Méthodes de Monte Carlo
et
Chaînes de Markov pour la simulation

Gersende FORT

LTCI, CNRS / TELECOM ParisTech

Soit une loi cible π sur X. On souhaite

- calculer des intégrales de la forme $\quad \mathbb{E}\left[\phi(X)\right] \qquad X \sim \pi$.
- explorer le support de la loi ex. pour la résolution de problèmes inverses bayésiens : $\pi \equiv$ loi a posteriori

Soit une loi cible π sur X. On souhaite

- calculer des intégrales de la forme $\quad \mathbb{E}\left[\phi(X)\right] \qquad X \sim \pi.$
- explorer le support de la loi ex. pour la résolution de problèmes inverses bayésiens : π $\equiv$ loi a posteriori

Comment obtenir des réalisations de v.a. $\{X_k, k \geq 1\}$ qui approchent π ?

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

Soit une loi cible π sur X. On souhaite

- calculer des intégrales de la forme $\mathbb{E}\left[\phi(X)\right] \qquad X \sim \pi.$
- explorer le support de la loi ex. pour la résolution de problèmes inverses bayésiens : $\pi \equiv$ loi a posteriori

Comment obtenir des réalisations de v.a. $\{X_k, k \geq 1\}$ qui approchent π ?

1. Méthodes Monte Carlo par Chaînes de Markov (MCMC)
   - produire une chaîne de Markov $\{X_k, k \geq 1\}$, stationnaire, de loi limite π.
   - qui vérifie des théorèmes limites (ergodicité; loi des grands nombres; TCL; $\cdots$)

   $\hookrightarrow$ quel noyau de transition ? rôle du noyau de transition sur l'efficacité de l'échantillonneur ?

Soit une loi cible π sur X. On souhaite

- calculer des intégrales de la forme $\mathbb{E}\left[\phi(X)\right] \qquad X \sim \pi.$
- explorer le support de la loi ex. pour la résolution de problèmes inverses bayésiens : $\pi \equiv$ loi a posteriori

Comment obtenir des réalisations de v.a. $\{X_k, k \geq 1\}$ qui approchent π ?

1. Méthodes Monte Carlo par Chaînes de Markov (MCMC)
2. Méthodes d'échantillonnage d'importance (IS)

Loi π de densité π w.r.t. Leb sur X $\subseteq \mathbb{R}^d$

$$\int \phi(x)\pi(x)dx = \int \phi(x)\frac{\pi(x)}{q(x)}\, q(x)dx \approx \frac{1}{n}\sum_{k=1}^{n}\frac{\pi(X_k)}{q(X_k)}\, \phi(X_k)$$

où $\quad X_k \overset{i.i.d.}{\sim} q \quad .$

Soit une loi cible π sur X. On souhaite

- calculer des intégrales de la forme $\mathbb{E}\left[\phi(X)\right] \qquad X \sim \pi$.
- explorer le support de la loi ex. pour la résolution de problèmes inverses bayésiens : $\pi \equiv$ loi a posteriori

Comment obtenir des réalisations de v.a. $\{X_k, k \geq 1\}$ qui approchent π ?

1. Méthodes Monte Carlo par Chaînes de Markov (MCMC)
2. Méthodes d'échantillonnage d'importance (IS)

Loi π de densité π w.r.t. Leb sur X $\subseteq \mathbb{R}^d$

$$\int \phi(x)\pi(x)dx = \int \phi(x)\frac{\pi(x)}{q(x)}\, q(x)dx \approx \frac{1}{n}\sum_{k=1}^{n}\frac{\pi(X_k)}{q(X_k)}\, \phi(X_k)$$

où $\quad X_k \overset{i.i.d.}{\sim} q \quad$.

$\hookrightarrow$ quelle loi instrumentale q ? rôle de la loi instrumentale sur l'efficacité de l'échantillonneur ?

Soit une loi cible π sur X. On souhaite

- calculer des intégrales de la forme $\mathbb{E}\left[\phi(X)\right] \qquad X \sim \pi.$
- explorer le support de la loi ex. pour la résolution de problèmes inverses bayésiens : $\pi \equiv$ loi a posteriori

Comment obtenir des réalisations de v.a. $\{X_k, k \geq 1\}$ qui approchent π ?

1. Méthodes Monte Carlo par Chaînes de Markov (MCMC)
2. Méthodes d'échantillonnage d'importance (IS)

Soit une loi cible π sur X. On souhaite

- calculer des intégrales de la forme $\quad \mathbb{E}\left[\phi(X)\right] \qquad X \sim \pi.$
- explorer le support de la loi ex. pour la résolution de problèmes inverses bayésiens : $\pi \equiv$ loi a posteriori

Comment obtenir des réalisations de v.a. $\{X_k, k \geq 1\}$ qui approchent π ?

1. Méthodes Monte Carlo par Chaînes de Markov (MCMC)
2. Méthodes d'échantillonnage d'importance (IS)

Axes de recherche

1. Identifier les paramètres d'implémentation
2. Critère d'efficacité des échantillonneurs $\longrightarrow$ paramètre optimal
3. procédures adaptatives pour guider le paramètre vers le paramètre optimal

Application I : Modèles à données cachées

- Modèle :
  - Etats cachés de loi jointe : $p(x_{1:n})$
  - Loi des observations conditionnellement aux états cachés : $g(y_{1:n}|x_{1:n})$

Application I : Modèles à données cachées

- Modèle :
  - Etats cachés de loi jointe : $p(x_{1:n}\ ;\theta)$
  - Loi des observations conditionnellement aux états cachés : $g(y_{1:n}|x_{1:n}\ ;\theta)$
  - Paramètres: θ

Application I : Modèles à données cachées

- Modèle :
  - Etats cachés de loi jointe : $p(x_{1:n}\ ;\theta)$
  - Loi des observations conditionnellement aux états cachés : $g(y_{1:n}|x_{1:n}\ ;\theta)$
  - Paramètres: θ
- Objectif :
  - estimation du paramètre θ
  - restauration des données cachées $x_{1:n}$

Application I : Modèles à données cachées

- Modèle :
  - Etats cachés de loi jointe : $p(x_{1:n}\ ;\theta)$
  - Loi des observations conditionnellement aux états cachés : $g(y_{1:n}|x_{1:n}\ ;\theta)$
  - Paramètres: θ
- Objectif :
  - estimation du paramètre θ
  - restauration des données cachées $x_{1:n}$
- Méthode : algorithme itératif *Expectation Maximization* (EM)
  - estimation du paramètre à chaque étape M.
  - restauration (possible) des données cachées à chaque étape E à partir de la loi a posteriori $p(x_{1:n}|y_{1:n};\theta^{(t)})$.

Application I : Modèles à données cachées

- Modèle :
  - Etats cachés de loi jointe : $p(x_{1:n}\ ;\theta)$
  - Loi des observations conditionnellement aux états cachés : $g(y_{1:n}|x_{1:n}\ ;\theta)$
  - Paramètres: θ
- Objectif :
  - estimation du paramètre θ
  - restauration des données cachées $x_{1:n}$
- Méthode : algorithme itératif *Expectation Maximization* (EM)
  - estimation du paramètre à chaque étape M.
  - restauration (possible) des données cachées à chaque étape E à partir de la loi a posteriori $p(x_{1:n}|y_{1:n};\theta^{(t)})$.

$\hookrightarrow$ L'EM ne s'applique pas lorsque la structure de dépendance entre les variables est trop complexe

- Ex. Segmentation d'images

Image vraie

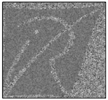


Image observée

- Données cachées : classes $\{1, \cdots, K\}$
- Attache aux données : $\prod_{k=1}^{n} \mathcal{N}(\mu_{x_k}, \sigma_{x_k}^2)[y_k]$
- Structure de dépendance sur les données cachées

- Ex. Segmentation d'images

Image vraie

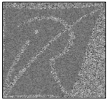


Image observée

Modèle indépendant

- Données cachées : classes $\{1, \cdots, K\}$
- Attache aux données : $\prod_{k=1}^{n} \mathcal{N}(\mu_{x_k}, \sigma^2_{x_k})[y_k]$
- Structure de dépendance sur les données cachées
  (i) i.i.d. : l'EM s'applique.

Ex. Segmentation d'images

Image vraie

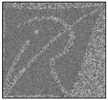


Image observée

Modèle indépendant

- Données cachées : classes $\{1, \cdots, K\}$
- Attache aux données : $\prod_{k=1}^{n} \mathcal{N}(\mu_{x_k}, \sigma^2_{x_k})[y_k]$
- Structure de dépendance sur les données cachées
  - (i) i.i.d. : l'EM s'applique.
  - (ii) champs de Markov : l'EM ne s'applique pas car
    - E : domaine d'intégration trop grand (taille K^n)
    - M : fonction non explicite (constante de normalisation)

▶ Ex. Segmentation d'images

Image vraie

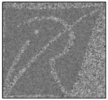


Image observée

Modèle indépendant

Modèle Markov

Résultats : (GF & F. Forbes, 2007); (GF & E. Moulines, 2003)

- algorithme qui combine EM de type stochastique et approche variationnelle.

$$q^{(t+1)} \in \mathrm{argmax}_{q \in \mathcal{I}} \sum_{x_{1:n} \in \mathsf{X}} \ln \left(\frac{p(y_{1:n}, x_{1:n}; \theta)}{q(x)} \right) q(x_{1:n})$$

$$\theta^{(t+1)} \in \mathrm{argmax}_{\theta} \sum_{x_{1:n} \in \mathsf{X}} \widehat{\ln p}(y_{1:n}, x_{1:n}; \theta)\, q^{(t+1)}(x_{1:n})$$

▶ Ex. Segmentation d'images


Image vraie

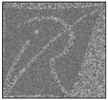


Image observée

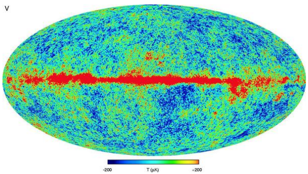

Modèle indépendant


Modèle Markov

Résultats : (GF & F. Forbes, 2007); (GF & E. Moulines, 2003)

- algorithme qui combine EM de type stochastique et approche variationnelle.
- caractérisation du comportement asymptotique (nbr itérations $\to +\infty$) de cet algorithme.
- *Contrib. MC* : (a) Contrôle de l'erreur d'approximation MC ; (b) Convergence d'algorithmes itératifs perturbés.

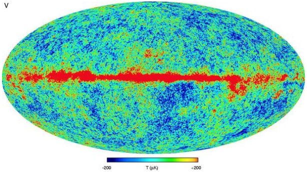
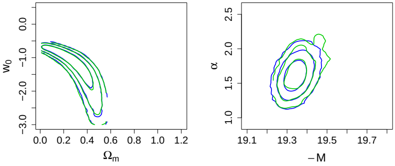
Application II : Inférence bayésienne



Anisotropie du CMB

- Modèle :
  - vraisemblance $p(y_{1:n}|\theta)$, très coûteuse en temps de calcul
  - un prior $p(\theta)$
- Objectif :
  - explorer la distribution a posteriori de θ Ω_m masse totale sous forme de matière noire; ω_0 paramètre décrivant l'énergie du vide; A_t amplitude primordiale des fluctuations tensorielles; $\cdots$

Application II : Inférence bayésienne

Anisotropie du CMB Lois marginales 2D

- Modèle :
  - vraisemblance $p(y_{1:n}|\theta)$, très coûteuse en temps de calcul
  - un prior $p(\theta)$
- Objectif :
  - explorer la distribution a posteriori de θ Ω_m masse totale sous forme de matière noire; ω_0 paramètre décrivant l'énergie du vide; A_t amplitude primordiale des fluctuations tensorielles; · · ·
- Originalité :
  - évaluation coûteuse de la loi cible $\pi(\theta) \propto p(y_{1:n}|\theta)\ p(\theta)$ (via "boîte noire")
  - structure de dépendance entre les composantes de θ

- Ex. Astro-statistique (D. Wraith, · · · GF et al., 2009)
  - Implémentation d'un algorithme d'IS adaptatif
  - Comparaison à un algorithme MCMC (Hastings-Metropolis à marche aléatoire symétrique) adaptatif

- Ex. Astro-statistique (D. Wraith, · · · GF et al., 2009)
  - Implémentation d'un algorithme d'IS adaptatif

$$\pi \approx \left(\sum_{j=1}^{n} \frac{\pi(X_j)}{q_{\theta_\star}(X_j)} \right)^{-1} \sum_{k=1}^{n} \frac{\pi(X_k)}{q_{\theta_\star}(X_k)} \; \delta_{X_k} \qquad\qquad X_k \overset{i.i.d}{\sim} q_{\theta_\star}$$

où $\theta_\star$ est solution de

$$q_{\theta_\star} \overset{\text{def}}{=} \text{argmin}_{q_\theta \in \mathcal{F}} \int \ln \frac{\pi(x)}{q_\theta(x)} \; \pi(x) \; dx$$

Lorsque $\mathcal{F} = \{$mélanges de gaussiennes (ou Student)$\}$, résolution du critère par un algorithme EM de type stochastique
Jeu de données simulées ▸ Voir

- Ex. Astro-statistique (D. Wraith, · · · GF et al., 2009)
  - Implémentation d'un algorithme d'IS adaptatif
  - Comparaison à un algorithme MCMC (Hastings-Metropolis à marche aléatoire symétrique) adaptatif

MCMC de noyau de transition

$$P_{\theta_\star}(x,dy) = \alpha(x,y)\ q_{\theta_\star}(x,y)dy + \delta_x(dy) \int (1-\alpha(x,z))\, q_{\theta_\star}(x,z)dz$$

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

La loi de proposition $q_\theta(x,\cdot) \equiv \mathcal{N}(x,\theta)$.
Il faut (Roberts,Gelman, Gilks (1997)) $\theta_\star \propto \Sigma_\pi$ inconnue en gal $\Rightarrow$ méthode itérative pour l'estimation de Σ_π

Comparaison sur données simulées ▸ Voir

Comparaison sur données réelles ▸ Voir

II. Théorie des Chaînes de Markov

Contributions :

1. Caractérisation de l'ergodicité des chaînes de Markov
2. Contrôle explicite de l'ergodicité
3. Applications à la convergence d'échantillonneurs MCMC

Ergodicité

Etant donnée une chaîne $\{X_k, k \geq 0\}$ de noyau de transition P sur $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$, à quelles conditions

- possède-t-elle une unique mesure de probabilité invariante π ?
- la loi de la chaîne converge-t-elle vers π ?
  - (ergodicité)

$$\lim_n \sup_{\{f, \sup_{\mathsf{X}} |f| \leq 1\}} |\mathbb{E}_x[f(X_n)] - \pi(f)| = 0$$

Ergodicité

Etant donnée une chaîne $\{X_k, k \geq 0\}$ de noyau de transition P sur $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$, à quelles conditions

- possède-t-elle une unique mesure de probabilité invariante π ?
- la loi de la chaîne converge-t-elle vers π ?
  - (ergodicité)

$$\lim_n \sup_{\{f, \sup_{\mathsf{X}} |f| \leq 1\}} |\mathbb{E}_x[f(X_n)] - \pi(f)| = 0$$

  - (ergodicité à la vitesse r)

$$\lim_n r(n) \sup_{\{f, \sup_{\mathsf{X}} |f| \leq 1\}} |\mathbb{E}_x[f(X_n)] - \pi(f)| = 0 \qquad \lim_{+\infty} r = +\infty$$

Ergodicité

Etant donnée une chaîne $\{X_k, k \geq 0\}$ de noyau de transition P sur $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$, à quelles conditions

- possède-t-elle une unique mesure de probabilité invariante π ?
- la loi de la chaîne converge-t-elle vers π ?
  - (ergodicité)

$$\lim_n \sup_{\{f, \sup_{\mathsf{X}} |f| \leq 1\}} |\mathbb{E}_x[f(X_n)] - \pi(f)| = 0$$

  - (ergodicité à la vitesse r)

$$\lim_n r(n) \sup_{\{f, \sup_{\mathsf{X}} |f| \leq 1\}} |\mathbb{E}_x[f(X_n)] - \pi(f)| = 0 \qquad \lim_{+\infty} r = +\infty$$

  - (V-ergodicité à la vitesse r)

$$\lim_n r(n) \sup_{\{f, \sup_{\mathsf{X}} |f| \leq V\}} |\mathbb{E}_x[f(X_n)] - \pi(f)| = 0 \qquad V : \mathsf{X} \to [1, +\infty)$$

Caractérisation de l'ergodicité

par une condition de drift (+ irréductibilité, apériodicité)

$$\mathbb{E}_x[V(X_1)] \leq V(x) - \underbrace{W(x)}_{\text{quantité positive}} + b\mathbb{1}_{\mathsf{C}}(x)$$

Caractérisation de l'ergodicité

par une condition de drift (+ irréductibilité, apériodicité)

$$\mathbb{E}_x[V(X_1)] \leq V(x) - \underbrace{W(x)}_{\text{quantité positive}} + b\mathbb{1}_{\mathsf{C}}(x)$$

- ergodicité à une vitesse polynomiale $r(n) \propto n^p, \quad p > 0$ (GF & E. Moulines, 2000); (GF & E. Moulines, 2003)
- ergodicité à une vitesse sous-géométrique : (GF et al. 2004)

$$\mathbb{E}_x[V(X_1)] \leq V(x) - \phi \circ V(x) + b\mathbb{1}_{\mathsf{C}}(x) \qquad \phi \text{ concave } \uparrow$$

Caractérisation de l'ergodicité

par une condition de drift (+ irréductibilité, apériodicité)

$$\mathbb{E}_x[V(X_1)] \leq V(x) - \underbrace{W(x)}_{\text{quantité positive}} + b\mathbb{1}_{\mathsf{C}}(x)$$

- ergodicité à une vitesse polynomiale $\quad r(n) \propto n^p, \quad p > 0$ (GF & E. Moulines, 2000); (GF & E. Moulines, 2003)
- ergodicité à une vitesse sous-géométrique : (GF et al. 2004)

$$\mathbb{E}_x[V(X_1)] \leq V(x) - \phi \circ V(x) + b\mathbb{1}_{\mathsf{C}}(x) \qquad \phi \text{ concave } \uparrow$$

- en norme variation totale, vitesse de convergence :

logarithmique	$\phi(t) \sim c(1+\ln t)^\alpha$		$r(t) \sim \log^\alpha(t)$
polynomiale	$\phi(t) \sim ct^{1-\alpha}$	$\alpha \in (0,1]$	$r(t) \sim t^{1/\alpha - 1}$
sous-expo	$\phi(t) \sim ct/[\ln t]^\alpha$	$\alpha > 0$	$\ln r(t) \sim t^{1/(1+\alpha)}$

Caractérisation de l'ergodicité

par une condition de drift (+ irréductibilité, apériodicité)

$$\mathbb{E}_x[V(X_1)] \leq V(x) - \underbrace{W(x)}_{\text{quantité positive}} + b\mathbb{1}_C(x)$$

- ergodicité à une vitesse polynomiale $r(n) \propto n^p, \quad p > 0$ (GF & E. Moulines, 2000); (GF & E. Moulines, 2003)
- ergodicité à une vitesse sous-géométrique : (GF et al. 2004)

$$\mathbb{E}_x[V(X_1)] \leq V(x) - \phi \circ V(x) + b\mathbb{1}_C(x) \qquad \phi \text{ concave } \uparrow$$

  - en norme variation totale, vitesse de convergence :

logarithmique	$\phi(t) \sim c(1 + \ln t)^\alpha$		$r(t) \sim \log^\alpha(t)$
polynomiale	$\phi(t) \sim ct^{1-\alpha}$	$\alpha \in (0,1]$	$r(t) \sim t^{1/\alpha - 1}$
sous-expo	$\phi(t) \sim ct/[\ln t]^\alpha$	$\alpha > 0$	$\ln r(t) \sim t^{1/(1+\alpha)}$

  - en norme $\phi \circ V$, vitesse minimale 1.
  - normes et vitesses intermédiaires, par interpolation.

Extensions

1. Condition de drift sur les itérés du noyau P, itérés dépendant de l'état courant, de la forme (GF & S. Connor, 2009)

$$\mathbb{E}_x\left[V(X_{n(X_0)})\right] \leq \lambda V(x) + b\mathbb{1}_{\mathsf{C}}(x) \qquad \lambda \in (0,1)$$

Extensions

1. Condition de drift sur les itérés du noyau P, itérés dépendant de l'état courant, de la forme (GF & S. Connor, 2009)

$$\mathbb{E}_x \left[V(X_{n(X_0)}) \right] \leq \lambda V(x) + b \mathbb{1}_C(x) \qquad \lambda \in (0,1)$$

2. Caractérisation de l'ergodicité de processus de Markov à temps continu : à l'aide d'une condition de drift sur le générateur $\mathcal{A}$

$$\mathcal{A}V(x) \leq -\phi \circ V(x) + b \mathbb{1}_C(x)$$

(GF & G.O. Roberts, 2005), (GF et al., 2009)

Contrôle explicite de l'ergodicité

Construire une borne $B(x)$ telle que

$$r(n) \sup_{\{f, \sup_{\mathsf{X}} |f|/W\}} |\mathbb{E}_x [f(X_n)] - \pi(f)| \leq B(x).$$

- Outils :
  - condition de minoration sur le noyau de transition
    $$P(x,A) \geq \epsilon \ \mathbb{1}_{\mathsf{C}}(x) \ \nu(A)$$
  - condition de drift sur le noyau de transition
    $$\mathbb{E}_x [V(X_1)] \leq V(x) - W(x) + b\mathbb{1}_{\mathsf{C}}(x)$$
  - irréductibilité, apériodicité
- Résultats : une expression de $B(x)$ en fonction de $\epsilon, \sup_C V, b, \nu(V)$ et $V(x)$ pour ergodicité géom et sous-géom (GF & E. Moulines, 2003), (GF, 2004), (GF & C. Andrieu, 2005)
- Applications : $\cdots$

III. Optimisation des méthodes de simulation

Contributions :

1. Identification du rôle des paramètres d'implémentation de certains échantillonneurs MCMC par la méthode de l'ODE
2. Convergence d'algorithmes de simulation adaptatifs à dynamique markovienne
3. Méthode de stratification adaptative

Exemple d'algorithmes de simulation adaptatifs à dynamique markovienne (I)

Soit une chaîne de Hastings-Metropolis à marche aléatoire symétrique sur $\mathbb{R}^d$, $\{X_k, k \geq 0\}$ de noyau de transition P_θ

$$P_\theta(x,A) = \int_A \alpha(x,y) q_\theta(x,y) dy + \mathbb{1}_A(x) \int (1 - \alpha(x,z)) \, q_\theta(x,z) dz$$

où α est le ratio d'acceptation-rejet $\quad \alpha(x,y) \stackrel{\text{def}}{=} 1 \wedge [\pi(y)/\pi(x)] \quad$.

Considérons le cas où $q_\theta(x,\cdot) \sim \mathcal{N}_d(x,\theta)$. Quel choix pour θ ? ▸ Voir

Exemple d'algorithmes de simulation adaptatifs à dynamique markovienne (I)

Soit une chaîne de Hastings-Metropolis à marche aléatoire symétrique sur $\mathbb{R}^d$, $\{X_k, k \geq 0\}$

Considérons le cas où $q_\theta(x, \cdot) \sim \mathcal{N}_d(x, \theta)$. Quel choix pour θ?

- Algorithme adaptatif (Haario et al. 1999)

$$\mu_n = \mu_{n-1} + \frac{1}{n}\left(X_n - \mu_{n-1}\right)$$

$$\Sigma_n = \Sigma_{n-1} + \frac{1}{n+1}\left\{\left(X_n - \mu_{n-1}\right)\left(X_n - \mu_{n-1}\right)^T - \Sigma_{n-1}\right\}$$

et simuler $\quad X_{n+1} \sim P_{\theta_n}(X_n, \cdot) \quad$ où $\theta_n \propto \Sigma_n$. Voir

Exemple d'algorithmes de simulation adaptatifs à dynamique markovienne (II)

- Algorithme Equi-Energy (Kou et al., 2006)
  - Etant donné
    - un processus auxiliaire $\{Y_n, n \geq 0\}$ tel $\theta_n \stackrel{\text{def}}{=} \frac{1}{n+1} \sum_{k=0}^{n} \delta_{Y_k}$ "tend vers $\tilde{\pi}$"
    - un noyau de transition P tel que $\pi P = \pi$
    - une probabilité d'échange ϵ
  - on construit le processus d'intérêt $\{X_n, n \geq 0\}$ selon la transition

$$P_{\theta_n}(x, A) = (1-\epsilon) P(x, A) + \epsilon \left[\int_A \alpha(x,y)\ \theta_n(dy) + \mathbb{1}_A(x) \int (1 - \alpha(x,y))\ \theta_n(dy) \right]$$

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

Exemple d'algorithmes de simulation adaptatifs à dynamique markovienne (II)

- Algorithme Equi-Energy (Kou et al., 2006)
  - Etant donné
    - un processus auxiliaire $\{Y_n, n \geq 0\}$ tel $\theta_n \stackrel{\text{def}}{=} \frac{1}{n+1} \sum_{k=0}^{n} \delta_{Y_k}$ "tend vers $\tilde{\pi}$"
    - un noyau de transition P tel que $\pi P = \pi$
    - une probabilité d'échange ϵ
  - on construit le processus d'intérêt $\{X_n, n \geq 0\}$ selon la transition

$$P_{\theta_n}(x, A) = (1-\epsilon) P(x, A) + \epsilon \left[\int_A \alpha(x, y)\ \theta_n(dy) + \mathbb{1}_A(x) \int (1 - \alpha(x, y))\ \theta_n(dy) \right]$$

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

  - version simplifiée $\cdots$

Voir

Deux exemples tels que

- construction d'un processus $\{(X_n,\theta_n),n \geq 0\}$
- algorithme à dynamique markovienne

$$\mathbb{P}\left(X_{n+1} \in A|\mathcal{F}_n\right) = P_{\theta_n}(X_n,A).$$

- algorithme à adaptation
  - interne : (exemple 1) θ_n est construit à l'aide du processus d'intérêt.
  - externe : (exemple 2) interactions entre processus.

Deux exemples tels que

- construction d'un processus $\{(X_n,\theta_n),n \geq 0\}$
- algorithme à dynamique markovienne

$$\mathbb{P}\left(X_{n+1} \in A|\mathcal{F}_n\right) = P_{\theta_n}(X_n,A).$$

- algorithme à adaptation
  - interne : (exemple 1) θ_n est construit à l'aide du processus d'intérêt.
  - externe : (exemple 2) interactions entre processus.

Etant donnée une famille de noyaux $\{P_\theta,\theta \in \Theta\}$ et un processus $\{(X_n,\theta_n),n \geq 0\}$ tel que

$$\mathbb{P}\left(X_{n+1} \in A|\mathcal{F}_n\right) = P_{\theta_n}(X_n,A)$$

A quelles conditions la loi de $\{X_n,n \geq 0\}$ converge-t-elle ?

Ergodicité

L'adaptation peut détruire la convergence vers π.

Exemple :

- Une famille de noyaux P_θ sur $\{0,1\}$ tels que pour tout θ, $\pi P_\theta = \pi$.

$$P_\theta = \begin{bmatrix} 1-\theta & \theta \\ \theta & 1-\theta \end{bmatrix} \qquad \theta \in]0,1[; \qquad \pi \propto (1,1)$$

- On construit un processus $(X_n,\theta_n) \in \{0,1\} \times]0,1[$: soient $t_1 \neq t_2$, on pose

$$\theta_n = t_1 \quad \text{ssi } X_n = 1 \qquad \theta_n = t_2 \quad \text{ssi } X_n = 0$$

- Alors $\{X_n, n \geq 0\}$ est une chaîne de noyau $\tilde{P}$ donné par

$$\tilde{P} = \begin{bmatrix} 1-t_1 & t_1 \\ t_2 & 1-t_2 \end{bmatrix}$$

et $\pi \tilde{P} \neq \pi$.

Ergodicité

L'adaptation peut détruire la convergence vers π.

Théorème (GF et al., 2010)

Si

- il existe une unique proba π_θ telle que $\pi_\theta P_\theta = \pi_\theta$
- adaptation décroissante :

$$\sup_{x \in \mathsf{X}} \|P_{\theta_n}(x,\cdot) - P_{\theta_{n-1}}(x,\cdot)\|_{\mathrm{VT}} \xrightarrow{\mathbb{P}} 0.$$

- comportement ergodique analogue, des noyaux P_θ :

$$\forall \epsilon > 0 \qquad \inf_k \{k \geq 1, \quad \|P^k_{\theta_n}(X_n,\cdot) - \pi_{\theta_n}\|_{\mathrm{VT}} \leq \epsilon\} \text{ borné en probabilité.}$$

- $\pi_{\theta_n}(f) \to \pi(f)$ en probabilité, pour une fonction f bornée.

Alors

$$\lim_n \mathbb{E}_x\left[f(X_n)\right] = \pi(f).$$

Loi des grands nombres

Théorème (GF et al., 2010)

Soit $f : \mathsf{X} \to \mathbb{R}$. Si

- $\exists!\ \pi_\theta$ telle que $\pi_\theta P_\theta = \pi_\theta$,
- $\pi_{\theta_n}(f) \to \pi(f)$ p.s.

Alors

$$\lim_n \frac{1}{n} \sum_{k=1}^{n} f(X_k) \xrightarrow{p.s.} \pi(f).$$

Idée de la demo On écrit

$$\frac{1}{n} \sum_{k=1}^{n} f(X_k) - \pi(f) = \frac{1}{n} \sum_{k=1}^{n} \left\{ f(X_k) - \pi_{\theta_{k-1}}(f) \right\} + \frac{1}{n} \sum_{k=1}^{n} \pi_{\theta_{k-1}}(f) - \pi(f)$$

Loi des grands nombres

Théorème (GF et al., 2010)

Soit $f : \mathsf{X} \to \mathbb{R}$. Si

- $\exists!$ π_θ telle que $\pi_\theta P_\theta = \pi_\theta$, et une solution $\hat{f}_\theta$
- Convergence p.s. du terme martingale
- Convergence p.s. des restes
- $\pi_{\theta_n}(f) \to \pi(f)$ p.s.

Alors

$$\lim_n \frac{1}{n} \sum_{k=1}^{n} f(X_k) \xrightarrow{p.s.} \pi(f).$$

Idée de la demo On écrit

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) - \pi(f) = \frac{1}{n}\sum_{k=1}^{n} \left\{ f(X_k) - \pi_{\theta_{k-1}}(f) \right\} + \frac{1}{n}\sum_{k=1}^{n} \pi_{\theta_{k-1}}(f) - \pi(f)$$

puis en introduisant $\hat{f}_\theta$ solution de $g - P_\theta g = f - \pi_\theta(f)$

$$\frac{1}{n}\sum_{k=1}^{n} \left\{ f(X_k) - \pi_{\theta_{k-1}}(f) \right\} = \frac{1}{n}\sum_{k=1}^{n} \left\{ \hat{f}_{\theta_{k-1}}(X_k) - P_{\theta_{k-1}} \hat{f}_{\theta_{k-1}}(X_k) \right\} = M_n + R_n$$

Loi des grands nombres

Théorème (GF et al., 2010)

Soit $f : \mathsf{X} \to \mathbb{R}$. Si

- $\exists!\ \pi_\theta$ telle que $\pi_\theta P_\theta = \pi_\theta$, et une solution $\hat{f}_\theta$
- Convergence p.s. du terme martingale
- Convergence p.s. des restes
- $\pi_{\theta_n}(f) \to \pi(f)$ p.s.

Alors

$$\lim_n \frac{1}{n} \sum_{k=1}^{n} f(X_k) \xrightarrow{p.s.} \pi(f).$$

Il suffit

- Comportement (V-)ergodique analogue des noyaux P_θ.
- Régularité en θ de $\theta \mapsto \hat{f}_\theta$; $\Leftarrow$ Régularité en θ de $\theta \mapsto P_\theta$; $\Leftarrow$ Adaptation décroissante (en norme V).
- $\pi_{\theta_n}(f) \to \pi(f)$ p.s.

Convergence des mesures invariantes

- On sait que π_θ existe mais on n'a pas nécessairement son expression.
- On peut utiliser cas où $\pi = \pi_{\theta_\star}$

$$\pi_\theta(f) - \pi_{\theta_\star}(f) = \left\{\pi_\theta(f) - P_\theta^k f(x)\right\} + \left\{P_{\theta_\star}^k f(x) - \pi_{\theta_\star}(f)\right\} + P_\theta^k f(x) - P_{\theta_\star}^k f(x)$$

puis

- comportement ergodique similaire des noyaux P_θ.
- régularité en θ de $\theta \mapsto P_\theta^k f(x)$

Convergence des mesures invariantes

- On sait que π_θ existe mais on n'a pas nécessairement son expression.
- On peut utiliser cas où $\pi = \pi_{\theta_\star}$

$$\pi_\theta(f) - \pi_{\theta_\star}(f) = \left\{\pi_\theta(f) - P_\theta^k f(x)\right\} + \left\{P_{\theta_\star}^k f(x) - \pi_{\theta_\star}(f)\right\} + P_\theta^k f(x) - P_{\theta_\star}^k f(x)$$

  puis
  - comportement ergodique similaire des noyaux P_θ.
  - régularité en θ de $\theta \mapsto P_\theta^k f(x)$
- Dans tous les cas,
  - régularité en θ de $\theta \mapsto \pi_\theta(f)$
  - Par ex. lorsque $\theta_n \equiv$ mesure empirique, résultats sur fonctionnelles de Von-Mises $+$ U-stat
- A suivre $\cdots$

IV. Programme de recherche

- Algorithmes de simulation adaptatifs
- Simulation de lois dont l'évaluation a un coût de calcul prohibitif.
  $\hookrightarrow$ traitement bayésien de données sans vraisemblance explicite
- Simulation de lois dynamiques.
  $\hookrightarrow$ SLAM (localisation et cartographie simultanées).

Lois dont l'évaluation a un coût prohibitif

- Approches envisagées :
  - Se ramener à de la simulation dans des espaces de plus petite dimension (simulation "par blocs") : identifier la structure de dépendance des variables; reparamétrisation; méthodes de simulation capables de se mouvoir entre les différents éléments de la partition $\hookrightarrow$ Méthodes transdimensionnelles adaptatives (Wang-Landau; Reversible Jump adaptatif).

Lois dont l'évaluation a un coût prohibitif

- Approches envisagées :
  - Se ramener à de la simulation dans des espaces de plus petite dimension (simulation "par blocs") : identifier la structure de dépendance des variables; reparamétrisation; méthodes de simulation capables de se mouvoir entre les différents éléments de la partition
    $\hookrightarrow$ Méthodes transdimensionnelles adaptatives (Wang-Landau; Reversible Jump adaptatif).
  - Approcher la loi cible par une loi auxiliaire et sélection de ces tirages pour n'évaluer la loi cible qu'en des points "pertinents" : combiner IS (tirages sous la loi auxiliaire) et MCMC de type HM (correction des tirages)
    $\hookrightarrow$ s'inspirer des méthodes d'optimisation d'une fonction coûteuse à évaluer.

Lois dont l'évaluation a un coût prohibitif

- Approches envisagées :
  - Se ramener à de la simulation dans des espaces de plus petite dimension (simulation "par blocs") : identifier la structure de dépendance des variables; reparamétrisation; méthodes de simulation capables de se mouvoir entre les différents éléments de la partition
    $\hookrightarrow$ Méthodes transdimensionnelles adaptatives (Wang-Landau; Reversible Jump adaptatif).
  - Approcher la loi cible par une loi auxiliaire et sélection de ces tirages pour n'évaluer la loi cible qu'en des points "pertinents" : combiner IS (tirages sous la loi auxiliaire) et MCMC de type HM (correction des tirages)
    $\hookrightarrow$ s'inspirer des méthodes d'optimisation d'une fonction coûteuse à évaluer.
  - Approches qui pourront exploiter la force de calcul disponible : méthodes parallélisables.
- Applications : astro-statistique.

Lois dynamiques

- Applications : problème inverse pour des données séquentielles $\Rightarrow$ famille de lois cibles "emboitées"
  $\hookrightarrow$ SLAM ▸ Voir

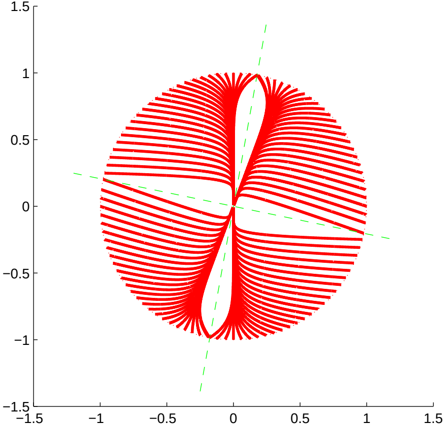
Lois dynamiques

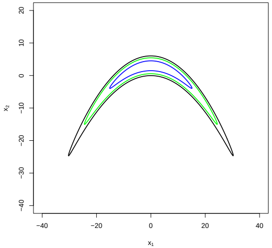
- Applications : problème inverse pour des données séquentielles $\Rightarrow$ famille de lois cibles "emboitées"
  $\hookrightarrow$ SLAM Voir
- Approches envisagées :
  - Méthodes *Sequential Monte Carlo*, plus robustes.
  - Combiner IS et MCMC : IS pour "mutation des particules" et MCMC pour correction des tirages. (Sequential Interacting MCMC)
    $\hookrightarrow$ inter-actions de plusieurs algorithmes de simulation.

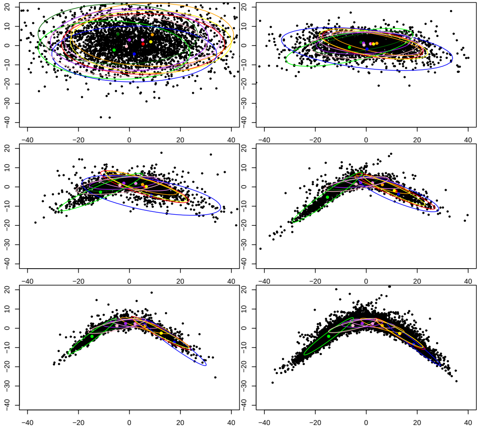
V. Récapitulatif des activités

- Recherche
  - Coordonnateur de deux ANR (ADAP'MC puis Big'MC)
  - Co-organisateur d'un workshop international *New developments in Monte Carlo methods* (Juin 2007)
  - Co-organisateur de séminaires (séminaire parisien de Stat, séminaire Big'MC)
  - Reviewer dans revues internationales
- Enseignement, Formation
  - Enseignement L3, M1-M2. ~ 95h par an (dont ~ 70h de cours)
  - Thèses co-dirigées.

- Production scientifique
  - Publications dans revues internationales
    - Ann. Appl. Proba, Ann. Stat, Stoch. Process. Appl., Bernoulli, J. Appl. Proba, Stat. Prob. Lett
    - Ann. Oper. Res., IEEE Trans. Image Processing, Phys. Rev.D, MNRAS, Bioinformatics
  - Collaborations
    - nationales (K. Benabed, O. Cappé, R. Douc, P. Etoré, F. Forbes, A. Guillin, B. Jourdain, S. Lambert-Lacroix, E. Moulines, P. Priouret, S. Prunet, P. Soulier, C. Robert).
    - internationales (C. Andrieu, Y. Atchadé, S. Connor, S.Meyn, G.O.Roberts, J.S. Rosenthal)
  - Conférences (sessions invitées, sessions contribuées), séminaires $\cdots$

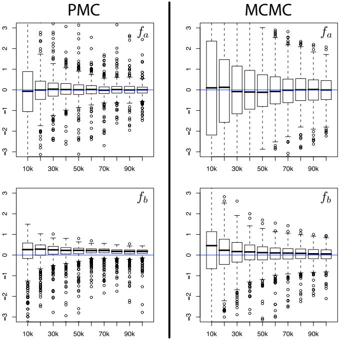

1.5
1
0.5
0
−0.5
−1
−1.5
−1.5
−1
−0.5
0
0.5
1
1.5

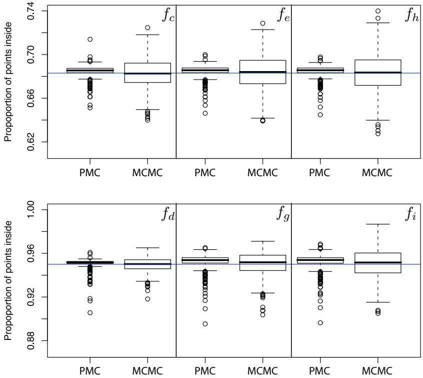




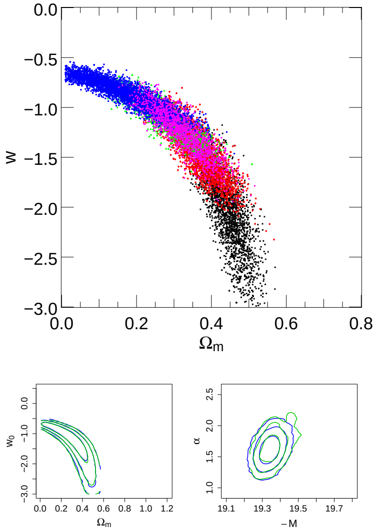
Retour

$$f_c(x) = \mathbb{1}_{68.3}(x) \qquad f_d(x) = \mathbb{1}_{95}(x)$$
$$f_e(x) = \mathbb{1}_{68.3}(x_1,x_2) \qquad f_g(x) = \mathbb{1}_{95}(x_1,x_2)$$
$$f_a(x) = x_1 \quad f_b(x) = x_2 \qquad f_h(x) = \mathbb{1}_{68.3}(x_1) \qquad f_i(x) = \mathbb{1}_{95}(x_1)$$



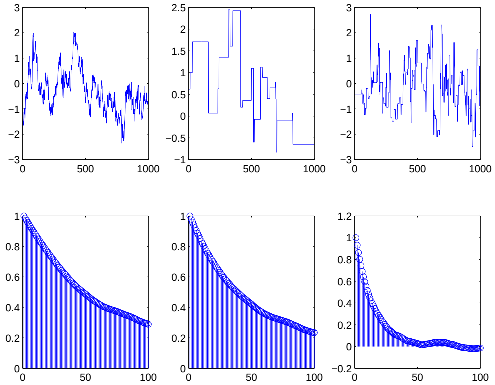


Retour



Param.	PMC	MCMC
Ω_b	$0.04424^{+0.00321}_{-0.00290}$	$0.04418^{+0.00321}_{-0.00294}$
Ω_m	$0.2633^{+0.0340}_{-0.0282}$	$0.2626^{+0.0359}_{-0.0280}$
τ	$0.0878^{+0.0181}_{-0.0160}$	$0.0885^{+0.0181}_{-0.0160}$
n_s	$0.9622^{+0.0145}_{-0.0143}$	$0.9628^{+0.0139}_{-0.0145}$
$10^9\Delta^2_R$	$2.431^{+0.118}_{-0.113}$	$2.429^{+0.123}_{-0.108}$
h	$0.7116^{+0.0271}_{-0.0261}$	$0.7125^{+0.0274}_{-0.0268}$

Suite

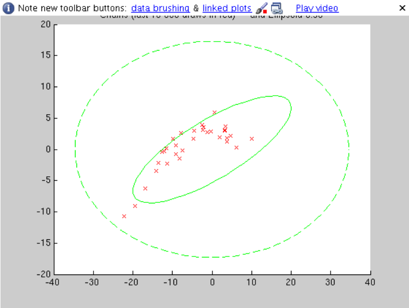


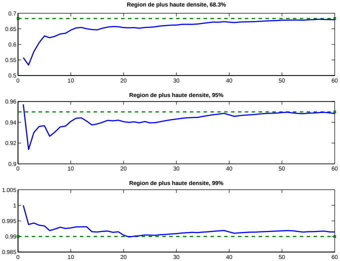
"goldilock principle"
Variance trop petite, trop grande, mieux adaptée

[dans $\mathbb{R}$] On montre, pour différentes valeurs de θ,

- l'évolution d'une trajectoire
- la fonction d'auto-corrélation $\longleftrightarrow$ variance dans le TCL pour $f(x) = x$

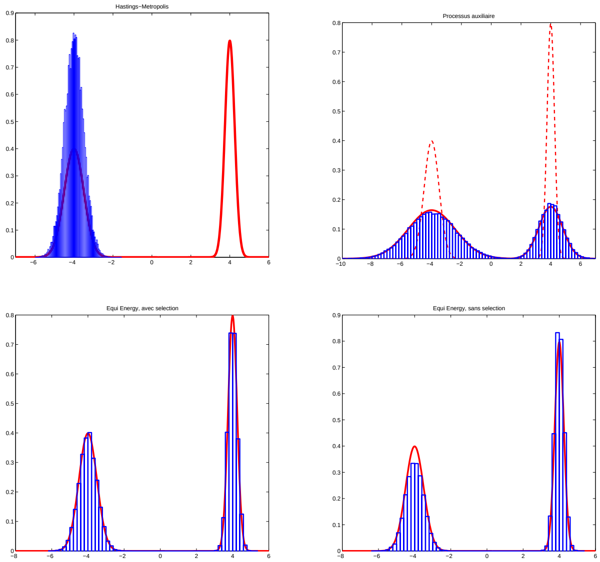
Retour





Adaptive HM

▸ Suite


Hastings−Metropolis
Processus auxiliaire
Equi Energy, avec selection
Equi Energy, sans selection

Suite

Montrer le robot qui bouge + quelques équations $\cdots$
Spécificité du problème :

- Etat caché : composante statique (la carte $\equiv$ landmarks identifiables) et une composante dynamique (position du robot).
- Equations d'état et d'observations : non linéaires, non gaussiennes.
- Résolution du problème inverse “en ligne”.

◂ Retour