The Metropolis Hastings algorithm : introduction and optimal scaling of the transient phase

Benjamin Jourdain

Joint work with T. Lelièvre and B. Miasojedow
Summer school CEMRACS 2017

July 19 2017

Outline of the talk

Motivation

Simulation according to a measure $\pi(dx) = \frac{\eta(x)\lambda(dx)}{\int_E \eta(y)\lambda(dy)}$ on E where

- λ is a reference measure on $(E, \mathcal{E})$,
- $\eta : E \to \mathbb{R}_+$ is measurable and such that $\int_E \eta(x)\lambda(dx) \in (0, \infty)$.

Examples

- Statistical physics : simulation according to the Boltzmann-Gibbs probability measure with density proportional to $\eta(x) = e^{-\frac{1}{k_B T} U(x)}$ w.r.t. the Lebesgue measure λ on $E = \mathbb{R}^n$ (k_B Boltzmann constant, T temperature, $U : \mathbb{R}^n \to \mathbb{R}$ potential function),
- Bayesian statistics : θ E-valued parameter with a priori density $p_\Theta(\theta)$ with respect to λ.
  Denoting by $p_{Y|\Theta}(y|\theta)$ the density of the observation Y when the parameter if θ, the a posteriori density of Θ is

$$\eta(\theta) = p_{\Theta|Y}(\theta|y) = \frac{p_{Y|\Theta}(y|\theta)p_\Theta(\theta)}{\int_E p_{Y|\Theta}(y|\vartheta)p_\Theta(\vartheta)\lambda(d\vartheta)}.$$

The computation of the normalizing constant is difficult in both cases.

The Metropolis Hastings algorithm

Let $q : E \times E \to \mathbb{R}_+$ be a mesurable function such that $\forall x \in E$,

- $\int_E q(x,y)\lambda(dy) = 1$,
- simulation according to the probability measure $q(x,y)\lambda(dy)$ is possible.

$$\text{Let } \alpha(x,y) = \begin{cases} \min\left(1, \frac{\eta(y)q(y,x)}{\eta(x)q(x,y)}\right) \text{ if } \eta(x)q(x,y) > 0 \\ 1 \text{ if } \eta(x)q(x,y) = 0 \end{cases}.$$

No need of the normalizing constant to compute α

Starting from an initial E-valued random variable X_0, construct a Markov chain $(X_k)_{k\in\mathbb{N}}$ by the following induction :

- Given $(X_0, \dots, X_k)$, one generates a proposal $Y_{k+1} \sim q(X_k, y)\lambda(dy)$ and an independent random variable $U_{k+1} \sim \mathcal{U}[0,1]$,
- One sets $X_{k+1} = Y_{k+1} 1_{\{U_{k+1} \leq \alpha(X_k, Y_{k+1})\}} + X_k 1_{\{U_{k+1} > \alpha(X_k, Y_{k+1})\}}$, i.e. the proposal is accepted with probability $\alpha(X_k, Y_{k+1})$ and otherwise the position X_k is kept.

Markov kernel of $(X_k)_k$

For $f : E \to \mathbb{R}$ measurable and bounded and $X_{0:k} = (X_0, X_1, \dots, X_k)$,

$$\begin{aligned}
&\mathbb{E}[f(X_{k+1})|X_{0:k}] \\
&= \mathbb{E}[\mathbb{E}[f(Y_{k+1})1_{\{U_{k+1}\leq\alpha(X_k,Y_{k+1})\}} + f(X_k)1_{\{U_{k+1}>\alpha(X_k,Y_{k+1})\}}|X_{0:k}, Y_{k+1}|X_{0:k}] \\
&= \mathbb{E}[f(Y_{k+1})\alpha(X_k, Y_{k+1}) + f(X_k)(1-\alpha(X_k, Y_{k+1}))|X_{0:k}] \\
&= \int_E f(y)\alpha(X_k, y)q(X_k, y)\lambda(dy) + f(X_k)\int_E (1-\alpha(X_k, y))q(X_k, y)\lambda(dy) \\
&= \int_E f(y)P(X_k, dy)
\end{aligned}$$

where
$$P(x, dy) = 1_{\{y\neq x\}}\alpha(x, y)q(x, y)\lambda(dy) + \left(\int_{E\setminus\{x\}} (1-\alpha(x, z))q(x, z)\lambda(dz) + q(x, x)\lambda(\{x\})\right)\delta_x(dy).$$

Thus $(X_k)_{k\in\mathbb{N}}$ is a Markov chain with kernel P.

Reversibility of π

For $y \neq x$,

$$\eta(x)q(x,y)\alpha(x,y) = \begin{cases} \eta(x)q(x,y)\min\left(1, \frac{\eta(y)q(y,x)}{\eta(x)q(x,y)}\right) \text{ if } \eta(x)q(x,y) > 0 \\ \eta(x)q(x,y) \times 1 \text{ if } \eta(x)q(x,y) = 0 \end{cases}$$
$$= \min(\eta(x)q(x,y), \eta(y)q(y,x)).$$

is a symmetric function of (x, y). As a consequence,

$$1_{\{x \neq y\}}\eta(x)\lambda(dx)P(x,dy) = 1_{\{x \neq y\}}\eta(x)q(x,y)\alpha(x,y)\lambda(dx)\lambda(dy)$$
$$= 1_{\{x \neq y\}}\eta(y)\lambda(dy)P(y,dx).$$

Since the equality clearly remains true with $1_{\{x=y\}}$ replacing $1_{\{x \neq y\}}$,

$$\pi(dx)P(x,dy) = \pi(dy)P(y,dx)$$

i.e. π is reversible for the Markov kernel P. This implies that $\int_{x \in E} \pi(dx)P(x,dy) = \int_{x \in E} \pi(dy)P(y,dx) = \pi(dy)\underbrace{P(y,E)}_{1} = \pi(dy)$.

Remarks

- the reversibility of π by the kernel P is preserved when

$$\alpha(x,y) = \begin{cases} a\left(\frac{\eta(y)q(y,x)}{\eta(x)q(x,y)}\right) \text{ if } \eta(x)q(x,y) > 0 \\ 1 \text{ if } \eta(x)q(x,y) = 0 \end{cases},$$

  where $a : \mathbb{R}_+ \to [0,1]$ satisfies $a(0) = 0$ and $a(u) = ua(1/u)$ for $u > 0$. The previous choice $a(u) = \min(u,1)$ leads to better asymptotic properties (Peskun 1973). Other ex: $a(u) = \frac{u}{1+u}$.
- When $E = \mathbb{R}^n$ et $q(x,y) = \varphi(y-x)$ for some symmetric probability density φ w.r.t. the Lebesgue measure λ (ex $\varphi(z) = e^{-\frac{|z|^2}{2\sigma^2}}/(2\pi\sigma^2)^{n/2}$), then

$$\frac{\eta(y)q(y,x)}{\eta(x)q(x,y)} = \frac{\eta(y)\varphi(y-x)}{\eta(x)\varphi(x-y)} = \frac{\eta(y)}{\eta(x)}.$$

  Algorithm called Random Walk Metroplis Hastings since the random variables $(Y_{n+1} - X_n)_{n\in N}$ are i.i.d. according to $\varphi(z)dz$.

Ergodic theory for Markov chains

Conditions on P and π ensuring that as $k \to \infty$,

- the law of X_k converges weakly to π,
- for $f : E \to \mathbb{R}$ measurable and such that $\int_E |f(x)|\pi(dx) < \infty$, $\frac{1}{k}\sum_{j=0}^{k-1} f(X_j)$ converges a.s. to $\int_E f(x)\pi(dx)$,
- $\sqrt{k}\left(\frac{1}{k}\sum_{j=0}^{k-1} f(X_j) - \int_E f(x)\pi(dx)\right)$ converges in law to $\mathcal{N}_1(0, \sigma_f^2)$

$$\text{where } \sigma_f^2 = \int_E \left(F^2(x) - \left(\int_E F(y)P(x,dy)\right)^2\right)\pi(dx)$$

with F solving the Poisson equation

$$\forall x \in E,\ F(x) - \underbrace{\int_E F(y)P(x,dy)}_{:=PF(x)} = f(x) - \underbrace{\int_E f(y)\pi(dy)}_{:=\pi(f)}$$

$\sum_{j=0}^{k-1}(f(X_j) - \pi(f)) =$
$\sum_{j=1}^{k-1}(F(X_j) - \mathbb{E}[F(X_j)|X_{0:j-1}]) + F(X_0) - PF(X_{k-1})$.

Random Walk Metropolis Hastings algorithm

- Sampling of a target probability measure with density η on $\mathbb{R}^n$
- $Y^n_{k+1} = X^n_k + \sigma G_{k+1}$ where $(G_k)_{k\geq 1}$ i.i.d. $\sim \mathcal{N}_n(0, I_n)$
- $q(x, y) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{|x-y|^2}{2\sigma^2}\right) = q(y, x)$
- Acceptance probability $\alpha(x, y) = \frac{\eta(y)}{\eta(x)} \wedge 1$.

How to choose σ in function of the dimension n?

Bad exploration of the space (and therefore poor ergodic properties) in the two opposite situations

- σ too large : large moves are proposed but almost always rejected,
- σ too small even if a large proportion of the proposed moves is then accepted.

Previous work: *Roberts, Gelman, Gilks 97*

Two fundamental assumptions:

- (H1) Product target: $\eta(x) = \eta(x_1, \dots, x_n) = \prod_{i=1}^n e^{-V(x_i)}$,
- (H2) Stationarity: $X_0^n = (X_0^{1,n}, \dots, X_0^{n,n}) \sim \eta(x)dx$ and thus $\forall k,\ X_k^n = (X_k^{1,n}, \dots, X_k^{n,n}) \sim \eta(x)dx$.

Then, pick the first component $X_k^{1,n}$, choose $\sigma_n = \frac{\ell}{\sqrt{n}}$, and rescale the time accordingly (diffusive scaling) by considering $(X_{\lfloor nt \rfloor}^{1,n})_{t\geq 0}$.

Under regularity assumptions on V, as $n \to \infty$, $(X_{\lfloor nt \rfloor}^{1,n})_{t\geq 0} \overset{(d)}{\Rightarrow} (X_t)_{t\geq 0}$ unique solution of the SDE

$$dX_t = \sqrt{h(\ell)}\, dB_t - \frac{h(\ell)}{2} V'(X_t)\, dt,$$

where $h(\ell) = 2\ell^2\, \Phi\left(-\frac{\ell\sqrt{\int_{\mathbb{R}} (V')^2 \exp(-V)}}{2}\right)$ with $\Phi(x) = \int_{-\infty}^x e^{-\frac{y^2}{2}} \frac{dy}{\sqrt{2\pi}}$.

Previous work: *Roberts, Gelman, Gilks 97*

Practical counterparts: scaling of the variance proposal.

Question: How to choose ℓ ?

- The function $\ell \mapsto h(\ell) = 2\ell^2\,\Phi\left(-\frac{\ell\sqrt{\int_{\mathbb{R}}(V')^2\exp(-V)}}{2}\right)$ is maximum at $\ell^\star \simeq \frac{2.38}{\sqrt{\int_{\mathbb{R}}(V')^2\exp(-V)}}$.
- Besides, the limiting average acceptance rate is

$$\mathbb{E}[\alpha(X_k^n, Y_{k+1}^n)] = \int_{\mathbb{R}^n\times\mathbb{R}^n} \underbrace{e^{\sum_{i=1}^n (V(x_i)-V(y_i))} \wedge 1}_{\alpha(x,y)}\, q_n(x,y) e^{-\sum_{i=1}^n V(x_i)} dxdy$$

$$\longrightarrow_{n\to\infty} a(\ell) = 2\Phi\left(-\frac{\ell\sqrt{\int_{\mathbb{R}}(V')^2\exp(-V)}}{2}\right) \in (0,1).$$

We observe that $a(\ell^\star) \simeq 0.234$, whatever V.

This justifies a constant acceptance rate strategy, with a target acceptance rate of approximately 25%.

Main result

Definition 1

A sequence $(\chi_1^n, \dots, \chi_n^n)_{n\geq 1}$ of exchangeable random variables is said to be ν-chaotic if for fixed $k \in \mathbb{N}^*$, the law of $(\chi_1^n, \dots, \chi_k^n)$ converges in distribution to $\nu^{\otimes k}$ as n goes to ∞.

Equivalent to the law of large numbers :

$$\nu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{\chi_i^n} \xrightarrow{Pr} \nu$$

Let $(G_k^i)_{i,k\geq 1}$ be i.i.d. $\sim \mathcal{N}_1(0,1)$ indep. $(U_k)_{k\geq 1}$ i.i.d. $\sim \mathcal{U}[0,1]$.

$$\begin{cases} X_{k+1}^{i,n} = X_k^{i,n} + \frac{\ell}{\sqrt{n}} G_{k+1}^i 1_{\mathcal{A}_{k+1}}, \ 1 \leq i \leq n, \\ \text{with } \mathcal{A}_{k+1} = \left\{ U_{k+1} \leq e^{\sum_{i=1}^n (V(X_k^{i,n}) - V(X_k^{i,n} + \frac{\ell}{\sqrt{n}} G_{k+1}^i))} \right\}. \end{cases}$$

Main result : RWMH target $\eta(x) = \prod_{i=1}^n \exp(-V(x_i))$

From now on, we assume that V is C^3 with V'' and $V^{(3)}$ bounded and m is a probability measure on $\mathbb{R}$ such that $\int_{\mathbb{R}} (V')^4(x)\, m(dx) < +\infty$.

Theorem 2

Assume that the initial positions $(X_0^{1,n}, \dots, X_0^{n,n})_{n\geq 1}$ are exchang., m-chaotic and s.t. $\lim_{n\to\infty} \mathbb{E}[(V'(X_0^{1,n}))^2] = \int_{\mathbb{R}} (V')^2(x)\, m(dx)$. Then the processes $((X_{\lfloor nt \rfloor}^{1,n}, \dots, X_{\lfloor nt \rfloor}^{n,n})_{t\geq 0})_{n\geq 1}$ are P-chaotic where P is the law of the unique solution to the SDE nonlinear in the sense of McKean with $X_0 \sim m$

$$dX_t = \sqrt{\Gamma}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)])dB_t - \mathcal{G}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)])V'(X_t)dt.$$

Moreover, $t \mapsto \mathbb{P}(\mathcal{A}_{\lfloor nt \rfloor})$ cv to $t \mapsto \frac{1}{\ell^2}\Gamma(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)])$.

Hypothesis satisfied if $\forall n \geq 1$, $X_0^{1,n}, \dots, X_0^{n,n}$ i.i.d. according to m.

The functions Γ and $\mathcal{G}$

$$\Gamma(a,b)=\begin{cases}\ell^2\Phi\left(-\frac{\ell b}{2\sqrt{a}}\right)+\ell^2 e^{\frac{\ell^2(a-b)}{2}}\Phi\left(\ell\left(\frac{b}{2\sqrt{a}}-\sqrt{a}\right)\right) \text{ if } a\in(0,+\infty),\\ \frac{\ell^2}{2} \text{ if } a=+\infty,\\ \ell^2 e^{-\frac{\ell^2 b^+}{2}} \text{where } b^+=\max(b,0) \text{ if } a=0,\end{cases}$$

$$\mathcal{G}(a,b)=\begin{cases}\ell^2 e^{\frac{\ell^2(a-b)}{2}}\Phi\left(\ell\left(\frac{b}{2\sqrt{a}}-\sqrt{a}\right)\right) \text{ if } a\in(0,+\infty),\\ 0 \text{ if } a=+\infty \text{ and } 1_{\{b>0\}}\ell^2 e^{-\frac{\ell^2 b}{2}} \text{ if } a=0.\end{cases}$$

For $a>0$, $\Gamma(a,a)=2\mathcal{G}(a,a)=2\ell^2\Phi\left(-\ell\sqrt{a}/2\right)$.

If $X_0^{i,n}$ i.i.d. $\sim e^{-V(x)}dx$, $\forall t\geq 0$, $X_t\sim e^{-V(x)}dx$ ($X_k^{i,n}\sim e^{-V(x)}dx$) and

$$\mathbb{E}[(V'(X_t))^2]=\int_{\mathbb{R}}V'(V'e^{-V})=\int_{\mathbb{R}}V'(-e^{-V})'=\int_{\mathbb{R}}V''e^{-V}=\mathbb{E}[V''(X_t)]$$

$$dX_t=\sqrt{h(\ell)}dB_t-\frac{h(\ell)}{2}V'(X_t)\,dt \text{ with } h(\ell)=2\ell^2\,\Phi\left(-\frac{\ell\sqrt{\int_{\mathbb{R}}(V')^2\exp(-V)}}{2}\right)$$

The functions Γ and $\mathcal{G}$

$$\Gamma(a,b) = \begin{cases} \ell^2\Phi\left(-\frac{\ell b}{2\sqrt{a}}\right) + \ell^2 e^{\frac{\ell^2(a-b)}{2}}\Phi\left(\ell\left(\frac{b}{2\sqrt{a}} - \sqrt{a}\right)\right) \text{ if } a \in (0,+\infty), \\ \frac{\ell^2}{2} \text{ if } a = +\infty, \\ \ell^2 e^{-\frac{\ell^2 b^+}{2}} \text{where } b^+ = \max(b,0) \text{ if } a = 0, \end{cases}$$

$$\mathcal{G}(a,b) = \begin{cases} \ell^2 e^{\frac{\ell^2(a-b)}{2}}\Phi\left(\ell\left(\frac{b}{2\sqrt{a}} - \sqrt{a}\right)\right) \text{ if } a \in (0,+\infty), \\ 0 \text{ if } a = +\infty \text{ and } 1_{\{b>0\}}\ell^2 e^{-\frac{\ell^2 b}{2}} \text{ if } a = 0. \end{cases}$$

For $a > 0$, $\Gamma(a,a) = 2\mathcal{G}(a,a) = 2\ell^2\Phi\left(-\ell\sqrt{a}/2\right)$.

If $X_0^{i,n}$ i.i.d. $\sim e^{-V(x)}dx$, $\forall t \geq 0$, $X_t \sim e^{-V(x)}dx$ ($X_k^{i,n} \sim e^{-V(x)}dx$) and

$$\mathbb{E}[(V'(X_t))^2] = \int_{\mathbb{R}} V'(V'e^{-V}) = \int_{\mathbb{R}} V'(-e^{-V})' = \int_{\mathbb{R}} V''e^{-V} = \mathbb{E}[V''(X_t)]$$

$$dX_t = \sqrt{h(\ell)}dB_t - \frac{h(\ell)}{2}V'(X_t)\,dt \text{ with } h(\ell) = 2\ell^2\,\Phi\left(-\frac{\ell\sqrt{\int_{\mathbb{R}}(V')^2\exp(-V)}}{2}\right)$$

Properties of Γ and $\mathcal{G}$

Lemma 3

1. $\forall (a,b) \in [0,+\infty] \times \mathbb{R},\ 0 \leq \mathcal{G}(a,b) \leq \Gamma(a,b) \leq l^2$,
2. *the function* Γ *is continuous on* $[0,+\infty] \times \mathbb{R}$ *and such that* $\inf_{(a,b)\in[0,+\infty]\times[\inf V'',\sup V'']} \Gamma(a,b) > 0$,
3. *the function* $\mathcal{G}$ *is continuous on* $\{[0,+\infty] \times \mathbb{R}\} \setminus \{(0,0)\}$,
4. $\exists C < +\infty,\ \forall (a,b)$ *and* $(a',b') \in [0,+\infty] \times [\inf V'', \sup V'']$,

$$|\Gamma(a,b) - \Gamma(a',b')| + (\sqrt{a} \wedge \sqrt{a'})|\mathcal{G}(a,b) - \mathcal{G}(a',b')| \leq C\left(|b'-b| + |a'-a| + |\sqrt{a'} - \sqrt{a}|\right).$$

$\sqrt{\Gamma}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)])$ and $\mathcal{G}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)])V'(X_t)$ the coefs of the SDE have the same regularity in terms of $(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)])$ by 2+4$\Rightarrow$ uniqueness by comp. $d(X_t - \tilde{X}_t)^2$

Mean field interaction

Let $(x_1, \dots, x_n) \in \mathbb{R}^n$, $\zeta^n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ and $(G^i)_{1\leq i\leq n}$ i.i.d. $\sim \mathcal{N}_1(0,1)$.

$$\mathcal{E}_{k+1} \stackrel{\text{def}}{=} \mathbb{E}\left(\left(\sum_{i=1}^n \left(V(x_i) - V(x_i + \frac{\ell G^i}{\sqrt{n}})\right) + \overbrace{\sum_{i=1}^n \left(V'(x_i)\frac{\ell G^i}{\sqrt{n}} + \frac{V''(x_i)\ell^2}{2n}\right)}^{\sim \mathcal{N}_1\left(\frac{\ell^2}{2}\langle \zeta^n, V''\rangle, \ell^2\langle \zeta^n, (V')^2\rangle\right)}\right)^2\right)$$

$$= \frac{\ell^4}{4n^2}\mathbb{E}\left(\left(\sum_{i=1}^n \left(V''(x_i)(1-(G^i)^2) - V^{(3)}(\chi_i)\frac{\ell}{3\sqrt{n}}(G^i)^3\right)\right)^2\right)$$

with $\chi_i \in [x_i, x_i + \frac{\ell G^i}{\sqrt{n}}]$ only depending on G^i. For $i \neq j$,

$$\mathbb{E}[V''(x_i)(1-(G^i)^2)\{V''(x_j)(1-(G^j)^2) - V^{(3)}(\chi_j)\frac{\ell}{3\sqrt{n}}(G^j)^3\}]$$
$$= V''(x_i)\mathbb{E}[1-(G^i)^2]\mathbb{E}[\dots] = 0$$

With boundedness of V'' and $V^{(3)}$, one concludes $\mathcal{E}_{k+1} \leq \frac{C}{n}$.

Mean field interaction

Let now $\mu_k^n = \frac{1}{n}\sum_{i=1}^n \delta_{X_k^{i,n}}$. The evolution of the RWM algorithm writes

$$X_{k+1}^{i,n} = X_k^{i,n} + \frac{\ell}{\sqrt{n}} G_{k+1}^i 1_{\left\{U_{k+1} \leq e^{\ell\sqrt{\langle \mu_k^n, (V')^2\rangle} G_{k+1} - \frac{\ell^2}{2}\langle \mu_k^n, V''\rangle + \mathcal{O}(n^{-1/2})}\right\}}, \quad 1 \leq i \leq n$$

where $G_{k+1} \sim \mathcal{N}_1(0,1)$ independent of the positions up to time k and such that

$$\mathbb{E}\left(G_{k+1}^i G_{k+1}\right) = \frac{\mathbb{E}(V'(X_k^{i,n}))}{\sqrt{n}}.$$

Gaussian calculations + diffusion approximation techniques lead to Theorem 1

Invariant measure

$$dX_t = -\mathcal{G}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)])V'(X_t)dt + \sqrt{\Gamma}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)])dB_t.$$

Proposition 4

The probability measure $e^{-V(x)}dx$ is the unique invariant measure for this SDE nonlinear in the sense of McKean.

- Choosing the $X_0^{i,n}$ i.i.d. according to $e^{-V(x)}dx$ in the main theorem, one obtains that this measure is invariant.
-
  - $\inf \Gamma > 0 \Rightarrow$ any invariant measure admits a density ψ_∞,
  - $\Gamma(+\infty, b) = \frac{\ell^2}{2}$ and $\mathcal{G}(+\infty, b) = 0 \Rightarrow a[\psi_\infty] \stackrel{\text{def}}{=} \int_{\mathbb{R}} (V')^2 \psi_\infty < +\infty$,
  - setting $b[\psi_\infty] \stackrel{\text{def}}{=} \int_{\mathbb{R}} V'' \psi_\infty$, one has $\psi_\infty \propto e^{-\frac{2\mathcal{G}}{\Gamma}(a[\psi_\infty], b[\psi_\infty])V}$ and $a[\psi_\infty] = \frac{\Gamma}{2\mathcal{G}}(a[\psi_\infty], b[\psi_\infty]) \int V'(-\psi_\infty)' = \frac{\Gamma}{2\mathcal{G}}(a[\psi_\infty], b[\psi_\infty])b[\psi_\infty]$ from which $a[\psi_\infty] = b[\psi_\infty]$ as $\frac{b\Gamma(a,b) - 2a\mathcal{G}(a,b)}{b-a} > 0$ when $a \neq b$.

Invariant measure

$$dX_t = -\mathcal{G}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)])V'(X_t)dt + \sqrt{\Gamma}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)])dB_t.$$

Proposition 4

The probability measure $e^{-V(x)}dx$ is the unique invariant measure for this SDE nonlinear in the sense of McKean.

- Choosing the $X_0^{i,n}$ i.i.d. according to $e^{-V(x)}dx$ in the main theorem, one obtains that this measure is invariant.
-
  - $\inf \Gamma > 0 \Rightarrow$ any invariant measure admits a density ψ_∞,
  - $\Gamma(+\infty, b) = \frac{\ell^2}{2}$ and $\mathcal{G}(+\infty, b) = 0 \Rightarrow a[\psi_\infty] \stackrel{\text{def}}{=} \int_{\mathbb{R}} (V')^2 \psi_\infty < +\infty$,
  - setting $b[\psi_\infty] \stackrel{\text{def}}{=} \int_{\mathbb{R}} V'' \psi_\infty$, one has $\psi_\infty \propto e^{-\frac{2\mathcal{G}}{\Gamma}(a[\psi_\infty], b[\psi_\infty])V}$ and $a[\psi_\infty] = \frac{\Gamma}{2\mathcal{G}}(a[\psi_\infty], b[\psi_\infty]) \int V'(-\psi_\infty)' = \frac{\Gamma}{2\mathcal{G}}(a[\psi_\infty], b[\psi_\infty])b[\psi_\infty]$ from which $a[\psi_\infty] = b[\psi_\infty]$ as $\frac{b\Gamma(a,b) - 2a\mathcal{G}(a,b)}{b-a} > 0$ when $a \neq b$.

Invariant measure

$$dX_t = -\mathcal{G}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)])V'(X_t)dt + \sqrt{\Gamma}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)])dB_t.$$

Proposition 4

The probability measure $e^{-V(x)}dx$ is the unique invariant measure for this SDE nonlinear in the sense of McKean.

- Choosing the $X_0^{i,n}$ i.i.d. according to $e^{-V(x)}dx$ in the main theorem, one obtains that this measure is invariant.
-
  - $\inf \Gamma > 0 \Rightarrow$ any invariant measure admits a density ψ_∞,
  - $\Gamma(+\infty, b) = \frac{\ell^2}{2}$ and $\mathcal{G}(+\infty, b) = 0 \Rightarrow a[\psi_\infty] \stackrel{\text{def}}{=} \int_{\mathbb{R}} (V')^2 \psi_\infty < +\infty$,
  - setting $b[\psi_\infty] \stackrel{\text{def}}{=} \int_{\mathbb{R}} V''\psi_\infty$, one has $\psi_\infty \propto e^{-\frac{2\mathcal{G}}{\Gamma}(a[\psi_\infty], b[\psi_\infty])V}$ and $a[\psi_\infty] = \frac{\Gamma}{2\mathcal{G}}(a[\psi_\infty], b[\psi_\infty]) \int V'(-\psi_\infty)' = \frac{\Gamma}{2\mathcal{G}}(a[\psi_\infty], b[\psi_\infty])b[\psi_\infty]$ from which $a[\psi_\infty] = b[\psi_\infty]$ as $\frac{b\Gamma(a,b) - 2a\mathcal{G}(a,b)}{b-a} > 0$ when $a \neq b$.

Fokker-Planck equation

Denoting by ψ_t the density of X_t, one has

$$\begin{cases} \partial_t \psi_t = \partial_x \Big(\mathcal{G}(a[\psi_t], b[\psi_t]) V' \psi_t + \frac{1}{2} \Gamma(a[\psi_t], b[\psi_t]) \partial_x \psi_t \Big), \\ a[\psi_t] = \int (V'(x))^2 \psi_t(x)\, dx, \\ b[\psi_t] = \int V''(x) \psi_t(x)\, dx. \end{cases} \tag{1}$$

Question 1: Does ψ_t converge to $\psi_\infty = \exp(-V)$?

Question 2: Is it possible to optimize the convergence, by appropriately choosing ℓ (recall that the variance of the proposal is ℓ^2/n, and thus that $\Gamma(a, b) = \Gamma(a, b, \ell)$ and $\mathcal{G}(a, b) = \mathcal{G}(a, b, \ell)$) ?

Fokker-Planck equation

To analyze the longtime behavior, we use entropy techniques.

Definition 5

The probability measure ν satisfies a log-Sobolev inequality with constant $\rho > 0$ (in short LSI(ρ)) if, for any probability measure μ absolutely continuous wrt ν,

$$H(\mu|\nu) \leq \frac{1}{2\rho} I(\mu|\nu) \text{ where} \tag{2}$$

- $H(\mu|\nu) = \int \ln\left(\frac{d\mu}{d\nu}\right) d\mu$ *is the Kullback-Leibler divergence (or relative entropy) of μ wrt ν,*
- $I(\mu|\nu) = \int \left|\nabla \ln\left(\frac{d\mu}{d\nu}\right)\right|^2 d\mu$ *is the Fisher information of μ wrt ν.*

Convergence to the invariant density $\psi_\infty = e^{-V}$

Theorem 6

If X_0 admits a density ψ_0 s.t. $\mathbb{E}[(V'(X_0))^2] < +\infty$ and $H(\psi_0|\psi_\infty) < \infty$, then

$$\frac{d}{dt}H(\psi_t|\psi_\infty) \leq -\frac{b[\psi_t]\Gamma(a[\psi_t], b[\psi_t]) - 2a[\psi_t]\mathcal{G}(a[\psi_t], b[\psi_t])}{2(b[\psi_t] - a[\psi_t])} I(\psi_t|\psi_\infty) < 0.$$

If moreover $\psi_\infty = e^{-V}$ satisfies LSI(ρ), then there exists a positive and non-increasing function $\lambda : [0, +\infty) \to (0, +\infty)$ such that $\forall t \geq 0$

$$H(\psi_t|\psi_\infty) \leq e^{-t\lambda(H(\psi_0|\psi_\infty))} H(\psi_0|\psi_\infty).$$

Roughly speaking, e^{-V} satisfies LSI(ρ) for some $\rho > 0$ if V has at least quadratic growth at ∞.

In the Gaussian case $V(x) = \frac{x^2 + \ln(2\pi)}{2}$, $\mathcal{N}_1(0,1)$ satisfies LSI(1).

Elements of proof

Writing a, b for $a[\psi_t], b[\psi_t]$, one has

$$\frac{d}{dt}H(\psi_t|\psi_\infty) = \int_{\mathbb{R}} \partial_t\psi_t \ln\psi_t + \int_{\mathbb{R}} V\partial_t\psi_t$$
$$= -\frac{\Gamma(a,b)}{2} I(\psi_t|\psi_\infty) + (a-b)^2 \times \left\{\frac{2\mathcal{G}(a,b)-\Gamma(a,b)}{2(b-a)}\right\}_{\geq 0}$$

$$(a-b)^2 = \left(\int_{\mathbb{R}} (V')^2\psi_t - \int_{\mathbb{R}} V''\psi_t\right)^2 = \left(\int_{\mathbb{R}} V'(V'\psi_t + \partial_x\psi_t)\right)^2$$
$$= \left(\int_{\mathbb{R}} V'\partial_x \ln(\psi_t/e^{-V})\psi_t\right)^2 \leq a \times I(\psi_t|\psi_\infty).$$

Hence $\frac{d}{dt}H(\psi_t|\psi_\infty) \leq -\frac{b\Gamma(a,b)-2a\mathcal{G}(a,b)}{2(b-a)} I(\psi_t|\psi_\infty)$. When ψ_∞ satisfies LSI(ρ), it satisfies the transport inequality $W_2^2(\psi_t,\psi_\infty) \leq \frac{2}{\rho}H(\psi_t|\psi_\infty)$.
With $t \mapsto H(\psi_t|\psi_\infty) \searrow \Rightarrow \sup_t a[\psi_t] < C(H(\psi_0|\psi_\infty))$ with $C \nearrow$.
$\lambda(H(\psi_0|\psi_\infty)) \stackrel{\text{def}}{=} \frac{1}{2\rho} \inf_{(a,b):a\leq C(H(\psi_0|\psi_\infty))} \frac{b\Gamma(a,b)-2a\mathcal{G}(a,b)}{2(b-a)} > 0.$

Decrease of the Kullback-Leibler divergence

When $b \leq 0$, one has $\frac{d}{dt}H(\psi_t|\psi_\infty) \leq -\frac{\Gamma(a,b)}{2}\int_{\mathbb{R}}(\partial_x \ln \psi_t)^2\psi_t$ with $\lim_{\ell\to\infty}\Gamma(a,b) = +\infty$. So one should choose ℓ as large as possible. From now on, suppose that $b > 0$ (recall that in the longtime limit $b = a > 0$).

$$\frac{d}{dt}H(\psi_t|\psi_\infty) \leq -\underbrace{\frac{b\Gamma(a,b) - 2a\mathcal{G}(a,b)}{2(b-a)}}_{\frac{1}{b}F(\frac{a}{b},\ell\sqrt{b})} I(\psi_t|\psi_\infty) < 0,$$

where

$$F(s,\ell) = \begin{cases} \ell^2 e^{-\frac{\ell^2}{2}} \text{ if } s = 0 \\ 2\ell^2\left(\left(1+\frac{\ell^2}{4}\right)\Phi\left(-\frac{\ell}{2}\right) - \frac{\ell}{2\sqrt{2\pi}}e^{-\frac{\ell^2}{8}}\right) \text{ if } s = 1 \\ \frac{\ell^2}{1-s}\left(\Phi\left(-\frac{\ell}{2\sqrt{s}}\right) + (1-2s)e^{\frac{\ell^2(s-1)}{2}}\Phi\left(\frac{\ell}{2\sqrt{s}} - \ell\sqrt{s}\right)\right) \text{ if } 0 < s \neq 1 \end{cases}$$

Choice of ℓ maximizing the exponential rate of cv

Lemma 7

Let $b > 0$. Then $\tilde{\ell}^\star(a,b) = \mathrm{argmax}_{\ell \geq 0} \frac{1}{b} F(\frac{a}{b}, \ell\sqrt{b}) = \frac{1}{\sqrt{b}} \ell^\star\left(\frac{a}{b}\right)$ where for any $s \geq 0$, $\ell^\star(s)$ realizes the unique maximum of $\ell \mapsto F(s,\ell)$. Moreover, $s \mapsto \ell^\star(s)$ is continuous on $[0,+\infty)$ and

- $\tilde{\ell}^\star(a,b) \sim_{a/b\to 0} \frac{\ell^\star(0)}{\sqrt{b}} = \frac{\sqrt{2}}{\sqrt{b}}$.
- $\tilde{\ell}^\star(a,b) \sim_{a/b\to 1} \frac{\ell^\star(1)}{\sqrt{b}}$.
- $\tilde{\ell}^\star(a,b) \sim_{a/b\to +\infty} \frac{x^\star\sqrt{a}}{b}$ *where* $x^\star \simeq 1.22$.

Notice that

$$dV(X_t) = V'(X_t)\left(\sqrt{\Gamma(a,b)}dB_t - \mathcal{G}(a,b)V'(X_t))dt\right) + \frac{1}{2}\Gamma(a,b)V''(X_t)dt$$

so that $\frac{d}{dt}\mathbb{E}[V(X_t)] = \frac{1}{2}(b\Gamma(a,b) - 2a\mathcal{G}(a,b)) = \frac{b-a}{b}F(\frac{a}{b}, \ell\sqrt{b})$ and $\tilde{\ell}^\star(a,b)$ also maximizes $|\frac{d}{dt}\mathbb{E}[V(X_t)]|$.

Comparison with constant acceptance rate strategies

The limiting mean acceptance rate in Theorem 2 is

$$acc(a,b,\ell) = \frac{1}{\ell^2}\Gamma(a,b) = H\left(\frac{a}{b}, \ell\sqrt{b}\right)$$

$$\text{where } H(s,\ell) = \Phi\left(-\frac{\ell}{2\sqrt{s}}\right) + e^{\frac{\ell^2(s-1)}{2}}\Phi\left(\ell\left(\frac{1}{2\sqrt{s}} - \sqrt{s}\right)\right).$$

Lemma 8

For $s > 0$, the function $\ell \mapsto H(s,\ell)$ is decreasing. Moreover, for $\alpha \in (0,1)$, the unique ℓ s.t. $acc(a,b,\ell) = \alpha$ is $\tilde{\ell}^\alpha(a,b) = \frac{1}{\sqrt{b}}\ell^\alpha\left(\frac{a}{b}\right)$ where $\ell^\alpha(s)$ is the unique solution to $H(s,\ell^\alpha(s)) = \alpha$. Last,

- $\tilde{\ell}^\alpha(a,b) \sim_{a/b\to 0} \frac{\sqrt{-2\ln(\alpha)}}{\sqrt{b}}$.
- $\tilde{\ell}^\alpha(a,b) \sim_{a/b\to 1} \frac{\ell^\alpha(1)}{\sqrt{b}}$.
- $\tilde{\ell}^\alpha(a,b) \sim_{a/b\to\infty} -2\Phi^{-1}(\alpha)\frac{\sqrt{a}}{b}$.

Comparison with constant acceptance rate strategies

Remark 1: Notice that $\tilde{\ell}^\star(a,b) = \frac{1}{\sqrt{b}}\ell^\star\left(\frac{a}{b}\right)$ and $\tilde{\ell}^\alpha(a,b) = \frac{1}{\sqrt{b}}\ell^\alpha\left(\frac{a}{b}\right)$ have the same scaling in (a,b).
$\longrightarrow$ Constant acceptance rate strategy seems sensible.

Remark 2: Choice of α: how to choose α to get $\tilde{\ell}^\star(a,b) \sim \tilde{\ell}^\alpha(a,b)$?

- $a/b \to 0$: $\alpha = \frac{1}{e} \simeq 0.37$.
- $a/b \to 1$: α such that $\ell^\alpha(1) = \ell^\star(1)$, namely $\alpha \simeq 0.35$.
- $a/b \to \infty$: $\alpha = \Phi(-x^\star/2) \simeq 0.27$.

(The standard choice for the RWM, under the stationarity assumption, is $\alpha = 0.234$.)
$\longrightarrow$ Constant acceptance rate with $\alpha \in (1/4, 1/3)$ seems sensible.

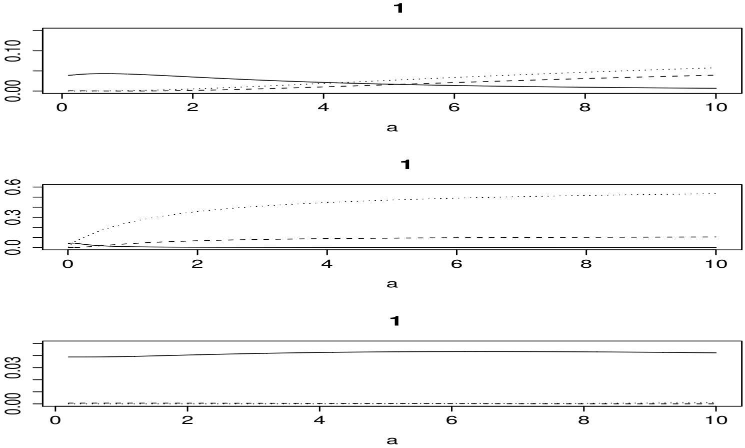

Figure : $\frac{F(\frac{a}{b},\tilde{l}^\star(a,b)\sqrt{b})-F(\frac{a}{b},\tilde{l}^\alpha(a,b)\sqrt{b})}{F(\frac{a}{b},\tilde{l}^\star(a,b)\sqrt{b})}$ as function of a for $b = 1, 0.1, 10$ and $\alpha = 0.27$ solid line, $\alpha = 0.35$ dashed line, $\alpha = e^{-1} = 0.37$ dotted line.

Gaussian target : $V(x) = \frac{1}{2}(x^2 + \ln(2\pi))$

We assume that ψ_0 Gaussian $\Rightarrow \psi_t$ Gaussian.

Setting $m(t) \stackrel{\text{def}}{=} \mathbb{E}[X_t] = \int_{\mathbb{R}} x\psi_t(x)dx$ and $s(t) \stackrel{\text{def}}{=} \mathbb{E}[(X_t)^2] = \int_{\mathbb{R}} x^2\psi_t(x)dx$, one has

$$H(\psi_t|\psi_\infty) = \frac{1}{2}\left(s(t) - \ln(s(t) - m(t)^2) - 1\right),$$

$$\frac{d}{dt}H(\psi_t|\psi_\infty) = \frac{1}{2}\left(F(s,\ell)(1-s) - \frac{F(s,\ell)(1-s) + 2m\mathcal{G}(s,1,\ell)}{s - m^2}\right).$$

It is possible to approximate $\ell^{ent}(m,s)$ maximizing $\left|\frac{d}{dt}H(\psi_t|\psi_\infty)\right|$.
To assess the convergence, we compute

$$t_0 \mapsto \hat{I}^m_{t_0,t_0+T} = \frac{1}{T}\sum_{k=t_0+1}^{t_0+T} \frac{X_k^{1,n} + \ldots + X_k^{n,n}}{n}$$

$$t_0 \mapsto \hat{I}^s_{t_0,t_0+T} = \frac{1}{T}\sum_{k=t_0+1}^{t_0+T} \frac{(X_k^{1,n})^2 + \ldots + (X_k^{n,n})^2}{n}.$$

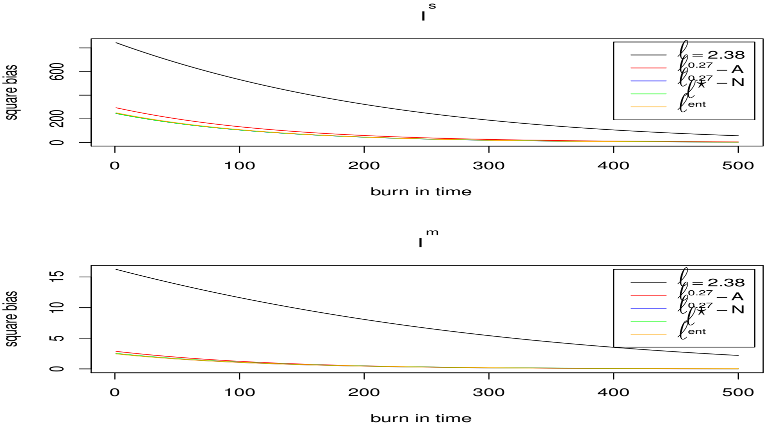

Figure : $t_0 \mapsto$square bias of $(\hat{I}^s_{t_0,T+t_0}, \hat{I}^m_{t_0,T+t_0})$, $(X_0^{1,n}, \dots, X_0^{n,n}) = (10, \dots, 10)$, $n = 100$($\ell^{0.27} - A \rightarrow$ adaptive scaling Metropolis algorithm and $\ell^{0.27} - N \rightarrow$ numerical approximation of $\ell^{0.27}(s, 1)$.)

Conclusions:

1. The constant ℓ strategy is bad ;
2. The constant average acceptance rate strategy (using ℓ^α) leads to very close convergence curves compared to the optimal exponential rate of convergence strategy (using $\ell^\star$) ;
3. The optimal exponential rate of convergence strategy is as good as the most optimal strategy one could design in terms of entropy decay (using ℓ^{ent}).