

Parallel Tempering with Equi-Energy Moves

Meïli Baragatti^{1,2}, Agnès Grimaud¹, Denys Pommeret¹

¹ Institut de Mathématiques de Luminy, Marseille.

² Ipsogen SA, Marseille.

Réunion Applibugs, Agro Paris Tech, 17 juin 2011

Motivation

Distributions multimodales ou modèles complexes ayant de nombreux paramètres : la chaîne générée par MCMC classique :

- se "mélange" mal.
- reste "bloquée" dans un maximum local de l'espace des états.

L'espace des paramètres n'est pas entièrement visité : des modes peuvent ne pas avoir été visités ou bien être représentés dans des proportions incorrectes.

Motivation

Développement d'algorithmes utilisant une seule chaîne :

- **delayed rejection** : Mira (1998) et Green et Mira (2001).
- **simulated annealing** : Kirkpatrick *et al.* (1983), Liu (2001).
- **simulated tempering** : Marinari et Parisi (1992) et Geyer et Thompson (1995).
- **tempered transition** : Neal (1996).

Motivation

Développement d'algorithmes utilisant une seule chaîne :

- **delayed rejection** : Mira (1998) et Green et Mira (2001).
- **simulated annealing** : Kirkpatrick *et al.* (1983), Liu (2001).
- **simulated tempering** : Marinari et Parisi (1992) et Geyer et Thompson (1995).
- **tempered transition** : Neal (1996).

Développement d'algorithmes population-based :

- **MCMCMC, Exchange Monte Carlo, Swapping Algorithm, Parallel Tempering** : Geyer (1991), Hukushima et Nemoto (1996).
- **Evolutionary Monte Carlo** : Liang et Wong, (2000 et 2001)
- **Equi-Energy Sampler** : Kou *et al.* (2006).

Parallel Tempering

Principe

Algorithme

Equi-Energy Sampler

PTEEM

Applications et perspectives

Notations

- π distribution d'intérêt définie sur $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.
- $\pi_1, \pi_2, \dots, \pi_N$ série de distributions liées entre elles définies sur $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ avec $\pi_1 = \pi \implies N$ chaînes en parallèle.
- Plus i augmente et plus π_i est facile à explorer.

Notations

- π distribution d'intérêt définie sur $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.
- $\pi_1, \pi_2, \dots, \pi_N$ série de distributions liées entre elles définies sur $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ avec $\pi_1 = \pi \implies N$ chaînes en parallèle.
- Plus i augmente et plus π_i est facile à explorer.

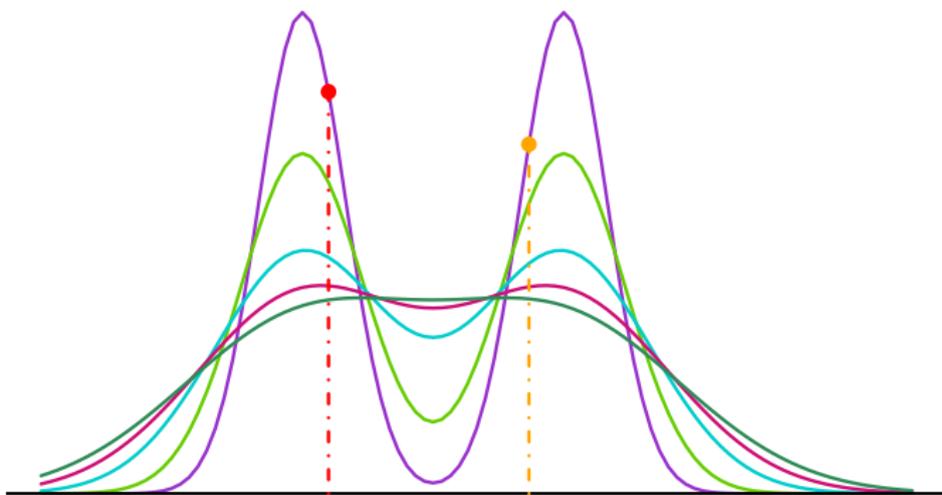
Choix des des distributions

En général, les π_i sont choisies comme étant des distributions tempérées de π .

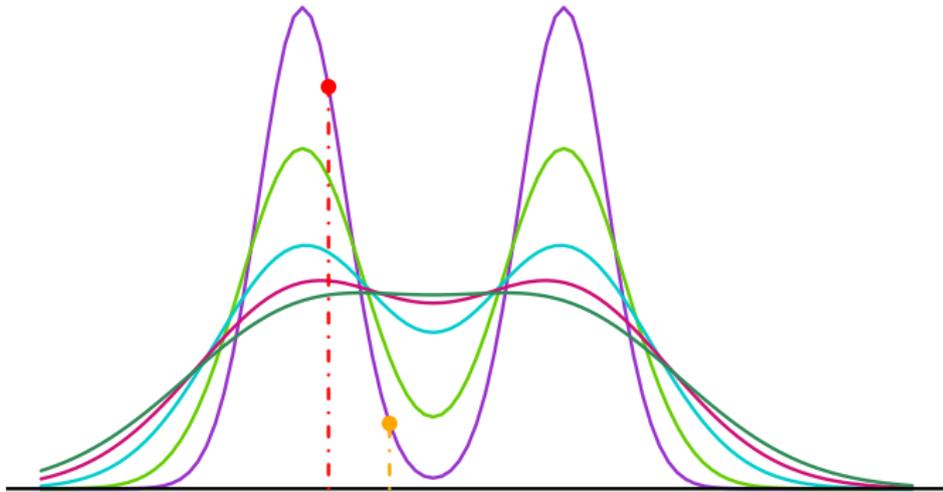
Avec $1 = T_1 < T_2 < \dots < T_i < \dots < T_N$,

$$\pi_i(x) \propto \pi(x)^{1/T_i}, \quad i \in \{1, 2, \dots, N\}$$

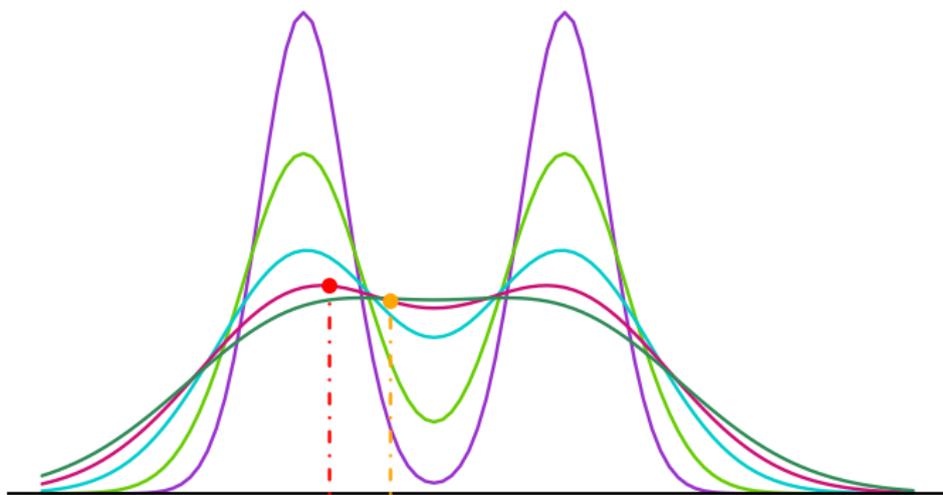
Distributions tempérées



Distributions tempérées



Distributions tempérées



Principe du Parallel Tempering

Principe

Permettre des échanges entre les différentes distributions.

Un mode exploré par une distribution chaude peut se propager jusqu'à une distribution froide et lui permettre de s'échapper d'une région modale dans laquelle elle était "coincée".

Algorithme Parallel Tempering (Geyer, 1991)

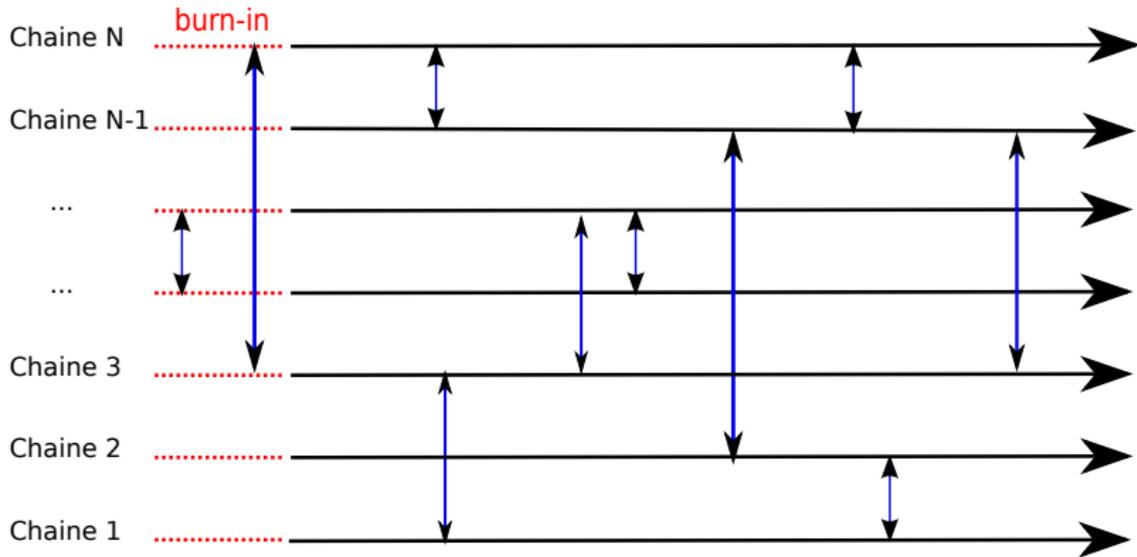
N chaînes en parallèle. Une itération est constituée de :

1. N mouvements locaux : pour $i = 1, 2, \dots, N$, actualiser la i^{eme} chaîne avec un algorithme MCMC ayant π_i comme distribution stationnaire.
2. Mouvement global d'échange classique : choisir deux chaînes uniformément dans $\{1, 2, \dots, N\}$ puis proposer d'échanger leurs états actuels.

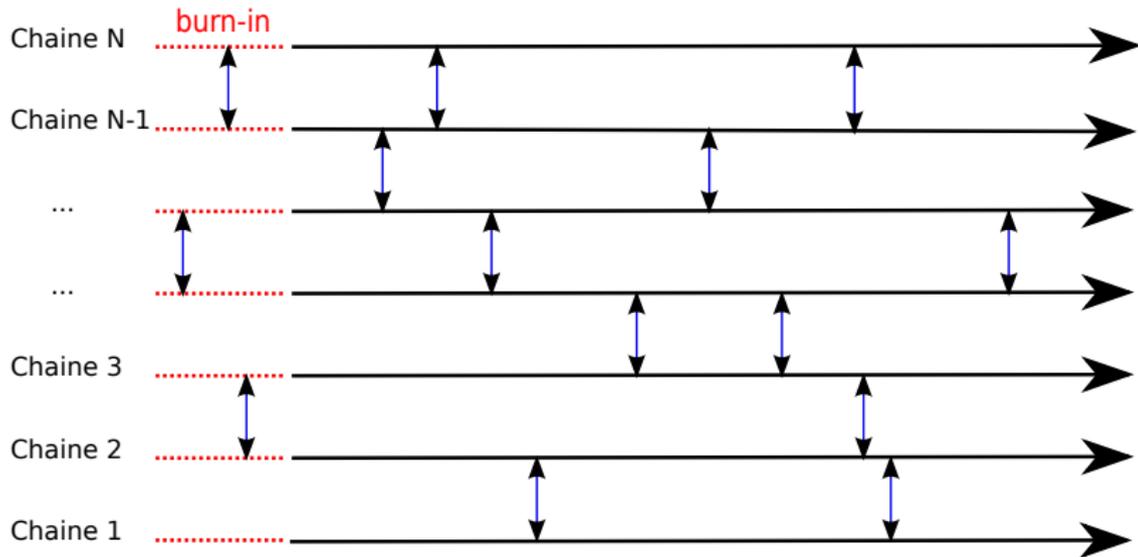
$$\theta = (x_1, \dots, x_j, \dots, x_k, \dots, x_N) \implies \theta' = (x_1, \dots, x_k, \dots, x_j, \dots, x_N)$$

Echange accepté avec le ratio MH suivant :

$$\rho_1(\theta, \theta') = 1 \wedge \frac{\pi_j(x_k)\pi_k(x_j)}{\pi_j(x_j)\pi_k(x_k)}$$



⇒ modification de l'algo Parallel Tempering



Parallel Tempering

Equi-Energy Sampler

Principe

Algorithme (Kou et al. 2006)

Avantages et inconvénients

PTEEM

Applications et perspectives

Notations de l'Equi-Energy sampler

- π distribution d'intérêt définie sur $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.
- $\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_K$ série de distributions définies sur $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$
 $\implies K$ chaînes.
- Niveaux d'énergie : $H_1 < \dots < H_K < H_{K+1} = \infty$,
 $H_1 \leq \inf_x \{-\ln(\pi(x))\}$.

Notations de l'Equi-Energy sampler

- π distribution d'intérêt définie sur $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.
- $\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_K$ série de distributions définies sur $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$
 $\implies K$ chaînes.
- Niveaux d'énergie : $H_1 < \dots < H_K < H_{K+1} = \infty$,
 $H_1 \leq \inf_x \{-\ln(\pi(x))\}$.

Choix des des distributions

Les $\tilde{\pi}_i$ sont des distributions tempérées de π + TRONCATURE.

$$\tilde{\pi}_i \propto \pi_i \wedge \exp\left(-\frac{H_i}{T_i}\right)$$

Principe, Kou *et al.* 2006

Partition de l'espace des états : $\mathcal{X} = \bigcup_{j=1}^K D_j$.

Dans un anneau se trouvent des états d'énergies similaires (vraisemblances similaires).

Mouvement equi-energy

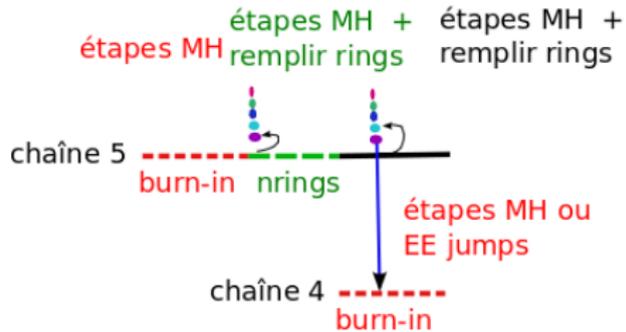
Une chaîne peut prendre un état du passé d'une autre chaîne qui est d'énergie similaire à son état actuel.

étapes MH

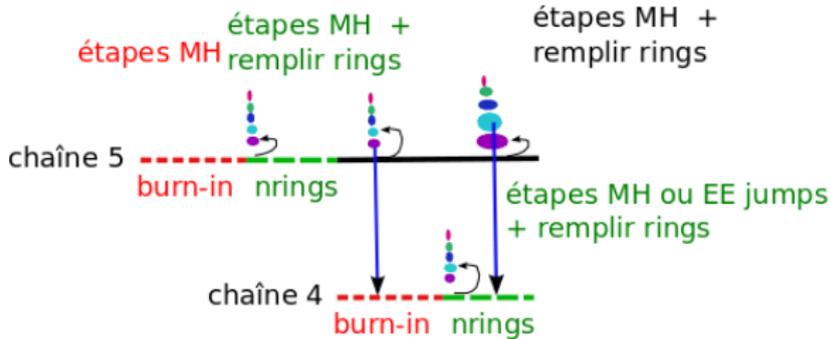
chaîne 5 -----
burn-in

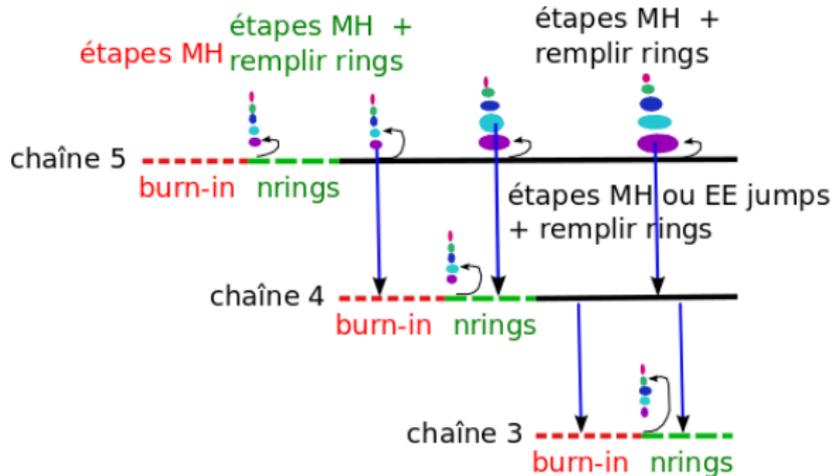
étapes MH +
étapes MH remplir rings

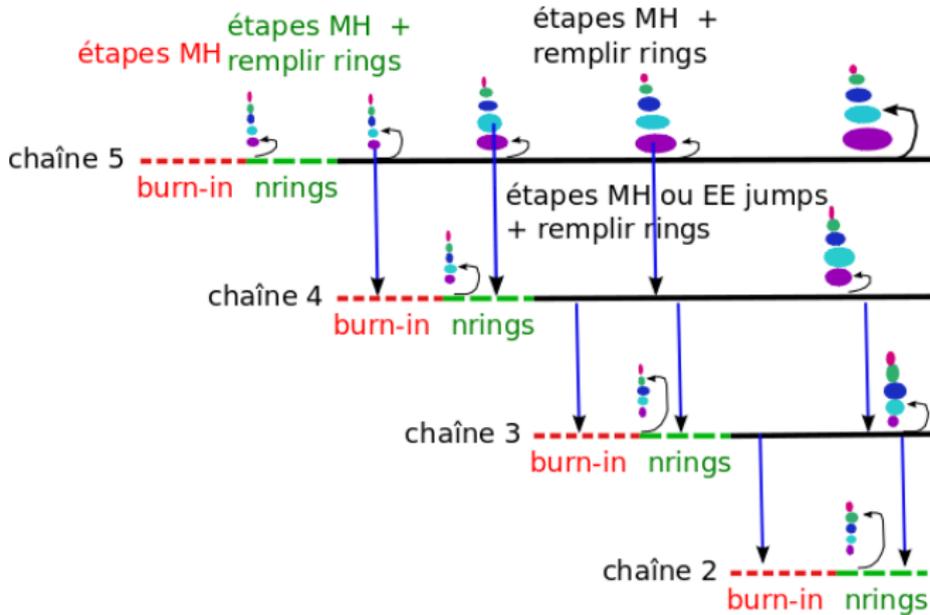


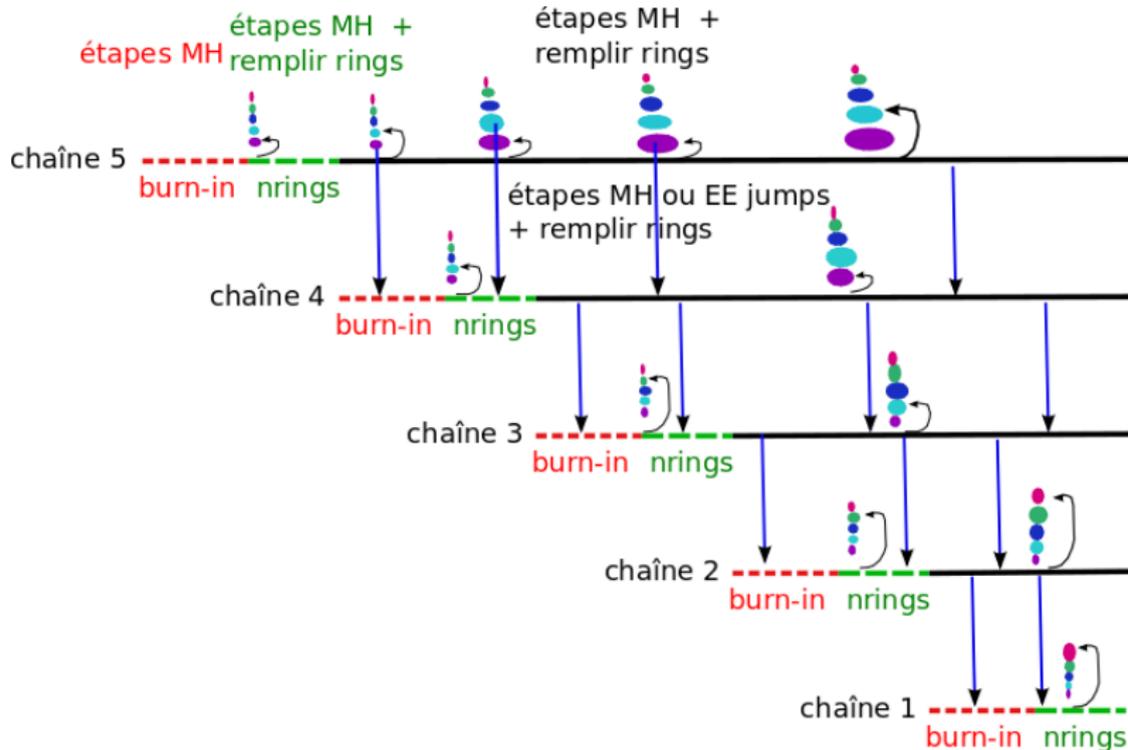


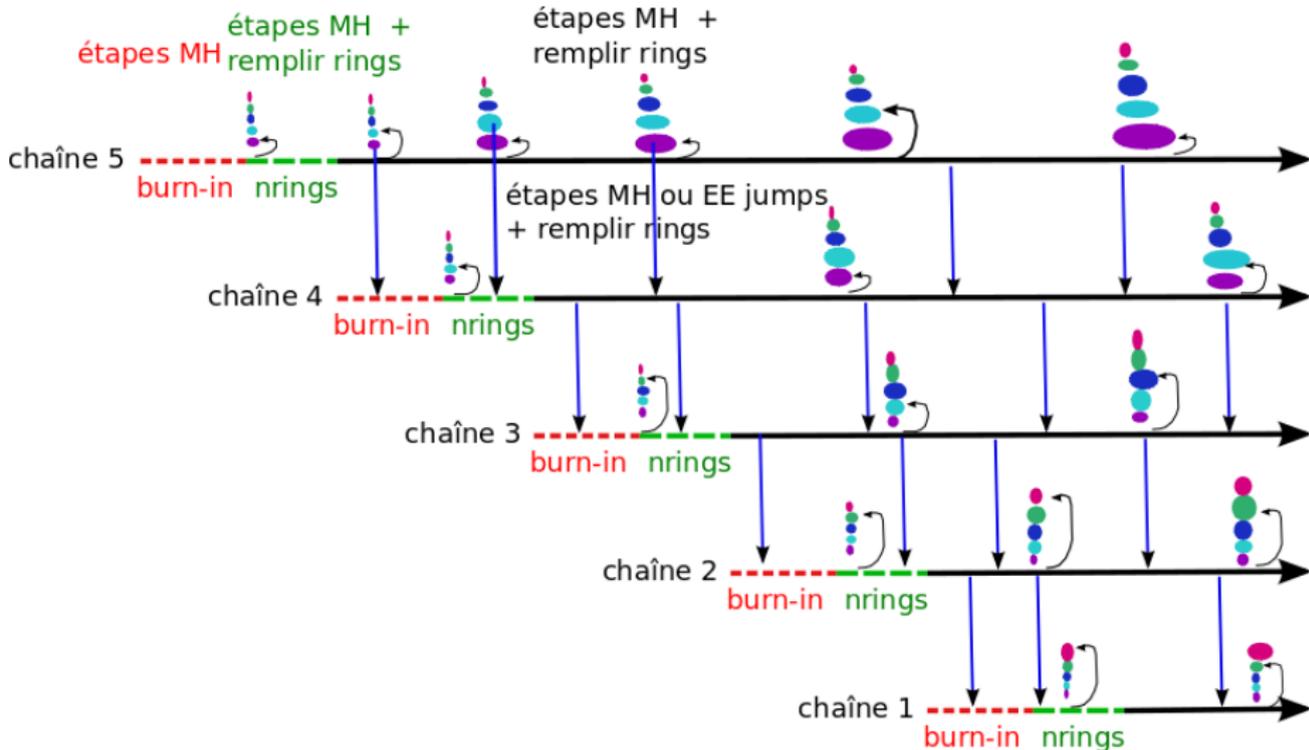
burn-in nrings







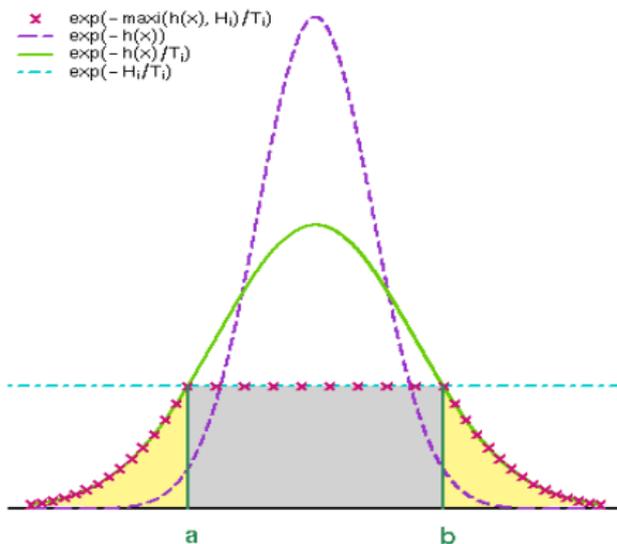




Avantages et inconvénients de l'EES

- Très bons résultats empiriques, meilleurs que ceux de PT.
- Processus généré non markovien \Rightarrow propriétés théoriques et convergence non évidentes :
 - ▶ Kou *et al.* (2006) : théorème incomplet.
 - ▶ Atchadé et Liu (2006) : proposition de preuve, hyp. fortes.
 - ▶ Andrieu *et al.* (2007) : preuve précédente non adaptée.
 - ▶ Andrieu *et al.* (2008) : une seule feeding chaîne, hyp. fortes.
 - ▶ Hua et Kou (2010) : espace des états discrets.
 - ▶ Fort, Priouret et Moulines (2010) : convergence of adaptive and interacting MCMC.
- Stockage important.
- **Utilisation de troncatures** : pb si Gibbs Sampler, RJMCMC, ou hybride.

$$\tilde{\pi}_i(x) \propto \pi_i \wedge \exp\left(-\frac{H_i}{T_i}\right)$$



Parallel Tempering

Equi-Energy Sampler

PTEEM

Principe

Algorithme

Propriétés théoriques

En pratique

Applications et perspectives

Principe

Objectif : algorithme ayant les propriétés suivantes :

- Bonnes propriétés théoriques comme le PT.
- S'appliquant facilement à un Gibbs sampler ou hybride.
- Donnant d'aussi bons résultats que l'EES.

Combiner PT avec l'idée equi-energy de l'EES

Un algorithme PT amélioré, utilisant des mouvements equi-energy réversibles.

Notations du PTEEM

- π distribution d'intérêt définie sur $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.
- $\pi_1, \pi_2, \dots, \pi_N$ série de distributions définies sur $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$
 $\implies N$ chaînes en parallèle.

Choix des des distributions

Distributions tempérées de π . Avec $1 = T_1 < T_2 < \dots < T_N$,

$$\pi_i(x) \propto \pi(x)^{1/T_i}, \quad i \in \{1, 2, \dots, N\}$$

Notations du PTEEM

- π distribution d'intérêt définie sur $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.
- $\pi_1, \pi_2, \dots, \pi_N$ série de distributions définies sur $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$
 $\implies N$ chaînes en parallèle.

Choix des des distributions

Distributions tempérées de π . Avec $1 = T_1 < T_2 < \dots < T_N$,

$$\pi_i(x) \propto \pi(x)^{1/T_i}, \quad i \in \{1, 2, \dots, N\}$$

- Partition de l'espace des états : $\mathcal{X} = \bigcup_{j=1}^K D_j$.

Avec $H_1 < \dots < H_K < H_{K+1} = \infty$, $H_1 \leq \inf_x h(x)$.

Algorithme PTEEM

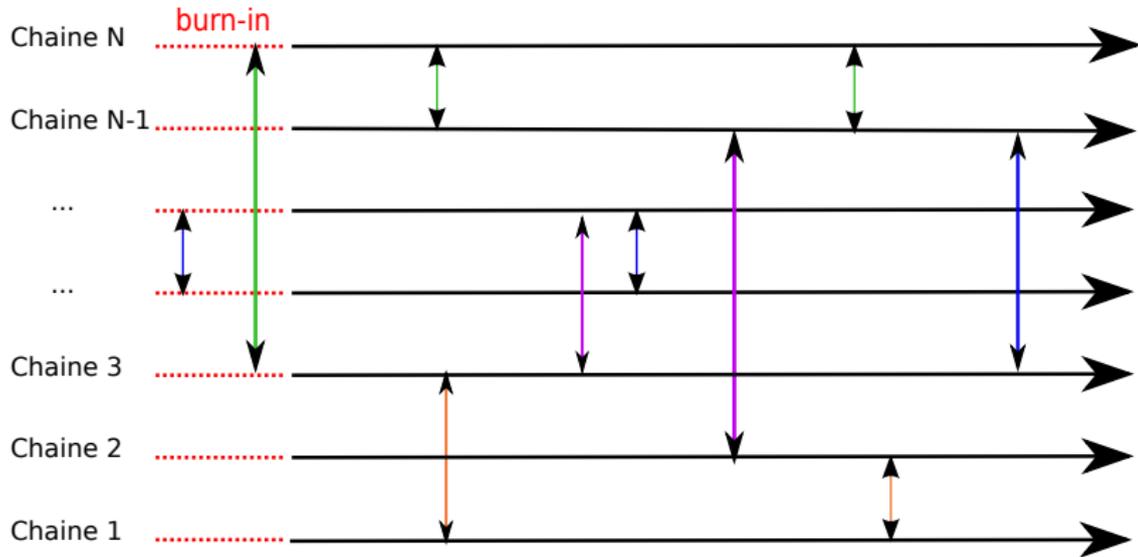
N chaînes en parallèle. Une itération est constituée de :

1. N mouvements locaux : pour $i = 1, 2, \dots, N$,
 - 1.1 actualiser la i^{eme} chaîne avec un algorithme MCMC ayant π_i comme distribution stationnaire.
 - 1.2 Actualisation des anneaux d'énergie.
2. Mouvement global d'échange equi-energy : Choisir un anneau contenant au moins deux chaînes, y choisir uniformément deux chaînes d'indices j et k , puis proposer d'échanger leurs états actuels.

$$\theta = (x_1, \dots, x_j, \dots, x_k, \dots, x_N) \implies \theta' = (x_1, \dots, x_k, \dots, x_j, \dots, x_N)$$

Cet échange est accepté avec le ratio MH suivant :

$$\rho_1(\theta, \theta') = 1 \wedge \frac{\pi_j(x_k)\pi_k(x_j)}{\pi_j(x_j)\pi_k(x_k)}$$



Proposition 1

Soit l'ensemble des N chaînes, P noyau de transition associé à une itération complète du PTEEM, et $\pi^* = \prod_{i=1}^N \pi_i$

Si les noyaux de transition des mouvements locaux sont réversibles, de distributions stationnaires π_i , apériodiques et λ -irréductibles ($i = 1, \dots, N$), alors :

$$\lim_{n \rightarrow \infty} \|P^n(s, \cdot) - \pi^*\|_{TV} = 0 \quad \pi^* - ps$$

La distribution stationnaire de S est donc π^ , et la chaîne associée à $T_1 = 1$ a des réalisations issues de $\pi_1 = \pi$ qui est la distribution d'intérêt.*

Proposition 2

Hypothèses :

- les noyaux de transition des mouvements locaux sont apériodiques et réversibles de distributions stationnaires π_i ($i = 1, \dots, N$).
- $\forall (x, x') \in \mathcal{X} \times \mathcal{X}, PL_i(x, x') > 0, i \in \{1, \dots, N\}$.

Lemme : S est λ_N -irréductible, positive et Harris-récurrente.

Proposition 2 :

$$\lim_{n \rightarrow \infty} \|P^n(s, \cdot) - \pi^*\|_{TV} = 0 \quad \forall s \in \mathcal{X}^N$$

Parallel Tempering

Equi-Energy Sampler

PTEEM

Principe

Algorithme

Propriétés théoriques

En pratique

Applications et perspectives

Vérification des choix de températures et de niveaux d'énergie

	Mauvaise répartition					Bonne répartition				
Anneaux	1	2	3	4	5	1	2	3	4	5
chaîne $i - 2$	990	10	0	0	0	990	10	0	0	0
chaîne $i - 1$	950	50	0	0	0	701	202	97	0	0
chaîne i	900	100	0	0	0	387	408	205	0	0
chaîne $i + 1$	0	2	237	511	250	45	312	355	288	0
chaîne $i + 2$	0	0	105	610	285	0	64	517	353	66

Parfois nécessaire d'ajuster les températures ou les niveaux d'énergie :

- Ajout de nouvelles températures.
- Nouvelle calibration des niveaux d'énergie.

Parallel Tempering

Equi-Energy Sampler

PTEEM

Applications et perspectives

Simulations avec mouvement local de MH

Estimation avec mouvement local Gibbs Sampler

Recherche de sites promoteurs de facteurs de transcription

Perspectives

Mélange 20 lois normales bi-dimensionnelles

$$f(x) = \sum_{i=1}^{20} \frac{w_i}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_i^2}(x - \mu_i)'(x - \mu_i)\right)$$

- Chacun des trois algos lancé 100 fois, utilisation de MH.
- EES : 5 chaînes.
- PT et PTEEM : 20 chaînes.
- PTEEM : 5 groupes d'énergies.

Estimations des premiers et seconds moments de la chaîne 1.

	$E(X_1)$	$E(X_2)$	$E(X_1)^2$	$E(X_2)^2$
Vraie valeur	4.478	4.905	25.605	33.920
PT	3.971 (0.809)	4.137 (1.114)	21.510 (7.741)	27.510 (10.407)
EES	4.448 (0.301)	4.953 (0.458)	25.229 (3.112)	34.226 (4.507)
PTEEM	4.483 (0.324)	4.912 (0.454)	25.556 (3.366)	33.889 (4.406)

Table 1: Estimations obtenues sur 100 runs par l'EES, PT et PTEEM. Les nombres entre parenthèses sont les écart-types obtenus sur ces différents runs.

Estimations des premiers et seconds moments de la chaîne 1.

	$E(X_1)$	$E(X_2)$	$E(X_1)^2$	$E(X_2)^2$
Vraie valeur	4.478	4.905	25.605	33.920
PT	3.971 (0.809)	4.137 (1.114)	21.510 (7.741)	27.510 (10.407)
EES	4.448 (0.301)	4.953 (0.458)	25.229 (3.112)	34.226 (4.507)
PTEEM	4.483 (0.324)	4.912 (0.454)	25.556 (3.366)	33.889 (4.406)

Table 1: Estimations obtenues sur 100 runs par l'EES, PT et PTEEM. Les nombres entre parenthèses sont les écart-types obtenus sur ces différents runs.

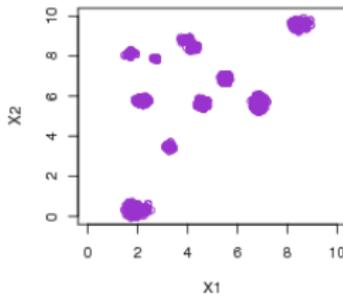
Nombre de modes du mélange visités par la chaîne 1.

PT	EES	PTEEM
De 10 à 18.	De 18 à 20.	De 19 à 20.
14.31 en moyenne.	19.92 en moyenne.	19.98 en moyenne.

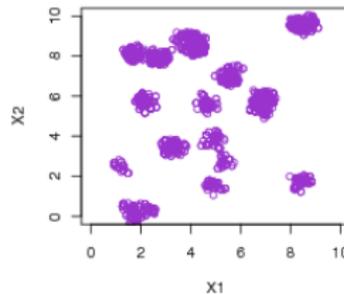
Table 2: Nombre de modes visités lors des 100 runs par l'EES et les algorithmes PT et PTEEM.

4 chaînes obtenues par PT

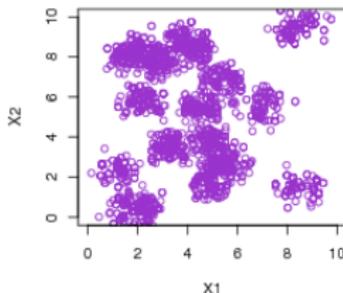
Simulations of chain 1 ($T=1$)



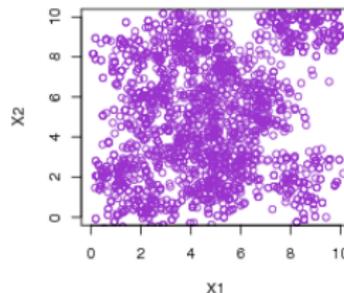
Simulations of chain 7 ($T=3.64$)



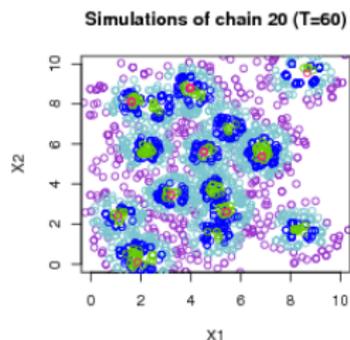
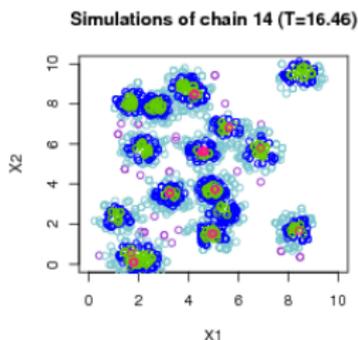
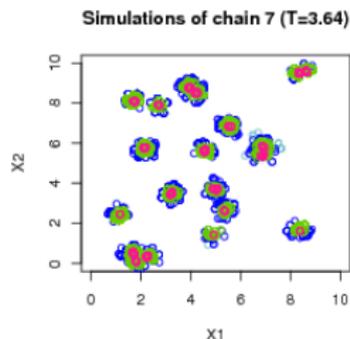
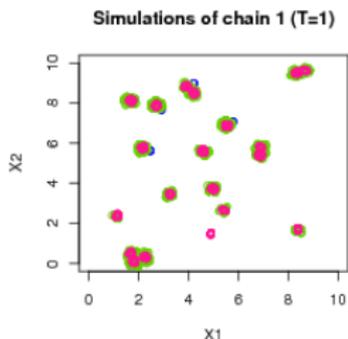
Simulations of chain 14 ($T=16.46$)



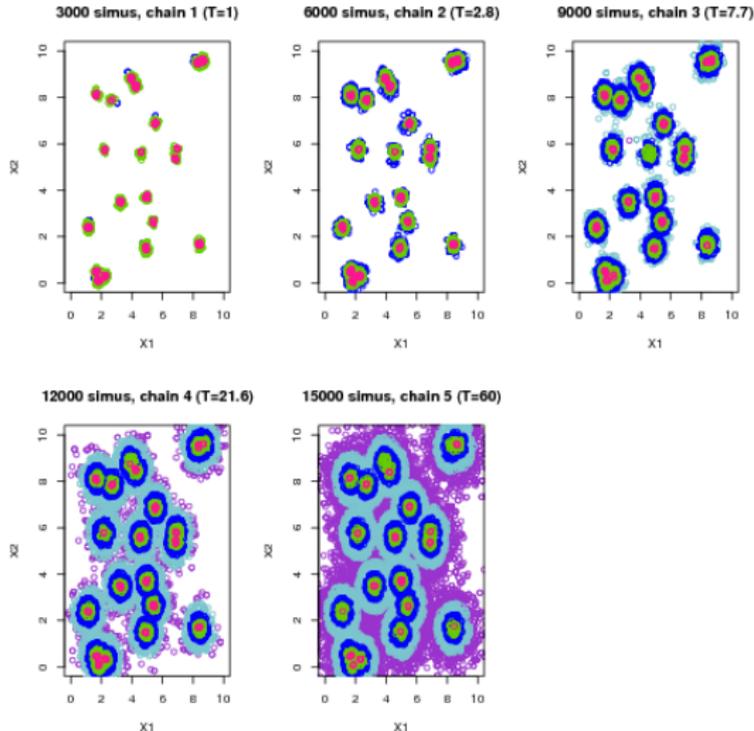
Simulations of chain 20 ($T=60$)



4 chaînes obtenues par PTEEM



Les 5 chaînes de l'EES



Parallel Tempering

Equi-Energy Sampler

PTEEM

Applications et perspectives

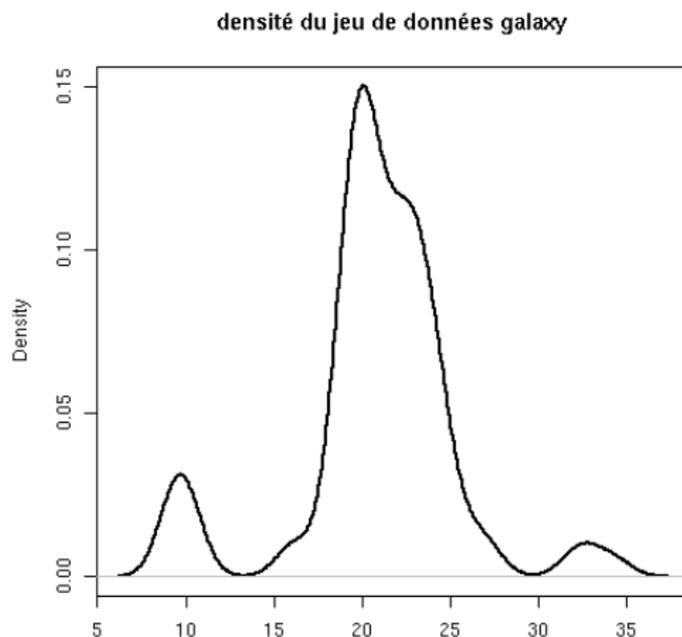
Simulations avec mouvement local de MH

Estimation avec mouvement local Gibbs Sampler

Recherche de sites promoteurs de facteurs de transcription

Perspectives

Exemple Galaxy à 6 composants



Exemple Galaxy à 6 composants

$$y_i \sim \sum_{p=1}^k w_p \mathcal{N}(\mu_p, \sigma_p^2), \quad i = 1, \dots, n,$$

- Vecteur latent d'allocation : $c_i = p$ si y_i est issue du p^{eme} composant.
- A priori classiques pour les c_i , μ_p , σ_p^{-2} et w , $p = 1, \dots, k$.
- L'ensemble des paramètres inconnus : $x = (\mu, \sigma^{-2}, w, c, \beta)$.

$$\pi(x) = p(\mu, \sigma^{-2}, w, c, \beta | y) \propto p(y | \mu, \sigma^{-2}, c) p(\mu, \sigma^{-2}, c, \beta, w).$$

On s'intéresse au phénomène de label-switching, et non pas à l'estimation des paramètres.

Résultats

- Nombre de modes visités lors de 100 runs de **PT** et **PTEEM** :

	moyenne	écart-type	min	max
PT	645.04	13.52	610	683
PTEEM	666.52	9.23	641	692

Parallel Tempering

Equi-Energy Sampler

PTEEM

Applications et perspectives

Simulations avec mouvement local de MH

Estimation avec mouvement local Gibbs Sampler

Recherche de sites promoteurs de facteurs de transcription

Perspectives

Recherche de sites promoteurs de facteurs de transcription

M séquences d'ADN (A, C, G, T)

$|A|$ sites de fixation de longueur w .

$|A|$ inconnu.

Background sequences : Markov
 d'ordre 1.

Sites de fixation : multinomiale
 produit de paramètre Θ .



Recherche de sites promoteurs de facteurs de transcription

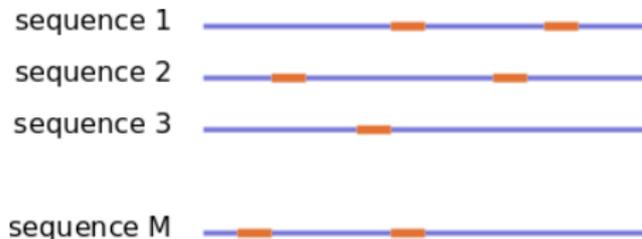
M séquences d'ADN (A, C, G, T)

$|A|$ sites de fixation de longueur w .

$|A|$ inconnu.

Background sequences : Markov
 d'ordre 1.

Sites de fixation : multinomiale
 produit de paramètre Θ .



$$\Theta = \begin{pmatrix} \theta_1 & \theta_2 & \dots & \theta_w \\ \theta_{1A} & \theta_{2A} & \dots & \theta_{wA} \\ \theta_{1C} & \theta_{2C} & \dots & \theta_{wC} \\ \theta_{1G} & \theta_{2G} & \dots & \theta_{wG} \\ \theta_{1T} & \theta_{2T} & \dots & \theta_{wT} \end{pmatrix}$$



Nos données simulées



10 séquences, 20 sites \Rightarrow 2 sites par séquences.

	1	2	3	4	5	6	7	8	9	10	11	12
1	T	T	C	A	A	T	A	G	A	T	A	T
2	A	A	C	A	G	G	A	G	T	C	A	C
3	A	T	C	T	G	G	T	T	T	T	C	T
4	T	A	C	A	G	G	T	G	T	C	A	C
5	A	T	C	G	G	T	T	G	A	C	A	C
6	A	G	T	A	A	G	T	G	T	T	C	C

Nos données simulées

La première séquence entière obtenue est la suivante :

```
TTTTAAAATGGGCCCCGTAAAAATTCAATAGATATTTAATG  
AAGAGGGGTCCGGAAGGGCCCCCGGTTCGGGGGGTTTTTGGAAA  
AGTAAAAGGGGTTTAAACCCCTTTAGGGGGGTGGGGCCCGCCA  
AACAGGAGT CACAACCTGGGGTAGGGGCCCGTTTTGGGTTGGCC  
CCTTTTGGGACCCTTTTTCCAAAACGGCCTTTTAATTTTTTGGG  
GGG.
```

Nos données simulées

La première séquence entière obtenue est la suivante :

TTTTAAAATGGGCCCCGTAAAAA**TTCAATAGATAT**TTAATG
AAGAGGGGTCCGGAAGGGCCCCCGGTTCGGGGGGTTTTTGGAAA
AGTAAAAGGGGTTTAAACCCCTTTAGGGGGGTGGGGCCCGCCA
AACAGGAGT CACAACCTGGGGTAGGGGCCCGTTTTGGGTTGGCC
CCTTTTGGGACCCTTTTT**CCAAAC**GGCCTTTAATTTTTTGGG
GGG.

Nos données simulées

La première séquence entière obtenue est la suivante :

TTTTAAAATGGGCCCGTAAAA **AATTC AATAGAT** ATTTAATG
AAGAGGGGTCCGGAAGGGCCCCCGGTTCGGGGGGTTTTTGGAAAA
AGTAAAAGGGGTTTTAAACCCCTTTAGGGGGGTGGGGCCCGCCA
A **ACAGGAGT CAC** AACCTGGGGTAGGGGCCCGTTTTGGGTTGGCC
CCTTTTGGGACCCTTTTT**CCAAAAC**GGCCTTTTAATTTTTTTGGG
GGG.

Approche classique : échantillonneur de Gibbs

Objectif : trouver les positions initiales des sites de fixations.

- M séquences les unes à la suite des autres \Rightarrow une longue séquence contenant $|A|$ sites de fixation.
- $A = (a_1, a_2, \dots, a_{L^*})$ constitué de 0 et de 1, avec $a_j = 1$ si la j^{eme} position de la grande séquence est le point de départ d'un site, et $a_j = 0$ sinon.

Approche classique : échantillonneur de Gibbs

Objectif : trouver les positions initiales des sites de fixations.

- M séquences les unes à la suite des autres \Rightarrow une longue séquence contenant $|A|$ sites de fixation.
- $A = (a_1, a_2, \dots, a_{L^*})$ constitué de 0 et de 1, avec $a_i = 1$ si la i^{eme} position de la grande séquence est le point de départ d'un site, et $a_i = 0$ sinon.

Formule d'actualisation dans l'échantillonneur de Gibbs

$$\frac{p(a_i = 1 | A_{[-i]}, S)}{p(a_i = 0 | A_{[-i]}, S)} = \frac{1}{\pi(S(a_i) | A, \theta_0)} \times \frac{|A_{[-i]}| + a}{L^* - |A_{[-i]}| - 1 + b} \times \prod_{k=1}^w \left(\frac{C_{k(-i)} + \beta_k}{|A_{[-i]}| + |\beta_k|} \right)^{C_{k(i)}}$$

EES : 9 chaînes, avec :

$$\tilde{\pi}_j(A) \propto \exp\left(-\frac{h(A) \vee H_j}{T_j}\right) \quad \text{et} \quad h(A) = -\log(\pi(A | S))$$

- Première chaîne actualisée avec échantillonneur de Gibbs.
- Les autres actualisées avec MH : un nouvel état A^* proposé à partir de A , et accepté suivant un ratio MH. **Loi de proposition non évidente.**

EES : 9 chaînes, avec :

$$\tilde{\pi}_j(A) \propto \exp\left(-\frac{h(A) \vee H_j}{T_j}\right) \quad \text{et} \quad h(A) = -\log(\pi(A | S))$$

- Première chaîne actualisée avec échantillonneur de Gibbs.
- Les autres actualisées avec MH : un nouvel état A^* proposé à partir de A , et accepté suivant un ratio MH. **Loi de proposition non évidente.**

PTEEM : 15 chaînes et 5 anneaux d'énergie, avec :

$$\pi_j(A | S) = \pi(A | S)^{\frac{1}{T_j}}$$

- Toutes les chaînes actualisées avec des échantillonneurs de Gibbs.
Utilisation de la formule d'actualisation tempérée.

Résultats

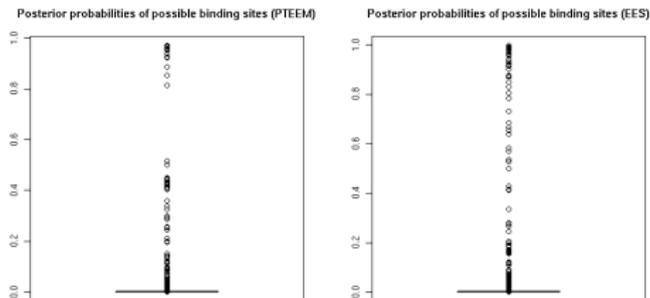


Figure 1: Boxplots des $P(a_i = 1 \mid S)$, $i = 1, \dots, L^*$ obtenues empiriquement lors d'un run PTEEM, et d'un run de EES.

Résultats

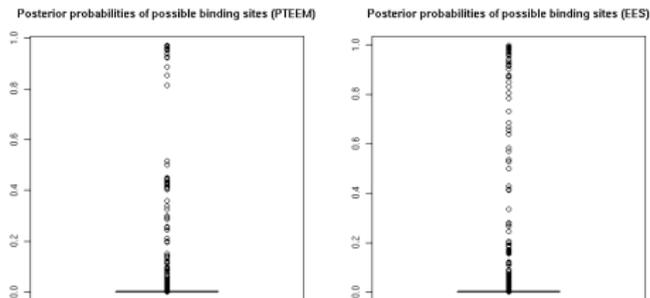


Figure 1: Boxplots des $P(a_i = 1 | S)$, $i = 1, \dots, L^*$ obtenues empiriquement lors d'un run PTEEM, et d'un run de EES.

	EES	PTEEM
Sites identifiés	15.6	16
Sites identifiés avec bonne position	9.6	15
Sites identifiés avec décalage ou plusieurs positions	6	1

Plusieurs positions :

877, 880 et 883 conservées,
 la vraie position étant 880.

Décalage :

27 conservée, la vraie
 position étant 26.

Discussion

- **Calibration** : difficile pour l'EES :
 - Températures : 1, 1.001, 1.002, 1.005, 1.01, 1.02, 1.06, 1.1 et 1.3.
 - Lois de proposition pour les algo MH.
- **Mélange des chaînes** : OK pour PTEEM. Moins bon pour l'EES :
 - Actualisation locale plus lente (entre 15% et 0.7%).
 - Peu de sauts equi-energy proposés entre chaînes 1 et 2 (26 en moyenne sur 1000 itérations).

L'EES a plus eu tendance à identifier les sites avec un décalage, ce qui signifie qu'il restait plus souvent bloqué dans des modes locaux.

⇒ **Avantage à utiliser PTEEM en combinaison avec un échantillonneur de Gibbs.**

Parallel Tempering

Equi-Energy Sampler

PTEEM

Applications et perspectives

Simulations avec mouvement local de MH

Estimation avec mouvement local Gibbs Sampler

Recherche de sites promoteurs de facteurs de transcription

Perspectives

Perspectives

- Alléger les hypothèses pour la Harris-récurrance.
- Vitesse de convergence : comparer PT et PTEEM, et éventuellement EES.
- Algorithme PTEEM adaptatif pour choix des températures et des niveaux d'énergie.

Bibliographie



C.J. Geyer et E.A. Thompson.

Annealing markov chain monte carlo with applications to ancestral inference.
[JASA, 90, 1995.](#)



S.C. Kou, Q. Zhou et W.H. Wong.

Equi-energy sampler with application in statistical inference and statistical mechanics.

[The Annals of Statistics, 33 \(4\), 2006.](#)



A. Jasra, D.A. Stephens, et C.C. Holmes

Population-based reversible jump Markov Chain Monte Carlo.
[Biometrika, 94, 2007](#)



L. Tierney.

Markov chains for exploring posterior distributions.
[Annals of Statistics, 22, 1994](#)



C. Robert et G. Casella.

Monte Carlo statistical methods, second edition.
[Springer, 2004.](#)



M. Baragatti, A. Grimaud et D. Pommeret .

Parallel Tempering with Equi-Energy Moves.
[Submitted, arxiv :1101.4743.](#)

Choix des niveaux d'énergie

Niveaux $H_1, H_2, \dots, H_d \Rightarrow d$ anneaux d'énergies.

Choix H_1 et H_d : utilisation d'un ou quelques runs d'un algo MCMC classique ayant $\pi_1 = \pi$ comme distribution stationnaire.

- H_d : énergie associée à un état ayant une énergie élevée par rapport aux autres (loin d'un mode). Soit l'énergie associée à un état après quelques itérations seulement de l'algo.
- H_1 : énergie associée à un mode observé. Soit l'énergie associée à un état obtenu après la période de burn-in (inutile si modes connus).

Choix des autres niveaux :

$$\ln(H_i) = \ln(H_1) + i \frac{\ln(H_d) - \ln(H_1)}{d - 1}$$

Choix des températures

- T_N : distribution associée suffisamment écrasée pour se déplacer facilement d'un mode à un autre : à vérifier.
- Une fois que T_1 et T_N ont été déterminés, plusieurs possibilités :
 - Les T_i régulièrement espacées sur une échelle logarithmique (Kou *et al.* 2006).
 - Les $1/T_i$ régulièrement espacés ou ayant une progression géométrique (Neal 1996, Nagata et Watanabe 2008).

Nombre de mouvements effectués

Soient B le nombre d'itérations burn-in, R le nombre d'itérations nécessaires au remplissage initial des anneaux dans l'EES, et M le nombre d'itérations voulu pour la chaîne d'intérêt après burn-in.

- Pour l'**EES**, le nombre total de mouvements locaux effectués est

$$\left[\frac{(K-1)K}{2}(B+R) + (K-1)(M-R) \right] (1 - p_{ee}) + K(B+R) + (M-R)$$

Et le nombre total de mouvements globaux effectués est

$$\left[\frac{(K-1)K}{2}(B+R) + (K-1)(M-R) \right] p_{ee}$$

- Pour **PTEEM**, le nombre total de mouvements locaux effectués est

$$NB + NM$$

Et le nombre total de mouvements globaux effectués est

$$B + M$$

Ex. sites promoteurs : loi de proposition pour MH dans EES

Soit A l'état actuel de la chaîne j .

- $C_k, k = 1, \dots, w$: vecteurs de comptage des différents nucléotides dans les sites de fixation \Rightarrow estimation de Θ .
- On ajoute des pseudocomptes en fonction de $j \Rightarrow \hat{\Theta}$.
- $\hat{\rho}_0 = 1/\bar{L}^*$.

Chaque élément de A est actualisé indépendamment : nous proposons $a_i^* = 1$ avec probabilité

$$q_i = \frac{\hat{\rho}_0 \prod_{k=1}^w \hat{\Theta}_{k \times k}}{\hat{\rho}_0 \prod_{k=1}^w \hat{\Theta}_{k \times k} + (1 - \hat{\rho}_0) P(x_1 x_2 \dots x_w | \theta_0)}$$

$\Rightarrow A^*$ est proposé et accepté avec la probabilité suivante :

$$\rho = \frac{\tilde{\pi}_l(A^*) P(A | \hat{\Theta}^*)}{\tilde{\pi}_l(A) P(A^* | \hat{\Theta})},$$

où $P(A^* | \hat{\Theta}) = \prod_{i=1}^{L^*} q_i^{a_i^*} (1 - q_i)^{1 - a_i^*}$ représente la probabilité de générer A^* à partir de l'état actuel A .

Notations (1/2)

Chaque chaîne séparément n'est pas Markovienne, c'est l'ensemble des N chaînes qui est markovien sur $(\mathcal{X}^N, \mathcal{B}(\mathcal{X})^N)$.

- S la chaîne de Markov obtenue par l'algo PTEEM, un état de S est noté s .
- P noyau de transition associé à une itération complète et P^k associé à k itérations, définis sur $\mathcal{X}^N \times \mathcal{B}(\mathcal{X})^N$.
- PL_i noyau de transition associé au mouvement local de la chaîne d'indice i , défini sur $\mathcal{X} \times \mathcal{B}(\mathcal{X})$.
- PL noyau associé à l'ensemble des mouvements locaux des N chaînes lors d'une itération, défini sur $\mathcal{X}^N \times \mathcal{B}(\mathcal{X})^N$.
- PE noyau associé au mouvement global equi-energy, défini sur $\mathcal{X}^N \times \mathcal{B}(\mathcal{X})^N$.

Notations (2/2)

$$\begin{aligned}s &= (x_1, \dots, x_i, \dots, x_k, \dots, x_N) \\ s' &= (x'_1, \dots, x'_i, \dots, x'_k, \dots, x'_N)\end{aligned}$$

Nous avons

$$\begin{aligned}PL(s, s') &= \prod_{i=1}^N PL_i(x_i, x'_i) \\ P(s, s') &= (PE * PL)(s, s') = \int_{\mathcal{X}^N} PE(\tilde{s}, s') PL(s, \tilde{s}) d\tilde{s}\end{aligned}$$

La norme en variation totale pour une mesure signée bornée μ sur $(\mathcal{X}^N, \mathcal{B}(\mathcal{X})^N)$ est définie par :

$$\|\mu\|_{TV} = \sup_{A \in \mathcal{B}(\mathcal{X})^N} |\mu(A)|$$

Preuve Proposition 1

1. PL réversible et de distribution stationnaire π^* .
2. PE réversible et de distribution stationnaire π^* (utilisation de la detailed balance condition).
3. P réversible et de distribution stationnaire π^* .
4. PL et P λ_N -irréductibles et apériodiques.
5. P est λ_N -irréductible, apériodique et de distribution stationnaire π^* . Théorème 1 de Tierney (1994) : S converge vers sa distribution stationnaire π^* au sens de la norme en variation totale, et pour presque tout $s \in \mathcal{X}^N$.

Preuve Lemme

1. D'après Prop 1, S est réversible de distribution stationnaire π^* , and fortement λ_N -irréductible \Rightarrow positivité.
2. Un état s' atteint depuis n'importe quel état de départ s ne peut pas appartenir à un ensemble $A \in \mathcal{X}^N$ tel que $\pi^*(A) = 0$.
3. Théorème 2 de Tierney (1994) : *S est Harris-récurrente si et seulement si les seules fonctions harmoniques bornées de la chaîne sont les fonctions constantes.*

Une fonction harmonique h pour S vérifie :

$$\mathbb{E}(h(S_n)|s_0) = \mathbb{E}(h(S_1)|s_0) = h(s_0) \quad \forall n \in \mathbb{N}$$

4. h est π^* -presque partout constante et égale à $\mathbb{E}_{\pi^*}(h(S))$.

Théorème 6.80 de Robert et Casella (2004), inspiré de Athreya *et al.* (1996) :

$$\lim_{n \rightarrow \infty} \sup_{A \in \mathcal{B}(\mathcal{X})^N} \left| \int_A P^n(s_0, s) ds - \int_A \pi^*(s) ds \right| = 0 \quad \pi^* - pp$$

En utilisant

$$\|\mu\|_{TV} = \sup_{A \in \mathcal{B}(\mathcal{X})^N} |\mu(A)| = \frac{1}{2} \sup_{|h| < 1} \left| \int h(x) \mu(dx) \right|,$$

on en déduit que pour h bornée,

$$\lim_{n \rightarrow \infty} \sup \left| \mathbb{E}[h(S_n)|s_0] - \mathbb{E}_{\pi^*}[h(s)] \right| = 0 \quad \pi^* - pp$$

5. h est partout constante et égale à $\mathbb{E}_{\pi^*}(h(S))$. Soit s_0 quelconque,

$$\begin{aligned} \mathbb{E}(h(S_1)|s_0) &= \int_{\mathcal{X}^N} P(s_0, s_1) h(s_1) ds_1 \\ &= \int_{\mathcal{X}^N} r(s_0) M(s_0, s_1) h(s_1) ds_1 + (1 - r(s_0)) h(s_0) \end{aligned}$$

Avec $E(h(S_1)|s_0) = h(s_0)$ et $h(s_1) = \mathbb{E}_{\pi^*}(h(S)) \Rightarrow h(s_0) = \mathbb{E}_{\pi^*}(h(S))$.