Rate of convergence, polynomial ergodicity and Computable bounds for Markov Chains

Eric Moulines, Ecole Nationale Supérieure des Télécommunications, Paris

Joint work with Randal Douc, Gersende Fort, G. Roberts, J. Rosenthal.

Computable bounds

- Let $\{X_n\}$ be a Markov chain on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ with transition kernel P and initial distribution ξ.
- Assume that $\{X_n\}$ admits a *single* invariant distribution.
- Let ρ be a distance between probability measure:

A *computable bound* $C_\rho(n, \xi)$ is a function of the iteration index n and of the initial distribution ξ, verifying

$$\rho(\xi P^n, \pi) \leq C_\rho(n, \xi), \tag{1}$$

that can be computed *easily* from quantities associated to the Markov kernel.

Applications

- Convergence assessment in MCMC : how many iterations are required to get a certain approximation of the target density ?..
- MCMC design: provides understanding on the impact of the choice of certain components on the rate of convergence
  1. Choice of the proposal,
  2. Choice of the methodology [Langevin vs MH vs RSMH...]

Outline of the talk

- General state space Markov chain (a bird's view !)
- Ergodicity [and stronger forms...]
- Foster-Lyapunov drift conditions and geometrical convergence,
- Polynomial convergence,
- Coupling constructions,
- Some applications

ψ-irreducibility

- Let ϕ be a probability on $\mathcal{B}(\mathcal{X})$. The chain P is said to be ϕ-irreducible if

$$\phi(A) > 0 \implies \forall x \in \mathcal{X},\ \exists n > 0,\ P^n(x, A) > 0,$$
$$\forall x \in \mathcal{X},\ \ \mathbb{P}_x(\tau_A < \infty) > 0 \qquad \text{where} \ \ \tau_A := \inf\{n \geq 1, X_n \in A\}.$$

- The probability measure ψ is said to be a **maximal irreducibility** measure if for all irreducibility measure ϕ, we have $\phi \prec \psi$.
- Let ϕ be an irreducibility measure. Then,

$$\psi(A) := \int \phi(dy) \sum_{n \geq 0} 2^{-(n+1)} P^n(y, A)$$

is a **maximal irreducibility measure**. All the maximal irreducibility measure are equivalent.

Aperiodicity

- Let P be a ψ-irreducible Markov kernel. Let $D_1, \cdots, D_d$ be a partition of $\mathcal{X}$. We say that $D_1, D_2, \ldots, D_d$ is a d-cycle if

  $$P\left(x, D_{(k+1) \bmod [d]}\right) = 1, \; \forall x \in D_k.$$

  $\mathcal{X}$ is a 1-cycle.

- **key result** Let ψ be a maximal irreducibible.

  Theorem 1 (hard). *There is an essentially unique largest d-cycle,*

  $$\mathcal{X} = D_1 \cup D_2 \cup \cdots \cup D_d \cup E$$

  where $\mathbb{P}(x, D_k) = 1 \; \forall x \in D_{(k-1)[d]}$ *and* $\psi(E) = 0$.

  The chain is **aperiodic** if there is no 2-cycle.

Small and petite sets

Let $m \geq 1$ be an integer, $\epsilon > 0$, and let ν_m be a probability measure on $\mathcal{B}(\mathcal{X})$. The set C is **small** if

$$\forall x \in C,\ P^m(x, \cdot) \geq \epsilon \nu_m(\cdot)$$

Theorem 2 (hard). *If P is ψ-irreducible, every accessible set contains an accessible ν_m-small set C, where ν_m can be chosen to be a maximal irreducibility measure.*

A set C is said to be ν_a-**petite** if there exists a probability measure $a = \{a(n)\}$ on $\mathbb{N}$, a constant $\epsilon > 0$ and a probability measure ν_a, such that

$$\forall x \in C,\ \forall A \in \mathcal{B}(\mathcal{X}),\ \sum_{n \geq 0} a(n) P^n(x, A) \geq \epsilon \nu_a(A).$$

Every small set is petite; conversely, if P is ψ-irreducible and aperiodic, every petite set is small.

Examples

- single points are small,
- for an irreducible aperiodic chain on a countable space, all finite sets are small,
- For a ϕ-irreducible aperiodic chain satisfying certain topological conditions (T-chains), all the compact sets are small,
- Finding small sets (and proving they are small) can be hard work in MCMC context !

Atom

The set α is said to be an **atom** if there exists a **probability measure** ν_α, such that

$$\forall A \in \mathcal{B}(\mathcal{X}),\ \forall x \in \alpha,\ \ P(x, A) = \nu_\alpha(A)$$

We denote (with a slight abuse of notation), $\nu_\alpha(A) =: P(\alpha, A)$. Of course, every atom is a small set. Conversely, if there is a small set, it is possible to construct a chain on an **extended probability** space which behaves marginally as the original chain and which has an atom. In some sense, it is always possible to reduce the study of ψ-irreducible chain with petite or small sets to the study of chain with atoms.

Return time. Ergodicity

The **return time** to the set C is defined as

$$\tau_C := \inf\{n \geq 1, X_n \in C\}, \qquad C \in \mathcal{B}(\mathcal{X})$$

Theorem 3 (*ergodicity*). *Let P be ψ-irreducible and aperiodic and let C be a small set such that $\sup_{x \in C} \mathbb{E}_x[\tau_C] < \infty$.*

1. *There exists an **unique** invariant probability measure π, $\pi P = \pi$,*
2. *π is a **maximal irreducibility measure***
3. *π a.s., $\quad \lim_n \|P^n(x, \cdot) - \pi(\cdot)\|_{TV} = 0$.*

Total Variation

Let μ be a signed measure on $\mathcal{B}(\mathcal{X})$.

$$\|\mu\|_{\mathrm{TV}} := \sup_{A \in \mathcal{B}(\mathcal{X})} |\mu(A)| = \inf_{|f| \leq 1} |\mu(f)|.$$

If μ and ν are two probability measures with densities f and g with respect to a common dominating measure ξ

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2} \int |f(x) - g(x)| \xi(dx).$$

If $\{X_n\}$ is a Markov chain with initial distribution ξ and transition P, for all $A \in \mathcal{B}(\mathcal{X})$

$$|\xi P^n(A) - \pi(A)| \leq \|\xi P^n - \pi\|_{\mathrm{TV}}$$

Return time. f-ergodicity

Let f be a positive function. The f-norm of the signed measure μ is defined as

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

Theorem 4. *Let P be ψ-irreducible and aperiodic and let C be a small set such that $\sup_{x \in C} \mathbb{E}_x[\sum_{k=0}^{\tau_C - 1} f(X_k)] < \infty$. Then,*

1. *there exists an unique invariant probability measure π and $\pi(f) < \infty$,*
2. *π a.s.* $\quad \lim_n \|P^n(x, \cdot) - \pi(\cdot)\|_f = 0.$

(f, r)-ergodicity

Let $r = \{r(n)\}$ be a non-decreasing sequence and let P be a ψ-irreducible and aperiodic chain with invariant probability measure π. We say that the chain is (f, r)-**ergodic** if

$$\pi \text{a.e.}, \quad \limsup r(n) \| P^n(x, .) - \pi(.) \|_f < \infty$$

1. **geometrical rate** $r(n) = \kappa^n$, $\kappa > 1$.
2. **sub-geometrical rate**: $\{r(n)\}$ non-decreasing, $r(0) = 1$, $\lim \downarrow \log r(n)/n = 0$.
   - sub-exponential $r(n) = \exp(\beta n^\gamma)$, $\beta > 0$, $0 < \gamma < 1$,
   - polynomial $r(n) = (n+1)^q$, $q \geq 0$

Sufficient conditions for geometric ergodicity

After Meyn and Tweedie (1993)

Let P be ψ-irreducible and aperiodic and let $f : \mathcal{X} \to [1, \infty[$. Assume there exist $\kappa > 1$ and a small set C such that

$$\sup_{x \in C} \mathbb{E}_x \Big[\sum_{k=0}^{\tau_C - 1} \kappa^k f(X_k) \Big] < \infty$$

Then,

- There exists an unique invariance probability measure π,
- There exist $\eta > 1$ and a constant $R < \infty$, such that π a.e.,

$$\sum_n \eta^n \| P^n(x, \cdot) - \pi(\cdot) \|_f \leq R f(x)$$

Sufficient conditions for sub-geometric ergodicity

After Tuominen and Tweedie (1994)

Let P be ψ-irreducible and aperiodic and let $f : \mathcal{X} \to [1, \infty[$. Assume there exists a non-decreasing sub-geometric sequence $r = \{r(n)\}$ and a small set C such that

$$\sup_{x \in C} \mathbb{E}_x \Big[\sum_{k=0}^{\tau_C - 1} r(k) f(X_k) \Big] < \infty$$

Then,

- There exists an unique invariant probability measure π,
- There exists a constant $R < \infty$, such that π a.e.,

$$r(n) \| P^n(x, \cdot) - \pi(\cdot) \|_f \leq R f(x)$$

Dynkin's inequality

Assume that there exists an adapted process $Z = \{Z_n\}$ verifying

$$\mathbb{E}[Z_{n+1}|\mathcal{F}_n] \leq Z_n - f_n(X_n) + s_n(X_n)$$

where f_n and s_n are positive (Borel) functions. Then for all $x \in \mathcal{X}$ and for all stopping time τ,

$$\mathbb{E}_x\Big[\sum_{k=0}^{\tau-1} f_k(X_k)\Big] \leq \mathbb{E}_x[Z_0] + \mathbb{E}_x\Big[\sum_{k=0}^{\tau-1} s_k(X_k)\Big]$$

Dynkin's inequality and drift conditions

Assume there exist some functions $V_n, f_n \geq 1$ and a set C such that

$$PV_{n+1}(x) \leq V_n(x) - r(n)f(x) + br(n)\mathbb{I}_C(x).$$

Then, by applying the Dynkin's formula, it holds

$$\mathbb{E}_x\left(\sum_{k=0}^{\tau_C-1} r(k)f(X_k)\right) \leq V_0(x) + br(0)\mathbb{I}_C(x)$$

and thus the f-modulated return time to C is bounded. Now, if **(i)** P is ψ-irreducible and aperiodic, **(ii)** C is small and $\sup_C V_0 < \infty$, then, π-a.s., we have

$$\lim_{n\to\infty} r(n)\|P^n(x,.) - \pi(.)\|_f = 0.$$

Dynkin's inequality and drift conditions

Assume that P is ψ-irreducible and that C is small. Then,

- For any accessible set B, there exists a finite constant $c(B)$ such that

$$\mathbb{E}_x \left[\sum_{k=0}^{\tau_B - 1} f(X_k) \right] \le V_0(x) + c(B)$$

- The set $\{V < \infty\}$ is **full and absorbing**,
- The set

$$\left\{ x \in \mathcal{X}, \forall B \in \mathcal{B}^+(\mathcal{X}),\ \mathbb{E}_x \left[\sum_{k=0}^{\tau_B - 1} f(X_k) \right] < \infty \right\}$$

is full and absorbing

Foster-Lyapunov drift conditions

Assume there exists a function $V \geq 1$, $\lambda \in (0,1)$, $b < \infty$ and a small set C such that

$$PV(x) \leq \lambda V(x) + b\mathbb{I}_C(x)$$

Then,

$$\sup_{x \in C} \mathbb{E}_x \left[\lambda^{-\tau_C} \right] < \infty$$

and thus, if P is ψ-irreducible and aperiodic and C is a small set, then P is V-geometrically ergodic, and there exists $\eta > 1$

$$\limsup \eta^n \|P^n(x,.) - \pi\|_V < \infty$$

Non-linear autoregressive model

Let $\{W_n\}$ be a sequence of i.i.d random vectors and let $F : \mathbb{R}^d \to \mathbb{R}^d$ be a Borel function. Consider the following recurrence equation

$$X_{n+1} = F(X_n) + W_{n+1}$$

- $W_n \sim \Gamma(dy) := \gamma(y)dy, \qquad |y| \leq \delta \Rightarrow \gamma(y) \geq \epsilon > 0, \qquad \mathbb{E}[|W_n|^s] < \infty, \quad s \geq 1.$
- F is continuous and is uniformly Lipshitz at infinity, *i.e.*there exists $0 < \rho < 1$, such that $|F(x)| \leq \rho\ |x|, \quad |x| \geq M$.

Then,

Theorem 5. • *P is ψ-irreducible and aperiodic and every compact set is small,*

- *there exists a unique invariant probability measure π,*
- $\exists r > 1 \quad \lim_n r^n\ \|P^n(x, \cdot) - \pi(\cdot)\|_{1+|x|^s} = 0.$

Metropolis-Hastings algorithm

Let $\pi(dx) = \pi(x)dx$ be a probability measure, with density $\pi(x)$ w.r.t the Lebesgue measure. The kernel of the random walk Metropolis-Hastings algorithm is defined as

$$P(x, dy) = \left(1 \wedge \frac{\pi(y)k(y;x)}{\pi(x)k(x;y)}\right) \; k(x;y) \; dy + \delta_x(dy) \int \left(1 - 1 \wedge \frac{\pi(y)k(y;x)}{\pi(x)k(x;y)}\right) k(x;y)dy.$$

Moves are proposed under a proposal kernel $K(x, dy) = k(x, y)dy$. Move are accepted with probability

$$\alpha(x, y) = \left(1 \wedge \frac{\pi(y)k(y;x)}{\pi(x)k(x;y)}\right)$$

and rejected otherwise, in which case the chain stay in the current position.

P satisfies the **detailed** balanced condition

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

and thus π is the unique invariant probability measure for P.

Metropolis-Hastings algorithm on $\mathbb{R}^d$

Assume that

- $k(x; y) = k(|y - x|)$ **and** $|z| \leq \delta \Rightarrow k(|z|) \geq \epsilon > 0$
- $\pi > 0$ and is sub-exponential

$$\lim_{|x| \to \infty} \left\langle \frac{x}{|x|} ; \nabla \log \pi(x) \right\rangle = -\infty.$$

- The level sets are regular

$$\limsup_{|x| \to \infty} \left\langle \frac{x}{|x|} ; \frac{\nabla \pi(x)}{|\nabla \pi(x)|} \right\rangle < 0$$

Then,

Theorem 6. *(Jarner, Hansen 1998)*

- *The Foster-Lyapunov drift condition is verified with* $V(x) = \pi(x)^{-1/2}$,
- $\exists r_* > 1 \quad \lim_n r_*^n \, \|P^n(x, \cdot) - \pi(\cdot)\|_V = 0, \quad x \in \mathbb{R}^d.$

Open questions

- What happens in the non-linear autoregressive setting when F is not Lipshitz at infinity ? In particular, we will consider the case where

  $$|F(x)| \leq |x|(1 - r|x|^{-d})$$

  and we will show that the NLAR is in such case ergodic at a polynomial rate when $0 < d \leq 2$.
- What happens for the RWMH algorithm when the target distribution is not sub-exponential in the tail ? In particular, we will consider the case where it is **regularly varying** in the tails or it is super-exponential in the tails.

Tuominen-Tweedie drift conditions

Assume there exist a sub-geometrical sequence $r = \{r(n)\}$ $(r(0) = 1)$, a set C, a function $f \geq 1$ and a sequence of functions $V_n \geq r(n) f$ verifying, for all n,

$$PV_{n+1}(x) \leq V_n(x) - r(n) f(x) + b\, r(n) \mathbb{I}_C(x)$$

Then, for all x we have

$$\mathbb{E}_x \left[\sum_{k=0}^{\tau_C - 1} r(k) f(X_k) \right] \leq V_0(x) + b \mathbb{I}_C(x)$$

If P is ψ-irreducible and aperiodic and if C is small, then, P has a unique invariant probability measure π and

$$\limsup_n r(n) \| P^n(x, .) - \pi(.) \|_f = 0.$$

Nevertheless, the Tuominen and Tweedie conditions, contrary to the Forster-Lyapunov condition is not extremely easy to handle !

Polynomial ergodicity

The objective is to find a **simple** condition upon guaranteeing that for some $f \geq 1$ and $l \geq 0$, we have

$$\sup_{x \in C} \mathbb{E}_x \Big[\sum_{k=0}^{\tau_C - 1} k^l f(X_k) \Big] < \infty$$

We know that this condition implies that, when P is ψ-irreducible and aperiodic, and C is a small set,

1. there exists a unique invariant probability π,
2. π a.s.,

$$\lim_n n^l \| P^n(x, \cdot) - \pi(\cdot) \|_f = 0$$

Nested drift conditions

Assume that there exist functions $1 \leq V_0 \leq V_1 \leq V_2$, constants b_0, b_1 and a set C such that

$$PV_1 \leq V_1 - V_0 + b_0 \, \mathbb{I}_C,$$
$$PV_2 \leq V_2 - V_1 + b_1 \, \mathbb{I}_C.$$

Then,

$$\mathbb{E}_x \Big[\sum_{k=0}^{\tau_C - 1} V_0(X_k) \Big] \leq V_1(x) + b_0 \mathbb{I}_C(x),$$
$$\mathbb{E}_x \Big[\sum_{k=0}^{\tau_C - 1} V_1(X_k) \Big] \leq V_2(x) + b_1 \mathbb{I}_C(x),$$
$$\mathbb{E}_x \Big[\sum_{k=0}^{\tau_C - 1} (k+1) V_0(X_k) \Big] \leq V_2(x) + (b_0 + b_1) \, \mathbb{I}_C(x).$$

Nested drift conditions

1. Assume that there exist $f := V_0 \leq \cdots \leq V_q$, and constants $b_k < \infty$, $k \in \{0, \cdots, q-1\}$ such that

$$PV_{k+1}(x) \leq V_{k+1}(x) - V_k(x) + b_k \mathbb{I}_C(x) \qquad k \in \{0, \cdots, q-1\}.$$

2. $\sup_C V_q < \infty$

Theorem 7.

$$\mathbb{E}_x \Big[\sum_{k=0}^{\tau_C - 1} \mathbb{I}^{*(q-1)}(k+1) f(X_k) \Big] \leq V_q(x) + \Big(\sum_{k=0}^{q-1} b_k \Big) \mathbb{I}_C(x),$$

where $\mathbb{I}^{*(q-1)}(k) \sim k^{q-1}/(q-1)!$ *are defined as*

$$\mathbb{I}^{*0}(k) := 1, \text{ and } \mathbb{I}^{*(j+1)}(k) := \sum_{l=1}^{k} \mathbb{I}^{*j}(l).$$

If P ψ*-irreducible and aperiodic and* C *is a small set, then, there exists a unique invariant probability* π *and* π*-a.s.,*

$$\lim_n (n+1)^{q-1} \| P^n(x, \cdot) - \pi(\cdot) \|_f = 0.$$

Note that we can trade speed against increased control.

Nested drift conditions

Assume that P is ψ-irreducible and that C is small. Then,

- For any accessible set B, there exists a finite constant $c(B)$ such that

$$\mathbb{E}_x \left(\sum_{k=0}^{\tau_B - 1} (k+1)^{q-1} f(X_k) \right) \le c(B) V_q(x)$$

- The set $\{V_q < \infty\}$ is **full and absorbing**,
- The set

$$\left\{ x \in \mathcal{X}, \forall B \in \mathcal{B}^+(\mathcal{X}),\ \mathbb{E}_x \left(\sum_{k=0}^{\tau_B - 1} (k+1)^{q-1} f(X_k) \right) < \infty \right\}$$

is full and absorbing

Non Linear Autoregressive model

$$X_{n+1} = F(X_n) + W_{n+1}$$

- P Lebesgue-irreducible, aperiodic and every compact set is small,
- there exist $0 < d \leq 2$, $r > 0$ and $M < \infty$ such that,

$$|F(x)| \leq |x| \left(1 - r|x|^{-d}\right), \qquad |x| \geq M \qquad \sup_{|x| \leq M} |F(x)| < \infty$$

- $\{W_n\}$ i.i.d. with distribution $\Gamma(dx) = \gamma(x)dx$, such that

$$\int y\Gamma(dy) = 0, \text{and} \int |y|^{s_*}\Gamma(dy) < \infty,$$

  for some $s_* \geq 4$.

Then, for all $d \leq s \leq s_*$, there exists $\lambda > 0$ such that

$$\int P(x, dy)|y|^s \leq |x|^s - \lambda|x|^{s-d}(1+\varepsilon(x)), \qquad \lim_{|x| \to \infty} \varepsilon(x) = 0.$$

NLAR model

Let $q = \lfloor (s_* - d)/d \rfloor$ and define, for $\eta \in [0, d)$,

$$V_0(x) := 1 + |x|^\eta,$$
$$V_1(x) := 1 + |x|^{\eta+d},$$
$$\vdots$$
$$V_q(x) := 1 + |x|^{\eta+qd}.$$

Then, we have for all $x \in \mathbb{R}^d$

$$\limsup_n n^{q-1} \|P^n(x, .) - \pi(.)\|_{V_0} = 0.$$

Random Walk Hasting-Metropolis

- The target distribution is continuous on $\mathbb{R}$, $\lim \downarrow p(x) = 0$ and for all $|x| \geq M$, for all $y \in \{p(x+y) \leq p(x)\}$

$$\left|\frac{p(x+y)}{p(x)} - 1 - syx^{-1}\right| \leq C|x|^{-1}\rho(x)y^2 \qquad s > 1 \qquad \lim_{x\to\infty} \rho(x) = 0$$

- The proposal is symmetric and verifies $|x| \leq \delta \Rightarrow k(x) \geq \epsilon$, and $\int |x|^{\zeta+3} k(x) dx < \infty, \zeta \geq 1$.

Set $s_* := \zeta \wedge s$. For all $2 \leq \beta < s_* + 1$, there exists $\lambda > 0$ such that

$$\int P(x, dy)|y|^\beta \leq |x|^\beta - \lambda |x|^{\beta-2}(1+\varepsilon(x)), \qquad \lim_{|x|\to\infty} \varepsilon(x) = 0.$$

Then, for all $0 \leq r < s_* - 1$, and all $0 \leq \gamma < (s_* - 1 - r)/2$,

$$\forall\, x \in \mathbb{R} \qquad \lim_n (n+1)^\gamma \|P^n(x,\cdot) - \pi(\cdot)\|_{1+|x|^r} = 0.$$

Random Walk Hastings-Metropolis

- target distribution $p(x) = \eta x^{\eta-1}\exp(-x^\eta)$, for $x > 0$ and $0 < \eta < 1$.

- proposal distribution k: k symmetric, $|x| \leq \delta \Rightarrow k(x) \geq \epsilon$, and

$$\int |x|^{\beta_* + 2\eta + 2} \exp(\alpha |x|^\eta) k(x) dx < \infty, \ \ \beta_* \in \mathbb{R}, 0 < \alpha < 1$$

Drift conditions There exists $\lambda > 0$, such that, for all $\beta \leq \beta_*$,

$$\int P(x, dy) y^\beta \exp(\alpha y^\eta) \leq x^\beta \exp(\alpha x^\eta) - \lambda x^{2(\eta-1)} x^\beta \exp(\alpha x^\eta)(1 + \varepsilon(x)),$$

where $\lim_{x\to\infty} \varepsilon(x) = 0$.

For all $\beta \leq \beta_*$, $q \in \mathbb{N}^*$, such that $\beta + q(1-\eta) \leq \beta_*$, $x > 0$

$$\lim_n (n+1)^q \left\| P^n(x, \cdot) - \pi(\cdot) \right\|_{(1\vee x)^\beta \exp(\alpha x^\eta)} = 0.$$

Reduction to a single drift condition

The set of drift conditions can be written sometimes in a more "compact" way.

`There exist a function` $V \geq 1$, `constants` $0 < c, b < \infty$, `a set` C `and` $\delta < 1$ `such that` $PV \leq V - cV^{1-\delta} + b1_C$

`There exists a function` $V \geq 1$, `a` q`-times differentiable function` $\phi : [1, \infty) \to [0, \infty)$, `a set` C `and` $b < \infty$ `such that` $PV \leq V - \phi \circ V + b1_C$.

All the examples known to date verify a condition of that type...

A convergence result

Theorem 8. *Suppose P is ψ-irreducible and aperiodic. Assume there exist a function $V \geq 1$, constants $0 < c, b < \infty$, a small set C and $0 < \delta < 1$ such that*

$$PV \leq V - cV^{(1-\delta)} + b1_C$$

Then P is (V_β, r_β)-regular for each $1 \leq \beta \leq 1/\delta$ where

$$V_\beta := V^{1-\beta\delta}, \qquad r_\beta(n) := (n+1)^{\beta-1}$$

In particular, the following polynomial convergence statements hold for all x

$$(n+1)^{(\beta-1)} \|P^n(x,.) - \pi(.)\|_{\mathrm{V}^{1-\beta\delta}} \to 0, \qquad n \to \infty.$$

Computable Bounds

The objective is here to find a function $B_f : \mathcal{X} \times \mathbb{N} \to \mathbb{R}^+$ verifying, for all $x \in \mathcal{X}$ and all $n \in \mathbb{N}$,

$$\|P^n(x, \cdot) - \pi(\cdot)\|_f \leq B_f(x, n)$$

The bound is said to be **computable** (admittedly a loosely defined expression) if B_f can be determined from quantities that can be explicitly determined from the kernel P, such as, the minoration constant ϵ on a small set, constants and functions appearing in drift conditions, etc..

Assumptions: geometric ergodicity

After Roberts and Tweedie (1998)

- There exists a ν_1-small set C:

$$P(x,\cdot) \geq \epsilon \nu_1(\cdot) \qquad x \in C.$$

- The Foster-Lyapunov drift condition is verified

$$PV \leq \lambda V + b \mathbb{I}_C, \qquad 0 < \lambda < 1.$$

- An (admittedly technical) condition: $\sup_C V \geq \frac{b}{2(1-\lambda)} - 1.$

Coupling construction

Coupling amounts to construct a Markov chain (Z_n, Z'_n, d_n) on an extended probability space $\mathcal{X} \times \mathcal{X} \times \{0,1\}$ such that, for any initial probability measure λ and μ, and $\forall A \in \mathcal{B}(\mathcal{X})$,

$$\mathbb{P}_{\lambda,\mu,0}(Z_n \in A) = \int \lambda(dx) P^n(x, A) \quad \text{and}$$

$$\mathbb{P}_{\lambda,\mu,0}(Z'_n \in A) = \int \mu(dx) P^n(x, A)$$

and such that, there exists a $P_{\lambda,\mu,0}$-a.s. finite stopping time T such that, for all $n \geq T$, $Z_n = Z'_n$. In the coupling terminology, T is a (strong) **coupling time**.

The random variable d_n is the *bell variable*, $d_n = 0$ if $n < T$ (before the coupling time), and $d_n = 1$ if $n \geq T$ (after the coupling time).

Lindvall's inequality

The Lindvall's inequality implies that

$$|\lambda P^n(A) - \mu P^n(A)| \leq \mathbb{P}_{\lambda,\mu,0}(T \geq n)$$

Assume that P has a stationary distribution π. Applying the relation above to $\lambda = \delta_x$ and $\mu = \pi$, we have

$$|P^n(x, A) - \pi(A)| \leq \mathbb{P}_{x,\pi,0}(T \geq n),$$

and thus

$$\|P^n(x, .) - \pi\|_{\mathrm{TV}} \leq \mathbb{P}_{x,\pi,0}(T \geq n).$$

Coupling Markov chain on discrete state space

Idea couple the Markov Chain the first time they meet [are in the same state at the same time]

- $d_n = 0$ **Before the coupling time**
  1. $Z_n \neq Z'_n$ then draw $Z_{n+1} \sim P(Z_n, .)$ and $Z'_{n+1} \sim P(Z'_n, .)$ independently, and set $d_{n+1} = 0$,
  2. $Z_n = Z'_n$, then draw $Z_{n+1} = Z'_{n+1} \sim P(Z_n, .)$ and set $d_{n+1} = 1$.
- $d_n = 1$ **After the coupling time**: draw $Z_{n+1} = Z'_{n+1} \sim P(Z_n, .)$ and set $d_{n+1} = 1$.

When the state space is finite and the chain is irreducible, then the coupling time will also be finite.

Coupling Markov chain on discrete state space

Assume that there exists a set $C \in \mathcal{B}(\mathcal{X})$ such that a **uniform minorization** condition holds: $\forall (x,y) \in C$, $P(x,y) \geq \epsilon > 0$. Set

$$\nu(A) := \frac{\sum_{x \in C} \delta_x(A)}{\#\{C\}} = \frac{\#\{C \cap A\}}{\#\{C\}}.$$

We have

$$P(x,A) \geq P(x, A \cap C) = \sum_{y \in A \cap C} P(x,y) \geq \epsilon \nu(A)$$

The set C is said to be a **small set**. Instead of coupling when the chains meet at a given point, we can couple when the chains meet in C.

Coupling within a set

- $d_n = 0$ **Before the coupling time**
  1. if $(Z_n, Z'_n) \notin C \times C$, then draw $Z_{n+1} \sim P(Z_n, .)$ and $Z'_{n+1} \sim P(Z'_n, .)$ independently and set $d_{n+1} = 0$.
  2. if $(Z_n, Z'_n) \in C \times C$, then draw an ϵ biased coin.
     - If the coin comes up head, then draw $Z_{n+1} = Z'_{n+1} \sim \nu(.)$ and set $d_{n+1} = 1$ (successful coupling)
     - Otherwise, draw $Z_{n+1} \sim R(Z_n, .)$ and $Z'_{n+1} \sim R(Z'_n, .)$ and set $d_{n+1} = 0$, where R is the residual kernel

$$R(x, y) := (1-\epsilon)^{-1}(P(x, y) - \epsilon\nu(y)).$$

- $d_n = 1$ **After the coupling time**: draw $Z_{n+1} = Z'_{n+1} \sim P(Z_n, .)$

Uniform ergodicity

- Assume that, $P(x, y) \geq \epsilon \nu(y) > 0$ for all $(x, y) \in \mathcal{X} \times \mathcal{X}$, for some probability measure ν (Doeblin condition).
- Set $C = \mathcal{X}$. We can try to couple the chain at every time instants. Hence,

$$\mathbb{P}_{\lambda,\pi,0}(T \geq n) = (1-\epsilon)^n$$

- Lindvall's inequality then implies that

$$\|\lambda P^n - \pi\|_{\mathrm{TV}} \leq (1-\epsilon)^n$$

Bounding the coupling time

We may decompose the coupling time on the return times of the chain to the set $C \times C$.

$$
\begin{aligned}
&T = \sum_{k=0}^{\infty} \tau_k I(d_{\tau_k} = 0) I(d_{\tau_k+1} = 1), \\
&\tau_k := \inf \{n \geq \tau_{k-1} + 1, (Z_n, Z'_n) \in C \times C\}, \\
&\tau_0 := \inf \{n \geq 0, (Z_n, Z'_n) \in C \times C\}.
\end{aligned}
$$

Hence

$$
\begin{aligned}
\mathbb{P}_{\lambda,\pi,0} \{T = n\} &= \sum_{k=0}^{n} \mathbb{P}_{\lambda,\pi,0} \{\tau_k = n, d_{\tau_k} = 0, d_{\tau_k+1} = 1\} \\
&= \epsilon \sum_{k=0}^{n} \mathbb{P}_{\lambda,\pi,0} \{\tau_k = n, d_{\tau_k} = 0\}.
\end{aligned}
$$

The key is thus to get bounds for the return times to $C \times C$ (which is -almost- a renewal process). Bounds for return times to sets are generally obtain using **drift conditions**.

Coupling construction

Let D be a small set ($P(x, dy) \geq \epsilon\nu_1(dy)\mathbb{I}_D(x)$) and $C \subset D$.
Denote $\Delta := C \times D \cup D \times C$ and $R(x, dy)$ the residual kernel

$$R(x, dy) := (1-\epsilon)^{-1} \left(P(x, dy) - \epsilon\nu_1(dy)\right) \mathbb{I}_D(x).$$

Set $(Z_0, Z_0', d_0) = (z, z', 0)$ and define recursively, (Z_n, Z_n', d_n) as follows

- $d_n = 0$:
  - If $(Z_n, Z_n') \notin \Delta$, $\quad Z_{n+1} \sim P(Z_n, \cdot)$, $\quad Z_{n+1}' \sim P(Z_n', \cdot)$, $\quad d_{n+1} = 0$.
  - If $(Z_n, Z_n') \in \Delta$,
    - with probability ϵ, $\quad Z_{n+1} \sim \nu_1$, $\quad Z_{n+1}' = Z_{n+1}$, $\quad d_{n+1} = 1$.
    - with probability $1 - \epsilon$, $\quad Z_{n+1} \sim R(Z_n, \cdot)$, $\quad Z_{n+1}' \sim R(Z_n', \cdot)$, $d_{n+1} = 0$.
- $d_n = 1$, draw $Z_{n+1} \sim P(Z_n, \cdot)$, and set $Z_{n+1}' = Z_{n+1}$ and $d_{n+1} = 1$.

Coupling time / Hitting time

Define σ_Δ the hitting time on Δ,

$$\sigma_\Delta := \inf\{n \geq 0, (Z_n, Z'_n) \in \Delta\}.$$

Theorem 9. $\forall \;\; l \in \mathbb{N}^*$,

$$\mathbb{E}_{z,z',0}\left[\mathbb{I}^{*l}(T-1)\right] \leq \epsilon \quad \sum_{j=0}^{l} B_{l,j} \; \mathbb{E}_{z,z',0}\left[\mathbb{I}^{*j}(\sigma_\Delta)\right]$$

where $B := (B_{i,j})_{1 \leq i,j \leq l}$ *is a lower triangular matrix defined as* $B := (I - A)^{-1}$, *where* $A := (A_{i,j})_{1 \leq i,j \leq l}$ *is a lower triangular Toëplitz matrix* $A_{i,j} := A(i-j)$ *and*

$$A(j) := (1-\epsilon) \sup_{(x,x') \in \Delta} \int \int R(x, dy) R(x', dy') \mathbb{E}_{y,y',0}\left[\mathbb{I}^{*j}(\sigma_\Delta)\right],$$

Note that the diagonal elements of matrix A are equal to $(1-\epsilon)$ and thus matrix $(I - A)$ is lower triangular with diagonal elements equal to $\epsilon > 0$ and is thus invertible.

Moments of the hitting time

It is convenient to control the moment of the hitting time using drift conditions on the original chain. In the polynomial case, return time can be bounded using the nested drift conditions, under a mild assumption on the growth of the drift function at infinity.

Drift There exist $1 \leq V_0 \leq \cdots \leq V_q < \infty$, $0 < a_k < 1$, $b_k < \infty$ such that $\sup_C V_q < \infty$

$$\begin{cases} PV_{k+1}(x) \leq V_{k+1}(x) - V_k(x) + b_k \mathbb{I}_C(x), \\ V_k(x) \geq b_k/(1-a_k) \qquad x \in D^c. \end{cases} \qquad k \in \{0, \cdots, q-1\}$$

Theorem 10. *For all* $l \in \{1, \cdots, q\}$,

$$\mathbb{E}_{z,z',0}\left[\mathbb{I}^{*l}(\sigma_\Delta)\right] \leq \left(\prod_{k=0}^{l-1} a_k\right)^{-1} \inf_{(x,x')\in\Delta^c} (V_0(x)+V_0(x')) \left(V_l(x)+V_l(x')\right).$$

Computable bound for polynomial ergodicity

$$\|P^n(x,\cdot) - \pi(\cdot)\|_{VT} \leq P_{x,\pi,0}(T > n)$$

$$P_{x,\pi,0}(T > n) \leq U_\epsilon(l,n)^{-1}\ \mathbb{E}_{x,\pi,0}\Big[\mathbb{I}^{*l}(T-1)\Big]$$

$$\mathbb{E}_{x,x',0}\Big[\mathbb{I}^{*l}(T-1)\Big] \leq \epsilon \quad \sum_{j=0}^{l} B_{l,j}\ \mathbb{E}_{x,x',0}\Big[\mathbb{I}^{*j}(\sigma_\Delta)\Big]$$

$$\mathbb{E}_{x,x',0}\Big[\mathbb{I}^{*l}(\sigma_\Delta)\Big] \leq \left(\prod_{k=0}^{l-1} a_k\right)^{-1} \inf_{(x,x')\in\Delta^c}(V_0(x)+V_0(x'))\ \Big(V_l(x)+V_l(x')\Big)$$

Theorem 11. *Assume that P is ψ-irreducible and aperiodic, and that the* ***nested drift*** *conditions are verified for some sets $C, D \in \mathcal{B}(\mathcal{X})$, $C \subset D$, D is a ν_1 small set. Then,*

1. *P has a unique invariant probability measure π,*
2. $\|P^n(x,\cdot) - \pi(\cdot)\|_{TV} \leq \min_{l\in\{0,\cdots,q-1\}} U_\epsilon(l,n)^{-1} W_l(x)$