

Méthodes de Monte Carlo par Chaînes de Markov

Eric MOULINES

Telecom Paris Tech
CNRS - LTCI

présenté par Gersende FORT (LTCI, CNRS & Telecom Paris Tech)

collaborations : S. Le Corff (PhD, Telecom ParisTech) et P. Priouret (Professeur, Paris 6).

I. Un problème de restauration d'images

Modèle d'Ising : énergie

- ▶ Un **système de spins** est une distribution de probabilité sur $X = \{0,1\}^{|V|}$, où V est l'ensemble des sommets d'un graphe $G = (V, E)$. E est l'ensemble des arêtes du graphe d'adjacence. La valeur $x_v \in \{0,1\}$ est le **spin** au sommet v .
- ▶ Le **modèle d'Ising** est le modèle de spin le plus couramment étudié. Dans ce système, l'**énergie** d'une configuration x est définie par

$$H(x) = -J \sum_{v,w \in V, v \sim w} \mathbb{1}_{\{x_v = x_w\}} - \sum_{v \in V} \{h_{v,1} \mathbb{1}_{\{1\}}(x_v) + h_{v,0} \mathbb{1}_{\{0\}}(x_v)\}$$

où $v \sim w$ signifie que $(v,w) \in E$ et $\{(h_{v,1}, h_{v,0})\}_{v \in V}$ est le champ externe, supposé anisotrope

- ▶ Nous considérons dans la suite le cas **ferromagnétique** où $J > 0$. Dans ce modèle, l'énergie augmente comme le nombre de paires (v,w) de sites voisins dont les spins sont différents...

Modèle d'Ising : distribution de Gibbs

- ▶ La **distribution de Gibbs** associée à l'énergie H est la distribution π sur X définie par

$$\pi(\mathbf{x}) = \frac{1}{Z(\beta)} e^{-\beta H(\mathbf{x})}, \quad (\beta \geq 0).$$

- ▶ $Z(\beta)$, la **fonction de partition** est la constante de normalisation,

$$Z(\beta) \stackrel{\text{def}}{=} \sum_{\mathbf{x} \in X} e^{-\beta H(\mathbf{x})}.$$

- ▶ β est l'inverse de la température.
 - ▶ A température infinie ($\beta = 0$), la fonction d'énergie H ne joue aucun rôle et π est la loi uniforme sur X .
 - ▶ Lorsque la température diminue, la distribution rend plus **probable** les configurations de faibles énergies.

Modèle d'Ising: influence de la température

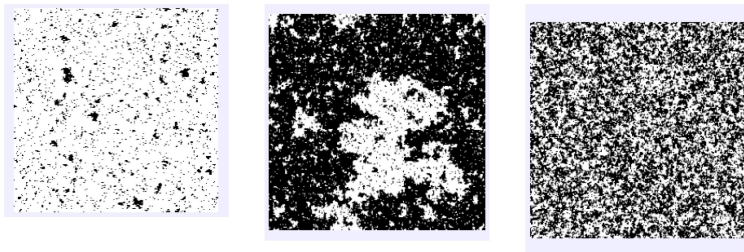


FIG.: Température croissante

Transition de phase

- ▶ Le cas le plus souvent utilisé en pratique est le cas d'un cristal ferromagnétique invariant par translation sur un réseau d -dimensionnel
- ▶ Dans ses travaux de doctorat, Ising (1924) a étudié le modèle en dimension 1. Dans ce cas, la corrélation $\text{Cov}_\pi(X_v, X_w)$ décroît exponentiellement vite avec la distance des sites $|v - w|$, indépendamment de la température.
- ▶ En dimension 2, Peierls (1930) a montré l'existence d'une transition de phase entre une phase désordonnée et une phase ordonnée. Le système est désordonné pour des petites valeurs de β et ordonné pour de grandes valeurs de β , au sens où $\text{Cov}_\pi(X_v, X_w) \geq C > 0$ pour tout $(v, w) \in V \times V$.

Modèle de Potts

- ▶ Le modèle de Potts est une généralisation naturelle du modèle d'Ising au cas où $X = \{1, 2, \dots, Q\}^{|V|}$.
- ▶ La fonction d'énergie est alors donnée par

$$H(\mathbf{x}) = -J \sum_{v,w \in V, v \sim w} \mathbb{1}_{\{x_v = x_w\}} + \sum_{v \in V} \sum_{j=1}^Q h_{v,j} \mathbb{1}_{\{j\}}(x_v).$$

L'énergie diminue lorsque deux voisins ont la même valeur. $h_{v,j}$ donne l'influence d'un champ externe.

- ▶ La distribution de Gibbs associée au modèle de Potts est donnée par

$$\pi(\mathbf{x}) = \frac{1}{Z(\beta)} e^{-\beta H(\mathbf{x})},$$

où $Z(\beta)$ est la **fonction de partition** $\beta \geq 0$.

Modèle élémentaire d'image bruitée

- ▶ $\mathbf{Y} = \{Y_v\}$ des observés sur un réseau régulier V à valeurs dans $\{0,1\}^{|V|}$.
- ▶ On modélise $Y_v = X_v \oplus Z_v$ où
 - (i) $\{Z_v\}_{v \in V}$ est une famille indépendante de variables de même loi de Bernoulli.
 - (ii) Le bruit $\{Z_v\}_{v \in V}$ est indépendant de l'image $\{X_v\}_{v \in V}$.
 - (iii) $\{X_v\}_{v \in V}$ est un modèle d'Ising ferromagnétique (sans champ externe).
 - (iv) \oplus est l'addition modulo 2.

$$p(\mathbf{y} \mid \mathbf{x}) = \frac{1}{Z_1(h)} \exp \left(-h \sum_{v \in V} \mathbb{1}_{\{y_v = x_v\}} \right)$$

$$p(\mathbf{x}) = \frac{1}{Z_2(\beta)} \exp \left(-\beta \sum_{v, w \in V, v \sim w} \mathbb{1}_{\{x_v = x_w\}} \right)$$

Restauration Bayésienne

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}) &= \frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{y})} \\ &= \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \\ &\propto \exp\left(-\beta \sum_{v,w \in V, v \sim w} \mathbb{1}_{\{x_v=x_w\}} - h \sum_{v \in V} \mathbb{1}_{\{y_v=x_v\}}\right) \end{aligned}$$

la loi de l'image **non bruitée** conditionnellement à l'image **bruitée** est un modèle d'Ising avec un champ anisotrope, **dépendant de l'image bruitée**.

Objectif de la restauration d'image

- ▶ **Observation:** une réalisation $\{\mathbf{y}_v\}_{v \in V}$ d'une image entachée d'un bruit additif.
- ▶ **Objectif:** Calculer
 1. La moyenne a posteriori de l'image non bruitée conditionnellement à l'image bruitée

$$\mathbb{E}[\mathbf{X} | \mathbf{Y} = \mathbf{y}] = \sum_{\mathbf{x}} \mathbf{x} p(\mathbf{x} | \mathbf{y})$$

2. Le maximum a posteriori

$$\hat{\mathbf{x}} = \operatorname{argmax}_{\mathbf{x}} p(\mathbf{x} | \mathbf{y}) .$$

II. Méthodes de Monte Carlo par Chaînes de Markov (MCMC)

Méthode de Monte Carlo par Chaînes de Markov (MCMC)

- ▶ La méthode MCMC consiste à construire une chaîne de Markov $\{X_n, n \in \mathbb{N}\}$ de noyau de transition P dont π soit la loi stationnaire.
- ▶ Les chaînes de Markov sont très simples à simuler
 1. **Simuler l'état initial X_0 suivant la loi initiale ξ .**
 2. **Simuler conditionnellement $X_{n+1} \sim P(X_n, \cdot)$ en utilisant le noyau de transition P**
- ▶ Lorsque la chaîne est **ergodique**, la loi de X_n donnée par $\xi_n = \xi P^n$ converge vers la loi stationnaire.
- ▶ Le théorème ergodique de Birkhoff montre aussi que la moyenne d'une réalisation de la chaîne converge vers la loi d'équilibre

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \rightarrow \int f d\pi .$$

Algorithmes de Hastings-Metropolis

- ▶ **Objectif:** Construire un noyau de transition admettant π comme loi stationnaire.
- ▶ **Idée:** De façon assez étonnante, il existe une façon universelle de procéder, introduite par Metropolis (circa 1950) et raffinée par Hastings (1970)
- ▶ **Algorithme:**
 1. **Proposer un candidat** Y_{k+1} **suivant un noyau de proposition de densité** $q(X_k, \cdot)$
 2. **Accepter ce candidat avec une probabilité égale à**

$$\alpha(X_k, Y_{k+1}) = \min \left(1, \frac{\pi(Y_{k+1})q(Y_{k+1}, X_k)}{\pi(X_k)q(X_k, Y_{k+1})} \right) .$$

- ▶ **Important:** le rapport d'acceptation ne dépend que du **rapport** $\pi(y)/\pi(x)$ et ne requiert pas la connaissance de la constante de normalisation !

Matrice de transition de la chaîne

- ▶ Matrice de transition de la chaîne

$$p(x, x') = q(x, x')\alpha(x, x') + \mathbb{1}_x(x') \sum_{\bar{x} \in X} (1 - \alpha(x, \bar{x}))q(x, \bar{x})$$

- ▶ Le taux d'acceptation vérifie

$$\begin{aligned} \pi(x)q(x, x')\alpha(x, x') &= \pi(x)q(x, x') \left(1 \wedge \frac{\pi(x')q(x', x)}{\pi(x)q(x, x')} \right) \\ &= \pi(x)q(x, x') \wedge \pi(x')q(x', x) \\ &= \pi(x')q(x', x)\alpha(x', x) \end{aligned}$$

- ▶ La matrice de transition satisfait la **condition de balance détaillée** par rapport à π

$$\pi(x)p(x, x') = \pi(x')p(x', x)$$

Réversibilité, balance détaillée et équilibre

$$\pi(x)p(x,x') = \pi(x')p(x',x)$$

- π est une loi stationnaire pour $P = (p(x,x'))$:

$$\begin{aligned} \sum_{x' \in X} \pi(x')p(x',x) &= \pi(x) \sum_{x' \in X} p(x,x') \\ &= \pi(x) \end{aligned}$$

car la matrice p est stochastique.

- Pour $f, g \in \mathcal{L}^2(X, \pi)$,

$$\begin{aligned} \langle f, Pg \rangle &= \sum_{x \in X} f(x)Pg(x)\pi(x) = \sum_{x \in X} \sum_{x' \in X} f(x)g(x')p(x,x')\pi(x) \\ &= \sum_{x \in X} \sum_{x' \in X} f(x)g(x')p(x',x)\pi(x') \\ &= \sum_{x' \in X} Pf(x')g(x')\pi(x') = \langle Pf, g \rangle \end{aligned}$$

Algorithmes de Hastings-Metropolis à marche aléatoire symétrique

- ▶ $Y_{k+1} = X_k + Z_{k+1}$ où $Z_{k+1} \sim_{\text{i.i.d.}} \bar{q}$, et \bar{q} est **symétrique**
- ▶ $q(x, x') = q(x', x) = \bar{q}(x' - x) = \bar{q}(x - x')$ et le taux d'acceptation est donné par

$$\alpha(x, x') = 1 \wedge \frac{\pi(x')}{\pi(x)}$$

- ▶ ... marche aléatoire de loi d'incrément q biaisée.
- ▶ Le noyau est **réversible** par rapport à π et la loi stationnaire de la chaîne est π .

Algorithme de Gibbs

- ▶ Supposons que $\mathbf{X} = X_1 \times \cdots \times X_m$ est un espace produit: $\mathbf{x} = (x_1, \dots, x_m)$, où $x_k \in X_k$. Pour tout $k \in \{1, \dots, m\}$, nous notons x_k le k ième composant de \mathbf{x} et par $\mathbf{x}_{-k} = \{x_\ell\}_{\ell \neq k}$ les autres composants.
- ▶ Soit $\pi_k(\cdot | \mathbf{x}_{-k})$ la probabilité conditionnelle définie par

$$\pi_k(x_k | \mathbf{x}_{-k}) = \frac{\pi(x_1, x_2, \dots, x_m)}{\sum_{x'_k \in X_k} \pi(x_1, \dots, x'_k, \dots, x_m)}$$

- ▶ L'algorithme de Gibbs consiste à mettre à jour une configuration $\xi^i = (\xi_1^i, \dots, \xi_m^i)$ de la façon suivante:
 pour $k = 1, 2, \dots, m$: Echantillonner X_k^{i+1} de loi $\pi_k(\cdot | X_1^{i+1}, \dots, X_{k-1}^{i+1}, X_{k+1}^i, \dots, X_m^i)$.
- ▶ **Réversibilité** Chaque noyau élémentaire m de l'algorithme de Gibbs est réversible par rapport à π et donc admet π comme loi d'équilibre.

Réversibilité de l'algorithme de Gibbs

- ▶ Appelons K_k le noyau de transition de la k ième étape de Gibbs

$$K_k(\mathbf{x}, \mathbf{x}') = \mathbb{1}_{\{x_{-k}\}}(\mathbf{x}'_{-k}) \pi_k(x'_k \mid \mathbf{x}_{-k}).$$

- ▶ Pour toutes fonctions $f, g \in \mathcal{L}^2(\mathcal{X}, \pi)$,

$$\begin{aligned} \langle f, K_k g \rangle &= \sum_{\mathbf{x}, \mathbf{x}'} f(\mathbf{x}) g(\mathbf{x}') \pi(\mathbf{x}) K_k(\mathbf{x}, \mathbf{x}') \\ &= \sum_{\mathbf{x}} \left\{ f(\mathbf{x}) \pi(\mathbf{x}) \sum_{\mathbf{x}'_k} g([\mathbf{x}'_k, \mathbf{x}_{-k}]) \pi_k(x'_k \mid \mathbf{x}_{-k}) \right\} \\ &= \sum_{[\mathbf{x}'_k, \mathbf{x}_{-k}]} \left\{ g([\mathbf{x}'_k, \mathbf{x}_{-k}]) \pi(\mathbf{x}'_k, \mathbf{x}_{-k}) \sum_{x_k} f(x_k, \mathbf{x}_{-k}) \pi_k(x_k \mid \mathbf{x}_{-k}) \right\} \\ &= \langle K_k f, g \rangle \end{aligned}$$

comme $\pi([\mathbf{x}_k, \mathbf{x}_{-k}]) \pi_k(x'_k \mid \mathbf{x}_{-k}) = \pi_k(x_k \mid \mathbf{x}_{-k}) \pi([\mathbf{x}'_k, \mathbf{x}_{-k}])$,

Gibbs comme cas particulier de HM

- ▶ Loi de proposition

$$q_k(\mathbf{x}, \mathbf{x}') = \mathbb{1}_{\mathbf{x}_{-k}}(\mathbf{x}'_{-k}) \pi_k(x'_k \mid \mathbf{x}_{-k})$$

- ▶ Rapport d'acceptation

$$\begin{aligned} \alpha(\mathbf{x}, [x'_k, \mathbf{x}_{-k}]) &= 1 \wedge \frac{\pi([x'_k, \mathbf{x}_{-k}]) \pi_k(x_k \mid \mathbf{x}_{-k})}{\pi(\mathbf{x}) \pi_k(x'_k \mid \mathbf{x}_{-k})} \\ &= 1 \end{aligned}$$

car $\pi([x_k, \mathbf{x}_{-k}]) \pi_k(x'_k \mid \mathbf{x}_{-k}) = \pi_k(x_k \mid \mathbf{x}_{-k}) \pi([x'_k, \mathbf{x}_{-k}])$,

- ▶ Pour l'algorithme de Gibbs, les mouvements sont acceptés avec une probabilité 1 !

Combinaison Metropolis-Gibbs

- ▶ Si l'on ne sait pas "simuler" suivant les lois conditionnelles $\pi_k(\cdot | \mathbf{x}_{-k})$ on peut utiliser un algorithme de Metropolis en gardant le cadre "Gibbs".
- ▶ Il suffit de proposer la k -ième composante x_k suivant un noyau de proposition sur X_k , $q_k(x_k, \cdot)$ et d'accepter ou refuser le mouvement avec la probabilité

$$\alpha(\mathbf{x}, [x'_k, \mathbf{x}_{-k}]) = 1 \wedge \frac{q_k(x'_k, x_k) \pi_k(x_k | \mathbf{x}_{-k})}{q_k(x_k, x'_k) \pi_k(x'_k | \mathbf{x}_{-k})}$$

- ▶ Base de très nombreux algorithmes MCMC!...

Noyaux Hybrides

- ▶ Supposons que K_1, \dots, K_m sont des matrices markoviennes admettant π comme loi stationnaire Alors
 - (a) $K_{\text{sys}} = K_1 K_2 \cdots K_m$ et
 - (b) $K_{\text{rand}} = \sum_{i=1}^m \alpha_i K_i$, où $\alpha_i > 0$ pour $i = 1, \dots, m$ $\sum_{i=1}^m \alpha_i = 1$, admettent π comme loi stationnaire.
- ▶ Si K_1, \dots, K_m sont π -réversible, K_{rand} est aussi π -reversible mais K_{sys} n'est pas nécessairement réversible.

Simulation dans un modèle d'Ising: Gibbs

- Loi conditionnelle d'un spin au sommet v

$$\pi_v(x_v \mid \mathbf{x}_{-v}) = \frac{e^{-S_v(x_v, \mathbf{x}_{-v})}}{e^{-S_v(0, \mathbf{x}_{-v})} + e^{-S_v(1, \mathbf{x}_{-v})}}$$

où

$$S_v(x_v, \mathbf{x}_{-v}) \stackrel{\text{def}}{=} \beta \sum_{u: u \sim v} \mathbb{1}_{\{x_v = x_u\}} + h_{v,0} \mathbb{1}_{\{0\}}(x_v) + h_{v,1} \mathbb{1}_{\{1\}}(x_v) .$$

... ne dépend que des valeurs des spins des sommet voisins de v dans le graphe

- **Algorithme:**

1. Choisir un sommet **au hasard** ou suivant un **balayage** prédéfini.
2. Mettre à jour la valeur du spin en simulant suivant la loi conditionnelle.

Simulation dans un modèle d'Ising: MH dans Gibbs

1. Choisir un sommet v **au hasard** ou suivant un **balayage** prédéfini.
2. Proposer $x'_v = x_v \oplus 1$
3. Calculer la différence d'énergie

$$\Delta S_v(\mathbf{x}) = S_v(x_v \oplus 1, \mathbf{x}_{-v}) - S_v(x_v, \mathbf{x}_{-v}) .$$

4. Accepter la nouvelle configuration avec une probabilité

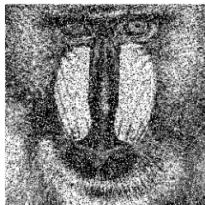
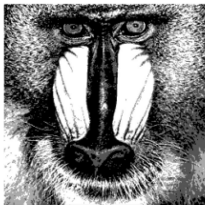
$$1 \wedge e^{-\Delta S_v(\mathbf{x})} .$$

On accepte toujours une configuration d'énergie plus basse...

Modèle d'Ising: influence de la température



Modèle d'Ising: influence de la température



III. Convergence d'une chaîne de Markov

Variation totale

- ▶ Si X est fini ou dénombrable, pour ξ une mesure signée finie sur $(X, \mathcal{P}(X))$ on définit

$$\|\xi\|_{\text{TV}} = \sum_{x \in X} |\xi(x)| .$$

- ▶ On dit qu'une matrice markovienne est **uniformément ergodique** si et seulement si

$$\lim_{k \rightarrow \infty} \sup_{x \in X} \|P^k(x, \cdot) - \pi\|_{\text{TV}} = 0$$

- ▶ On cherche à quantifier le plus précisément possible $\|P^k(x, \cdot) - \pi\|_{\text{TV}}$ pour obtenir une **garantie** de convergence !

III-1. Convergence d'une chaîne de Markov

Approche opérateur

Coefficient de Dobrushin

Définition (Coefficient de Dobrushin)

Notons

$\mathbb{M}_0(\mathcal{X}, \mathcal{X})$ l'ensemble des mesures signées finies de masse nulle.

Soit P une matrice markovienne sur \mathcal{X} . Le **coefficient de Dobrushin** $\Delta(P)$ de P est donné par

$$\Delta(P) \stackrel{\text{def}}{=} \sup \{ \|\xi P\|_{\text{TV}} : \xi \in \mathbb{M}_0(\mathcal{X}, \mathcal{X}), \|\xi\|_{\text{TV}} \leq 1 \} .$$

Notons que $\Delta(P) \leq 1$... Le coefficient de Dobrushin est vraiment intéressant lorsqu'il existe $m \in \mathbb{N}$ tel que $\Delta(P^m) < 1$!

Rq: $\Delta(P) \leq 1$ car

$$\begin{aligned}\|\xi P\|_{\text{TV}} &= \sum_y |\xi P(y)| = \sum_y \left| \sum_x \xi(x) P(x,y) \right| \\ &\leq \sum_y \sum_x |\xi(x)| P(x,y) = \sum_x \left(\sum_y P(x,y) \right) |\xi(x)| \\ &\leq \sum_x |\xi(x)| = \|\xi\|_{\text{TV}}\end{aligned}$$

donc

$$\Delta(P) \stackrel{\text{def}}{=} \sup \{ \|\xi P\|_{\text{TV}} : \xi \in \mathbb{M}_0(\mathbf{X}, \mathcal{X}), \|\xi\|_{\text{TV}} \leq 1 \} \leq 1.$$

sous-multiplicativité

Proposition

Si P et R sont des matrices markoviennes, alors $\Delta(PR) \leq \Delta(P)\Delta(R)$.

sous-multiplicativité

Proposition

Si P et R sont des matrices markoviennes, alors $\Delta(PR) \leq \Delta(P)\Delta(R)$.

Démonstration.

Soit $\xi \in \mathbb{M}_0(\mathcal{X}, \mathcal{X})$. Alors $\xi P(\mathcal{X}) = \xi(\mathcal{X}) = 0$ et donc $\xi P \in \mathbb{M}_0(\mathcal{X}, \mathcal{X})$. En posant $\nu = \xi P$, nous avons $\|\nu R\|_{\text{TV}} \leq \Delta(R) \|\nu\|_{\text{TV}}$. Par conséquent :

$$\|\xi PR\|_{\text{TV}} = \|(\xi P)R\|_{\text{TV}} \leq \Delta(R) \|\xi P\|_{\text{TV}} \leq \Delta(P)\Delta(R) \|\xi\|_{\text{TV}} .$$



Ergodicité uniforme

Notons

$\mathbb{M}_1(X, \mathcal{X})$ l'ensemble des probabilités sur (X, \mathcal{X}) .

Théorème

Soit P une matrice markovienne telle que pour un m entier, $\Delta(P^m) \leq \rho < 1$. Alors, P admet une unique probabilité invariante π . De plus, pour tout $\nu \in \mathbb{M}_1(X, \mathcal{X})$,

$$\|\nu P^n - \pi\|_{\text{TV}} \leq \rho^{\lfloor n/m \rfloor} \|\nu - \pi\|_{\text{TV}} ,$$

où $\lfloor u \rfloor$ est la partie entière de u .

Démonstration.

Pour $n \in \mathbb{N}$, on note $q = \lfloor n/m \rfloor$, et $r = n - qm \in \{0, \dots, m-1\}$. La sous-additivité du coefficient de Dobrushin et la propriété $\Delta(R) \leq 1$ montre que pour toute probabilité ν, ν'

$$\begin{aligned} \|\nu P^n - \nu' P^n\|_{\text{TV}} &\leq \Delta(P^n) \|\nu - \nu'\|_{\text{TV}} \\ &\leq \Delta(P^r) [\Delta(P^m)]^q \|\nu - \nu'\|_{\text{TV}} \leq \rho^{\lfloor n/m \rfloor} \|\nu - \nu'\|_{\text{TV}} . \end{aligned}$$

En prenant $\nu' = \nu P$, et en notant $\|\nu - \nu P\|_{\text{TV}} \leq 2$, nous avons donc

$$\|\nu P^n - \nu P^{n+1}\|_{\text{TV}} \leq 2\rho^{\lfloor n/m \rfloor} ,$$

qui montre que la série $\sum_n \|\nu P^n - \nu P^{n+1}\|_{\text{TV}}$ est convergente. Comme $(\mathbb{M}_1(X, \mathcal{X}), \|\cdot\|_{\text{TV}})$ est un espace de Banach (facile) à, ceci montre que la suite $\{\nu P^n : n \in \mathbb{N}\}$ converge vers une mesure signée π , dont on montre qu'elle est une probabilité (facile) . □

Condition de Doeblin

Définition (Doeblin condition)

Une matrice Markovienne P satisfait la **condition de Doeblin** s'il existe un entier $m \geq 1$, $\epsilon > 0$, et une mesure de probabilité μ sur X telle que, pour tout $x \in X$ et $A \in \mathcal{X}$,

$$P^m(x, A) \geq \epsilon \mu(A) .$$

De Doeblin à Dobrushin

1. Etape 1 : établir que

$$\Delta(P) = \frac{1}{2} \sup_{x,y} \|P(x,\cdot) - P(y,\cdot)\|_{\text{TV}} .$$

2. Etape 2 : établir que pour toute probabilités ν, ν' ,

$$\|\nu - \nu'\|_{\text{TV}} = \sum_{x \in X} |\nu(x) - \nu'(x)| = 2 \left(1 - \sum_{x \in X} \nu(x) \wedge \nu'(x) \right) .$$

De Doeblin à Dobrushin (suite)

- ▶ En posant $\nu = P^m(x, \cdot)$ et $\nu' = P^m(x', \cdot)$, nous avons

$$\Delta(P^m) = 1 - \sum_{y \in X} P^m(x, y) \wedge P^m(x', y)$$

- ▶ Si la matrice P satisfait la condition de Doeblin,

$$P^m(x, y) \wedge P^m(x', y) \geq \epsilon \mu(y)$$

$$\sum_{y \in X} P^m(x, y) \wedge P^m(x', y) \geq \epsilon,$$

car μ est une probabilité.

- ▶ Par conséquent, le coefficient de Dobrushin est borné par

$$\Delta(P^m) \leq 1 - \epsilon.$$

La vitesse de convergence est **explicite** et dépend d'une condition aisément vérifiable.

Cas particulier d'une matrice markovienne primitive

- ▶ Lorsque l'espace d'état X est fini, nous avons

$$\Delta P^m \leq 1 - |X| \min_{(x,x') \in X^2} P^m(x,x')$$

- ▶ Par conséquent, s'il existe un entier m tel que $P^m(x,x') > 0$ pour tout $(x,x') \in X^2$, alors la chaîne est uniformément géométriquement ergodique: il existe une unique probabilité $\pi \in \mathbb{M}_1(X, \mathcal{X})$ invariante et, pour tout $x \in X$,

$$\|P^n(x, \cdot) - \pi\|_{\text{TV}} \leq 2 \left(1 - |X| \min_{(x,x') \in X^2} P^m(x,x') \right)^{\lfloor n/m \rfloor} .$$

Démonstration étape 2

Démonstration.

Découle de l'identité élémentaire $|a - b| = a + b - 2(a \wedge b)$:

$$\begin{aligned}\|\nu - \nu'\|_{\text{TV}} &= \sum_{x \in X} (\nu(x) + \nu'(x) - 2(\nu(x) \wedge \nu'(x))) \\ &= 2 \left(1 - \sum_{x \in X} \nu(x) \wedge \nu'(x) \right).\end{aligned}$$



III-2. Convergence d'une chaîne de Markov

Approche par couplage

Approche par couplage

- ▶ Une alternative à la démonstration de l'ergodicité uniforme sous l'hypothèse de condition de Doeblin cas $m = 1$ pour simplifier

$$P(x, A) \geq \epsilon \mu(A) .$$

- ▶ On introduit le noyau résiduel

$$R(x, x') \stackrel{\text{def}}{=} \frac{P(x, x') - \epsilon \mu(x')}{1 - \epsilon}$$

- ▶ On construit un processus bivarié $\{(X_n, X'_n), n \geq 0\}$ de sorte que le **temps de couplage** est fini avec probabilité 1

$$T \stackrel{\text{def}}{=} \inf\{n \geq 0 : X_q = X'_q \quad \forall q \geq n\} .$$

Construction du processus bivarié

- ▶ Initialisation : $X_0 \sim \nu$ $X'_0 \sim \pi$ $d_0 = 0$
- ▶ Etant donné l'état courant (X_n, X'_n, d_n)
 - ▶ Si $d_n = 1$ alors $X_{n+1} \sim P(X_n, \cdot)$ $X'_{n+1} = X_{n+1}$ $d_{n+1} = 1$.
 - ▶ Si $d_n = 0$,
 - ▶ avec probabilité ϵ , $X_{n+1} = X'_{n+1} \sim \mu$ $d_{n+1} = 1$
 - ▶ sinon, $X_{n+1} \sim R(X_n, \cdot)$ $X'_{n+1} \sim R(X'_n, \cdot)$ $d_{n+1} = 0$

Propriétés de ce processus bivarié

- ▶ $T = \inf\{n \geq 1 : d_n = 1\}$.
- ▶ $X_n \sim \nu P^n \quad X'_n \sim \pi P^n = \pi$.

Par suite: (inégalité de couplage)

$$\begin{aligned}
 \|\nu P^n - \pi\|_{\text{TV}} &= \sum_y |\nu P^n(y) - \pi(y)| = \sum_y |\mathbb{P}(X_n = y) - \mathbb{P}(X'_n = y)| \\
 &= \sum_y |\mathbb{P}(X_n = y, T \leq n) - \mathbb{P}(X'_n = y, T \leq n)| \\
 &\quad + |\mathbb{P}(X_n = y, T > n) - \mathbb{P}(X'_n = y, T > n)| \\
 &= \sum_y |\mathbb{P}(X_n = y, T > n) - \mathbb{P}(X'_n = y, T > n)| \\
 &\leq 2 \mathbb{P}(T > n) \\
 &\leq 2 (1 - \epsilon)^n .
 \end{aligned}$$

IV. Conclusion

Conclusion

- ▶ Utilisation des chaînes de Markov (discrètes) pour la simulation Monte Carlo.
- ▶ Convergence des méthodes de simulation Monte Carlo par Chaînes de Markov (par ex. via approche opérateur, approche couplage).
- ▶ Illustration : application (simple) au traitement statistique des images.