Adaptive and non-linear MCMC algorithms

Eric MOULINES

Telecom Paris Tech
CNRS - LTCI

Joint work with Y. ATCHADE (Univ. Michigan, USA), G. FORT (TELECOM ParisTech, France) and P. PRIOURET (Univ. Paris VI, France)

Outline

1. MCMC algorithms are a flexible family of algorithms to sample distributions, known up to a normalisation factor,
2. This flexibility comes at a price... badly tuned MCMC can be very slow to converge and the convergence may be difficult to diagnose.
3. In the last 10 years, several classes of algorithms have been introduced to *increase* the sampling efficiency of the MCMC, without demanding much additional user supervision. The common idea is to let the algorithms **self-learned** from the past simulations by **adapting** its parameters
4. **Problem :** the Markov property is not retained and the convergence is more difficult to study
5. **Today :** the basic ingredients of successful adaptations.

Adaptive SRWM

- All MCMC algorithms depend on some design parameters...
- for the **S**ymmetric **R**andom **W**alk **M**etropolis (SRWM) with normal proposal distribution, the design parameter is the covariance Σ_q of the Gaussian proposal.
- **Idea** : Tune these design parameters **on the fly**

Adaptive SRWM

- The **scaling** technique (the dimension d of the space $\to \infty$), suggests to set the covariance of the proposal Σ_q proportional to Σ_π.
- In practice Σ_π is unknown : at iteration n, replace it by an estimate obtained from the last simulated samples $\{X_k, \ k \leq n\}$.
- P_θ : kernel of a SRWM algorithm with proposal $\mathcal{N}_d(0, \theta)$,
- Iteration n
  - draw : $X_{n+1} \sim P_{\theta_n}(X_n, \cdot)$
  - update : $\theta_{n+1} = \phi_n\,(\theta_n, X_{n+1})$.

Adaptive MCMC

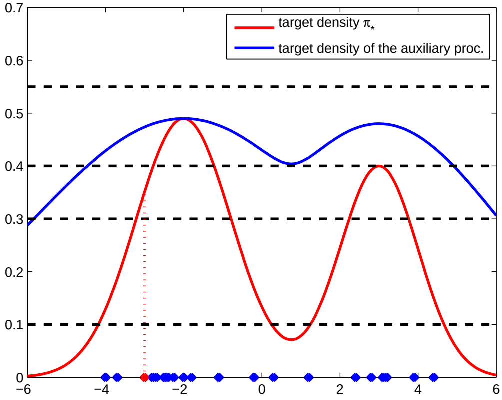
- A family of transition kernels $\{P_\theta, \theta \in \Theta\}$ such that, for all $\theta \in \Theta$, the target distribution $\pi_\star$ is the stationary distribution of P_θ : $\pi_\star P_\theta = \pi_\star$.
- An adaptive MCMC algorithm : process $\{(X_n, \theta_n), n \geq 0\}$ on the product space $\mathsf{X} \times \Theta$:
  - **Sampling** : given the past, draw

$$X_{n+1} \sim P_{\theta_n}(X_n, \cdot)$$

  - **Internal adaptation** : update the **parameter** θ_n from the **past** values of the X and θ

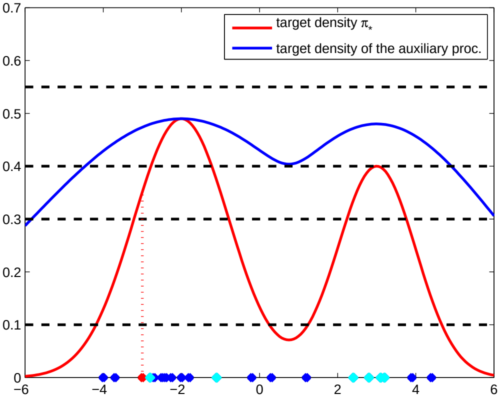
Interacting MCMC

- a transition kernel P s.t. $\pi_\star P = \pi_\star$
- a probability of swap $\epsilon \in (0,1)$
- an auxiliary process $\{Y_n, n \geq 0\}$ targeting a **tempered** version $\pi_\star^\beta$



Interacting MCMC

- a transition kernel P s.t. $\pi_\star P = \pi_\star$
- a probability of swap $\epsilon \in (0,1)$
- an auxiliary process $\{Y_n, n \geq 0\}$ targeting a **tempered** version $\pi_\star^\beta$



Interacting MCMC

- a transition kernel P s.t. $\pi_\star P = \pi_\star$
- a probability of swap $\epsilon \in (0,1)$
- an auxiliary process $\{Y_n, n \geq 0\}$ targeting a **tempered** version $\pi_\star^\beta$

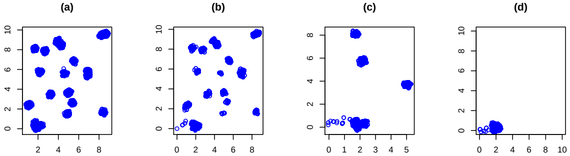


FIGURE: Example : Mixture of a 2D-Normal distribution [target / EE / Parallel Tempering / SRWM]

Interacting MCMC

- a transition kernel P s.t. $\pi_\star P = \pi_\star$
- a probability of swap $\epsilon \in (0,1)$
- an auxiliary process $\{Y_n, n \geq 0\}$ targeting a **tempered** version $\pi_\star^\beta$
- Iteration n :

(a) with probability $(1-\epsilon)$ draw $X_{n+1} \sim P(X_n, \cdot)$

$$P_{\theta_n}(X_n, A) = (1-\epsilon) P(X_n, A) + \cdots$$

Interacting MCMC

- a transition kernel P s.t. $\pi_\star P = \pi_\star$
- a probability of swap $\epsilon \in (0,1)$
- an auxiliary process $\{Y_n, n \geq 0\}$ targeting a **tempered** version $\pi_\star^\beta$
- Iteration n :

(b) with probability ϵ, **draw** a point $Y_\star$ among $\{Y_1, \cdots, Y_n\}$ and **accept/reject** with probability $\alpha(X_n, Y_\star)$

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

where

$$\theta_n(dy) = \frac{1}{n} \sum_{k=1}^{n} \delta_{Y_k}(dy).$$

Non-linear MCMC

- Construct an auxiliary process $\{Y_n, n \geq 0\}$ s.t. its empirical process $\lim_n \theta_n$ converges in some appropriate sense to $\tilde{\pi}(\cdot)$ so that asymptotically,

$$P_{\theta_n} \approx P_{\tilde{\pi}}$$

- The acceptance ratio $\alpha(x, y)$ of the interaction is chosen s.t. $\pi_\star \; P_{\tilde{\pi}} = \pi_\star$
- **Heuristic** :
  1. if these two conditions are satisfied, then the distribution of $(X_k)_{k \geq 0}$ converges to $\pi_\star$ as $k \to \infty$.
  2. **wishful thinking** The convergence might be faster in $\tilde{\pi}$ is well chosen and if the convergence of the auxiliary process is fast.

Non-linear MCMC : refinements

When sampling from the past of the auxiliary process, select the points : introduce a **selection** $g(x, y)$ function (satisfying $g(x, y) = g(y, x)$)

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

where

$$\theta_n(dy) = \frac{1}{n} \sum_{k=1}^{n} \delta_{Y_k}(dy) \qquad \alpha(x, y) = 1 \wedge \frac{\pi(y)\ \tilde{\pi}(x)}{\tilde{\pi}(y)\ \pi(x)}$$

Non-linear MCMC : refinements

When sampling from the past of the auxiliary process, select the points : introduce a **selection** $g(x,y)$ function (satisfying $g(x,y) = g(y,x)$)
This yields :

$$P_{\theta_n}(X_n, A) = (1-\epsilon_{\theta_n}(x))P(X_n, A)+\epsilon_{\theta_n}(x) \left\{ \int_A \frac{g(x,y)\theta_n(dy)}{\int g(x,y)\theta_n(dy)} \alpha(X_n, y) + \mathbb{1}_A(X_n) \int \frac{g(x,y)\theta_n(dy)}{\int g(x,y)\theta_n(dy)} \{1 - \alpha(X_n, y)\} \right\}$$

where

$$\theta_n(dy) = \frac{1}{n}\sum_{k=1}^{n} \delta_{Y_k}(dy) \qquad \alpha(x,y) = 1 \wedge \frac{\pi(y)\ \tilde{\pi}(x)}{\tilde{\pi}(y)\ \pi(x)} \qquad \epsilon_\theta(x) := \epsilon \mathbb{1}_{\int \theta(dy) g(x,y)}$$

Non-linear MCMC

- A family of transition kernels $\{P_\theta, \theta \in \Theta\}$ with invariant probability distribution π_θ : $\pi_\theta P_\theta = \pi_\theta$
- A non-linear MCMC is a process $\{(X_n, \theta_n), n \geq 0\}$ on the product space $\mathsf{X} \times \Theta$ defined as
  - **Simulation** Given the past , draw

$$X_{n+1} \sim P_{\theta_n}(X_n, \cdot)$$

  - **External adaptation** update the **parameter** θ_n (here, a probability distribution) according to

$$\theta_{n+1} \longleftrightarrow \text{computed from } \underline{\text{an auxiliary}} \text{ process } \{Y_k, k \leq n\}$$

Adaptive and non-linear MCMC in a nutshell

- A family of transition kernels $\{P_\theta, \theta \in \Theta\}$ with invariant distribution : $\pi_\star$ (**internal adaptation**) or π_θ (**external adaptation**).

We define a filtration $\mathcal{F}_n$, and a process $\{(X_n, \theta_n), n \geq 0\}$ s.t.

- component θ_n : $\mathcal{F}_n$ adapted with **internal / external** adaptation
- component X_n (process of interest) :

$$\mathbb{E}\left[f(X_{n+1}) \middle| \mathcal{F}_n \right] = \int P_{\theta_n}(X_n, dy) \; f(y).$$

Convergence of the marginals

- **Key ingredients** to prove the ergodicity of an MCMC algorithms :
  1. Markov Chain
  2. the transition kernel is **reversible** w.r.t the target distribution
- These properties are lost when adapting the algorithms...
- **Questions :** Conditions to guarantee that the **adaptation does not destroy the convergence** ?

Adaptive MCMC : $\pi_\theta = \pi_\star$

$$
\begin{aligned}
\mathbb{E}\left[f(X_n)\right] &= \mathbb{E}\left[\mathbb{E}\left[f(X_n)|\mathcal{F}_{n-N}\right]\right] \\
&= \mathbb{E}\Bigg[\underbrace{\mathbb{E}\left[f(X_n)|\mathcal{F}_{n-N}\right] - P^N_{\theta_{n-N}} f(X_{n-N})}_{\text{comparison with a frozen chain with transition } P_{\theta_{n-N}}} \\
&\qquad + \underbrace{P^N_{\theta_{n-N}} f(X_{n-N}) - \pi_\star(f)}_{\text{ergodicity of the frozen chain}}\Bigg] + \pi_\star(f).
\end{aligned}
$$

Diminishing adaptation

$$\sup_x \|P_{\theta_n}(x,\cdot) - P_{\theta_{n-1}}(x,\cdot)\|_{\mathrm{TV}} \longrightarrow_{\mathbb{P}} 0$$

- Generally problem specific
- ... But most often, amounts to check a condition of the type

$$\sup_x \|P_{\theta_n}(x,\cdot) - P_{\theta_{n-1}}(x,\cdot)\|_{\mathrm{TV}} \leq C \; \|\theta_n - \theta_{n-1}\|_{\mathbf{xxx}}$$

so that convergence in probability is implied by the **adaptation scheme**.

Containment condition

$$\lim_M \limsup_n \mathbb{P}\left(M_\epsilon(X_n, \theta_n) \geq M\right) = 0,$$
$$M_\epsilon(x, \theta) := \inf\{n \geq 1, \|P_\theta^n(x, \cdot) - \pi_\star\|_{\mathrm{TV}} \leq \epsilon\}$$

- Most often, deduced from **ergodicity** + homogeneity
- The easy case is when the ergodicity is **uniform** in θ :

$$\sup_\theta \|P_\theta^n(x, \cdot) - \pi_\star\|_{\mathrm{TV}} \leq \rho(n)\ U(x) \qquad \lim_n \rho(n) = 0$$

then

$$M_\epsilon(x, \theta) \leq \rho^{-1}\left(\epsilon C^{-1} U^{-1}(x)\right).$$

Adaptive MCMC

Theorem

Assume

1. **(Diminishing adaptation)**

$$\sup_x \|P_{\theta_n}(x,\cdot) - P_{\theta_{n-1}}(x,\cdot)\|_{\mathrm{TV}} \longrightarrow_{\mathbb{P}} 0$$

2. **(Containment condition)**

$$\lim_M \limsup_n \mathbb{P}\left(M_\epsilon(X_n,\theta_n) \geq M\right) = 0.$$

Then

$$\lim_n \sup_{f,|f|_\infty \leq 1} |\mathbb{E}\left[f(X_n)\right] - \pi_\star(f)| = 0$$

Non-Linear MCMC : $\pi_\theta P_\theta = \pi_\theta$

$$
\begin{aligned}
\mathbb{E}\left[f(X_n)\right] &= \mathbb{E}\left[\mathbb{E}\left[f(X_n)|\mathcal{F}_{n-N}\right]\right] \\
&= \mathbb{E}\Bigg[\underbrace{\mathbb{E}\left[f(X_n)|\mathcal{F}_{n-N}\right] - P^N_{\theta_{n-N}} f(X_{n-N})}_{\text{comparison with a frozen chain with transition } P_{\theta_{n-N}}} \\
&\quad + \underbrace{P^N_{\theta_{n-N}} f(X_{n-N}) - \pi_{\theta_{n-N}}(f)}_{\text{ergodicity of the frozen chain}} \\
&\quad + \pi_{\theta_{n-N}}(f) - \pi_\star(f)\Bigg] + \pi_\star(f).
\end{aligned}
$$

Non-Linear MCMC : $\pi_\theta P_\theta = \pi_\theta$

$$
\begin{aligned}
\mathbb{E}\left[f(X_n)\right] &= \mathbb{E}\left[\mathbb{E}\left[f(X_n)|\mathcal{F}_{n-N}\right]\right] \\
&= \mathbb{E}\Bigg[\underbrace{\mathbb{E}\left[f(X_n)|\mathcal{F}_{n-N}\right] - P^N_{\theta_{n-N}} f(X_{n-N})}_{\text{comparison with a frozen chain with transition } P_{\theta_{n-N}}} \\
&\quad + \underbrace{P^N_{\theta_{n-N}} f(X_{n-N}) - \pi_{\theta_{n-N}}(f)}_{\text{ergodicity of the frozen chain}} \\
&\quad + \pi_{\theta_{n-N}}(f) - \pi_\star(f)\Bigg] + \pi_\star(f).
\end{aligned}
$$

- (same) : Diminishing adaptation, Containment condition
- **Convergence of the invariant measures** $\{\pi_{\theta_n}, n \geq 0\}$ to some $\pi_\star$

Non linear MCMC

Theorem

Assume

1. *(Diminishing adaptation)*

$$\sup_x \|P_{\theta_n}(x,\cdot) - P_{\theta_{n-1}}(x,\cdot)\|_{\mathrm{TV}} \longrightarrow_{\mathbb{P}} 0$$

2. **(Containment condition)**

$$\lim_M \limsup_n \mathbb{P}\left(M_\epsilon(X_n,\theta_n) \geq M\right) = 0.$$

3. *(Convergence of the invariant distributions)*

$$\pi_{\theta_n}(f) - \pi_\star(f) \to_{\mathbb{P}} 0.$$

Then

$$\lim_n \; \left|\mathbb{E}\left[f(X_n)\right] - \pi_\star(f)\right| = 0$$

How to check these conditions ?

- (Convergence of the invariant distributions)

$$\pi_{\theta_n}(f) - \pi_\star(f) \to_{\mathbb{P}} 0.$$

We proved that if

(i) there exist x s.t.

$$\lim_n \sup_\theta \|P_\theta^n(x,\cdot) - \pi_\theta\|_{\mathrm{TV}} = 0,$$

(ii) there exist $\theta_\star \in \Theta$ and a set A such that $\mathbb{P}(A) = 1$ and

$$\forall \omega \in A, x \in \mathsf{X}, B \in \mathcal{B}(\mathsf{X}) \qquad \lim_n P_{\theta_n(\omega)}(x,B) = P_{\theta_\star}(x,B)$$

(iii) the state space X is Polish

then for any bounded function f,

$$\pi_{\theta_n}(f) \longrightarrow_{a.s.} \pi_{\theta_\star}(f)$$

Back to the Interacting MCMC

Let $\pi_\star$ be positive and continuous on X s.t. $\sup_{\mathsf{X}} \pi_\star < +\infty$.
Let $\beta \in (0, 1)$.

- **On the auxiliary process :**
- **On the transition kernel P :**
- **On the probability of swap ϵ :**

Back to the Interacting MCMC

Let $\pi_\star$ be positive and continuous on X s.t. $\sup_{\mathsf{X}} \pi_\star < +\infty$.
Let $\beta \in (0,1)$.

- **On the auxiliary process :** for any bounded function f,

$$\frac{1}{n}\sum_{k=1}^{n} f(Y_k) \longrightarrow_{a.s.} \pi_\star^\beta(f).$$

- **On the transition kernel P :**
- **On the probability of swap ϵ :**

Back to the Interacting MCMC

Let $\pi_\star$ be positive and continuous on X s.t. $\sup_{\mathsf{X}} \pi_\star < +\infty$.
Let $\beta \in (0,1)$.

- **On the auxiliary process :** for any bounded function f,

$$\frac{1}{n}\sum_{k=1}^{n} f(Y_k) \longrightarrow_{a.s.} \pi_\star^{\beta}(f).$$

- **On the transition kernel P :** P is phi-irreducible, $\pi_\star P = \pi_\star$, the level sets $\{\pi \geq p\}$ are 1-small and

$$PV(x) \leq \lambda V(x) + b\mathbb{1}_{\mathcal{C}}(x) \qquad V(x) = \left(\frac{\pi(x)}{\sup_{\mathsf{X}} \pi}\right)^{-\tau(1-\beta)}$$

for some $\lambda \in (0,1)$, $b < +\infty$, a set $\mathcal{C}$, $\tau \in (0,1]$.

- **On the probability of swap ϵ :**

Back to the Interacting MCMC

Let $\pi_\star$ be positive and continuous on X s.t. $\sup_{\mathsf{X}} \pi_\star < +\infty$.
Let $\beta \in (0,1)$.

- **On the auxiliary process :** for any bounded function f,

$$\frac{1}{n}\sum_{k=1}^{n} f(Y_k) \longrightarrow_{a.s.} \pi_\star^\beta(f).$$

- **On the transition kernel P :** P is phi-irreducible, $\pi_\star P = \pi_\star$, the level sets $\{\pi \geq p\}$ are 1-small and

$$PV(x) \leq \lambda V(x) + b\mathbb{1}_{\mathcal{C}}(x) \qquad V(x) = \left(\frac{\pi(x)}{\sup_{\mathsf{X}} \pi}\right)^{-\tau(1-\beta)}$$

for some $\lambda \in (0,1)$, $b < +\infty$, a set $\mathcal{C}$, $\tau \in (0,1]$.

- **On the probability of swap ϵ :**

$$0 \leq \epsilon < \frac{1-\lambda}{1-\lambda+\tau(1-\tau)^{(1-\tau)/\tau}}$$

Under these conditions,

- the diminishing adaptation condition holds
- a <u>uniform-in-θ drift</u> condition holds

$$\tilde{\lambda} \in (0,1), \qquad P_\theta V(x) \leq \tilde{\lambda} V(x) + b \mathbb{1}_{\mathcal{C}}(x),$$

  and we prove the containment condition.
- the invariant measures a.s. converge : $\lim_n \pi_{\theta_n}(f) = \pi_\star(f)$ a.s. for any bounded function.

Hence, for any bounded function f

$$\mathbb{E}\left[f(X_n)\right] \longrightarrow_n \pi_\star(f).$$

Strong LLN

Sufficient Conditions for the existence of $\pi_\star$ s.t. the strong LLN

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \longrightarrow_{a.s.} \pi_\star(f)$$

is satisfied for any function f in a (hopefully large) class of functions.

Idea : use the Poisson equation

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) - \pi_\star(f) = \underbrace{\frac{1}{n}\sum_{k=1}^{n}\{f(X_k) - \pi_{\theta_{k-1}}(f)\}}_{\text{“Poisson term”}} + \underbrace{\frac{1}{n}\sum_{k=1}^{n}\pi_{\theta_{k-1}}(f) - \pi_\star(f)}_{\text{Cesaro mean (is null when } \pi_\theta = \pi_\star)}$$

Idea : use the Poisson equation

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) - \pi_\star(f) = \underbrace{\frac{1}{n}\sum_{k=1}^{n}\{f(X_k) - \pi_{\theta_{k-1}}(f)\}}_{\text{“Poisson term”}} + \underbrace{\frac{1}{n}\sum_{k=1}^{n}\pi_{\theta_{k-1}}(f) - \pi_\star(f)}_{\text{Cesaro mean (is null when } \pi_\theta = \pi_\star)}$$

if

(i) uniform-in-θ V-ergodicity for some x,

$$\lim_n \sup_\theta \|P_\theta^n(x,\cdot) - \pi_\theta\|_V = 0,$$

(ii) There exist $\theta_\star \in \Omega_0$ and A s.t. $\mathbb{P}(\Omega_0) = 1$ and

$$\forall \omega \in A, x, B \qquad P_{\theta_n(\omega)}(x,B) \longrightarrow P_{\theta_\star}(x,B)$$

(iii) Polish state space X

then

$$\pi_{\theta_n}(f) \longrightarrow_{a.s.} \pi_{\theta_\star}(f) \qquad \text{for any } f \in \mathcal{L}_{V^\alpha},\ \alpha \in [0,1)$$

Decomposition

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

$$= \underbrace{n^{-1}\sum_{k=1}^{n}\{\hat{f}_{\theta_{k-1}}(X_k) - P_{\theta_{k-1}}\hat{f}_{\theta_{k-1}}(X_{k-1})\}}_{\text{martingale term}}$$

$$+ \underbrace{\frac{1}{n}\sum_{k=1}^{n-1}\{P_{\theta_k}\hat{f}_{\theta_k}(X_k) - P_{\theta_{k-1}}\hat{f}_{\theta_{k-1}}(X_k)\}}_{\text{Remainder term (I)}}$$

$$+ \underbrace{n^{-1}\{P_{\theta_0}f_{\theta_0}(X_0) - P_{\theta_{n-1}}f_{\theta_{n-1}}(X_{n-1})\}}_{\text{Remainder term (II)}}$$

where $\hat{f}_\theta$ solves $$f - \pi_\theta(f) = \hat{f}_\theta - P_\theta\hat{f}_\theta.$$

Decomposition

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

$$= n^{-1}\underbrace{\sum_{k=1}^{n}\{\hat{f}_{\theta_{k-1}}(X_k)-P_{\theta_{k-1}}\hat{f}_{\theta_{k-1}}(X_{k-1})\}}_{\text{martingale term}}$$

$$+\underbrace{\frac{1}{n}\sum_{k=1}^{n-1}\{P_{\theta_k}\hat{f}_{\theta_k}(X_k)-P_{\theta_{k-1}}\hat{f}_{\theta_{k-1}}(X_k)\}}_{\text{Remainder term (I)}}$$

$$+\underbrace{n^{-1}\{P_{\theta_0}f_{\theta_0}(X_0)-P_{\theta_{n-1}}f_{\theta_{n-1}}(X_{n-1})\}}_{\text{Remainder term (II)}}$$

where $\hat{f}_\theta$ solves $$f-\pi_\theta(f)=\hat{f}_\theta-P_\theta\hat{f}_\theta.$$

- a.s. convergence of the martingale : conditions on the L^p-moments of the increment $\hookrightarrow$ **uniform-in-θ drift conditions** on the

Decomposition

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

$$= n^{-1}\underbrace{\sum_{k=1}^{n}\{\hat{f}_{\theta_{k-1}}(X_k)-P_{\theta_{k-1}}\hat{f}_{\theta_{k-1}}(X_{k-1})\}}_{\text{martingale term}}$$

$$+\underbrace{\frac{1}{n}\sum_{k=1}^{n-1}\{P_{\theta_k}\hat{f}_{\theta_k}(X_k)-P_{\theta_{k-1}}\hat{f}_{\theta_{k-1}}(X_k)\}}_{\text{Remainder term (I)}}$$

$$+\underbrace{n^{-1}\{P_{\theta_0}f_{\theta_0}(X_0)-P_{\theta_{n-1}}f_{\theta_{n-1}}(X_{n-1})\}}_{\text{Remainder term (II)}}$$

where $\hat{f}_\theta$ solves $$f-\pi_\theta(f)=\hat{f}_\theta-P_\theta\hat{f}_\theta.$$

- a.s. convergence of the martingale : conditions on the L^p-moments of the increment $\hookrightarrow$ **uniform-in-θ drift conditions** on the

Define

$$D_V(\theta, \theta') := \sup_x \frac{\|P_\theta(x, \cdot) - P_{\theta'}(x, \cdot)\|_V}{V(x)}$$

Theorem

Assume

(i) **(uniform ergodic behavior)** P_θ *is phi-irreducible,*

$$P_\theta V \leq \lambda V + b \mathbb{1}_\mathcal{C} \qquad \lambda \in (0,1), b < +\infty,$$

and level sets of V *are* 1*-small.*

(ii) *(strenghtened D.A.)* $\sum_k \frac{1}{k} V^\alpha(X_k) \;\; D_{V^\alpha}(\theta_k, \theta_{k-1}) < +\infty$ *a.s.*

(iii) *(convergence of the invariant measures)*

Then : if $\mathbb{E}[V(X_0)] < \infty$*, for any* $\alpha \in [0,1)$ *and any* $f \in \mathcal{L}_{V^\alpha}$

$$\frac{1}{n} \sum_{k=1}^n f(X_k) \longrightarrow_{a.s.} \pi_\star(f),$$

Conclusion : when applied to the Equi-Energy sampler

Let $\pi_\star$ be positive and continuous on X s.t. $\sup_{\mathsf{X}} \pi_\star < +\infty$.
Let $\beta \in (0,1)$.

- On the transition kernel P :
- On the probability of swap ϵ :
- On the auxiliary process :

Conclusion : when applied to the Equi-Energy sampler

Let $\pi_\star$ be positive and continuous on X s.t. $\sup_{\mathsf{X}} \pi_\star < +\infty$.
Let $\beta \in (0, 1)$.

- On the transition kernel P : (same as those for the convergence of the marginals)
- On the probability of swap ϵ :
- On the auxiliary process :

Conclusion : when applied to the Equi-Energy sampler

Let $\pi_\star$ be positive and continuous on X s.t. $\sup_{\mathsf{X}} \pi_\star < +\infty$.
Let $\beta \in (0, 1)$.

- On the transition kernel P : (same as those for the convergence of the marginals)
- On the probability of swap ϵ : (same as those for the convergence of the marginals)
- On the auxiliary process :

Conclusion : when applied to the Equi-Energy sampler

Let $\pi_\star$ be positive and continuous on X s.t. $\sup_{\mathsf{X}} \pi_\star < +\infty$.
Let $\beta \in (0,1)$.

- On the transition kernel P : (same as those for the convergence of the marginals)
- On the probability of swap ϵ : (same as those for the convergence of the marginals)
- On the auxiliary process : **for any** $\alpha \in [0,1)$ **and** $f \in \mathcal{L}_{V^\alpha}$

$$\frac{1}{n}\sum_{k=1}^{n} f(Y_k) \longrightarrow_{a.s.} \pi_\star^\beta(f).$$

Conclusion : when applied to the Equi-Energy sampler

Let $\pi_\star$ be positive and continuous on X s.t. $\sup_{\mathsf{X}} \pi_\star < +\infty$.
Let $\beta \in (0, 1)$.

- On the transition kernel P : (same as those for the convergence of the marginals)
- On the probability of swap ϵ : (same as those for the convergence of the marginals)
- On the auxiliary process : **for any $\alpha \in [0, 1)$ and $f \in \mathcal{L}_{V^\alpha}$**

$$\frac{1}{n} \sum_{k=1}^{n} f(Y_k) \longrightarrow_{a.s.} \pi_\star^\beta(f).$$

Note that : it is assumed that a strong LLN holds for the auxiliary process and any function $f \in \mathcal{L}_{V^\alpha}$, $\alpha \in (0, 1)$; in order to prove a strong LLN for the process of interest and any function $f \in \mathcal{L}_{V^\alpha}$, $\alpha \in (0, 1)$.

$\hookrightarrow$ repeat the mecanism and prove the convergence of the marginals + a strong LLN for the K-levels Equi-Energy sampler

Conclusion of the talk

- We prove convergence of the marginals for general adaptive MCMC samplers with the main ingredients
  - diminishing adaptation
  - ergodicity of the kernels + some form of uniformity in θ
  - For external adaptation : a.s. convergence of the invariant measures π_{θ_n}
- Under the same assumptions, a L.L.N can be established.