マルコフ連鎖と混合時間 ー カード・シャッフルの数理 ー

熊谷　隆（数理解析研究所）

August 4, 2011

Abstract

トランプをするとき、ゲームの前にカードがしっかり混ざるようにカードを切り（シャッフルし）ますが、何回くらい切ればカードがよく混ざってくれるでしょうか？この問題は、一般にマルコフ連鎖が定常分布に近づくのにかかる時間（混合時間）を調べる問題として設定することができます。このような定式化の下、トランプを大体７回切ればよく、しかも７回前後で急に「よく混ざった」状態になることが、今から２０年ほど前に P. Diaconis 教授らによって証明されました。

この講義では、カード・シャッフルなどの具体例を念頭に置きながら、マルコフ連鎖とその混合時間について解説します。「急によく混ざった状態になる」という現象（cut-off 現象）についても定式化し、具体例を通じてどのようなときにこの現象が起こるかについても触れたいと思います。

1　シャッフリングの表現

d 枚のカードがあるとき、これをシャッフルする操作は数学的にどのように定式化できるだろうか？カードを下から上に一列に並べ、下から順に $1, 2, \cdots, d$ と番号付けすると、カードをシャッフルすることで（例えば）番号１のカードが下から i_1 番目に移り、2 のカードが下から i_2 番目に移り $\cdots$ という風にカードの並び替えが行われる。このように $1, 2, \cdots, d$ を並び替える操作のことを、置換という。（正確には、$\pi : \{1, 2, \cdots, d\} \to \{1, 2, \cdots, d\}$ が全単射のとき、π を置換という。）つまり、シャッフリングとは数学的には置換のことである。

そこで、少し置換について述べよう。

$$\pi(1) = i_1, \pi(2) = i_2, \cdots, \pi(d) = i_d$$

のとき、

$$\pi = \begin{pmatrix} 1 & 2 & \cdots & d \\ i_1 & i_2 & \cdots & i_d \end{pmatrix} \tag{1.1}$$

と書く。$\pi : \{1, 2, \cdots, d\}$ 上の置換全体を d 次対称群と呼び、$\mathcal{S}_d$ と書く。$\mathcal{S}_d$ の元の数は $d!$ であることは容易に分かる。その名前が示す通り、$\mathcal{S}_d$ は（合成を積とする）群をなす。例えば $d = 3$ として

$$\pi = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix},\ \tau = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$$

ならば、$\tau \circ \pi$（π を施した後 τ を施すという操作）を $\tau\pi$ と書くことにより、

$$\tau\pi = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix},\ \pi\tau = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

である。この例から分かるように、一般に $\tau\pi \neq \pi\tau$ である。（このような群を非アーベル群という。）この積は結合律を満たす、つまり $(\pi_1\pi_2)\pi_3 = \pi_1(\pi_2\pi_3)$ を満たす。d 次対称群の単位元

は、どの文字も動かさない置換（単位置換）である。π が (1.1) で与えられるとき、その逆変換は $\begin{pmatrix} i_1 & i_2 & \cdots & i_d \\ 1 & 2 & \cdots & d \end{pmatrix}$ である（これを π^{-1} と書く）。以上から、$\mathcal{S}_d$ が群の定義を満たすことが分かる。

カード・シャッフルに話を戻そう。カードをシャッフルする時には、上の置換をランダムに行う、つまり何らかの確率分布に従って $\mathcal{S}_d$ の元を選ぶという操作をしているのである。いくつかの具体例を通じて、この意味をもう少し明確にしてみよう。

例 1.1. （ランダム・カット）下から上に並んだ d 枚のカードをランダムに２つの山に分け、上の山をもう下の山の前に移動するという操作をランダム・カットという（いわゆる、「トランプをきる」[1]操作に当たる）。これを数学的に表現すると、1 から d までの間の数字 k を等確率で選び (つまり、すべての $1 \le i \le d$ について $P(k=i)=1/d$、一様分布に従って選ぶという)、

$$\pi_k := \begin{pmatrix} 1 & 2 & \cdots & k & k+1 & \cdots & d \\ d-k+1 & d-k+2 & \cdots & d & 1 & \cdots & d-k \end{pmatrix} \tag{1.2}$$

という置換を行うことである。言い換えるとランダム・カットは、$1 \le i \le d$ の各 k について確率 $1/d$ で $\pi_k \in \mathcal{S}_d$ という置換を選ぶという操作である。

ここで注意しておくこととして、実際にトランプを切るとき常に同じ確率分布できることは人間業では不可能であるし、分布が一様分布であることが妥当かどうかも、明らかではないという事実である。数学的モデル化をする際には、この辺は大胆にやってしまう訳で、現実の問題に話を戻すときにはそのモデルが本当に実際の現象を反映しているかどうかの検証が常に必要になる。

例 1.2. （トップ **to** ランダム）一番下のカードを取って、ランダムにカードの山の中に入れる操作である。(一枚ずつきるとは、カード・シャッフルとしてはあまり効率がよくない感じだが、あくまでモデルということで考えよう。) ランダムな選択を、再び一様分布に従うものと考えると、この数学的表現は、$1 \le k \le d$ の各 k について確率 $1/d$ で

$$k \ge 2 \text{ のとき } \tau_k := \begin{pmatrix} 1 & 2 & \cdots & k & k+1 & \cdots & d \\ k & 1 & \cdots & k-1 & k+1 & \cdots & d \end{pmatrix}, \quad k=1 \text{ のとき単位置換} \tag{1.3}$$

を行うことである。

例 1.3. （ランダム **to** トップ）こちらは、ランダムにカードの山の中からカードを取って、一番下に入れる操作である。カード・シャッフルとしてはあまり現実的ではないかも知れないが、本棚に一列に本が並んでいて、必要な本を取り出して読み終わったり一番端に戻すという（“超整理法”的な）操作がこれにあたる。これも、ランダムな選択を一様分布に従うものと考えると、数学的表現は、$1 \le k \le d$ の各 k について確率 $1/d$ で

$$k \ge 2 \text{ のとき } \hat{\tau}_k := \begin{pmatrix} 1 & \cdots & k-1 & k & k+1 & \cdots & d \\ 2 & \cdots & k & 1 & k+1 & \cdots & d \end{pmatrix}, \quad k=1 \text{ のとき単位置換} \tag{1.4}$$

を行うことである。

例 1.4. （ランダム・トランスポジション）d 枚のカードから、ランダムに２枚を選び（同じカードを選ぶことも許す)、その２枚を交換して戻す操作をランダム・トランスポジションという。ここで２枚のカードの選び方は、それぞれ一様分布に従い独立である（一方の取り出し方が他方に影響しない）とし、この操作を数学的に表現してみよう。まず、確率 $1/d$ で同じカードを選ぶ事になるので、この操作は確率 $1/d$ で単位置換を選ぶことになる。さらに、すべての $i<j$ なるペアについて、以下の置換を確率 $2/d^2$ で選ぶのがランダム・トランスポジションである。(ペアと

[1]私が子供の頃は「トランプをくる」と言っていたが、どうやら「きる」の方が標準的のようなので、ここでは「きる」と書く。

して (i,j) が選ばれるのは、i,j の順で選ばれる場合と j,i の順で選ばれる場合があるので、$2/d^2$ となる。)

$$\begin{pmatrix} 1 & 2 & \cdots & i & \cdots & j & \cdots & d \\ 1 & 2 & \cdots & j & \cdots & i & \cdots & d \end{pmatrix}.$$

例 **1.5.** (リフル・シャッフル) 最後に最も典型的なカード・シャッフルであるリフル・シャッフルを挙げよう。マジシャンがよくやるシャッフルで、カードを二つの山に分け、パラパラパラと混ぜていくやり方である。リフル・シャッフルの数学的表現はどうなるだろうか？順を追ってみて行こう。

i) まず、カードを二つの山に分けるところだが、シンプルでかつこのモデルにふさわしいと考えられるのは二項分布である。つまり、$1 \le k \le d$ なる k について、山の大きさが k と $d-k$ になる確率が ${}_dC_k 2^{-d}$ であるとする。(二項分布は、$k=[d/2]$ の周辺の確率が最も高いので、リフル・シャッフルするときに "二つの山の大きさがある程度同じように分ける" という感覚に近い。)

ii) 山の大きさが k と $d-k$ である二つの山に分けられたとして、次にパラパラパラと混ぜて行くステップを数学的に表現しよう。右手に k 枚、左手に $d-k$ 枚持っているとして、最初にパラッと机に落ちるカードが右手のカードである確率は k/d、左手のカードである確率は $(d-k)/d$ とするのが自然だろう。帰納的に考えて、右手に a 枚、左手に b 枚残っている段階では、次に落ちるカードが右手のカードである確率は $a/(a+b)$、左手のカードである確率は $b/(a+b)$ とする。この操作を、両手のカードがすべて机に落ちるまで続けたとき、起こりうる「二つの山の混ざり方」はすべて当確率で起こり、その確率は $k!(d-k)!/d! = 1/{}_dC_k$ となる。実際、「最初に右手から、次に左手から $\cdots$」といった具合に起こりうるそれぞれの混ざり方の確率を計算してみると、常に分母は $d\cdot(d-1)\cdots 1 = d!$ となり、分子は (かけ算の順番を無視すると) $k\cdot(k-1)\cdots 1\cdot(d-k)\cdot(d-k-1)\cdots 1 = k!(d-k)!$ となることが分かる。なお、k 枚のカードを一列に並んだ d 個の箱に入れ (この入れ方の場合の数は ${}_dC_k$ 通り)、残りの箱に右手の $d-k$ 枚のカードを順に入れることにより、カードが混ざった状態を作ることができるが、このそれぞれの状態が等確率で起きるという設定と上で述べた設定は同じ確率分布を与えるということになる。

iii) 上の操作により、結局混ざったカードの並び方はどういう確率分布をもつだろうか？まず、このようなシャッフルの結果、結局カードの順番が全く変わらなかった (つまり単位置換が選ばれた) 確率は、

$$\begin{aligned}
&\sum_{k=0}^{d} P\Big(\text{一つ目の山 (つまり左手にもつカード) が } k \text{ 枚}\Big) P\Big(\text{順番変わらず} \mid \text{一つ目の山が } k \text{ 枚}\Big) \\
&= \sum_{k=0}^{d} \frac{{}_dC_k}{2^d} \cdot \frac{1}{{}_dC_k} = \frac{d+1}{2^d} \qquad (1.5)
\end{aligned}$$

である。次に、単位置換ではないとすると、混ざったカードの順番はどうなっているだろうか？二つの山に分けたとき、それぞれの番号は $1,2,\cdots,k$ と $k+1,\cdots,d$ であり、これらが順に混じり合う (しかも単位置換にはならない) のだから、混ざったカードの中に $1,2,\cdots,k$ は下からこの順で入っており、また $k+1,\cdots,d$ もこの順で入っているが、$k+1$ は k より下に入っている。言い換えると、以下の $\mathcal{R}$ の元に属する置換を選んだ事になる。

$$\begin{aligned}
\mathcal{R} \;:=\; &\{\pi \in \mathcal{S}_d : 1 \le l \le d-1 \text{ なる } l \text{ があって、} \pi(1) < \cdots < \pi(l) > \pi(l+1), \\
&\qquad \pi(l+1) < \cdots < \pi(d)\}
\end{aligned}$$

では、$\pi_l \in \mathcal{R}$ はどんな確率で選ばれるだろうか？(ここで π_l の l は、$\mathcal{R}$ の元の定義に出て来る l を表すものとする。) π_l が選ばれるには、はじめに二つの山に分けるときに一つ目の山が l 枚でなくてはいけないので、求める確率は

$$P\Big(\text{一つ目の山 (つまり左手にもつカード) が } l \text{ 枚}\Big) P\Big(\pi_l \mid \text{一つ目の山が } l \text{ 枚}\Big) = \frac{{}_dC_l}{2^d} \cdot \frac{1}{{}_dC_l} = \frac{1}{2^d}.$$

つまり $\mathcal{R}$ の元はどれも等確率で選ばれた事になる！

大分長くなったが、以上、いろいろなカード・シャッフルで、どのような確率分布に従って $\mathcal{S}_d$ の元が選ばれるかを具体的に見た。ところで、実際にはカード・シャッフルは一回で終わりではなく、何回も繰り返して行うものである。何回も繰り返す操作は数学的にどのように記述すればよいだろうか？$\mathcal{S}_d$ は群であることを思い出してもらうと、答えは容易に見つかる。

$Y_1, Y_2, \cdots, Y_n$ を、それぞれ $\mathcal{S}_d$ に値を持つ独立な確率変数としよう。各 Y_i が、上の例で見た一回のカード・シャッフル（ある確率分布に従って $\mathcal{S}_d$ の元を選ぶ操作）にあたる。このとき、

$$X_n := Y_n \cdot Y_{n-1} \cdots Y_1 \tag{1.6}$$

とすると、X_n が n 回のカード・シャッフルの結果に相当する。($\mathcal{S}_d$ の群の作用は、右から左に順に行うことにしていたことに注意。）ここで、X_n は X_{n-1} の情報と Y_n の情報があれば決まることに注意しておこう。

これで、カード・シャッフルを繰り返すという操作の数学的記述、つまり $\{X_i\}_{i=0,1,2,\cdots}$ という確率変数の族によるランダムな時間発展ができた。(X_0 は単位置換とする。）このような確率変数の族（ランダムな時間発展）を確率過程と呼ぶ。特に、X_n が X_{n-1} の情報と、推移確率と呼ばれる確率分布のみから決まり、それ以前の過去の情報（$X_1, \cdots, X_{n-2}$ の情報）に依らないとき、マルコフ連鎖と呼ぶ。次の節で見るように、(1.6) はマルコフ連鎖である。

2　マルコフ連鎖と混合時間

この節ではマルコフ連鎖の定義と簡単な性質、そして本講のテーマである混合時間について述べる。抽象的な話が中心になるが、例を頭におきながら、抽象論が意味するところを理解して欲しい。

2.1　マルコフ連鎖と混合時間

S を有限集合とし、その元の数を d とする。S 上の確率測度全体を $\mathcal{P}(S)$ と置く。

$d \times d$ 行列 $P = (P(x,y))_{x,y \in S}$ が、

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

を満たすとき、P を確率行列という。

定理 2.1. $\mu \in \mathcal{P}(S)$ と確率行列 $P = (P(x,y))_{x,y \in S}$ が与えられたとき、次の性質を持つ S に値を取る確率過程 $\{X_i\}_{i=0,1,2,\cdots}$ が存在する。

1. X_0 の分布が μ に従う。(つまり $\mathbb{P}(X_0 = x) = \mu(x)$ がすべての $x \in S$ で成り立つ。)
2. すべての $n \geq 0$ と $x, y, x_0, \cdots, x_{n-1} \in S$ について、以下が成り立つ。

$$\begin{aligned}\mathbb{P}(X_{n+1} = y | X_0 = x_0, X_1 = x_1, \cdots, X_{n-1} = x_{n-1}, X_n = x) &= \mathbb{P}(X_{n+1} = y | X_n = x) \\ &= P(x,y).\end{aligned}$$

上の定理で定まる確率過程を、S 上の推移確率 P、初期分布 μ のマルコフ連鎖と呼ぶ。

前節の例で具体的に見てみよう。前節の例では、$S = \mathcal{S}_d$、初期分布 μ は δ_{id}、つまり確率 1 で単位置換となる分布であり、X_n は (1.6) で定まるのであるが、では推移確率はどうなっているだろうか？まず例 1.1 では、(1.2) で決まる π_k について $\mathcal{U}_1 := \{\pi_k : 1 \leq k \leq d\}$ と置くと、$\sigma, \tau \in \mathcal{S}_d$ に対して $\tau\sigma^{-1} \in \mathcal{U}_1$ なら $P(\sigma,\tau) = 1/d$、そうでなければ $P(\sigma,\tau) = 0$ となる。例 1.2 では (1.3) で決まる τ_k について $\{\tau_k : 1 \leq k \leq d\}$ を $\mathcal{U}_1$ と置き、例 1.3 では (1.4) で決まる $\hat{\tau}_k$ について $\{\hat{\tau}_k : 1 \leq k \leq d\}$ を $\mathcal{U}_1$ と置くことで、同様に推移確率を定めるとよい。例 1.4 のため、$\mathcal{E} := \{(\sigma,\tau) \in \mathcal{S}_d \times \mathcal{S}_d : \tau\sigma^{-1}$が互換$\}$ とする。(なお、$\pi \in \mathcal{S}_d$ が、ある $1 \leq i < j \leq d$

について $\pi(i)=j,\pi(j)=i$ となり、$k\neq i,j$ のとき $\pi(k)=k$ となるとき、π は互換であるという。）すると、例 1.4 の推移確率は、$\sigma,\tau\in\mathcal{S}_d$ に対して $\sigma=\tau$ なら $P(\sigma,\tau)=1/d$、$(\sigma,\tau)\in\mathcal{E}$ なら $P(\sigma,\tau)=2/d^2$、そうでなければ $P(\sigma,\tau)=0$ となる。最後に例 1.5 の推移確率は、$\sigma,\tau\in\mathcal{S}_d$ に対して $\sigma=\tau$ なら $P(\sigma,\tau)=(d+1)/2^d$、$\tau\sigma^{-1}\in\mathcal{R}$ なら $P(\sigma,\tau)=1/2^d$、そうでなければ $P(\sigma,\tau)=0$ となる。

定義 **2.2.** *i)* $\nu=(\nu(x))_{x\in S}\in\mathcal{P}(S)$ が、すべての $x\in S$ について

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

を満たすとき、ν は定常分布であるという。
ii) $\nu=(\nu(x))_{x\in S}\in\mathcal{P}(S)$ が、すべての $x,y\in S$ について

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

を満たすとき、ν は可逆分布であるという。このとき、マルコフ連鎖は可逆であるという。

可逆分布ならば定常分布であることが、定義から容易に確認できる。

例 1.1–1.5 は、いずれも $\mathcal{S}_d$ 上の一様分布を定常分布として持つ。中でも例 1.1 と例 1.4 は、$\mathcal{S}_d$ 上の一様分布を可逆分布として持つ、可逆マルコフ連鎖である。

以下、推移確率 P に対してその n 乗である P^n の (x,y) 成分を $P^n(x,y)$ と表す。

定義 **2.3.** *i)* すべての $x,y\in S$ に対し、ある $n>0$ が存在して $P^n(x,y)>0$ となるとき、マルコフ連鎖は既約であるという。
ii) $x\in S$ に対し、$\{n\geq 1: P^n(x,x)>0\}$ の最大公約数を、マルコフ連鎖の x における周期という。すべての $x\in S$ の周期が 1 であるとき、マルコフ連鎖は非周期的であるという。

マルコフ連鎖が既約とは、どの点から出発してもいつかはすべての点を（正の確率で）通ることである。既約なマルコフ連鎖の周期は、$x\in S$ の取り方によらず定まることが知られている。

定義 **2.4.** $\mu,\nu\in\mathcal{P}(S)$ に対して、これらの全変動距離を以下で定める。

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

このとき次の等式が成り立つ。

命題 **2.5.** $\|\mu-\nu\|_{TV}=\frac{1}{2}\sum_{x\in S}|\mu(x)-\nu(x)|.$

証明：$B:=\{x\in S:\mu(x)>\nu(x)\}$ とおくと、任意の $A\subset S$ に対して

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

となる。ここで二つの不等号が成り立つのは、B の定義から $\mu(A\cap B^c)-\nu(A\cap B^c)\leq 0$, $\mu(A^c\cap B)-\nu(A^c\cap B)\geq 0$ だからである。μ と ν を入れ替えて同様の議論を行うと、

$$\nu(A)-\mu(A)\leq\nu(B^c)-\mu(B^c)$$

となり、$\mu(B)-\nu(B)=1-\mu(B^c)-(1-\nu(B^c))=\nu(B^c)-\mu(B^c)$ であることに注意すると、

$$|\mu(A)-\nu(A)|\leq\mu(B)-\nu(B)=\frac{1}{2}\Big(\mu(B)-\nu(B)+\nu(B^c)-\mu(B^c)\Big)=\frac{1}{2}\sum_{x\in S}|\mu(x)-\nu(x)| \quad (2.1)$$

を得る。A について左辺の最大値を取ると、(2.1) の不等号が等号になることが分かるので、結論を得る。 □

既約かつ非周期的なマルコフ連鎖について、次の収束定理が知られている（例えば、[4, Theorem 4.9] 参照）。

定理 2.6. 推移確率 P を持つマルコフ連鎖が既約かつ非周期的とし、ν を P の定常分布とする。このとき $\theta \in (0,1)$ と $C > 0$ が存在して、以下が成り立つ。

$$\max_{x \in S} \|P^n(x,\cdot) - \nu\|_{TV} \le C\theta^n, \qquad n \ge 1.$$

系 2.7. 既約なマルコフ連鎖は定常分布を唯一つ持つ。

例 1.1 は既約ではない。(実際、$\mathcal{U}_1$ が群をなすので、δ_{id} から出発するマルコフ連鎖は、$\mathcal{U}_1$ の中だけを動く。既約ではないので、前のページで例 1.1 が $\mathcal{S}_d$ 上の一様分布を可逆分布として持つ、と書いたのは多少不適切で、$\mathcal{U}_1$ 上の一様分布を可逆分布として持つというべきであった。) 例 1.2–1.5 は既約である。(マルコフ連鎖を繰り返すことで正の確率で任意の互換を作り出せることを示すと(互換は $\mathcal{S}_d$ を生成するので)よい。)

上で見たように、既約かつ非周期的なマルコフ連鎖は唯一つの定常分布(ν と書く)を持ち、

$$\lim_{n \to \infty} P^n(x,y) = \nu(y), \qquad x,y \in S$$

を満たす。混合時間は、この定常分布に "ほぼ収束" するのにかかる時間である。以下、きちんと定義しよう。

定義 2.8. $n \ge 0$ に対して $d(n) = \max_{x \in S} \|P^n(x,\cdot) - \nu\|_{TV}$ とおく。このとき、$\varepsilon \in (0,1]$ に対して

$$t_{mix}(\varepsilon) = \inf\{n \ge 0 : d(n) \le \varepsilon\}, \quad t_{mix} = t_{mix}(1/4)$$

と定め、t_{mix} をこのマルコフ連鎖の混合時間という。

t_{mix} の定義に出て来る $1/4$ には特段の意味はない。実際 $\varepsilon \in (0,1/2)$ なら $t_{mix}(\varepsilon) \le \lceil \log_2 \varepsilon^{-1} \rceil t_{mix}$ が成り立つことが知られており、$t_{mix}(\varepsilon)$ を用いても定数倍の違いであることが分かる。

2.2 混合時間の上下評価

具体的なマルコフ連鎖について、混合時間が正確に求められる例は限られており、多くの場合混合時間の評価を行うことになる。この小節では、混合時間の上からの評価と下からの評価について、典型的なものを一つずつ紹介する。

(1) 上からの評価: 上からの評価を紹介するために、まず確率分布や確率過程のカップリングについて説明する。

S 上の二つの確率分布 μ,ν に対して、(X,Y) が μ,ν のカップリングであるとは、(X,Y) が一つの確率空間(その確率分布を $\tilde{P}$ と書く)に定義された確率変数であり、X,Y それぞれの周辺分布がそれぞれ μ,ν である(つまり $\tilde{P}(X=x) = \mu(x), \tilde{P}(Y=y) = \nu(y)$ がすべての $x,y \in S$ で成り立つ)ことを言う。特に、S 上の二つのマルコフ連鎖 X_n, Y_n に対して、$(\tilde{X}_n, \tilde{Y}_n)$ が X_n, Y_n のカップリングであり、かつ $S \times S$ 上のマルコフ連鎖でもあるとき、$(\tilde{X}_n, \tilde{Y}_n)$ を X_n, Y_n のマルコビアン・カップリングであるという。このとき、

$$\tau_{\mathrm{couple}} := \min\{n \ge 0 : \tilde{X}_n = \tilde{Y}_n\}$$

(つまり初めて $\tilde{X}_n = \tilde{Y}_n$ となった時刻)を、カップリングタイムという。マルコビアン・カップリング $(\tilde{X}_n, \tilde{Y}_n)$ については、$n \ge \tau_{\mathrm{couple}}$ では $\tilde{X}_n = \tilde{Y}_n$ となる(つまり一旦粒子が出会ったら、それ以降はくっついて動く)ものを考えることが多い。以下本講では、常にそのようなマルコビアン・カップリングを考える。

命題 2.9. X_n, Y_n を、ともに推移確率が P であるマルコフ連鎖とし、$(\tilde{X}_n, \tilde{Y}_n)$ を X_n, Y_n のマルコビアン・カップリングとする。このとき

$$\|P^n(x,\cdot) - P^n(y,\cdot)\|_{TV} \leq \tilde{P}_{(x,y)}(\tau_{\text{couple}} > n) \tag{2.2}$$

が成り立つ（ここで、$\tilde{P}_{(x,y)}$ は、$(\tilde{X}_0, \tilde{Y}_0) = (x,y)$ を意味する）。特に以下が成り立つ。

$$d(n) \leq \max_{x,y \in S} \tilde{P}_{(x,y)}(\tau_{\text{couple}} > n). \tag{2.3}$$

証明： $(\tilde{X}_n, \tilde{Y}_n)$ が X_n, Y_n のカップリングだから

$$\begin{aligned} P^n(x,z) - P^n(y,z) &= \tilde{P}_{(x,y)}(\tilde{X}_n = z) - \tilde{P}_{(x,y)}(\tilde{Y}_n = z) \\ &= \tilde{P}_{(x,y)}(\tilde{X}_n = z, \tau_{\text{couple}} > n) + \tilde{P}_{(x,y)}(\tilde{X}_n = z, \tau_{\text{couple}} \leq n) \\ &\quad - \tilde{P}_{(x,y)}(\tilde{Y}_n = z, \tau_{\text{couple}} > n) - \tilde{P}_{(x,y)}(\tilde{Y}_n = z, \tau_{\text{couple}} \leq n) \\ &= \tilde{P}_{(x,y)}(\tilde{X}_n = z, \tau_{\text{couple}} > n) - \tilde{P}_{(x,y)}(\tilde{Y}_n = z, \tau_{\text{couple}} > n). \end{aligned}$$

ここで、最後の等号は、$\{\tilde{X}_n = z, \tau_{\text{couple}} \leq n\} = \{\tilde{Y}_n = z, \tau_{\text{couple}} \leq n\}$ から従う。よって

$$|P^n(x,z) - P^n(y,z)| \leq \tilde{P}_{(x,y)}(\tilde{X}_n = z, \tau_{\text{couple}} > n) + \tilde{P}_{(x,y)}(\tilde{Y}_n = z, \tau_{\text{couple}} > n)$$

となり、z について和を取り 2 で割ることにより命題2.5 から (2.2) を得る。さらに

$$|P^n(x,A) - \nu(A)| = |P^n(x,A)\sum_{y \in S}\nu(y) - \sum_{y \in S}\nu(y)P^n(y,A)| \leq \sum_{y \in S}\nu(y)|P^n(x,A) - P^n(y,A)|$$

に注意して、(2.2) を x, y について max を取ることにより、(2.3) を得る。 □

この命題から、混合時間の上からの評価を出すためには、$\tilde{P}_{(x,y)}(\tau_{\text{couple}} > n)$ を上から評価すればよいことが分かる。$\tilde{P}_{(x,y)}(\tau_{\text{couple}} > n)$ が小さくなるように、いかにうまくマルコビアン・カップリングを取るかが、腕の見せ所となる。

(2) 下からの評価： ここでは等周定数を用いた下からの評価を紹介する。

定義 2.10. S とその上のマルコフ連鎖 (P, ν) に対して、S の等周定数 Φ^* を以下で定義する。

$$\Phi^* := \min_{D \subset S, \nu(D) \leq 1/2} \frac{\sum_{x \in D}\sum_{y \in D^c}\nu(x)P(x,y)}{\nu(D)}$$

このとき、以下の評価が知られている。

命題 2.11. S 上のマルコフ連鎖 (P, ν) について、以下が成り立つ。

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

3 トランプは何回きればよいか？（カット・オフ現象）

3.1 リフル・シャッフルは７回すればよい

今から 20 年ほど前に P. Diaconis 教授ら（[1]）は、カードゲームを行うときトランプは大体７回きればよく（正確には、リフル・シャッフルは大体７回すればよく）、しかも７回前後で急に「よく混ざった」状態になることを証明した。この結果は 1990 年 1 月 9 日付けの New York Times 誌の記事にもなるなど話題になったが、ここではその証明の流れを [3, 5] に沿って述べる。

リフル・シャッフルについては例 1.5 で述べたので、思い出して欲しい。「よく混ざった状態」とは、数学的には定常状態を指すので、今の場合一様分布（つまり $\mathcal{S}_d$ の各々の元が確率 $1/(d!)$ で（つまり等確率で）現れる分布、以下では U と書く）を指す。どのくらいよく混ざっているかの尺度は、前節で導入した全変動距離を用いることにしよう。初期分布は、確率 1 で単位置換を与える分布としよう（以下この分布を δ_{id} と書く。要は最初は $1, 2, \cdots, d$ の順で整然と並んでいるということ）。P をリフル・シャッフルを独立に繰り返すマルコフ連鎖の推移確率として、n 回のシャッフルにおける「混ざり具合」は以下の量で表すことができる。

$$\|P^n(\delta_{id}, \cdot) - U\|_{TV} = \frac{1}{2}\sum_{\pi \in \mathcal{S}_d} \Big| P^n(\delta_{id}, \pi) - \frac{1}{d!} \Big|. \tag{3.1}$$

ここで、等号は命題 2.5 による。

では、$P^n(\delta_{id}, \pi)$ を計算するにはどうすればよいだろうか？そのために、多少天下り的だが自然数 a に対して a-シャッフルという「きり方」を定義しよう（リフル・シャッフルは、2-シャッフルに当たる）。まずカードを大きさがそれぞれ $k_1, k_2, \cdots, k_a$ 枚の a 個の山に分ける（ただし $k_1 + \cdots + k_a = d$）。それぞれの山にあるカードの番号は、一つ目の山が $1, 2, \cdots, k_1$、次の山が $(k_1+1), \cdots, (k_1+k_2)$ という具合で最後の山のカードは $(k_1+\cdots+k_{a-1}+1), \cdots, d$ である。ここで $k_1, k_2, \cdots, k_a$ は以下のような多項分布に従うとする：

$$P\Big(\text{山の大きさがそれぞれ } k_1, k_2, \cdots, k_a\Big) = \frac{d!}{k_1!k_2!\cdots k_a!}a^{-d}.$$

次に、これら a 個の山のカードをパラパラと混ぜて行く（カードをきる人が a 本の手を持っていて、器用にカードをパラパラ混ぜる様子を思い描いてみよう）。$a = 2$ の場合と同様、「a 個の山の混ざり方」がすべて等確率で起こるように混ぜるものとする。「a 個の山の混ざり方」の場合の数は、d 個の箱に a 種類の（各々の個数が $k_1, \cdots, k_a$ 個の）ボールを入れる入れ方の総数に等しいので $d!/(k_1!k_2!\cdots k_a!)$ となる。従って、各々の混ざり方が実現する確率は

$$\frac{k_1!k_2!\cdots k_a!}{d!}$$

である。以上が a-シャッフルである。$a > d$ でもこのシャッフルは意味を持つ（その場合、「空」の山が必ず出て来る）ことに注意して欲しい。

では、a-シャッフルを行った結果、混ざったカードの並び方はどういう確率分布をもつだろうか？今、$\pi(i) > \pi(i+1)$ となる i が丁度 s 個あるような置換の全体を

$$\begin{aligned}\mathcal{R}_s \quad := \quad & \{\pi \in \mathcal{S} : 1 \le j_1 < j_2 < \cdots < j_s \le d-1 \text{ があって、} \pi(1) < \cdots < \pi(j_1) > \pi(j_1+1), \\ & \pi(j_1+1) < \cdots < \pi(j_2) > \pi(j_2+1), \cdots, \pi(j_s+1) < \cdots < \pi(d)\}\end{aligned}$$

と置き、$r = s+1$ とする（例 1.5 の $\mathcal{R}$ はここでの $\mathcal{R}_1$ に当たる）。すると、a-シャッフルの結果選ばれる π は、$\mathcal{R}_s$ の元で $a \ge r$ となることが容易に分かる。まず、$a = r$ の場合はシャッフルの結果が π であるには、はじめに a 個の山に分けるときに山の大きさがそれぞれ $j_1, (j_2 - j_1), \cdots, (j_s - j_{s-1}), d - j_s$ でなくてはいけないので、求める確率は

$$\begin{aligned}&P\Big(\text{山の大きさが } j_1, (j_2-j_1), \cdots, (j_s-j_{s-1}), d-j_s\Big) P\Big(\pi \,|\, \text{山の大きさが } j_1, (j_2-j_1), \cdots, (j_s-j_{s-1}), d-j_s\Big) \\ &= \frac{d!}{j_1!(j_2-j_1)!\cdots(j_s-j_{s-1})!(d-j_s)!}a^{-d} \cdot \frac{j_1!(j_2-j_1)!\cdots(j_s-j_{s-1})!(d-j_s)!}{d!} = \frac{1}{a^d}\end{aligned} \tag{3.2}$$

となる。$a > r$ の場合は、シャッフルの結果が $\mathcal{R}_s$ の元になるには、まず a 個の山に分け、そのうち $(a-r)$ 個の山についてはシャッフルの際にカードの相対的な順序が変わらないようにシャッフルされなければならない。そこでその $(a-r)$ 個の山については最初から "山をくっつ

ける”こととすると、シャッフルの結果が π になるにはくっつけた結果の山の大きさがそれぞれ $j_1, (j_2 - j_1), \cdots, (j_s - j_{s-1}), d - j_s$ でないといけない。くっつけた結果がこのようになる a 個の山の分け方は、(d 個のボールの間に $(a-r)$ 個の敷居を入れる入れ方の数に等しいから) ${}_{d+(a-r)}C_d$ 通りある。一旦 $j_1, (j_2 - j_1), \cdots, (j_s - j_{s-1}), d - j_s$ 個の山が出来たら、その山をシャッフルした結果 π になる確率は (3.2) に等しいから、結局シャッフルの結果が π である確率は

$$ {}_{d+(a-r)}C_d a^{-d} \tag{3.3}$$

である。ちなみに (1.5) は、ここで $a = 2, r = 1$ の場合に当たる。また、$a = r$ とすると (3.2) と一致することに注意する。

以上で、a-シャッフルとその確率分布が分かった。ここで、鍵となる命題を紹介する。

命題 3.1. a, b を自然数とする。このとき、a-シャッフルの後 b-シャッフルを行うと、全体として ab-シャッフルを行ったと同じことになる（同じ確率分布を与える）。

この命題の証明は、本講では行わない。証明に興味のある方は、例えば [5] をご覧頂きたい。この命題から、独立なリフル・シャッフルを n 回行うことは、2^n-シャッフルを行うことと同じ事になる（同じ確率分布を与える）。また、$\pi \in \mathcal{R}_{r-1}$ とすると (3.3) より

$$P^n(\delta_{id}, \pi) = {}_{d+2^n-r}C_d \cdot 2^{-nd}$$

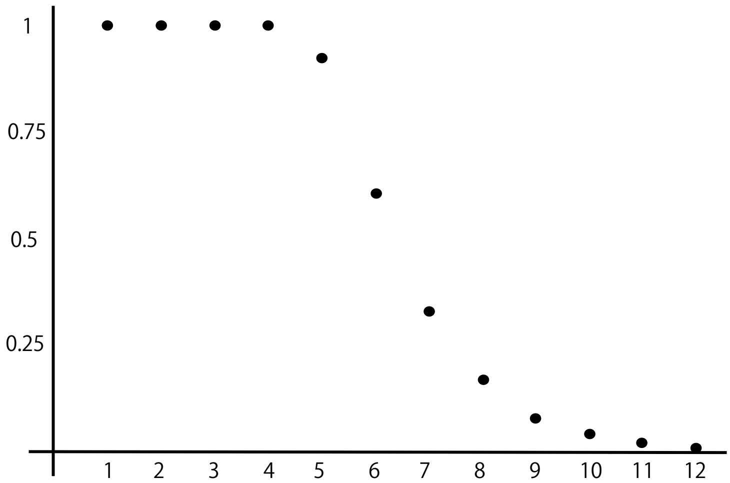
となり、従って (3.1) より、

$$\|P^n(\delta_{id}, \cdot) - U\|_{TV} = \frac{1}{2}\sum_{r=1}^{d} A_{d,r}\left|{}_{d+2^n-r}C_d \cdot 2^{-nd} - \frac{1}{d!}\right| \tag{3.4}$$

となる。ただし $A_{d,r}$ は、$\mathcal{R}_{r-1}$ に属する元の数（これは $\mathcal{S}_d$ の元のうち、昇数字列が丁度 r 個あるものの数に等しい）とする。これは組合せ論で Eulerian number と呼ばれる数で、例えば次のような帰納的な関係式を満たす。

$$A_{d,1} = 1, \quad A_{d,r} = r^d - \sum_{j=1}^{r-1} {}_{d+r-j}C_d \cdot A_{d,j}.$$

以上を用いて、$d = 52$ のトランプの場合に (3.4) の左辺を計算すると以下のようになる ([4] の 111 ページ参照)。この図表を見ると、7 回前後で急に「よく混ざった」状態になると分かってもらえるだろう。



1	2	3	4	5	6	7	8	9	10	11	12
1.0000	1.0000	1.0000	1.0000	0.9237	0.6135	0.3341	0.1672	0.0854	0.0429	0.0215	0.0108

3.2 カット・オフ現象

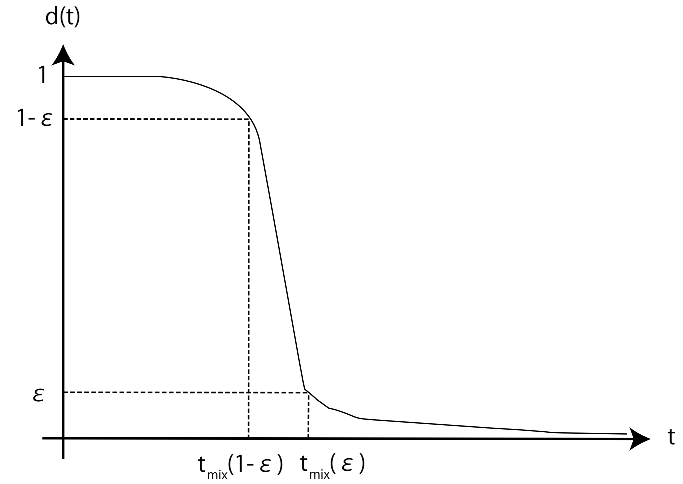
前小節の図から、どのような現象をカット・オフ現象と呼ぼうとしているか、およその雰囲気は分かると思うが、数学的にきちんとこの現象を定式化するには、一つの固定した d を見るのでは不十分である。

定義 3.2. n 個の元からなる有限集合 $S(n)$ 上に、それぞれマルコフ連鎖が与えられているとし、その混合時間を $t^{(n)}_{mix}(\varepsilon)$ と書くことにする。このとき、このマルコフ連鎖の列がカット・オフを持つとは、任意の $\varepsilon > 0$ に対して以下が成り立つことである。

$$\lim_{n\to\infty} \frac{t^{(n)}_{mix}(\varepsilon)}{t^{(n)}_{mix}(1-\varepsilon)} = 1. \tag{3.5}$$

(3.5) のような現象をカット・オフ現象という。このように、数学的には、カット・オフ現象とは無限個のマルコフ連鎖の列に関する性質なのである。(なお、前節までは考えている有限集合の個数を固定していたので d と表していたが、本節以降は n と書く。)

以下の図で分かるように、カット・オフ現象は（n が大きいとき）混合時間の周辺で、急激に $d(t)$ が小さくなる（つまり「良く混ざる」）現象のことなのである。n が十分大きいという点に注意して欲しい。



[1] で Diaconis 達は、以下の定理を示した。

定理 3.3. $\mathcal{S}_n$ 上の独立なリフル・シャッフルを $m = \frac{3}{2}\log_2 n + \theta$ 回繰り返すと、n が大きいとき以下が成り立つ。

$$\Big\| P^m(\sigma,\cdot) - U \Big\|_{TV} = 1 - 2\Phi\Big(\frac{-1}{4\sqrt{3}\cdot 2^\theta}\Big) + O\Big(n^{-1/4}\Big), \qquad \sigma \in \mathcal{S}_n. \tag{3.6}$$

ここで $\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} \exp(-t^2/2)dt$ で、U は $\mathcal{S}_n$ 上の一様分布とする。

(3.6) は $\theta \to -\infty$ のとき 1 で $\theta \to \infty$ のとき 0 になることから任意の $\varepsilon > 0$ について (3.5) が成り立つことがわかり、リフル・シャッフルについてカット・オフ現象が証明されたとこになる。n が十分大きいとき、混合時間は $\frac{3}{2}\log_2 n$ である。($n = 52$ の場合、$\frac{3}{2}\log_2(52) = 8.55\cdots$ なので前小節の 7 回とは少しずれるが、これは 52 という数がまだ “十分大きくない” ために生じるずれである。)

4　カット・オフ現象はいつ起きるか？

少し大胆なタイトルの節だが、答えを一言でいうと「一般には分からない」である。一般のマルコフ連鎖について、カット・オフ現象が起こるための満足のいく必要十分条件は知られていないと言ってよい。ただ、それでは収まりが悪いので、いくつかの例を見ながら、[2] に沿ってカット・オフ現象が起こるメカニズムを調べてみよう。

まず、いくつかの例でカット・オフ現象が起こるかどうかについて見てみる。以下、U は $\mathcal{S}_n$ 上の一様分布とする。

定理 4.1. （例 1.3（ランダム to トップ）の場合）
i) $\theta > 0,\ m = n(\log n + \theta)$ のとき、以下が成り立つ。

$$\limsup_{n\to\infty} \left\| P^m(\sigma,\cdot) - U \right\|_{TV} \le \frac{1}{2}\exp(-2\theta), \quad \sigma \in \mathcal{S}_n.$$

ii) $\theta > 0,\ m = n(\log n - \theta)$ のとき、以下が成り立つ。

$$\lim_{n,\theta\to\infty} \left\| P^m(\sigma,\cdot) - U \right\|_{TV} = 1, \quad \sigma \in \mathcal{S}_n.$$

証明: i) うまいマルコビアン・カップリングを取って、命題2.9を適用することを目指す。$\mathcal{S}_n \times \mathcal{S}_n$ 上のマルコフ連鎖の推移確率 $\tilde{P}((\cdot,\cdot),(\cdot,\cdot))$ を以下のように定める。まず (1.4) で（$d = n$ として）決まる $\hat{\tau}_k$ について $\mathcal{U}_1 := \{\hat{\tau}_k : 1 \le k \le n\}$ と置き、$\tau\sigma^{-1} = \tau'\sigma'^{-1} \in \mathcal{U}_1$ のとき $\tilde{P}((\sigma,\sigma'),(\tau,\tau')) = 1/n$（この値は $P(\sigma,\tau) = P(\sigma',\tau')$ に等しい）と定め、それ以外の場合 $\tilde{P}((\sigma,\sigma'),(\tau,\tau')) = 0$ と定める。このマルコフ連鎖の周辺分布はランダム to トップを繰り返すマルコフ連鎖になるので、(2.3) から

$$d(n) \le \max_{x,y\in S} \tilde{P}_{(x,y)}(\tau_{\text{couple}} > n)$$

となる。$1 \le i < j \le n$ という２つのカードに着目しよう。i, j のどちらかが選ばれてトップに移動したら、その時点でそれぞれのマルコフ連鎖 $\tilde{X}, \tilde{Y}$ における i と j の相対的な位置は同じになる（例えば i が選ばれた場合、i がトップに移動した時点で i の方が j より相対的に前になる）。従って、すべての i, j のペアについてペアのどちらかが選ばれてトップに移動した時点（この時刻を $\hat{\sigma}$ と書くことにしよう）までには、$\tilde{X}, \tilde{Y}$ はそれぞれの初期分布に関わらずカップルしていることになる。m 回のシャッフリングのまでに i, j のペアが選ばれない確率は $(1-2/n)^m$ だから

$$\max_{x,y\in S} \tilde{P}_{(x,y)}(\tau_{\text{couple}} > n) \le P(\hat{\sigma} > n) \le \sum_{i<j}(1-2/n)^m = \frac{n(n-1)}{2}(1-2/n)^{n\log n + n\theta}$$

となり、右辺は $n \to \infty$ のときに $\frac{1}{2}\exp(-2\theta)$ に収束するので、結論を得る。

ii) $A_j := \{\pi \in \mathcal{S}_n :$ 初期分布から π にシャッフルされて行く際、$1, 2, \cdots, n$ のうち選ばれていないカードが j 枚以上ある $\}$ と置くと、 A_j の元の数は相異なる n 個の中から $n-j$ 個を並べる並べ方の数に等しいので、$U(A_j) = n(n-1)\cdots(j+1)/n! = 1/j!$ となる。よって

$$\|P^m(\sigma,\cdot) - U\|_{TV} \ge \max_j |P^m(\sigma,A_j) - U(A_j)| = \max_j |P^m(\sigma,A_j) - 1/j!|$$

となり、従って任意の σ と j について $m = n(\log n - \theta)$ として $\lim_{\theta\to\infty} P^m(\sigma,A_j) = 1$ が示されればよい。さて、n 種類のカードのうち l 種類目のカードが初めて選ばれた時刻を T_l としよう（$1 \le l \le n$ である）。すると

$$P^m(\sigma,A_j) \ge P(T_{n-j} > m) = 1 - P(T_{n-j} \le m)$$

なので、$\lim_{\theta\to\infty} P(T_{n-j} \le m) = 0$ を示せばよいことになる。

そこで $\{T_l\}$ について詳しく見て行こう。(T_n は、n 種類のクーポンをそろえるために買わなければいけないクーポンの枚数を表しており、T_n を計算する問題はクーポンコレクターの問題と呼ばれる。) 明らかに $T_1 = 1$ である。また、$T_{l+1} - T_l$ は l 種類目のクーポンが手に入ってから $l+1$ 種類目のクーポンが手に入るまでにかかる時間を表している。これらは l について独立であり、引いたカードが既に持っている l 種類のどれかである確率は l/n だから、

$$P(T_{l+1} - T_l = s) = (l/n)^{s-1}(1 - l/n), \qquad s = 1, 2, \cdots \tag{4.1}$$

となる（このような分布をパラメータ $(1 - l/n)$ の幾何分布という）。(4.1) を用いて計算すると、

$$E[T_{l+1} - T_l] = \frac{n}{n-l}, \quad \mathrm{Var}\,[T_{l+1} - T_l] := E\Big[(T_{l+1} - T_l - E[T_{l+1} - T_l])^2\Big] = (\frac{n}{n-l})^2(1 - \frac{n-l}{n})$$

となる。$\{T_{l+1} - T_l\}$ の l に関する独立性を用いると、

$$\begin{aligned} E[T_{n-j}] &= \sum_{l=0}^{n-j-1} \frac{n}{n-l} = n\log n + O(n) \\ \mathrm{Var}\,[T_{n-j}] &= \sum_{l=0}^{n-j-1} (\frac{n}{n-l})^2(1 - \frac{n-l}{n}) = O(n^2) \end{aligned}$$

と計算できる。さて、α を十分大きく（以下に出る $O(1)$ の絶対値より大きくなるように）取り、$\theta > \alpha$ と取ると、

$$\begin{aligned} & P(\frac{1}{n}|T_{n-j} - E[T_{n-j}]| \geq \theta - \alpha) = P(|\frac{1}{n}(T_{n-j} - n\log n) - O(1)| \geq \theta - \alpha) \\ \geq\ & P(\frac{1}{n}(T_{n-j} - n\log n) \leq -\theta) = P(T_{n-j} \leq n\log n - \theta n) = P(T_{n-j} \leq m) \end{aligned} \tag{4.2}$$

となる。また、$\beta > 0$ のとき $\beta^2 P(|X| \geq \beta) \leq E[X^2]$ だから $P(|X| \geq \beta) \leq E[X^2]/\beta^2$ となる（チェビシェフの不等式と呼ばれる）ことに気をつけて

$$P(\frac{1}{n}|T_{n-j} - E[T_{n-j}]| \geq \theta - \alpha) \leq \frac{\mathrm{Var}\,[T_{n-j}]}{n^2(\theta-\alpha)^2} \leq \frac{M}{(\theta-\alpha)^2}. \tag{4.3}$$

(4.2), (4.3) をまとめると

$$P(T_{n-j} \leq m) \leq \frac{M}{(\theta-\alpha)^2}$$

となり、$\theta \to \infty$ のとき右辺は 0 に収束するから、$\lim_{\theta\to\infty} P(T_{n-j} \leq m) = 0$ が示された。 □

上の定理から、独立なランダム to トップを繰り返すシャッフルは、時刻 $n\log n$ でカット・オフ現象を起こすことが分かる。

定理 4.2. （例 1.4（ランダム・トランスポジション ）の場合）
i) $\theta > 0$, $m = \frac{n}{2}(\log n + \theta)$ のとき、ある $A > 0$ が存在して以下を満たす。

$$\Big\| P^m(\sigma, \cdot) - U \Big\|_{TV} \leq A\exp(-\theta/2), \quad \sigma \in \mathcal{S}_n.$$

ii) $\theta > 0$, $m = \frac{n}{2}(\log n - \theta)$ のとき、以下が成り立つ。

$$\lim_{n,\theta\to\infty} \Big\| P^m(\sigma, \cdot) - U \Big\|_{TV} = 1, \quad \sigma \in \mathcal{S}_n.$$

つまり、独立なランダム・トランスポジションを繰り返すシャッフルは、時刻 $\frac{n}{2}\log n$ でカット・オフ現象を起こす。

カット・オフ現象を起こすマルコフ連鎖ばかり見てしまったので、カット・オフ現象を起こさないマルコフ連鎖も見ておこう。

例 **4.3.** （サイクル上のランダムウォーク） $S(n) = \{1, 2, \cdots, n\}$ とし、この上に次のような推移確率をもつマルコフ連鎖を考える。

$$P(i, i-1) = 1/4, \quad P(i, i) = 1/2, \quad P(i, i+1) = 1/4, \qquad i \in S(n).$$

ただし、$i = 1$ のとき $i - 1 = n$ とし、$i = n$ のとき $i + 1 = 1$ であるとする。このようなマルコフ連鎖を、サイクル上のレイジー・ランダムウォークという。

このように、$P(i, i) > 0$ であるランダムウォークを一般にレイジー・ランダムウォーク（lazy random walk、"動きの鈍い"ランダムウォーク）と言う。$P(i, i) = 0$ として左右対称に動かしてもよいのだが、敢えてこうするのは、非周期的なマルコフ連鎖を考えたいためである。この例 4.3 は、$S(n)$ 上の一様分布を可逆分布として持つ、可逆マルコフ連鎖である。

定理 **4.4.** （例 4.3（サイクル上のランダムウォーク）の場合）
単調減少する正の連続関数 f で、$\lim_{\theta\to 0} f(\theta) = 1, \lim_{\theta\to\infty} f(\theta) = 0$ なるものが存在して、以下を満たす。

$$\Big\| P^m(\sigma, \cdot) - U \Big\|_{TV} \sim f\Big(\frac{m}{n^2}\Big), \quad \sigma \in S(n).$$

つまり、このマルコフ連鎖の列は、n^2 のオーダーで徐々に一様分布に近づく。(もう少し数学的にいうと、n が大きいとき $S(n)$ 上のこのマルコフ連鎖は $d(tn^2)$ が t を 0 付近から無限大に動かすに連れ、1 から 0 に動く。）従ってカット・オフ現象を起こさない。

では、カット・オフ現象が起こるメカニズムはどうなっているのであろうか？この節の始めにも書いたように、一般のマルコフ連鎖についてはカット・オフ現象が起こるための有用な必要十分条件は知られていない。そこで、以下では可逆マルコフ連鎖に限って、カット・オフ現象が起こる一つの要因について観察する。

元の数が n 個である有限集合 $S(n)$ 上の既約な可逆マルコフ連鎖について、P をその推移確率とする。このとき P の固有値 $\{\beta_i\}_{i=1}^n$ は、$1 = \beta_1 > \beta_2 \geq \cdots \geq \beta_n \geq -1$ と並べることができる。V_i を β_i の直交する固有関数（つまり $PV_i = \beta_i V_i$, $i \neq j$ のとき $\sum_{x\in S(n)} V_i(x)V_j(x) = 0$ なる関数）で $\sum_{x\in S(n)} V_i(x)^2 \cdot \frac{1}{n} = 1$ なるものとする。このとき、「第二固有値が高い重複度を持つ」ことが、カット・オフ現象が起こるための一つの要因となる。このことを見ていこう。

補題 **4.5.** 上の仮定のもと、以下が成り立つ。

$$4\Big\| P^m(x, \cdot) - U \Big\|_{TV}^2 \leq \sum_{i=2}^n \beta_i^{2m} V_i(x)^2, \qquad x \in S(n). \tag{4.4}$$

この補題の証明に、以下の二つの事実を（認めて）使う。

i) （シュワルツの不等式） 任意の実数 a_i, b_i, $(i = 1, 2, \cdots, n)$ について以下が成り立つ。

$$(\sum_{i=1}^n a_i b_i)^2 \leq (\sum_{i=1}^n a_i^2)(\sum_{i=1}^n b_i^2).$$

ii) （パーセバルの等式）$(\cdot, \cdot)$ を内積とする n 次元ベクトル空間で、$\{\phi_k(\cdot)\}_{k=1}^n$ を正規直交系（つまり、任意の $1 \leq i, j \leq n$ について $(\phi_i, \phi_j) = \delta_{ij}$ となるベクトルの列）とすると、

$$(u, u) = \sum_{k=1}^n (u, \phi_k)^2$$

が、任意のベクトル u について成り立つ。
補題 4.5 の証明: まず、既約で可逆なマルコフ連鎖の場合、可逆分布は一様分布である（つまり、(4.4) の U は、任意の $x \in S(n)$ について $U(x) = 1/n$ である）ことに注意する。命題 2.5 とシュワルツの不等式から

$$4\left\|P^m(x,\cdot) - U\right\|_{TV}^2 = \Big(\sum_{y \in S(n)} |P^m(x,y) - 1/n|\Big)^2 \le n\Big(\sum_{y \in S(n)} |P^m(x,y) - 1/n|^2\Big) \tag{4.5}$$

となる。今、$(u,v) = \sum_{x \in S(n)} u(x)v(x) \cdot \frac{1}{n}$ という内積について $(V_i, V_j) = \delta_{ij}$ となり、さらに、簡単な線形代数の計算で $P^m(x,y) = \frac{1}{n}\sum_{i=1}^n \beta_i^m V_i(x)V_i(y)$ と分かるので、$u(y) = \frac{1}{n}\sum_{i=1}^n \beta_i^m V_i(x)V_i(y) - 1/n$ として（$\beta_1 = 1, V_1 \equiv 1$ だから $u(y) = \frac{1}{n}\sum_{i=2}^n \beta_i^m V_i(x)V_i(y)$ となる）、パーセバルの等式を用いると

$$\frac{1}{n}\sum_{y \in S(n)} |P^m(x,y) - 1/n|^2 = \sum_{k=1}^n (u, V_k)^2 = \frac{1}{n^2}\sum_{k=2}^n \beta_k^{2m} V_k(x)^2$$

となり、(4.5) と合わせて結論を得る。 □

第二固有値の重複度を l とすると、さまざまな例で、$\beta_2^{2m}\sum_{j=2}^{l+1} V_j(x)^2$ が (4.4) の左辺の主要項となることが知られている。そこで、いくつかの例で $\beta_2^{2m}\sum_{j=2}^{l+1} V_j(x)^2$ の挙動を見てみよう。（以下の例では、$\beta_2^{2m}\sum_{j=2}^{l+1} V_j(x)^2$ が主要項である。）

例 4.3（サイクル上のランダムウォーク）の場合： 固有値は

$$\beta_j = 1/3 + (2/3)\cdot\cos(2\pi(j-1)/n) \sim 1 - \frac{4\pi^2(j-1)^2}{3n^2}, \qquad j = 1, 2, \cdots$$

となり、従って $\beta_2 \sim 1 - 4\pi^2/(3n^2)$ となり、その重複度は 1 である。固有関数は有界であるので、$m = \theta n^2$ とすると

$$\beta_2^{2m} V_2^2(x) \asymp C(1 - \frac{4\pi^2}{3n^2})^{2m} \stackrel{n\to\infty}{\longrightarrow} C\exp(-4\pi^2\theta/3)$$

となり、確かに θn^2 で θ が 0 から ∞ に動く間にゆっくりと 0 になる様子が見られる。

例 1.4（ランダム・トランスポジション ）の場合： 第二固有値は $(1 - 2/n)$、その重複度は $(n-1)^2$ であり、対応する固有関数はすべて有界となる。従って $m = \frac{n}{2}(\log n + \theta)$ とすると

$$\beta_2^{2m}\sum_{j=2}^{(n-1)^2+1} V_j^2(x) \asymp (n-1)^2(1 - 2/n)^{2m} \stackrel{n\to\infty}{\longrightarrow} \exp(-c\theta)$$

となり、時刻 $\frac{n}{2}\log n$ で急激に減少している。

References

[1] D. Bayer and P. Diaconis, *Trailing the dovetail shuffle to its lair*, Ann. Appl. Probab. **2** (1992), 294–313.

[2] P. Diaconis, *The cuttoff phenomenon in finite Markov chains*, Proc. Natl. Acad. Sci. USA **93** (1996), 1659–1664.
`http://www-stat.stanford.edu/~cgates/PERSI/papers/cutoff.pdf` で入手可能。

[3] G.F. Lawler and L.N. Coyle, *Lectures on Contemporary Probability*, Amer. Math. Soc. 1999.

[4] D. Levin, Y. Peres and E. Wilmer, *Markov chains and mixing times*, Amer. Math. Soc., Providence, RI, 2009.
`http://pages.uoregon.edu/dlevin/MARKOV/markovmixing.pdf` で入手可能。

[5] B. Mann, *How many times should you shuffle a deck of cards?*, in Snell (1995), 261–289.
`http://www.dartmouth.edu/~chance/teaching_aids/Mann.pdf` で入手可能。

[6] 白井朋之, マルコフ連鎖と混合時間, 大学院 GP 数学レクチャーノートシリーズ, 東北大学大学院理学研究科
`http://imi.kyushu-u.ac.jp/~shirai/pdf/shirai2007-tohoku-markovmixing.pdf` で入手可能。