

Méthodes de Monte Carlo par Chaînes de Markov adaptatives

Gersende FORT

LTCI, CNRS / TELECOM ParisTech

Méthodes MCMC

Introduction

Méthodes MCMC

Vers les procédures MCMC adaptatives

Algorithmes MCMC adaptatifs

Paramètres d'implémentation du SRW-HM

Choix des paramètres d'implémentation

Exemples d'algorithmes MCMC adaptatifs

Conclusion

Convergence des algorithmes MCMC adaptatifs

L'adaptation peut détruire la convergence

Conditions pour la convergence des marginales

Adaptation et Ergodicité

Résultat

En pratique, comment vérifier les conditions?

Doit-on avoir convergence du processus d'adaptation?

Conclusion

Références

Introduction

- Les méthodes **de Monte Carlo** permettent d'obtenir des réalisations de v.a. de loi $d\pi$ spécifiée par l'utilisateur.
- Elles sont en particulier utilisées comme
 - méthodes numériques probabilistes pour l'approximation d'intégrales

$$\int \phi d\pi = \mathbb{E}[\phi(X)] \quad X \sim d\pi$$

- méthodes d'exploration d'une loi (recherche des maxima, de la médiane, des zones de plus haute densité, ...)

On peut regrouper les méthodes de Monte Carlo en deux grandes familles

1. **méthodes par pondération** : se donner une loi instrumentale dg ,

- tirer des v.a. i.i.d. de loi dg
- et approcher la loi cible par

$$\pi \approx \frac{1}{n} \sum_{k=1}^n \frac{\pi(X_k)}{g(X_k)} \delta_{X_k} \quad \text{abus: densité / loi}$$

On peut regrouper les méthodes de Monte Carlo en deux grandes familles

1. **méthodes par pondération** : se donner une loi instrumentale dg ,

- tirer des v.a. i.i.d. de loi dg
- et approcher la loi cible par

$$\pi \approx \frac{1}{n} \sum_{k=1}^n \frac{\pi(X_k)}{g(X_k)} \delta_{X_k} \quad \text{abus: densité / loi}$$

2. **méthodes par acceptation-rejet** : se donner une loi instrumentale sg ,

- tirer des v.a. i.i.d. de loi dg
- et accepter chaque tirage que si \dots - (critère faisant intervenir $d\pi$ et dg) -
- approcher la loi cible par

$$\pi \approx \frac{1}{n} \sum_{k=1}^n \delta_{X_k} \quad (\text{moyenne des points acceptés})$$

Comment choisir la loi instrumentale g ?

$$\sup_x \frac{\pi}{g} < +\infty$$

C'est une condition

- suffisante pour les méthodes par pondération; et nécessaire en pratique pour éviter la dégénérescence des poids d'importance $\pi(X_k)/g(X_k)$.
- nécessaire pour les méthodes d'acceptation-rejet classiques.

Comment choisir la loi instrumentale g ?

$$\sup_x \frac{\pi}{g} < +\infty$$

C'est une condition

- suffisante pour les méthodes par pondération; et nécessaire en pratique pour éviter la dégénérescence des poids d'importance $\pi(X_k)/g(X_k)$.
- nécessaire pour les méthodes d'acceptation-rejet classiques.

C'est une condition d'autant plus contraignante que

- la dimension de l'espace de simulation X est grande.
- la loi cible est multimodale

Comment choisir la loi instrumentale g ?

$$\sup_x \frac{\pi}{g} < +\infty$$

C'est une condition

- suffisante pour les méthodes par pondération; et nécessaire en pratique pour éviter la dégénérescence des poids d'importance $\pi(X_k)/g(X_k)$.
- nécessaire pour les méthodes d'acceptation-rejet classiques.

C'est une condition d'autant plus contraignante que

- la dimension de l'espace de simulation X est grande.
- la loi cible est multimodale

Les méthodes par acceptation-rejet sont plus robustes à la dimension :

- méthodes d'A/R classiques: coût de simulation trop élevé.
- méthodes d'A/R par [Chaînes de Markov](#)

Méthodes Monte Carlo par Chaînes de Markov

Etant donnée une loi cible $\pi d\nu$, il s'agit de construire une chaîne de Markov de sorte que

- $\pi d\nu$ soit la probabilité stationnaire de la chaîne.
- la chaîne vérifie des propriétés de convergence :
 - de type **ergodicité** :

$$\sup_{\{\phi, |\phi|_{\infty} \leq 1\}} \left| \mathbb{E}_x [\phi(X_n)] - \int \phi \pi d\nu \right| \rightarrow 0 \quad \forall x \in X$$

- ou de type **loi des grands nombres** :

$$\frac{1}{n} \sum_{k=1}^n \phi(X_k) \xrightarrow{\text{p.s.}} \int \phi \pi d\nu$$

- et éventuellement de type **TLC**

$$\frac{\sqrt{n}}{\sigma} \left(\frac{1}{n} \sum_{k=1}^n \phi(X_k) - \int \phi \pi d\nu \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

Principalement, deux grandes familles d'algorithmes MCMC

④ algorithmes Hastings-Metropolis

- se donner un noyau de proposition $q(x,y) d\nu(y)$.
- Mise à jour $X_n \rightarrow X_{n+1}$: proposer un candidat $Y \sim q(X_n, \cdot) d\nu$ et

$$\begin{cases} X_{n+1} = Y & \text{avec probabilité } \alpha(X_n, Y) \\ X_{n+1} = X_n & \text{sinon} \end{cases}$$

où le ratio d'acceptation-rejet est donné par

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

Principalement, deux grandes familles d'algorithmes MCMC

① algorithmes Hastings-Metropolis

② algorithmes de Gibbs

- Décomposer l'espace d'état en sous-blocs $x = (x^{(1)}, \dots, x^{(d)})$
- Mise à jour $X_n \rightarrow X_{n+1}$: simuler le sous-bloc $X_{n+1}^{(i)}$ sous la loi conditionnelle du bloc i sachant les autres blocs

$$\propto \pi \left(X_{n+1}^{(1)}, \dots, X_{n+1}^{(i-1)}, y, X_n^{(i+1)}, \dots, X_n^{(d)} \right)$$

Principalement, deux grandes familles d'algorithmes MCMC

① algorithmes Hastings-Metropolis

② algorithmes de Gibbs

③ et les versions hybrides : *Metropolis within Gibbs*

dans l'algorithme de Gibbs, remplacer la simulation sous la loi conditionnelle par une étape de Hastings-Metropolis dont la loi cible est cette loi conditionnelle.

Vers les procédures MCMC adaptatives

Dans la mise en oeuvre de ces algorithmes, différents “degrés de liberté” laissés à l'utilisateur : par exemple,

- Hastings-Metropolis : quelle loi de proposition choisir ?
- Gibbs : quelle partition de l'espace d'état X ? quel ordre de mise à jour des composantes et avec quelle fréquence ?

⇒ définition d'algorithmes **adaptatifs** construits pour

- que l'algorithme apprenne de son propre comportement passé, comment modifier les **paramètres d'implémentation**
- et les modifie au fur et à mesure de son déroulement : chaque itération comporte une étape de mise à jour de la chaîne, et de mise à jour des paramètres d'implémentation.

II. Algorithmes MCMC adaptatifs

Dans la suite de cette section : illustrations à partir

- de l'algorithme de Hastings-Metropolis à marche aléatoire symétrique (SRW-HM)
- pour une densité cible π par rapport à la mesure de Lebesgue sur \mathbb{R}^d

Paramètres d'implémentation du SRW-HM (marche aléatoire symétrique Hastings-Metropolis)

Lorsque le noyau de proposition est de la forme

$$q(x,y) = q(y,x) \quad (\text{symétrie})$$

$$q(x,y) = q(y-x) \quad (\text{marche aléatoire})$$

une itération $X_n \rightarrow X_{n+1}$ devient :

- tirer $Z \sim q(z)dz$

- Poser

$$\begin{cases} X_{n+1} = X_n + Z & \text{avec probabilité } 1 \wedge \frac{\pi(X_n+Z)}{\pi(X_n)} \\ X_{n+1} = X_n & \text{sinon} \end{cases}$$

Paramètres d'implémentation du SRW-HM (marche aléatoire symétrique Hastings-Metropolis)

Lorsque le noyau de proposition est de la forme

$$q(x,y) = q(y,x) \quad (\text{symétrie})$$

$$q(x,y) = q(y-x) \quad (\text{marche aléatoire})$$

une itération $X_n \rightarrow X_{n+1}$ devient :

- tirer $Z \sim q(z)dz$

- Poser

$$\begin{cases} X_{n+1} = X_n + Z & \text{avec probabilité } 1 \wedge \frac{\pi(X_n+Z)}{\pi(X_n)} \\ X_{n+1} = X_n & \text{sinon} \end{cases}$$

En particulier, toute configuration $X_n + Z$ qui augmente la valeur de la densité cible est systématiquement acceptée.

Paramètres d'implémentation du SRW-HM (marche aléatoire symétrique Hastings-Metropolis)

Lorsque le noyau de proposition est de la forme

$$q(x,y) = q(y,x) \quad (\text{symétrie})$$

$$q(x,y) = q(y-x) \quad (\text{marche aléatoire})$$

une itération $X_n \rightarrow X_{n+1}$ devient :

- tirer $Z \sim q(z)dz$
- Poser

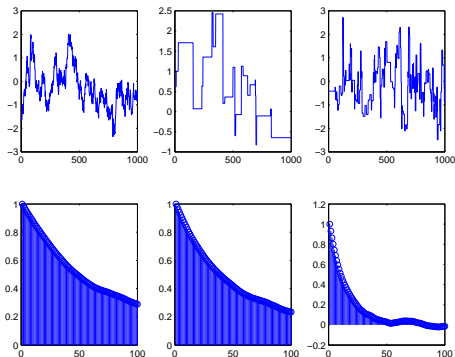
$$\begin{cases} X_{n+1} = X_n + Z & \text{avec probabilité } 1 \wedge \frac{\pi(X_n+Z)}{\pi(X_n)} \\ X_{n+1} = X_n & \text{sinon} \end{cases}$$

En particulier, toute configuration $X_n + Z$ qui augmente la valeur de la densité cible est systématiquement acceptée.

↔ En pratique, quel choix pour la loi de proposition q ?

Situation 1: "Vitesse de déplacement"

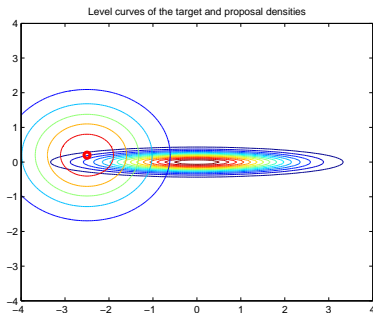
- Chaîne de Markov SRW-HM à valeur dans \mathbb{R}
- Avec loi de proposition $q = \mathcal{N}(0, \sigma^2)$ pour différentes valeurs de σ^2 .



Pour différentes valeurs de σ : [haut] trajectoire de la chaîne. [bas] fonction d'auto-corrélation.

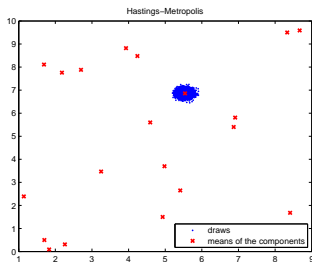
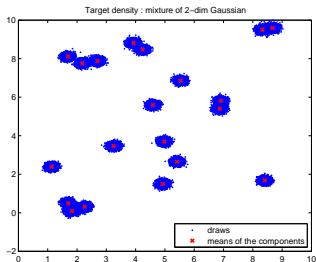
Situation 1' : Structure de covariance de la densité cible

- Chaîne de Markov SRW-HM à valeur dans \mathbb{R}^2 de densité cible $\mathcal{N}_2(0, \Sigma)$
- Avec loi de proposition $\mathcal{N}_d(0, \text{Id})$



Situation 2: Multimodalité de la densité cible

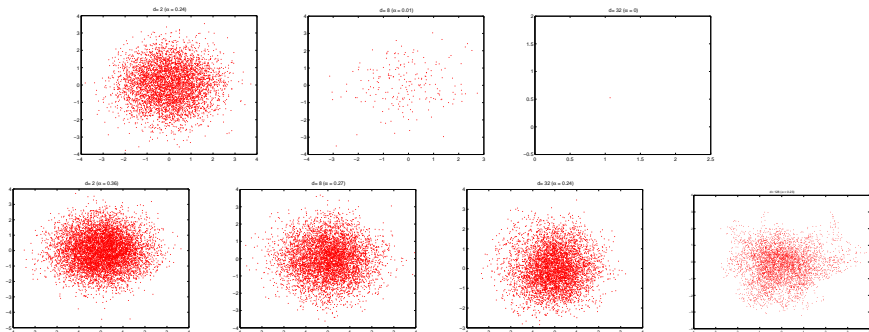
- Chaîne de Markov SRW-HM à valeur dans \mathbb{R}^2 de loi cible $\sum_{i=1}^{20} \mathcal{N}_2(\mu_i, \Sigma_i)$
- Avec loi de proposition $\mathcal{N}_d(0, \sigma^2 \text{Id})$



[gauche] Tirages i.i.d. sous la loi cible. [droite] Chaîne SRW-HM

Situation 3: Dimension de l'espace de simulation $X \subseteq \mathbb{R}^d$

- Chaîne de Markov SRW-HM à valeur dans \mathbb{R}^d de loi cible $\mathcal{N}_d(0, \text{Id})$
- Avec loi de proposition $\mathcal{N}_d(0, \sigma^2 \text{Id})$



Pour différentes valeurs de d : chaîne projetée dans \mathbb{R}^2 [haut] σ est indépendant de d
 [bas] σ est de la forme c/\sqrt{d}

Choix des paramètres d'implémentation

Quelques réponses dans le cas d'une loi de proposition $\mathcal{N}_d(0, \Sigma)$:

- choisir pour matrice de var-covar: $\Sigma \propto \Sigma_\pi$.
- avec un facteur d'échelle qui tient compte de la dimension

$$\Sigma = \frac{2.38^2}{d} \Sigma_\pi$$

Choix des paramètres d'implémentation

Quelques réponses dans le cas d'une loi de proposition $\mathcal{N}_d(0, \Sigma)$:

- choisir pour matrice de var-covar: $\Sigma \propto \Sigma_\pi$.
- avec un facteur d'échelle qui tient compte de la dimension

$$\Sigma = \frac{2.38^2}{d} \Sigma_\pi$$

Réponses apportées

- sous des conditions plus ou moins restrictives sur la densité cible.
- pour différents algorithmes de type Hastings-Metropolis.
- basées sur la vitesse de convergence optimale de "limites diffusives" de la chaîne (renormalisée).

Algorithme de Metropolis adaptatif

En pratique, la matrice de var-covar Σ_π est inconnue \implies coupler simulation et estimation

Notons

- θ_n l'estimation courante de la matrice Σ_π .
- P_θ le noyau de transition d'une chaîne SRW-HM de loi de proposition $\mathcal{N}_d(0, \theta)$

Une itération $(\theta_n, X_n) \rightarrow (\theta_{n+1}, X_{n+1})$ devient

- **Simuler** $X_{n+1} \sim P_{\theta_n}(X_n, \cdot)$
- **Mettre à jour l'estimation** de Σ_π par la formule

$$\theta_{n+1} = \theta_n + \frac{1}{n+1} \left((X_{n+1} - \mu_{n+1})(X_{n+1} - \mu_{n+1})^T - \theta_n \right)$$

Algorithmes MCMC en interaction de type *tempering*

Algorithmes basés sur du *tempering*:

- simuler des processus auxiliaires qui “tendent vers” la loi cible π^β ($\beta \in]0,1[$)
- les densités π^β sont moins “piquées” que π : les déplacements entre les différents modes sont facilités.
- construire le processus d'intérêt (de loi cible π) de façon à favoriser les déplacements locaux et autoriser quelques sauts en des points éloignés par *sélection* d'un point des processus auxiliaires.

Algorithmes MCMC en interaction de type *tempering*

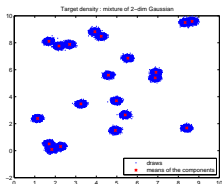
Algorithmes basés sur du *tempering*:

- simuler des processus auxiliaires qui “tendent vers” la loi cible π^β ($\beta \in]0,1[$)
- les densités π^β sont moins “piquées” que π : les déplacements entre les différents modes sont facilités.
- construire le processus d'intérêt (de loi cible π) de façon à favoriser les déplacements locaux et autoriser quelques sauts en des points éloignés par *sélection* d'un point des processus auxiliaires.

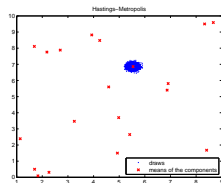
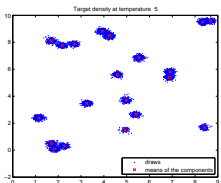
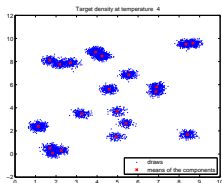
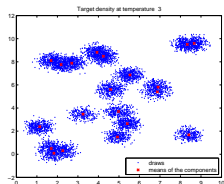
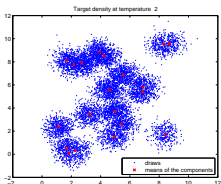
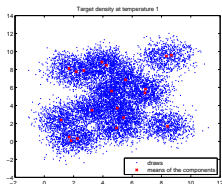
Par exemple: échantillonneur *Equi-Energy*, à l'itération n

$$X_{n+1} \sim \epsilon P(X_n, \cdot) + (1 - \epsilon) \left\{ \begin{array}{l} \text{sélection d'un point } Z \sim \theta_n = \frac{1}{n} \sum_{k=1}^n \delta_{Y_k} \\ \text{et correction par mécanisme acceptation-rejet} \end{array} \right.$$

$$\theta_{n+1} = \left(1 - \frac{1}{n+1} \right) \theta_n + \frac{1}{n+1} \delta_{Y_{n+1}}$$



- densité cible: $\pi = \sum_{i=1}^{20} \mathcal{N}_2(\mu_i, \Sigma_i)$
- K processus auxiliaires de loi cible π^{1/T_i}
 $T_1 > T_2 > \dots > T_{K+1} = 1$



Conclusion

- ④ Les algorithmes MCMC classiques (non adaptatifs) :

$$\mathbb{E}[\phi(X_{n+1})|\mathcal{F}_n] = \int P(X_n, dy)\phi(y)$$

avec $\pi P = \pi$.

Conclusion

- ① Les algorithmes MCMC classiques (non adaptatifs) :

$$\mathbb{E}[\phi(X_{n+1})|\mathcal{F}_n] = \int P(X_n, dy)\phi(y)$$

avec $\pi P = \pi$.

- ② Les algorithmes MCMC adaptatifs

$$\mathbb{E}[\phi(X_{n+1})|\mathcal{F}_n] = \int P_{\theta_n}(X_n, dy)\phi(y)$$

avec

- un processus d'adaptation θ_n tel que $\theta_{n+1} = \Xi_n(\theta_n, X_{n+1}, U)$
- une famille de noyaux de transition $\{P_\theta, \theta \in \Theta\}$
- avec $\pi_\theta P_\theta = \pi_\theta$ pour tout θ mais PAS nécessairement $\pi_\theta = \pi$.

↪ A quelles conditions le processus adapté $\{X_n, n \geq 0\}$ converge-t-il vers la loi cible π ?

III. Convergence des algorithmes MCMC adaptatifs

Résultats obtenus en collaboration avec E. MOULINES (TELECOM ParisTech)
et P. PRIOURET (Paris 6).

Par “convergence”, on entend

- convergence des marginales

$$\mathbb{E}[\phi(X_n)] \rightarrow \int \phi d\pi \quad \text{pour toute fonction bornée } \phi$$

- existence d'une loi des grands nombres pour une famille de fonctions (bornées ou pas)
- existence d'un théorème central limite.

Nous apportons des réponses dans le cas

- $\pi P_\theta = \pi$ pour tout θ
- $\pi_\theta P_\theta = \pi_\theta$ et $\pi_\theta \neq \pi$

sous forme de conditions **suffisantes** sur

- le noyau de transition P_θ
- et la façon dont est faite l'adaptation.

L'adaptation peut détruire la convergence

- Considérons la famille de noyaux de transition sur $\{0,1\}$:

$$P_\theta = \begin{pmatrix} 1 - \theta & \theta \\ \theta & 1 - \theta \end{pmatrix} \quad \theta \in (0,1)$$

Alors pour tout $\theta \in (0,1)$, $\pi P_\theta = \pi$ avec $\pi = (1/2; 1/2)$.

- Fixons $t_0, t_1 \in (0,1)$. On définit le processus adaptatif $\{X_n, n \geq 0\}$ par

$$\begin{cases} X_{n+1} \sim P_{\theta_n}(X_n, \cdot) \\ \theta_{n+1} = t_{X_{n+1}} \end{cases}$$

Alors le noyau de transition de ce processus adaptatif est

$$\begin{pmatrix} 1 - t_0 & t_0 \\ t_1 & 1 - t_1 \end{pmatrix}$$

qui a pour mesure invariante $\pi \propto (t_1, t_0)$.

Conditions pour la convergence des marginales

On écrit

$$\begin{aligned}\mathbb{E}[f(X_n)] - \pi(f) &= \mathbb{E}\left[f(X_n) - P_{\theta_{n-N}}^N f(X_{n-N})\right] \\ &\quad + \mathbb{E}\left[P_{\theta_{n-N}}^N f(X_{n-N}) - \pi(f)\right]\end{aligned}$$

Conditions pour la convergence des marginales

On écrit

$$\begin{aligned} \mathbb{E}[f(X_n)] - \pi(f) &= \mathbb{E}\left[f(X_n) - P_{\theta_{n-N}}^N f(X_{n-N})\right] \\ &\quad + \mathbb{E}\left[P_{\theta_{n-N}}^N f(X_{n-N}) - \pi(f)\right] \end{aligned}$$

► [1] condition d'ergodicité des noyaux de transition

“Habituellement”, les noyaux de transition $\{P_\theta, \theta \in \Theta\}$ sont géométriquement ergodiques :

$$\sup_{f, |f|_\infty \leq 1} |P_\theta^n f(x) - \pi(f)| \leq C_\theta \rho_\theta^n V(x) \quad \rho_\theta \in (0,1)$$

MAIS la vitesse de convergence peut dépendre de θ de telle sorte que

$$\rho_\theta \rightarrow 1 \quad \text{quand } \theta \rightarrow \partial\Theta .$$

Par suite, il est nécessaire de contrôler la vitesse à laquelle $\theta_n \rightarrow \partial\Theta$

Conditions pour la convergence des marginales

On écrit

$$\mathbb{E}[f(X_n)] - \pi(f) = \mathbb{E}\left[f(X_n) - P_{\theta_{n-N}}^N f(X_{n-N})\right] + \mathbb{E}\left[P_{\theta_{n-N}}^N f(X_{n-N}) - \pi(f)\right]$$

► [2] condition sur le mécanisme d'adaptation puisque

$$\left| \mathbb{E}\left[f(X_n) - P_{\theta_{n-N}}^N f(X_{n-N})\right] \right| \leq \sum_{j=1}^{N-1} (N-j) \mathbb{E} \left[\underbrace{\sup_x \left\| P_{\theta_{n-N+j}}(x, \cdot) - P_{\theta_{n-N+j-1}}(x, \cdot) \right\|_{\text{TV}}}_{\text{"distance" entre deux noyaux successifs}} \right]$$

Par la suite, l'adaptation doit être "décroissante".

Adaptation et Ergodicité

$$\mathbb{E}[f(X_n)] - \pi(f) = \mathbb{E}\left[f(X_n) - P_{\theta_{n-N}}^N f(X_{n-N})\right] + \mathbb{E}\left[P_{\theta_{n-N}}^N f(X_{n-N}) - \pi(f)\right]$$

► Exemple: $X_{n+1} \sim P_{\theta_n}(X_n, \cdot)$ $\theta_n = n^{-1/4}$ $P_\theta = \begin{pmatrix} 1 - \theta & \theta \\ \theta & 1 - \theta \end{pmatrix}$

Dans ce cas, puisque $\theta_n \rightarrow 0$

$$\left| \mathbb{E}\left[P_{\theta_{n-N}}^N f(X_{n-N}) - \pi(f)\right] \right| (\leq) = |1 - 2\theta_{n-N}|^N \rightarrow 1 \quad \text{pour } N \text{ fixé}$$

Adaptation et Ergodicité

$$\mathbb{E} [f(X_n)] - \pi(f) = \mathbb{E} \left[f(X_n) - P_{\theta_{n-N}}^N f(X_{n-N}) \right] + \mathbb{E} \left[P_{\theta_{n-N}}^N f(X_{n-N}) - \pi(f) \right]$$

► Exemple: $X_{n+1} \sim P_{\theta_n}(X_n, \cdot)$ $\theta_n = n^{-1/4}$ $P_{\theta} = \begin{pmatrix} 1 - \theta & \theta \\ \theta & 1 - \theta \end{pmatrix}$

Donc prendre N fonction de n

$$\left| \mathbb{E} \left[P_{\theta_{n-N_n}}^{N_n} f(X_{n-N_n}) - \pi(f) \right] \right| \leq |1 - 2\theta_{n-N_n}|^{N_n} \rightarrow 0$$

et l'adaptation doit être telle que

$$\sum_{j=1}^{N_n-1} (N_n - j) \mathbb{E} \left[\underbrace{\sup_x \left\| P_{\theta_{n-N_n+j}}(x, \cdot) - P_{\theta_{n-N_n+j-1}}(x, \cdot) \right\|_{\text{TV}}}_{\text{"distance" entre deux noyaux consécutifs}} \right] \rightarrow 0$$

↔ La "vitesse" d'adaptation dépend du comportement ergodique des noyaux

Résultat

Si

- $\exists \pi$ tq $\pi P_\theta = \pi$ pour tout $\theta \in \Theta$.
- (**Ergodicité**) pour tout $\epsilon > 0$, il existe une suite positive croissante $\{r_\epsilon(n), n \geq 0\}$ tq $\limsup_{n \rightarrow \infty} r_\epsilon(n)/n = 0$ et

$$\limsup_{n \rightarrow \infty} \mathbb{E} \left[\left\| P_{\theta_{n-r_\epsilon(n)}}^{r_\epsilon(n)}(X_{n-r_\epsilon(n)}, \cdot) - \pi \right\|_{\text{TV}} \right] \leq \epsilon .$$

- (**Adaptation décroissante**) Pour tout $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \sum_{j=0}^{r_\epsilon(n)-1} \mathbb{E} \left[\sup_x \left\| P_{\theta_{n-r_\epsilon(n)+j}}(x, \cdot) - P_{\theta_{n-r_\epsilon(n)}}(x, \cdot) \right\|_{\text{TV}} \right] = 0$$

Alors

$$\sup_{\{\phi, \|\phi\|_\infty \leq 1\}} |\mathbb{E}[\phi(X_n)] - \int \phi d\pi| \rightarrow 0 .$$

En pratique, comment vérifier les conditions?

Dans un contexte MCMC et plus précisément Hastings-Metropolis :

- 1 la condition d'ergodicité qui nécessite le contrôle de $\|P_\theta^N(x, \cdot) - \pi\|_{\text{TV}}$ s'obtient
 - en vérifiant une condition de **drift** : $P_\theta V \leq \lambda_\theta V + b_\theta \mathbb{1}_{\mathcal{C}_\theta}$
 - en vérifiant une condition de **minoration** : $P_\theta(x, \cdot) \geq \epsilon_\theta \mathbb{1}_{\mathcal{C}_\theta}(x) \nu_\theta(\cdot)$
 - et en utilisant des résultats sur **le calcul explicite de bornes d'ergodicité** à partir des quantités $b_\theta, \lambda_\theta, \epsilon_\theta, \mathcal{C}_\theta$.

En pratique, comment vérifier les conditions?

Dans un contexte MCMC et plus précisément Hastings-Metropolis :

- 1 la condition d'ergodicité qui nécessite le contrôle de $\|P_\theta^N(x, \cdot) - \pi\|_{\text{TV}}$ s'obtient
 - en vérifiant une condition de **drift** : $P_\theta V \leq \lambda_\theta V + b_\theta \mathbb{1}_{\mathcal{C}_\theta}$
 - en vérifiant une condition de **minoration** : $P_\theta(x, \cdot) \geq \epsilon_\theta \mathbb{1}_{\mathcal{C}_\theta}(x) \nu_\theta(\cdot)$
 - et en utilisant des résultats sur **le calcul explicite de bornes d'ergodicité** à partir des quantités $b_\theta, \lambda_\theta, \epsilon_\theta, \mathcal{C}_\theta$.
- 2 la condition d'adaptation décroissante s'obtient en établissant une inégalité de la forme

$$\|P_\theta(x, \cdot) - P_{\theta'}(x, \cdot)\|_{\text{TV}} \leq C d(\theta, \theta') .$$

Doit-on avoir convergence du processus d'adaptation ?

Selon la situation

- 1 **Non** : dans la situation que l'on vient d'étudier, la seule condition qui apparaît porte sur
 - $d(\theta_n, \theta_{n-1}) \rightarrow 0$
 - le contrôle de la vitesse à laquelle θ_n s'approche du bord de l'ensemble $\partial\Theta$.

Mais c'est une situation particulière dans laquelle les noyaux P_θ ont TOUS même mesure invariante.

Doit-on avoir convergence du processus d'adaptation ?

Selon la situation

- ① **Non** : dans la situation que l'on vient d'étudier, la seule condition qui apparaît porte sur
- $d(\theta_n, \theta_{n-1}) \rightarrow 0$
 - le contrôle de la vitesse à laquelle θ_n s'approche du bord de l'ensemble $\partial\Theta$.

Mais c'est une situation particulière dans laquelle les noyaux P_θ ont TOUS même mesure invariante.

- ② **Oui** : dans la situation où chaque noyau P_θ a sa propre mesure invariante π_θ . Dans ce cas, il faut des conditions pour que

$$\pi_{\theta_n}(\phi) \xrightarrow{\text{p.s.}} \pi(\phi) \quad \text{par ex. pour l'obtention d'une LGN}$$

et ces conditions implicitement demandent la convergence de θ_n vers θ_* tel que $\pi_{\theta_*} = \pi$. En pratique, on n'a pas d'expression explicite de π_θ et cela s'établit à l'aide de conditions sur P_θ .

Conclusion

- ▶ Les procédures MCMC adaptatives sont basées sur la construction conjointe d'un processus d'adaptation et du processus d'intérêt.
- ▶ Les procédures d'adaptation peuvent détruire la convergence de l'algorithme vers la loi cible.

Conclusion

- ▶ Les procédures MCMC adaptatives sont basées sur la construction conjointe d'un processus d'adaptation et du processus d'intérêt.
- ▶ Les procédures d'adaptation peuvent détruire la convergence de l'algorithme vers la loi cible.
- ▶ On peut néanmoins montrer la convergence de procédures au sens de adaptatives
 - convergence des marginales
 - loi des grands nombres
 - théorème central limite

dès lors que celles-ci vérifient des conditions

- (i) d'ergodicité de chaque noyau de transition
- (ii) d'adaptation "décroissante"
- (iii) et dans certains cas - de convergence des noyaux P_{θ_n} vers un noyau P_{θ_*} .

Méthodes MCMC adaptatives : méthodologie

- C. Andrieu, J. Thoms. *Statistics and Computing*, 2008.
- J.S. Rosenthal. *MCMC handbook*, 2009.
- Y. Atchadé, G. Fort, E. Moulines, P. Priouret. *Times Series book*, 2010.

Méthodes MCMC adaptatives : résultats de convergence

- C. Andrieu, E. Moulines. *Ann. Appl. Probab.*, 2006.
- G.O. Roberts, J.S. Rosenthal. *J. Appl. Probab.* 2007.
- Y. Atchadé, G. Fort. *Stoch. Processes Appl.*, 2010.
- G. Fort, E. Moulines, P. Priouret. *Convergence of adaptive MCMC algorithms: ergodicity and law of large numbers*. Preprint, 2010.
- C. Andrieu, A. Jasra, A. Doucet, P. DelMoral *Preprint 2007*
- Y. Atchadé *Statistica Sinica*, 2010
- E. Saksman, M. Vihola, *Ann. Appl. Probab.* 2010
- M. Vihola *Preprint*, 2010