Coupling methods for the convergence rate of Markov processes

Joint work with Nicolas Fournier (SU)

Armand Bernou

LPSM, Sorbonne Université

January 13, 2020

Outline

Motivations

Consider a process satisfying the strong Markov property $(X_t)_{t \in T}$ with $T = \mathbb{N}$ (Markov chain) or $T = \mathbb{R}_+$ (Markov process)

Assume $(X_t)_{t \in T}$ takes value in a complete metric space E, endowed with a σ-algebra $\mathcal{E}$.

Motivations

Consider a process satisfying the strong Markov property $(X_t)_{t \in T}$ with $T = \mathbb{N}$ (Markov chain) or $T = \mathbb{R}_+$ (Markov process)

Assume $(X_t)_{t \in T}$ takes value in a complete metric space E, endowed with a σ-algebra $\mathcal{E}$.

Denote μ_0 for the distribution of X_0, and assume that $\mathcal{L}(X_t) \to \pi$, π invariant distribution.

Question for today: how can we compute the rate of this convergence ?

Motivations II

Typical example: you buy a new deck of cards, hence cards are ordered. How long do you need to shuffle them before the distribution is really uniform ?

This question is strongly related to the study of cutoff phenomenon for Markov chains (Diaconis, Shahshahani, Aldous)

Motivations II

Typical example: you buy a new deck of cards, hence cards are ordered. How long do you need to shuffle them before the distribution is really uniform ?

This question is strongly related to the study of cutoff phenomenon for Markov chains (Diaconis, Shahshahani, Aldous)

For the general question, lots of techniques for Markov chains: Geometric methods (Dirichlet energy), spectral methods, Meyn-Tweedie theory, Wilson's notions of mixing...

Today: focus on one method which translates in continuous time, the coupling method.

Total variation distance

Consider S the state space, countable for simplicity (but this is absolutely not a restriction). Let $\mathcal{M}(S)$ the set of probability measures on S.

Definition

Let $\mu, \nu \in \mathcal{M}(S)$. We define the total variation distance as

$$\|\mu - \nu\|_{TV} = \sup_{A \subset S} |\mu(A) - \nu(A)|.$$

Idea: you try to approximate μ with ν, and you evaluate your approximation with the worst possible case.

Coupling: basic definition

Definition

Given $\mu, \nu \in \mathcal{M}(S)$*, a coupling of* μ *and* ν *is a couple* (X, Y) *of random variables such that* $X \sim \mu$*,* $Y \sim \nu$*.*

Example (Dummy examples)

- *If* $X \sim \mu$*,* $Y \sim \nu$ *with* X *independent from* Y*,* (X, Y) *is a coupling of* μ *and* ν*.*
- *If* $\mu = \nu$*, letting* $X \sim \mu$ *and taking* $Y = X$ *gives a coupling.*

Coupling: A more interesting example

Example

Consider a sequence $(B_i)_{i\geq 1}$ *of independent events with* $\mathbb{P}(B_i) \geq c > 0$ *for all* i*. Set* $N = \inf\{i \geq 1, B_i \text{ is realized}\}$*. Then, coupling gives us a way of saying that there exists* $G \sim \mathcal{G}(c)$ *such that "*$N \leq G$*".*

Indeed, consider a sequence $(U_i)_{i\geq 1}$ *of i.i.d. uniform random variables on* $[0,1]$ *and set*

$$\alpha_i = \mathbf{1}_{\{U_i \leq \mathbb{P}(B_i)\}}, \quad \beta_i = \mathbf{1}_{\{U_i \leq c\}}.$$

Clearly $N \overset{\mathcal{L}}{=} \inf\{i \geq 1, \alpha_i = 1\}$ *and setting* $G = \inf\{i \geq 1, \beta_i = 1\}$*,* $G \sim \mathcal{G}(c)$*. Moreover,* $\alpha_i \geq \beta_i$ *for all* i *so "*$N \leq G$*".*

Coupling and total variation distance

We have the following

Theorem (Relation between coupling and total variation)

Let $\mu, \nu \in \mathcal{M}(S)$. For all coupling (X, Y) of μ and ν,

$$\|\mu - \nu\|_{TV} \leq \mathbb{P}(X \neq Y).$$

Moreover, there exists a coupling (X_*, Y_*) such that we have equality.

An application: Random-to-top shuffling

Assume we have a deck of cards of size n. We label the cards from 1 to n. Hence the state space is

$$S_n = \{\text{permutations of } \{1, \dots, n\}\}.$$

We perform a random-to-top shuffling, i.e. at each step we pick a card in the deck at random and put it on top.

By symmetry, assume that $X_0 = Id$. Take another chain $(Y_n)_{n \geq 0}$ such that $Y_0 \sim \mathcal{U}(S_n) = \pi$. Since π is invariant, $Y_n \sim \pi$ for all $n \geq 0$.

Random-to-top shuffling II

At each step, pick $i \in \{1, \ldots, n\}$ at random, find the card i and put in on top in both decks .

This is indeed a random-to-top dynamic. Advantage: when card number i is picked, it remains in the same position in both decks for ever .

To have $X_n = Y_n$, we thus just need to have selected all cards at time n. This is the "coupon-collector" problem: the time to collect all elements of a collection of size n when picking uniformly converges in probability towards $n \ln(n)$. Set

$$\tau_n = \inf\{t > 0, \text{ all cards have been selected }\}.$$

Random-to-top shuffling III

We conclude that if $t = (1 + \epsilon)n \ln(n)$ for some $\epsilon > 0$, writing μ_k the distribution of X_k,

$$\|\mu_t - \pi\|_{TV} \leq \mathbb{P}(X_t \neq Y_t) \leq \mathbb{P}(\tau_n > t) \to 0,$$

as $n \to \infty$.

Random-to-top shuffling III

We conclude that if $t = (1+\epsilon)n\ln(n)$ for some $\epsilon > 0$, writing μ_k the distribution of X_k,

$$\|\mu_t - \pi\|_{TV} \leq \mathbb{P}(X_t \neq Y_t) \leq \mathbb{P}(\tau_n > t) \to 0,$$

as $n \to \infty$.

Key ideas to keep in mind: to analyse the convergence towards equilibrium,

- We considered two instances of the process, one of which was distributed according to π.
- We have exhibited a realisation of the dynamic leading to equality (here, selecting all cards).
- We have estimated the time necessary for this realisation to occur.

Application to collisionless kinetic theory

Consider a gas enclosed in a vessel (bounded domain) $D \subset \mathbb{R}^n$ with $n \in 2, 3$. We study the phase space, hence the density of the gas depends on $t \geq 0$, $x \in D$, $v \in \mathbb{R}^n$.

We focus on Knudsen (rarefied) gas: no collisions between particles. Particle in position (x, v) in the phase space just evolves along v until it collides with the boundary.

The randomness comes from the boundary, where particles are reflected stochastically (diffuse reflection).

Free-transport equation with boundary condition

The corresponding PDE is given, for $f_0 \in L^1(D \times \mathbb{R}^n)$, with n_x the unit inward normal vector at $x \in \partial D$,

$$\begin{cases} \partial_t f(t,x,v) + v \cdot \nabla_x f(t,x,v) = 0, & (t,x,v) \in (0,\infty) \times D \times \mathbb{R}^n, \\ f(0,x,v) = f_0(x,v), & x \in D, v \in \mathbb{R}^n, \\ f(t,x,v) = cM(v)Kf(t,x), & t > 0, x \in \partial D, v \cdot n_x > 0, \end{cases}$$

where c is a renormalization constant and the flux $Kf(t,x)$ at the point $x \in \partial D$ at time t is given by

$$Kf(t,x) = \int_{v' \cdot n_x < 0} f(t,x,v')|v' \cdot n_x| dv'.$$

Typically $M(v) = \exp(-|v|^2)\frac{1}{(2\pi)^{n/2}}$. In all cases, M has radial symmetry.

Main result

We write $L^1(D \times \mathbb{R}^n)$ for $L^1(D \times \mathbb{R}^n, \mathrm{Leb})$. We know that there exists $f_\infty \propto M(v)$.

Main result

We write $L^1(D \times \mathbb{R}^n)$ for $L^1(D \times \mathbb{R}^n, \mathrm{Leb})$. We know that there exists $f_\infty \propto M(v)$.

Theorem (B., Fournier)

Let $f_0 \in L^1(D \times \mathbb{R}^n)$, writing f_t for the unique solution of the problem at time $t \geq 0$, if $r : \mathbb{R}_+ \to \mathbb{R}_+$ is increasing with $r(x+y) \leq C_1(r(x)+r(y))$ for some $C_1 > 0$ and such that

$$\int_{D\times\mathbb{R}^n} r\Big(\frac{1}{|v|}\Big) f_0(x,v)dvdx + \int_{D\times\mathbb{R}^n} r\Big(\frac{1}{|v|}\Big) M(v)dvdx < \infty,$$

then for all $t \geq 0$, for some constant $C > 0$,

$$\|f_t - f_\infty\|_{L^1} \leq \frac{C}{r(t)}.$$

Remark

The solution is a solution in the weak sense of measures.

Proof strategy

One can show that if $f_0 \in L^1(D \times \mathbb{R}^n)$, for all $t > 0$, $f_t \in L^1(D \times \mathbb{R}^n)$. Then the L^1 distance and the total variation distance are equivalent

$$\|f_t - f_\infty\|_{L^1} \propto \|f_t - f_\infty\|_{TV},$$

where we recall that, for μ, ν two probability measures from $(E, \mathcal{E})$ to $(F, \mathcal{F})$, by the previous theorem,

$$\|\mu - \nu\|_{TV} = \inf_{X \sim \mu, Y \sim \nu} \mathbb{P}(X \neq Y).$$

Proof strategy II

Idea: Construct $(X_t, V_t, \tilde{X}_t, \tilde{V}_t)_{t\geq 0}$ such that for all $t \geq 0$, $(X_t, V_t) \sim f_t$ and $(\tilde{X}_t, \tilde{V}_t) \sim f_\infty$. Then set

$$\tau = \inf\{t > 0 : (X_{t+s}, V_{t+s})_{s\geq 0} = (\tilde{X}_{t+s}, \tilde{V}_{t+s})_{s\geq 0}\}.$$

Proof strategy II

Idea: Construct $(X_t, V_t, \tilde{X}_t, \tilde{V}_t)_{t\geq 0}$ such that for all $t \geq 0$, $(X_t, V_t) \sim f_t$ and $(\tilde{X}_t, \tilde{V}_t) \sim f_\infty$. Then set

$$\tau = \inf\{t > 0 : (X_{t+s}, V_{t+s})_{s\geq 0} = (\tilde{X}_{t+s}, \tilde{V}_{t+s})_{s\geq 0}\}.$$

Using Markov's inequality we then get

$$\|f_t - f_\infty\|_{L^1} \lesssim \mathbb{P}\Big((X_t, V_t) \neq (\tilde{X}_t, \tilde{V}_t)\Big) \lesssim \mathbb{P}(\tau > t) \lesssim \frac{\mathbb{E}[r(\tau)]}{r(t)}.$$

Stochastic formulation

To build $(X_t, V_t)_{t\geq 0}$, we introduce two functions:

$$\zeta(x,v) = \begin{cases} \inf\{t > 0, x + tv \in \partial D\}, & x \in D \text{ or } x \in \partial D, v \cdot n_x > 0, \\ 0 & x \in \partial D, v \cdot n_x \leq 0, \end{cases}$$

$$q(x,v) = x + \zeta(x,v)v.$$

Stochastic formulation

To build $(X_t, V_t)_{t\geq 0}$, we introduce two functions:

$$\zeta(x,v) = \begin{cases} \inf\{t>0, x+tv \in \partial D\}, & x \in D \text{ or } x \in \partial D, v\cdot n_x > 0, \\ 0 & x \in \partial D, v \cdot n_x \leq 0, \end{cases}$$

$$q(x,v) = x + \zeta(x,v)v.$$

- Pick $(X_0, V_0) \sim f_0$, set $T_0 = 0$, $T_1 = T_0 + \zeta(X_0, V_0)$.

Stochastic formulation

To build $(X_t, V_t)_{t\geq 0}$, we introduce two functions:

$$\zeta(x,v) = \begin{cases} \inf\{t > 0, x + tv \in \partial D\}, & x \in D \text{ or } x \in \partial D, v \cdot n_x > 0, \\ 0 & x \in \partial D, v \cdot n_x \leq 0, \end{cases}$$

$$q(x,v) = x + \zeta(x,v)v.$$

- Pick $(X_0, V_0) \sim f_0$, set $T_0 = 0$, $T_1 = T_0 + \zeta(X_0, V_0)$.
- For $t \in (T_0, T_1)$, set $X_t = X_0 + tV_0$, $V_t = V_0$.

Stochastic formulation

To build $(X_t, V_t)_{t\geq 0}$, we introduce two functions:

$$\zeta(x,v) = \begin{cases} \inf\{t > 0, x + tv \in \partial D\}, & x \in D \text{ or } x \in \partial D, v \cdot n_x > 0, \\ 0 & x \in \partial D, v \cdot n_x \leq 0, \end{cases}$$

$$q(x,v) = x + \zeta(x,v)v.$$

- Pick $(X_0, V_0) \sim f_0$, set $T_0 = 0$, $T_1 = T_0 + \zeta(X_0, V_0)$.
- For $t \in (T_0, T_1)$, set $X_t = X_0 + tV_0$, $V_t = V_0$.
- Set $X_{T_1} = q(X_0, V_0)$, $V_{T_1} = R_1\vartheta(X_{T_1}, \Theta_1)$. Set $T_2 = T_1 + \zeta(X_{T_1}, V_{T_1})$.

Stochastic formulation

To build $(X_t, V_t)_{t\geq 0}$, we introduce two functions:

$$\zeta(x,v) = \begin{cases} \inf\{t>0, x+tv \in \partial D\}, & x \in D \text{ or } x \in \partial D, v\cdot n_x > 0, \\ 0 & x \in \partial D, v \cdot n_x \leq 0, \end{cases}$$

$$q(x,v) = x + \zeta(x,v)v.$$

- Pick $(X_0, V_0) \sim f_0$, set $T_0 = 0$, $T_1 = T_0 + \zeta(X_0, V_0)$.
- For $t \in (T_0, T_1)$, set $X_t = X_0 + tV_0$, $V_t = V_0$.
- Set $X_{T_1} = q(X_0, V_0)$, $V_{T_1} = R_1 \vartheta(X_{T_1}, \Theta_1)$. Set $T_2 = T_1 + \zeta(X_{T_1}, V_{T_1})$.
- ...

R_1 random variable in $\mathbb{R}_+$ (new norm), Θ_1 random variable in $(-\frac{\pi}{2}, \frac{\pi}{2}) \times [0,\pi]^{n-2}$ such that, for $x \in \partial D$,

$$R_1 \vartheta(x, \Theta_1) \sim M(v)|v\cdot n_x| \mathbf{1}_{\{v\cdot n_x > 0\}}.$$

A key remark

1. If $V \sim M(v)$,
2. or if $V \sim \int_D f_0(x,v)dx$,
3. or if $V \sim M(v)|v \cdot n_x|\mathbf{1}_{\{v \cdot n_x > 0\}}$ for some $x \in \partial D$,

with the hypothesis of Theorem 1,

$$\mathbb{E}\Big[r\Big(\frac{d(D)}{\|V\|}\Big)\Big] < \infty,$$

where $d(D)$ denotes the diameter of D.

Coupling from the stochastic formulation

With the previous construction, we get two i.i.d. sequences $(R_i)_{i\geq 1}$ and $(\Theta_i)_{i\geq 1}$. At all time $t > 0$, the couple (X_t, V_t) has law f_t.

Coupling from the stochastic formulation

With the previous construction, we get two i.i.d. sequences $(R_i)_{i\geq 1}$ and $(\Theta_i)_{i\geq 1}$. At all time $t > 0$, the couple (X_t, V_t) has law f_t.

The same construction with f_∞ instead of f_0 gives a process $(\tilde{X}_t, \tilde{V}_t)_{t\geq 0}$ such that, at all $t > 0$, $(\tilde{X}_t, \tilde{V}_t) \sim f_\infty$. We also have two i.i.d. sequences $(\tilde{R}_i)_{i\geq 1}$, $(\tilde{\Theta}_i)_{i\geq 1}$.

Coupling from the stochastic formulation

With the previous construction, we get two i.i.d. sequences $(R_i)_{i\geq 1}$ and $(\Theta_i)_{i\geq 1}$. At all time $t > 0$, the couple (X_t, V_t) has law f_t.

The same construction with f_∞ instead of f_0 gives a process $(\tilde{X}_t, \tilde{V}_t)_{t\geq 0}$ such that, at all $t > 0$, $(\tilde{X}_t, \tilde{V}_t) \sim f_\infty$. We also have two i.i.d. sequences $(\tilde{R}_i)_{i\geq 1}$, $(\tilde{\Theta}_i)_{i\geq 1}$.

We want to correlate the sequences $(R_i)_{i\geq 1}$, $(\tilde{R}_i)_{i\geq 1}$, $(\Theta_i)_{i\geq 1}$ and $(\tilde{\Theta}_i)_{i\geq 1}$ to obtain that $\mathbb{E}[r(\tau)] < \infty$.

Coupling from the stochastic formulation

We assume D to be strictly convex. We claim that $\tau \leq T_N$ with $N \sim \mathcal{G}(c)$ for some $c > 0$. Indeed, at some time T_i, $i \geq 1$, we have two possible situations:

Coupling from the stochastic formulation

We assume D to be strictly convex. We claim that $\tau \leq T_N$ with $N \sim \mathcal{G}(c)$ for some $c > 0$. Indeed, at some time T_i, $i \geq 1$, we have two possible situations:

if $\|\tilde{V}_{T_i}\| < 1$, we choose the corresponding R, $\tilde{R}$, Θ, $\tilde{\Theta}$ independently until the data are independent of those at time T_i. Then we test again.

Coupling from the stochastic formulation

We assume D to be strictly convex. We claim that $\tau \leq T_N$ with $N \sim \mathcal{G}(c)$ for some $c > 0$. Indeed, at some time T_i, $i \geq 1$, we have two possible situations:

if $\|\tilde{V}_{T_i}\| < 1$, we choose the corresponding R, $\tilde{R}$, Θ, $\tilde{\Theta}$ independently until the data are independent of those at time T_i. Then we test again.

if $\|\tilde{V}_{T_i}\| > 1$ (which happens after a geometric number of iterations of the previous step), we can couple (R_i, Θ_i) with $(\tilde{R}_i, \tilde{\Theta}_i)$ such that, for some $c_1 > 0$,

$$\mathbb{P}\Big((X_{T_{i+1}}, V_{T_{i+1}}) = (\tilde{X}_{T_{i+1}}, \tilde{V}_{T_{i+1}})\Big) \geq c_1.$$

We do so using the maximal coupling from Theorem 1: if $Z \sim \mu, \tilde{Z} \sim \nu$, we can correlate Z and $\tilde{Z}$ such that

$$\mathbb{P}(Z = \tilde{Z}) = \int \mu \wedge \nu.$$

Coupling from the stochastic formulation II

The second step is “successful” after a geometric number of iterations, and we come back to the second step after a geometric number of iterations of the first step (remember the example above !).

A geometric sum of geometric random variables is geometric. The claim follows.

Conclusion

Once the claim is established we use the previous remark. Upon modifying slightly the rate r, we might use Hölder's inequality: for some $\epsilon > 0$,

$$\begin{aligned}
\mathbb{E}[r(T_N)] &\lesssim \sum_{i=1}^{\infty}\sum_{j=1}^{i-1}\mathbb{E}[r(T_{j+1}-T_j)\mathbf{1}_{\{N=i\}}] \\
&\lesssim \sum_{i=1}^{\infty}\sum_{j=1}^{i-1}\mathbb{E}\left[r\left(\frac{d(D)}{\|V_{T_j}\|}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}\mathbb{P}(N=i)^{\frac{\epsilon}{1+\epsilon}} \\
&\lesssim \sum_{i=1}^{\infty} i(1-c)^{\frac{i\epsilon}{1+\epsilon}} < \infty.
\end{aligned}$$

For the physically relevant case $M(v) = \frac{e^{-\|v\|^2}}{(2\pi)^{n/2}}$, we find the optimal rate $\frac{1}{t^{n-}}$.

Extensions

To treat the case where D is not strictly convex, we use a result from Evans (2001): for every C^1 domain there exists $N \in \mathbb{N}^*$ so that two points at the boundary can be joined in at most N collisions with the boundary.

One possible extension is to add specular (deterministic) reflection at the boundary for a fraction of the particles. This changes the construction of the stochastic process, but the proof strategy remains the same.

Thank you

Thank you for your attention !