Chaînes de Markov, Génération Aléatoire, et Couplage Depuis le Passé (7)

Petite École de Combinatoire

P. Duchon

LaBRI

Couplages

- (Encore une) technique pour majorer le temps de mélange
- “Si on arrive à faire tourner deux copies de la chaîne de telle sorte qu’elles aient tendance à se retrouver dans le même état, alors la chaîne mélange vite”
- Technique avancée : couplage de chemins (*path coupling*)
- Couplage Venu du Passé (*Coupling From the Past, CFTP*) : une technique pour tirer selon la distribution limite *exacte*.

Distance variationnelle

$$D(\pi, \pi') = \frac{1}{2} \sum_{x \in \Omega} |\pi(x) - \pi'(x)|$$

Distance variationnelle

$$D(\pi, \pi') = \frac{1}{2} \sum_{x \in \Omega} |\pi(x) - \pi'(x)|$$

Autre caractérisation :

Distance variationnelle

$$D(\pi, \pi') = \frac{1}{2} \sum_{x \in \Omega} |\pi(x) - \pi'(x)|$$

Autre caractérisation :

- si X et X' sont deux variables aléatoires, définies sur le même espace, et de lois respectives π et π', alors on a

$$\mathbb{P}(X \neq X') \geq D(\pi, \pi')$$

Distance variationnelle

$$D(\pi, \pi') = \frac{1}{2} \sum_{x \in \Omega} |\pi(x) - \pi'(x)|$$

Autre caractérisation :

- si X et X' sont deux variables aléatoires, définies sur le même espace, et de lois respectives π et π', alors on a

$$\mathbb{P}(X \neq X') \geq D(\pi, \pi')$$

- on peut avoir l'égalité

	0.2	0.2	0.2	0.2	0.2
0.1					
0.15					
0.18					
0.22					
0.35					

	0.1	0.05	0.02	0	0
0	0.1				
0		0.15			
0			0.18		
0.02				0.2	
0.15					0.2

$$
\begin{array}{c|c|c|c|c|c|}
 & \color{blue}{0.08} & 0.05 & 0.02 & 0 & 0 \\
\hline
0 & 0.1 & & & & \\
\hline
0 & & 0.15 & & & \\
\hline
0 & & & 0.18 & & \\
\hline
\color{blue}{0} & \color{red}{0.02} & & & 0.2 & \\
\hline
0.15 & & & & & 0.2 \\
\hline
\end{array}
$$

$$\begin{array}{c|ccccc|}
 & 0.08 & {\color{blue}0} & 0.02 & 0 & 0 \\
\hline
0 & 0.1 & & & & \\
0 & & 0.15 & & & \\
0 & & & 0.18 & & \\
0 & 0.02 & & & 0.2 & \\
{\color{blue}0.1} & & {\color{red}0.05} & & & 0.2 \\
\hline
\end{array}$$

	0.08	0	0	0	0
0	0.1				
0		0.15			
0			0.18		
0	0.02			0.2	
0.08		0.05	0.02		0.2

	0	0	0	0	0
0	0.1				
0		0.15			
0			0.18		
0	0.02			0.2	
0	0.08	0.05	0.02		0.2

Idée du couplage de chaînes

On définit un couplage de la chaîne avec elle-même : un processus $(X_t, Y_t)_{t \geq 0}$, tel que

- $(X_t)_{t \geq 0}$ est une copie de la chaîne, partant de l'état x_0;
- $(Y_t)_{t \geq 0}$ est une copie de la chaîne, partant de la distribution limite.

Automatiquement, on a $\delta_{x_0}(t) = D(p_{x_0}^{(t)}, \pi) \leq \mathbb{P}(X_t \neq Y_t)$.
On ajoute comme contrainte (sur le couplage) : si $X_t = Y_t$, alors $X_{t+1} = Y_{t+1}$.
Alors, $(\mathbb{P}(X_t \neq Y_t))_{t \geq 0}$ décroît, et on a automatiquement

$$\tau_{x_0}(\epsilon) \leq \min\{t : \mathbb{P}(X_t \neq Y_t) \leq \epsilon\}$$

(De plus, dans quasiment tous les exemples, le couplage lui-même est markovien)

Définition du couplage

Pour définir le couplage (markovien), on définit une procédure qui

- examine les deux états (x, y);
- [simule X] choisit le nouvel état x', selon la loi $p(x, .)$;
- [simule Y] si $y = x$, choisit $y' = x'$;
- sinon, choisit y' selon une loi $p'(x, x', y, .)$ qui doit vérifier :

$$\forall x, \forall y, \forall y', \sum_{x'} p(x, x') p'(x, x', y, y') = p(y, y')$$

On n'est évidemment pas obligé de regarder la valeur de x' pour choisir y', mais on en a le droit.

Cas particulier : fonction de mise à jour

Une façon bien pratique de visualiser le couplage : on simule la chaîne de base par une procédure de mise à jour :

- Tirer U aléatoire (selon une loi connue, e.g. uniforme sur $[0, 1]$)
- Si l'état courant est x, le nouvel état est $x' = f(U, x)$

Cas particulier : fonction de mise à jour

Une façon bien pratique de visualiser le couplage : on simule la chaîne de base par une procédure de mise à jour :

- Tirer U aléatoire (selon une loi connue, e.g. uniforme sur $[0, 1]$)
- Si l'état courant est x, le nouvel état est $x' = f(U, x)$

La fonction f doit vérifier : pour tous x et x',

$$\mu(\{u : f(u, x) = x') = p(x, x')$$

(définir f revient à décrire un moyen de simulation de la chaîne)

Cas particulier : fonction de mise à jour

Une façon bien pratique de visualiser le couplage : on simule la chaîne de base par une procédure de mise à jour :

- Tirer U aléatoire (selon une loi connue, e.g. uniforme sur $[0, 1]$)
- Si l'état courant est x, le nouvel état est $x' = f(U, x)$

Alors le couplage peut se définir par :

- Tirer U aléatoire;
- $x' = f(U, x)$;
- $y' = f(U, y)$.

Cas particulier : fonction de mise à jour

Une façon bien pratique de visualiser le couplage : on simule la chaîne de base par une procédure de mise à jour :

- Tirer U aléatoire (selon une loi connue, e.g. uniforme sur $[0,1]$)
- Si l'état courant est x, le nouvel état est $x' = f(U, x)$

Alors le couplage peut se définir par :

- Tirer U aléatoire;
- $x' = f(U, x)$;
- $y' = f(U, y)$.

(les deux chaînes sont simulées en utilisant les mêmes jets de dés)

Cas particulier : fonction de mise à jour

Une façon bien pratique de visualiser le couplage : on simule la chaîne de base par une procédure de mise à jour :

- Tirer U aléatoire (selon une loi connue, e.g. uniforme sur $[0, 1]$)
- Si l'état courant est x, le nouvel état est $x' = f(U, x)$

Alors le couplage peut se définir par :

- Tirer U aléatoire;
- $x' = f(U, x)$;
- $y' = f(U, y)$.

La fonction $f(U, .)$ représente une fonction (aléatoire, car U est aléatoire) de mise à jour de l'état.

Vitesse de convergence

Exemple des colorations d'un graphe de degré maximum Δ, avec $C \geq 2\Delta + 1$ couleurs

Vitesse de convergence

Exemple des colorations d'un graphe de degré maximum Δ, avec $C \geq 2\Delta + 1$ couleurs
La chaîne simple : X la coloration courante

- Choisir un sommet u au hasard;
- le recolorier avec une couleur aléatoire uniforme parmi celles qui n'apparaissent pas parmi les voisins de u dans X.

Vitesse de convergence

Exemple des colorations d'un graphe de degré maximum Δ, avec $C \geq 2\Delta + 1$ couleurs
Le couplage : X, Y les deux colorations courantes

- Choisir un sommet u au hasard;
- Soit A l'ensemble des couleurs qui n'apparaissent pas parmi les voisins de u dans X, et B l'ensemble des couleurs qui n'apparaissent pas parmi les voisins de u dans Y;
- Choisir un couple $(a, b) \in A \times B$, aléatoire, avec a uniforme sur A et b uniforme sur B, et $\mathbb{P}(a = b) = |A \cap B| / \max(|A|, |B|)$
- Dans X, recolorer u en a; dans Y, recolorer u en b.

Soit D_t l'ensemble des sommets pour lesquels X_t et Y_t diffèrent : on veut majorer $\mathbb{P}(D_t \neq \emptyset)$.

Soit D_t l'ensemble des sommets pour lesquels X_t et Y_t diffèrent : on veut majorer $\mathbb{P}(D_t \neq \emptyset)$. Pour cela, on majore $\mathbb{E}(|D_t|)$.

Soit D_t l'ensemble des sommets pour lesquels X_t et Y_t diffèrent : on veut majorer $\mathbb{P}(D_t \neq \emptyset)$. Pour cela, on majore $\mathbb{E}(|D_t|)$. Soit m le nombre d'arêtes du graphe qui séparent D_t de $\overline{D_t}$: $m = \sum_{v \in D_t} d(v) = \sum_{v \notin D_t} d(v)$, où $d(v)$ est le nombre de voisins de v qui ne sont pas dans D_t (si $v \in D_t$), ou qui sont dans D_t (si $v \notin D_t$).

Soit D_t l'ensemble des sommets pour lesquels X_t et Y_t diffèrent : on veut majorer $\mathbb{P}(D_t \neq \emptyset)$. Pour cela, on majore $\mathbb{E}(|D_t|)$. Soit m le nombre d'arêtes du graphe qui séparent D_t de $\overline{D_t}$: $m = \sum_{v \in D_t} d(v) = \sum_{v \notin D_t} d(v)$, où $d(v)$ est le nombre de voisins de v qui ne sont pas dans D_t (si $v \in D_t$), ou qui sont dans D_t (si $v \notin D_t$).
On aura $|D_{t+1}| = |D_t| + 1$ si

Soit D_t l'ensemble des sommets pour lesquels X_t et Y_t diffèrent : on veut majorer $\mathbb{P}(D_t \neq \emptyset)$. Pour cela, on majore $\mathbb{E}(|D_t|)$. Soit m le nombre d'arêtes du graphe qui séparent D_t de $\overline{D_t}$: $m = \sum_{v \in D_t} d(v) = \sum_{v \notin D_t} d(v)$, où $d(v)$ est le nombre de voisins de v qui ne sont pas dans D_t (si $v \in D_t$), ou qui sont dans D_t (si $v \notin D_t$).

On aura $|D_{t+1}| = |D_t| + 1$ si

- le sommet u est dans $\overline{D_t}$

Soit D_t l'ensemble des sommets pour lesquels X_t et Y_t diffèrent : on veut majorer $\mathbb{P}(D_t \neq \emptyset)$. Pour cela, on majore $\mathbb{E}(|D_t|)$. Soit m le nombre d'arêtes du graphe qui séparent D_t de $\overline{D_t}$: $m = \sum_{v \in D_t} d(v) = \sum_{v \notin D_t} d(v)$, où $d(v)$ est le nombre de voisins de v qui ne sont pas dans D_t (si $v \in D_t$), ou qui sont dans D_t (si $v \notin D_t$).
On aura $|D_{t+1}| = |D_t| + 1$ si

- le sommet u est dans $\overline{D_t}$
- les deux couleurs choisies a et b ne sont pas égales

Soit D_t l'ensemble des sommets pour lesquels X_t et Y_t diffèrent : on veut majorer $\mathbb{P}(D_t \neq \emptyset)$. Pour cela, on majore $\mathbb{E}(|D_t|)$. Soit m le nombre d'arêtes du graphe qui séparent D_t de $\overline{D_t}$: $m = \sum_{v \in D_t} d(v) = \sum_{v \notin D_t} d(v)$, où $d(v)$ est le nombre de voisins de v qui ne sont pas dans D_t (si $v \in D_t$), ou qui sont dans D_t (si $v \notin D_t$).
On aura $|D_{t+1}| = |D_t| + 1$ si

- le sommet u est dans $\overline{D_t}$
- les deux couleurs choisies a et b ne sont pas égales

et $|D_{t+1}| = |D_t| - 1$ si

Soit D_t l'ensemble des sommets pour lesquels X_t et Y_t diffèrent : on veut majorer $\mathbb{P}(D_t \neq \emptyset)$. Pour cela, on majore $\mathbb{E}(|D_t|)$. Soit m le nombre d'arêtes du graphe qui séparent D_t de $\overline{D_t}$: $m = \sum_{v \in D_t} d(v) = \sum_{v \notin D_t} d(v)$, où $d(v)$ est le nombre de voisins de v qui ne sont pas dans D_t (si $v \in D_t$), ou qui sont dans D_t (si $v \notin D_t$).
On aura $|D_{t+1}| = |D_t| + 1$ si

- le sommet u est dans $\overline{D_t}$
- les deux couleurs choisies a et b ne sont pas égales

et $|D_{t+1}| = |D_t| - 1$ si

- le sommet u est dans D_t

Soit D_t l'ensemble des sommets pour lesquels X_t et Y_t diffèrent : on veut majorer $\mathbb{P}(D_t \neq \emptyset)$. Pour cela, on majore $\mathbb{E}(|D_t|)$. Soit m le nombre d'arêtes du graphe qui séparent D_t de $\overline{D_t}$: $m = \sum_{v \in D_t} d(v) = \sum_{v \notin D_t} d(v)$, où $d(v)$ est le nombre de voisins de v qui ne sont pas dans D_t (si $v \in D_t$), ou qui sont dans D_t (si $v \notin D_t$).
On aura $|D_{t+1}| = |D_t| + 1$ si

- le sommet u est dans $\overline{D_t}$
- les deux couleurs choisies a et b ne sont pas égales

et $|D_{t+1}| = |D_t| - 1$ si

- le sommet u est dans D_t
- les couleurs a et b sont égales

Cas $|D_{t+1}| = |D_t| + 1$

Supposons u fixé : $\alpha = |A|$, $\beta = |B|$, $\gamma = |A \cap B|$.

Cas $|D_{t+1}| = |D_t| + 1$

Supposons u fixé : $\alpha = |A|$, $\beta = |B|$, $\gamma = |A \cap B|$.
On a, si $u \in \overline{D_t}$,

$$\begin{aligned} \alpha - \gamma &\leq d(u) \\ \beta - \gamma &\leq d(u) \\ \gamma &\geq C - \Delta - d(u) \end{aligned}$$

Cas $|D_{t+1}| = |D_t| + 1$

Supposons u fixé : $\alpha = |A|$, $\beta = |B|$, $\gamma = |A \cap B|$.
On a, si $u \in \overline{D_t}$,

$$\begin{aligned} \alpha - \gamma &\leq d(u) \\ \beta - \gamma &\leq d(u) \\ \gamma &\geq C - \Delta - d(u) \end{aligned}$$

donc $\mathbb{P}(a \neq b) \leq 1 - \frac{\gamma}{\gamma + d(u)} \leq \frac{d(u)}{C - \Delta}$

Cas $|D_{t+1}| = |D_t| + 1$

Supposons u fixé : $\alpha = |A|$, $\beta = |B|$, $\gamma = |A \cap B|$.
On a, si $u \in \overline{D_t}$,

$$\begin{aligned} \alpha - \gamma &\leq d(u) \\ \beta - \gamma &\leq d(u) \\ \gamma &\geq C - \Delta - d(u) \end{aligned}$$

donc $\mathbb{P}(a \neq b) \leq 1 - \frac{\gamma}{\gamma + d(u)} \leq \frac{d(u)}{C - \Delta}$, d'où

$$\mathbb{P}(|D_{t+1}| = |D_t| + 1) \leq \frac{1}{n} \sum_{u \in \overline{D_t}} \frac{d(u)}{C - \Delta} = \frac{m}{(C - \Delta)n}.$$

Cas $|D_{t+1}| = |D_t| - 1$

Supposons u fixé : $\alpha = |A|$, $\beta = |B|$, $\gamma = |A \cap B|$.

Cas $|D_{t+1}| = |D_t| - 1$

Supposons u fixé : $\alpha = |A|$, $\beta = |B|$, $\gamma = |A \cap B|$.
On a, si $u \in D_t$,

$$\begin{aligned} \alpha - \gamma &\leq \Delta - d(u) \\ \beta - \gamma &\leq \Delta - d(u) \\ \gamma &\geq C - 2\Delta + d(u) \end{aligned}$$

Cas $|D_{t+1}| = |D_t| - 1$

Supposons u fixé : $\alpha = |A|$, $\beta = |B|$, $\gamma = |A \cap B|$.
On a, si $u \in D_t$,

$$\begin{aligned} \alpha - \gamma &\leq \Delta - d(u) \\ \beta - \gamma &\leq \Delta - d(u) \\ \gamma &\geq C - 2\Delta + d(u) \end{aligned}$$

donc $\mathbb{P}(a = b) \geq \frac{\gamma}{\Delta + \gamma - d(u)} \geq \frac{C - 2\Delta + d(u)}{C - \Delta}$

Cas $|D_{t+1}| = |D_t| - 1$

Supposons u fixé : $\alpha = |A|$, $\beta = |B|$, $\gamma = |A \cap B|$.
On a, si $u \in D_t$,

$$\begin{array}{rcl} \alpha - \gamma & \leq & \Delta - d(u) \\ \beta - \gamma & \leq & \Delta - d(u) \\ \gamma & \geq & C - 2\Delta + d(u) \end{array}$$

donc $\mathbb{P}(a = b) \geq \frac{\gamma}{\Delta + \gamma - d(u)} \geq \frac{C - 2\Delta + d(u)}{C - \Delta}$, d'où

$$\begin{array}{rcl} \mathbb{P}(|D_{t+1}| & = & |D_t| - 1) \geq \dfrac{1}{n} \displaystyle\sum_{u \in D_t} \dfrac{C - 2\Delta}{C - \Delta} + \dfrac{d(u)}{C - \Delta} \\ & \leq & \dfrac{C - 2\Delta}{(C - \Delta)n} |D_t| + \dfrac{m}{(C - \Delta)n}. \end{array}$$

Majoration du temps de mélange

Au total, $\mathbb{E}(\#D_{t+1}|D_t) \leq \left(1 - \frac{C-2\Delta}{(C-\Delta)n}\right)|D_t|$, et donc

$$\mathbb{E}(|D_t|) \leq \left(1 - \frac{C - 2\Delta}{(C - \Delta)n}\right)^t |D_0| \leq n \exp\left(-t\frac{C - 2\Delta}{(C - \Delta)n}\right)$$

Majoration du temps de mélange

Au total, $\mathbb{E}(\#D_{t+1}|D_t) \leq \left(1 - \frac{C-2\Delta}{(C-\Delta)n}\right) |D_t|$, et donc

$$\mathbb{E}(|D_t|) \leq \left(1 - \frac{C - 2\Delta}{(C - \Delta)n}\right)^t |D_0| \leq n \exp\left(-t \frac{C - 2\Delta}{(C - \Delta)n}\right)$$

Par conséquent, $\mathbb{E}(|D_t|) \leq \epsilon$ dès lors que

$$t \geq \ln\left(\frac{n}{\epsilon}\right) \frac{(C - \Delta)n}{C - 2\Delta}.$$

Majoration du temps de mélange

Au total, $\mathbb{E}(\#D_{t+1}|D_t) \leq \left(1 - \frac{C-2\Delta}{(C-\Delta)n}\right)|D_t|$, et donc

$$\mathbb{E}(|D_t|) \leq \left(1 - \frac{C-2\Delta}{(C-\Delta)n}\right)^t |D_0| \leq n\exp\left(-t\frac{C-2\Delta}{(C-\Delta)n}\right)$$

Par conséquent, $\mathbb{E}(|D_t|) \leq \epsilon$ dès lors que

$$t \geq \ln\left(\frac{n}{\epsilon}\right)\frac{(C-\Delta)n}{C-2\Delta}.$$

Comme par ailleurs on a $\mathbb{P}(D_t \neq \emptyset) \leq \mathbb{E}(|D_t|)$, on a donc

$$\tau_\epsilon \leq \ln\left(\frac{n}{\epsilon}\right)\frac{(C-\Delta)n}{C-2\Delta}.$$

C'est typique des preuves de mélange par couplage : on identifie une distance (entière) entre états $D_t = f(X_t, Y_t)$, et on cherche à obtenir une inégalité de la forme : pour tous X_t, Y_t,

$$\mathbb{E}(D_{t+1}|X_t, Y_t) \leq aD_t$$

pour un $a < 1$, ce qui implique

$$\mathbb{P}(D_t \neq 0) \leq \mathbb{E}(D_t) \leq a^t \max_{x,y\in\Omega} f(x,y) \quad (C)$$

C'est typique des preuves de mélange par couplage : on identifie une distance (entière) entre états $D_t = f(X_t, Y_t)$, et on cherche à obtenir une inégalité de la forme : pour tous X_t, Y_t,

$$\mathbb{E}(D_{t+1} | X_t, Y_t) \leq a D_t$$

pour un $a < 1$, ce qui implique

$$\mathbb{P}(D_t \neq 0) \leq \mathbb{E}(D_t) \leq a^t \max_{x,y \in \Omega} f(x,y) \quad (C)$$

Problème : il se peut qu'on arrive à assurer (C) pour des paires d'états "proches", mais pas de manière globale. C'est là qu'intervient la technique du couplage de chemins : on construit le couplage localement de telle sorte qu'il vérifie (C), et le "couplage de chemins" montre qu'on peut étendre la construction globalement.