Chaînes de Markov, Génération Aléatoire, et Couplage Depuis le Passé (4)

Petite École de Combinatoire

P. Duchon

LaBRI

Résumé (1): Distance variationnelle

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

Version relative (non symétrique) :

$$D_r(\pi, \pi') = \max_{A \subset \Omega} \frac{|\pi(A) - \pi'(A)|}{\pi'(A)}$$

Résumé (2) : temps de mélange

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

Temps de mélange en partant de l'état u :

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

Pour uniformiser les choses, on pose

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

Résumé (3) : chaîne inversible

Une chaîne (ergodique) est inversible si la distribution stationnaire π satisfait la condition d'équilibre local : pour tous états u et v,

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

L'inversibilité revient à dire que, dans la distribution stationnaire, la chaîne $(X_n)_{0 \leq n \leq N}$ est (probabilistiquement) indistinguable de la chaîne à temps inversé $(X_{N-n})_{0 \leq n \leq N}$.

Inversibilité

Comment reconnaît-on une chaîne inversible ?

- Si π est uniforme : $p_{u,v} = p_{v,u}$ (matrice de transitions symétrique);
- Sinon, en posant $R_{u,v} = p_{u,v}/p_{v,u}$ pour toute arête, on doit avoir, pour tout chemin $(u_0, u_1, \dots, u_k)$,

$$R_{u_0,u_1} R_{u_1,u_2} \dots R_{u_{k-1},u_k} = \frac{\pi(u_k)}{\pi(u_0)};$$

  (ne dépend pas du chemin, seulement des extrêmités)
- On peut donc définir $R_{u,v}$ pour tous états u et v, par produit le long de n'importe quel chemin de u à v;
- C'est clairement une CNS (si Ω est infini, rajouter la convergence de la série $\sum_v R_{u,v}$).

Remarque

Si on connaît une chaîne inversible et sa distribution stationnaire π, on peut modifier les probabilités de transition sans changer π : il suffit de multiplier $p_{u,v}$ et $p_{v,u}$ par la même constante (en augmentant ou diminuant les probabilités de boucles, à condition que cela soit compatible).

Remarque

Si on connaît une chaîne inversible et sa distribution stationnaire π, on peut modifier les probabilités de transition sans changer π : il suffit de multiplier $p_{u,v}$ et $p_{v,u}$ par la même constante (en augmentant ou diminuant les probabilités de boucles, à condition que cela soit compatible).
(Il se trouve que le fait d'augmenter certains $Q(e)$, améliore certaines des majorations de la vitesse de convergence)

Proposition : si la chaîne est inversible et si sa matrice de probabilités de transitions a pour valeurs propres $1 = \lambda_0 > \lambda_1 \geq \ldots \geq \lambda_{N-1}$, en posant $\lambda = \max(|\lambda_1|, |\lambda_{N-1}|)$ on a, indépendamment de l'état de départ,

$$|\mathbb{P}(X_t = v) - \pi(v)| \leq \lambda^t \frac{\pi(v)}{\min_u \pi(u)}$$

Proposition : si la chaîne est inversible et si sa matrice de probabilités de transitions a pour valeurs propres $1 = \lambda_0 > \lambda_1 \geq \ldots \geq \lambda_{N-1}$, en posant $\lambda = \max(|\lambda_1|, |\lambda_{N-1}|)$ on a, indépendamment de l'état de départ,

$$|\mathbb{P}(X_t = v) - \pi(v)| \leq \lambda^t \frac{\pi(v)}{\min_u \pi(u)}$$

Conséquence : $\delta(t) \leq \lambda^t / \min_u \pi(u)$, et

$$\tau(\epsilon) \leq \frac{1}{1-\lambda} \left(\ln \left(\frac{1}{\epsilon} \right) + \ln \left(\frac{1}{\min_u \pi(u)} \right) \right)$$

Proposition : si la chaîne est inversible et si sa matrice de probabilités de transitions a pour valeurs propres $1 = \lambda_0 > \lambda_1 \geq \ldots \geq \lambda_{N-1}$, en posant $\lambda = \max(|\lambda_1|, |\lambda_{N-1}|)$ on a, indépendamment de l'état de départ,

$$|\mathbb{P}(X_t = v) - \pi(v)| \leq \lambda^t \frac{\pi(v)}{\min_u \pi(u)}$$

Conséquence : $\delta(t) \leq \lambda^t / \min_u \pi(u)$, et

$$\tau(\epsilon) \leq \frac{1}{1-\lambda}\left(\ln\left(\frac{1}{\epsilon}\right) + \ln\left(\frac{1}{\min_u \pi(u)}\right)\right)$$

En d'autres termes, mélange rapide dès que $\lambda = 1 - \Omega(n^{-\alpha})$ et $\pi(u) = \Omega(e^{-n^{\beta}})$

Preuve

Soit D la matrice diagonale $D_{u,v} = \delta_{uv}\sqrt{\pi(u)}$; la matrice $A = D.P.D^{-1}$ est symétrique, avec les mêmes valeurs propres que P, donc diagonalisable dans une base orthonormale :

$$A = D.P.D^{-1} = \sum_{0 \leq i < N} \lambda_i \mathbf{E}^{(i)}$$

Preuve

Soit D la matrice diagonale $D_{u,v} = \delta_{uv}\sqrt{\pi(u)}$; la matrice $A = D.P.D^{-1}$ est symétrique, avec les mêmes valeurs propres que P, donc diagonalisable dans une base orthonormale :

$$A = D.P.D^{-1} = \sum_{0 \le i < N} \lambda_i \mathbf{E}^{(i)}$$

(chaque matrice $\mathbf{E}^{(i)}$ est la projection sur une direction propre, $\mathbf{E}^{(0)}$ projette sur le vecteur π)

Preuve

Soit D la matrice diagonale $D_{u,v} = \delta_{uv}\sqrt{\pi(u)}$; la matrice $A = D.P.D^{-1}$ est symétrique, avec les mêmes valeurs propres que P, donc diagonalisable dans une base orthonormale :

$$A = D.P.D^{-1} = \sum_{0 \leq i < N} \lambda_i \mathbf{E}^{(i)}$$

$$P^t_{u,v} = \pi(v) + \sqrt{\frac{\pi(v)}{\pi(u)}} \sum_{0 < i < N} \lambda_i^t \mathbf{e}_u^{(i)} \mathbf{e}_v^{(i)}$$

Preuve

Soit D la matrice diagonale $D_{u,v} = \delta_{uv}\sqrt{\pi(u)}$; la matrice $A = D.P.D^{-1}$ est symétrique, avec les mêmes valeurs propres que P, donc diagonalisable dans une base orthonormale :

$$A = D.P.D^{-1} = \sum_{0 \leq i < N} \lambda_i \mathbf{E}^{(i)}$$

$$P_{u,v}^t = \pi(v) + \sqrt{\frac{\pi(v)}{\pi(u)}} \sum_{0 < i < N} \lambda_i^t \mathbf{e}_u^{(i)} \mathbf{e}_v^{(i)}$$

d'où

$$|P_{u,v}^t - \pi(v)| \leq \frac{\pi(v)}{\sqrt{\pi(u)\pi(v)}} \lambda^t \sum_{1 < i < N} |\mathbf{e}_u^{(i)}|.|\mathbf{e}_v^{(i)}| \leq \frac{\lambda^t \pi(v)}{\min_w \pi(w)}$$

Conductance

Pour une arête $e = (u, v)$, $Q(e) = \pi(u)p_{u,v}$ représente la probabilité (stationnaire) d'être en u et de passer en v en une transition. Pour deux parties S et S' de Ω, on définit

$$Q(S, S') = \sum_{u \in S, v \in S'} Q(u, v)$$

qui est la probabilité de passer de S à S' en une étape.
La conductance de la chaîne est définie par

$$\Phi = \min_{\pi(S) \leq 1/2} \frac{Q(S, \overline{S})}{\pi(S)}$$

(le minimum, sur tous les “petits” ensembles S, de la probabilité de sortir de S sachant qu'on est dans S)

Conductance

"Intuitivement" (Jerrum & Sinclair, 1989), si Φ n'est pas trop petite, aucune partie S de Ω n'est telle qu'il est "difficile" d'en sortir, donc la chaîne devrait mélanger assez rapidement. . .

Théoreme (Jerrum, Sinclair) La seconde valeur propre d'une chaîne inversible, de probabilités de boucle au moins $1/2$, vérifie

$$1 - 2\Phi \leq \lambda \leq 1 - \frac{\Phi^2}{2}$$

Chemins canoniques

Une technique d'évaluation (minoration) de la conductance, est celle des chemins canoniques : pour chaque paire d'états (u, v), on choisit dans le graphe un chemin γ_{uv} de u à v (i.e., une suite de transitions possibles de la chaîne permettant de passer de l'état u à l'état v).

Chemins canoniques

Une technique d'évaluation (minoration) de la conductance, est celle des chemins canoniques : pour chaque paire d'états (u, v), on choisit dans le graphe un chemin γ_{uv} de u à v (i.e., une suite de transitions possibles de la chaîne permettant de passer de l'état u à l'état v). La "charge" d'une arête $e = (u, v)$ est alors

$$\rho(e) = \frac{1}{Q(e)} \sum_{(u,v):e \in \gamma_{uv}} \pi(u)\pi(v)$$

et on pose

$$\rho = \max_e \rho(e)$$

Chemins canoniques et conductance

Theorème (Jerrum, Sinclair) Pour toute chaîne inversible et tout choix de chemins canoniques,

$$\Phi \geq \frac{1}{2\rho}$$

Chemins canoniques et conductance

Theorème (Jerrum, Sinclair) Pour toute chaîne inversible et tout choix de chemins canoniques,

$$\Phi \geq \frac{1}{2\rho}$$

Conséquence :

$$1 - \lambda \geq \frac{1}{8\rho^2}$$

Chemins canoniques et conductance

Theorème (Jerrum, Sinclair) Pour toute chaîne inversible et tout choix de chemins canoniques,

$$\Phi \geq \frac{1}{2\rho}$$

Preuve : soit S, $\pi(S) \leq 1/2$, qui minimise $Q(S, \overline{S})/\pi(S) = \Phi$.

Chemins canoniques et conductance

Theorème (Jerrum, Sinclair) Pour toute chaîne inversible et tout choix de chemins canoniques,

$$\Phi \geq \frac{1}{2\rho}$$

Preuve : soit S, $\pi(S) \leq 1/2$, qui minimise $Q(S,\overline{S})/\pi(S) = \Phi$. On a donc $\pi(\overline{S}) \geq 1/2$, et $\pi(S)\pi(\overline{S}) = \frac{Q(S,\overline{S})\pi(\overline{S})}{\Phi} \geq \frac{Q(S,\overline{S})}{2\Phi}$.

Chemins canoniques et conductance

Theorème (Jerrum, Sinclair) Pour toute chaîne inversible et tout choix de chemins canoniques,

$$\Phi \geq \frac{1}{2\rho}$$

Preuve : soit S, $\pi(S) \leq 1/2$, qui minimise $Q(S,\overline{S})/\pi(S) = \Phi$.
On a donc $\pi(\overline{S}) \geq 1/2$, et $\pi(S)\pi(\overline{S}) = \frac{Q(S,\overline{S})\pi(\overline{S})}{\Phi} \geq \frac{Q(S,\overline{S})}{2\Phi}$.
Pour tout couple (u,v) avec $u \in S$ et $v \in \overline{S}$, le chemin γ_{uv} emprunte au moins une des arêtes de la coupe $(S,\overline{S})$.

Chemins canoniques et conductance

Theorème (Jerrum, Sinclair) Pour toute chaîne inversible et tout choix de chemins canoniques,

$$\Phi \geq \frac{1}{2\rho}$$

Preuve : soit S, $\pi(S) \leq 1/2$, qui minimise $Q(S,\overline{S})/\pi(S) = \Phi$. On a donc $\pi(\overline{S}) \geq 1/2$, et $\pi(S)\pi(\overline{S}) = \frac{Q(S,\overline{S})\pi(\overline{S})}{\Phi} \geq \frac{Q(S,\overline{S})}{2\Phi}$. Pour tout couple (u,v) avec $u \in S$ et $v \in \overline{S}$, le chemin γ_{uv} emprunte au moins une des arêtes de la coupe $(S,\overline{S})$. Donc on a

$$\sum_{e=(u,v):u\in S, v\in\overline{S}} \rho(e)Q(e) \geq \sum_{w\in S, z\in\overline{S}} \pi(w)\pi(z) = \pi(S)\pi(\overline{S}) \geq \frac{Q(S,\overline{S})}{2\Phi}.$$