北京師範大學 数学科学学院
Beijing Normal University • School of Mathematical Sciences

Dirichlet Principles of Hitting Times for Non-reversible Markov Chains

Lu-Jing Huang

based on a joint work with prof. Y.-H. Mao

July 16, 2018

Overview

1. Markov chain: reversibility vs non-reversibility

2. Main results

3. Applications

Markov chains

- Let V be a finite state space. Let $K = (K_{ij})$ be a probability transition matrix(PTM), reversible w.r.t a probability measure π:

$$K_{ij} \geq 0, \quad \sum_j K_{ij} = 1, \quad \pi_i K_{ij} = \pi_j K_{ji}.$$

- Let P be also PTM, with π its stationary distribution:

$$\sum_i \pi_i P_{ij} = \pi_j.$$

- In general, P is not reversible, but we can get K from P:

$$K_{ij} = \frac{1}{2}[P_{ij} + P^*_{ji}], \quad P^*_{ij} = \frac{\pi_j P_{ji}}{\pi_i}.$$

Toward a same target

- $\lim_{n\to\infty} K_{ij}^{(n)} = \lim_{n\to\infty} P_{ij}^{(n)} = \pi_j$.
- Which is better or faster?
- In MCMC, specially in classical Metropolis-Hastings algorithm, it proceeds by constructing a reversible Markov chain towards a given but implicit stationary distribution.
- Many authors recently found that non-reversible Markov chain is better in some respects.
  Hwang C.-R. et al (1993-2018): Non-reversible Markov chain, diffusion

Comparison criteria

- Asymptotic variance related to CLT
- Large deviation
- Spectral gap
- Mixing times

Asymptotic variance

- Asymptotic variance related to CLT: Let X_k is the Markov chain of P with stationary distribution π. Then for $\pi(f) = 0$,

$$\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} f(X_k) \Rightarrow N(0, \sigma^2(P, f))$$

with

$$\sigma^2(P, f) = \lim_{n \to \infty} \mathsf{Var}_\pi[\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} f(X_k)].$$

- The smaller $\sigma^2(P, f)$, the better!
Sun, Gomez and Schmidhuber(2010); Chen and Hwang(2013); Pai and Hwang(2013); Hwang, Normanda and Wu(2015)

Large deviation

- Let the occupation measure $L_n = \frac{1}{n}\sum_{k=1}^{n}\delta_{X_k}$. Then $L_n \in$ LDP with rate function

$$I_P(\mu) := -\inf_{\phi>0}\sum_{i\in V}\mu_i \log\frac{P\phi_i}{\phi_i}.$$

- Roughly, for large n,

$$\mathbb{P}(L_n \in \cdot) \simeq \exp\{-n \inf_{\mu\cdot} I_P(\mu)\}.$$

- Thus the bigger $I_P(\mu)$, the better!
  Bierkens(2016): Continuous time Markov chain
- Question is that $I_P(\mu) \geq I_K(\mu)$?

Spectral gap

- In $L^2(\pi)$:

$$\lambda(P) = \text{spectral radius of } \sigma(P) \setminus \{1\}.$$

- By total variance:

$$\rho(P) = \inf\{\epsilon : \sum_j |P_{ij}^{(n)} - \pi_j| \leq C\epsilon^n\}.$$

- In the reversible case, $\lambda(P) = \rho(P)$ and we have the Poincaré inequality

$$1 - \lambda(P) = \inf\{\langle f, (I - P)f\rangle_\pi : \pi(f) = 0, \pi(f^2) = 1\}.$$

- Also the smaller $\lambda(P)$ or $\rho(P)$, the better.
  Hwang, Hwang-Ma and Sheu(1993, 2005): diffusions

Hitting times

- Let

$$\tau_i = \inf\{n \geq 0 : X_n = i\}.$$

- For $i, j \in V$, define the commute time

$$T_{ij}(P) = \mathbb{E}_i \tau_j + \mathbb{E}_j \tau_i.$$

- The average hitting time

$$T_0(P) = \sum_i \sum_j \pi_i \pi_j \mathbb{E}_i \tau_j = \frac{1}{2} \sum_i \sum_j \pi_i \pi_j T_{ij}(P).$$

→ Strong ergodicity

- In general, if $T_0(P) < \infty$, then the chain is strongly ergodic: there exist $C < \infty$ and $\rho < 1$ such that

$$\sup_i \sum_j |P_{ij}^{(n)} - \pi_j| \leq C\rho^n.$$

  Moreover, we have $\rho \leq 1 - 1/T_0(P)$.
- To see that the smaller $T_0(P)$, the better.
- Question: Which one is better between non-reversible Makov chain and its reversibility?

Previous results

- Let $\tilde{f}$ be the solution of

$$\begin{cases} Pf(k) = f_k, & k \neq i,\ j; \\ f_i = 1,\ f_j = 0. \end{cases}$$

- The classical Thompson's principle(reversible case)

$$\begin{aligned} \mathsf{Cap}_{ij} &:= \pi_i \mathbb{P}_i(\tau_j < \tau_i^+) \\ &= \inf\{\langle f, (I-P)f \rangle_\pi :\ f_i = 1, f_j = 0\} \end{aligned}$$

- Gaudillière-Landim(2014) extended Thompson's principle to the non-reversible case: For every pair of points $i \neq j$,

$$\mathsf{Cap}_{ij} = \inf\{\langle f, (I-P)(I-K)^{-1}(I-P)^* f \rangle_\pi :\ f_i = 1, f_j = 0\}$$

attains at $(\tilde{f} + \tilde{f}^*)/2$.

A comparison result

- But

$$T_{ij}(P)(= \mathbb{E}_i\tau_j + \mathbb{E}_j\tau_i) = 1/\mathsf{Cap}_{ij}.$$

We have a result on comparison:

Theorem (H.-Mao, 2017)

Let K be the reversible part of P. Fix any pair of $i \neq j$ and let $T_{ij}(K)$, $T_{ij}(P)$ respectively be the commute time between i, j of chains K and P. Then

$$T_{ij}(P) \leq T_{ij}(K).$$

Consequently, the average hitting times ($T_0 = \frac{1}{2}\sum_{k,l} \pi_k\pi_l T_{kl}$) of chains satisfy

$$T_0(P) \leq T_0(K).$$

Question

- $T_0(P) = \sum_i \sum_j \pi_i \pi_j \mathbb{E}_i \tau_j = \sum_j \pi_j \mathbb{E}_\pi \tau_j$.
- So we decide to study the properties (variational formula) for the mean hitting time.

Notations

- Let P be a irreducible PTM on a denumerable state space V.
- It admits a unique stationary distribution π.
- Define

$$D_\lambda(f,g) = \langle f, (I - \mathrm{e}^{-\lambda} P)g\rangle_\pi, \ f,g \in L^2(\pi), \ \lambda \geq 0,$$

with the natural convention $D := D_0$.

- Note that P maybe non-reversible, so $D_\lambda(f,g) \neq D_\lambda(g,f)$.

Variational formula for the mean hitting time

- $\phi^* := (\mathbb{E}_i {\tau^*}_A : \ i \in V)$ is a solution of Poisson equation

$$\begin{cases} (I - P^*)x(i) = 1, & i \in A^c; \\ x_i = 0, & i \in A. \end{cases}$$

Theorem (H.-Mao, 2018)

For any non-trivial subset $A \subseteq V$,

$$1/\mathbb{E}_\pi \tau_A = D(\phi^*, \phi) = \inf_{f|_A=0,\pi(f)=1} \sup_{g|_A=0,\pi(g)=0} D(f-g, f+g).$$

- For a finite reversible P (Aldous-Fill's book [Chap 3, Prop 41]),

$$1/\mathbb{E}_\pi \tau_A = \inf_{f|_A=0,\pi(f)=1} D(f, f).$$

Variational formula for the Laplace transform of τ_A

- $\psi^* := (\mathbb{E}_i[\exp(-\lambda {\tau^*}_A)] : \ i \in V)$ is a solution of Poisson equation

$$\begin{cases} (I - \mathrm{e}^{-\lambda} P^*)x(i) = 0, & i \in A^c; \\ x_i = 1, & i \in A. \end{cases}$$

Theorem (H.-Mao, 2018)

For any non-trivial subset $A \subseteq V$ and $\lambda > 0$,

$$\frac{1 - \mathrm{e}^{-\lambda}}{1 - \mathbb{E}_\pi[\exp(-\lambda \tau_A)]} = \inf_{f|_A = 0, \pi(f) = 1} \sup_{g|_A = 0, \pi(g) = 0} D_\lambda(f - g, f + g).$$

- This is new even for reversible P:

$$\frac{1 - \mathrm{e}^{-\lambda}}{1 - \mathbb{E}_\pi[\exp(-\lambda \tau_A)]} = \inf_{f|_A = 0, \pi(f) = 1} D_\lambda(f, f).$$

Monotonicity law

- Peskun ordering: For two PTMs K and P, $K \preceq P$, if

$$K_{ij} \leq P_{ij}, \quad i \neq j.$$

Theorem

Assume that $K,\ P$ be irreducible PTMs with same stationary distribution π. If $K \preceq P$ and K is reversible, then for any A,

$$\mathbb{E}_\pi[exp(-\lambda\tau_A(K))] \leq \mathbb{E}_\pi[exp(-\lambda\tau_A(P))], \quad \lambda > 0.$$

In particular, $\mathbb{E}_\pi[\tau_A(K)] \geq \mathbb{E}_\pi[\tau_A(P)]$.

- Under peskun ordering, similar results for asymptotic variance, Peskun (1973), Tierney (1998)

Comparison theorem

- Let K be a reversible PTM with stationary distribution π.
- Γ is a vorticity matrix, i.e., $\Gamma 1 = 0$ and $\Gamma^T = -\Gamma$. Assume that

$$\Gamma_{ij} > \pi_i K_{ij}, \quad i \neq j.$$

- Define

$$P_\alpha = K + \alpha \text{diag}(\pi)^{-1}\Gamma, \quad -1 \leq \alpha \leq 1.$$

Theorem

For any subset A and $\lambda > 0$, denote by $R(\alpha)$ either $\mathbb{E}_\pi[\tau_A(\alpha)]$ or $T_0(\alpha)$. Then
(1) $\mathbb{E}_\pi[\text{exp}(-\lambda\tau_A(\alpha))] = \mathbb{E}_\pi[\text{exp}(-\lambda\tau_A(-\alpha))]$ and $R(\alpha) = R(-\alpha)$, $\alpha \in [-1, 1]$.
(2) $\mathbb{E}_\pi[\text{exp}(-\lambda\tau_A(\alpha))]$ is non-increasing for $\alpha \in [-1, 0]$ and $R(\alpha)$ is non-decreasing for $\alpha \in [-1, 0]$.

Aldous-Fill's conjecture

- Let
$$Z_{ij} = \sum_{n=0}^{\infty} [P_{ij}^{(n)} - \pi_j]$$
be the fundamental matrix of P.
- Aldous-Fill in their book (1995++)[Chapter 9, conjecture 22] conjectured that
$$\text{trace}[Z^2(P^* - P)] \geq 0.$$
- Let $P_\lambda = \lambda P + (1-\lambda)P^*$, $\lambda \in [0,1]$, then (we proved) this conjecture is equivalent to
$$T_0(P_\lambda) \leq T_0(P_{1/2}).$$
- Previous result (H. and Mao, 2017): This is true!

Stronger version of Aldous-Fill's conjecture

Theorem

Assume that X *is an irreducible Markov chain on the finite state space with PTM* P. *Then* $[Z(P^* - P)Z]_{ii} \geq 0$ *for any* i. *In particular, trace*$[Z^2(P^* - P)] \geq 0$.

Thank you for your attention!