Adaptive MCMC : theory and methods

Gersende FORT

LTCI
CNRS - TELECOM ParisTech

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

We will provide sufficient conditions so that

the process $\{X_n, n \geq 0\}$ produced by an adaptive MCMC sampler approximates a target density $\pi_\star$ i.e.

- (convergence of the marginals) for any bounded function f

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

We will provide sufficient conditions so that

the process $\{X_n, n \geq 0\}$ produced by an adaptive MCMC sampler approximates a target density $\pi_\star$ i.e.

- (convergence of the marginals) for any bounded function f

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

- (strong LLN) for any function f in a large class of functions

$$\frac{1}{n} \sum_{k=1}^{n} f(X_k) \longrightarrow \pi_\star(f) \qquad \mathbb{P} - a.s.$$

1. Examples
2. Sufficient conditions for convergence of the marginals
3. Sufficient conditions for strong LLN

Example 1 : Adaptive SRWM

- MCMC depends on some design parameters.

Ex. for the **S**ymmetric **R**andom **W**alk **M**etropolis (SRWM) with normal proposal distribution, the design parameter is the variance Σ_q of the Gaussian proposal.

Example 1 : Adaptive SRWM

- MCMC depends on some design parameters.

Ex. for the **S**ymmetric **R**andom **W**alk **M**etropolis (SRWM) with normal proposal distribution, the design parameter is the variance Σ_q of the Gaussian proposal.

- Tune these design parameters "on the fly", during the run of the algorithm.

Ex. (to follow)

based on results obtained by the scaling technique, choose $\Sigma_q \propto \Sigma_\pi$.

usually, Σ_π is unknown : at iteration n, replace it by an estimation computed with the samples $\{X_k, \, k \leq n\}$.

This yields the adaptive SRWM

- 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 the estimate of Σ_π : $\qquad \theta_{n+1} = \phi_n(\theta_n,X_{n+1})$.

This is an example of the following general framework :

- let a family of transition kernels $\{P_\theta, \theta \in \Theta\}$
- with the same invariant probability distribution $\pi_\star$.

- define a process $\{(X_n, \theta_n), n \geq 0\}$ as follows
  - given the past (a filtration $\mathcal{F}_n$), draw

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

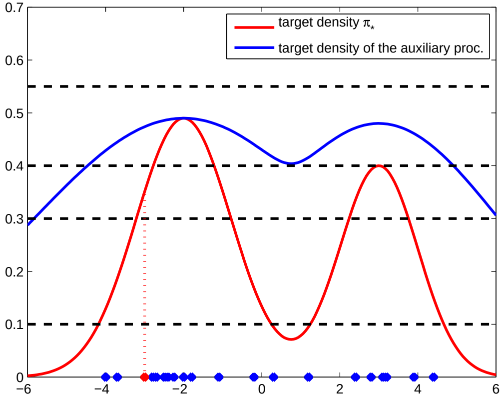
  - update the 'parameter' with an "internal adaptation" scheme

$$\theta_{n+1} \longleftrightarrow \text{built from the process } \{X_k, k \leq n\} \text{ itself}$$

Example 2: Equi-Energy sampler

- Given

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



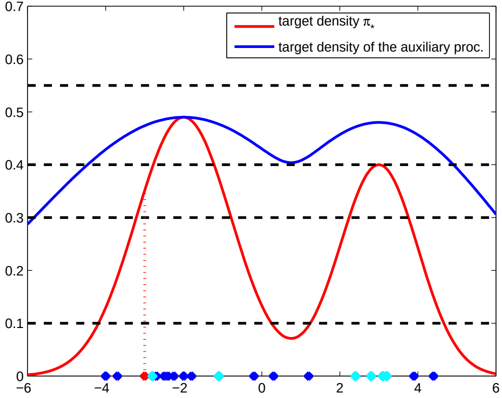
Example 2: Equi-Energy sampler

- Given

  a transition kernel P s.t. $\pi_\star P = \pi_\star$

  a probability of swap $\epsilon \in (0,1)$

  an auxiliary process $\{Y_n, n \geq 0\}$ target : $\pi_\star^\beta$



Example 2: Equi-Energy sampler

- Given

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

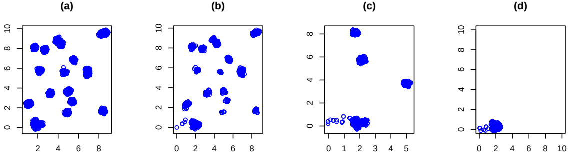


FIG.: Example : Mixture of a 2D-Normal distribution [target / EE / Parallel Tempering / SRWM]

Example 2: Equi-Energy sampler

- Given

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

- Iteration n :

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

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

Example 2: Equi-Energy sampler

- Given

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

- Iteration n :

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

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

where

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

In practice

- Choose the auxiliary process $\{Y_n, n \geq 0\}$ such that $\lim_n \theta_n = \tilde{\pi}$ in some sense, so that asymptotically, "$P_{\theta_n} \approx P_{\tilde{\pi}}$".
- Choose the acceptation-rejection mecanism $\alpha(x,y)$ so that $\pi_\star P_{\tilde{\pi}} = \pi_\star$, so that asymptotically, "$\pi_\star$ is invariant for P_{θ_n}".
- When sampling in the past of the auxiliary process, select the points : introduce a selection $g(x,y)$ (such that $g(x,y) = g(y,x)$)

In practice

- Choose the auxiliary process $\{Y_n, n \geq 0\}$ such that $\lim_n \theta_n = \tilde{\pi}$ in some sense, so that asymptotically, "$P_{\theta_n} \approx P_{\tilde{\pi}}$".
- Choose the acceptation-rejection mecanism $\alpha(x,y)$ so that $\pi_\star \, P_{\tilde{\pi}} = \pi_\star$, so that asymptotically, "$\pi_\star$ is invariant for P_{θ_n}".
- When sampling in the past of the auxiliary process, select the points : introduce a selection $g(x,y)$ (such that $g(x,y) = g(y,x)$)

This yields :

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

where

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

In practice

- Choose the auxiliary process $\{Y_n, n \geq 0\}$ such that $\lim_n \theta_n = \tilde{\pi}$ in some sense, so that asymptotically, "$P_{\theta_n} \approx P_{\tilde{\pi}}$".
- Choose the acceptation-rejection mecanism $\alpha(x,y)$ so that $\pi_\star \; P_{\tilde{\pi}} = \pi_\star$, so that asymptotically, "$\pi_\star$ is invariant for P_{θ_n}".
- When sampling in the past of the auxiliary process, select the points : introduce a selection $g(x,y)$ (such that $g(x,y) = g(y,x)$)

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

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

This is an example of the following general framework :

- let a family of transition kernels $\{P_\theta, \theta \in \Theta\}$
- with their own invariant probability distribution π_θ : $\pi_\theta P_\theta = \pi_\theta$

- define a process $\{(X_n, \theta_n), n \geq 0\}$ as follows
  - given the past (a filtration $\mathcal{F}_n$), draw

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

  - update the 'parameter' with an "external adaptation" scheme

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

Conclusion of Section I

We have

- a family of transition kernels $\{P_\theta, \theta \in \Theta\}$,
- with invariant distribution : π_θ or $\pi_\star$.

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**):

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

1. Examples
2. Sufficient conditions for convergence of the marginals
3. Sufficient conditions for strong LLN

Suff Cond for : the existence of $\pi_\star$ s.t.

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

for any **bounded** function f.

Idea, when $\forall \theta, \quad \pi_\theta = \pi_\star$

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

Idea, when $\forall\theta, \quad \pi_\theta = \pi_\star$

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

Conditions on

- (Diminishing adaptation) two successive transition kernels are similar : "$\|P_{\theta_n}(x,\cdot) - P_{\theta_{n-1}}(x,\cdot)\|_{\mathrm{TV}} \to 0$"
- (Containment condition) ergodicity of the transition kernel " $\|P^n_\theta(x,\cdot) - \pi_\star\|_{\mathrm{TV}} \to 0$ *uniformly*"

Result, when $\forall \theta, \quad \pi_\theta = \pi_\star$

Define

$$M_\epsilon(x,\theta) := \inf\{n \geq 1, \|P_\theta^n(x,\cdot) - \pi_\star\|_{\mathrm{TV}} \leq \epsilon\}$$

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

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

How to check these conditions?

- (Diminishing adaptation)

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

$\hookrightarrow$ Problem specific.
For ex. we can have

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

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

How to check these conditions?

- (Containment condition)

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

$\hookrightarrow$ usually, deduced from **uniform-in-θ ergodicity** : if

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

then

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

Hence : Containment Cond is proved if $\tilde{U}(X_n)$ **is bounded in probability**.

It can thus be proved from uniform-in-θ conditions of the form :

- $\exists \varepsilon > 0, \nu, \mathcal{C} \qquad P_\theta(x,\cdot) \geq \varepsilon \nu(\cdot) \mathbb{1}_\mathcal{C}(x)$
- $P_\theta V(x) \leq V(x) - \phi \circ V(x) + b \mathbb{1}_\mathcal{C}(x)$.

Idea, when $\pi_\theta P_\theta = \pi_\theta$

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

Idea, when $\pi_\theta P_\theta = \pi_\theta$

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

Conditions on

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

Result when $\pi_\theta P_\theta = \pi_\theta$

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,

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

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

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 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:

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 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:

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 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

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

1. Examples
2. Sufficient conditions for convergence of the marginals
3. Sufficient conditions for strong LLN

Suff Cond for : the existence of $\pi_\star$ s.t.

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

for any function f in a 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

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

About the convergence of the invariant measures: we prove that

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 $\theta_\star \in \Theta$ and A s.t. $\mathbb{P}(A) = 1$ 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

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

About the "Poisson" term we write

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

$$+\underbrace{\frac{1}{n}\sum_{k=1}^{n-1}\{P_{\theta_k}\hat{f}_{\theta_k}(X_k)-P_{\theta_{k-1}}\hat{f}_{\theta_{k-1}}(X_k)\}}_{\text{Remainder term (I)}}+\underbrace{n^{-1}\{P_{\theta_0}f_{\theta_0}(X_0)-P_{\theta_{n-1}}f_{\theta_{n-1}}(X_{n-1})\}}_{\text{Remainder term (II)}}$$

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

About the "Poisson" term we write

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

$$+\underbrace{\frac{1}{n}\sum_{k=1}^{n-1}\{P_{\theta_k}\hat{f}_{\theta_k}(X_k)-P_{\theta_{k-1}}\hat{f}_{\theta_{k-1}}(X_k)\}}_{\text{Remainder term (I)}}+\underbrace{n^{-1}\{P_{\theta_0}f_{\theta_0}(X_0)-P_{\theta_{n-1}}f_{\theta_{n-1}}(X_{n-1})\}}_{\text{Remainder term (II)}}$$

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 $\hookrightarrow$ in practice, **uniform-in-θ drift conditions** on the kernels P_θ.

About the "Poisson" term we write

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

$$+\underbrace{\frac{1}{n}\sum_{k=1}^{n-1}\{P_{\theta_k}\hat f_{\theta_k}(X_k)-P_{\theta_{k-1}}\hat f_{\theta_{k-1}}(X_k)\}}_{\text{Remainder term (I)}}+\underbrace{n^{-1}\{P_{\theta_0}f_{\theta_0}(X_0)-P_{\theta_{n-1}}f_{\theta_{n-1}}(X_{n-1})\}}_{\text{Remainder term (II)}}$$

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 $\hookrightarrow$ in practice, **uniform-in-θ drift conditions** on the kernels P_θ.
- a.s. convergence of the remainder terms : regularity in θ of the solution to the Poisson equation $\hookrightarrow$ in practice, **strenghtened diminishing adaptation condition**.

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 + strong LLN for general adaptive MCMC samplers with the main ingredients
  - (strenghtened) diminishing adaptation
  - "uniform" ergodic behavior of the kernels
  - `when` $\pi_\theta \neq \pi_\star$: a.s. convergence of the invariant measures π_{θ_n}

  .
- And illustrate the conditions by considering the Equi-Energy sampler.

Conclusion of the talk

- We prove convergence of the marginals + strong LLN for general adaptive MCMC samplers with the main ingredients
  - (strenghtened) diminishing adaptation
  - "uniform" ergodic behavior of the kernels
  - when $\pi_\theta \neq \pi_\star$: a.s. convergence of the invariant measures π_{θ_n}

  .
- And illustrate the conditions by considering the Equi-Energy sampler.
- Extensions (not discussed here) : uniform-in-θ ergodicity conditions have been proved by showing that the transition kernels are **geometrically ergodic**. We also provide examples in which they are only **sub-geometrically ergodic**. For ex. in the case

$$P_\theta V \leq V - c\, V^{1-\alpha} + b\mathbb{1}_{\mathcal{C}}$$

  we prove a strong LLN for functions increasing like V^β for any $\beta \in [0,1-\alpha)$.