Asymptotic Behavior of Markov Processes: a Dive into the Sub-Geometric Case

Armand Bernou

LJLL, Sorbonne Université

February 1, 2021

Markov processes and stability issues

Markov process: a stochastic process $(X_t)_{t\geq 0}$ which depends on its past only through its present.

Some questions for a number of Markov processes are focused on the stability structure:

1. is there an invariant measure ?
2. do we have a form of convergence towards it ?
3. at which rate does this convergence occur ?

The many dimensions of the problem

- structure of the state space: countable or not,
- structure of the time (Markov chains and Markov processes),
- sub-geometric or geometric nature of the convergence.

The many dimensions of the problem

- structure of the state space: countable or not,
- structure of the time (Markov chains and Markov processes),
- sub-geometric or geometric nature of the convergence.

Today: non-countable state space, continuous time, geometric and sub-geometric convergence. We will only focus on the convergence in the total variation distance: if μ, ν are two measures on E,

$$\|\mu - \nu\|_{TV} = \sup_{A \in \mathcal{B}(E)} |\mu(A) - \nu(A)|,$$

but many results are available for norms of the form

$$\|\mu\|_f = \sup_{|g| \leq f} |\mu(g)|.$$

For $f \equiv 1 \rightarrow$ total variation norm.

Applications in statistics: convergence rate of the MCMC algorithms

Those norms are especially useful for statisticians. Suppose you want to compute $\mathbb{E}[f(Y)]$ with $Y \sim \pi$ and find a process $(X_t)_{t\geq 0}$ with $X_0 = x$ such that its law $\mathcal{P}_t(x, \cdot)$ converges to π.

Applications in statistics: convergence rate of the MCMC algorithms

Those norms are especially useful for statisticians. Suppose you want to compute $\mathbb{E}[f(Y)]$ with $Y \sim \pi$ and find a process $(X_t)_{t \geq 0}$ with $X_0 = x$ such that its law $\mathcal{P}_t(x, \cdot)$ converges to π.

The techniques of today allow you to understand how fast

$$\|\mathcal{P}_t(x, \cdot) - \pi\|_f = |\mathbb{E}_x[f(X_t)] - \mathbb{E}[f(Y)]|,$$

converges towards 0, which may save you a lot of time.

Applications in statistics: convergence rate of the MCMC algorithms

Those norms are especially useful for statisticians. Suppose you want to compute $\mathbb{E}[f(Y)]$ with $Y \sim \pi$ and find a process $(X_t)_{t\geq 0}$ with $X_0 = x$ such that its law $\mathcal{P}_t(x,\cdot)$ converges to π.

The techniques of today allow you to understand how fast

$$\|\mathcal{P}_t(x,\cdot) - \pi\|_f = |\mathbb{E}_x[f(X_t)] - \mathbb{E}[f(Y)]|,$$

converges towards 0, which may save you a lot of time.

In particular, this allows you to understand the asymptotic behavior of Langevin tampered distribution (Fort-Roberts 2005, Douc-Fort-Guillin 2009).

First assumptions

Consider a process $(X_t)_{t\geq 0}$ on a locally compact, separable metric space E, σ-field $\mathcal{B}(E)$. We assume that $(X_t)_{t\geq 0}$ is time-homogeneous, strong Markov, càdlàg, write $(\mathcal{P}_t)_{t\geq 0}$ its associated semigroup, $\mathcal{L}$ the corresponding generator.

Definition

A non-empty measurable set C is petite if there exist a probability measure a on $\mathcal{B}(\mathbb{R}_+)$ and a non-trivial σ-finite measure μ on $\mathcal{B}(E)$ such that

$$\forall x \in C, \int_0^{\infty} \mathcal{P}_t(x,\cdot)a(dt) \geq \mu(\cdot).$$

For many cases, when $(X_t)_{t\geq 0}$ is Feller (i.e. $\lim \mathit{limits}_{t\to 0^+}\mathbb{E}_x[f(X_t)] = f(x)$ for all $f \in C_0(E)$) all compact sets are petite. Often, when we try to identify a petite set, we consider a compact one.

Stability structure assumptions

We will require the following properties

Stability structure assumptions

We will require the following properties

- Harris-recurrence (implies irreducibility): there exists a measure ν on $\mathcal{B}(E)$ such that $\nu(A) > 0$ implies

$$\mathbb{P}_x\Big[\int_0^\infty \mathbf{1}_A(X_s)ds = \infty\Big] = 1, \quad \text{for all } x \in E.$$

  This implies the existence of an invariant measure $\tilde{\pi}$.

Stability structure assumptions

We will require the following properties

- Harris-recurrence (implies irreducibility): there exists a measure ν on $\mathcal{B}(E)$ such that $\nu(A) > 0$ implies

  $$\mathbb{P}_x\Big[\int_0^\infty \mathbf{1}_A(X_s)ds = \infty\Big] = 1, \quad \text{for all } x \in E.$$

  This implies the existence of an invariant measure $\tilde{\pi}$.
- Positive Harris-recurrence: there exists an invariant probability measure π.

Stability structure assumptions

We will require the following properties

- Harris-recurrence (implies irreducibility): there exists a measure ν on $\mathcal{B}(E)$ such that $\nu(A) > 0$ implies

$$\mathbb{P}_x\Big[\int_0^\infty \mathbf{1}_A(X_s)ds = \infty\Big] = 1, \quad \text{for all } x \in E.$$

  This implies the existence of an invariant measure $\tilde{\pi}$.
- Positive Harris-recurrence: there exists an invariant probability measure π.
- Aperiodicity: there exists a μ_{δ_m} petite set C, $t_0 > 0$ such that for all $x \in C$, $t \geq t_0$, $P^t(x, C) > 0$.

Stability structure assumptions

We will require the following properties

- Harris-recurrence (implies irreducibility): there exists a measure ν on $\mathcal{B}(E)$ such that $\nu(A) > 0$ implies

$$\mathbb{P}_x\left[\int_0^\infty \mathbf{1}_A(X_s)ds = \infty\right] = 1, \quad \text{for all } x \in E.$$

  This implies the existence of an invariant measure $\tilde{\pi}$.
- Positive Harris-recurrence: there exists an invariant probability measure π.
- Aperiodicity: there exists a μ_{δ_m} petite set C, $t_0 > 0$ such that for all $x \in C$, $t \geq t_0$, $P^t(x, C) > 0$.
- The process is non-explosive: let $(O_n)_{n\geq 0}$ be a sequence of precompact sets with $O_n \uparrow E$, T^m be the first entrance time into O_m^c, and let

$$\zeta := \lim_{m\to\infty} T^m.$$

  Then $\mathbb{P}_x(\zeta = \infty) = 1$ for all $x \in E$.

A first tool: delayed stopping times

Define, for all set C, $\delta > 0$,

$$\tau_C(\delta) = \inf\{t > \delta, X_t \in C\}.$$

Theorem (Meyn-Tweedie 1993)

Assume $(X_t)_{t \geq 0}$ is irreducible, non-explosive and aperiodic. Let $C \in \mathcal{B}(E)$ be a petite set, assume $\mathbb{P}_x(\tau_C < \infty) \equiv 1$, and that for some $\delta > 0$,

$$\sup_{x \in C} \mathbb{E}_x[\tau_C(\delta)] < \infty.$$

Then $(X_t)_{t \geq 0}$ is positive Harris recurrent. In fact, we also have ergodicity (convergence towards the invariant probability measure at infinity).

A second tool: Lyapunov inequalities

Some inequalities for the generator applied to a norm-like function $V : E \to \mathbb{R}_+$ with $V(x) \to \infty$ as $|x| \to \infty$ (i.e. $\{x : V(x) \leq B\}$) is precompact for all $B > 0$) also gives key properties. For instance, if there exists $c, d \geq 0$ constant such that for all $x \in E$,

$$\mathcal{L}V(x) \leq cV(x) + d,$$

then we have non-explosion.

Proposition (Meyn-Tweedie 1998)

Assume that $(X_t)_{t\geq 0}$ is non-explosive, irreducible and aperiodic. Then if there exists C a petite set and $V \geq 0$ a test function with $V(x_0) < \infty$ for some $x_0 \in E$, $b > 0$ constant satisfying, for all $x \in E$,

$$\mathcal{L}V(x) \leq -1 + b\mathbf{1}_C(x),$$

the process is positive Harris-recurrent. This also implies ergodicity.

The geometric case

Theorem (Meyn-Tweedie 1998)

Assume that $(X_t)_{t\geq 0}$ *is non-explosive, irreducible and aperiodic. The two following conditions are equivalent:*

1. *there exists some compact petite set* $C \in \mathcal{B}(E)$*, some* $\delta > 0$ *and* $\kappa > 1$ *such that for all* $x \in E$*,*

$$\mathbb{E}_x[\kappa^{\tau_C(\delta)}] < \infty \text{ and } \sup_{x\in C} \mathbb{E}_x[\kappa^{\tau_C(\delta)}] < \infty;$$

The geometric case

Theorem (Meyn-Tweedie 1998)

Assume that $(X_t)_{t\geq 0}$ is non-explosive, irreducible and aperiodic. The two following conditions are equivalent:

1. *there exists some compact petite set $C \in \mathcal{B}(E)$, some $\delta > 0$ and $\kappa > 1$ such that for all $x \in E$,*

$$\mathbb{E}_x[\kappa^{\tau_C(\delta)}] < \infty \text{ and } \sup_{x\in C} \mathbb{E}_x[\kappa^{\tau_C(\delta)}] < \infty;$$

2. *there exists a compact petite set C, constants $b < \infty, \beta > 0$ and $V \geq 1$ function finite at some $x_0 \in E$ such that*

$$\mathcal{L}V(x) \leq -\beta V(x) + b\mathbf{1}_C(x), \qquad x \in E.$$

The geometric case

Theorem (Meyn-Tweedie 1998)

Assume that $(X_t)_{t\geq 0}$ *is non-explosive, irreducible and aperiodic. The two following conditions are equivalent:*

1. *there exists some compact petite set* $C \in \mathcal{B}(E)$, *some* $\delta > 0$ *and* $\kappa > 1$ *such that for all* $x \in E$,
$$\mathbb{E}_x[\kappa^{\tau_C(\delta)}] < \infty \text{ and } \sup_{x\in C} \mathbb{E}_x[\kappa^{\tau_C(\delta)}] < \infty;$$
2. *there exists a compact petite set* C, *constants* $b < \infty, \beta > 0$ *and* $V \geq 1$ *function finite at some* $x_0 \in E$ *such that*
$$\mathcal{L}V(x) \leq -\beta V(x) + b\mathbf{1}_C(x), \qquad x \in E.$$

Both conditions imply that for some $\rho < 1$, *for all* $x \in E$ *with* $V(x) < \infty$
$$\lim_{t\to\infty} \rho^{-t} \|\mathcal{P}_t(x, \cdot) - \pi(\cdot)\|_{TV} = 0.$$

Example: the Ornstein-Uhlenbeck process

Let $(X_t)_{t\geq 0}$ be solution to the following SDE on $\mathbb{R}$:

$$dX_t = \sqrt{2}dB_t - X_t dt,$$

with $(B_t)_{t\geq 0}$ the standard Brownian motion. The stochastic generator is given, for all $f \in C^2(\mathbb{R})$, by

$$\mathcal{L}f = \partial^2_{xx} f - x\partial_x f.$$

We can show that this equation has an invariant distribution given by $\mu_\infty(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

Example: the Ornstein-Uhlenbeck process II

Let $V : \mathbb{R} \to [1, \infty)$ defined by $V(x) = e^{a|x|}$ for some $a > 0$. Then, for $x > 0$,

$$\begin{aligned}\mathcal{L}V(x) = a^2 e^{ax} - xae^{ax} &\leq (a^2 - xa)e^{ax}(\mathbf{1}_{x\in(0,2a]} + \mathbf{1}_{x\in(2a,\infty)}) \\ &\leq -a^2 V(x) + a^2 e^{2a^2} \mathbf{1}_{\{|x|\leq 2a\}}.\end{aligned}$$

A similar computation can be done for $x \leq 0$. For this process $\{x : |x| \leq 2a\}$ is petite (this is based on the Feller property). This proves the exponential convergence.

The sub-geometric case

Theorem (Douc-Fort-Guillin, Hairer)

Assume that $(X_t)_{t\geq 0}$ *is non-explosive, irreducible, aperiodic. Let* $\phi : [1,\infty) \to \mathbb{R}_+^*$ C^1, *strictly increasing, strictly concave (+ technical properties). Define* $H_\phi(u) = \int_1^u \frac{ds}{s}$ *for all* $u \geq 1$, *and let* $H_\phi^{-1} : [0,\infty) \to [1,\infty)$ *be its inverse function. Consider the two following conditions:*

1. *there exists* C *compact, petite,* $\delta > 0$ *such that*
$$\mathbb{E}_x[H_\phi^{-1}(\tau_C(\delta))] < \infty \text{ for all } x \in E, \qquad \sup_{x\in C} \mathbb{E}_x[H_\phi^{-1}(\tau_C(\delta))] < \infty;$$
2. *there exists a compact petite subset* C *of* E, $K > 0$ *constant and* $V : E \to [1,\infty)$ *continuous with precompact sublevel sets such that for all* $x \in E$,
$$\mathcal{L}V(x) \leq -\phi(V(x)) + K\mathbf{1}_C(x).$$

In those two cases, there exists an invariant probability measure π *on* E *such that for all* $x \in E$,

$$\lim_{t\to\infty} \phi(H_\phi^{-1}(t))\|\mathcal{P}_t(x,\cdot) - \pi(\cdot)\|_{TV} = 0.$$

Typical rates that one obtains are of the form

$$r(t) = t^{\alpha} \ln(t)^{\beta} \exp(\gamma t^{\eta}), \qquad \text{with } \eta \in (0,1) \text{ and } \begin{cases} \gamma > 0, \alpha, \beta \in \mathbb{R} \text{ or,} \\ \gamma = 0, \alpha > 0, \beta \in \mathbb{R} \text{ or,} \\ \gamma = \alpha = 0, \beta > 0. \end{cases}$$

Typical rates that one obtains are of the form

$$r(t) = t^{\alpha} \ln(t)^{\beta} \exp(\gamma t^{\eta}), \qquad \text{with } \eta \in (0,1) \text{ and } \begin{cases} \gamma > 0, \alpha, \beta \in \mathbb{R} \text{ or,} \\ \gamma = 0, \alpha > 0, \beta \in \mathbb{R} \text{ or,} \\ \gamma = \alpha = 0, \beta > 0. \end{cases}$$

Example: if $\alpha \in (0,1)$ and $\phi(x) = x^{\alpha}$, then $\phi(H_{\phi}^{-1}(x)) \sim x^{\frac{\alpha}{1-\alpha}}$.

Typical rates that one obtains are of the form

$$r(t) = t^{\alpha} \ln(t)^{\beta} \exp(\gamma t^{\eta}), \qquad \text{with } \eta \in (0,1) \text{ and } \begin{cases} \gamma > 0, \alpha, \beta \in \mathbb{R} \text{ or,} \\ \gamma = 0, \alpha > 0, \beta \in \mathbb{R} \text{ or,} \\ \gamma = \alpha = 0, \beta > 0. \end{cases}$$

Example: if $\alpha \in (0,1)$ and $\phi(x) = x^{\alpha}$, then $\phi(H_{\phi}^{-1}(x)) \sim x^{\frac{\alpha}{1-\alpha}}$.

Remark

In constrast with the exponential case, there is no equivalence between the two conditions in the sub-geometric theorem.

A typical example for the sub-geometric case: the gradient dynamic on $\mathbb{R}$

We consider the process $(X_t)_{t\geq 0}$ solution to the SDE

$$dX_t = -\partial_x V(X_t)dt + \sqrt{2}dB_t,$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion, and

$$V(x) = 2(1+|x|^2)^{\frac{1}{4}}, \qquad x \in \mathbb{R}.$$

The stochastic generator is given by

$$\mathcal{L} = \partial^2_{xx} - \partial_x V \partial_x,$$

and the equilibrium distribution is $\mu_\infty(x) \propto e^{-V(x)}$.

A typical example for the sub-geometric case: the gradient dynamic on $\mathbb{R}$ II

Let $W(x) = e^{\alpha V(x)}$ with $\alpha \in (0,1)$ constant. We have

$$\mathcal{L}W(x) = \alpha W(x)(1+x^2)^{-\frac{7}{4}}\big(1 - \frac{1}{2}x^2 + (\alpha - 1)x^2(1+x^2)^{\frac{1}{4}}\big).$$

In the bracket $\to$ a negative quantity upper bounded by some constant outside a compact set $C := \{x : W(x) \leq \bar{W}\}$, $\bar{W} > 0$ constant. Hence, for two constants $\beta, K > 0$, we have

$$\mathcal{L}W \leq -\beta \frac{W}{\ln(W)^7} + K\mathbf{1}_C.$$

A typical example for the sub-geometric case: the gradient dynamic on $\mathbb{R}$ II

Let $W(x) = e^{\alpha V(x)}$ with $\alpha \in (0,1)$ constant. We have

$$\mathcal{L}W(x) = \alpha W(x)(1+x^2)^{-\frac{7}{4}}\big(1 - \frac{1}{2}x^2 + (\alpha - 1)x^2(1+x^2)^{\frac{1}{4}}\big).$$

In the bracket $\to$ a negative quantity upper bounded by some constant outside a compact set $C := \{x : W(x) \leq \bar{W}\}$, $\bar{W} > 0$ constant. Hence, for two constants $\beta, K > 0$, we have

$$\mathcal{L}W \leq -\beta \frac{W}{\ln(W)^7} + K\mathbf{1}_C.$$

Once again, all compact sets are petite, and for $\phi(x) = \frac{x}{\ln(x)^7}$, we find a final rate

$$r(t) = t^{-\frac{7}{8}} e^{ct^{\frac{1}{8}}}$$

for some constant $c > 0$.

The few things I haven't mentioned

1. To go from Condition 2 in the subgeometric theorem to f-ergodicity $\rightarrow$ Young's functions and interpolation (you can find this in DFG 2009 or Fort-Roberts 2005).

The few things I haven't mentioned

1. To go from Condition 2 in the subgeometric theorem to f-ergodicity $\rightarrow$ Young's functions and interpolation (you can find this in DFG 2009 or Fort-Roberts 2005).

2. A recent result (B. 2020) provides two new conditions for the sub-geometric case, one with a randomized hitting time, one with a Lyapunov inequalities for a function depending also on times, that are equivalent and lie between conditions 2 and 1 of the previous theorem.

The few things I haven't mentioned

1. To go from Condition 2 in the subgeometric theorem to f-ergodicity $\rightarrow$ Young's functions and interpolation (you can find this in DFG 2009 or Fort-Roberts 2005).

2. A recent result (B. 2020) provides two new conditions for the sub-geometric case, one with a randomized hitting time, one with a Lyapunov inequalities for a function depending also on times, that are equivalent and lie between conditions 2 and 1 of the previous theorem.

3. The results presented here are not optimal (compactness assumptions can be relaxed).

Thank you

Thank you for your attention !