Chaînes de Markov, Génération Aléatoire, et Couplage Depuis le Passé (3)

Petite École de Combinatoire

P. Duchon

LaBRI

Résumé (1): Récurrence et positivité

Une chaîne de Markov (homogène) a potentiellement trois types d'états (de composantes fortement connexes) :

- **transients** on n'y revient pas toujours;

Résumé (1): Récurrence et positivité

Une chaîne de Markov (homogène) a potentiellement trois types d'états (de composantes fortement connexes) :

- **transients** on n'y revient pas toujours;
- **récurrents nuls** on y revient toujours, mais au bout d'un temps moyen infini;

Résumé (1): Récurrence et positivité

Une chaîne de Markov (homogène) a potentiellement trois types d'états (de composantes fortement connexes) :

- **transients** on n'y revient pas toujours;
- **récurrents nuls** on y revient toujours, mais au bout d'un temps moyen infini;
- **récurrents positifs** on y revient infiniment souvent, à intervalles de temps finis en moyenne.

Résumé (1): Récurrence et positivité

Une chaîne de Markov (homogène) a potentiellement trois types d'états (de composantes fortement connexes) :

- **transients** on n'y revient pas toujours;
- **récurrents nuls** on y revient toujours, mais au bout d'un temps moyen infini;
- **récurrents positifs** on y revient infiniment souvent, à intervalles de temps finis en moyenne.

Si la chaîne est à espace d'états fini, elle ne peut pas avoir d'états récurrents nuls.

Résumé (2): Stationnarité et convergence

- sur chaque CFC positive, il existe une (unique) distribution de probabilités laissée invariante par l'évolution;

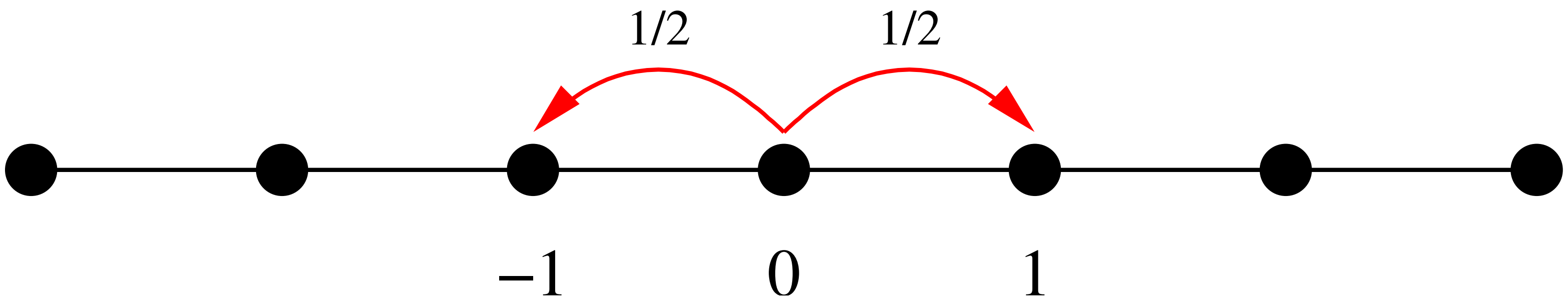
Résumé (2): Stationnarité et convergence

- sur chaque CFC positive, il existe une (unique) distribution de probabilités laissée invariante par l'évolution;
- quel que soit le point de départ dans cette CFC, les probabilités d'être dans chaque état convergent vers la loi stationnaire au sens de Cesaro;

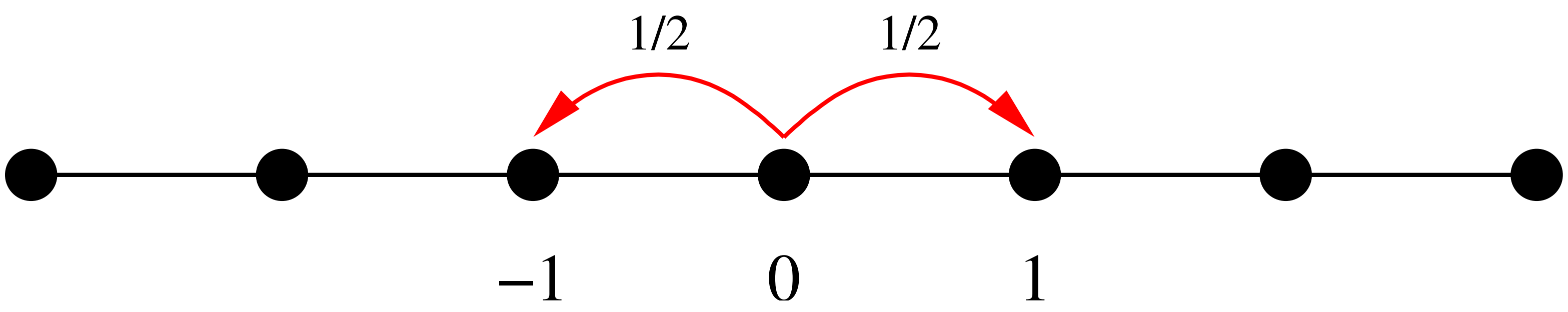
Résumé (2): Stationnarité et convergence

- sur chaque CFC positive, il existe une (unique) distribution de probabilités laissée invariante par l'évolution;
- quel que soit le point de départ dans cette CFC, les probabilités d'être dans chaque état convergent vers la loi stationnaire au sens de Cesaro;
- il y a convergence simple si la CFC est de période 1.

Exemple: marche aléatoire sur $\mathbb{Z}$

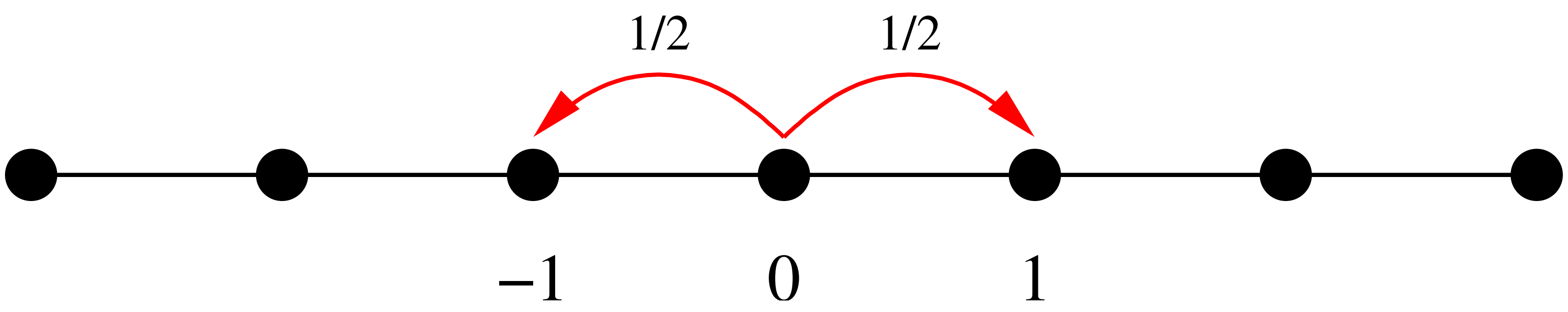


Exemple: marche aléatoire sur $\mathbb{Z}$



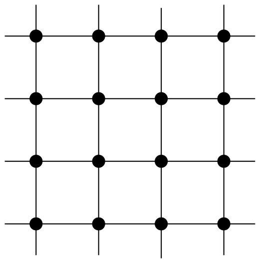
(On a déjà vérifié que) cette chaîne est récurrente nulle.

Exemple: marche aléatoire sur $\mathbb{Z}$

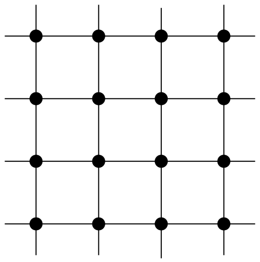


(On a déjà vérifié que) cette chaîne est récurrente nulle.
On peut se convaincre facilement qu'elle ne peut pas être positive : sinon, il y aurait existence et unicité de loi stationnaire, mais une telle loi devrait être invariante par translation sur $\mathbb{Z}$, ce qui est impossible…

Exemple: marche aléatoire sur $\mathbb{Z}^2$

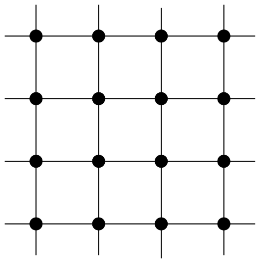


Exemple: marche aléatoire sur $\mathbb{Z}^2$



Comme pour $\mathbb{Z}$, elle ne peut être récurrente positive; est-elle transiente, ou récurrente nulle?

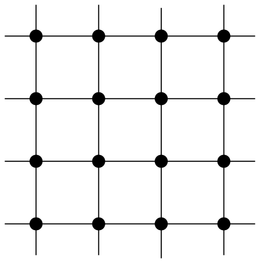
Exemple: marche aléatoire sur $\mathbb{Z}^2$



La probabilité, partant de 0 au temps 0, d'y être de nouveau au temps $2n$, est

$$p_{0,0}(2n) = \sum_{k=0}^{n} \frac{(2n)!}{k!^2(n-k)!^2} \left(\frac{1}{4}\right)^{2n}$$

Exemple: marche aléatoire sur $\mathbb{Z}^2$



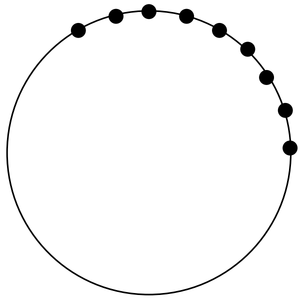
La probabilité, partant de 0 au temps 0, d'y être de nouveau au temps $2n$, est

$$p_{0,0}(2n) = \sum_{k=0}^{n} \frac{(2n)!}{k!^2 (n-k)!^2} \left(\frac{1}{4}\right)^{2n}$$

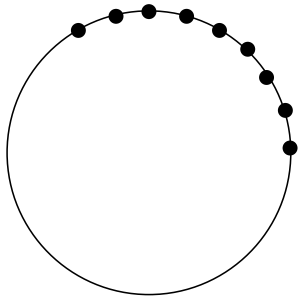
Donc on a

$$P_{0,0}(1) = \sum_{k,\ell \geq 0} \frac{(2k+2\ell)!}{k!^2 \ell!^2} \left(\frac{1}{4}\right)^{2(k+\ell)}$$

Exemple: Marche aléatoire sur un n-cycle

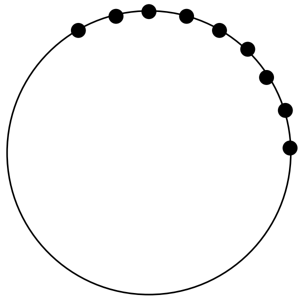


Exemple: Marche aléatoire sur un n-cycle



Chaîne finie, graphe non orienté : une seule distribution stationnaire, qui ne peut être qu'uniforme par symétries.

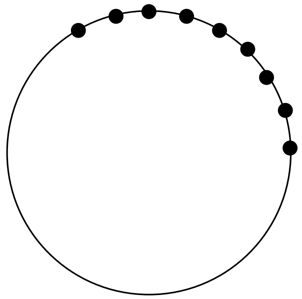
Exemple: Marche aléatoire sur un n-cycle



Chaîne finie, graphe non orienté : une seule distribution stationnaire, qui ne peut être qu'uniforme par symétries.

- Si n est pair : la chaîne est périodique de période 2, et il ne peut pas y avoir de convergence vers l'uniforme pour la chaîne partant d'un sommet donné.

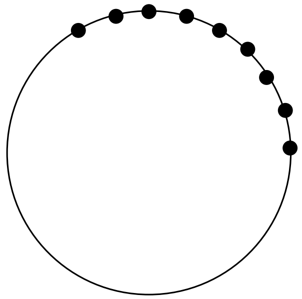
Exemple: Marche aléatoire sur un n-cycle



Chaîne finie, graphe non orienté : une seule distribution stationnaire, qui ne peut être qu'uniforme par symétries.

- Si n est pair : la chaîne est périodique de période 2, et il ne peut pas y avoir de convergence vers l'uniforme pour la chaîne partant d'un sommet donné.
- Si n est impair : la chaîne est apériodique, il y a bien convergence vers l'uniforme;

Exemple: Marche aléatoire sur un n-cycle



Chaîne finie, graphe non orienté : une seule distribution stationnaire, qui ne peut être qu'uniforme par symétries.

- Si n est pair : la chaîne est périodique de période 2, et il ne peut pas y avoir de convergence vers l'uniforme pour la chaîne partant d'un sommet donné.
- Si n est impair : la chaîne est apériodique, il y a bien convergence vers l'uniforme; on se doute que, si n est grand, il va falloir "longtemps" pour que la chaîne "oublie" la parité de son sommet de départ.

Apériodicité

Si on a une chaîne $\mathcal{M}$, de probabilités de transition p_{uv}, on peut définir une nouvelle chaîne $\mathcal{M}_\epsilon$ $(0 < \epsilon < 1)$ par ses probabilités de transition :

$$p'_{uv} = \epsilon\delta_{uv} + (1-\epsilon)p_{uv}$$

Apériodicité

Si on a une chaîne $\mathcal{M}$, de probabilités de transition p_{uv}, on peut définir une nouvelle chaîne $\mathcal{M}_\epsilon$ $(0 < \epsilon < 1)$ par ses probabilités de transition :

$$p'_{uv} = \epsilon\delta_{uv} + (1-\epsilon)p_{uv}$$

(à chaque étape, la chaîne $\mathcal{M}_\epsilon$ tire au sort : avec probabilité ϵ, elle ne fait rien, sinon elle fait un mouvement de la chaîne $\mathcal{M}$)

Apériodicité

Si on a une chaîne $\mathcal{M}$, de probabilités de transition p_{uv}, on peut définir une nouvelle chaîne $\mathcal{M}_\epsilon$ $(0 < \epsilon < 1)$ par ses probabilités de transition :

$$p'_{uv} = \epsilon\delta_{uv} + (1-\epsilon)p_{uv}$$

(à chaque étape, la chaîne $\mathcal{M}_\epsilon$ tire au sort : avec probabilité ϵ, elle ne fait rien, sinon elle fait un mouvement de la chaîne $\mathcal{M}$)

- La chaîne $\mathcal{M}_\epsilon$ a les mêmes lois stationnaires que $\mathcal{M}$;

Apériodicité

Si on a une chaîne $\mathcal{M}$, de probabilités de transition p_{uv}, on peut définir une nouvelle chaîne $\mathcal{M}_\epsilon$ $(0 < \epsilon < 1)$ par ses probabilités de transition :

$$p'_{uv} = \epsilon\delta_{uv} + (1-\epsilon)p_{uv}$$

(à chaque étape, la chaîne $\mathcal{M}_\epsilon$ tire au sort : avec probabilité ϵ, elle ne fait rien, sinon elle fait un mouvement de la chaîne $\mathcal{M}$)

- La chaîne $\mathcal{M}_\epsilon$ a les mêmes lois stationnaires que $\mathcal{M}$;
- elle est automatiquement apériodique;

Apériodicité

Si on a une chaîne $\mathcal{M}$, de probabilités de transition p_{uv}, on peut définir une nouvelle chaîne $\mathcal{M}_\epsilon$ $(0 < \epsilon < 1)$ par ses probabilités de transition :

$$p'_{uv} = \epsilon\delta_{uv} + (1-\epsilon)p_{uv}$$

(à chaque étape, la chaîne $\mathcal{M}_\epsilon$ tire au sort : avec probabilité ϵ, elle ne fait rien, sinon elle fait un mouvement de la chaîne $\mathcal{M}$)

- La chaîne $\mathcal{M}_\epsilon$ a les mêmes lois stationnaires que $\mathcal{M}$;
- elle est automatiquement apériodique;
- la convergence pour $\mathcal{M}_\epsilon$ (Cesaro ou simple) est grosso modo $1/(1-\epsilon)$ fois plus lente pour $\mathcal{M}_\epsilon$ que pour $\mathcal{M}$.

Recherche des lois limite

Problème : étant donnée une chaîne de Markov (qu'on peut supposer apériodique, avec une loi stationnaire unique; par exemple, irréductible positive), identifier sa loi stationnaire.

Recherche des lois limite

Problème : étant donnée une chaîne de Markov (qu'on peut supposer apériodique, avec une loi stationnaire unique; par exemple, irréductible positive), identifier sa loi stationnaire.
C'est naturellement l'unique vecteur propre gauche (à coefficients positifs, normalisé) de la matrice des probabilités de transition, pour la valeur propre 1.

Recherche des lois limite

Problème : étant donnée une chaîne de Markov (qu'on peut supposer apériodique, avec une loi stationnaire unique; par exemple, irréductible positive), identifier sa loi stationnaire.
C'est naturellement l'unique vecteur propre gauche (à coefficients positifs, normalisé) de la matrice des probabilités de transition, pour la valeur propre 1.
Condition d'équilibre : pour tout v,

$$\sum_u \pi(u)p_{uv} = \pi(v)$$

Recherche des lois limite

Problème : étant donnée une chaîne de Markov (qu'on peut supposer apériodique, avec une loi stationnaire unique; par exemple, irréductible positive), identifier sa loi stationnaire.
C'est naturellement l'unique vecteur propre gauche (à coefficients positifs, normalisé) de la matrice des probabilités de transition, pour la valeur propre 1.
Condition d'équilibre : pour tout v,

$$\sum_u \pi(u) p_{uv} = \pi(v)$$

En particulier, si (et seulement si) le graphe (non orienté) est régulier, la marche aléatoire non biaisée laisse invariante la distribution uniforme. Dans le cas d'un graphe orienté, la condition devient que les degrés entrants soient uniformes.

Lois limites : le cas inversible

Proposition : si π est une distribution de probabilités qui vérifie, pour tout couple de sommets (u, v), la condition d'équilibre local (*detailed balance condition*) :

$$\pi(u)p_{u,v} = \pi(v)p_{v,u},$$

alors π est une distribution stationnaire.

Lois limites : le cas inversible

Proposition : si π est une distribution de probabilités qui vérifie, pour tout couple de sommets (u, v), la condition d'équilibre local (*detailed balance condition*) :

$$\pi(u)p_{u,v} = \pi(v)p_{v,u},$$

alors π est une distribution stationnaire.

Preuve : en sommant sur u, le membre droit devient $\pi(v)$.$\square$

Lois limites : le cas inversible

Proposition : si π est une distribution de probabilités qui vérifie, pour tout couple de sommets (u, v), la condition d'équilibre local (*detailed balance condition*) :

$$\pi(u)p_{u,v} = \pi(v)p_{v,u},$$

alors π est une distribution stationnaire.

Preuve : en sommant sur u, le membre droit devient $\pi(v)$.$\square$

(chaîne inversible - *time-reversible*)

Lois limites : le cas inversible

Proposition : si π est une distribution de probabilités qui vérifie, pour tout couple de sommets (u, v), la condition d'équilibre local (*detailed balance condition*) :

$$\pi(u)p_{u,v} = \pi(v)p_{v,u},$$

alors π est une distribution stationnaire.

Preuve : en sommant sur u, le membre droit devient $\pi(v)$.$\square$

(chaîne inversible - *time-reversible*)

Conséquence : si la matrice des probabilités de transition est symétrique, la distribution uniforme est stationnaire.

Choix d'une chaîne

Étant donné un ensemble fini Ω, créer une chaîne de Markov d'ensemble d'états Ω dont la distribution converge vers la distribution uniforme, c'est "facile" :

Choix d'une chaîne

Étant donné un ensemble fini Ω, créer une chaîne de Markov d'ensemble d'états Ω dont la distribution converge vers la distribution uniforme, c'est “facile” :

- définir un graphe non orienté, connexe, dont les sommets sont les éléments de Ω;

Choix d'une chaîne

Étant donné un ensemble fini Ω, créer une chaîne de Markov d'ensemble d'états Ω dont la distribution converge vers la distribution uniforme, c'est "facile" :

- définir un graphe non orienté, connexe, dont les sommets sont les éléments de Ω;
- choisir arbitrairement des probabilités de transition symétriques $p_{u,v} = p_{v,u}$, strictement positives, le long des arêtes;

Choix d'une chaîne

Étant donné un ensemble fini Ω, créer une chaîne de Markov d'ensemble d'états Ω dont la distribution converge vers la distribution uniforme, c'est “facile” :

- définir un graphe non orienté, connexe, dont les sommets sont les éléments de Ω;
- choisir arbitrairement des probabilités de transition symétriques $p_{u,v} = p_{v,u}$, strictement positives, le long des arêtes;
- compléter avec des boucles : $p_{u,u} = 1 - \sum_{v \neq u} p_{u,v}$;

Choix d'une chaîne

Étant donné un ensemble fini Ω, créer une chaîne de Markov d'ensemble d'états Ω dont la distribution converge vers la distribution uniforme, c'est "facile" :

- définir un graphe non orienté, connexe, dont les sommets sont les éléments de Ω;
- choisir arbitrairement des probabilités de transition symétriques $p_{u,v} = p_{v,u}$, strictement positives, le long des arêtes;
- compléter avec des boucles : $p_{u,u} = 1 - \sum_{v \neq u} p_{u,v}$;
- pour peu qu'au moins un des $p_{u,u}$ soit strictement positif, c'est gagné, sinon il faut vérifier l'apériodicité.

Distance variationnelle

Pour évaluer l'écart entre une ditribution et une autre (par exemple, entre la distribution au temps n et la distribution limite), on utilise la distance variationnelle :

$$D(\pi, \pi') = \max_{A \subset \Omega} |\pi(A) - \pi'(A)| = \frac{1}{2} \sum_{u \in \Omega} |\pi(u) - \pi'(u)|$$

Clairement, on a aussi, en posant
$A = \{u \in \Omega : \pi(u) > \pi'(u)\}$,

$$D(\pi, \pi') = \pi(A) - \pi'(A) = \sum_{u \in A} \pi(u) - \pi'(u).$$

(C'est bien une distance sur l'ensemble des distributions de probabilités sur Ω)

Convergence et temps de mélange

On considère une chaîne de Markov ergodique, de distribution stationnaire π. Pour $u \in \Omega$ et $t \in \mathbb{N}$, on pose

$$\delta_u(t) = D(p_{u,\cdot}^{(t)}, \pi).$$

$\delta_u(t)$ mesure combien la chaîne, partant de u, est encore loin de sa distribution limite au temps t.

Convergence et temps de mélange

On considère une chaîne de Markov ergodique, de distribution stationnaire π. Pour $u \in \Omega$ et $t \in \mathbb{N}$, on pose

$$\delta_u(t) = D(p^{(t)}_{u,\cdot}, \pi).$$

On définit alors le temps de mélange :

$$\tau_u(\epsilon) = \min\{t : \delta_u(t') \leq \epsilon \text{ pour tout } t' \geq t\}$$

Convergence et temps de mélange

On considère une chaîne de Markov ergodique, de distribution stationnaire π. Pour $u \in \Omega$ et $t \in \mathbb{N}$, on pose

$$\delta_u(t) = D(p^{(t)}_{u,\cdot}, \pi).$$

On définit alors le temps de mélange :

$$\tau_u(\epsilon) = \min\{t : \delta_u(t') \leq \epsilon \text{ pour tout } t' \geq t\}$$

(à partir de $\tau_u(\epsilon)$, la chaîne partie de u ne sera plus jamais à distance plus grande que ϵ de π)

Convergence et temps de mélange

On considère une chaîne de Markov ergodique, de distribution stationnaire π. Pour $u \in \Omega$ et $t \in \mathbb{N}$, on pose

$$\delta_u(t) = D(p^{(t)}_{u,\cdot}, \pi).$$

On définit alors le temps de mélange :

$$\tau_u(\epsilon) = \min\{t : \delta_u(t') \leq \epsilon \text{ pour tout } t' \geq t\}$$

Pour uniformiser les choses, on pose

$$\tau(\epsilon) = \max_{u \in \Omega} \tau_u(\epsilon)$$

Convergence et temps de mélange

On considère une chaîne de Markov ergodique, de distribution stationnaire π. Pour $u \in \Omega$ et $t \in \mathbb{N}$, on pose

$$\delta_u(t) = D(p_{u,\cdot}^{(t)}, \pi).$$

On définit alors le temps de mélange :

$$\tau_u(\epsilon) = \min\{t : \delta_u(t') \leq \epsilon \text{ pour tout } t' \geq t\}$$

Pour uniformiser les choses, on pose

$$\tau(\epsilon) = \max_{u \in \Omega} \tau_u(\epsilon)$$

(quelle que soit sa distribution de départ, à partir du temps $\tau(\epsilon)$ la chaîne ne sera plus jamais à distance $> \epsilon$ de π)

Mélange rapide

Typiquement, $\Omega = \Omega_n$ (on définit une *famille*, indexée par des *tailles*, de chaînes sur des ensembles de plus en plus grands). On dira que cette famille de chaînes de Markov est rapidement mélangeante, si $\tau^{(n)}(\epsilon)$ ne croît pas trop vite lorsque n tend vers l'infini et ϵ vers 0 :

$$\tau^{(n)}(\epsilon) \leq \text{poly}(n, \log(1/\epsilon)).$$

Mélange rapide

Typiquement, $\Omega = \Omega_n$ (on définit une *famille*, indexée par des *tailles*, de chaînes sur des ensembles de plus en plus grands). On dira que cette famille de chaînes de Markov est rapidement mélangeante, si $\tau^{(n)}(\epsilon)$ ne croît pas trop vite lorsque n tend vers l'infini et ϵ vers 0 :

$$\tau^{(n)}(\epsilon) \leq \text{poly}(n, \log(1/\epsilon)).$$

Deux questions naturelles :

Mélange rapide

Typiquement, $\Omega = \Omega_n$ (on définit une *famille*, indexée par des *tailles*, de chaînes sur des ensembles de plus en plus grands). On dira que cette famille de chaînes de Markov est rapidement mélangeante, si $\tau^{(n)}(\epsilon)$ ne croît pas trop vite lorsque n tend vers l'infini et ϵ vers 0 :

$$\tau^{(n)}(\epsilon) \leq \text{poly}(n, \log(1/\epsilon)).$$

Deux questions naturelles :

- comment prouve-t-on qu'une chaîne est, ou non, rapidement mélangeante ?

Mélange rapide

Typiquement, $\Omega = \Omega_n$ (on définit une *famille*, indexée par des *tailles*, de chaînes sur des ensembles de plus en plus grands). On dira que cette famille de chaînes de Markov est rapidement mélangeante, si $\tau^{(n)}(\epsilon)$ ne croît pas trop vite lorsque n tend vers l'infini et ϵ vers 0 :

$$\tau^{(n)}(\epsilon) \leq \text{poly}(n, \log(1/\epsilon)).$$

Deux questions naturelles :

- comment prouve-t-on qu'une chaîne est, ou non, rapidement mélangeante ?
- comment construire des chaînes les plus rapidement mélangeantes possible ?