

マルコフ連鎖と混合時間

白井朋之

2008年1月8–11日
於 東北大学

1 序

“52枚のトランプのカードをシャッフルする際、何回程度で「ばらばら」になるか”という問いについて考える。この問いに対する答えとして、1990年1月9日付けのNew York Times紙に次のような見出しの記事がある [9]: “In Shuffling Cards, 7 Is Winning Number”. この記事はまず、52枚のトランプのカードをシャッフルする際7回シャッフルすれば十分で、それより少なければ十分にばらばらにならず、またそれより多くシャッフルしても混ざり具合が大きく改善されることはないという、P. Diaconis (Stanford 大学教授, 当時 Harvard 大学. マジシャンでもある) らによる発見について触れ、続けて “The mathematical proof, discovered after studies of results from elaborate computer calculations and careful observation of card games, confirms the intuition of many gamblers, bridge enthusiasts and casual players that most shuffling is inadequate.” と紹介している。なお Diaconis らがそのための論証に用いたシャッフルの方法は riffle shuffle と呼ばれるもので、カジノなどでよく用いられる方法である。

それでは「十分ばらばら」な状態とは一体どのような状態であろうか。カードのシャッフリングの場合、数学的には $52!$ の元からなる対称群上に値を取る Markov 連鎖の定常分布をもって「最もばらばら」な状態と定義し、「十分ばらばら」という状態を定常分布との全変動距離による差をもって定義する。この講義では、より一般の Markov 連鎖について、混合時間と呼ばれる「十分ばらばら」になるまでにかかる時間に焦点を当て解説する。また、上記のカードのシャッフリングに見られるような、ある値を境に混ざり具合に大きな変化が現れる現象は cut-off 現象と呼ばれ、対称性の強い Markov 連鎖においてこのような現象がしばしば観測されている。講義ではこの現象についても例を挙げて紹介する。

2 基礎概念の準備

本節では、確率変数や分布などの確率論の基礎概念についての解説を行う。 $(\Omega, \mathcal{F}, \mathbb{P})$ を確率空間、すなわち $\mathbb{P}(\Omega) = 1$ を満たす測度空間とする。 (S, \mathcal{S}) は可測空間とし、 S 上の確率測度全体を $\mathcal{P}(S)$ で表すことにする。

定義 2.1 (確率変数). \mathcal{F} 可測な函数 $X : \Omega \rightarrow S$ を S 値確率変数と云う。特に $S = \mathbb{R}$ のとき、 X を実確率変数と云う。

定義 2.2 (分布). S 値確率変数 X に対し,

$$\begin{aligned}\mathbb{P}^X(A) &= \mathbb{P}(X^{-1}(A)) \\ &= \mathbb{P}(X \in A), \quad A \in \mathcal{S},\end{aligned}$$

で与えられる S 上の測度 \mathbb{P}^X を X の分布と云う. 定義から明らかに $\mathbb{P}^X \in \mathcal{P}(S)$ である.

例 2.1. $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}([0, 1])$, \mathbb{P} を $[0, 1]$ 上の Lebesgue 測度とすると $(\Omega, \mathcal{F}, \mathbb{P})$ は確率空間である. ただし, $\mathcal{B}([0, 1])$ は $[0, 1]$ 上の Borel σ 加法族である. $X_1(\omega) = \omega$, $X_2(\omega) = 1 - \omega$, $X_3(\omega) = 1 - |2\omega - 1|$ はともに $S = [0, 1]$ 値確率変数である. また, X_1, X_2, X_3 の分布 $\mathbb{P}^{X_1}, \mathbb{P}^{X_2}, \mathbb{P}^{X_3}$ は S 上の Lebesgue 測度となる.

定義 2.3 (同分布). S 値確率変数 X, Y に対し $\mathbb{P}^X = \mathbb{P}^Y$ となるとき, X と Y は同分布であると言いい, $X \stackrel{d}{=} Y$ で表す.

定義 2.4 (一様確率変数). $S = [0, 1]$ のとき, \mathbb{P}^X が S 上の Lebesgue 測度となる S 値確率変数 X を $[0, 1]$ 上に値を取る一様確率変数と云う.

注 2.1. S が有限集合 ($S = \{1, 2, \dots, n\}$) のとき, $\mu \in \mathcal{P}(S)$ は

$$\sum_{x \in S} \mu(x) = 1, \quad \mu(x) \geq 0$$

と表すことができる. $[0, 1]$ 上に値を取る一様確率変数 U に対し, X を

$$X(\omega) = i, \quad \sum_{k=1}^{i-1} \mu(k) \leq U(\omega) < \sum_{k=1}^i \mu(k)$$

と定義することで $\mathbb{P}^X = \mu$ となり, U から分布 μ を持つ S 値確率変数 X を構成することができる.

定義 2.5 (平均, 分散). 確率変数 X の値域を $\text{Ran}X$ で表し, いま $\text{Ran}X$ は高々可算であるとする. このとき,

$$\mathbb{E}[X] = \sum_{k \in \text{Ran}X} k \mathbb{P}(X = k)$$

を X の平均と云い, $f : \text{Ran}X \rightarrow \mathbb{R}$ について,

$$\mathbb{E}[f(X)] = \sum_{k \in \text{Ran}X} f(k) \mathbb{P}(X = k)$$

を $f(X)$ の平均と云う. また,

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

を X の分散と云う.

定義 2.6 (条件付き確率). X, Y は確率変数とし, $\text{Ran}X, \text{Ran}Y$ は高々可算かつ $\mathbb{P}(Y = y) > 0$ であるとする. このとき,

$$\mathbb{P}(X = x \mid Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}$$

を X の Y による条件付き確率と云う.

3 Markov 連鎖

以降断りのない限り S は有限集合とする. また S には順序構造が入っているものとする. 本節では, 時間の経過とともに S 上にある確率法則に従ってランダムに移動する点の運動の内, Markov 連鎖と呼ばれるものの定義を与える. ある時刻において Markov 連鎖が $x \in S$ の位置にいるとき, Markov 連鎖はその時刻において x の状態にある (あるいは x の値を取る) と云い, 集合 S を Markov 連鎖の状態空間と呼ぶ.

例 3.1. $S = \{1, 2, 3, 4\}$ とし, S 上を動く点が図 1 のような確率で各状態間を推移する状況を考える. この図のように, 状態空間での推移の様子を表した図を状態遷移図という. 4×4 行列 P を, 状態 x から状態 y ($x, y \in S$) へ推移する確率を (x, y) 成分に持つ行列とすると,

$$P = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

となる. このようにして定まる行列を推移確率行列と云う.

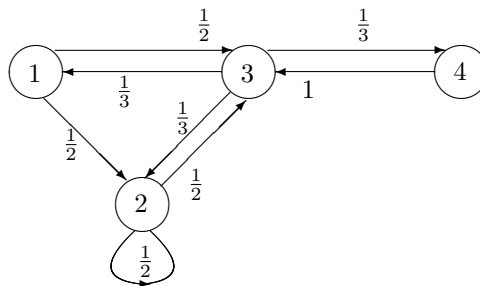


図 1: 状態遷移図

定義 3.1 (確率行列). $|S| \times |S|$ 行列 $P = (P(x, y))_{x, y \in S}$ が,

$$\sum_{y \in S} P(x, y) = 1, \quad \forall x \in S,$$

$$P(x, y) \geq 0, \quad \forall x, y \in S,$$

を満たすとき, 行列 P を確率行列と云う. ただし, $|S|$ (または $\#S$) で S の濃度を表す. また $P(x, \bullet)$ で行列 P の第 x 行目ベクトルを表すことにする. 特に, $P(x, \bullet) \in \mathcal{P}(S)$ である.

注 3.1. 確率行列は必ず固有値 1 をもち, その固有ベクトルは $\mathbf{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ となる.

定義 3.2 (確率過程). S 値確率変数の列 $\{X_t\}_{t=0,1,2,\dots}$ のことを S 値確率過程と云う.

次の定理の証明については補遺参照.

定理 3.1. 任意の $\mu \in \mathcal{P}(S)$ と任意の確率行列 $P = (P(x, y))_{x, y \in S}$ に対して, 次の性質を持つ確率過程 $\{X_t\}_{t=0,1,2,\dots}$ が存在する.

1. X_0 の分布が μ である. つまり $\mathbb{P}^{X_0} = \mu$.
2. 任意の $t \geq 0, x, y, x_0, \dots, x_{t-1} \in S$ に対し, $\mathbb{P}(X_{t+1} = y \mid X_0 = x_0, X_1 = x_1, \dots, X_{t-1} = x_{t-1}, X_t = x) = P(x, y)$.

定義 3.3 (Markov 連鎖). 定理 3.1 で得られる確率過程を初期条件 μ , 推移確率 P の S 値 Markov 連鎖 (あるいは S 上の Markov 連鎖) と云い, (X_t, \mathbb{P}_μ) と書く. 特に $\mu = \delta_x$ のとき x から出発する Markov 連鎖と云い, (X_t, \mathbb{P}_x) と書く. ここで, δ_x は x 成分が 1 で他の成分が 0 であるような行ベクトルである.

例 3.2 (例 3.1 の続き). X_0 の分布を $\mu = (0, 0, 1, 0)$ とする. X_1 の分布は $(\frac{1}{3}, \frac{1}{3}, 0, \frac{1}{3})$ となり, これは μP に等しい. X_2 の分布は $(0, \frac{1}{3}, \frac{2}{3}, 0)$ となり, これは μP^2 に等しい.

一般に次が成り立つ.

命題 3.1. 初期分布を μ , 推移確率を P とする Markov 連鎖の時刻 t での分布は μP^t で与えられる.

例 3.3 (Simple Random Walk on a Graph). $G = (V, E)$ を有限グラフ, $S = V$ とする.

$$\begin{aligned} \deg(x) &= \text{点 } x \text{ から出る辺の本数} \\ &= |N_x| \end{aligned}$$

とする. ただし $N_x = \{y \in V : y \text{ は } x \text{ の隣接点}\}$ とする. このとき,

$$P(x, y) = \begin{cases} \frac{1}{\deg(x)}, & y \in N_x, \\ 0, & y \notin N_x, \end{cases}$$

と定めると $P = (P(x, y))$ は確率行列となる. この P を推移確率とする Markov 連鎖を, このグラフ上の simple random walk (SRW) と云う.

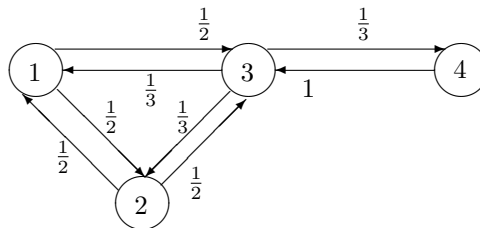


図 2: $|S| = 4$ のときのグラフの例とその上の SRW

グラフが複雑な場合, 確率行列を与えるよりも, ある時刻での Markov 連鎖の状態を次にどのように推移させるかそのルールを具体的に書き下した方が分かりやすい場合がある. そのような例を二つ挙げよう.

例 3.4 (Simple Random Walk on Hypercube). $S = \{0, 1\}^n$ とし, $\rho : S \times S \rightarrow \mathbb{R}$ を次で定義する:

$$x = (x_1, x_2, \dots, x_n), y = (y_1, y_2, \dots, y_n) \in S \text{ に対し, } \rho(x, y) = \#\{1 \leq i \leq n : x_i \neq y_i\}.$$

ρ は S 上の距離となり, Hamming 距離と呼ばれる. 例えば $n = 4$ のとき, $\rho((0, 1, 1, 1), (1, 0, 1, 1)) = 2$ である. $x, y \in S$ に対し $\rho(x, y) = 1$ のとき, $xy \in E$ (辺の集合) と定義することで, (S, ρ) はグラフとみなせる. このグラフを hypercube と云う. S 上を動く点のある時刻での状態を $x = (x_1, x_2, \dots, x_n)$ とするとき, 次の時刻での状態への推移を以下のルールで与える.

1. $1/n$ の確率で第 i 座標を選ぶ.
2. 選んだ第 i 座標に対し, $x_i \mapsto 1 - x_i$ とする.

すると, このルールにより定まる S 上を動く点の運動は, hypercube 上の SRW となる.

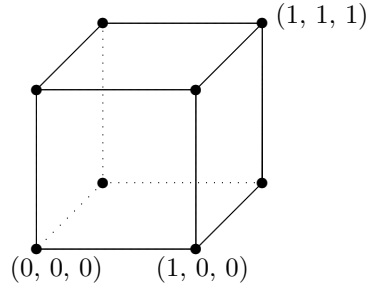


図 3: $n = 3$ のときの hypercube

例 3.5 (Markov Chain on q -coloring). $G = (V, E)$ を有限連結グラフ, q を色の数とする. $x, y \in V$ に対し, x と y が隣接している, すなわち $xy \in E$ となるとき $x \sim y$ と書くことにし, $S = \{c : V \rightarrow [q] = \{1, 2, \dots, q\} : c(x) \neq c(y), x \sim y\}$ とする. また $c \in S$ のとき, グラフの彩色 c が proper であると云う. このとき以下のルールでグラフの彩色を推移させる.

1. $c \in S$ とし, V から確率 $1/|V|$ で点 v を選ぶ.
2. $\mathcal{A}_v(c) = [q] \setminus \{c(x) \in [q] : x \in N_v\}$ (v に彩色可能な色の集合) とし, この集合から v に置く色を確率 $1/|\mathcal{A}_v(c)|$ で一様に選んで更新する.

すると, この推移で得られた新たな彩色 \tilde{c} についても $\tilde{c} \in S$ となる. このルールにより得られる G の proper な彩色の状態の推移は S 上の Markov 連鎖となる.

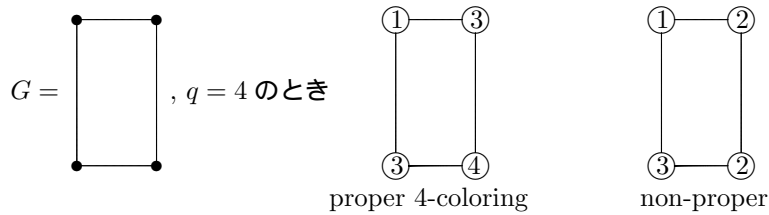


図 4: proper と non-proper

本節の最後に Markov 連鎖に関する重要な等式を述べる.

定理 3.2 (Chapman-Kolmogorov の等式). (X_t, \mathbb{P}_x) を Markov 連鎖とする. このとき, 以下の等式が成立する.

$$\mathbb{P}_x(X_{t+s} = y) = \sum_{z \in S} \mathbb{P}_x(X_t = z) \mathbb{P}_z(X_s = y), \quad \forall x, y \in S, \forall t \geq 0.$$

この等式を Chapman-Kolmogorov の等式と呼ぶ.

証明. 任意の $x, y \in S$ と任意の $t \geq 0$ について, $\mathbb{P}_x(X_t = y) = (\delta_x P^t)(y)$ と表されることに注意すれば,

$$\begin{aligned} \mathbb{P}_x(X_{t+s} = y) &= (\delta_x P^{t+s})(y) \\ &= ((\delta_x P^t)P^s)(y) \\ &= \sum_{z \in S} (\delta_x P^t)(z) P^s(z, y) \\ &= \sum_{z \in S} \mathbb{P}_x(X_t = z) \mathbb{P}_z(X_s = y). \end{aligned}$$

よって結論を得る. □

4 Markov 連鎖に関する基本的な概念

本節では, 定常分布や既約性といった, Markov 連鎖に関する基本的な概念について解説する.

4.1 定常分布と可逆分布

推移確率 $P = (P(x, y))$ を持つ S 上の Markov 連鎖 $\{X_t\}_{t=0,1,2,\dots}$ が与えられているものとする.

定義 4.1 (定常分布). $\pi = (\pi(x)) \in \mathcal{P}(S)$ が $\pi P = \pi$ すなわち, 全ての $x \in S$ に対し,

$$\sum_{y \in S} \pi(y) P(y, x) = \pi(x)$$

を満たすとき, π は定常分布であると言う.

定義 4.2 (可逆分布). $\pi = (\pi(x)) \in \mathcal{P}(S)$ が全ての x, y に対し,

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

を満たすとき, π は可逆分布であると言う.

注 4.1. 1. π が定常分布ならば, 全ての $t = 0, 1, 2, \dots$ に対し $\pi P^t = \pi$ である. すなわち $\mathbb{P}_\pi(X_t = x) = \pi(x)$ となる.

2. π が可逆分布ならば定常分布である. なぜなら π が可逆とすると, 任意の $y \in S$ に対し,

$$\sum_{x \in S} \pi(x) P(x, y) = \sum_{x \in S} \pi(y) P(y, x) = \pi(y)$$

だからである.

3. P が対称行列, すなわち任意の $x, y \in S$ に対し, $P(x, y) = P(y, x)$ であるとき, $\pi(x) = \frac{1}{|S|}$ は可逆分布である.

例 4.1 (例 3.3 の続き). グラフ $G = (V, E)$ 上の SRW の推移確率

$$P(x, y) = \begin{cases} \frac{1}{\deg(x)}, & y \in N_x, \\ 0, & y \notin N_x, \end{cases}$$

について, $\pi(x) = \frac{\deg(x)}{2|E|}$ は可逆分布である.

4.2 既約性と周期性

定義 4.3 (既約性). 推移確率 $P = (P(x, y))$ を持つ S 上の Markov 連鎖が, 互いに全ての点に行き渡り得る, すなわち, 任意の $x, y \in S$ に対しある $t > 0$ が存在して, $P^t(x, y) > 0$ となるとき, Markov 連鎖は既約であると言う.

例 4.2. 図 5 のような状態遷移図で与えられる $S = \{1, 2, 3\}$ 上の Markov 連鎖を考える. 3 から 1 にはいつまで経っても行けないので, この Markov 連鎖は既約でない.

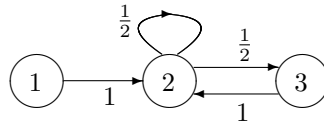


図 5: 既約でない Markov 連鎖

例 4.3 (Random Knight Move). S をチェスボード (8×8 のマス目状) とし, ナイトの駒を動かすという Markov 連鎖を考える. ナイトは前(後)へ 2 マス, 横へ 1 マスの位置または, 横へ 2 マス, 前(後)へ 1 マスの位置に移動できる. ナイトはいずれはどのマス目にも移動できるのでこの Markov 連鎖は既約である (おおよそ 500 回で全てのマス目を訪問できることが知られている).

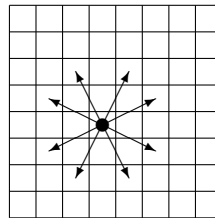


図 6: チェスボード

定義 4.4 (周期). $x \in S$ に対して, $\text{Per}(x) = \{t \geq 1 : P^t(x, x) > 0\}$ とし, $d_x = \text{G.C.D}(\text{Per}(x))$ (G.C.D は最大公約数を表す) を点 x の周期と言う.

注 4.2. Markov 連鎖が既約ならば d_x は x に依らないということが知られている. 特に $d_x \equiv 1$ であるとき, 非周期的と言う.

例 4.4 (Simple Random Walk on \mathbb{Z}_n). $\mathbb{Z}_n = \{0, 1, \dots, n-1\}$ (modulo n) 上の SRW を考える.

- n が奇数のときは各 $i \in \mathbb{Z}_n$ に対し, $i \rightarrow i+1 \rightarrow i$ のように偶数回の移動で戻ってくることも, $i \rightarrow i+1 \rightarrow \dots \rightarrow i-1 \rightarrow i$ と, n 回で戻ってくることもできるため周期は 1, つまり非周期的である.
- n が偶数のときは元の点に戻るには必ず偶数回移動 (最低でも 2 回移動) しないと行けないので, 周期は 2 である.

例 4.5. 図 8 のグラフ $G = (V, E)$ 上の SRW を考える. $V = \{\text{正六角形の頂点と重心}\}$ であり, 任意の $x \in V$ に対し, 周期 d_x は 3 である.

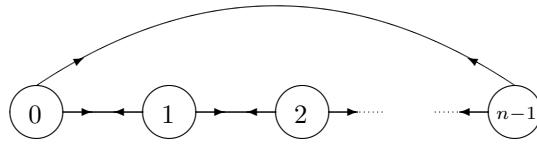


図 7: \mathbb{Z}_n 上の Markov 連鎖

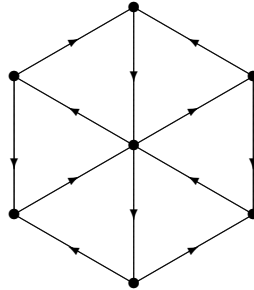


図 8: 正六角形の頂点と重心を点とするグラフ

例 4.4 のような場合, n の偶奇に関わらず, Markov 連鎖の時間無限大における挙動には大差がないことが知られている. そこで偶奇の場合分けの煩わしさを除くためにしばしば次のような lazy version の Markov 連鎖を代わりに用いることがある.

定義 4.5 (Lazy Version Markov Chain). グラフ $G = (V, E)$ 上の推移確率 P を持つ Markov 連鎖があるとする. 各 $x \in V$ にループ (x から x へ行けるようにする辺) を加え, 各点ループする確率を $\frac{1}{2}$, 他の点に推移する確率をループを加える前の $\frac{1}{2}$ 倍とする. このようにして得られる新たな Markov 連鎖を元の Markov 連鎖の lazy version と云う. すなわちその推移確率を P_{lazy} とすると,

$$P_{\text{lazy}} = \frac{1}{2}(I + P) \quad (I \text{ は } |V| \text{ 次単位行列})$$

である. なお P の定常分布を π とすると,

$$\begin{aligned} \pi P_{\text{lazy}} &= \frac{1}{2}(\pi + \pi P) \\ &= \pi \end{aligned}$$

なので P_{lazy} の定常分布も π となる.

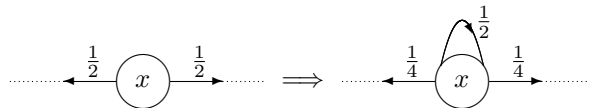


図 9: Lazy version

注 4.3. 上記の例 4.4 の場合, n の偶奇に関わらず lazy version の Markov 連鎖は非周期的となる.

4.3 定常分布の存在

ここではどのような場合に定常分布が存在するかについて紹介する。いま初期分布 μ について μP^t が $t \rightarrow \infty$ としてある分布 π に収束したと仮定する。このとき、

$$\pi = \lim_{t \rightarrow \infty} \mu P^{t+1} = \lim_{t \rightarrow \infty} (\mu P^t) P = \pi P$$

となる。従って、上で仮定した極限が存在するならば P には定常分布が存在することが分かる。

定義 4.6 (到達時刻). (X_t, \mathbb{P}_x) を状態空間 S 上の Markov 連鎖とする。確率変数 τ_x を以下で定義する。

$$\tau_x = \begin{cases} \inf\{t \geq 1 : X_t = x\}, & \{\dots\} \neq \emptyset, \\ \infty, & \{\dots\} = \emptyset. \end{cases}$$

すなわち、 τ_x とは始点が x 以外の場合は初めて x に到達する時刻であり、始点が x の場合は、最初に x に帰って来る時刻である。この τ_x を到達時刻と云う。

定義 4.7. $\mathbb{P}_x(\tau_x < \infty) = 1$ となるとき、 $x \in S$ は再帰的であると云う。

命題 4.1. Markov 連鎖は既約かつ非周期的であるとすると、このとき任意の $x, y \in S$ に対し、

$$\mathbb{E}_x[\tau_y] < \infty$$

が成り立つ。特に S の任意の点は再帰的である。

次の定理については、例えば [6, Theorem 1.8.2] 参照。

定理 4.1. $x \in S$ を再帰的であるとして

$$\mu_x(y) = \mathbb{E}_x \left[\sum_{t=0}^{\tau_x-1} 1_{\{y\}}(X_t) \right]$$

と定義する。ただし 1_A は $A \subset S$ の定義関数を表す。このとき、 $\pi(y) = \mu_x(y) / \mathbb{E}_x[\tau_x]$ と定義すると $\pi(y)$ は定常分布となる。

注 4.4. $\sum_{t=0}^{\tau_x-1} 1_{\{y\}}(X_t)$ は y での滞在時間 (occupation time) と呼ばれ、Markov 連鎖 (X_t, \mathbb{P}_x) が時刻 0 から τ_x-1 までに y に何回到達したかをカウントする確率変数である。

注 4.5. 定常分布についてはその一意性も含め第 7 節で再び議論する。

5 周辺分布とカップリング

$\mu \in \mathcal{P}(S \times S)$ に対して

$$\begin{aligned} \mu_1(x) &= \sum_{y \in S} \mu(x, y), & x \in S, \\ \mu_2(y) &= \sum_{x \in S} \mu(x, y), & y \in S, \end{aligned}$$

と定義すると、 $\mu_1, \mu_2 \in \mathcal{P}(S)$ となる。この μ_1 と μ_2 を μ の周辺分布と呼ぶ (もしくは、第 1, 第 2 座標への μ の射影と呼ぶ)。

逆に $\mu_1, \mu_2 \in \mathcal{P}(S)$ に対して、周辺分布が μ_1, μ_2 となる $\mu \in \mathcal{P}(S \times S)$ は常に存在するだろうかという問題が生じる。その答えは「少なくとも 1 つ存在する」である。それは次に例としてあげる直積測度である。

例 5.1 (直積測度 $\mu_1 \otimes \mu_2$).

$$(\mu_1 \otimes \mu_2)(x, y) = \mu_1(x)\mu_2(y), \quad x, y \in S.$$

しかし、上の例のような直積測度以外にも周辺分布となる組み合わせは存在し得る。そこで、いくつかあるものの中で目的に応じて最適なものを選ぶことになる。

定義 5.1. $\mu_1, \mu_2 \in \mathcal{P}(S)$ に対して、ある確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ 上に定義された確率変数 (X, Y) の組で $\mathbb{P}^X = \mu_1, \mathbb{P}^Y = \mu_2$ (これを $X \sim \mu_1, Y \sim \mu_2$ と書く) を満たすものを (μ_1, μ_2) のカップリングと呼ぶ。

定義 5.2. 確率変数 X と Y が独立であるとは、 $\{X = x\}$ と $\{Y = y\}$ が同時に起こる確率が、各々の起こる確率の積であることである。すなわち、任意の $x, y \in S$ に対して、 $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$ となることを云う。

例 5.2 (公平なコイン投げ). $S = \{0, 1\}$ とし、 S 上の確率測度 μ を $\mu(0) = \mu(1) = \frac{1}{2}$ と定める。 (X, Y) を (μ, μ) のカップリングとし (すなわち $X \sim \mu, Y \sim \mu$),

$$\mu(i, j) = \mathbb{P}(X = i, Y = j), \quad i, j = 0, 1,$$

と定義する。これを

$$\begin{bmatrix} \mu(0, 0) & \mu(0, 1) \\ \mu(1, 0) & \mu(1, 1) \end{bmatrix}$$

と行列表示するとき、次のようなカップリングを考えることができる。ただし以下では二人のプレイヤー A, B がいるとして、ある公平なコイン投げの試行の結果 ($\frac{1}{2}$ の確率で表か裏が出る) に応じて A は X の値を、B は Y の値を 0 か 1 のいずれかにそれぞれ定めるものとする。

1. $\begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}$: A, B が別々にコインを投げ、それぞれの結果に応じて各々 X, Y の値を定める。二人のコイン投げの試行は互いに独立なので、得られるカップリングは直積測度に対応するものとなる。
2. $\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$: C という第三者によるコイン投げの結果に応じて A, B は X, Y の値を $X = Y$ となるように定める。
3. $\begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix}$: C によるコイン投げの結果に応じて A, B は X, Y の値を $Y = 1 - X$ となるように定める。

このように定めた (X, Y) は実際 (μ, μ) のカップリングとなっている。そのことを確かめるには次のように考えればよい。例えばプレイヤー B の存在を忘れて A による X の値の定め方にのみ注目すれば、上のいずれの場合も、一つのコイン投げの試行の結果に応じて A は X の値を 0 か 1 にそれぞれ $\frac{1}{2}$ の確率で定めているだけであるので、 $X \sim \mu$ である。また逆に A の存在を忘れて B へのみ注目すれば、同様に $Y \sim \mu$ となることが見てとれる。

6 全変動距離 (total variation) とカップリング

定義 6.1. μ と ν を S 上の確率測度とする. このとき, 以下のように全変動距離を定義する.

$$\|\mu - \nu\|_{TV} = \max_{A \subseteq S} |\mu(A) - \nu(A)|.$$

ただし, $\mu(A) = \sum_{x \in A} \mu(x)$ と定義する, $\nu(A)$ についても同様である.

注 6.1. $0 \leq \|\bullet\|_{TV} \leq 1$ が成立する.

定理 6.1. 上で定義した全変動距離に対して, 以下の等式が成り立つ

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in S} |\mu(x) - \nu(x)| \quad (6.1)$$

$$= \sum_{x \in S, \mu(x) > \nu(x)} (\mu(x) - \nu(x)) \quad (6.2)$$

$$= \inf\{\mathbb{P}(X \neq Y) : (X, Y) \text{ は } (\mu, \nu) \text{ のカップリング}\}. \quad (6.3)$$

証明. S の部分集合 B を $B = \{x \in S : \mu(x) > \nu(x)\}$ とおく. すると, S の任意の部分集合 A に対して

$$\begin{aligned} \mu(A) - \nu(A) &= \mu(A \cap B) - \nu(A \cap B) \\ &\quad + \mu(A \cap B^c) - \nu(A \cap B^c) \end{aligned}$$

となる. ここで, $\mu(A \cap B) - \nu(A \cap B) \geq 0$ かつ $\mu(A \cap B^c) - \nu(A \cap B^c) \leq 0$ であることから

$$\begin{aligned} \mu(A) - \nu(A) &\leq \mu(A \cap B) - \nu(A \cap B) \\ &\leq \mu(B) - \nu(B). \end{aligned}$$

μ と ν の役割を入れ替えると, 同様にして

$$\nu(A) - \mu(A) \leq \nu(B) - \mu(B).$$

ゆえに

$$|\mu(A) - \nu(A)| \leq \mu(B) - \nu(B)$$

である. ここで, A について最大値をとると,

$$\|\mu - \nu\|_{TV} = \mu(B) - \nu(B)$$

となり, (6.2) が成り立つ. また, $\mu(B) - \nu(B) = \nu(B^c) - \mu(B^c)$ より (6.1) が成り立つ. 次に (X, Y) を (μ, ν) の任意のカップリングとする. A を S の任意の部分集合として,

$$\begin{aligned} \mu(A) - \nu(A) &= \mathbb{P}^X(A) - \mathbb{P}^Y(A) \\ &= \mathbb{P}(X \in A) - \mathbb{P}(Y \in A) \\ &= \mathbb{P}(X \in A, Y \in A^c) + \mathbb{P}(X \in A, Y \in A) - \mathbb{P}(Y \in A) \\ &\leq \mathbb{P}(X \in A, Y \in A^c) \\ &\leq \mathbb{P}(X \neq Y). \end{aligned}$$

従って, $\max_{A \subset S} |\mu(A) - \nu(A)| \leq P(X \neq Y)$ となる. これにより,

$$\|\mu - \nu\|_{TV} \leq \inf\{P(X \neq Y) : X \sim \mu, Y \sim \nu\}$$

が成立する. 従ってあとは

$$\|\mu - \nu\|_{TV} = \mathbb{P}(X \neq Y) \tag{6.4}$$

を満たすカップリング (X, Y) が構成できればよい.

$$p = \sum_{x \in S} \mu(x) \wedge \nu(x)$$

と定義する. (6.2) から $\|\mu - \nu\|_{TV} = 1 - p$ となることに注意する. 更に, S 上の確率測度 $\gamma_1, \gamma_2, \gamma_3$ を以下のように構成する.

$$\gamma_1(x) = \begin{cases} \frac{\mu(x) - \nu(x)}{1 - p}, & \mu(x) > \nu(x), \\ 0, & \text{上記以外,} \end{cases}$$

$$\gamma_2(x) = \begin{cases} \frac{\nu(x) - \mu(x)}{1 - p}, & \mu(x) \leq \nu(x), \\ 0, & \text{上記以外,} \end{cases}$$

$$\gamma_3(x) = \frac{\mu(x) \wedge \nu(x)}{p}.$$

カップリングの構成

確率 p で表が出るコインを投げるとする. 表が出た場合, $X = Y \sim \gamma_3$ とする. 裏が出た場合は, X, Y を独立で $X \sim \gamma_1, Y \sim \gamma_2$ となるように定める. このとき,

$$\begin{aligned} \mathbb{P}(X = x) &= p\gamma_3(x) + (1 - p)\gamma_1(x) \\ &= \mu(x) \end{aligned}$$

である. 同様にして, $\mathbb{P}(Y = y) = \nu(y)$ である. また定義から $\mathbb{P}(X \neq Y) = 1 - p = \|\mu - \nu\|_{TV}$ である. よって, (6.4) を満たすカップリングが構成できた. \square

定義 6.2. (6.3) の \inf の値を達成するカップリングを最適カップリングと云う.

7 定常分布への収束

本節では, Markov 連鎖の定常分布への収束について, 定常分布の一意性も含めて議論する.

補題 7.1. 推移確率 P を持つ Markov 連鎖が既約かつ非周期的とする. このとき, ある $r > 1$ が存在し, 全ての $x, y \in S$ と $t \geq r$ に対して $P^t(x, y) > 0$ が成立する.

定理 7.1. 既約かつ非周期的な Markov 連鎖について, 定数 r を上の補題により与えられるものとす. また π を推移確率 P の定常分布とする. このとき, ある定数 $0 < \alpha < 1$ が存在して,

$$\max_{x \in S} \|P^t(x, \bullet) - \pi\|_{TV} \leq \alpha^{t-r} \wedge 1, \quad \forall t \geq r,$$

が成り立つ.

証明. 補題の r をとると,

$$P^r(x, y) \geq \delta\pi(y)$$

を満たす δ が存在する. この δ を用いて行列 Q を

$$P^r = \delta\Pi + (1 - \delta)Q$$

によって定義する. ただし $\Pi(x, y) = \pi(y)$ である. このとき Q は確率行列である. いま,

$$P\Pi = \Pi P = \Pi, Q\Pi = \Pi, \Pi^2 = \Pi$$

となることに注意する.

$$\begin{aligned} ((1 - \delta)Q)^k &= (P^r - \delta\Pi)^k \\ &= P^{rk} + \sum_{j=1}^k \binom{k}{j} (P^r)^j (-\delta\Pi)^{k-j} \\ &= P^{rk} + \sum_{j=1}^k \binom{k}{j} (-\delta)^{k-j} \Pi \\ &= P^{rk} + ((1 - \delta)^k - 1)\Pi. \end{aligned}$$

よって

$$P^{rk} - \Pi = (1 - \delta)^k(Q^k - \Pi).$$

両辺に P^j をかけると ($0 \leq j \leq r - 1$)

$$P^{rk+j} - \Pi = (1 - \delta)^k(Q^k P^j - \Pi).$$

x 行の L^1 ノルムの $1/2$ をとると (6.1) から

$$\begin{aligned} \|P^{rk+j}(x, \bullet) - \pi\|_{TV} &= (1 - \delta)^k \|Q^k P^j(x, \bullet) - \pi\|_{TV} \\ &\leq (1 - \delta)^k. \end{aligned}$$

ゆえに定理の主張が従う. □

系 7.1. Markov 連鎖が既約ならば定常分布は存在して一意である (非周期的でない場合は lazy version を考えることにより同様に議論できる).

8 混合時間

S を有限集合とする. (X_t, \mathbb{P}_x) を S 上の既約, 非周期的な Markov 連鎖とする. このとき, 前節でみたように, 対応する推移確率 $P = (P(x, y))_{x, y \in S}$ について

$$\lim_{t \rightarrow \infty} P^t(x, y) = \pi(x), \quad \forall x, y \in S$$

が成り立つ.

定義 8.1. $t = 0, 1, 2, \dots$ について

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

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

とおく.

注 8.1. 1. $0 \leq \|\bullet\|_{TV} \leq 1$ より $0 \leq d(t), \bar{d}(t) \leq 1$ である. また, これらは t について単調非増加である (補題 8.3).

2. $P^t(x, \bullet)$ は x から出発した Markov 連鎖の時刻 t での分布なので, $d(t)$ は定常分布 π へ時刻 t でどれくらい近付いているかを表す.

定義 8.2 (混合時間). $0 < \epsilon \leq 1$ に対して,

$$t_{\text{mix}}(\epsilon) = \inf\{t \geq 0 : d(t) \leq \epsilon\},$$

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

とおく. この t_{mix} を混合時間 (mixing time) と云う.

注 8.2. 1. n 次対称群上の Markov 連鎖の混合時間を直感的に捉えると, 一組の枚数が n 枚であるようなトランプのカードをシャッフルする際に, “十分よく混ぜる” までに必要なシャッフルの回数である.

2. t_{mix} を定義する $\epsilon = 1/4$ には本質的な意味はなく, 命題 8.2 で見るように, $0 < \epsilon < 1/2$ なる ϵ ならばどれでも良い.

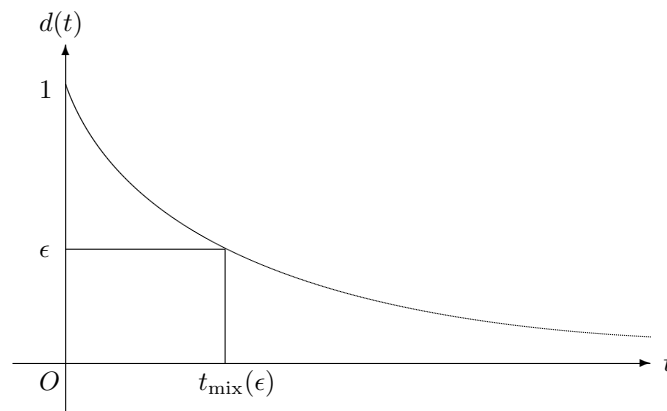


図 10: $d(t)$ と $t_{\text{mix}}(\epsilon)$

問題. $|S_n|$ が n とともに増大する状態空間 S_n 上の Markov 連鎖の混合時間を $t_{\text{mix}}^{(n)}$ で表す. $n \rightarrow \infty$ としたとき $t_{\text{mix}}^{(n)}$ はどのような振る舞いをするだろうか?

例 8.1. n 次元 hypercube $S_n = \{0, 1\}^n$ 上の lazy simple random walk ¹ では $t_{\text{mix}}^{(n)} = O(n \log n)$ となる. さらに, 十分大きな n のとき, n に依存したある時刻までは, $d(t)$ はほとんど変わらないが, その時刻を超えると $d(t)$ が急激に減少する. このような現象を cut-off 現象と云う.

¹p.20 例 9.1 で定義を与える.

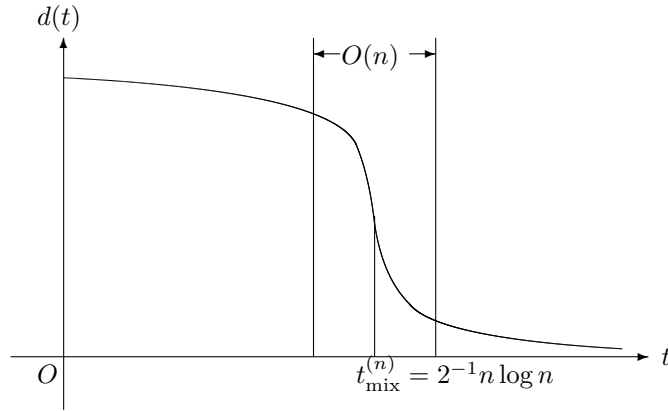


図 11: cut-off 現象の様子

補題 8.1. d, \bar{d} について次が成り立つ.

$$d(t) \leq \bar{d}(t) \leq 2d(t), \quad \forall t \geq 0.$$

証明. 推移確率行列 $P = (P(x, y))_{x, y \in S}$ の行ベクトル $P^t(x, \bullet)$ は $\mathcal{P}(S)$ の元であり, 定常分布 π は $\pi = \pi P^t$ を満たすことに注意すれば, $A \subset S$ に対して,

$$\begin{aligned} \pi(A) &= \sum_{z \in A} (\pi P^t)(z) \\ &= \sum_{z \in A} \sum_{y \in S} \pi(y) P^t(y, z) = \sum_{y \in S} \pi(y) \sum_{z \in A} P^t(y, z) \\ &= \sum_{y \in S} \pi(y) P^t(y, A) \end{aligned}$$

が成り立つことが分かる. これより,

$$\begin{aligned} |P^t(x, A) - \pi(A)| &= \left| P^t(x, A) - \sum_{y \in S} \pi(y) P^t(y, A) \right| \\ &= \left| \sum_{y \in S} \pi(y) P^t(x, A) - \sum_{y \in S} \pi(y) P^t(y, A) \right| \\ &\leq \sum_{y \in S} \pi(y) |P^t(x, A) - P^t(y, A)| \end{aligned}$$

となる. 従って, 全変動距離の定義より,

$$\begin{aligned} \|P^t(x, \bullet) - \pi\|_{TV} &= \max_{A \subset S} |P^t(x, A) - \pi(A)| \\ &\leq \sum_{y \in S} \pi(y) \max_{A \subset S} |P^t(x, A) - P^t(y, A)| \\ &= \sum_{y \in S} \pi(y) \|P^t(x, \bullet) - P^t(y, \bullet)\|_{TV} \\ &\leq \sum_{y \in S} \pi(y) \times \bar{d}(t) = \bar{d}(t) \end{aligned}$$

が得られる. 左辺を $x \in S$ について最大化すれば最初の不等式が得られる.

2 番目の不等式は

$$\|P^t(x, \bullet) - P^t(y, \bullet)\|_{TV} \leq \|P^t(x, \bullet) - \pi\|_{TV} + \|P^t(y, \bullet) - \pi\|_{TV}$$

を $x, y \in S$ について最大化することで得られる. □

補題 8.2. d, \bar{d} について次の表現が成り立つ.

$$d(t) = \sup_{\mu \in \mathcal{P}(S)} \|\mu P^t - \pi\|_{TV}, \quad (8.1)$$

$$\bar{d}(t) = \sup_{\mu, \nu \in \mathcal{P}(S)} \|\mu P^t - \nu P^t\|_{TV}. \quad (8.2)$$

証明. (8.1) について. 第 x 成分のみが 1 である行ベクトルを δ_x とおくと, $\delta_x \in \mathcal{P}(S)$ であり, $P^t(x, \bullet) = \delta_x P^t$ が成り立つので,

$$\begin{aligned} \|P^t(x, \bullet) - \pi\|_{TV} &= \|\delta_x P^t - \pi\|_{TV} \\ &\leq \sup_{\mu \in \mathcal{P}(S)} \|\mu P^t - \pi\|_{TV} \end{aligned}$$

が得られる. $x \in S$ について最大化すれば

$$d(t) \leq \sup_{\mu \in \mathcal{P}(S)} \|\mu P^t - \pi\|_{TV}$$

が従う. 次に逆側の不等式を示す. まず全変動距離の表現定理 6.1 の (6.1) により,

$$\begin{aligned} \|\mu P^t - \pi\|_{TV} &= \frac{1}{2} \sum_{y \in S} |(\mu P^t)(y) - \pi(y)| \\ &= \frac{1}{2} \sum_{y \in S} \left| \sum_{x \in S} \mu(x) P^t(x, y) - \sum_{x \in S} \mu(x) \pi(y) \right| \\ &\leq \frac{1}{2} \sum_{y \in S} \sum_{x \in S} \mu(x) |P^t(x, y) - \pi(y)| \end{aligned}$$

となる. 最後の式で y について和をとることで

$$\begin{aligned} \|\mu P^t - \pi\|_{TV} &\leq \sum_{x \in S} \|P^t(x, \bullet) - \pi\|_{TV} \mu(x) \\ &\leq \max_{x \in S} \|P^t(x, \bullet) - \pi\|_{TV} \sum_{x \in S} \mu(x) \\ &= d(t) \end{aligned}$$

となる. $\mu \in \mathcal{P}(S)$ について最大化することで

$$d(t) \geq \sup_{\mu \in \mathcal{P}(S)} \|\mu P^t - \pi\|_{TV}$$

が従う. よって第 1 の等式が示された. 第 2 の等式 (8.2) も同様に示される. □

補題 8.3. $\mu, \nu \in \mathcal{P}(S)$ に対して

$$\|\mu P - \nu P\|_{TV} \leq \|\mu - \nu\|_{TV}$$

が成り立つ. 特に $d(t), \bar{d}(t)$ は単調非増加である.

証明. 全変動距離の表現定理 6.1 の (6.1) を用いて,

$$\begin{aligned}
\|\mu P - \nu P\|_{TV} &= \frac{1}{2} \sum_{y \in S} |(\mu P)(y) - (\nu P)(y)| \\
&= \frac{1}{2} \sum_{y \in S} \left| \sum_{x \in S} \mu(x) P(x, y) - \sum_{x \in S} \nu(x) P(x, y) \right| \\
&\leq \frac{1}{2} \sum_{y \in S} \sum_{x \in S} |\mu(x) - \nu(x)| P(x, y) \\
&= \frac{1}{2} \sum_{x \in S} |\mu(x) - \nu(x)| \\
&= \|\mu - \nu\|_{TV}
\end{aligned}$$

となる. そして

$$\begin{aligned}
d(t+1) &= \sup_{\mu \in \mathcal{P}(S)} \|\mu P^{t+1} - \pi\|_{TV} \\
&= \sup_{\mu \in \mathcal{P}(S)} \|(\mu P^t)P - \pi P\|_{TV} \\
&\leq \sup_{\mu \in \mathcal{P}(S)} \|\mu P^t - \pi\| \\
&= d(t)
\end{aligned}$$

となるので $d(t)$ の単調非増加性も分かる. $\bar{d}(t)$ の単調非増加性も同様である. \square

命題 8.1. \bar{d} は劣乗法的である. つまり,

$$\bar{d}(s+t) \leq \bar{d}(s)\bar{d}(t), \quad s, t \geq 0$$

が成り立つ.

証明. $P^s(x, \bullet), P^s(y, \bullet) \in \mathcal{P}(S)$ の最適カップリングを (X_s, Y_s) とする. Chapman-Kolmogorov の等式 (定理 3.2) より,

$$\begin{aligned}
P^{s+t}(x, w) &= \sum_{z \in S} P^s(x, z) P^t(z, w) \\
&= \sum_{z \in S} \mathbb{P}(X_s = z) P^t(z, w) \\
&= \mathbb{E}[P^t(X_s, w)].
\end{aligned}$$

同様にして $P^{s+t}(y, w) = \mathbb{E}[P^t(Y_s, w)]$ も得られる. これより

$$\begin{aligned}
\|P^{s+t}(x, \bullet) - P^{s+t}(y, \bullet)\|_{TV} &= \frac{1}{2} \sum_{w \in S} |P^{s+t}(x, w) - P^{s+t}(y, w)| \\
&= \frac{1}{2} \sum_{w \in S} |\mathbb{E}[P^t(X_s, w) - P^t(Y_s, w)]| \\
&\leq \frac{1}{2} \sum_{w \in S} \mathbb{E}[|P^t(X_s, w) - P^t(Y_s, w)|]
\end{aligned}$$

$$= \mathbb{E} \left[\frac{1}{2} \sum_{w \in S} |P^t(X_s, w) - P^t(Y_s, w)|; X_s \neq Y_s \right]$$

となる. ここで, $X_s, Y_s \in S$ より,

$$\begin{aligned} \frac{1}{2} \sum_{w \in S} |P^t(X_s, w) - P^t(Y_s, w)| &\leq \max_{x, y \in S} \left[\frac{1}{2} \sum_{w \in S} |P^t(x, w) - P^t(y, w)| \right] \\ &= \max_{x, y \in S} \|P^t(x, \bullet) - P^t(y, \bullet)\|_{TV} \\ &= \bar{d}(t) \end{aligned}$$

となるので,

$$\begin{aligned} \|P^{s+t}(x, \bullet) - P^{s+t}(y, \bullet)\|_{TV} &\leq \bar{d}(t) \mathbb{P}(X_s \neq Y_s) \\ &= \bar{d}(t) \|P^s(x, \bullet) - P^s(y, \bullet)\|_{TV} \end{aligned}$$

が得られる. 等号は (X_s, Y_s) が最適カップリングであることによる. 両辺を $x, y \in S$ について最大化することで結論が従う. \square

命題 8.2. $0 < \epsilon < 1/2$ のとき,

$$t_{\text{mix}}(\epsilon) \leq \lceil \log_2 \epsilon^{-1} \rceil t_{\text{mix}}$$

が成り立つ. ここで $\lceil a \rceil$ は a 以上の最小の整数を表す.

証明. c を正の数とし, $\lfloor c \rfloor$ で c 以下の最大の整数を表す. すると, $\bar{d}(t)$ は単調非増加で劣乗法的であり, $d(t) \leq \bar{d}(t) \leq 2d(t)$ が成り立つので,

$$\begin{aligned} d(ct_{\text{mix}}) &\leq \bar{d}(ct_{\text{mix}}) \\ &\leq \bar{d}(\lfloor c \rfloor t_{\text{mix}}) \\ &\leq (\bar{d}(t_{\text{mix}}))^{\lfloor c \rfloor} \\ &\leq (2d(t_{\text{mix}}))^{\lfloor c \rfloor} \\ &\leq (2 \cdot 4^{-1})^{\lfloor c \rfloor} = 2^{-\lfloor c \rfloor} \end{aligned}$$

となる. これより, $t_{\text{mix}}(2^{-\lfloor c \rfloor}) \leq ct_{\text{mix}}$ が得られる. $c = \log_2 \epsilon^{-1}$ とすれば結論が従う. \square

9 カップリング

S 値確率変数 (X, Y) のカップリング (\tilde{X}, \tilde{Y}) , つまり $\tilde{X} \stackrel{d}{=} X, \tilde{Y} \stackrel{d}{=} Y$ となるもので, “付加的な良い性質を持つもの” を構成することを考える. 例として次のような問題を挙げる.

問題. $S = \{0, 1, 2, \dots, n\}$ とし, S 上を動く粒子の運動を次のルールで定める.

- 状態 0 の粒子は確率 $1/2$ で状態 0 に留まるか, 状態 1 へ動く.
- 状態 k ($1 \leq k \leq n-1$) の粒子は確率 $1/2$ で状態 $(k-1)$ か状態 $(k+1)$ へ動く.
- 状態 n の粒子は確率 $1/2$ で状態 $n-1$ へ動くか, 状態 n に留まる.

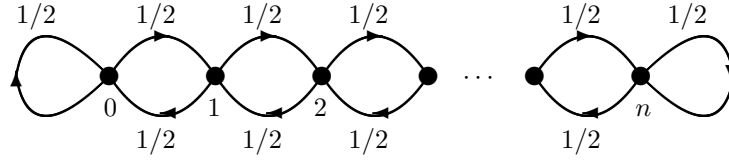


図 12: 両端で lazy な振る舞いをする SRW

つまり、この粒子は、 S の両端の点を除いたところでは SRW のように振る舞い、端点では lazy に振る舞う。この粒子の運動は S 上の Markov 連鎖を定める。それでは、その推移確率行列を $P = (P(x, y))_{x, y \in S}$ とするとき、 $x \leq y$ を満たす任意の $x, y \in S$ について

$$P^t(x, n) \leq P^t(y, n), \quad \forall t \geq 0 \quad (9.1)$$

は成り立つであろうか？

この問いに対する答えは「成り立つ」である。このことは、Markov 連鎖のカップリングを用いることにより容易に確かめることができる。

注 9.1. (9.1) は、 y 出発の粒子の方が早く n に到達することを示す。一見自明のように見えるが、 n が $n-1$ のときは成立しないので、それほど自明ではない。

証明. $x, y \in S$ は $x \leq y$ を満たすとする。 $\{X_t\}_{t \geq 0}, \{Y_t\}_{t \geq 0}$ を、上記のルールにより定まる、それぞれ x, y を出発する Markov 連鎖とする。 $(\{X_t\}, \{Y_t\})$ のカップリング $(\{\tilde{X}_t\}, \{\tilde{Y}_t\})$ を次のように構成する: 公平なコインを投げて、

- 表ならば、 \tilde{X}_t, \tilde{Y}_t はともに右に動き、
- 裏ならば、 \tilde{X}_t, \tilde{Y}_t はともに左に動く。

ただし、状態 0 で左とは状態 0 に留まることを意味し、状態 n で右とは状態 n に留まることを意味する。この $(\tilde{X}_t, \tilde{Y}_t)$ は (X_t, Y_t) のカップリングで

- $\tilde{X}_t \leq \tilde{Y}_t, t \geq 0$
- $\tilde{Y}_t - \tilde{X}_t$ は単調非増加

である。この性質より $\{\tilde{X}_t = n\} \subset \{\tilde{Y}_t = n\}$ となるから

$$\begin{aligned} P^t(x, n) &= \mathbb{P}(\tilde{X}_t = n) \\ &\leq \mathbb{P}(\tilde{Y}_t = n) \\ &= P^t(y, n) \end{aligned}$$

が成り立つ。 □

証明の中で構成したカップリング $(\tilde{X}_t, \tilde{Y}_t)$ は $S \times S$ 上の Markov 連鎖となっている。このように S 上の Markov 連鎖 X_t, Y_t から作られるカップリング $(\tilde{X}_t, \tilde{Y}_t)$ で $(\tilde{X}_t, \tilde{Y}_t)$ が $S \times S$ 上の Markov 連鎖となるものを Markov カップリングと云う。そして $\tilde{X}_t = \tilde{Y}_t$ となる時刻

$$\tau_{\text{couple}} = \inf\{t \geq 0 : \tilde{X}_t = \tilde{Y}_t\}$$

をカップリングタイムと云う。

命題 9.1. $x, y \in S$ を出発する二つの Markov 連鎖のカップリング $(\tilde{X}_t, \tilde{Y}_t)$ を考え, τ_{couple} をそのカップリングタイムとする. そして, $t \geq \tau_{\text{couple}}$ のとき $\tilde{X}_t = \tilde{Y}_t$ を仮定する. このとき

$$\|P^t(x, \bullet) - P^t(y, \bullet)\|_{TV} \leq \mathbb{P}_{(x,y)}(\tau_{\text{couple}} > t) \quad (9.2)$$

が成り立つ. 特に,

$$d(t) \leq \bar{d}(t) \leq \max_{x,y \in S} \mathbb{P}_{(x,y)}(\tau_{\text{couple}} > t) \quad (9.3)$$

が成り立つ.

証明. カップリングの定義より,

$$\begin{aligned} P^t(x, w) - P^t(y, w) &= \mathbb{P}_{(x,y)}(\tilde{X}_t = w) - \mathbb{P}_{(x,y)}(\tilde{Y}_t = w) \\ &= \mathbb{P}_{(x,y)}(\tilde{X}_t = w, \tau_{\text{couple}} > t) + \mathbb{P}_{(x,y)}(\tilde{X}_t = w, \tau_{\text{couple}} \leq t) \\ &\quad - \mathbb{P}_{(x,y)}(\tilde{Y}_t = w, \tau_{\text{couple}} > t) - \mathbb{P}_{(x,y)}(\tilde{Y}_t = w, \tau_{\text{couple}} \leq t) \end{aligned}$$

となる. ところが, 仮定により $\{\tilde{X}_t = w, \tau_{\text{couple}} \leq t\} = \{\tilde{Y}_t = w, \tau_{\text{couple}} \leq t\}$ なので,

$$\begin{aligned} |P^t(x, w) - P^t(y, w)| &= |\mathbb{P}_{(x,y)}(\tilde{X}_t = w, \tau_{\text{couple}} > t) - \mathbb{P}_{(x,y)}(\tilde{Y}_t = w, \tau_{\text{couple}} > t)| \\ &\leq \mathbb{P}_{(x,y)}(\tilde{X}_t = w, \tau_{\text{couple}} > t) + \mathbb{P}_{(x,y)}(\tilde{Y}_t = w, \tau_{\text{couple}} > t) \end{aligned}$$

となる. w について和をとり $1/2$ 倍することで (9.2) が従う. (9.3) は (9.2) を x, y について最大化することで得られる. \square

この命題から, できるだけ早くカップルする (つまり τ_{couple} の小さい) カップリングが構成できれば, 混合時間の上からの良い評価も得られることになる. このことを次のような Markov 連鎖の混合時間の評価を通して考察する.

例 9.1 (Lazy Simple Random Walk on Hypercube). $S_n = \{0, 1\}^n$ に Hamming 距離

$$\rho(x, y) = \#\{1 \leq i \leq n : x_i \neq y_i\}, \quad x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in S,$$

によりグラフ構造を入れる. このグラフ上を動く Markov 連鎖を以下のルールで定める:

1. 確率 $1/n$ で座標を一つ選ぶ.
2. 第 i 座標が選ばれたら確率 $1/2$ で第 i 座標をフリップさせる.

ここで, フリップとは 0 と 1 を入れ替えることを云う. このルールにより得られる Markov 連鎖を, hypercube S_n 上の lazy simple random walk (LSRW) と呼ぶ.

この S_n 上の LSRW の混合時間を $t_{\text{mix}}^{(n)}$ とすると

定理 9.1.

$$t_{\text{mix}}^{(n)} \leq 4n \lceil 1 + \log n \rceil$$

が成り立つ.

証明. カップリングの構成

$x, y \in S$ を出発する LSRW のカップリング (X_t, Y_t) を次のルールで構成する:

1. 確率 $1/n$ で座標を一つ選ぶ.
2. 第 i 座標が選ばれたとき, 確率 $1/2$ で X, Y の第 i 座標をともに 0 または 1 とする.

例えば,

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \underline{1} & 0 \\ 0 & 1 & 1 & \underline{0} & 1 \end{pmatrix}$$

という状態において第 4 座標が選ばれた場合, $\begin{pmatrix} X_{t+1} \\ Y_{t+1} \end{pmatrix}$ として,

$$\begin{pmatrix} 1 & 0 & 0 & \underline{1} & 0 \\ 0 & 1 & 1 & \underline{1} & 1 \end{pmatrix} \quad \text{および} \quad \begin{pmatrix} 1 & 0 & 0 & \underline{0} & 0 \\ 0 & 1 & 1 & \underline{0} & 1 \end{pmatrix}$$

が考えられ, 各々に推移する確率は $1/2$ である. このカップリングでは, 一度選ばれた座標の値はそれ以降同じなので, $\rho(X_t, Y_t)$ は単調非増加である (値の揃った座標が再び選ばれる可能性があるため, ワンステップで必ず 1 減るとは限らない).

クーポンコレクターの問題

n 種類のクーポンをランダムに集める人 (クーポンコレクター) が, 全てのクーポンを集めるまでに要する時間 τ_{coupon} を考える. これは, まだ値の揃っていない座標をこれから集めねばならないクーポンと見なすことにより, (X_t, Y_t) のカップリングタイムを τ_{coupon} を用いて評価することができると考えられるからである.

クーポンコレクターを $\{0, 1, 2, \dots, n\}$ 上の Markov 連鎖と見なし,

Z_t = 時刻 t で集まったクーポンの種類数,

τ_i = Z が状態 $(i-1)$ から i へ移るのに要する時間,

$$\tau_{\text{coupon}} = \sum_{i=1}^n \tau_i$$

とする. このとき $i = 1, 2, \dots, n$ として,

$$\begin{aligned} \mathbb{P}(Z_{t+1} = i-1 | Z_t = i-1) &= \frac{i-1}{n}, \\ \mathbb{P}(Z_{t+1} = i | Z_t = i-1) &= \frac{n-(i-1)}{n} \end{aligned}$$

より τ_i はパラメータ $(n-(i-1))/n$ の幾何分布に従う. つまり

$$\mathbb{P}(\tau_i = k) = \frac{n-(i-1)}{n} \left(\frac{i-1}{n} \right)^{k-1}, \quad k = 1, 2, \dots$$

が成り立つ. よって,

$$\begin{aligned} \mathbb{E}[\tau_i] &= \frac{n}{n-(i-1)}, \\ \mathbb{E}[\tau_{\text{coupon}}] &= \sum_{i=1}^n \mathbb{E}[\tau_i] \\ &= n \sum_{k=1}^n k^{-1} \end{aligned}$$

$$\leq n(1 + \log n)$$

を得る.

仕上げ

クーポンコレクターは始めに 0 個のクーポンを持つが, x, y を出発する LSRW のカップリングは始めからいくつかの座標が同じ可能性があるので $\tau_{\text{coupon}} \geq \tau_{\text{couple}}$ である. これより

$$\begin{aligned} \mathbb{E}_{(x,y)}[\tau_{\text{coupon}}] &\geq \mathbb{E}_{(x,y)}[\tau_{\text{couple}}] \\ &\geq \mathbb{E}_{(x,y)}[\tau_{\text{couple}}; \tau_{\text{couple}} > t] \\ &\geq t\mathbb{P}_{(x,y)}(\tau_{\text{couple}} > t) \end{aligned}$$

となる. 従って, $\mathbb{P}_{(x,y)}(\tau_{\text{couple}} > t) \leq t^{-1}n(1 + \log n)$ が全ての $x, y \in S$ で成り立つ. これと命題 9.1 より,

$$d(t) \leq \frac{n(1 + \log n)}{t}$$

となる. 特に $t = 4n[1 + \log n]$ として $d(4n[1 + \log n]) \leq 1/4$ が従い, 結果が得られた. \square

注 9.2. 実際は $n \rightarrow \infty$ として $t_{\text{mix}}^{(n)} \sim 2^{-1}n \log n$ であることが知られている. 定理 9.1 では, 定数倍の差異はあるものの, $n \log n$ のオーダーは正しく得られていることが分る.

10 混合時間の下からの評価

前節では, 混合時間の上からの評価を行った. 本節では, 混合時間の下からの評価についての考察を行う.

以下, 状態空間 S 上の Markov 連鎖が与えられたものとする. このとき, 与えられた Markov 連鎖の推移確率行列 P から自然にグラフ構造を S に与えることができる. すなわち, $P(x, y) > 0$ となるとき, $x \sim y$ ($xy \in E$) と定義する. このとき, 最短経路距離 ρ ((11.1) 式参照) にもとづいて S の直径

$$\text{diam}(S) = \max_{x,y \in S} (\rho(x, y))$$

を定義する.

10.1 混合時間の下からの自明な評価

以下の命題が成立する.

命題 10.1. $0 < \epsilon < 1/2$ のとき,

$$t_{\text{mix}}(\epsilon) \geq \left\lfloor \frac{\text{diam}(S) - 1}{2} \right\rfloor.$$

証明. 以下, 中心 x , 半径 r なる球を $B(x, r)$ と表記する. $\rho(x_0, y_0) = \text{diam}(S)$ となる x_0, y_0 をとり, $t_0 = \lfloor (\text{diam}(S) - 1)/2 \rfloor$ とおく. いま明らかに $\text{supp}[P^{t_0}(x_0, \bullet)] \subset B(x_0, t_0)$, $\text{supp}[P^{t_0}(y_0, \bullet)] \subset B(y_0, t_0)$ である. $\text{supp}[P^{t_0}(x_0, \bullet)] \cap \text{supp}[P^{t_0}(y_0, \bullet)] \subset B(x_0, t_0) \cap B(y_0, t_0) = \emptyset$ より,

$$\bar{d}(t_0) \geq |P^{t_0}(x_0, B(x_0, t_0)) - P^{t_0}(y_0, B(x_0, t_0))| = 1,$$

すなわち $\bar{d}(t_0) = 1$. 従って, $1 = \bar{d}(t_0) \leq 2d(t_0)$ より, $d(t_0) \geq 1/2$ である. これにより, $0 < \epsilon < 1/2$ ならば $t_{\text{mix}}(\epsilon) \geq t_0$ が成立する. \square

例 10.1 (Lazy Simple Random Walk on Hypercube). S_n を hypercube $\{0, 1\}^n$, また, Markov 連鎖を S_n 上の LSRW とする. このとき, $\text{diam}(S_n) = n$ で,

$$t_{\text{mix}}^{(n)} \geq \left\lfloor \frac{n-1}{2} \right\rfloor = O(n)$$

となる. 注 9.2 でも触れたように, 実際は $t_{\text{mix}}^{(n)} \sim 2^{-1}n \log n$ であるので, 命題 10.1 からはそれほど良い評価は得られない.

10.2 等周定数 (Bottleneck ratio) による混合時間の評価

本節では, 等周定数と呼ばれる, S に Markov 連鎖とは無関係に予め与えられたグラフ構造に依存して決定される定数に基づいて, 混合時間の下からのより良い評価を行う. 推移確率 P , 定常分布 π を持つ S 上の Markov 連鎖 (P, π) に対して, 以下のような $S \times S$ 上の測度を定義する.

$$Q(x, y) = \pi(x)P(x, y),$$

$$Q(A, B) = \sum_{x \in A} \sum_{y \in B} Q(x, y).$$

定義 10.1 (等周定数). S の等周定数 (Bottleneck ratio) Φ^* を以下で定義する.

$$\Phi^* = \min_{D \subset S, \pi(D) \leq 1/2} \Phi(D).$$

ここで,

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

である.

例 10.2 (Simple Random Walk on a Graph). 状態空間 S をグラフ $G = (V, E)$ とする. G 上の (P, π) を SRW, すなわち

$$\pi(x) = \frac{\deg(x)}{2|E|},$$

$$P(x, y) = \begin{cases} \frac{1}{\deg(x)}, & y \sim x, \\ 0, & \text{その他,} \end{cases}$$

により定まる Markov 連鎖とすると,

$$\Phi(D) = \frac{|\partial D|}{\sum_{x \in D} \deg(x)}$$

である. ここで, ∂D は D と D^c を結ぶ辺の集合である

注 10.1. 一般のグラフ $G = (V, E)$ について, D が V の部分集合であれば, ∂D は E の部分集合となる.

次の命題は混合時間が非常に大きい場合の下からの評価によく用いられる.

命題 10.2. S 上の Markov 連鎖 (P, π) が与えられたものとする. このとき

$$t_{\text{mix}} \geq \frac{1}{4\Phi^*}$$

が成立する. すなわち, $\pi(D) \leq 1/2$ なる任意の $D \subset S$ について $t_{\text{mix}} \geq 1/(4\Phi(D))$ が成り立つ.

この命題の証明のために以下の補題を準備する.

補題 10.1. D を S の部分集合, μ を S 上の確率測度とする.

$$\mu_D(A) = \frac{\pi(A \cap D)}{\pi(D)}$$

とおくと,

$$\Phi(D) = \|\mu_D P - \mu_D\|_{TV}$$

である. 特に,

$$\|\mu_D P^t - \mu_D\|_{TV} \leq t\Phi(D)$$

が成立する.

注 10.2. $\mu_D \in \mathcal{P}(S)$ かつ $\text{supp}[\mu_D] \subset D$ である.

証明. $D^c = \{y \in S : \mu_D P(y) \geq \mu_D(y)\}$ となることに注意すれば,

$$\begin{aligned} \|\mu_D P - \mu_D\|_{TV} &= \sum_{y \in S, \mu_D P(y) > \mu_D(y)} (\mu_D P(y) - \mu_D(y)) \\ &= \sum_{y \in D^c} (\mu_D P(y) - \mu_D(y)) \\ &= \sum_{y \in D^c} \sum_{x \in S} \mu_D(x) P(x, y) \\ &= \sum_{y \in D^c} \sum_{x \in D} \frac{1}{\pi(D)} \pi(x) P(x, y) \\ &= \Phi(D) \end{aligned}$$

が成立する. 従って, 補題 8.3 と三角不等式により,

$$\|\mu_D P^t - \mu_D\|_{TV} \leq \sum_{i=1}^t \|\mu_D P^i - \mu_D P^{i-1}\|_{TV} \leq t\Phi(D)$$

が成立する. □

命題 10.2 の証明. $D \subset S$, $\pi(D) \leq 1/2$ とする. $\text{supp}[\mu_D] \subset D$ に注意すれば, $\|\mu_D - \pi\|_{TV} \geq |\mu_D(D^c) - \pi(D^c)| \geq 1/2$ である. 従って, 補題 10.1 より

$$\begin{aligned} \frac{1}{2} &\leq \|\mu_D - \pi\|_{TV} \leq \|\mu_D - \mu_D P^t\|_{TV} + \|\mu_D P^t - \pi\|_{TV} \\ &\leq t\Phi(D) + d(t) \end{aligned}$$

となる. ここで, $t = t_{\text{mix}}$ とおけば, $1/2 \leq t_{\text{mix}}\Phi(D) + 1/4$, つまり $t_{\text{mix}}\Phi(D) \geq 1/4$ である. これは任意の $D \subset S$ に対して成立するので, 命題の不等式を得る. □

ここで, “良い条件を満たす”(つまり, Φ^* に近い値を取る) D について考察する.

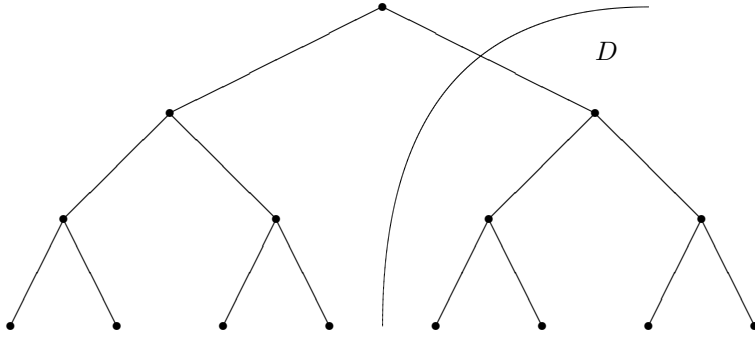


図 13: Binary Tree

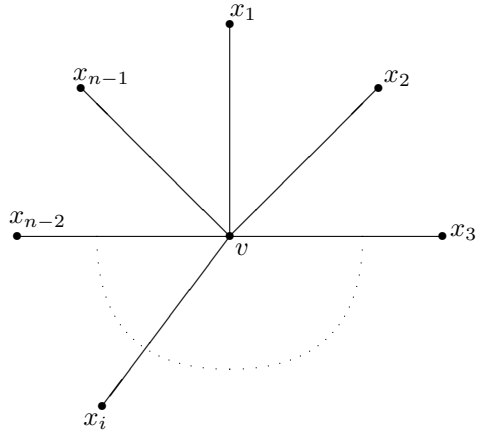


図 14: Star Graph

例 10.3 (Simple Random Walk on Binary Tree). $G = (V, E)$ を binary tree (図 13), (P, π) を binary tree 上の SRW とすると, 等周定数を取る D は弧の右側を含む部分となる.

それでは混合時間が非常に大きくなる例として, 図 14 のようなグラフについて, その proper な彩色集合を状態空間とする Markov 連鎖を考えよう.

例 10.4 (q -coloring on Star Graph). G を star graph $K_{1,n-1} = (V, E)$ として (図 14), S を G 上の proper な彩色集合, すなわち $S = \{c : V \rightarrow [q] : c(x) \neq c(v), x \in N_v\}$ とする. このとき $|S| = q(q-1)^{n-1}$ である. ここで, S 上の Markov 連鎖を以下の手順で構成する:
ある時刻で Markov 連鎖が $c \in S$ という状態にあったとき, 次の時刻での状態を次のルールで定める.

1. V から一様に点 w を選ぶ (つまり, 各点が選ばれる確率は $1/n$ である).
2. 選ばれた $w \in V$ に対して, w に彩色可能な色の集合

$$\mathcal{A}_w(c) = \begin{cases} [q] - \{c(x) : x \in N_v\}, & w = v, \\ [q] - \{c(v)\}, & w \neq v, \end{cases}$$

から一様に一つの色を選んで w に塗る.

すなわち, 1 度の手順において, 1 つの点を選択して, proper な彩色であるという性質を保存したままその点の色を塗り変えるという作業を行うものである. このとき, $q \geq 3$ ならば, この Markov 連鎖は既約かつ非周期的であり, 定常分布は一様分布, すなわち $\pi(c) = 1/|S|$ である. 実際, $c = (1, c(x_1), c(x_2), \dots, c(x_{n-1}))$ として一般性を失わず,

$$\begin{aligned} (1, c(x_1), c(x_2), \dots, c(x_{n-1})) &\rightarrow \cdots \rightarrow (1, 3, \dots, 3) \\ &\rightarrow (2, 3, \dots, 3) \\ &\rightarrow \cdots \rightarrow (2, c'(x_1), c'(x_2), \dots, c'(x_{n-1})) \end{aligned}$$

と色を塗り変える事ができて, v 以外の点は互いに隣り合うことがないので, v 以外の任意の色に彩色可能であることから既約性は示される. また, 非周期性に関しては,

$$\begin{aligned} (1, c(x_1), c(x_2), \dots, c(x_{n-1})) &\rightarrow (1, 2, c(x_2), \dots, c(x_{n-1})) \\ &\rightarrow (1, c(x_1), c(x_2), \dots, c(x_{n-1})) \end{aligned}$$

とすれば 2 回で元の彩色に戻り,

$$\begin{aligned} (1, c(x_1), c(x_2), \dots, c(x_{n-1})) &\rightarrow (1, 2, c(x_2), \dots, c(x_{n-1})) \\ &\rightarrow (1, 3, c(x_2), \dots, c(x_{n-1})) \\ &\rightarrow (1, c(x_1), c(x_2), \dots, c(x_{n-1})) \end{aligned}$$

とすれば 3 回で元の彩色に戻ることから示される. 定常分布が一様分布であることは Markov 連鎖の定義の方法から明らかである. ここで, $D = \{c \in S : c(v) = 1\}$ とおく. $D^c = \{c \in S : c(v) \in \{2, \dots, q\}\}$ であることに注意すれば,

$$\begin{aligned} Q(D, D^c) &= \sum_{c \in D} \sum_{c' \in D^c} \pi(c) P(c, c') \\ &= \sum_{c \in D} \pi(c) P(c, D^c) \\ &= \sum_{c \in D} \frac{1}{|S|} \cdot \frac{1}{n} \cdot \frac{|\mathcal{A}_v(c)| - 1}{|\mathcal{A}_v(c)|} \\ &\leq \frac{1}{|S|n} (q-1)(q-2)^{n-1} \end{aligned} \tag{10.1}$$

となる. 実際, $P(c, D^c)$ は, 上記のルールのもと 1 度の手順で v の色が 1 からそれ以外の色に変更される確率であり, $1/n$ の確率で v を選び, 更に $\mathcal{A}_v(c)$ から 1 以外の色を選ぶ確率を考えれば,

$$P(c, D^c) = \frac{1}{n} \cdot \frac{|\mathcal{A}_v(c)| - 1}{|\mathcal{A}_v(c)|}$$

である. また, $|\mathcal{A}_v(c)| = 1$ のとき $(|\mathcal{A}_v(c)| - 1)/|\mathcal{A}_v(c)| = 0$ であり, その他の場合は $(|\mathcal{A}_v(c)| - 1)/|\mathcal{A}_v(c)| \leq 1$ と評価することで

$$\sum_{c \in D} \frac{|\mathcal{A}_v(c)| - 1}{|\mathcal{A}_v(c)|} \leq \#\{c \in S : c(v) = 1, |\mathcal{A}_v(c)| \geq 2\} \tag{10.2}$$

となる. ここで, (10.2) の右辺は $|\mathcal{A}_v(c)|$ の定義により, 1 を除いた $(q-1)$ 色の内, v 以外の点を塗るのに $(q-2)$ 色以下しか使わない塗り方の総数となることに注意しよう. このような塗り方において

は、まず 1 以外のどの色を使わないかで $(q-1)$ 通りの選び方が考えられる。次に x_1, x_2, \dots, x_{n-1} を残りの $(q-2)$ 色で塗ることになるが、 $|\mathcal{A}_v(c)| \geq 2$ との条件から使わない色の存在を許すので、これらの塗り方には $(q-2)^{n-1}$ 通りの方法が考えられる。ところがこの数え方には、“最初に使わないと決定された色”と“(q-2)色の中から選ばれなかった色”の違いが反映されないため、重複が含まれる。従って、

$$\#\{c \in S : c(v) = 1, |\mathcal{A}_v(c)| \geq 2\} \leq (q-1)(q-2)^{n-1}$$

が成立する。更に π が一様分布であることにより、(10.1) が成り立つことが分かる。よって、

$$\begin{aligned} \Phi(D) &= \frac{Q(D, D^c)}{\pi(D)} \leq \frac{\frac{1}{|S|^n} \cdot (q-1)(q-2)^{n-1}}{\frac{1}{|S|^n} (q-1)^{n-1}} \\ &= \frac{1}{n} \frac{(q-2)^{n-1}}{(q-1)^{n-2}} = \frac{1}{n} \frac{(q-1)^2}{(q-2)} \left(\frac{q-2}{q-1}\right)^n \\ &= \frac{1}{n} \frac{(q-1)^2}{(q-2)} \left(1 - \frac{1}{q-1}\right)^n. \end{aligned}$$

いま、 $1-x \leq e^{-x}$ より、

$$\Phi(D) \leq \frac{(q-1)^2}{n(q-2)} e^{-n/(q-1)}$$

となるから、命題 10.2 より

$$t_{\text{mix}} \geq \frac{q-2}{4(q-1)^2} n e^{n/(q-1)}$$

である。すなわち、この例では n について指数のオーダーでしか mixing が起こらず、従ってコンピュータによるシミュレーションでは観測することができない(つまりシミュレーション上はいつまで経っても mixing が起こらないように見える)。

11 Kantorovich 距離と混合時間

本節では状態空間上の確率測度間の距離を用いて混合時間の評価について考察を行う。

定義 11.1. (S, ρ) を距離空間とする。函数 $\rho_K : \mathcal{P}(S) \times \mathcal{P}(S) \rightarrow [0, \infty)$ を

$$\begin{aligned} \rho_K(\mu, \nu) &= \inf \{ \mathbb{E}[\rho(X, Y)] : (X, Y) \text{ は } (\mu, \nu) \text{ のカップリング} \} \\ &= \inf \left\{ \sum_{x, y \in S} \rho(x, y) q(x, y) : q \text{ は } (\mu, \nu) \text{ のカップリング} \right\} \end{aligned}$$

で定義する。この ρ_K を Kantorovich 距離と云う。

命題 11.1. ρ_K は $\mathcal{P}(S)$ 上の距離となる。

証明. 三角不等式を示す。他の性質は明らかであろう。 $\mu_1, \mu_2, \mu_3 \in \mathcal{P}(S)$ とする。 $p(x, y), q(x, y) \in \mathcal{P}(S \times S)$ をそれぞれ (μ_1, μ_2) のカップリング、 (μ_2, μ_3) のカップリングとする。これを用いて、

$$r(x, z) = \sum_{y \in S} \frac{p(x, y)q(y, z)}{\mu_2(y)}$$

と定義すると、 $r(x, z)$ は (μ_1, μ_3) のカップリングである。ここで、簡単のため、全ての $y \in S$ に対して $\mu_2(y) > 0$ とする。実際、

$$\begin{aligned}\sum_{z \in S} r(x, z) &= \sum_{z \in S} \sum_{y \in S} \frac{p(x, y)q(y, z)}{\mu_2(y)} = \sum_{y \in S} p(x, y) = \mu_1(x), \\ \sum_{x \in S} r(x, z) &= \sum_{x \in S} \sum_{y \in S} \frac{p(x, y)q(y, z)}{\mu_2(y)} = \sum_{y \in S} q(y, z) = \mu_3(z)\end{aligned}$$

となる。よって、

$$\begin{aligned}\rho_K(\mu_1, \mu_3) &\leq \sum_{x, z \in S} \rho(x, z)r(x, z) \\ &= \sum_{x, z \in S} \sum_{y \in S} \rho(x, z) \frac{p(x, y)q(y, z)}{\mu_2(y)} \\ &\leq \sum_{x, z \in S} \sum_{y \in S} \{\rho(x, y) + \rho(y, z)\} \frac{p(x, y)q(y, z)}{\mu_2(y)} \\ &= \sum_{x, y \in S} \frac{\rho(x, y)p(x, y)}{\mu_2(y)} \sum_{z \in S} q(y, z) + \sum_{y, z \in S} \frac{\rho(y, z)q(y, z)}{\mu_2(y)} \sum_{x \in S} p(x, y) \\ &= \sum_{x, y \in S} \rho(x, y)p(x, y) + \sum_{y, z \in S} \rho(y, z)q(y, z)\end{aligned}$$

が成立する。従って、 p, q に関して \inf を取れば、三角不等式を得る。 \square

注 11.1. 1. $S \ni x \mapsto \delta_x \in \mathcal{P}(S)$ という写像を考えれば、 ρ_K は S 上の距離 ρ を $\mathcal{P}(S)$ 上に“リフト”したものと捉えることができる。実際、 (δ_x, δ_y) のカップリングは $\delta_{(x, y)}$ のみであるから、 $\rho_K(\delta_x, \delta_y) = \rho(x, y)$ である。

2. Kantorovich 距離は Wasserstein 距離とも呼ばれる。なお、Kantorovich によるこの距離の概念の発見は 1942 年であり、Wasserstein による“再発見”はそのおよそ 30 年後であった。[12] では、なぜ Kantorovich による発見が忘れられ、これまで後者の名前が冠せられることが多かったかについて、その歴史的経緯が紹介されている。

3. Kantorovich 距離は最適輸送問題 (例えば [14, 15] 参照) に関連する。これは、 μ, ν を“物資”が点 x に存在する割合、 $\rho(x, y)$ は x から y へ物資を運ぶコストと捉えれば、

$$\sum_{x, y \in S} \rho(x, y)q(x, y)$$

は μ の分布に従って配置されている物資を ν の分布に従った配置に移動するコストと考えられ、カップリング $q(x, y)$ はそのための戦略と捉えられるからである。

以下、状態空間 S はグラフ構造を持つものとする。さらに、 S 上の距離 ρ を最短経路距離すなわち、

$$\begin{aligned}\rho(x, y) &= \min\{k : x = x_0x_1x_2 \cdots x_k = y, x_{i-1}x_i \in E\} \\ &= \text{“}x \text{ と } y \text{ を結ぶ経路の長さの最小値”}\end{aligned}\tag{11.1}$$

で定義する。

注 11.2. 最短経路距離は Markov 連鎖 (P, π) から誘導されるグラフ構造とは一般には一致しないが、Markov 連鎖をグラフ上の SRW とした場合一致することが知られている。

次に離散距離 d を以下のように定義する:

$$d(x, y) = \begin{cases} 1, & x \neq y, \\ 0, & x = y. \end{cases}$$

これらの定義から明らかに $d(x, y) \leq \rho(x, y)$ である. 従って任意の S 値確率変数 X, Y について,

$$\begin{aligned} \mathbb{P}(X \neq Y) &= \mathbb{E}[d(X, Y)] \\ &\leq \mathbb{E}[\rho(X, Y)] \end{aligned}$$

となる. そこで (X, Y) を (μ, ν) のカップリングとすれば, (6.3) と ρ_K の定義から

$$\|\mu - \nu\|_{TV} \leq \rho_K(\mu, \nu) \quad (11.2)$$

が成立する.

定理 11.1 (Bubley-Dyer [3]). (S, ρ) 上に Markov 連鎖 (P, π) を与える. ある定数 $\alpha > 0$ に対し $\rho(x, y) = 1$ なる任意の $x, y \in S$ について $(P(x, \bullet), P(y, \bullet))$ のカップリング (X, Y) が存在して,

$$\mathbb{E}[\rho(X, Y)] \leq e^{-\alpha}$$

を満たすと仮定する. このとき,

$$\rho_K(\mu P, \nu P) \leq e^{-\alpha} \rho_K(\mu, \nu) \quad \forall \mu, \nu \in \mathcal{P}(S) \quad (11.3)$$

である. さらに,

$$d(t) \leq e^{-\alpha t} \text{diam}(S) \quad (11.4)$$

となる. 特に,

$$t_{\text{mix}}(\epsilon) \leq \left\lceil \frac{-\log \epsilon + \log \text{diam}(S)}{\alpha} \right\rceil \quad (11.5)$$

が成立する.

証明. $x, y \in S$ を任意に取る. $\rho(x, y) = k$ として, $x = x_0 x_1 x_2 \cdots x_k = y$ となる路を 1 つ固定する. このとき,

$$\begin{aligned} \rho_K(P(x, \bullet), P(y, \bullet)) &\leq \sum_{i=1}^k \rho_K(P(x_{i-1}, \bullet), P(x_i, \bullet)) \\ &\leq k \mathbb{E}[\rho(X, Y)] \\ &\leq e^{-\alpha} \rho(x, y) \end{aligned} \quad (11.6)$$

が成立することに注意しておく. $\mu, \nu \in \mathcal{P}(S)$ に対し, $P_0 \in \mathcal{P}(S \times S)$ を (μ, ν) のカップリング, $Q((x, y), (\bullet, \bullet))$ を $(P(x, \bullet), P(y, \bullet))$ のカップリングとする. このとき, $P_0 Q \in \mathcal{P}(S \times S)$ を

$$(P_0 Q)(z, w) = \sum_{x, y \in S} P_0(x, y) Q((x, y), (z, w)), \quad \forall z, w \in S,$$

で定義すると, $P_0 Q$ は $(\mu P, \nu P)$ のカップリングとなっている. 実際,

$$\begin{aligned} \sum_{w \in S} (P_0 Q)(z, w) &= \sum_{w \in S} \sum_{x, y \in S} P_0(x, y) Q((x, y), (z, w)) \\ &= \sum_{x, y \in S} P_0(x, y) P(x, z) \\ &= \sum_{x \in S} \mu(x) P(x, z) \\ &= (\mu P)(z) \end{aligned}$$

であり, νP についても同様に示される. ここで ρ_K の定義から

$$\begin{aligned}\rho_K(\mu P, \nu P) &\leq \sum_{z, w \in S} \rho(z, w)(P_0 Q)(z, w) \\ &= \sum_{z, w, x, y \in S} \rho(z, w) P_0(x, y) Q((x, y), (z, w))\end{aligned}$$

であり, いま P_0 と Q はそれぞれ最適カップリングであるとすれば, まず Q の最適性から

$$= \sum_{x, y \in S} P_0(x, y) \rho_K(P(x, \bullet), P(y, \bullet)).$$

ゆえに (11.6) より

$$\rho_K(\mu P, \nu P) \leq \sum_{x, y \in S} P_0(x, y) e^{-\alpha} \rho(x, y) = e^{-\alpha} \rho_K(\mu, \nu)$$

が成立し (11.3) が示された. ただし等号は P_0 の最適性による. 従って, (11.2) により,

$$\begin{aligned}\|\mu P^t - \nu P^t\|_{TV} &\leq \rho_K(\mu P^t, \nu P^t) \\ &\leq e^{-\alpha t} \rho_K(\mu, \nu)\end{aligned}$$

が成立する. また, (μ, ν) のカップリングを (X, Y) とすると,

$$\begin{aligned}\rho_K(\mu, \nu) &\leq \mathbb{E}[\rho(X, Y)] \\ &\leq \text{diam}(S)\end{aligned}$$

であるから, $d(t) \leq \bar{d}(t) \leq e^{-\alpha t} \text{diam}(S)$ が成立する. (11.5) は (11.4) から明らかである. \square

一般の有限連結グラフ G について, q 色による頂点の彩色全体の集合 S を状態空間とする Markov 連鎖を考えよう. ただし, ここでは proper でない彩色も含めて考えるものとする.

例 11.1 (Markov Chain on q -coloring). $G = (V, E)$ をグラフとし, $\Delta = \max_{x \in G} \deg(x)$ (最大次数) とおく. S を (proper とは限らない) q -coloring 全体の集合 $S = \{c : V \rightarrow [q]\}$ とし, S の部分集合 S_{proper} を proper な彩色全体の集合とする. S 上の Markov 連鎖を以下の手順で構成する. ある時刻での Markov 連鎖の状態を $c \in S$ として,

1. 点 $v \in V$ と色 $\alpha \in [q]$ を一様に選ぶ (これらは各々 $1/|V|, 1/q$ の確率で選ばれる);
2. α が N_v の色と異なれば, $c(v)$ を α に更新する.

いま, 2 つ目の手順によって更新された彩色が S_{proper} に属せば, それ以降は S_{proper} 内で彩色が更新されていくことに注意しておく. $q \geq \Delta + 2$ であれば, この Markov 連鎖は必ず S_{proper} の内部に入ることが知られている. すなわち, 任意の連結なグラフは $(\Delta + 2)$ -coloring が可能であるということである. このとき, $P|_{S_{\text{proper}}}$ は既約かつ非周期的な Markov 連鎖となる.

上の例の仮定 $q \geq \Delta + 2$ に関連して, Brooks [2] による 1941 年の以下の定理が知られている (例えば [5, p.115, Theorem 5.2.4] を参照のこと): C_n を n 個の頂点から成る cycle 状のグラフ, K_n を n 個の頂点から成る完全グラフ (つまり n 角形の頂点を全て結んで得られるグラフ) とすると,

定理 11.2 (Brooks [2]). グラフ G について,

1. $G \neq C_{2n-1}, K_n$ であれば, Δ -coloring 可能である.
2. $G = C_{2n-1}$ または K_n であれば, $(\Delta + 1)$ -coloring 可能である.

上の例で構成した Markov 連鎖の混合時間の評価について, 次のような定理が知られている.

定理 11.3 (Jerrum [8]). $G = (V, E)$ を, $|V| = n$ で最大次数が Δ であるグラフとし, S を G の q -coloring 全体の集合とする. $q \geq 2\Delta + 1$ ならば, 例 11.1 で構成した S 上の Markov 連鎖の混合時間について, $t_{\text{mix}} = O(n \log n)$ が成り立つ.

頂点数 n について指数以上のオーダーでしか mixing が起こらなかった例 10.4 の Markov 連鎖と比べ (グラフの最大次数は $n - 1$ であった), 次数の低いグラフにおいては, n について高々多項式のオーダーで mixing が起こることをこの定理は主張している. これを fast mixing (あるいは rapid mixing) と呼ぶ. この評価は path coupling の方法 (定理 11.1) を用いることにより示される. なお, この定理に現れた条件 $q \geq 2\Delta + 1$ を $q > (1 + \epsilon)(\Delta + 2)$ に変えても fast mixing が起こるであろうと予想されている. すなわち,

問題 (Open problem). 定理 11.3 と同様の設定のもと, $q > (1 + \epsilon)(\Delta + 2)$ ならば, ある $\alpha > 0$ が存在して, $t_{\text{mix}} = O(n^\alpha)$ となる.

12 補遺

講義中紹介できなかった定理 3.1 の証明をここに付しておく.

定理 3.1 の証明. $S = \{1, 2, \dots, n\}$ と番号をつけて, P を $n \times n$ の確率行列とする. $f : S \times [0, 1) \rightarrow S$ を

$$f(i, u) = j, \quad \sum_{k=1}^{j-1} P(i, k) \leq u < \sum_{k=1}^j P(i, k)$$

によって定義する. $\{\xi_t\}_{t=0,1,2,\dots}$ を $[0, 1)$ 上の一様分布に従う独立確率変数列とし,

$$X_{t+1} = f(X_t, \xi_t), \quad t = 0, 1, 2, \dots$$

と定義すると, これが求める確率過程となることはすぐに確かめられる. □

参考文献

- [1] D. Aldous and P. Diaconis, *Shuffling cards and stopping times*, Amer. Math'l Monthly **93**, 333–348, 1986, available at <http://www-stat.stanford.edu/~cgates/PERSI>
- [2] R. L. Brooks, *On colouring the nodes of a network*, Proc. Cambridge Philos. Soc. **37**, 194–197, 1941.
- [3] R. Bubley, M. Dyer, *Path coupling: A technique for proving rapid mixing in Markov chains*, Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 223–231, 1997.
- [4] P. Diaconis, *The cutoff phenomenon in finite Markov chains*, Proc. Natl. Acad. Sci. USA **93**, 1659–1664, 1996.

- [5] R. Diestel, *Graph Theory*, Electronic Edition 2005, available at <http://www.math.uni-hamburg.de/home/diestel/books/graph.theory>
- [6] R. Durrett, *Essentials of Stochastic Processes*, Springer, 1999.
- [7] A. Frieze, E. Vigoda, *A survey on the use of Markov chains to randomly sample colorings*, in *Combinatorics, Complexity and Chance, A tribute to Dominic Welsh* (G. Grimmett, C. McDiarmid eds.), 53–71, Oxford University Press, 2007, available at <http://www.math.cmu.edu/~af1p>
- [8] M. Jerrum, *A very simple algorithm for estimating the number of k -colourings of a low-degree graph*, available at <http://www.lfcs.inf.ed.ac.uk/reports/94/ECS-LFCS-94-290/index.html>
- [9] G. Kolata, *In Shuffling Cards, 7 Is Winning Number*, The New York Times, Jan. 9, 1990, available at <http://query.nytimes.com/gst/fullpage.html?res=9C0CE0DC103AF93AA35752C0A966958260>
- [10] T. Lindvall, *Lectures on the Coupling Method*, Dover, Mineola, New York, 2002.
- [11] Y. Peres, *Mixing for Markov Chains and Spin Systems*, DRAFT Lecture notes for summer school at UBC on Mixing for Markov chains and spin systems, 2005, available at <http://www.stat.berkeley.edu/~peres>
- [12] A. M. Vershik, *Kantorovich metric: initial history and little-known applications*, *J. Math. Sci.* **133**, 1410–1417, 2006.
- [13] E. Vigoda, *Improved bounds for sampling colorings*, *J. Math. Phys.* **41**, 1555–1569, 2000.
- [14] C. Villani, *Topics in Optimal Transportation*, Graduate Studies in Mathematics **58**, Amer. Math. Soc., 2003.
- [15] C. Villani, *Optimal Transport, Old and New*, version of Feb. 25, 2008, available at <http://www.umpa.ens-lyon.fr/~cvillani/surveys.html>