Markov Chain

竹田航太

2021 年 6 月 28 日

目次

概要

Markov Chain とその収束概念についてまとめる．基本的な測度論を用いた確率論の知識は前提とする．基本的に [1]，[2] による．

1 Preliminary

確率論の基礎的な定義や Notation を記述する．

Notation

- 確率空間を $(\Omega, \mathcal{F}, P)$ と書く．
- Ω 上の確率測度全体の集合を $\mathcal{M}_1(\Omega)$ と書く．
- 確率測度と確率密度関数を同じ記号で書くことがある．
- 確率変数は r.v. と略記する．特に断りがない場合は $\mathbb{R}$-値とする．
- 確率過程は s.p. と略記する．

1.1 確率変数

確率空間 $(\Omega, \mathcal{F}, P)$ 上の $\mathbb{R}^n$-値確率変数 X のことを，前提となる確率空間を明示的に書かずに，$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P)$ 上の確率変数 X と書くことがある．

1.2 確率変数の収束

Definition 1.1. $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P)$ 上の *r.v.* の列 $(X_m)_{m\in\mathbb{N}}$ と *r.v.* X を考える．以下の *4* つの収束を定義する．

(1) X_m が X に概収束する．*(*$X_m \overset{a.s.}{\to} X$ などと書く．*)*
$\overset{def}{\Leftrightarrow} \lim_{m\to\infty} X_m(\omega) = X(\omega)$ *a.s.* ω

(2) X_m が X に L^p 収束する．*(*$X_m \overset{L^p}{\to} X$ などと書く．*)*
$\overset{def}{\Leftrightarrow} \lim_{m\to\infty} \mathbb{E}[|X_m - X|^p]^{1/p} = 0$

(3) X_m が X に確率収束する．*(*$X_m \overset{P}{\to} X$ などと書く．*)*
$\overset{def}{\Leftrightarrow} \forall\epsilon > 0, \lim_{m\to\infty} P(|X_m - X| > \epsilon) = 0$

(4) X_m が X に法則収束する．*(*$X_m \overset{d}{\to} X$ や $X_m \overset{\mathcal{L}}{\to} X$ などと書く．*)*
$\overset{def}{\Leftrightarrow}$ 任意の F_X の連続点 x で $\lim_{m\to\infty} F_{X_m}(x) = F_X(x)$*(*ただし，$F_{X_m}, F_X$ はそれぞれ X_m, X の分布関数*)*
($\Leftrightarrow$ 像測度 P^{X_m} が P^X に弱収束する．*)*

Definition 1.2 (測度の弱収束)**.** $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P)$ 上の確率測度 μ_m が μ に弱収束する．

$\overset{def}{\Leftrightarrow}$ 任意の有界連続関数 $f \in C_b(\mathbb{R}^n)$ に対して

$$\lim_{m\to\infty}\int_{\mathbb{R}^n} f d\mu_m = \int_{\mathbb{R}^n} f d\mu$$

1.3 確率過程

Definition 1.3 (確率過程)**.** 確率空間 $(\Omega, \mathcal{F}, P)$ と $T \subset \mathbb{R}$ を与える．各 $t \in T$ に対して X_t が $(\Omega, \mathcal{F}, P)$ 上の $r.v.$ であるとき $(X_t)_{t\in T}$ を確率過程という．

2 Overview

[1, Markov chains and stochastic stability] にある前書きをまとめる．

2.1 Theme

(1) splitting technique:
(2) ”Foster-Lyapunov” drift condition:
(3) continuity condition
(4) control model

鍵となる概念は splitting technique である．Part1 は上記のテーマや関連概念の紹介，Part2 は splitting method の使い方，Part3 は収束に関する結果を示す．splitting technique のアプローチは”geometrically ergodic” chain の考えられる最良な例を与える．

τ_C を任意の集合 C の到達時刻 (hitting time) とする．”n-step transition probability”を次のように表現する；$P^n(x, A) = \mathbb{P}(\Phi_n \in A|\Phi_0 = x)$．Part2 と Part3 の目標は以下の 2 つの条件を繋げること：

(1) chain が”small” set C に早く回帰する条件
(2) $P^n(x, A)$ が定常分布に収束する条件

2.2 Results

Chapter 15 で示される美しい結果は以下.
ϕ-irreducible, aperiodic chain に対して以下は同値

A) ある”small” set$C$ に対して，return time の分布は geometric tail をもつ:

i.e. $\exists r > 1$ s.t.

$$\sup_{x \in C} \mathbb{E}_x[r^{\tau_C}] < \infty$$

B) ある"small" set$C$ に対して，transition probability は geometrically quick に収束する:
i.e. $\exists M < \infty, P^\infty(C) > 0, \rho_C < 1$ s.t.

$$\sup_{x \in C} |P^n(x, C) - P^\infty| \leq M\rho_C^n$$

C) ある"small" set$C$ に対して，"geometric drift"が存在する．:
i.e. $\exists V \geq 1, \exists \beta > 0$ s.t.

$$\int P(x, dy)V(y) \leq (1-\beta)V(x) + 1_C(x)$$

それぞれが極限の確率測度 π の存在を示し，π は次を満たす．$\exists R < \infty, \exists \rho < 1$ s.t.

$$\sup_{|f| \leq V} |\int P^n(x, dy)f(y) - \int \pi(dy)f(y)| \leq RV(x)\rho^n$$

3 Markov 連鎖の基礎

離散時間の確率過程で遷移確率が現在の状態にのみ依存するものを考える．初めに状態空間が有限 (もしくは可算) なものを考えるの自然であるが，サンプリング問題に繋げていくために確率変数は連続的な状態を取り得る定義をする．

確率空間を $(\Omega, \mathcal{F}, P)$ とし，状態空間を Borel 空間 $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ とする．$\mathcal{B}(\mathcal{X})$ は可分とする．

3.1 Markov 連鎖の定義

Markov 遷移核 (Markov transition kernel) と初期分布を指定することで Markov 連鎖を定義できる．

Definition 3.1 (Markov 遷移核)**.** $P : \mathcal{X} \times \mathcal{B}(\mathcal{X}) \to [0, 1]$ が *Markov* 遷移核である．
$\overset{def}{\Leftrightarrow}$ 以下が成り立つ．

A) $\forall x \in \mathcal{X}, P(x, \cdot)$ が $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ 上の確率測度である．
B) $\forall A \in \mathcal{B}(\mathcal{X}), P(\cdot, A)$ は $\mathcal{B}(\mathcal{X})$-可測．

確率過程の有限時間のパスを Markov 遷移核により生成する．確率過程 $(X_k)_{k=0}^m$ をパス空間 $(\mathcal{X}^{m+1}, \bigvee_{k=0}^m \mathcal{B}(\mathcal{X}_k))$ 上に考え，$x \in \mathcal{X}$ から出発した確率過程のパス空間上の確率測度 $P_x^m(\cdot)$ を考える．$A_k \subset \mathcal{X}$ $(k = 1, 2, \cdots)$ に対して，以下のようにしてパス空間上の cylinder set に確

率を定める．

$$P_x^1(A_1) = P(x, A)$$
$$P_x^2(A_1 \times A_2) = \int_{A_1} P(x, dx_1) P(x_1, A_2)$$
$$\vdots$$
$$P_x^m(A_1 \times A_2 \times \cdots \times A_m) = \int_{A_1} P(x, dx_1) \int_{A_2} P(x_1, dx_2) \cdots P(x_{m-1}, A_m)$$

パス空間上の一般の Borel 集合に測度 P_x^m を拡張することで有限時間の Markov Chain を構成できる．次に m を無限大にしたときの Markov Chain の存在について以下の定理がある．

Theorem 3.2 (Markov 連鎖の存在)**.** 初期分布 $\mu \in \mathcal{M}_1(\mathcal{X})$，*Markov* 遷移核 P に対して，$(\prod_{m=1}^\infty \mathcal{X}_m, \bigvee_{m=1}^\infty \mathcal{B}(\mathcal{X}_m))$ 上の確率過程 $(X_m)_{m=1}^\infty$ と確率測度 P_μ が存在して次を満たす．任意の $m \in \mathbb{N}, (A_k)_{k=1}^m \subset \mathcal{B}(\mathcal{X})$

$$\begin{aligned} P_\mu(A_0 \times A_1 \times \cdots \times A_m) & \ (= \mathbb{P}(X_0 \in A_0, X_1 \in A_1, \cdots, X_m \in A_m)) \\ &= \int_{A_0} \mu(dy_0) \int_{A_1} P(y_0, dy_1) \cdots P(y_{m-1}, A_m) \end{aligned} \tag{3.1}$$

Proof. [1, Theorem 3.4.1] による．Kolmogorov の拡張定理を使う． □

Definition 3.3 (Markov 連鎖)**.** *Theorem 3.2* の確率過程 $(X_m)_{m=0}^\infty$ を *Markov* 連鎖という．

Remark 3.4. *Markov* 連鎖を初期分布 μ，*Markov* 遷移核 P，*(*確率空間の*)* 確率測度 P の組 $((X_m)_{m=0}^\infty, \mu, P, P)$ で表すこともある．

3.2 Markov 性

Definition 3.5 (Markov 性)**.** 確率変数の列 $(X_m)_{m\geq 0} : \Omega \to \mathcal{X}$ が以下を満たすとき *Markov* 性をもつという．任意の $m \in \mathbb{N}$ と $(A_k)_{k=0}^{m+1} \subset \mathcal{B}(\mathcal{X})$ に対して，

$$\mathbb{P}(X_{m+1} \in A_{m+1} | X_m \in A_m, \cdots, X_0 \in A_0) = \mathbb{P}(X_{m+1} \in A_{m+1} | X_m \in A_m)$$

3.3 Markov 遷移核の作用

Definition 3.6 (Markov 遷移核の作用)**.** $\mu \in \mathcal{M}_1(\mathcal{X})$ に対して，$\mu P \in \mathcal{M}_1(\mathcal{X})$ を次で定める．

$$\mu P(\cdot) \coloneqq \int_{\mathcal{X}} \mu(dx) P(x, \cdot)$$

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

$$Pf(\cdot) = \int_{\mathcal{X}} f(y) P(\cdot, dy)$$

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

$$P^m(x,A) \coloneqq (P(x,\cdot)P^{m-1})(A) = \int_{\mathcal{X}} P(x,dy)P^{m-1}(y,A) \tag{3.2}$$

Remark 3.7 (m-step Markov 遷移核)**.** *m-step Markov* 遷移核は *(3.1)* で $\mu = \delta_x$，$A_m = A, A_k = \mathcal{X}$ $(k = 0, \cdots, m-1)$ とした確率に等しい．

Theorem 3.8 (Chapman-Kolmogorov)**.** 任意の自然数 $m \geq k$ に対して，以下が成り立つ．

$$P^m(x,A) = \int_{\mathcal{X}} P^m(x,dy)P^{n-m}(y,A) \tag{3.3}$$

Proof. (3.1) と (3.2) から従う． □

4 MCMC

状態空間などの Notation は前節と同様．MCMC(Markov Chain Monte Carlo) の中でも，Metroplis-Hasting Algorithm についてまとめる．MH Algorithm は提案確率に従って現在の点から候補点を提案し，ある確率で次の点として採択するという流れで目的分布に従う Chain を構成するアルゴリズムである．次のような採択確率を考えることで pi から外れた候補点を棄却する．

Definition 4.1. 目的分布の *density* π と *proposal density* $q : \mathcal{X} \times \mathcal{X} \to [0,1]$ に対して採択確率 $a(x,y)$ を次で定める．

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

(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．Q : *proposal transition kernel, density* $q(x,y)$ *for* $Q(x,dy)$ と採択確率 a に対して，

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

ただし，$r(x) = 1 - \int a(x,y)Q(x,dy)$

Remark 4.3. 提案確率 $Q(x,dy)$（もしくはその *density* $q(x,y)$）はユーザーが設定する自由度がある．

5 定常分布

状態空間などの Notation は前節と同様.

5.1 定常分布と可逆性

Definition 5.1 (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

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

Lemma 5.2. *Markov kernel* P *が* π *に対して reversible のとき* π *は P-invariant となる.*

Proof.

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

□

5.2 Total Variation

Definition 5.3 (Total Variation)**.** $\nu_1, \nu_2 \in \mathcal{M}(\mathcal{X})$ に対し，*total variation distance* を以下で定める.

$$\|\nu_1 - \nu_2\|_{TV} = \sup_A |\nu_1(A) - \nu_2(A)|$$

Theorem 5.4. *total variation distance* は確率測度の空間 $\mathcal{M}(\mathcal{X})$ 上の距離となる.

この確率測度の距離に関する収束を議論する.

Lemma 5.5. 確率空間 *total variation distance* と *Markov kernel* P について以下が成り立つ.

a) $\|\nu_1 - \nu_2\|_{TV} = \sup_{f:\mathcal{X}\to[0,1] \text{ in } L^1(\nu_1)\cap L^1(\nu_2)} |\int f d\nu_1 - \int f d\nu_2|$

b) $\forall a < b,\ \|\nu_1 - \nu_2\|_{TV} = \frac{1}{b-a}\sup_{f:\mathcal{X}\to[a,b]} |\int f d\nu_1 - \int f d\nu_2|$

c) π が *P-invariant* であるとき，$\|P^n(x,\cdot) - \pi(\cdot)\|_{TV}$ は n に関して非増加.

d) より一般に，$\|(\nu_1 P) - (\nu_2 P)\|_{TV} \le \|\nu_1 - \nu_2\|_{TV}$

e) *P-invariant* な π に対して，$t(n) = 2\sup_{x\in\mathcal{X}} \|P^n(x,\cdot) - \pi\|$ とおくと，t は劣乗法的である.

f) ある σ-有限測度 ρ に関して，μ, ν がそれぞれ g, h という *density* をもつとき，$M = \max(g,h), m = \min(g,h)$ とおくと次が成り立つ.

$$\|\mu - \nu\|_{TV} = \frac{1}{2}\int_{\mathcal{X}} (M - m)d\rho = 1 - \int_{\mathcal{X}} m d\rho$$

g) $\forall \mu, \nu$ に対して，$\exists X, Y$*: jointly defined r.v. s.t.* $X \sim \mu, Y \sim \nu, \mathbb{P}[X = Y] = 1 - \|\mu - \nu\|_{TV}$

5.3 定常分布への収束

Definition 5.6 (ϕ-irridicible)**.** *Markov kernel* P *が* $\exists \phi$*:* σ*-finite measure on* $\mathcal{X}$ *に対して，* ϕ*-irridicible(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 5.1 (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* を $q(x, y)$ 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*-irridicible* 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，[*1]

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

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*-irridicible* である．

Definition 5.7 (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 5.2 (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，

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

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

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 [2] に譲る．

Theorem 5.8 (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 5.9. *Theorem 5.8* 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．

6 Ergodicity

定常分布への収束の速さを考えるための定式化を行う．

6.1 Uniform Ergodicity

Definition 6.1 (Uniform Ergodicity)**.** *Markov Chain P with stationary* π が *Uniform Ergodic* とは以下を満たすこと．
$\exists\rho < 1, \exists M > 0$*; π-a.e. finite s.t.*

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

Definition 6.2 (small set)**.** $C \subset \mathcal{X}$ が *'small'(*もしくは *(n_0, ϵ, ν)-small)*
$\overset{def}{\Leftrightarrow}$ $\exists n_0 \in \mathbb{N}, \epsilon > 0, \nu$*: meas on* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)$ *s.t.* $\forall x \in C, A \in \mathcal{B}(\mathbb{R}^d)$

$$P^{n_0}(x, A) \ge \epsilon\nu(A) \tag{6.1}$$

(6.1) を *minorisation condition* という．

Remark 6.3. *[1]* では *small set* の定義に $\pi(C) > 0$ を仮定しているが，多くの場合次の *Drift condition* を加えるとこれを示すことができるのでここでは省略する．

Theorem 6.4. *Markov Chain P with stationary* π が *(n_0, ϵ, ν)-small set* $C = \mathcal{X}$ をもつとする．このとき，P は *Uniform Ergodic* であり，次を満たす．

$$\|P^n(x, \cdot) - \pi(\cdot)\|_{TV} \le (1-\epsilon)^{\lfloor n/n_0 \rfloor} \ \forall x \in \mathcal{X}$$

Example 6.1 (Running Exmaple 続き)**.** 上で紹介した例で q に連続性を仮定したがこれだけでは任意の *compact set* が *small* にはならない．次の反例がある．$d = 1$ で $\pi(x) = 1_{0<|x|<1}|x|^{-1/2}, q(x, y) \propto \exp\{-(x-y)^2/2\}$ とすると，任意の 0 の近傍は *not small set*

しかしながら，*Running Exmaple* の一般的な設定でも任意の $\pi(x)$-*bounded* な *compact set* は *small* となる．C を *compact set* で $\pi(x) < k < \infty$ *on* C とする．$x \in C$ をとって，D を *compact set* で $|D|, \pi(D) > 0, \inf_{x \in C, y \in D} q(x,y) > \epsilon$ を満たすとする．このとき，

$$P(x, dy) \geq q(x,y)dy \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\} \geq \epsilon dy \min\left\{1, \frac{\pi(y)}{k}\right\}$$

6.2 Geometric Ergodicity

次は定常分布への収束の速さを評価したい．そのための概念が Ergodicity である．

Definition 6.5 (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 6.6 (Lyapnov function, Drift condition). *Markov kernel* P に対して，*Lyapnov 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.* $\exists 0 < \lambda < 1, \exists b < \infty, \exists \omega < \infty$

$$PV(x) \leq \lambda V(x) + b 1_{C_\omega}(x) \tag{6.2}$$

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

Theorem 6.7 (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，*Lyapnov 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.*

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

6.3 Quantitive Bounds of Convergence

上記の例のように Geomertic Ergodicity だけではサンプリングのパフォーマンスを保証できない．明示的に g を与えて，次のような評価をしたい．

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

目標は以下の bivaiate drift condition である．

$$\bar{P}h(x,y) \leq \alpha^{-1}h(x,y) \ (x,y) \notin C \times C \tag{6.3}$$

ただし，$\alpha > 1, \bar{P}h(x,y) = \int\int h(v,w)P(x,dv)P(y,dw)$.

Lemma 6.8. *drift condition(6.2)* が $V:\mathcal{X}\to[1,\infty], C\subset\mathcal{X}, \lambda<1, b<\infty$ で成立し，$d=\inf_{x\in C^c}V(x)$ とおくと，$d>b/(1-\lambda)-1$ を満たすとする．このとき *(6.3)* が同じ C と

$$h(x,y)=\frac{1}{2}[V(x)+V(y)]$$
$$\alpha^{-1}=\lambda+\frac{b}{(d+1)}<1$$

また，次の量を考える．

$$B_{n_0}=\max[1,\alpha^{n_0}(1-\epsilon)\sup_{C\times C}\bar{R}h] \tag{6.4}$$

ただし，$\bar{R}h(x,y)=\int\int(1-\epsilon)^2h(z,w)(P^{n_0}(x,dz)-\epsilon\nu(dz))(P^{n_0}(y,dw)-\epsilon\nu(dw))$ on $C\times C$.

Theorem 6.9. $\mathcal{X}$ 上の *Markov Chain* P を考え，$C\subset\mathcal{X}, h:\mathcal{X}\times\mathcal{X}\to[1,\infty), \alpha>1, n_0\in\mathbb{N}, \epsilon>0$ があって *minorisation condition(6.1)* と *bi-drift condition(6.3)* が成立するとする．*(6.4)* で B_{n_0} を定める．

このとき，任意の初期分布 $\mathcal{L}(X_0,X_0')$ と $\forall k\in\mathbb{N}$ と $1\le\forall j\le k$ に対し，$\mathcal{L}(X_0,X_0')$ から始まる *P-chain*$(x_n),(X_n')$ は次を満たす．

$$\|P^{X_k}-P^{X_k'}\|_{TV}\le(1-\epsilon)^j+\alpha^{-k}(B_{n_0})^{j-1}\mathbb{E}[h(X_k,X_k')]$$

特に，$j=\lfloor rk\rfloor$ を十分大きな $r>0$ と撮ると，上の右辺は $k\to\infty$ のとき $\to 0$ となる．

6.4 Convergence Proofs using Coupling Constructions

ここでは [2] で使われた Coupling を使った MCMC の解析を紹介し，いくつかの定理を改めて証明する．[1] にあるような長い解析議論をせずに証明することができる．

6.4.1 Coupling 不等式

coupling の基本的なアイデアは次のようなものである．X,Y を jointly definded r.v. on $\mathcal{X}$ とする．像測度をそれぞれ P^X,P^Y と書くと

$$\begin{aligned}\|P^X-P^Y\|_{TV}=&\sup_A|\mathbb{P}(X\in A)-\mathbb{P}(Y\in A)|\\=&\sup_A|\mathbb{P}(X\in A,X=Y)+\mathbb{P}(X\in A,X\neq Y)\\&-\mathbb{P}(Y\in A,X=Y)-\mathbb{P}(Y\in A,X\neq Y)|\\=&\sup_A|\mathbb{P}(X\in A,X\neq Y)-\mathbb{P}(Y\in A,X\neq Y)|\\\le&\mathbb{P}(X\neq Y)\end{aligned}$$

つまり

$$\|P^X - P^Y\|_{TV} \le \mathbb{P}(X \ne Y) \tag{6.5}$$

2 つの r.v. の像測度の total variation distance は 2 者が等しくない確率で上から抑えられる．

6.4.2 Small Sets and Couplling

Coupling 不等式を使って $P^n(x,\cdot)$ と π の差を評価するために以下のような Coupling construction を行う．P を chain, C を (n_0, ϵ, ν)-small set とする．$(X_n)_n, (X'_n)_n$ を 2 つの P-chain の複製とする．

Definition 6.10 (Coupling Construction)**.** $X_0 = x, X'_0 \sim \pi(\cdot), n = 0$
loop given X_n, X'_n*:*
1. if $X_n == X'_n$*, then* $X_{n+1} = X'_{n+1} \sim P(X_n, \cdot)$, $n \leftarrow n+1$
2. else if $(X_n, X'_n) \in C \times C$ *then:*

- *w.p.* $\epsilon, X_{n+n_0} = X'_{n+n_0} \sim \nu(\cdot)$
- *w.p.* $1-\epsilon$*, conditionally independent*
  $X_{n+n_0} \sim \frac{1}{1-\epsilon}[P^{n_0}(X_n,\cdot) - \epsilon\nu(\cdot)]$, $X'_{n+n_0} \sim \frac{1}{1-\epsilon}[P^{n_0}(X_n,\cdot) - \epsilon\nu(\cdot)]$

$n_0 > 1$ の場合は $X_{n+1}, \cdots, X_{n+n_0-1}$ を後ろ向きに構成し X_n と X_{n+n_0} に対する条件付き分布が合うようにする．(X'_n) についても同様.
3. else:
$X_{n+1} \sim P(X_n,\cdot), X'_{n+1} \sim P(X'_n,\cdot)$ *conditionally independent*

Remark 6.11. *Coupling Construction* は以下のような性質から P^n と π の差をうまく評価できる．

- 上記の手順で構成される $(X_n)_n, (X'_n)_n$ は正しく P で *update* されている，
- さらに，$\mathbb{P}(X_n \in A) = P^n(x, A), \mathbb{P}(X'_n \in A) = \pi(A)$ が従う．
- *2* つの *chain* は両者が C に入るまで独立に生成される．
- *Coupling* 不等式から $\|P^n(x,\cdot) - \pi(\cdot)\|_{TV} \le \mathbb{P}[X_n \ne X'_n]$ が成り立ち，これを使って定理を示す．

Proof of Theorem 6.4. $C = \mathcal{X}$ なので n_0 回の反復毎にと X_n と X'_n は少なくとも ϵ の確率で等しくなる．このため，$n = n_0 m$ とすると $\mathbb{P}[X_n \ne X'_n] \le (1-\epsilon)^m$ が成り立つ．よって Coupling 不等式 (6.5) から

$$\|P^n(x,\cdot) - \pi(\cdot)\| \le (1-\epsilon)^m = (1-\epsilon)^{n/n_0}$$

Lemma 5.5 c) から任意の $n \in \mathbb{N}$ に対して

$$\|P^n(x,\cdot) - \pi(\cdot)\| \leq (1-\epsilon)^m = (1-\epsilon)^{\lfloor n/n_0 \rfloor}$$

が成り立つ. □

Proof of Theorem 6.7. まず, $h(x,y) = \frac{1}{2}[V(x) + V(y)]$ とおく. 次の lemma から V が C 上有界を仮定する.

Lemma 6.12. *w.l.o.g.* 以下を仮定できる.

$$\sup_{x \in C} V(x) < \infty \tag{6.6}$$

i.e. given small set C*, drift* 関数 V *satisfies (6.1), (6.2)* に対して, *(6.1), (6.2)* を保ったまま, *(6.6)* を満たす $C_0 \subset C$ を取れる.

(6.6) を仮定できたので drift (6.2) と合わせて

$$\sup_{(x,y) \in C \times C} \bar{R}h(x,y) < \infty \tag{6.7}$$

が成り立ち, (6.7) から $B_{n_0} < \infty$ がわかる. ここで

$$d = \inf_{C^c} V(x)$$

とおいて,

$$d > \frac{b}{(1-\lambda)} - 1$$

のとき, Lemma 6.8 から bi-drift condition (6.3) が成立.

今, $d > \frac{b}{(1-\lambda)} - 1$ が成り立つなら, Theorem 6.7 は Lemma 6.8 と Theorem 6.9 となり証明は終わる. しかし, $d \leq \frac{b}{(1-\lambda)} - 1$ の場合は必要な条件を保ったまま C を大きくとって, $d > \frac{b}{(1-\lambda)} - 1$ が成り立つように調整する必要がある. 具体的には任意の $d' \leq \frac{b}{(1-\lambda)} - 1$ に対して, $S = \{x \in \mathcal{X}; V(x) \leq d'\}$ とおき, $C' = C \cup S$ と定めると $d_{new} = \inf_{C'^c} V \geq d' > \frac{b}{1-\lambda} - 1$ となる. さらに, V は S 上有界なので (6.6) は C' でも成立し, (6.7) と (6.2) から B_{n_0} も成り立つ. 残りは C' が small set であることを示すのみ. 以下の Lemma から保証される.

Lemma 6.13. C' は *small set*

証明は [2] の 4 章を参照. □

参考文献

[1] Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability.* Springer Science & Business Media, 2005.

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

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