P-10

マルコフ連鎖に対する混合時間
Mixing Times for Markov Chains

○高木　司[1]
*Tsukasa Takagi[1],

Abstract: We discuss the upper and lower bound for mixing times for Markov chains and present some applications.

1. はじめに

ランダムな時間発展であるマルコフ連鎖について、一定の条件下では極限分布を持つことが知られている。具体的には、時刻 t の位置の分布は、$t \to \infty$ のとき時間発展による不動点とも言える、定常分布へ収束する。このことを用いた数値計算手法が、マルコフ連鎖モンテカルロ法 (MCMC) である。

この収束の速さは、少なくとも指数的であることが既知であるが、あくまで一般論によるものであり、個々のマルコフ連鎖の特徴を反映したものではないため、その最適性には再考の余地がある。MCMC における、必要な精度を得るために必要な反復回数に対する評価とも言え、応用上でも重要な問題である。

本講演では、上記で述べた定常分布への収束の速さについて、参考文献 [1] に基づいて概説する。

2. マルコフ連鎖と既約性・非周期性

以降、状態空間 $\mathcal{X}$ は有限集合とする。

定義 1 (マルコフ連鎖) $\mathcal{X}$ 値の確率変数列 $X = (X_0, X_1, \dots)$ がマルコフ性

任意の $x, y \in \mathcal{X}, t \geq 1$ と $P(H_{t-1} \cap \{X_t = x\}) > 0$ を満たす全ての事象 $H_{t-1} = \bigcap_{s=0}^{t-1} \{X_s = x_s\}$ に対して

$$\begin{aligned} P\{X_{t+1} = y | H_{t-1} \cap \{X_t = x\}\} \\ = P\{X_{t+1} = y | X_t = x\} = P(x, y) \end{aligned}$$

を満たす時、X は推移確率を P とするマルコフ連鎖であるという。

以下の議論において、マルコフ連鎖に対して課す条件について述べる。

定義 2 (既約) 任意の $x, y \in \mathcal{X}$ に対して,$P^t(x, y) > 0$ となる整数 t が存在するとき、推移確率 P は既約的と言う。ただし、$t > 0$ に対して

$$P^t(x, y) = P\{X_t = y | X_0 = x\}$$

である。

定義 3 (周期) $x \in \mathcal{X}$ に対して,

$$T(x) = \{t \geq 1 : P^t(x, x) > 0\}$$

とする。$T(x)$ の最大公約数を状態 x の周期と呼ぶ。すべての状態が周期 1 であるならば、その連鎖は非周期的と言い、そうでないならば周期的と言う。

本稿では、推移確率 P は既約的かつ非周期的と仮定する。

3. 定常分布への収束と混合時間

今の仮定の下では、次で定める定常分布がただ一つ存在し、時刻無限大の極限分布として現れる。

定義 4 (定常分布)

$$\pi(y) = \sum_{x \in \mathcal{X}} \pi(x) P(x, y), \quad y \in \mathcal{X}$$

を満たす $\mathcal{X}$ 上の確率分布 π を定常分布と呼ぶ。

この収束についてより詳しく述べると、定常分布 π に対し、

$$\|P^t(x, \cdot) - \pi\|_{TV} \leq C_1 e^{-C_2 t}$$

となる定数 $C_1, C_2 > 0$ が存在することが知られている。ただし、$\|\cdot\|_{TV}$ は全変動距離、すなわち $\mathcal{X}$ 上の分布 μ, ν に対し、

$$\|\mu - \nu\|_{TV} = \max_{A \subseteq \mathcal{X}} |\mu(A) - \nu(A)| \tag{1}$$

とする。なお、証明の中で与えられる定数 C_1, C_2 は最良とは限らないことに注意する。本稿では、この収束の速さについて、個々の連鎖の特徴を踏まえつつ、より詳細な形で考察する。特に興味があるのは、

$$t_{mix}(\varepsilon) = \min\{t; d(t) \leq \varepsilon\}, \quad \varepsilon > 0$$

で与えられる混合時間である。ただし、

$$d(t) = \max_{x \in \mathcal{X}} \|P^t(x, \cdot) - \pi\|_{TV}$$

1: 日大理工・院 (前)・数学

である。ここで、

$$t_{mix} = t_{mix}(1/4)$$

とすると、

$$t_{mix}(\varepsilon) \leq \lceil \log_2 \varepsilon^{-1} \rceil t_{mix}, \quad \varepsilon > 0$$

が得られる。よって、記号を簡略化するため t_{mix} を評価することにする。また、

$$\bar{d}(t) = \max_{x,y \in \mathcal{X}} \|P^t(x,\cdot) - P^t(y,\cdot)\|_{TV}$$

とすると、

$$d(t) \leq \bar{d}(t) \leq 2d(t) \tag{2}$$

が成り立つ。このことから、d の代わりに $\bar{d}$ を評価すればよいことが分かる。

4. t_{mix} の上界

t_{mix} の上界を与える方法の一つに coupling を用いる方法がある。本節ではその概略を述べる。

定義 5 (Coupling) $\mathcal{X}$ 上の確率分布 μ と ν の *coupling* とは、一つの確率空間上で定義され、*X* の分布が μ, *Y* の分布が ν であるような確率変数 *(X,Y)* の組のことである。

(1) で与えた全変動距離を coupling を用いて表すことができる。

命題 6 μ, ν を $\mathcal{X}$ 上の確率分布とする。この時、

$$\begin{aligned} &\|\mu - \nu\|_{TV} \\ &= \inf\{P(X \neq Y) : (X,Y) \text{ は } \mu \text{ と } \nu \text{ の } \textit{coupling}\} \end{aligned}$$

が成り立つ。

この命題と (2) から、$P^t(x,\cdot)$ と $P^t(y,\cdot)$ の coupling を適切にとることにより $\bar{d}$, ひいては d の評価を得られることがわかる。ここでは、時刻 t だけではなく、連鎖としての coupling を導入する。

定義 7 (Markovian coupling) 次の条件を満たす状態空間 $\mathcal{X} \times \mathcal{X}$ のマルコフ連鎖 $\{(X_t, Y_t)\}_{t \geq 0}$ を *Markovian coupling* と言う。

任意の $t \geq 0, x, y, x', y' \in \mathcal{X}$ に対して、

$$\begin{aligned} P(X_{t+1} = x' | X_t = x, Y_t = y) = P(x, x') \\ P(Y_{t+1} = y' | X_t = x, Y_t = y) = P(y, y') \end{aligned}$$

Markovian coupling が与えられると、それに基づいて t_{mix} の上界を与えることができる。

定理 8 $x, y \in \mathcal{X}$ に対して、$X_0 = x, Y_0 = y$ であるような *Markovian coupling* $\{(X_t, Y_t)\}_{t \geq 0}$ の分布を $P_{x,y}$ とする。また、

$$\tau_{couple} = \min\{s : X_s = Y_s\}$$

とすると、

$$d(t) \leq \max_{x,y \in \mathcal{X}} P_{x,y}(\tau_{couple} > t)$$

であり、従って

$$t_{mix} \leq 4 \max_{x,y \in \mathcal{X}} E_{x,y}(\tau_{couple})$$

5. t_{mix} の下界

本節では、bottleneck ratio という量を用いて t_{mix} の下界を与える方法について紹介する。

定義 9 (Bottleneck ratio) $S \subset \mathcal{X}$ に対し、*bottleneck ratio* $\Phi(S)$ を

$$\Phi(S) = \frac{Q(S, S^c)}{\pi(S)}$$

により定義する。ただし、$x, y \in \mathcal{X}$ と $A, B \subset \mathcal{X}$ に対し、

$$\begin{aligned} Q(x,y) &= \pi(x)p(x,y) \\ Q(A,B) &= \sum_{x \in A} \sum_{y \in B} Q(x,y) \end{aligned}$$

とする。また、連鎖全体に対する *bottleneck ratio* Φ_* を

$$\Phi_* = \min_{S: \pi(S) \leq 1/2} \Phi(S)$$

により定める。

この bottleneck ratio から t_{mix} の下界が得られる。

定理 10

$$t_{mix} = t_{mix}(1/4) \geq \frac{1}{4\Phi_*}$$

が成り立つ。

6. 例

本節では、定理 8, 定理 11 を用いて、混合時間の評価を与えられるような例を示す。

例 11 (Lazy random walk on binary tree) 深さ $k \geq 1$ の *binary tree* $T_{2,k}$ 上 *lazy random walk* に対して、

$$\frac{n-2}{2} \leq t_{mix} \leq 24n$$

となる。ただし、n は $T_{2,k}$ の頂点数である。

他の応用例については講演の中で述べる。

7. 参考文献

[1] D. A. Levin and Y.l Peres, Markov Chains and Mixing Times (Second Edition), AMS, 2017.