image pixelisée

L'ensemble d'arêtes E peut être résumé par une matrice M, appelée *matrice d'adjacence*, et définie par :

$$\forall u, v \in V, \quad M(u, v) = \mathbb{1}(uv \in E).$$



Figure: Un exemple de graphe, sa matrice d'adjacence et son image pixelisée.

Un grand graphe dense est un graphe fini (non-orienté) G avec un nombre de sommets n grand, et un nombre d'arêtes $\Omega(n^2)$.

• Problème : le coût pour représenter entièrement ces graphes est prohibitif

- Problème : le coût pour représenter entièrement ces graphes est prohibitif
- Propriété importante : l'échantillonnage de sous-graphes
 k ≪ n, X₁, · · · , X_k iid distribués uniformément sur V
 G(k, G) = G[X₁, · · · , X_k] le sous-graphe induit de G



- Problème : le coût pour représenter entièrement ces graphes est prohibitif
- Propriété importante : l'échantillonnage de sous-graphes
 k ≪ n, X₁, · · · , X_k iid distribués uniformément sur V
 G(k, G) = G[X₁, · · · , X_k] le sous-graphe induit de G



- Problème : le coût pour représenter entièrement ces graphes est prohibitif
- Propriété importante : l'échantillonnage de sous-graphes
 k ≪ n, X₁, · · · , X_k iid distribués uniformément sur V
 G(k, G) = G[X₁, · · · , X_k] le sous-graphe induit de G



- Problème : le coût pour représenter entièrement ces graphes est prohibitif
- Propriété importante : l'échantillonnage de sous-graphes
 k ≪ n, X₁, · · · , X_k iid distribués uniformément sur V
 G(k, G) = G[X₁, · · · , X_k] le sous-graphe induit de G



Un grand graphe dense est un graphe fini (non-orienté) G avec un nombre de sommets n grand, et un nombre d'arêtes $\Omega(n^2)$.

- Problème : le coût pour représenter entièrement ces graphes est prohibitif
- Propriété importante : l'échantillonnage de sous-graphes
 k ≪ n, X₁, · · · , X_k iid distribués uniformément sur V
 G(k, G) = G[X₁, · · · , X_k] le sous-graphe induit de G



 Question : Soit (G_n)_{n∈N} suite de grands graphes denses telle que G(k, G_n) ^(d)→ H_k un graphe aléatoire pour tout k ≥ 2. Peut-on trouver un objet limite (continu) pour cette suite ?

Quelques exemples de convergence (i)

Graphe d'Erdős-Rényi :

- *n* sommets, i.e. $V = \{1, \dots, n\}$,
- chaque arête (u, v) est présente indépendamment des autres avec probabilité p = 1/2.



Figure: De gauche à droite : images pixelisées de simulations du graphe d'Erdős-Rényi avec n = 10, 50 et 100 sommets, et la limite attendue.

Quelques exemples de convergence (ii)

Modèle à blocs stochastiques (SBM) :

- *n* sommets, i.e. $V = \{1, \cdots, n\}$,
- chaque sommet u reçoit indépendamment des autres un type aléatoire X_u uniformément dans {1,2},
- chaque arête (u, v) est présente indépendamment des autres avec probabilité p = 3/4 si X_u = X_v, et q = 1/4 sinon.



Figure: De gauche à droite : images pixelisées de simulations du modèle à blocs stochastiques avec n = 10, 50 et 100 sommets, et la limite attendue.

La limite ne dépend pas de l'ordre des sommets

Graphes bipartites avec différents étiquetages des sommets :



SBMs avec étiquetage par classe ou aléatoire :



Graphons : objets limites des grands graphes denses

Un graphon est une fonction mesurable symétrique $w : [0,1]^2 \rightarrow [0,1].$

Pour deux « sommets » $x, y \in [0, 1]$, w(x, y) corresponds à la densité moyenne d'arêtes entre x et y.

Idée : un graphon est la limite de matrices d'adjacences de graphes.



Les graphons ont été introduits et étudiés par Lovász et ses co-auteurs, voir [Lovász, 2012].

Graphes pondérés



Figure: Exemple d'un graphe pondéré étiqueté.

Un graphe pondéré est :

- un graphe orienté (V, E),
- où chaque arête e ∈ E possède une décoration z_e dans un espace polonais Z (i.e. un espace métrique complet séparable).

Définition d'un graphon de probabilités

Un graphon de probabilités est un noyau de probabilités W de $[0, 1]^2$ dans l'espace des mesures de probabilités $\mathcal{M}_1(\mathbf{Z})$.

Pour deux « sommets » $x, y \in [0, 1]$, on a :

Graphons (classiques) :

w(x, y) corresponds à la densité moyenne d'arêtes entre x et y.

Graphons de probabilités :

W(x, y; dz) corresponds à la distribution des poids z des arêtes entre x et y.

Cas non-pondéré, i.e. $\mathbf{Z} = \{0, 1\}$:

$$W(x,y;\cdot) = w(x,y)\delta_1 + (1-w(x,y))\delta_0.$$

Définition d'un graphon de probabilités

Un graphon de probabilités est un noyau de probabilités W de $[0, 1]^2$ dans l'espace des mesures de probabilités $\mathcal{M}_1(\mathbf{Z})$.

Pour deux « sommets » $x, y \in [0, 1]$, on a :

Graphons (classiques) :

w(x, y) corresponds à la densité moyenne d'arêtes entre x et y. Graphons de probabilités :

W(x, y; dz) corresponds à la distribution des poids z des arêtes entre x et y.

Cas discret, i.e. $\mathbf{Z} = \{0, 1, \cdots, D\}$:

$$W(x,y;\cdot) = \sum_{i=0}^D w_i(x,y) \delta_i$$
 avec $\sum_{i=0}^D w_i(x,y) = 1$

Ré-étiquetage pour les graphons de probabilités

On dit qu'une fonction $\varphi : [0,1] \rightarrow [0,1]$ préserve la mesure (de Lebesgue) si pour tout fonction mesurable bornée f sur [0,1] :

$$\int_{[0,1]} f(\varphi(x)) \, \mathrm{d}x = \int_{[0,1]} f(x) \, \mathrm{d}x.$$

Pour un graphon de probabilités W, on définit son *ré-étiquetage* par une fonction préservant la mesure φ comme :

$$W^{\varphi}(x,y;\cdot) = W(\varphi(x),\varphi(y);\cdot).$$

Ré-étiquetage pour les graphons de probabilités

On dit qu'une fonction $\varphi : [0,1] \rightarrow [0,1]$ préserve la mesure (de Lebesgue) si pour tout fonction mesurable bornée f sur [0,1] :

$$\int_{[0,1]} f(\varphi(x)) \, \mathrm{d}x = \int_{[0,1]} f(x) \, \mathrm{d}x.$$

Pour un graphon de probabilités W, on définit son *ré-étiquetage* par une fonction préservant la mesure φ comme :

$$W^{\varphi}(x,y;\cdot) = W(\varphi(x),\varphi(y);\cdot).$$

On note \widetilde{W}_1 l'espace des graphons de probabilité où on identifie W, U s'il existe des fonctions préservant la mesure φ, ψ telles que :

$$W^{arphi} = U^{\psi}$$
 presque partout sur $[0,1]^2$.

La distance de coupe pour les graphons

• La distance de coupe pour les matrices d'adjacence de graphes non-pondérés est définie par :

$$d_{\Box}(M,N) = \sup_{S,T \subset \{1,\cdots,n\}} \left| \sum_{i \in S, j \in T} M_{i,j} - N_{i,j} \right|.$$

1

La distance de coupe pour les graphons

• La *distance de coupe* pour les matrices d'adjacence de graphes non-pondérés est définie par :

$$d_{\Box}(M,N) = \sup_{S,T \subset \{1,\cdots,n\}} \left| \sum_{i \in S, j \in T} M_{i,j} - N_{i,j} \right|.$$

• La distance de coupe pour les graphons est définie par :

$$\delta_{\Box}(w, u) = \inf_{\varphi, \psi} \sup_{S, T \subset [0,1]} \left| \int_{S \times T} w^{\varphi}(x, y) - u^{\psi}(x, y) \, \mathrm{d}x \mathrm{d}y \right|,$$

où φ et ψ préservent la mesure sur [0, 1].

La distance de coupe pour les graphons

• La *distance de coupe* pour les matrices d'adjacence de graphes non-pondérés est définie par :

$$d_{\Box}(M,N) = \sup_{S,T \subset \{1,\cdots,n\}} \left| \sum_{i \in S, j \in T} M_{i,j} - N_{i,j} \right|.$$

• La distance de coupe pour les graphons est définie par :

$$\delta_{\Box}(w, u) = \inf_{\varphi, \psi} \sup_{S, T \subset [0,1]} \left| \int_{S \times T} w^{\varphi}(x, y) - u^{\psi}(x, y) \, \mathrm{d}x \mathrm{d}y \right|,$$

où φ et ψ préservent la mesure sur [0, 1].

Théorème 1 ([Lovász et Szegedy, 2007])

L'espace des graphons quotienté par la relation d'équivalence $\delta_{\Box}(u, w) = 0$ et muni de la distance de coupe δ_{\Box} est compact.

Soit d_m distance générant la topologie faible (convergence en loi) sur l'espace des mesures de sous-probabilitiés $\mathcal{M}_{\leq 1}(\mathbf{Z})$.

Soit d_m distance générant la topologie faible (convergence en loi) sur l'espace des mesures de sous-probabilitiés $\mathcal{M}_{\leq 1}(\mathbf{Z})$.

Exemples: la distance de Prohorov $d_{\mathcal{P}}$:

$$\begin{split} d_{\mathcal{P}}(\mu,\nu) &= \inf\{\varepsilon > 0: \forall A \in \mathcal{B}(\mathbf{Z}), \ \mu(A) \leq \nu(A^{\varepsilon}) + \varepsilon \\ & \text{et} \quad \nu(A) \leq \mu(A^{\varepsilon}) + \varepsilon\}, \end{split}$$

et la norme de Kantorovitch-Rubinstein $\|\cdot\|_{\mathsf{KR}}$:

$$\|\mu\|_{\mathsf{KR}} = \sup\left\{\int_{\mathbf{Z}} f \, \mathrm{d}\mu : f \text{ est 1-Lipschitz et } \|f\|_{\infty} \leq 1
ight\}.$$

Soit $d_{\rm m}$ distance générant la topologie faible (convergence en loi) sur l'espace des mesures de sous-probabilitiés $\mathcal{M}_{<1}(\mathbf{Z})$.

Pour $S, T \subset [0, 1]$, on définit la mesure de sous-probabilité $W(S, T; \cdot) = \int_{S \times T} W(x, y; \cdot) \, dx dy.$

 \bullet La distance de coupe pour les graphons de probabilités est définie sur $\widetilde{\mathcal{W}_1}$ par :

 $\delta_{\Box,\mathrm{m}}(W,U) = \min_{\varphi,\psi} \sup_{S,T \subset [0,1]} d_{\mathrm{m}}(W^{\varphi}(S,T;\cdot), U^{\psi}(S,T;\cdot)),$

où φ et ψ préservent la mesure sur [0,1].



Si d_m est complète, alors $(\mathcal{W}_1, \delta_{\Box,m})$ est un espace polonais.

Remarque : Par la suite, on suppose $d_{\rm m}$ non-complète.

Propriétés topologique de $\delta_{\Box,m}$ (ii)

On dit qu'une distance d_m sur $\mathcal{M}_{\leq 1}(\mathbf{Z})$ est quasi-convexe si pour tous $\mu_1, \mu_2, \nu_1, \nu_2 \in \mathcal{M}_{\leq 1}(\mathbf{Z})$ et $\alpha \in [0, 1]$, on a :

 $d_{\mathrm{m}}(\alpha\mu_1+(1-\alpha)\mu_2,\alpha\nu_1+(1-\alpha)\nu_2)\leq \max(d_{\mathrm{m}}(\mu_1,\nu_1),d_{\mathrm{m}}(\mu_2,\nu_2)).$

• Les distances de Prohorov $d_{\mathcal{P}}$ est quasi-convexe. La norme de Kantorovitch-Rubinstein induit une distance quasi-convexe.

Propriétés topologique de $\delta_{\Box,m}$ (ii)

On dit qu'une distance d_m sur $\mathcal{M}_{\leq 1}(\mathbf{Z})$ est quasi-convexe si pour tous $\mu_1, \mu_2, \nu_1, \nu_2 \in \mathcal{M}_{\leq 1}(\mathbf{Z})$ et $\alpha \in [0, 1]$, on a :

 $d_{\mathrm{m}}(\alpha\mu_1+(1-\alpha)\mu_2,\alpha\nu_1+(1-\alpha)\nu_2)\leq \max(d_{\mathrm{m}}(\mu_1,\nu_1),d_{\mathrm{m}}(\mu_2,\nu_2)).$

• Les distances de Prohorov $d_{\mathcal{P}}$ est quasi-convexe. La norme de Kantorovitch-Rubinstein induit une distance quasi-convexe.

Théorème 3 (Équivalence des topologies induites par $\delta_{\Box,m}$ sur \widetilde{W}_1 , [Abraham, Delmas, W., 2023])

La topologie sur \widetilde{W}_1 induite par la distance $\delta_{\Box,m}$ ne dépend pas du choix de la distance d_m sur $\mathcal{M}_{\leq 1}(\mathbf{Z})$, dès que d_m est quasi-convexe.

Tension et compacité pour les graphons de probabilités

Un sous-ensemble de graphons de probabilités \mathcal{K} est dit *tendu* si le sous-ensemble de mesures de probabilités $\{M_W : W \in \mathcal{K}\} \subset \mathcal{M}_1(\mathbf{Z})$ est tendu, où :

$$M_W(dz) = W([0,1]^2; dz) = \int_{[0,1]^2} W(x,y; dz) dxdy.$$

Théorème 4 (Théorème de tension-compacité pour W_1 , [Abraham, Delmas, W., 2023])

Considérons la topologie du Théorème 3 sur \widetilde{W}_1 .

- Si une suite de graphons de probabilités est tendue, alors elle possède une sous-suite convergente.
- Si Z est compact, alors l'espace $\widehat{\mathcal{W}}_1$ est compact.

Échantillonnage à partir d'un graphon de probabilités

Soit X_1, \ldots, X_k des variables aléatoires indépendantes et uniformément distribuées sur [0, 1]. On définit le graphe pondéré aléatoire $\mathbb{G}(k, W)$ avec :

- k sommets, i.e. $V = \{1, \cdots, k\}$,
- chaque arête (i, j) est décorée indépendamment avec une décoration aléatoire Z_{i,j} distribuée selon W(X_i, X_j; dz).

Échantillonnage à partir d'un graphon de probabilités

Soit X_1, \ldots, X_k des variables aléatoires indépendantes et uniformément distribuées sur [0, 1]. On définit le graphe pondéré aléatoire $\mathbb{G}(k, W)$ avec :

- k sommets, i.e. $V = \{1, \cdots, k\}$,
- chaque arête (i, j) est décorée indépendamment avec une décoration aléatoire Z_{i,j} distribuée selon W(X_i, X_j; dz).



Proposition 5 (Convergence des graphes échantillonnés, [Abraham, Delmas, W., 2023])

Soit W un graphon de probabilités. Alors, p.s. $(\mathbb{G}(k, W))_{k \in \mathbb{N}}$ converge vers W pour la topologie du Théorème 3.

En particulier, tout graphon de probabilités est une $\delta_{\Box,m}\text{-limite}$ de graphes pondérés.

Convergence pour $\delta_{\Box,m}$ et échantillonnage

Théorème 6 (Caractérisation de la topologie induite par $\delta_{\Box,m}$, [Abraham, Delmas, W., 2023])

Soient $(W_n)_{n \in \mathbb{N}}$ et W des graphons de probabilités dans \widetilde{W}_1 . Les propriétés suivantes sont équivalentes :

- $\mathbb{G}(k, W_n) \xrightarrow{(d)} \mathbb{G}(k, W)$ pour tout $k \ge 2$.
- **2** $W_n \rightarrow W$ pour la topologie du Théorème 3.

Convergence pour $\delta_{\Box,m}$ et échantillonnage

Théorème 6 (Caractérisation de la topologie induite par $\delta_{\Box,m}$, [Abraham, Delmas, W., 2023])

Soient $(W_n)_{n \in \mathbb{N}}$ et W des graphons de probabilités dans \widetilde{W}_1 . Les propriétés suivantes sont équivalentes :

- $\mathbb{G}(k, W_n) \xrightarrow{(d)} \mathbb{G}(k, W)$ pour tout $k \ge 2$.
- **2** $W_n \rightarrow W$ pour la topologie du Théorème 3.

Théorème 7 (Les suites faiblement convergentes convergent pour $\delta_{\Box,m}$, [Abraham, Delmas, W., 2023])

Supposons que
$$\mathbb{G}(k, W_n) \xrightarrow{(d)} H_k$$
 pour tout $k \ge 2$. Alors :

• $\exists W \in \widetilde{W}_1$, tel que $W_n \to W$ pour la topologie du Théorème 3,

• et
$$\mathbb{G}(k, W) \stackrel{(d)}{=} H_k$$
 pour tout $k \ge 2$.

Ainsi, les graphons de probabilités sont les objets limites recherchés pour les suites convergentes de grands graphes pondérés denses.

Perspectives

• L'équivalence des topologies pour la distance de coupe $\delta_{\Box,m}$ est elle valable pour tous les choix de la distance d_m sur $\mathcal{M}_{\leq 1}(\mathbf{Z})$?
Perspectives

- L'équivalence des topologies pour la distance de coupe $\delta_{\Box,m}$ est elle valable pour tous les choix de la distance d_m sur $\mathcal{M}_{\leq 1}(\mathbf{Z})$?
- Limites locale et d'échelle de l'arbre couvrant minimal (MST) d'une suite de graphes pondérés aléatoires.
 - limite poids total MST [Hladký et Viswanathan, 2023]
 - limite locale et d'échelle arbre couvrant uniforme (UST) [Hladký et al., 2018, Archer et al., 2024]

- L'équivalence des topologies pour la distance de coupe $\delta_{\Box,m}$ est elle valable pour tous les choix de la distance d_m sur $\mathcal{M}_{\leq 1}(\mathbf{Z})$?
- Limites locale et d'échelle de l'arbre couvrant minimal (MST) d'une suite de graphes pondérés aléatoires.
 - limite poids total MST [Hladký et Viswanathan, 2023]
 - limite locale et d'échelle arbre couvrant uniforme (UST) [Hladký et al., 2018, Archer et al., 2024]
- Détection de communauté pour le modèle à blocs stochastiques pondérés ou décorés (weighted/labeled SBMs).
 - régime *sparse* avec degré moyen constant ou log *n*

- L'équivalence des topologies pour la distance de coupe $\delta_{\Box,m}$ est elle valable pour tous les choix de la distance d_m sur $\mathcal{M}_{\leq 1}(\mathbf{Z})$?
- Limites locale et d'échelle de l'arbre couvrant minimal (MST) d'une suite de graphes pondérés aléatoires.
 - limite poids total MST [Hladký et Viswanathan, 2023]
 - limite locale et d'échelle arbre couvrant uniforme (UST) [Hladký et al., 2018, Archer et al., 2024]
- Détection de communauté pour le modèle à blocs stochastiques pondérés ou décorés (weighted/labeled SBMs).
 - régime *sparse* avec degré moyen constant ou log *n*
- Estimation de graphons de probabilités à partir de graphes échantillonnés. [Dufour et Olhede, 2024]

- L'équivalence des topologies pour la distance de coupe $\delta_{\Box,m}$ est elle valable pour tous les choix de la distance d_m sur $\mathcal{M}_{\leq 1}(\mathbf{Z})$?
- Limites locale et d'échelle de l'arbre couvrant minimal (MST) d'une suite de graphes pondérés aléatoires.
 - limite poids total MST [Hladký et Viswanathan, 2023]
 - limite locale et d'échelle arbre couvrant uniforme (UST) [Hladký et al., 2018, Archer et al., 2024]
- Détection de communauté pour le modèle à blocs stochastiques pondérés ou décorés (weighted/labeled SBMs).
 - régime *sparse* avec degré moyen constant ou log *n*
- Estimation de graphons de probabilités à partir de graphes échantillonnés. [Dufour et Olhede, 2024]
- Application pour des systèmes mean-field. [Ayi et Duteil, 2023]

Chaînes de Markov branchantes cachées

I Graphons de probabilités et limites de graphes pondérés aléatoires

2 Chaînes de Markov branchantes cachées

De la détection de communauté aux processus de Markov cachés sur des arbres

Il existe un couplage (voir [Lelarge et al., 2015]) entre :

- la détection de communauté sur un SBM avec degré moyen constant,
- la reconstruction sur un arbre de Bienaymé-Galton-Watson.



Figure: Exemple d'arbre enraciné planaire avec notation de Neveu.

De la détection de communauté aux processus de Markov cachés sur des arbres

Il existe un couplage (voir [Lelarge et al., 2015]) entre :

- la détection de communauté sur un SBM avec degré moyen constant,
- la reconstruction sur un arbre de Bienaymé-Galton-Watson.



Figure: Exemple d'arbre enraciné planaire avec notation de Neveu.

De la détection de communauté aux processus de Markov cachés sur des arbres

Il existe un couplage (voir [Lelarge et al., 2015]) entre :

- la détection de communauté sur un SBM avec degré moyen constant,
- la reconstruction sur un arbre de Bienaymé-Galton-Watson.



Figure: Exemple d'arbre enraciné planaire avec notation de Neveu.

Définition : Chaîne de Markov cachée (HMM)

Une chaîne de Markov cachée est composée de deux processus :

- un processus caché X = (X_n)_{n∈ℕ} qui est une chaîne de Markov,
- un processus observé Y = (Y_n)_{n∈ℕ} tel que, conditionnellement à X, les variables (Y_n)_{n∈ℕ} sont indépendantes et Y_n dépend seulement de X_n.

$$Q(x_{i-1}, x_i) = \mathbb{P}(X_i = x_i | X_{i-1} = x_{i-1})$$

$$G(x_i, y_i) = \mathbb{P}(Y_i = y_i | X_i = x_i)$$

$$\mathbb{P}(X_{0:n} = x_{0:n}, Y_{0:n} = y_{0:n}) = \nu(x_0)G(x_0, y_0)\prod_{i=1}^n Q(x_{i-1}, x_i)G(x_i, y_i)$$



Définition : Chaîne de Markov branchante

Une chaîne de Markov branchante $X = (X_u, u \in T)$ est un processus aléatoire tel que pour tout sous-arbre fini $T_0 \subset T$, nous avons :

$$\mathbb{P}(X_{T_0} \in \mathrm{d} x_{T_0}) = \nu(\mathrm{d} x_{\partial}) \prod_{u \in T_0 \setminus \{\partial\}} Q(x_{\mathrm{p}(u)}; \mathrm{d} x_u),$$

où Q est le noyau de transition et ν la distribution initiale. Notation: p(u) est le sommet parent du sommet u. ∂ est la racine.



Définition : Chaîne de Markov branchante

Une chaîne de Markov branchante $X = (X_u, u \in T)$ est un processus aléatoire tel que pour tout sous-arbre fini $T_0 \subset T$, nous avons :

$$\mathbb{P}(X_{\mathcal{T}_0} \in \mathrm{d} x_{\mathcal{T}_0}) = \nu(\mathrm{d} x_{\partial}) \prod_{u \in \mathcal{T}_0 \setminus \{\partial\}} Q(x_{\mathrm{p}(u)}; \mathrm{d} x_u),$$

où Q est le noyau de transition et ν la distribution initiale. Notation: p(u) est le sommet parent du sommet u. ∂ est la racine.



Chaîne de Markov branchante et variance minimale

- Soit μ une mesure invariante pour Q.
- Soit f une fonction dans $L^2(\mu)$.
- Pour un arbre fini *T* et une chaîne de Markov branchante (X_u, u ∈ T) de paramètre (Q, μ), nous définissons la moyenne empirique normalisée :

$$\bar{M}_T(f) = \frac{1}{|T|} \sum_{u \in T} f(X_u).$$

Théorème 8 (Théorème ergodique pour les chaînes de Markov branchantes pour des arbres de formes arbitraires, [W., 2024b])

Sous certaines hypothèses sur Q, f et $(T_n)_{n \in \mathbb{N}}$, et en particulier que $\mathbb{P}(d(U_n, V_n) \leq k) \xrightarrow[n \to \infty]{} 0$ pour tout $k \in \mathbb{N}$ où U_n, V_n sont indépendants et uniformément distribués sur T_n , nous avons :

$$ar{M}_{\mathcal{T}_n}(f) \mathop{\longrightarrow}\limits_{n o \infty} \int f \, \mathrm{d} \mu \quad dans \; L^2(\mu).$$

Chaîne de Markov branchante et variance minimale

- Soit μ une mesure invariante pour Q.
- Soit f une fonction dans $L^2(\mu)$.
- Pour un arbre fini T et une chaîne de Markov branchante $(X_u, u \in T)$ de paramètre (Q, μ) ,

nous définissons la moyenne empirique normalisée :

$$\bar{M}_T(f) = \frac{1}{|T|} \sum_{u \in T} f(X_u).$$

Proposition 8 (Le graphe ligne a une variance minimale, [W., 2024b])

Supposons que (Q, μ) induise une chaîne de Markov réversible.

• Le minimum de l'application $T \mapsto \operatorname{Var}(\overline{M}_T(f))$ parmi les arbres finis de cardinal donné est atteint par le graphe ligne (i.e. la chaîne de Markov).

De plus, on a $\mathbb{E}\left[\bar{M}_T(f)\right] = \int f \, \mathrm{d}\mu$.

Définition: Chaîne de Markov cachée branchante (HMT)

Une chaîne de Markov cachée branchante (HMT) est composée de deux processus :

- un processus caché X = (X_u, u ∈ T) qui est une chaîne de Markov branchante indexée par l'arbre binaire T,
- un processus observé Y = (Y_u, u ∈ T) tel que, conditionnellement à X, les variables Y = (Y_u, u ∈ T) sont indépendantes et Y_u dépend seulement de X_u.

$$\mathbb{P}(Y_{T_n} \in \mathrm{d} y_{T_n} \,|\, X_{T_n} = x_{T_n}) = \Pi_{u \in T_n} G(x_u; \mathrm{d} y_u).$$



Définition: Chaîne de Markov cachée branchante (HMT)

Une chaîne de Markov cachée branchante (HMT) est composée de deux processus :

- un processus caché X = (X_u, u ∈ T) qui est une chaîne de Markov branchante indexée par l'arbre binaire T,
- un processus observé Y = (Y_u, u ∈ T) tel que, conditionnellement à X, les variables Y = (Y_u, u ∈ T) sont indépendantes et Y_u dépend seulement de X_u.

$$\mathbb{P}(Y_{T_n} \in \mathrm{d} y_{T_n} \,|\, X_{T_n} = x_{T_n}) = \prod_{u \in T_n} G(x_u; \mathrm{d} y_u).$$



Considérons le cas où le HMT est paramétré par un certain $\theta \in \Theta \subset \mathbb{R}^d$ qui détermine les noyaux de transition Q_{θ} et G_{θ} mais pas la distribution initiale ν , c'est-à-dire :

Considérons le cas où le HMT est paramétré par un certain $\theta \in \Theta \subset \mathbb{R}^d$ qui détermine les noyaux de transition Q_{θ} et G_{θ} mais pas la distribution initiale ν , c'est-à-dire :

$$\mathbb{P}_{\theta}(X_{T_n} \in \mathrm{d} x_{T_n}) = \nu(\mathrm{d} x_{\partial}) \prod_{u \in T_n \setminus \{\partial\}} Q_{\theta}(x_{\mathrm{p}(u)}, x_u)$$
$$\mathbb{P}_{\theta}(Y_{T_n} \in \mathrm{d} y_{T_n} | X_{T_n} = x_{T_n}) = \prod_{u \in T_n} G_{\theta}(x_u, y_u).$$

Le processus X (resp. Y) est à valeurs dans un espace métrique général \mathcal{X} (resp. \mathcal{Y}).

Considérons le cas où le HMT est paramétré par un certain $\theta \in \Theta \subset \mathbb{R}^d$ qui détermine les noyaux de transition Q_{θ} et G_{θ} mais pas la distribution initiale ν , c'est-à-dire :

$$\mathbb{P}_{\theta}(X_{T_n} \in \mathrm{d} x_{T_n}) = \nu(\mathrm{d} x_{\partial}) \prod_{u \in T_n \setminus \{\partial\}} Q_{\theta}(x_{\mathrm{p}(u)}, x_u)$$
$$\mathbb{P}_{\theta}(Y_{T_n} \in \mathrm{d} y_{T_n} | X_{T_n} = x_{T_n}) = \prod_{u \in T_n} G_{\theta}(x_u, y_u).$$

Le processus X (resp. Y) est à valeurs dans un espace métrique général \mathcal{X} (resp. \mathcal{Y}).

Objectif : À partir de l'observation seule des $(Y_u, u \in T_n)$, nous voulons estimer le vrai paramètre θ^* . Nous analysons la log-vraisemblance partant de $X_{\partial} = x$ (non observé), c'est-à-dire :

$$\ell_{n,x}(\theta; Y_{T_n}) := \log p_{\theta}(Y_u, u \in T_n | X_{\partial} = x).$$

L'estimateur du maximum de vraisemblance est alors défini par :

$$\hat{\theta}_{n,x} \in \operatorname{argmax}_{\theta \in \Theta} \ell_{n,x}(\theta).$$

Cas du HMM [Douc et al., 2004] :

$$\ell_n(\theta) = \frac{1}{n} \sum_{k=1}^n \log p_\theta(Y_k \mid Y_{1:k-1})$$
$$\approx \frac{1}{n} \sum_{k=1}^n \log p_\theta(Y_k \mid Y_{-\infty:k-1})$$
$$\approx \mathbb{E}_{\theta^*}[\log p_\theta(Y_1 \mid Y_{-\infty:0})]$$

 $1 \longrightarrow 2 \longrightarrow 3 \longrightarrow 4 \longrightarrow 5 \longrightarrow \cdots$

$$\ell_n(heta) = rac{1}{|\mathcal{T}_n|} \sum_{u \in \mathcal{T}_n} \log p_ heta(Y_u \mid Y_{\{v:v < u\}}) \ pprox rac{1}{|\mathcal{T}_n|} \sum_{u \in \mathcal{T}_n} \log p_ heta(Y_u \mid Y_{\Delta(u,\infty)}) \ pprox \mathbb{E}_{\Delta(\partial,\infty)} \otimes \mathbb{E}_{ heta^*}[\log p_ heta(Y_\partial \mid Y_{\Delta(\partial,\infty)})]$$



Cas du HMM [Douc et al., 2004] :

$$\ell_n(\theta) = \frac{1}{n} \sum_{k=1}^n \log p_\theta(Y_k \mid Y_{1:k-1})$$
$$\approx \frac{1}{n} \sum_{k=1}^n \log p_\theta(Y_k \mid Y_{-\infty:k-1})$$
$$\approx \mathbb{E}_{\theta^*}[\log p_\theta(Y_1 \mid Y_{-\infty:0})]$$

$$\ell_n(heta) = rac{1}{|\mathcal{T}_n|} \sum_{u \in \mathcal{T}_n} \log p_ heta(Y_u \mid Y_{\{v:v < u\}}) \ pprox rac{1}{|\mathcal{T}_n|} \sum_{u \in \mathcal{T}_n} \log p_ heta(Y_u \mid Y_{\Delta(u,\infty)}) \ pprox \mathbb{E}_{\Delta(\partial,\infty)} \otimes \mathbb{E}_{ heta^*}[\log p_ heta(Y_\partial \mid Y_{\Delta(\partial,\infty)})]$$



Cas du HMM [Douc et al., 2004] :

$$\ell_n(\theta) = \frac{1}{n} \sum_{k=1}^n \log p_\theta(Y_k \mid Y_{1:k-1})$$
$$\approx \frac{1}{n} \sum_{k=1}^n \log p_\theta(Y_k \mid Y_{-\infty:k-1})$$
$$\approx \mathbb{E}_{\theta^*}[\log p_\theta(Y_1 \mid Y_{-\infty:0})]$$

$$\ell_n(\theta) = \frac{1}{|T_n|} \sum_{u \in T_n} \log p_{\theta}(Y_u \mid Y_{\{v: v < u\}})$$
$$\approx \frac{1}{|T_n|} \sum_{u \in T_n} \log p_{\theta}(Y_u \mid Y_{\Delta(u,\infty)})$$
$$\approx \mathbb{E}_{\Delta(\partial,\infty)} \otimes \mathbb{E}_{\theta^*}[\log p_{\theta}(Y_\partial \mid Y_{\Delta(\partial,\infty)})]$$



Cas du HMM [Douc et al., 2004] :

$$\ell_n(\theta) = \frac{1}{n} \sum_{k=1}^n \log p_{\theta}(Y_k \mid Y_{1:k-1})$$
$$\approx \frac{1}{n} \sum_{k=1}^n \log p_{\theta}(Y_k \mid Y_{-\infty:k-1})$$
$$\approx \mathbb{E}_{\theta^*}[\log p_{\theta}(Y_1 \mid Y_{-\infty:0})]$$

$$\ell_n(heta) = rac{1}{|\mathcal{T}_n|} \sum_{u \in \mathcal{T}_n} \log p_ heta(Y_u \mid Y_{\{v:v < u\}}) \ pprox rac{1}{|\mathcal{T}_n|} \sum_{u \in \mathcal{T}_n} \log p_ heta(Y_u \mid Y_{\Delta(u,\infty)}) \ pprox \mathbb{E}_{\Delta(\partial,\infty)} \otimes \mathbb{E}_{ heta^*}[\log p_ heta(Y_\partial \mid Y_{\Delta(\partial,\infty)})]$$



Cas du HMM [Douc et al., 2004] :

$$\ell_n(\theta) = \frac{1}{n} \sum_{k=1}^n \log p_\theta(Y_k \mid Y_{1:k-1})$$
$$\approx \frac{1}{n} \sum_{k=1}^n \log p_\theta(Y_k \mid Y_{-\infty:k-1})$$
$$\approx \mathbb{E}_{\theta^*}[\log p_\theta(Y_1 \mid Y_{-\infty:0})]$$

71

$$\ell_n(heta) = rac{1}{|\mathcal{T}_n|} \sum_{u \in \mathcal{T}_n} \log p_ heta(Y_u \mid Y_{\{v:v < u\}})$$
 $pprox rac{1}{|\mathcal{T}_n|} \sum_{u \in \mathcal{T}_n} \log p_ heta(Y_u \mid Y_{\Delta(u,\infty)})$
 $pprox \mathbb{E}_{\Delta(\partial,\infty)} \otimes \mathbb{E}_{ heta^*}[\log p_ heta(Y_\partial \mid Y_{\Delta(\partial,\infty)})]$



Nous prouvons nos résultats sous les mêmes hypothèses utilisées que dans [Douc, Moulines et Rydén, 2004].

Ceci implique les propriétés de pertes de mémoire exponentielles uniformes avec un **taux de mélange** ρ de la distribution initiale conditionnée aux observations ($Y_u, u \in T_n$).

Nous prouvons nos résultats sous les mêmes hypothèses utilisées que dans [Douc, Moulines et Rydén, 2004].

Ceci implique les propriétés de pertes de mémoire exponentielles uniformes avec un **taux de mélange** ρ de la distribution initiale conditionnée aux observations ($Y_u, u \in T_n$).

Théorème 9 (Consistance forte du MLE, [W., 2024a])

Sous les hypothèses de [Douc et al., 2004], pour tout $x \in \mathcal{X}$, le MLE $\hat{\theta}_{n,x}$ est fortement consistant, c'est-à-dire que la suite $(\hat{\theta}_{n,x})_{n\in\mathbb{N}}$ converge \mathbb{P}_{θ^*} -presque sûrement vers le vrai paramètre $\theta^* \in \Theta$.

Normalité asymptotique du MLE

Notons $\mathcal{I}(\theta^{\star})$ la matrice d'information de Fisher du modèle.

Théorème 10 (Normalité asymptotique du score normalisé, [W., 2024a])

Sous les hypothèses de [Douc et al., 2004], et sous l'hypothèse que $\rho < 1/\sqrt{2}$ pour le taux de mélange ρ du processus HMT, dans le cas stationnaire nous avons :

$$|T_n|^{-1/2} \,
abla_{ heta} \ell_{n,x}(heta^\star) \stackrel{(d)}{\longrightarrow} \mathcal{N}(0,\mathcal{I}(heta^\star)) \quad sous \ \mathbb{P}_{ heta^\star}.$$

Normalité asymptotique du MLE

Notons $\mathcal{I}(\theta^{\star})$ la matrice d'information de Fisher du modèle.

Théorème 10 (Normalité asymptotique du score normalisé, [W., 2024a])

Sous les hypothèses de [Douc et al., 2004], et sous l'hypothèse que $\rho < 1/\sqrt{2}$ pour le taux de mélange ρ du processus HMT, dans le cas stationnaire nous avons :

$$|T_n|^{-1/2} \nabla_{\theta} \ell_{n,x}(\theta^\star) \xrightarrow[n \to \infty]{(d)} \mathcal{N}(0, \mathcal{I}(\theta^\star)) \quad \text{sous } \mathbb{P}_{\theta^\star}.$$

Théorème 11 (Normalité asymptotique du MLE, [W., 2024a])

Sous les hypothèses de [Douc et al., 2004], et sous l'hypothèse que $\rho < 1/2$ pour le taux de mélange ρ du processus HMT, nous avons la convergence en distribution suivante :

$$|T_n|^{1/2}(\hat{\theta}_n - \theta^{\star}) \xrightarrow[n \to \infty]{(d)} \mathcal{N}(0, \mathcal{I}(\theta^{\star})^{-1}) \quad sous \mathbb{P}_{\theta^{\star}}.$$

- Peut-on avoir ces résultats pour d'autres valeurs de ρ ?
 - $\rho \in (0,1)$ pour le TCL avec des incréments martingales [Guyon, 2007]
 - $ho < 1/\sqrt{2}$ pour le TCL en général [Bitseki Penda et Delmas, 2022]

- Peut-on avoir ces résultats pour d'autres valeurs de ρ ?
 - $\rho \in (0,1)$ pour le TCL avec des incréments martingales [Guyon, 2007]
 - $ho < 1/\sqrt{2}$ pour le TCL en général [Bitseki Penda et Delmas, 2022]
- HMT avec noyau de transition asymétrique, agrégats et division cellulaire

- Peut-on avoir ces résultats pour d'autres valeurs de ρ ?
 - $\rho \in (0,1)$ pour le TCL avec des incréments martingales [Guyon, 2007]
 - $ho < 1/\sqrt{2}$ pour le TCL en général [Bitseki Penda et Delmas, 2022]
- HMT avec noyau de transition asymétrique, agrégats et division cellulaire
- HMT pour des arbres aléatoires, HMT pour des lois de reproduction dépendant du processus

- Peut-on avoir ces résultats pour d'autres valeurs de ρ ?
 - $\rho \in (0,1)$ pour le TCL avec des incréments martingales [Guyon, 2007]
 - $ho < 1/\sqrt{2}$ pour le TCL en général [Bitseki Penda et Delmas, 2022]
- HMT avec noyau de transition asymétrique, agrégats et division cellulaire
- HMT pour des arbres aléatoires, HMT pour des lois de reproduction dépendant du processus
- Limites locale et d'échelle de l'arbre couvrant minimal (MST) d'une suite de graphes pondérés aléatoires.

- Peut-on avoir ces résultats pour d'autres valeurs de ρ ?
 - $\rho \in (0,1)$ pour le TCL avec des incréments martingales [Guyon, 2007]
 - $ho < 1/\sqrt{2}$ pour le TCL en général [Bitseki Penda et Delmas, 2022]
- HMT avec noyau de transition asymétrique, agrégats et division cellulaire
- HMT pour des arbres aléatoires, HMT pour des lois de reproduction dépendant du processus
- Limites locale et d'échelle de l'arbre couvrant minimal (MST) d'une suite de graphes pondérés aléatoires.
- Détection de communauté pour le modèle à blocs stochastiques pondérés ou décorés (en anglais, weighted SBMs et labeled SBMs).

Merci de votre attention !

Abraham, R., Delmas, J.-F., et Weibel, J. (2023).

Probability-graphons: Limits of large dense weighted graphs.

Weibel, J. (2024b).

Ergodic theorem for branching Markov chains indexed by trees with arbitrary shape.



Weibel, J. (2024a).

Asymptotic properties of the maximum likelihood estimator for Hidden Markov Models indexed by binary trees.

Archer, E., Nachmias, A., et Shalev, M. (2024). The GHP Scaling Limit of Uniform Spanning Trees in High Dimensions. *Communications in Mathematical Physics*, 405(3):73.

Ayi, N. et Duteil, N. P. (2023).

Graph Limit for Interacting Particle Systems on Weighted Random Graphs.



Bitseki Penda, S. V. et Delmas, J.-F. (2022).

Central limit theorem for bifurcating Markov chains under pointwise ergodic conditions.

The Annals of Applied Probability, 32(5).

Douc, R., Moulines, É., et Rydén, T. (2004).

Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime.

The Annals of Statistics, 32(5).



Dufour, C. et Olhede, S. C. (2024).

Inference for decorated graphs and application to multiplex networks.



Guyon, J. (2007).

Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging.

The Annals of Applied Probability, 17(5-6):1538-1569.

Hladký, J., Nachmias, A., et Tran, T. (2018). The local limit of the uniform spanning tree on dense graphs. *Journal of Statistical Physics*, 173(3-4):502–545.



Hladký, J. et Viswanathan, G. (2023).

Random minimum spanning tree and dense graph limits.



Lelarge, M., Massoulie, L., et Xu, J. (2015). Reconstruction in the Labelled Stochastic Block Model. *IEEE Transactions on Network Science and Engineering*, 2(4):152–163.



Lovász, L. (2012).

Large networks and graph limits, volume 60.

American Mathematical Society, Colloquium Publications.



Lovász, L. et Szegedy, B. (2007).

Szemerédi's Lemma for the Analyst.

GAFA Geometric And Functional Analysis, 17(1):252–270.