Simulation parfaite des chaînes de Markov

Ana Bušić

INRIA - ENS

http://www.di.ens.fr/~busic/

ana.busic@inria.fr

Master AMIS - UVSQ

Versailles, 28 octobre 2016

Chaînes de Markov reversibles

Soit $\{X_t\}_{t\geq 0}$ une chaîne de Markov avec l'espace d'états $S = \{s_1, \ldots, s_n\}$ et la matrice de transition P. Une distribution de probabilités π sur S est dite reversible si pour tout $i, j \in \{1, \ldots, n\}$,

$$\pi_i P_{i,j} = \pi_j P_{j,i}.$$

Chaînes de Markov reversibles

Soit $\{X_t\}_{t\geq 0}$ une chaîne de Markov avec l'espace d'états $S = \{s_1, \ldots, s_n\}$ et la matrice de transition P. Une distribution de probabilités π sur S est dite reversible si pour tout $i, j \in \{1, \ldots, n\}$,

$$\pi_i P_{i,j} = \pi_j P_{j,i}.$$

Théorème 1.

Si π est une distribution reversible, alors elle est aussi une distribution stationnaire.

Chaînes de Markov reversibles

Soit $\{X_t\}_{t\geq 0}$ une chaîne de Markov avec l'espace d'états $S = \{s_1, \ldots, s_n\}$ et la matrice de transition P. Une distribution de probabilités π sur S est dite reversible si pour tout $i, j \in \{1, \ldots, n\}$,

$$\pi_i P_{i,j} = \pi_j P_{j,i}.$$

Théorème 1.
Si π est une distribution reversible, alors elle est aussi une distribution stationnaire.

Démonstration.
Pour tout $j \in \{1, \ldots, n\}$, on a

$$\pi_j = \pi_j \sum_{i=1}^{n} P_{j,i} = \sum_{i=1}^{n} \pi_j P_{j,i} = \sum_{i=1}^{n} \pi_i P_{i,j}.$$

$\square$

Temps inverse

Théorème 2.
Soit $\{X_t\}_{t\geq 0}$ une chaîne de Markov avec l'espace d'états S et la matrice de transition P et la distribution stationnaire π reversible.

Si $X_0 \sim \pi$, alors pour tout $k \in \mathbb{N}$ et tous $s_{i_0}, s_{i_1}, \ldots, s_{i_k} \in S$,

$$P(X_0 = s_{i_0}, X_1 = \ldots, X_k = s_{i_k}) = P(X_0 = s_{i_k}, X_1 = s_{i_{k-1}}, \ldots, X_k = s_{i_0}).$$

La probabilité d'un chemin dans un sens est égale à la probabilité du même chemin dans le sens inverse.

Exemple : processus de naissance et de mort

Matrice de transition P telle que :

- $P_{i,j} > 0$ si $|i-j| = 1$,
- $P_{i,j} = 0$ si $|i-j| \geq 2$.

Notons par

$$\pi_i^* = \frac{\prod_{k=1}^{i-1} P_{k,k+1}}{\prod_{k=1}^{i-1} P_{k+1,k}}.$$

Alors,

$$\pi = (\pi_1, \ldots, \pi_n) = (\pi_1^*, \ldots, \pi_n^*)/(\sum_{i=1}^{n} \pi_i^*)$$

est une distribution de probabilités reversible.

Exemple : marche aléatoire sur un graphe

Un graphe $G = (V, E)$ avec les sommets $V = \{v_1, \ldots, v_n\}$ et les arêtes $E = \{e_1, \ldots, e_m\}$. Soit d_i le degré du sommet v_i.

Exemple : marche aléatoire sur un graphe

Un graphe $G = (V, E)$ avec les sommets $V = \{v_1, \ldots, v_n\}$ et les arêtes $E = \{e_1, \ldots, e_m\}$. Soit d_i le degré du sommet v_i.

Un marcheur qui est au sommet v_i à l'instant t change sa position pour un sommet voisin de v_i avec la même probabilité pour tous les voisins.

Exemple : marche aléatoire sur un graphe

Un graphe $G = (V, E)$ avec les sommets $V = \{v_1, \dots, v_n\}$ et les arêtes $E = \{e_1, \dots, e_m\}$. Soit d_i le degré du sommet v_i.

Un marcheur qui est au sommet v_i à l'instant t change sa position pour un sommet voisin de v_i avec la même probabilité pour tous les voisins.

Matrice de transition P :

$$P_{i,j} = 1/d_i, \text{ si } (v_i, v_j) \in E$$

et sinon $P_{i,j} = 0$.

Exemple : marche aléatoire sur un graphe

Un graphe $G = (V, E)$ avec les sommets $V = \{v_1, \dots, v_n\}$ et les arêtes $E = \{e_1, \dots, e_m\}$. Soit d_i le degré du sommet v_i.

Un marcheur qui est au sommet v_i à l'instant t change sa position pour un sommet voisin de v_i avec la même probabilité pour tous les voisins.

Matrice de transition P :

$$P_{i,j} = 1/d_i, \ \mathrm{si}\ (v_i, v_j) \in E$$

et sinon $P_{i,j} = 0$.

Alors,

$$\pi = (d_1, \dots, d_n)/(\sum_{i=1}^{n} d_i)$$

est une distribution de probabilités reversible.

Echantillonneur de Gibbs

$G = (V, E)$ un graphe avec les sommets $V = \{v_1, \ldots, v_n\}$ et les arêtes $E = \{e_1, \ldots, e_m\}$. $N(v)$ les voisins du sommet v.
C un ensemble fini - l'ensemble des valeurs pour les sommets.

Echantillonneur de Gibbs

$G = (V, E)$ un graphe avec les sommets $V = \{v_1, \dots, v_n\}$ et les arêtes $E = \{e_1, \dots, e_m\}$. $N(v)$ les voisins du sommet v.
C un ensemble fini - l'ensemble des valeurs pour les sommets.

L'éspace d'états : $\mathcal{X} = C^V$.
But : echantilloner une distribution π sur $\mathcal{X}$.

Echantillonneur de Gibbs

$G = (V, E)$ un graphe avec les sommets $V = \{v_1, \ldots, v_n\}$ et les arêtes $E = \{e_1, \ldots, e_m\}$. $N(v)$ les voisins du sommet v.
C un ensemble fini - l'ensemble des valeurs pour les sommets.

L'éspace d'états : $\mathcal{X} = C^V$.
But : echantilloner une distribution π sur $\mathcal{X}$.

Idée : créer une chaîne de Markov sur $\mathcal{X}$ qui a la distribution stationnaire π.

Echantillonneur de Gibbs

$G = (V, E)$ un graphe avec les sommets $V = \{v_1, \ldots, v_n\}$ et les arêtes $E = \{e_1, \ldots, e_m\}$. $N(v)$ les voisins du sommet v.
C un ensemble fini - l'ensemble des valeurs pour les sommets.

L'éspace d'états : $\mathcal{X} = C^V$.
But : echantilloner une distribution π sur $\mathcal{X}$.

Idée : créer une chaîne de Markov sur $\mathcal{X}$ qui a la distribution stationnaire π.

Algorithme

1. Choisir un sommet $v \in V$ uniformement.

Echantillonneur de Gibbs

$G = (V, E)$ un graphe avec les sommets $V = \{v_1, \dots, v_n\}$ et les arêtes $E = \{e_1, \dots, e_m\}$. $N(v)$ les voisins du sommet v.
C un ensemble fini - l'ensemble des valeurs pour les sommets.

L'éspace d'états : $\mathcal{X} = C^V$.
But : echantilloner une distribution π sur $\mathcal{X}$.

Idée : créer une chaîne de Markov sur $\mathcal{X}$ qui a la distribution stationnaire π.

Algorithme

1. Choisir un sommet $v \in V$ uniformement.
2. Choisir $X_{t+1}(v)$ selon la loi π sachant les valeurs de tous les sommets voisins :

$$P\big(X_{t+1}(v)\big) = \pi\big(X_{t+1}(v) \mid X_{t+1}(w) = X_t(w), \forall w \in N(v)\big)$$

Echantillonneur de Gibbs

$G = (V, E)$ un graphe avec les sommets $V = \{v_1, \ldots, v_n\}$ et les arêtes $E = \{e_1, \ldots, e_m\}$. $N(v)$ les voisins du sommet v.
C un ensemble fini - l'ensemble des valeurs pour les sommets.

L'éspace d'états : $\mathcal{X} = C^V$.
But : echantilloner une distribution π sur $\mathcal{X}$.

Idée : créer une chaîne de Markov sur $\mathcal{X}$ qui a la distribution stationnaire π.

Algorithme

1. Choisir un sommet $v \in V$ uniformement.
2. Choisir $X_{t+1}(v)$ selon la loi π sachant les valeurs de tous les sommets voisins :

$$P\big(X_{t+1}(v)\big) = \pi\big(X_{t+1}(v) \mid X_{t+1}(w) = X_t(w), \forall w \in N(v)\big)$$

3. $X_{t+1}(w) = X_t(w), \forall w \neq v$.

Exemple : coloriages propres

$G = (V, E)$ un graphe avec les sommets $V = \{v_1, \ldots, v_n\}$ et les arêtes $E = \{e_1, \ldots, e_m\}$. $N(v)$ les voisins du sommet v.

Exemple : coloriages propres

$G = (V, E)$ un graphe avec les sommets $V = \{v_1, \ldots, v_n\}$ et les arêtes $E = \{e_1, \ldots, e_m\}$. $N(v)$ les voisins du sommet v.

$q \geq 2$, $C = \{1, \ldots, q\}$. L'espace d'états : $\mathcal{X} \subset C^V$.
Un coloriage propre : pas de sommets voisins avec la même couleur.

Exemple : coloriages propres

$G = (V, E)$ un graphe avec les sommets $V = \{v_1, \dots, v_n\}$ et les arêtes $E = \{e_1, \dots, e_m\}$. $N(v)$ les voisins du sommet v.

$q \geq 2$, $C = \{1, \dots, q\}$. L'espace d'états : $\mathcal{X} \subset C^V$.
Un coloriage propre : pas de sommets voisins avec la même couleur.

Algorithme

1. Choisir un sommet $v \in V$ uniformément.

Exemple : coloriages propres

$G = (V, E)$ un graphe avec les sommets $V = \{v_1, \ldots, v_n\}$ et les arêtes $E = \{e_1, \ldots, e_m\}$. $N(v)$ les voisins du sommet v.

$q \geq 2$, $C = \{1, \ldots, q\}$. L'espace d'états : $\mathcal{X} \subset C^V$.
Un coloriage propre : pas de sommets voisins avec la même couleur.

Algorithme

1. Choisir un sommet $v \in V$ uniformement.
2. Choisir $X_{t+1}(v)$ uniformement parmi toutes les couleurs autorisées (pas de sommet voisin de cette couleur).

Exemple : coloriages propres

$G = (V, E)$ un graphe avec les sommets $V = \{v_1, \ldots, v_n\}$ et les arêtes $E = \{e_1, \ldots, e_m\}$. $N(v)$ les voisins du sommet v.

$q \geq 2$, $C = \{1, \ldots, q\}$. L'espace d'états : $\mathcal{X} \subset C^V$.
Un coloriage propre : pas de sommets voisins avec la même couleur.

Algorithme

1. Choisir un sommet $v \in V$ uniformement.
2. Choisir $X_{t+1}(v)$ uniformement parmi toutes les couleurs autorisées (pas de sommet voisin de cette couleur).
3. $X_{t+1}(w) = X_t(w), \forall w \neq v$.

Question

Quelle est la loi π et montrer qu'elle est reversible.

Exemple : ensembles indépendants d'un graphe

$G = (V, E)$ un graphe avec les sommets $V = \{v_1, \ldots, v_n\}$ et les arêtes $E = \{e_1, \ldots, e_m\}$. $N(v)$ les voisins du sommet v.

Un ensemble indépendant $U \subset V$ est un ensemble tel qu'il n'y a pas d'arêtes dans E reliant deux sommets dans U.

Exemple : ensembles indépendants d'un graphe

$G = (V, E)$ un graphe avec les sommets $V = \{v_1, \ldots, v_n\}$ et les arêtes $E = \{e_1, \ldots, e_m\}$. $N(v)$ les voisins du sommet v.

Un ensemble indépendant $U \subset V$ est un ensemble tel qu'il n'y a pas d'arêtes dans E reliant deux sommets dans U.

$C = \{0, 1\}$. L'espace d'états : $\mathcal{X} \subset C^V$.
L'état $x \in C^V$, $x(v) = 1$ ssi $v \in U$.

Exemple : ensembles indépendants d'un graphe

$G = (V, E)$ un graphe avec les sommets $V = \{v_1, \ldots, v_n\}$ et les arêtes $E = \{e_1, \ldots, e_m\}$. $N(v)$ les voisins du sommet v.

Un ensemble indépendant $U \subset V$ est un ensemble tel qu'il n'y a pas d'arêtes dans E reliant deux sommets dans U.

$C = \{0, 1\}$. L'espace d'états : $\mathcal{X} \subset C^V$.
L'état $x \in C^V$, $x(v) = 1$ ssi $v \in U$.

Algorithme

1. Choisir un sommet $v \in V$ uniformément.

Exemple : ensembles indépendants d'un graphe

$G = (V, E)$ un graphe avec les sommets $V = \{v_1, \ldots, v_n\}$ et les arêtes $E = \{e_1, \ldots, e_m\}$. $N(v)$ les voisins du sommet v.

Un ensemble indépendant $U \subset V$ est un ensemble tel qu'il n'y a pas d'arêtes dans E reliant deux sommets dans U.

$C = \{0, 1\}$. L'espace d'états : $\mathcal{X} \subset C^V$.
L'état $x \in C^V$, $x(v) = 1$ ssi $v \in U$.

Algorithme

1. Choisir un sommet $v \in V$ uniformément.
2. Jeter une pièce non-biaisée. Si "tête" et tous les sommets voisins de v ont la valeur 0, alors $X_{t+1}(v) = 1$, sinon $X_{t+1}(v) = 0$.

Exemple : ensembles indépendants d'un graphe

$G = (V, E)$ un graphe avec les sommets $V = \{v_1, \ldots, v_n\}$ et les arêtes $E = \{e_1, \ldots, e_m\}$. $N(v)$ les voisins du sommet v.

Un ensemble indépendant $U \subset V$ est un ensemble tel qu'il n'y a pas d'arêtes dans E reliant deux sommets dans U.

$C = \{0, 1\}$. L'espace d'états : $\mathcal{X} \subset C^V$.
L'état $x \in C^V$, $x(v) = 1$ ssi $v \in U$.

Algorithme

1. Choisir un sommet $v \in V$ uniformement.
2. Jeter une pièce non-biaisée. Si "tête" et tous les sommets voisins de v ont la valeur 0, alors $X_{t+1}(v) = 1$, sinon $X_{t+1}(v) = 0$.
3. $X_{t+1}(w) = X_t(w), \forall w \neq v$.

Exemple : ensembles indépendants d'un graphe

Question

Quelle est la loi π et montrer qu'elle est reversible.

Question

Modifier l'algorithme pour favoriser les ensembles indépendants de grande cardinalité :

$$\pi_\lambda(x) = \frac{\lambda^{|x|}}{Z_\lambda}$$

pour toutes les configurations admissibles $x \subset C^V$ (et $\pi_\lambda(x) = 0$ si x n'est pas un ensemble indépendant). $|x|$ est la cardinalité de x (nombre de sommets avec la valeur 1) et Z_λ est la constante de normalisation.

Chaînes bornantes et la simulation parfaite

[Huber, 2004]

L'espace d'états : $\mathcal{X} \subset C^V$.

Chaîne $\{X_n\}$ evolue dans $\mathcal{X}$.
Chaîne $\{Y_n\}$ evolue dans $(2^C)^V$.

Chaîne $\{Y_n\}$ est une chaîne bornante pour $\{X_n\}$ s'il existe un couplage entre $\{X_n\}$ et $\{Y_n\}$ tel que :

$$X_n(v) \in Y_n(v), \quad \forall v \in V \Rightarrow X_{n+1}(v) \in Y_{n+1}(v), \quad \forall v \in V.$$

Construction :

$$Y_{n+1}(v) = \cup_{x \in_V Y_n}(\phi(x))(v),$$

avec $x \in_V y$ si $x(v) \in y(v), \forall v \in V$.

Example : coloriages propres

Coloriage propre d'un graphe (V, E) : $x \in C^V$ tel que pour tout $v, w \in E$, $x(v) \neq x(w)$.

Gibbs sampler :

1. Choisir v uniformement dans $V = \{1, \ldots, n\}$
2. Choisir c uniformement dans $C = \{1, \ldots, k\}$ tant que $c \notin x(N_v)$
3. $x(v) \leftarrow c$

Example : coloriages propres

Coloriage propre d'un graphe (V, E) : $x \in C^V$ tel que pour tout $v, w \in E$, $x(v) \neq x(w)$.

Gibbs sampler :

1. Choisir v uniformement dans $V = \{1, \ldots, n\}$
2. Choisir c uniformement dans $C = \{1, \ldots, k\}$ tant que $c \notin x(N_v)$
3. $x(v) \leftarrow c$

Borne :

1. Choisir v uniformement dans $V = \{1, \ldots, n\}$, $y(v) \leftarrow \emptyset$
2. Répéter :

   Choisir c uniformement dans $C = \{1, \ldots, k\}$
   Si pas de $w \in N_v$ tel que $y(w) = \{c\}$,
   alors $y(v) \leftarrow y(v) \cup \{c\}$

   tant que $c \notin \cup_{w \in N_v} y(w)$ ou $|y(v)| > \Delta$

Example : coloriages propres

Soit Δ le degré maximum du graphe.

Théorème. Si $k \geq \Delta(\Delta + 2)$, alors la probabilité que la chaîne bornante detecte le couplage en $\log_\beta n + \theta$ est au moins $1 - \beta^\theta$, avec

$$\beta = 1 - \frac{1 - (\Delta + 1)\Delta/(k - \Delta + 1)}{n}.$$

Variante

L'espace d'états : $\mathcal{X} \subset C^V$.

Chaîne $\{X_n\}$ evolue dans $\mathcal{X}$.
Chaîne $\{Y_n\}$ evolue dans $C^{(2^V)}$. Décrite par $\{(B_n, D_n)\}, B_n, D_n \in C^{(2^V)}$.

Chaîne $\{Y_n\}$ est une chaîne bornante pour $\{X_n\}$ s'il existe un couplage entre $\{X_n\}$ et $\{Y_n\}$ tel que :
$B_n(c) \subset A_n(c) \subset B_n(c) \cup D_n(c), \; \forall c \in C$
$\Rightarrow B_{n+1}(c) \subset A_{n+1}(c) \subset B_{n+1}(c) \cup D_{n+1}(c), \; \forall c \in C.$

Notation : $a \in_C (b, d)$ si $b(c) \subset a(c) \subset b(c) \cup d(c), \forall c \in C.$

Construction :

$$B_{n+1}(c) = \cap_{a \in_C (B_n, D_n)} (\phi(a))(c),$$

$$D_{n+1}(c) = \left(\cup_{a \in_C (B_n, D_n)} (\phi(a))(c)\right) \setminus B_{n+1},$$

Example : ensembles indépendants

Distribution sur les ensembles indépendants :

$$\pi(A) = \frac{\lambda^{|A|}}{Z_\lambda}.$$

Dyer et Greenhil (2000) chaîne de Markov
(variante de Gibbs sampler) :

1. Choisir v uniformement dans $\{1, \ldots, n\}$
2. Choisir U uniformement dans $[0, 1]$
3. Cas I : Si $U > \frac{\lambda}{(\lambda+1)}$, $A \leftarrow A \backslash \{v\}$
4. Cas II : Si $U < \frac{\lambda}{(\lambda+1)}$ et pas de voisins de v dans A, alors $A \leftarrow A \cup \{v\}$
5. Cas III : Si $U < p_{swap} \frac{\lambda}{(\lambda+1)}$ et exactement un voisin w de v dans A, alors $A \leftarrow A \backslash \{w\} \cup \{v\}$

Example : ensembles indépendants

Soit Δ le degré maximum du graphe.

Théorème. Si $\lambda < \frac{2}{\Delta - 2}$, le temps de couplage de la chaîne bornante est inférieur à $\log_\beta n + \theta$ avec probabilité au moins $1 - \beta^\theta$, avec

$$\beta = \frac{\Delta\lambda}{2(\lambda + 1)}.$$

Algorithme Metropolis

Objectif : Echantilloner une distribution π sur $S = \{s_1, \ldots, s_n\}$.
L'idée : Construir un graphe de voisinage entre les états.
Hypothèses :

- Le graphe G est connexe.
- Les degrés des sommets sont petits.

$$P_{i,j} = \begin{cases} \frac{1}{d_i} \min\{\frac{\pi_j d_i}{\pi_i d_j}, 1\}, & (v_i, v_j) \in E \\ 0, & i \neq j, (v_i, v_j) \in E \\ 1 - \sum_{\ell \mid (s_\ell, s_i) \in E} \min\{\frac{\pi_j d_i}{\pi_i d_j}, 1\}, & i = j \end{cases}$$

Question

Monterer que π est la distribution reversible (et donc stationnaire) de P.

Algorithme Metropolis

Algorithme :
Soit $X_t = s_i$. Alors,

1. Choisir un voisin de s_i uniformement parmi les voisins.
2. $$X_{t+1} = \begin{cases} s_j, & \text{avec probabilité } \min\{\frac{\pi_j d_i}{\pi_i d_j}, 1\} \\ s_i, & \text{avec probabilité } 1 - \min\{\frac{\pi_j d_i}{\pi_i d_j}, 1\} \end{cases}$$

L'optimisation : recuit simulé

Problème :

- $\mathcal{X} = \{s_1, \ldots, s_n\}$.
- On cherche à minimiser une fonction $f : \mathcal{X} \to \mathbb{R}$.

$T > 0$ temperature.
Distribution de Boltzmann :

$$\pi_{f,T}(s) = \frac{1}{Z_{f,T}} exp(\frac{-f(s)}{T}).$$

avec $Z_{f,T}$ constante de normalisation.

L'optimisation : recuit simulé

Théorème 3.
Soit $\alpha(T)$ la probabilité qu'un élément choisi selon $\pi_{f,T}$ véfifie

$$f(Y) = \min_{s \in \mathcal{X}} f(s)$$

Alors

$$\lim_{T \to 0} \alpha(T) = 1.$$

Question

Montrer le théorème sous l'hypothèse que f a un unique minimum.
Et si le minimum n'est pas unique ?