6 - An introduction to diffusion processes

Here we consider "continuous state spaces" and Markov processes whose trajectories are continuous.

The most famous example is the real Brownian motion. One way to define it is through martingale problems:

We take $V = \mathbb{R}$ and consider the operator $L := \frac{\partial^2}{2}$ defined on

$$\mathcal{D}(L) = C_b^2(\mathbb{R}),$$ the set of

C^2-functions on $\mathbb{R}$ which are bounded as well as their first and second derivatives.

A process $X := (X(t))_{t \geq 0}$ with continuous trajectories is said to be a Brownian motion when for any $f \in \mathcal{D}(L)$, $M^f :=$

$$(M^f_t)_{t \geq 0} := \left(f(X(t)) - f(X(0)) - \int_0^t L[f](X(s))\,ds \right)_{t \geq 0}$$

is a martingale (in the filtration generated by X).

In the theory of stochastic calculus, M^f is given by the Itô's integrals

$$\forall t \geq 0, \quad M_t^f = \int_0^t f'(X(s))\, dX(s)$$

but this knowledge is not necessary for Markov theory, as the bracket $\langle M^f \rangle$ can be directly deduced from the martingale problem. Indeed, as soon as the domain $\mathcal{D}(L)$ of a Markov generator L is an algebra (assuming $\mathcal{D}(L) \subset C_b(V)$ and $L(\mathcal{D}(L)) \subset C_b(V)$), we can compute that

$$\forall t \geq 0, \quad M_t^f = \int_0^t \Gamma_L(f,f)(X(s))\, ds$$

where the carré du champ Γ is given by

$$\forall f,g \in \mathcal{D}(L) \quad \Gamma_L(f,g) = L[fg] - fLg - gLf$$

In the case of the Brownian motion, we have

$$\forall f \in \mathcal{D}(L), \quad \Gamma_L(f,f) = (f')^2$$

which gave the name carré du champ

(field square).

When L is a finite Markov generator, $\forall x \in V$,

$$\Gamma_L(f,f)(x) = \sum_y L(x,y)\left(f(y)-f(x)\right)^2$$

Note that the Dirichlet form is given by

$$\forall f,g \in \mathcal{F}(V), \quad \mathcal{E}(f,g) = \frac{1}{2}\pi\left[\Gamma(f,f)\right]$$

This equality also holds in more general situations.

The continuous equivalent of a finite state space is a compact differential manifold V. It allows us not to worry about the behavior of processes at infinity. This is also the reason why we did not consider denumerable state spaces. Problem at infinity are usually treated via other methods such as Lyapounov functions.

It is convenient to endow V with a Riemannian structure: For any $x \in V$, we are give a scalar product $\langle \cdot, \cdot \rangle_x$ on $T_x V$, smoothly depending on x. It enables us to define a gradient via

$$\forall f \in C^1(V), \forall x \in V, \forall j \in [\![d]\!],$$

$$\left(\nabla f(x) \right)_j = \sum_{i \in [\![d]\!]} g_{j,i}^{-1}(x) \, \partial_i f(x)$$

(in a basis $(e_i)_{i \in [\![d]\!]}$ of $T_x V$, where $g_{i,j}(x) := \langle e_i, e_j \rangle_x$ and d is the dimension of V)

and the Laplace-Beltrami operator:

$$\forall f \in C^2(V), \forall x \in V$$

$$\Delta f(x) = \frac{1}{\sqrt{|g(x)|}} \sum_{i,j \in [\![d]\!]} \partial_i \left(\sqrt{|g(x)|} \, g_{i,j}^{-1}(x) \, \partial_j f(x) \right)$$

where $|g(x)|$ is the determinant of $g(x)$ " $\left(g_{i,j}(x) \right)_{i,j \in [\![d]\!]}$

A Markov process $(X(t))_{t \geq 0}$ with continuous trajectories is said to be a Brownian motion on V when it is a solution the martingale problem associated to $L = \frac{\Delta}{2}$ with domain $\mathcal{D}(L) = C^2(V)$.

Remark

The continuity of the trajectories is in fact a consequence of the martingale problem, if we had only assumed them to be càdlàg.

This continuity can be seen to be equivalent to the property that

$\forall \varphi \in C^2(\mathbb{R})$, $\forall f \in \mathcal{D}(L)$,

$\varphi \circ f \in \mathcal{D}(L)$ and

$$L[\varphi \circ f] = \varphi'(f) L f + \frac{\varphi''(f)}{2} \Gamma(f, f)$$

or again equivalently

$\forall \varphi \in C^2(\mathbb{R})$, $\forall f, g \in \mathcal{D}(L)$

$\varphi \circ f \in \mathcal{D}(L)$ and

$$\Gamma(\varphi(f), g) = \varphi'(f)\,\Gamma(f, g).$$

Markov processes with continuous trajectories are called <u>diffusions</u>.

Let be given a smooth potential

$$U : V \longrightarrow \mathbb{R}$$

whose global minimal we would like to find.

For any inverse temperature $\beta \geq 0$, the equivalent of the Metropolis algorithm is the Langevin generator:

$$L_\beta := \frac{\Delta \cdot}{2} - \beta \langle \nabla U, \nabla \cdot \rangle$$

on $\mathcal{D}(L_\beta) = C^2(V)$.

To compute its reversible measure, let us recall two other Riemannian objects: • <u>the measure</u> ℓ defined in any coordinate system $(x^1, x^2, \ldots, x^d)$ by

$$d\ell = \sqrt{|g(x)|}\, dx^1 \cdots dx^d$$

<u>Stokes' theorem</u> enables us to see

that ℓ is reversible for Δ:

$\forall f, g \in C^2(V)$:

$$\int f \, \Delta g \, d\ell = - \int \langle \nabla f, \nabla g \rangle \, d\ell$$

$$= \int g \, \Delta f \, d\ell$$

- The <u>divergence operator</u> acting on vector fields b, which writes in any coordinate system $(x^1, x^2, \dots x^d)$

$$\operatorname{div}(b)(x) = \frac{1}{\sqrt{|g(x)|}} \sum_{i \in [d]} \partial_i \left(\sqrt{|g|} \, b^i \right)(x)$$

We have $\forall b$ vector field and any $f \in C^1(V)$,

$$\int \langle \nabla f, b \rangle \, d\ell = - \int f \operatorname{div}(b) \, d\ell$$

It follows that

$$\Delta = -\operatorname{div}(\nabla f)$$

Define for $\beta \geq 0$, the Gibbs probability measure

$$\forall x \in V, \quad \Pi_\beta(dx) = \frac{1}{Z_\beta} e^{-2\beta U(x)} \ell(dx)$$

where Z_β is the normalizing constant

Lemma 46

Π_β is reversible for L_β

Proof

Let us check that

$$L_\beta = \frac{1}{2} e^{2\beta U} \operatorname{div}\left(e^{-2\beta U} \nabla \cdot \right)$$

Indeed the r.h.s. is equal to

$$\frac{1}{2} e^{2\beta U} \Big(e^{-2\beta U} \operatorname{div}(\nabla \cdot)$$

$$- 2\beta e^{-2\beta U} \langle \nabla U, \nabla \cdot \rangle \Big)$$

$$= \frac{\operatorname{div}(\nabla \cdot)}{2} - \beta \langle \nabla U, \nabla \cdot \rangle$$

$$= L_\beta$$

It follows that for any $f, g \in \mathcal{D}(L)$

$$\Pi_\beta[f L_\beta g] = \frac{1}{Z_\beta} \int f L_\beta g \; e^{-2\beta U} d\ell$$

$$= \frac{1}{Z_\beta} \int f \, \frac{1}{2} e^{2\beta U} \operatorname{div}\left(e^{-2\beta U} \nabla g\right) e^{-2\beta U} d\rho$$

$$= \frac{1}{2Z_\beta} \int f \operatorname{div}\left(e^{-2\beta U} \nabla g\right) d\rho$$

$$= \frac{1}{2Z_\beta} \int \langle \nabla f, e^{-2\beta U} \nabla g \rangle \, d\rho$$

$$= \frac{1}{2Z_\beta} \int \langle \nabla f, \nabla g \rangle e^{-2\beta U} d\rho$$

$$= \Pi_\beta [g L f]$$

$\square$

In particular, for large $\beta \geq 0$, Π_β concentrates on neighborhoods of $\mathcal{U} = \{ x \in V ; U(x) = \min_V U \}$:

for any such neighorhood $\mathcal{N}$,

$$\lim_{\beta \to +\infty} \Pi_\beta(\mathcal{N}) = 1$$

As in the finite case, the global optimization of U is performed by resorting to simulated annealing.

Let $(\beta_t)_{t \geq 0}$ be an inverse temperature schedule, non-negative, increasing, smooth and satisfying $\lim \beta_t = +\infty$

We consider a corresponding time-inhomogeneous Markov process $(X(t))_{t \geq 0}$ associated to the family of generators $(L_{\beta_t})_{t \geq 0}$.

Define

$\forall t \geq 0, \quad m_t = \mathcal{L}(X_t)$.

It can be shown that for any $t > 0$, $m_t \ll \ell$ with a smooth density f_t.

We want to show that

$$\lim_{t\to+\infty} \| m_t - \Pi_{\beta_t} \|_{tv} = 0 \qquad (35)$$

It will follow that for any neighborhood $\mathcal{N}$ of $\mathcal{U}$,

$$\lim_{t\to+\infty} m_t(\mathcal{N}) = 1$$

which is the wanted concentration.

The proof of (35) is similar to the one provided in the finite case and relies on log-Sob inequalities.

For any $\beta \geq 0$, consider $\alpha(\beta)$ the largest constant $\alpha \geq 0$ such that

$\forall f \in C^2(V)$, $f > 0$, $\Pi_\beta[f] = 1$

$$\alpha \quad \mathrm{Ent}_{\Pi_\beta}(f) \leq \mathcal{E}_{L_\beta}(\sqrt{f}, \sqrt{f})$$

$$\overset{\shortparallel}{-\Pi_\beta[\sqrt{f}\, L_\beta(\sqrt{f})]}$$

$$= \frac{1}{2} \Pi_\beta[\Gamma_{L_\beta}(\sqrt{f}, \sqrt{f})]$$

In the diffusion case there is

no difference between the mod log Sob and log Sob inequalities, up to a factor 4.

Indeed, we have for any $f \in C^2(V)$

$$\Gamma(\sqrt{f}, \sqrt{f}) = \frac{1}{2\sqrt{f}} \Gamma(f, \sqrt{f})$$

$$= \frac{1}{4f} \Gamma(f, f)$$

while

$$\Gamma(f, \ln f) = \frac{1}{f} \Gamma(f, f)$$

Consider the critical constant c defined as follows:

A path is a continuous function $p : [0, T] \longrightarrow V$, for some $T \geq 0$. $p(0)$ and $p(T)$ are called the initial and final points.

The elevation of such a path p is $\max_{0 \leq t \leq T} U(p(t))$.

The absolute communication cost between $x, y \in V$ is

$$h(x,y) := \inf_{\substack{p \text{ path} \\ \text{from } x \\ \text{to } y}} e(p)$$

$$c(U) := \max_{x,y \in V} h(x,y) - U(x) - U(y) + \min_V U$$

$$= \max_{x \in V} \max_{y \in \mathcal{U}} h(x,y) - U(x)$$

Theorem 43 is still valid:

<u>Theorem</u> 47

Consider the evolution

$\forall t \geq 0, \quad \beta_t = \frac{1}{h} \ln(1+t)$

- If $h > c(U)$, then whatever the initial distribution, (35) is satisfied

- If $h < c(U)$, there exists initial conditions such that (35) does not hold.

Here is the equivalent of lemma 45,

due to Holley and Stroock (1988):

Lemma 48

There exists a constant $K>0$ such that for any $\beta \geq 0$,

$$\alpha(\beta) \geq \frac{K}{(1+\beta\|\nabla U\|_\infty)^{5d-1}} e^{-\beta c(U)}$$

The proof is much more technical than in the finite case, even if it follows a similar pattern. It also requires the existence of dimension-dependent Sobolev inequalities.

So let us admit it.

The proof the first point of Theorem 47 is then essentially identical to that of Theorem 43:

Consider for any $t>0$,

$$I_t = \mathrm{Ent}(m_t \mid \pi_{\beta_t})$$

$$= \int \ln f_t \, f_t \, d\pi_{\beta_t}$$

$<+\infty$

(even if we can have $I_0 = +\infty$).

Differentiating, we get

$$\dot{I}_t = \int (1 + \ln f_t)\, \dot{f}_t \, d\Pi_{\beta_t} - \int \ln f_t \, f_t \left(\frac{d}{dt} \ln \Pi_{\beta_t} \right) d\Pi_{\beta_t}$$

$$= \int L_{\beta_t}[\ln f_t]\, f_t \, d\Pi_{\beta_t}$$

$$+ \dot{\beta}_t \underbrace{\int \ln f_t \, f_t \, (U - \Pi_{\beta_t}(U))\, d\Pi_{\beta_t}}_{\int \big(\underbrace{f_t \ln f_t + \frac{1}{e}}_{\geq 0}\big)(U - \Pi_{\beta_t}[U])\, d\Pi_{\beta_t}}$$

$$\leq \operatorname{osc}(U) \int f_t \ln f_t + \frac{1}{e} \, d\Pi_{\beta_t}$$

$$= \operatorname{osc}(U) \left(I_t + \frac{1}{e} \right)$$

Thus for any $t > 0$

$$\dot{I}_t \leq -\left(\alpha(L_{\beta_t}) - \frac{\operatorname{osc}(U)}{e} \dot{\beta}_t \right) I_t$$

$$+ \frac{osc(u)}{e} \dot{\beta}_t$$

For $h > c$,

$$\alpha(L_{\beta_t}) - \frac{osc(u)}{e} \dot{\beta}_t \sim \alpha(L_{\beta_t})$$

$$\gtrsim \frac{K'}{\ln^{5d-1}(1+t)} \left(\frac{1}{1+t}\right)^{-c/h}$$

For a certain constant $K' > 0$,

and
$$\lim_{t \to +\infty} \frac{osc(u)\dot{\beta}_t}{e\left(\alpha(L_{\beta_t}) - \frac{osc(u)}{e} \dot{\beta}_t\right)}$$

$$= 0$$

Gronwall's lemma implies

$$\lim_{t \to +\infty} I_t = 0$$

as wanted.

This method enables to get quantitative bounds.

All the extensions alluded in the finite case are also valid for compact Riemannian manifolds.

Remarks

(a) An advantage of the spectrum of the (opposite of the) Laplace-Beltrami operator on a compact Riemannian manifold is that its spectrum is discrete and diverging to $+\infty$, so it has a spectral gap. This is no longer necessarily true on a (complete) non-compact Riemannian manifold. For instance for $-\partial^2$ on $\mathbb{R}$, the spectrum is continuous and consists of $\mathbb{R}_+$.

For a Langevin operator on a non-compact manifold, to get a spectral gap requires that U explodes sufficiently fast at infinity.

E.g. on $\mathbb{R}$, with

$U(x) = |x|^a$ for $|x|$ sufficiently large, $L = \frac{\partial^2}{2} - U'\partial$

has a (positive) spectral gap if and only if $a \geq 1$ and there is a (positive) log-Sob constant if and only if $a \geq 2$.

The first example of log-Sob inequality is due to Gross (1975) for the Ornstein-Uhlenbeck operator

$$\frac{\Delta}{2} - \langle \frac{x}{2}, \nabla \cdot \rangle$$

(corresponding to $U(x) = \frac{\|x\|^2}{4}$ on $\mathbb{R}^d$).

Its reversible measure is the standard Gaussian distribution.

The proof used the log-Sob inequality for the symmetric random walk on $\mathbb{Z}_2$, the tensorization property and the central limit theorem to approximate the one-dimensional Gaussian.

(b) The Riemannian geometry structure is entirely encapsulated in the Laplace-Beltrami operator (or equivalently in the Brownian motion!).

Indeed instead of defining the scalar product on vectors, it is sufficient to define it on 1-forms.

So let $x \in V$ and $\xi_1, \xi_2 \in T_x^* V$. Find two functions f, g such that

$$df(x) = \xi_1, \quad dg(x) = \xi_2$$

Then we have

$$\langle \xi_1, \xi_2 \rangle = \Gamma_{\frac{\Delta}{2}}(f, g)(x)$$

$$= \langle \nabla f, \nabla g \rangle$$

(c) The analytico-probabilist approach of log-Sob initiated by Bakry-Emery (1985) has led to a lot of developments in Riemannian geometry and its generalisations, as well as in optimal transport theory.

This field of functional inequalities has fully benefited from being at the interface of analysis, probability and geometry.