On characterizations of Metropolis type algorithms in continuous time

Persi Diaconis and Laurent Miclo

24 novembre 2008

Plan of the talk

1. Introduction and results
2. Finite state space Girsanov formula
3. Entropy minimization
4. Other Metropolis type projections
5. On the compact Riemannian diffusion situation

The classical Metropolis algorithm

Very popular algorithm for simulation.

- S finite set endowed with a Markovian kernel K
- π positive probability measure on S

Goal : construct K_π Markov kernel reversible wrt π.

$$K_\pi(x,y) \quad := \quad \begin{cases} \min\left\{K(x,y), \frac{\pi(y)}{\pi(x)}K(y,x)\right\} & \text{, if } x \neq y \\ 1 - \sum_{z\in S\setminus\{x\}} K_\pi(x,z) & \text{, if } x = y \end{cases}$$

If it is furthermore ergodic, use a corresponding Markov chain $(\widetilde{X}_n)_{n\in\mathbb{N}}$.

Billera and Diaconis' characterization

$\mathcal{K}$ set of Markov kernels, $\mathcal{K}(\pi)$ subset of those reversible wrt π, $\mathcal{K}(\pi, K)$ subset of $\mathcal{K}(\pi)$ consisting of Markov matrices M whose off-diagonal entries are less or equal than those of K.

$$\forall\, K, K' \in \mathcal{K}, \quad d(K', K) \;:=\; \sum_{x \in S} \pi(x) \sum_{y \in S \setminus \{x\}} \left| K'(x, y) - K(x, y) \right|$$

Theorem 1 (Billera and Diaconis [2001])

With respect to d, the Metropolis kernel K_π minimizes the distance from K to $\mathcal{K}(\pi)$ and it is the unique minimizer of the distance from K to $\mathcal{K}(\pi, K)$.

Our purpose

Extension to continuous time framework ?

Immediate for finite jump processes ! What for diffusion processes ?

How to recover the well-know fact : the Metropolis algorithm associated to $\triangle/2$ and $\pi \# \exp(-U)$ is

$$(\triangle \cdot - \langle \nabla U, \nabla \cdot \rangle)/2$$

Incantation : use entropy.
First work with S finite set endowed with a generator L, Markov process $X^{(\mu)} := (X^{(\mu)}(t))_{t \geq 0}$ starting from μ (M other generator, Markov process $Y^{(\mu)}$).

Entropy

Recall that for any probability measures μ, ν,

$$\mathrm{Ent}(\mu|\nu) \quad := \quad \begin{cases} \int \frac{d\mu}{d\nu} \ln\left(\frac{d\mu}{d\nu}\right) d\nu \ \leq \ +\infty & \text{, if } \mu \ll \nu \\ +\infty & \text{otherwise} \end{cases}$$

Interpretation for simulation ?

Proposition 2

If π is invariant for M, for any $T \geq 0$, we have

$$\mathrm{Ent}(\mathcal{L}(Y^{(\pi)}([0,T]))|\mathcal{L}(X^{(\pi)}([0,T]))) \quad = \quad T\widetilde{d}(M,L) \ \leq \ +\infty$$

where is the discrepancy $\widetilde{d}(M,L)$ is given by

$$\sum_{x\in S} \pi(x) \sum_{y\in S\setminus\{x\}} M(x,y) \ln\left(\frac{M(x,y)}{L(x,y)}\right) - M(x,y) + L(x,y)$$

Another Metropolis generator

Disturbing the initial distribution π ?

$$\lim_{T\to+\infty} \mathrm{Ent}(\mathcal{L}(Y^{(\mu)}([0,T]))|\mathcal{L}(X^{(\mu)}([0,T])))/T \quad = \quad \widetilde{d}(M,L)$$

Definition of another Metropolis algorithm :

Theorem 3

The mapping $\mathcal{L}(\pi) \ni M \mapsto \widetilde{d}(M,L)$ admits a unique minimizer $\widetilde{L}_\pi$ which is given by

$$\widetilde{L}_\pi(x,y) \quad := \quad \begin{cases} \sqrt{\frac{\pi(y)}{\pi(x)}}\sqrt{L(x,y)L(y,x)} & , \text{ if } x \neq y \\ -\sum_{z\in S\setminus\{x\}} \widetilde{L}_\pi(x,z) & , \text{ if } x = y \end{cases}$$

φ-relative entropy

$\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ convex function satisfying $\varphi(1) = 0$, $\varphi'(1) = 0$ and whose growth is at most of polynomial order

$$\mathrm{Ent}_\varphi(\mu|\nu) := \begin{cases} \int \varphi\left(\frac{d\mu}{d\nu}\right) d\nu & \text{, if } \mu \ll \nu \\ +\infty & \text{otherwise} \end{cases}$$

Proposition 4

Without any assumption on M, we have

$$\lim_{T \to 0_+} \mathrm{Ent}_\varphi(\mathcal{L}(Y^{(\pi)}([0,T]))|\mathcal{L}(X^{(\pi)}([0,T])))/T = d_\varphi(M, L)$$

where the discrepancy $d_\varphi(M, L)$ is given by

$$\begin{cases} \sum_{x \in S} \pi(x) \sum_{y \in S \setminus \{x\}} L(x,y) \varphi\left(\frac{M(x,y)}{L(x,y)}\right) & \text{, if } M \ll L \\ +\infty & \text{otherwise} \end{cases}$$

Even more Metropolis generators...

Proposition 5

If φ is furthermore assumed to be strictly convex then the mapping $\mathcal{L}(\pi) \ni M \mapsto d_\varphi(M, L)$ admits a unique minimizer $L_{\varphi,\pi}$.

Example of a kind of Huber loss function : for $\epsilon \in (0, 1/2]$, let φ_ϵ satisfy $\varphi_\epsilon(x) = (x-1)^2$ for any $x \in [1-\epsilon, 1+\epsilon]$, $\varphi'_\epsilon = -1$ on $[0, 1-\epsilon)$ and $\varphi'_\epsilon = 1+\epsilon$ on $(1+\epsilon, +\infty)$. Then there is a unique minimizer $L_{\varphi_\epsilon,\pi}$ and

$$\lim_{\epsilon \to 0_+} L_{\varphi_\epsilon,\pi} \quad = \quad L_\pi$$

This "definition" can be extended to the diffusion situation.

Martingale problems

S finite set, two generators L and $\widetilde{L}$, μ initial distribution, corresponding trajectorial laws $\mathbb{P}_{\mu,[0,T]}$ and $\widetilde{\mathbb{P}}_{\mu,[0,T]}$ on the canonical probability space.
Natural martingale associated to a function f :

$$\mathcal{M}_t^{(f)} \quad := \quad f(X(t)) - f(X(0)) - \int_0^t L[f](X(s))\, ds$$

Extension for functions on S^2 : if $F = f \otimes g$:

$$\mathcal{M}_t^{(F)} \quad := \quad \int_0^t f(X^{(\mu)}(s-))\, d\mathcal{M}_s^{(g)}$$

For $x \neq y$ and $t \geq 0$,

$$\mathcal{M}_t^{(x,y)} \quad = \quad N_t^{(x,y)} - \int_0^t L(x,y) \mathbb{1}_x(X(s))\, ds$$

where $N_t^{(x,y)} =$ number of jumps from x to y up to time $t \geq 0$.

Girsanov's formula

Theorem 6

Under the assumption that $\widetilde{L} \ll L$, for any initial condition μ and any finite time horizon $T \geq 0$, we have $\widetilde{\mathbb{P}}_{\mu,[0,T]} \ll \mathbb{P}_{\mu,[0,T]}$ and the corresponding Radon-Nikodym derivative is given by

$$\begin{aligned}
\frac{d\widetilde{\mathbb{P}}_{\mu,[0,T]}}{d\mathbb{P}_{\mu,[0,T]}} &= \exp\left(\mathcal{M}_T^{(V)} + \int_0^T v(X(s))\,ds\right) \\
&= \exp\left(\sum_{x \neq y \in S} V(x,y) N_T^{(x,y)} + \int_0^T g(X(s))\,ds\right)
\end{aligned}$$

Formulas...

where

$$\begin{aligned} V(x,y) &:= \ln\left(\frac{\widetilde{L}(x,y)}{L(x,y)}\right) \\ v(x) &:= \sum_{y\neq x} L(x,y)(V(x,y) - \exp(V(x,y)) + 1) \\ g(x) &:= \widetilde{L}(x,x) - L(x,x) \end{aligned}$$

This comes from a more general abstract formulation :
Under technical conditions, the Markov process $\widetilde{X}$ is absolutely continuous on compact time interval with respect to X iff there exists a (pseudo-)function V of two variables such that

$$\widetilde{L}\cdot = L\cdot + \Gamma[\exp(V), \cdot]$$

where Γ is the carré du champ associated to L (acting on the second variable).

This corresponds to modify the intensity of jumps and adding drifts in the directions permitted by the diffusion coefficients (result due to Kunita [1969]).
The Radon-Nikodym density is given on the time interval $[0, T]$ by

$$\exp\left(\mathcal{M}_T^{(V)} + \int_0^T I[V](X(s))\, ds\right)$$

where I is a kind of exponentielle du champ :

$$I[V](x) \quad := \quad \exp(-V(x,x))L[\exp(V(x,\cdot)](x) - L[V(x,\cdot)](x)$$

(it coincides with $\Gamma[V, V](x)/2$ for diffusions processes).

Relative entropy computation

Let μ_t the law of $Y_t^{(\mu)}$ and

$$F(x) := \sum_{y \in S \setminus \{x\}} M(x,y) \ln\left(\frac{M(x,y)}{L(x,y)}\right) - M(x,y) + L(x,y)$$

Using that $(N_t^{(x,y)} - \int_0^t M(x,y) \mathbb{1}_x(Y^{(\mu)}(s))\, ds)_{t \in [0,T]}$ is a martingale, we get (if $M \ll L$)

$$\mathrm{Ent}(\mathcal{L}(Y^{(\mu)}([0,T])) | \mathcal{L}(X^{(\mu)}([0,T]))) = \int_0^T \mu_t[F]\, dt$$

Proposition 2 follows from $\pi[F] = \widetilde{d}(M,L)$.

Minimization of $d(\cdot, L)$ on $\mathcal{L}(\pi)$

Let $\preceq$ be a total ordering on S, using the reversibility of M wrt π,

$$\begin{aligned}\widetilde{d}(M, L) &= \sum_{x \prec y} \pi(x)M(x,y)\ln\left(\frac{(\pi(x)M(x,y))^2}{\pi(x)L(x,y)\pi(y)L(y,x)}\right) \\ &\quad -2\pi(x)M(x,y) + \pi(x)L(x,y) + \pi(y)L(y,x)\end{aligned}$$

and each summand can be minimized independently, to prove Theorem 3.

Note due to the inequality $\forall\, a, b \geq 0$, $\min(a,b) \leq \sqrt{ab}$, $\widetilde{L}_\pi$ makes the corresponding process go faster to the equilibrium π than L_π.

Jumps rate

The probabilistic description of $X^{(\mu)}$ leads to

$$\lim_{t\to 0_+} t^{-1}\mathbb{E}_\mu\left[\varphi\left(\exp\left(\sum_{(x,y)\in S^{(2)}} a(x,y)N_t^{(x,y)}\right)\right)\right]$$

$$= \sum_{(x,y)\in S^{(2)}} \mu(x)L(x,y)\varphi(\exp(a(x,y)))$$

for any function φ locally bounded and whose growth is at most of polynomial order and any $a(x,y) \in \mathbb{R} \sqcup \{-\infty\}$.
Proposition 4 follows, because the addition of a term $\int_0^t h(X^{(\mu)}(s))\,ds$ doesn't change this behavior if $\varphi'(1) = 0$.

Convex minimization

As in the entropy case, we are led to minimize in α the convex expression (with $\beta = \pi(x)L(x,y)$, $\beta' = \pi(y)L(y,x)$, for $x \prec y$),

$$\Phi_{\beta,\beta'}(\alpha) \quad := \quad \beta\varphi\left(\frac{\alpha}{\beta}\right) + \beta'\varphi\left(\frac{\alpha}{\beta'}\right)$$

and there is a unique solution if "there is no opposite slopes of flat parts of φ".
Not the case for $|\cdot - 1|$: M is minimizing for $d(\cdot, L)$ on $\mathcal{L}(\pi)$ iff $\forall\ x \prec y$, $\pi(x)M(x,y) = \pi(y)M(y,x) \in [\pi(x)L(x,y), \pi(y)L(y,x)]$ and we get the announced convergence of $L_{\varphi_\epsilon,\pi}$.

Diffusion generators

S smooth compact manifold of dimension $n \in \mathbb{N}^*$

$$L[f](x) \quad := \quad \frac{1}{2} \sum_{i,j \in [\![1,n]\!]} a_{i,j}(x) \partial_{i,j} f(x) + \sum_{i \in [\![1,n]\!]} b_i(x) \partial_i f(x)$$

in any chart, a and b smooth and a invertible.

$X^{(\mu)}$ corresponding diffusion process and $\widetilde{X}^{(\widetilde{\mu})}$ associated to another generator of the same kind $\widetilde{L}$.

Ellipticity and iterated logarithm law enable to show that for any $T > 0$,

$$\mathcal{L}(\widetilde{X}^{(\widetilde{\mu})}([0,T])) \ll \mathcal{L}(X^{(\mu)}([0,T])) \;\Rightarrow\; \widetilde{a} = a$$

Riemannian setting

Since we will be interested in absolute continuous diffusions, consider the Riemannian structure generated by a^{-1}. Usual notations : $|\cdot|$, ∇, $\triangle$, λ etc.

Intrinsic writting : $L\cdot = \triangle/2 \cdot + \langle b, \nabla \cdot \rangle$, with b a vector field.

The traditional Girsanov's formula can be formally recovered by the same approach as in the finite set case, using a two variables function F such that

$$\forall\, x \in S, \qquad \nabla_y F(x,y)_{|y=x} \quad = \quad \widetilde{b}(x) - b(x)$$

Girsanov's formula

For any initial distribution μ and any finite time horizon $T \geq 0$, the law $\mathcal{L}(\widetilde{X}^{(\mu)}([0,T]))$ is absolutely continuous with respect to $\mathcal{L}(X^{(\mu)}([0,T]))$ and the corresponding Radon-Nikodym density is equal to

$$\frac{\mathcal{L}(\widetilde{X}^{(\mu)}([0,T]))}{\mathcal{L}(X^{(\mu)}([0,T]))} \quad = \quad \exp\left(\mathcal{M}_T^{(\widetilde{b}-b)} - \frac{1}{2}\int_0^T \left|\widetilde{b}-b\right|^2 (X^{(\mu)}(t))\,dt\right)$$

where $(\mathcal{M}_t^{(\widetilde{b}-b)})_{t\geq 0}$ is a martingale whose bracket is given by

$$\forall\, t \geq 0, \qquad \langle \mathcal{M}^{(\widetilde{b}-b)}\rangle_t \quad = \quad \int_0^t \left|\widetilde{b}-b\right|^2 (X^{(\mu)}(s))\,ds$$

Discrepancy on the set of generators

For the behavior of φ-relative entropies, we need that $\varphi''(1)$ exists :

$$\lim_{T\to 0_+} \mathrm{Ent}_\varphi(\mathcal{L}(\widetilde{X}^{(\pi)}([0,T]))|\mathcal{L}(X^{(\pi)}([0,T])))/T \quad = \quad \frac{\varphi''(1)}{2} d(\widetilde{L}, L)$$

where the discrepancy $d(\widetilde{L}, L)$ doesn't depend on φ :

$$\begin{cases} \int \left|\widetilde{b} - b\right|^2 (x)\, \pi(dx) & \text{, if } \widetilde{L} \sim L \\ +\infty & \text{, otherwise} \end{cases}$$

Entropy minimization

Let π a probability measure with a smooth and positive density wrt λ. The Metropolis algorithm associated to L and π should be the minimizer M of $\mathcal{L}(\pi) \ni \widetilde{L} \mapsto d(\widetilde{L}, L)$. This is $M = (\triangle \cdot - \langle \nabla \ln(\pi), \nabla \cdot \rangle)/2$, since it is the unique generator of the form $\widetilde{L} \cdot = \triangle/2 \cdot + \langle \widetilde{b}, \nabla \cdot \rangle$ which is reversible wrt π.

Are you not fed up yet ?

Jumps can be added by considering generator of the kind $H = L + Q$ with

$$Q[f](x) \;:=\; \int (f(y) - f(x))\, q(x,y)\lambda(dy)$$

where q a smooth and positive function on S^2.
Using that diffusive and jump parts don't interact (diffusive and jump martingales are orthogonal, Girsanov density splits into distinct factors, H is reversible wrt π iff L and Q are reversible wrt π), similar computations can be done ...