Geometric Ergodicity of Hamiltonian Monte Carlo

竹田航太

2022 年 11 月 6 日

概要

MCMC などのアルゴリズムにおいて生成される estimator の有効性を考える上で Markov Chain の収束について考えることは重要である．収束性は (Geometric) Ergodicity という概念で表現されるが，何らかのノルムによる定常分布への収束性を指すものである．ここでは主に HMC に関する収束定理についてまとめる．確率微分方程式やマルコフ連鎖，凸最適化に関する知識が求められる．

1　ノルム

収束定理において使われる測度空間のノルムを整理する．現在確認しているものは以下

(1) total variaion:
$d_{TV}(\nu_1, \nu_2) = \|\nu_1 - \nu_2\|_{TV} = \sup_{A \in \mathcal{F}}(\nu_1(A) - \nu_2(A)) = \int_A |\frac{d\nu_1}{d\nu_2} - 1| d\nu_2$

(2) χ^2-距離:(これは距離ではない)
$\|\nu_1 - \nu_2\|_{\chi^2} = \int_A |\frac{d\nu_1}{d\nu_2} - 1|^2 d\nu_2$

(3) Wasserstein-k 距離:
$d_{W_k}(\nu_1, \nu_2)^k = W_k(\nu_1, \nu_2)^k = \inf_{(X,Y) \in \mathcal{C}(\nu_1,\nu_2)} \mathbb{E}[\|X - Y\|^k]$

1.1　ノルム間の不等式

ノルム間に成り立つ関係については次を参照 [1, ON CHOOSING AND BOUNDING PROBABILITY]

例えば

$$d_{min} \cdot d_{TV} \leq W_1 \leq diam(\Omega) \cdot d_{TV}$$

2 証明にあたっての仮定

HMC の収束性を証明するためによく仮定される条件をまとめる．分布を $\pi = e^{-U}$ と表し，ポテンシャル U について条件を課す．典型的なものは log-concave と gradient-Lipschitz である．

2.1 Log-Concave

HMC の収束において density の log-concave 性 (もしくはポテンシャルの convex 性) が必要とされる．この背景には convex optimization の理論がある．しかし，応用上は non-convex な最適化を必要とする場合があり，HMC の non-concave density への適用の需要も高まっている．

2.2 Gradient-Lipschitz

後述の通り，gradient-Lipschitz 性から density から誘導される確率微分方程式の解の存在が示される．

2.3 Bounded

状態空間がコンパクトでポテンシャルが有界な時，Log-concave 性を使わず簡単に HMC の収束を示すことができる．

3 SDE との関係

一般に MCMC に対する収束定理は対応する SDE の離散化解のノルム収束評価として捉えることができる．

3.1 Langevin diffusion

HMC の簡易版のアルゴリズムとして Langevin Monte Carlo がある．これは π から定められる Langevin 方程式と呼ばれる以下の確率微分方程式を離散化 (Euler scheme) したアルゴリズムである．

$$dY_t = -\nabla U(Y_t)dt + \sqrt{2}dB_t \tag{3.1}$$

[2, Durmus] によると U に関するの convex 性や gradient-Lipschitz 性の仮定の元で π に対する log-Sobolev 不等式や Y_t に対する状態空間の次元に依存しない指数収束性が示される．また，[Karatzas, Shreve][*1] の Chapter5,Thm2.5 によると gradient-Lipschitz 性から (3.1) と 2 次モー

[*1] https://link.springer.com/content/pdf/

メントをもつ初期分布 μ_0 に対する一意な強解の存在を示せる．さらに Thm2.9 によれば (3.1) から定まる半群[*2] P_t は π と可逆になり，π を不変測度としてもつ．[*3]

3.1.1　SDE の知識

以下などを参考にして SDE に関する知識をつける必要がある．

Karatzas, Shreve; `https://link.springer.com/content/pdf/`

Ergodicity for SDEs; `https://www.sciencedirect.com/science/article/pii/S0304414902001503`

Eponential convergence for LD; `https://projecteuclid.org/journals/bernoulli/volume-2/issue-4/Exponential-convergence-of-Langevin-distributions-and-their-discrete-approximations/bj/1178291835.full`

4　収束・Ergodicity

Markov kernel の定義は Markov Chain に譲る．

4.1　Markov kernel の表現

$(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ 上の Markov Chain$(X_n)_{n\geq 0}$ を考える．この Markov kernel を定式化を行う．

$(\Theta, \mathcal{B}_\Theta)$ 上の r.v. θ を考え，これにより Chain を生成する際のランダム性を表現する．

Definition 4.1. $f_\theta : \mathbb{R}^d \to \mathbb{R}^d$ に対して，次の形の *Markov kernel* を考える．

$$P(x, A) = \int 1_A(f_\theta(x))\gamma(d\theta) \qquad\qquad x \in \mathbb{R}^d, A \in \mathcal{B}(\mathbb{R}^d)$$

今，分布 π への chain を Metropolis-Hasting Algorithm によって構成するので f_θ の形は以下のようになる．

$$f_\theta(x) = \begin{cases} g_\xi(x) \ (u < a(x, g_\xi(x))) \\ x \text{ other} \end{cases}$$

ここで $\theta = (\xi, u)$ $u \sim U[0,1]$ であり，$\xi \sim \mu \in \mathcal{M}_1(\mathbb{R}^d)$ として，g_ξ は candidate(proposal) map と呼ばれる．

g_ξ を用いて proposal kernel $Q : \mathbb{R} \times \mathcal{B}(\mathbb{R}^d) \to [0,1]$ は次で与えられる．

$$Q(x, A) = \int 1_A(g_\xi(x))\mu(d\xi) \qquad\qquad x \in \mathbb{R}^d, A \in \mathcal{B}(\mathbb{R}^d)$$

[*2] transion kernel とも呼ばれる

[*3] マルコフ連鎖については `https://litharge3141.github.io/blog_pdf/markov_chain/markovchain.pdf` なども参照

さらに, $\pi, Q(x, dy)$ にそれぞれ density$q(x, y), \pi(x)$ の存在を仮定すると, acceptance probability $a(x, y)$ は次のようにかける

$$a(x, y) = 1 \wedge \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} 1_S$$

(ただし，$S = \{\pi(x)q(x, y) > 0\}$) 特に $\pi(x)q(x, y) = 0$ の場合は $a(x, y) = 1$ となる.

Definition 4.2 (MH-type kernel). *Markov kernel P が Metropolis-Hasting type(MH-type) であるとは以下を満たすこと.*

$$P(x, A) = a(x, y)Q(x, dy) + r(x)\delta_x(dy)$$

ただし，$r(x) = 1 - \int a(x, y)Q(x, dy)$

Lemma 4.3 (MH-type の導出). 上で f_θ から定義した P は *MH-type* となる.

4.2 Exmaple

g_ξ の例を典型的なアルゴリズムで紹介する.

表 1　g_ξ の例

Algorithm	$g_\xi(x)$	$\xi \sim$
RWM	$x + \xi$	mean zero symmetric measure on $\mathbb{R}^d$
MALA	$x + \frac{h}{2}\nabla \log(\pi(x)) + \sqrt{h}\xi$	$N(0, I)$
HMC	$\mathrm{Pr}_x \circ \phi_T(x, p)$	$\xi = (p, T)$; $p \sim N(0, M), T \sim$ measure on $\mathbb{N}$

Remark 4.4. *MALA は以下の Langevin diffusion の Euler-Maruyama scheme による離散化.*

$$dX_t = \nabla \log(\pi(X_t))dt + \sqrt{2}dW_t$$

ただし，W_t は *Brawnian Montion.*

4.3 Conditions

目的の定常分布 π に収束するための Markov kernel P に対する条件を考える.

4.3.1 定常分布と可逆性

まず，定常分布の定義について.

Definition 4.5 (Markov 遷移核の作用)**.** $\mu \in \mathcal{M}_1(\mathbb{R}^d)$ に対して，$\mu P \in \mathcal{M}_1(\mathbb{R}^d)$ を次で定める．

$$\mu P(\cdot) \coloneqq \int_{\mathbb{R}^d} \mu(dx) P(x, \cdot)$$

$f : \mathbb{R}^d \to \mathbb{R}$ *s.t.* 有界，$\mathcal{B}(\mathbb{R}^d)$-可測に対して，$Tf : \mathbb{R}^d \to \mathbb{R}$ を次で定める．

$$Pf(\cdot) = \int_{\mathbb{R}^d} f(y) P(\cdot, dy)$$

また，*m-step Markov* 遷移核 P^m $(m \geq 2)$ を再帰的に以下で定義する．

$$P^m(x, A) \coloneqq (P(x, \cdot) P^{m-1})(A) = \int_{\mathbb{R}^d} P(x, dy) P^{m-1}(y, A) \tag{4.1}$$

Definition 4.6 (定常分布と可逆性)**.** $\pi \in \mathcal{M}(\mathbb{R}^d)$ が *P-stationary(invariant)* とは以下が成り立つこと

$$\pi P = \pi$$

また，π と P が *reversible* とは以下が成り立つこと

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

Lemma 4.7. *Markov kernel* P が π に対して *reversible* のとき π は *P-invariant* となる．

Proof.

$$\int_{x \in \mathcal{X}} \pi(dx) P(x, dy) = \int_{x \in \mathcal{X}} \pi(dy) P(y, dx) = \pi(dy) \int_{x \in \mathcal{X}} P(y, dx) = \pi(dy)$$

□

Lemma 4.8. π に対して定義した *MH-type kernel* は π に *reversible* な *chain* を作る．

4.3.2 定常分布への収束

次に一意的な定常分布への収束を保証するための条件を考える．

Definition 4.9 (ϕ-irriducible)**.** *Markov kernel* P が $\exists \phi$*: σ-finite measure on* $\mathcal{X}$ に対して，*φ-irriducible(*既約*)*
$\overset{def}{\Leftrightarrow}$ 任意の $x \in \mathcal{X}$ と $\phi(A) > 0$ となる $A \in \mathcal{B}(\mathcal{X})$ に対し，ある $n = n(x, A) \in \mathbb{N}$ *s.t.* $P^n(x, A) > 0$

Example 4.1 (Running Example)**.** $\pi \in \mathcal{M}(\mathbb{R}^d)$ の *density* を同様に π と書き，π に対する *MH-type Algorithm* を考える．*proposal density* $q(x, y)$ は $\mathbb{R}^d \times \mathbb{R}^d$ 上正値，連続とし，π は至

る所有限とする．
このとき，このアルゴリズムは π-*irridiucible* となる．
確かに，$\forall x \in \mathbb{R}^d$ と $\pi(A) > 0$ となる A に対して，$\exists R > 0$ *s.t.* $\pi(A_R) > 0$（ただし，$A_R = A \cap B_R(0)$）．次に連続性から，$\exists \epsilon$ *s.t.* $\forall x \in \mathbb{R}^d, \inf_{y \in \mathbb{R}^d} \min\{q(x,y), q(y,x)\} \geq \epsilon$ $\pi(x) > 0$ として，[*4]

$$\begin{aligned}
P(x,A) \geq P(x,A_R) &\geq \int_{A_R} q(x,y) \min\left[1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right] dy \\
&\geq \epsilon \int_{A_R} \frac{\pi(y)}{\pi(x)} dy \\
&\geq \epsilon |\{y \in A_R; \pi(y) \geq \pi(x)\}| + \frac{\epsilon}{\pi(x)} \pi(\{y \in A_R; \pi(y) < \pi(x)\})
\end{aligned}$$

π は *Lebesgue* 測度に対して絶対連続なので $|A_R| > 0$ であり，そのため最右辺は同時に 0 にならない．以上から $P(x,A) > 0$ となり，P は π-*irridiucible* である．

Definition 4.10 (aperiodic)**.** *P with* 定常分布 π が *appriodic* とは次を満たす $d \geq 2$ が存在しないこと．*disjoint* $A_1, \cdots, A_d \in \mathcal{B}(\mathbb{R}^d)$ *s.t.*
$P(x, A_{i+1}) = 1$ *if* $x \in A_i$ *and* $P(x, A_1) = 1$ *if* $x \in A_d$ *and* $\pi(A_1) > 0$

Example 4.2 (Running Example 続き)**.** 追加の過程なしで *aperiodic* が成り立つ．背理法により示す．*disjoint* な $\mathcal{X}_1, \mathcal{X}_2 \subset \mathcal{X}$ *s.t.* $\pi(\mathcal{X}_1), \pi(\mathcal{X}_2) > 0$ かつ $P(x, \mathcal{X}_2) = 1$ $\forall x \in \mathcal{X}_1$ を考える．π の絶対連続性から $|\mathcal{X}_1| > 0$ となるので，

$$P(x, X_1) \geq \int_{y \in \mathcal{X}_1} q(x,y) a(x,y) dy > 0$$

次に asymptotic convergence theorem を述べる．状態空間には可算生成な σ 代数を仮定するが $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ はこれを満たす．定理の証明は [3] に譲る．

Theorem 4.11 (定常分布への収束)**.** *Markov* 連鎖 (T, δ_x) は ϕ-*irriducible, aperiodic* で定常分布 $\pi \in \mathcal{M}_1(\mathcal{X})$ をもつとする．このとき π *a.e.* $x \in \mathcal{X}$ で

$$\lim_{m \to \infty} \|P^m(x, \cdot) - \pi(\cdot)\|_{TV} = 0$$

Remark 4.12. *Theorem 4.11* は P に ϕ-*irriducible* と *aperiodic* を要請するが，π が定常分布となるアルゴリズム *(MH-type* など*)* を使えば大きな問題ではなくなる．

4.4 Geometric Ergodicity

次は定常分布への収束の速さを評価したい．べき的 (geometric) な収束を保証するのが (Geometric)Ergodicity である．

[*4] そうでない場合は $a(x,y) = 1$ となり，直ちに $P(x,A)$ が従う．

Definition 4.13 (Geometric Ergodicity)**.** *Markov Chain* P *with stationary* π が *Geometrically Ergodic* とは以下を満たすこと．
$\exists \rho < 1, \exists M : \mathbb{R}^d \to [0,1]$*;* π*-a.e. finite s.t.*

$$\|P^n(x,\cdot) - \pi(\cdot)\|_{TV} \leq M(x)\rho^n$$

Definition 4.14 (Lyapunov function, Drift condition)**.** *Markov kernel* P に対して*, Lyapunov function* V が存在する *(*もしくは *Drift condition* を満たす*)*
$\overset{def}{\Leftrightarrow}$ $V : \mathbb{R}^d \to [1,\infty]$ *s.t.* $\exists 0 < \lambda < 1, \exists b < \infty, \exists \omega < \infty$

$$PV(x) \leq \lambda V(x) + b1_{C_\omega}(x) \tag{4.2}$$

ただし，$C_\omega = \{x; V(x) \leq \omega\}$

Theorem 4.15 (Geometrically Ergodic Theorem)**.** $x \in \mathbb{R}^d$ を初期値として *Markov kernel* P により生成される *Chain* が π*-irriducible, aperiodic* であり，*Lyapunov function* V をもつとする．このとき P は *Geometicaly Ergodic* である．*i.e.*
$\exists \rho < 1, \exists M : \mathbb{R}^d \to [0,1]$*;* π*-a.e. finite s.t.*

$$\|P^n(x,\cdot) - \pi(\cdot)\|_{TV} \leq M(x)\rho^n$$

Example 4.3 (not irreducibile chain)**.** $U(x) = x, L = 2, \epsilon = \sqrt{2}$ の時，$\phi(x_0, p_0) = x_0$ となり *chain* は *not irreducibile.*

Remark 4.16 ([4])**.** *MH-type* について，$\operatorname{ess\,sup} r(x) = 1$ のとき，*geometric ergodicity* は成り立たない．

Example 4.4 ([4])**.** 以下の形の目的分布に対して，*Random Walk Metropolis* は *not geometrically ergodic.*

$$\pi(x,y) \propto \exp -(x^2 + x^2y^2 + y^2).$$

参考文献

[1] Alison L. Gibbs and Francis Edward Su. On choosing and bounding probability metrics. *International Statistical Review / Revue Internationale de Statistique*, 70(3):419–435, 2002.

[2] Alain Durmus and Éric Moulines. Sampling from a strongly log-concave distribution with the Unadjusted Langevin Algorithm. Preliminary version, April 2016.

[3] Gareth O. Roberts and Jeffrey S. Rosenthal. General state space markov chains and mcmc algorithms. *Probability Surveys*, 1(none), Jan 2004.

[4] GO Roberts and RL Tweedie. Geometric convergence and central limit theorems for multidimensional hastings and metropolis algorithms. *BIOMETRIKA*, 83(1):95–110, MAR 1996.

[5] Samuel Livingstone, Michael Betancourt, Simon Byrne, and Mark Girolami. On the geometric ergodicity of hamiltonian monte carlo, 2018.

[6] Oren Mangoubi and Aaron Smith. Rapid mixing of hamiltonian monte carlo on strongly log-concave distributions, 2017.

[7] Arnak S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities, 2016.

[8] Gareth O. Roberts and Richard L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341 – 363, 1996.