マルコフ連鎖入門

永幡幸生
新潟大学
nagahata@ie.niigata-u.ac.jp

2016 January 4-8

Markov 連鎖

Markov 連鎖は確率過程の中でも取り扱いやすく、現実問題への応用という観点から考えると、物理現象との相性が良い。
取り扱いやすいと言ってもやはり奥が深く、いろいろな取り扱い方があるが、一つの方法として線形代数をよく知っていると分かってくることがある。
応用の観点から Markov 連鎖の族を考えそのスケーリング (時空間に関するスケール極限) を考えたい。このスケーリングをどう決めればよいのかということの一つの指標がスペクトルギャップもしくはスペクトルギャップのオーダーである。また逆にこのオーダーの評価があるとその後の解析ができるようになる例がある。
(例えば流体力学極限の証明)
この講義ではなるべく線形代数の観点からこれらおよび関連する話をする。

Markov 連鎖

この講義では条件付確率は高校の数 I の範囲の割り算で定義するもの、すなわち

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) \neq 0$$

で与えられるもので十分であるが、$P(B) = 0$ になるところでも (ダミーで) 定義されているものと思う。但し基本的には $P(B) = 0$ になるような時には $P(A|B)$ が単独で表れることはなく必ず $P(B)P(A|B)$ のように 0 になる部分が入っており、表記を簡単にするために導入するだけである。
表記を簡単にするため $P(A, B) = P(A \cap B)$ 等と $\cap$ を省略して書くことにする。

問題

$P(B, C) > 0$ のとき $P(A, B|C) = P(A|B, C)P(B|C)$ を示せ。

Markov 連鎖

Markov 過程は確率過程の特別なものである。
確率過程とは時間パラメータの入った確率変数であり
$(X_n)_{n\geq 0}$ を確率過程とした時 $X_n \in S$ となる S のことを状態空間と呼ぶ。
通常確率過程 $(X_n)_{n\geq 0}$ に対して X_n の確率法則はそれ以前の状態 $X_0, X_1, \ldots, X_{n-1}$ に依存する。
(離散時間)Markov 過程とは次の Markov 性を満たす確率過程のことである
$P(X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_1 = x_1, X_0 = x_0) > 0$ であれば

$$
\begin{aligned}
&P(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_1 = x_1, X_0 = x_0) \\
&= P(X_n = x_n | X_{n-1} = x_{n-1})
\end{aligned}
$$

すなわち X_n の確率法則が 1 歩前の X_{n-1} の状態だけで決定されるものである。

Markov 連鎖

用語として Markov 過程と Markov 連鎖とあるが本講義では使い分けをしないが以下の設定では Markov 連鎖のみを取り扱っている。数学辞典によると Markov 過程は状態空間を (d-次元) 実数空間もしくは局所コンパクト Hausdorff 空間で第 2 可算公理を満たすものとし、Markov 連鎖はこのうち状態空間が高々可算集合のものである。
本講義では基本的には状態空間は有限集合を考えるが、可算集合の場合でも成り立つ主張が多くある。同時に特に線形代数、行列の話で有限集合、有限サイズの行列に限った話も出てくる。状態空間が可算集合の場合にも成り立つ主張を状態空間が有限集合の場合に適用すると自明な結果しか得られないことが多々ある。

Markov 連鎖

(離散時間)Markov 過程は X_n の確率法則が 1 歩前の X_{n-1} の状態だけで決定されるものであるが次のように考えるとランダムウォークの自然な拡張と思える。
ランダムウォーカーはコイン (サイコロ) を投げて出た面 (目) によって進む方向を決めて行く。特にこのコイン (サイコロ) はランダムウォーカーの持ち物でありいつも同じものを使っている。
一方で Markov 過程ではウォーカーは自分自身のコイン (サイコロ) をもっておらず、そのたどり着いた場所に備え付けのコイン (サイコロ) を使っている。そのためコイン (サイコロ) 達が場所によって違うことを考える必要がある。

Markov 連鎖

(人間の考える) 多くの確率過程は (状態空間を大きく取れば) Markov 過程になる。一方でその Markov 過程の一部分だけを取り出して確率過程と捉えると Markov 過程にならない (Markov 性を持たない) 例が多くある。

Markov 連鎖

例　倍賭け法のもうけ

まず 2 次元の確率過程 $(X_n, Y_n)_{n\geq 0}$ として勝つか負けるかだけが確率的に決まることに次のような戦略で賭けを行なった時の時刻 n でのもうけを X_n, 次にいくら賭けるかを Y_n で表すものとする。

- 勝てば掛け金を 2 倍に、負ければ 1 にリセットする。

問題を単純にするため $X_0 = 0$、$Y_0 = 1$ とし、毎時独立で表が出る確率が $0 < p < 1$ のコインを投げ、表が出た時に勝ちとする。

$$(X_{n+1}, Y_{n+1}) = \begin{cases} (X_n + Y_n, 2Y_n) & \text{表が出た場合：確率 } p \\ (X_n - Y_n, 1) & \text{裏が出た場合：確率 } 1-p \end{cases}$$

と解釈する。

この設定では当然 $(X_n, Y_n)_{n\geq 0}$ は Markov 過程である。

さらに $(Y_n)_{n\geq 0}$ だけを見ていても、Y_{n+1} は Y_n の 2 倍になるか (確率 p)、1 になるか (確率 $1-p$) だけなので Markov 過程である。

Markov 連鎖

一方コインが「表表裏」「裏裏表」と出た場合を考えると
$(X_n, Y_n)_{n=0,1,2,3}$ は

	(X_0, Y_0)		(X_1, Y_1)		(X_2, Y_2)		(X_3, Y_3)
表表裏	$(0, 1)$	$\to$	$(1, 2)$	$\to$	$(3, 4)$	$\to$	$(-1, 1)$
裏裏表	$(0, 1)$	$\to$	$(-1, 1)$	$\to$	$(-2, 1)$	$\to$	$(-1, 2)$

と推移するので共に $X_3 = -1$ である。
また $Y_n, Y_{n-1}, \dots, Y_1$ の値を知らなくても X_n, X_{n-1} の値を知れば $Y_{n-1} = |X_n - X_{n-1}|$ と分かり、$X_n - X_{n-1}$ の符号を見ることで n 回目の賭けの勝敗も分かる。（正ならば勝ち、負ならば負け。）
最後にこの 2 つの情報から Y_n を決定することが可能になり、従って X_{n+1} の分布は計算することができる。

Markov 連鎖

このことから X_4 に関しての条件付確率は計算できて

$$P(X_4 = k | X_3 = -1, X_2 = 3, X_1 = 1, X_0 = 0) = \begin{cases} p & k = 0 \\ 1-p & k = -2 \\ 0 & \text{その他} \end{cases}$$

$$P(X_4 = k | X_3 = -1, X_2 = -2, X_1 = -1, X_0 = 0) = \begin{cases} p & k = 1 \\ 1-p & k = -3 \\ 0 & \text{その他} \end{cases}$$

であるため $(X_n)_{n\geq 0}$ だけみて Markov 過程になるのであれば

$$\begin{aligned} & P(X_4 = k | X_3 = -1, X_2 = 3, X_1 = 1, X_0 = 0) \\ & = P(X_4 = k | X_3 = -1) \\ & = P(X_4 = k | X_3 = -1, X_2 = -2, X_1 = -1, X_0 = 0) \end{aligned}$$

であるはずであるが明らかに成立していない。

例　Markov 過程とは思えないが Markov 過程に帰着できる例 (ランダムな) 力学モデルの離散化として $V_n = X_n - X_{n-1}$ を速度 (の対応物) ととらえ、X_{n+1} の確率法則は X_n, V_{n-1} で与えられるようなモデルを考えるがこれは X_{n+1} の確率法則は X_n, X_{n-1} で与えられるようなモデルと考え直すことができる。
例として

$$\begin{aligned} X_{n+1} &= X_n + f(X_n, V_n) + \text{ランダム項} \\ &= X_n + f(X_n, X_n - X_{n-1}) + \text{ランダム項} \end{aligned}$$

を考えるが、ランダム項は X_n, X_{n-1} の値には依存してもよい。
このとき X_n は Markov 性を満たさないが
$Y_n := (X_n, V_n) = (X_n, X_n - X_{n-1})$ とおけば $(Y_n)_n$ が Markov 性を持つことを確認することは容易である。

Markov 連鎖

Markov 過程のうち特に

$$P(X_{n+1} = y | X_n = x) = P(X_1 = y | X_0 = x), \quad \forall n$$

を満たすものを時間的に一様な Markov 過程と呼ぶ。
以下では特に断らない限り、時間的に一様な Markov 過程を取り扱うことにする。
時間的に一様な Markov 過程において

$$P_{x,y} := P(X_1 = y | X_0 = x)$$

とおき (巨大なサイズの) 行列を考える。
この行列を推移確率もしくは推移確率行列と呼ぶ。

行列による表記？

$P(X_1 = y | X_0 = x)$ だけを考えたが一般に $P(X_{n+m} = y | X_m = x)$ はどうなるのか？$P_{x,y}$ もしくはこの行列を使って書けるのか？

Markov 連鎖

次のような計算をする。

$$
\begin{aligned}
&P(X_{n+m} = y | X_m = x) \\
&= P(X_{n+m} = y, \bigcup_{y_{n+m-1}} X_{n+m-1} = y_{n+m-1} | X_m = x) \\
&= \sum_{y_{n+m-1}} P(X_{n+m} = y, X_{n+m-1} = y_{n+m-1} | X_m = x) \\
&= \sum_{y_{n+m-1}} P(X_{n+m} = y | X_{n+m-1} = y_{n+m-1}) \\
&\qquad \times P(X_{n-1} = y_{n+m-1} | X_m = x) \\
&= \sum_{y_{n+m-1}} P(X_{n+m-1} = y_{n+m-1} | X_m = x) P_{y_{n+m-1}, y}
\end{aligned}
$$

Markov 連鎖

以下帰納的に計算することで

$$
\begin{aligned}
&P(X_{n+m} = y | X_m = x) \\
&= \sum_{y_{n+m-1}} P(X_{n+m-1} = y_{n+m-1} | X_m = x) P_{y_{n+m-1},y} \\
&\vdots \\
&= \sum_{y_{m+1}} \sum_{y_{m+2}} \cdots \sum_{y_{n+m-1}} P_{x,y_{m+1}} P_{y_{m+1},y_{m+2}} \cdots P_{y_{n+m-1},y} \\
&= (P^n)_{x,y}
\end{aligned}
$$

と行列 P の n 乗に帰着できる。

Markov 連鎖

定義 (a random mapping representation)

推移確率行列 P の関数 f と確率変数 Z による表現とは、関数 $f : S \times \Lambda \to S$ と Λ を状態空間とする確率変数 Z (その確率を Q とする) を用いて

$$Q(\{f(x, Z) = y\}) = P_{x,y}$$

を満たすこととする。

命題 (a random mapping representation)

状態空間 S を有限集合とすると、推移確率行列 P はある関数 f と、ある確率変数 Z による表現が可能である。

実際にコンピュータシミュレーションを行なう上ではこの命題は重要であり、この命題 (の証明) に従ってプログラムを書いていると言っても過言ではない。
証明でも用いるが $\Lambda = [0, 1]$ として Z を $[0, 1]$ 上の一様乱数として実現するのが有効。
またこの命題では状態空間 S を有限集合としたが可算集合でも（もっと広いクラスでも）証明可能である。一方で、コンピュータシミュレーションを行なうことを前提にしていて、現在のコンピュータでは有限集合しか取り扱えないことを考えるとこれで十分でもある。

証明

$S=\{s_1,s_2,\ldots,s_n\}$ として $\Lambda=[0,1]$ とし Z を $\Lambda=[0,1]$ 上の一様分布に従う確率変数とする。
$\{F_{j,k};1\le j\le n,0\le k\le n\}$ を $F_{j,0}=0$, $F_{j,k}=\sum_{i=1}^{k}P_{s_j,s_i}$ ととり $F_{j,k-1}<z\le F_{j,k}$ のとき $f(s_j,z):=s_k$ とする。 但し $f(s,0)=s_1$ とする。
このとき明らかに

$$Q(\{f(x,Z)=y\})=P_{x,y}$$

を満たす。 □

Markov 連鎖

推移確率は $P_{x,y} = P(X_1 = y | X_0 = x)$ で与えたので

$$0 \le P_{x,y} \le 1, \quad \forall x, y$$
$$\sum_y P_{x,y} = 1, \quad \forall x$$

を満たしている。
逆にこの 2 つの条件を満たしていれば対応する Markov 過程が存在する。

Markov 連鎖

一般に $P(X_{n+m} = y | X_m = x)$ が推移確率行列 P の n 乗で書けるので、行列の計算、特に固有値、固有ベクトル、対角化が重要になる。

$$Q^{-1}PQ = \Lambda, \quad P = Q\Lambda Q^{-1}$$

$$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

と書けたとき、全ての固有値が実数だと取り扱いやすい。
この十分条件として P が対称行列であることがあげられるが、これの拡張として

Markov 連鎖

定義 (π に対して対称)

$\pi = (\pi_1, \pi_2, \ldots, \pi_n) > 0$ に対して P が対称であるとは全ての x, y に対して以下を満たすこと

$$\pi_x P_{x,y} = \pi_y P_{y,x}$$

$$\Pi = \begin{pmatrix} \pi_1 & & \\ & \ddots & \\ & & \pi_n \end{pmatrix}, \quad \sqrt{\Pi} = \begin{pmatrix} \sqrt{\pi_1} & & \\ & \ddots & \\ & & \sqrt{\pi_n} \end{pmatrix}$$

等とおくとき

$$S = \sqrt{\Pi} P \sqrt{\Pi}^{-1}$$

とおくと

Markov 連鎖

$$S = \sqrt{\Pi} P \sqrt{\Pi}^{-1} = \sqrt{\Pi}^{-1} (\Pi P) \sqrt{\Pi}^{-1}$$

と書けるが右辺に出てくる 3 つの行列はすべて対称行列であるため S も対称行列になる。
S が実対称行列なので全ての固有値が実数で対角化できるので Λ を対角行列として $S = R \Lambda R^{-1}$ とすると

$$P = \sqrt{\Pi}^{-1} S \sqrt{\Pi} = (\sqrt{\Pi}^{-1} R) \Lambda (R^{-1} \sqrt{\Pi}) = (\sqrt{\Pi}^{-1} R) \Lambda (\sqrt{\Pi}^{-1} R)^{-1}$$

と書けるため P と S は同じ固有値を持っていることが分かる。

逆に全ての固有値が実数でも、どのような π に対しても対称にならない例も作れる。

<u>例</u>

$$\begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}$$

の固有値は $1, 0, -1/2$ になる。

一方 $P_{x,y}, P_{y,x}$ のうち一方が 0 もう一方が 0 でない場合にはどのような $\pi > 0$ を持ってきても対称にすることは不可能。

$$\frac{1}{12}\begin{pmatrix} 2 & 8 & 2 \\ 5 & 2 & 5 \\ 5 & 2 & 5 \end{pmatrix}$$

の固有値も $1, 0, -1/4$ になり、計算するとやはりどのような $\pi > 0$ を持ってきても対称にすることは不可能。(後で解説)

Markov 連鎖

同値類 ↔、既約

定義 (→、↔)

推移確率行列 P が与えられているとき、関係 $\to$、$\leftrightarrow$ を次のように定義する；
(1) $x \to x$.
(2) $\exists n$, $x = x_0, x_1, x_2, \ldots, x_n = y$ s.t. $P_{x_{i-1},x_i} > 0$, $1 \leq \forall i \leq n$ ならば $x \to y$
(3) $x \to y$ かつ $y \to x$ ならば $x \leftrightarrow y$.

この定義では推移確率行列を用いたが一般に行列で定義してよい。

問題

(2) の条件は $\exists n$ s.t. $P(X_n = y | X_0 = x) > 0$ と同値であることを示せ。

Markov 連鎖

補題 (同値関係 $\leftrightarrow$)

関係 $\leftrightarrow$ は同値関係である。すなわち
(1) $x \leftrightarrow x$
(2) $x \leftrightarrow y$ ならば $y \leftrightarrow x$
(3) $x \leftrightarrow y$ かつ $y \leftrightarrow z$ ならば $x \leftrightarrow z$

問題

ほぼ自明ではあるが、この補題を証明せよ。

定義 (既約)

$\forall x, y (\in S)$ に対して $x \leftrightarrow y$ ならば状態空間 S は既約である。
また P が既約であるとも呼ぶ。

既約でない場合は $\leftrightarrow$ で分類される同値類毎に考えればよいので
以後既約の場合だけを考えることにする。

周期

定義 (周期)

x 毎に $N(x) := \{n; (P^n)_{x,x} > 0\}$ とおき、$N(x)$ の最大公約数を d_x と書き、これを周期と呼ぶ。

命題 (周期)

$x \leftrightarrow y$ ならば $d_x = d_y$

証明

$x \leftrightarrow y$ なので $x \to y$ かつ $y \to x$ なので

$$\exists n_1, n_2 \text{ s.t. } P(X_{n_1} = y | X_0 = x) > 0,\ P(X_{n_2} = x | X_0 = y) > 0$$

また $n \in N(x)$ とすると ($n = 0$ も可能であることに注意する)

$$P(X_n = x | X_0 = x) > 0.$$

従って

$$\begin{aligned}
&P(X_{n_1+n+n_2} = y | X_0 = y) \\
&\geq P(X_{n_1+n+n_2} = y, X_{n+n_2} = x, X_{n_2} = x | X_0 = y) \\
&= P(X_{n_1} = x | X_0 = y) P(X_{n_2} = x | X_0 = y) P(X_{n_2} = x | X_0 = y) > 0
\end{aligned}$$

すなわち $n_1 + n + n_2 \in N(y)$ である。

Markov 連鎖

$n = 0$ でも成立しているので $n_1 + n_2 \in N(y)$ なので

$$\exists k \text{ s.t. } n_1 + n_2 = kd_y$$

また $n \in N(x)$ に対して同様に $n_1 + n + n_2 \in N(y)$ なので

$$\exists l \text{ s.t. } n_1 + n + n_2 = ld_y$$

より

$$n = (l - k)d_y$$

と書ける。従って $d_x \geq d_y$ である。
今の議論は x, y を入れ替えても構わないので $d_y \geq d_x$ でもある。 $\square$

Markov 連鎖

補題

$\exists n_x$ s.t. $\forall k \geq n_x,\ kd_x \in N(x)$

問題

この補題を証明せよ。実質的には $d_x = 1$ の場合だけを考えればよい。

注意：$d_x = 1$ で $\#S = n$ とすると $n_x \leq n$ のような気がするが

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

とおくと $n_3 = 5 > 4 = n$ になる。

Markov 連鎖

系

状態空間 S は既約とする。このとき全ての点の周期は等しくなるので d とする。状態空間の番号付けを変更して推移確率行列を書き直す必要があるが
$\exists n$ s.t.

$$P^n = \begin{pmatrix} P_1 & & & \\ & P_2 & & \\ & & \ddots & \\ & & & P_d \end{pmatrix}$$

とブロック行列に分解できて $P_i > 0$ である。更に各 P_i, $1 \leq i \leq d$ は推移確率行列になる。

Markov 連鎖

注意：P_i のサイズが i によって異なっても構わない。例えば次は $d=2$, $n=2$ になるケース

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{pmatrix}, \quad P^2 = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Markov 連鎖

定常測度、不変測度、定常分布、不変分布

定義 (定常測度、定常分布)

S が有限集合 ($\#S = n$ とする) の場合は $\mu = (\mu_{x_1}, \mu_{x_2}, \ldots, \mu_{x_n})$、$S$ が可算集合の場合は $\mu = (\mu_{x_1}, \mu_{x_2}, \ldots,)$ が $\forall x$ に対して $\mu_x \geq 0$ かつ $\mu \neq 0$ とする。
$\mu P = \mu$ を満たすものを定常測度、不変測度と呼び特に
$\sum_x \mu_x = 1$ を満たすものを定常分布、定常確率測度、不変分布、不変確率測度と呼ぶ。

注意：定常測度を定数倍したものは定常測度になる。特に S が有限集合の場合は $\sum_x \mu_x$ で割って規格化して不変分布に変換することができる。

P のサイズが大きい時には不変測度 μ を探すのは困難になるが以下の事実が使える時がある。

問題

以下を示せ
$\pi > 0$ に対して P が対称であれば π は P の定常測度になる。
($\pi \geq 0$ でも (対称の定義を拡張した上で) この主張は成立する。)

Markov 連鎖

$$\frac{1}{12}\begin{pmatrix} 2 & 8 & 2 \\ 5 & 2 & 5 \\ 5 & 2 & 5 \end{pmatrix}$$

の固有値も $1, 0, -1/4$ になり、計算するとやはりどのような $\pi > 0$ を持ってきても対称にすることは不可能。の解説
これを示すには

$$\pi_x P_{x,y} = \pi_y P_{y,x}, \quad x, y = a, b, c,$$

という方程式系を満たす自明でない π を探すことなのでこの方程式系に対応する行列を考えそのランクを考えるのが正統的。
一方で P を対称にする π があったとするとその π は不変測度であるが、不変測度は固有値 1 に対する左固有ベクトルでなければいけない。この例では固有値を先に計算していて固有値 1 は単純特性根なので固有ベクトル $\pi = (1, 1, 1)$ に対して対称になっているかを確かめればよい。

Markov 連鎖

測度 $\mu > 0$ に対して (重み付き) 内積 $\langle \cdot, \cdot \rangle_\mu$ を

$$\langle x, y \rangle_\mu = \sum_i x_i y_i \mu_i$$

とする。通常の内積は $1 = (1, 1, \ldots, 1)$ に対応する、すなわち $x \cdot y = \langle x, y \rangle_1$ と思える。
全ての x, y に対して $\langle x, Ay \rangle_\mu = \langle y, A^* x \rangle_\mu$ を満たす $A^* = A^*_\mu$ を A の μ に対する共役行列と呼ぶことにする。

問題

$(A^*_\mu)_{i,j} = \mu_i^{-1} A_{j,i} \mu_j$ を示せ。
また A が μ 対称ならば $(A^*_\mu) = A$ を示せ。

この問題の前半部分から次のことが分かる。

Markov 連鎖

命題 (P^* と定常測度)

P を確率行列とする。P^*_μ が再び確率行列になるのは μ が P の定常測度の時のみ。

証明

先の問題から $(P^*_\mu)_{i,j} = \mu_i^{-1} P_{j,i} \mu_j$ と書けているので $(P^*_\mu)_{i,j} \geq 0$ は分かる。

$\forall i, \sum_j (P^*_\mu)_{i,j} = 1$ となるのが μ が定常測度の時だけであることをいえばよい。

$$\sum_j (P^*_\mu)_{i,j} = \sum_j \mu_i^{-1} P_{j,i} \mu_j = \mu_i^{-1} \sum_j \mu_j P_{j,i}$$

であることに注意すれば μ が定常測度の場合は明らかに成り立っており、定常測度でない場合は $\sum_j \mu_j P_{j,i} \neq \mu_i$ となる i が少なくとも 1 つあるので成り立たない。 □

Markov 連鎖

定理 (Perron-Frobenius)

A を非負の正方行列として既約とする。このとき以下が成り立つ。
(1) A はスペクトル半径 (固有値の絶対値の最大値) $\rho(A)$ と等しい固有値を持つ。
(2) $\rho(A)$ に対する正の固有ベクトル u が存在する。
(3) $\rho(A)$ は A の全ての成分に対して単調増大。
(4) $\rho(A)$ は単純特性根。

注意：Perron-Frobenius の定理はいくつかの書き方があるようである。基本的には [V] の教科書に従っている。

<u>証明</u>

定理の証明のために用語、補題等を準備する。

Markov 連鎖

x を非負のベクトルとして

$$r_x := \min\{\frac{(Ax)_i}{x_i}; x_i > 0\}$$

と定義する。さらに

$$r := \sup\{r_x; x \geq 0, x \neq 0\}$$

と定義する。
注意：r の定義で全ての非負の x に関する上限を取ったが実際は $\|x\| = 1$ だけに制限しても構わない。
また

$$Az \geq rz$$

を満たす非負のベクトル $z(\neq 0)$ を極ベクトル (extremal vector) と呼ぶ。

Markov 連鎖

A は既約なので $\exists n$ s.t. $(I+A)^n > 0$ とできる。$\|y\|=1$ となる非負のベクトル y に対して

$$T(y) := y' = \tilde{y}'/\|\tilde{y}'\|, \quad \tilde{y}' = (I+A)^n y$$

と決めると $y' > 0$ かつ $\|y'\| = 1$ となる。また $Ay \geq r_y y$ に注意すると

$$\frac{(Ay')_i}{y_i} = \frac{((I+A)^n Ay)_i}{(I+A)^n y)_i} \geq \frac{(r_y(I+A)^n y)_i}{(I+A)^n y)_i} \geq r_y$$

なので $B := \{y;$ 非負のベクトルで $\|y\| = 1\}$ とすると $TB \subset \{y;$ 正のベクトルで $\|y\| = 1\} (= B') \subset B$ となり r を定義するベクトルは TB の上で上限をとっても構わない。B' の上では r_x の定義は連続であり、 TB を含み B' の部分集合であるような閉集合がとれるので極ベクトルは必ず存在する。

補題 (極ベクトル)

A を非負の正方行列として既約とする。このとき以下が成り立つ。
(1) $r > 0$
(2) 極ベクトル z は正で A の r に対する固有ベクトル。すなわち
$Az = rz$

証明

(1) 正のベクトル x に対して Ax も正のベクトルになることから明らかに $r_x > 0$ になっているので定義より $r > 0$

(2) まず既約なので十分大きく n をとれば $(I + A)^n$ は正の正方行列になる。

極ベクトルの定義から $y = Az - rz \geq 0$ になるが $y \neq 0$ とすると $w = (I + A)^n z > 0$ に対して

$$Aw - rw = A(I + A)^n z - r(I + A)^n z = (I + A)^n y > 0$$

このとき

$$r_w = \min_i \{ \frac{Aw_i}{w_i} \} = \min_i \{ \frac{rw_i + ((I + A)^n y)_i}{w_i} \} > r$$

より r の定義に矛盾する。すなわち $Az = rz$

この下で $w = (I + A)^n z = (1 + r)^n z > 0$ なので z は正。 $\square$

Markov 連鎖

補題 (行列の比較)

A を非負の正方行列として既約とする。B を複素正方行列とし $|B| \leq A$ とする。
β を B の固有値とすると

$$|\beta| \leq r$$

さらに等号が成立する必要十分条件は $|B| = A$ で $\beta = re^{i\phi}$ と書けたとき対角行列 D でその対角成分 d_i は $|d_i| = 1$ を満たすものを用いて

$$B = e^{i\phi} DAD^{-1}$$

と書ける。

証明

$y \neq 0$ として $By = \beta y$ とすると各成分を比較することで

$$|\beta||y| \leq |B||y| \leq A|y|$$

すなわち $|\beta| \leq r_{|y|} \leq r$ である。
$|\beta| = r$ を仮定すると $|\beta||y| \leq A|y|$ なので $|y|$ は極ベクトル。先の補題より $|y|$ は A の固有値 r に対応する固有ベクトル。すなわち $r|y| = A|y|$。上の式と比較すると

$$r|y| = |B||y| = A|y|$$

でなければならない。また先の補題より $|y|$ は正のベクトルなので $|B| \leq A$ とこの式から $|B| = A$ でなければいけない。

Markov 連鎖

$|y|>0$ となるベクトル y に対して

$$D = \begin{pmatrix} y_1/|y_1| & & \\ & \ddots & \\ & & y_n/|y_n| \end{pmatrix}$$

は補題の条件を満たしていて $y = D|y|$ と書ける。
$\beta = re^{i\phi}$ として $By = \beta y$ をこれらを使って書き換えると

$$BD|y| = By = \beta y = e^{i\phi} rD|y|,\ \text{すなわち}\ e^{-i\phi}D^{-1}BD|y| = r|y|$$

と書ける。前ページから証明してきたことにより

$$|e^{-i\phi}D^{-1}BD| = |B| = A, \quad e^{-i\phi}D^{-1}BD|y| = r|y| = A|y|$$

となるが次のページの問題から $B = e^{i\phi}DAD^{-1}$ である。 □

問題

C を複素行列とする。$\exists y > 0$ s.t.

$$|C|y = Cy$$

であれば $|C| = C$ を示せ。

Markov 連鎖

補題 (行列の比較) で $B = A$ とすることで

系 (行列の比較)

A を非負の正方行列として既約とする。このとき r はスペクトル半径にもなる。

正方行列から j 行と j 列を取り除いてできた行列を B とする。(取り除く行、列は 2 つ以上でもよい。) このような行列を主部分行列 (principal submatrix) と呼ぶ。
行の交換に対応する行列 P を使えば PAP^{-1} で取り除きたい行、列を下 (右) から順に並べ直すことが可能なので B は A のブロック行列の左上の部分と思ってよい。

補題 (スペクトル半径の比較)

A を非負の正方行列として既約とする。このとき A の主部分行列 B に対して $\rho(A) > \rho(B)$

証明

前ページに書いたように行の交換に対応する行列 P を用いて

$$A' = PAP^{-1} = \begin{pmatrix} B & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix}$$

と変換して $\rho(A')$ と $\rho(B)$ を比較をしているものと思い直してよい。

$$C = \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix}$$

と定義すると明らかに $\rho(B) = \rho(C)$ でかつ $0 \leq C \leq A'$, $C \neq A$ が成り立つ。従って補題 (行列の比較) が適用でき
$\rho(B) < \rho(A)$ □

Markov 連鎖

問題

各成分が微分可能な関数で与えられる行列 $A(t)$ を
$A(t) = (a_1(t), a_2(t), \ldots, a_n(t))$ の形で書く。このとき

$$\frac{d}{dt} detA(t) = \sum_i \det(a_1(t), \ldots, \frac{d}{dt} a_i(t), \ldots, a_n(t))$$

を示せ、但し行列 (ベクトル) の微分は各成分の微分とする。
特に B を定数行列として $A(t) = tI - B$ を考えることで

$$\frac{d}{dt} \det(tI - B) = \sum_i \det(tI - B_i)$$

を示せ、但し B_i は B の i 行、列を取り除いた主部分行列とする。

Markov 連鎖

定理の証明に戻る。
(1) A はスペクトル半径 (固有値の絶対値の最大値) $\rho(A)$ と等しい固有値を持つ。
(2) $\rho(A)$ に対する正の固有ベクトル u が存在する。
この 2 つは補題 (極ベクトル) と系 (行列の比較) から出る。
(3) $\rho(A)$ は A の全ての成分に対して単調増大。
は補題 (行列の比較) から出る。
(4) $\rho(A)$ は単純特性根。
は補題 (スペクトル半径の比較) より A_i を A の i 行、列を取り除いた主部分行列とするとき $\rho(A_i) < \rho(A)$ であるので $\det(\rho(A)I - A_i) > 0$ $(\forall i)$ である。問題とあわせることにより

$$\frac{d}{dt}\det(tI - A)|_{t=\rho(A)} = \sum_i \det(tI - A_i)|_{t=\rho(A)} > 0$$

より特性方程式において $\rho(A)$ は単純特性根であることが分かる。 □

Markov 連鎖

命題

A を非負の正方行列として既約として正のベクトル u が A の固有ベクトルであるとする。すなわち $Au = \lambda u$ とすると $\lambda = \rho(A)$ である。

注意：[Sai] の教科書の中ではこの命題は Perron-Frobenius の定理の一部になっている。確率論的にはこの事実から $\rho(P) = 1$ がすぐに分かる。(ブロック行列に分けて考える毎に $\rho(P_i) = 1$ が分かる。)

証明

${}^tA$ に関しても Perron-Frobenius の定理は適用可能なので正のベクトル v があって ${}^tvA = \rho(A){}^tv$ を満たす。

$$\rho(A){}^tvu = ({}^tvA)u = {}^tv(Au) = \lambda{}^tvu$$

を満たしていて $u, v > 0$ であるから $\rho(A) = \lambda$ になる。 □

Markov 連鎖

命題

A を正の正方行列とする。このとき A の固有値 λ のうち $\lambda \neq \rho(A)$ となるものは $|\lambda| < \rho(A)$ である。

注意：[Sai] の教科書の中ではこの命題は Perron-Frobenius の定理の一部になっている。

証明

${}^tA$ に関しても Perron-Frobenius の定理は適用可能なので正のベクトル v があって ${}^tvA = \rho(A){}^tv$ を満たす。一方 A の固有値 $\lambda(\neq \rho(A))$ とその固有ベクトル x に対して

$$\lambda x_i = \sum_i a_{i,j}x_j, \qquad |\lambda||x_i| = |\lambda x_i| = |\sum_i a_{i,j}x_j| \leq \sum_i a_{i,j}|x_j|$$

が成り立つ。

Markov 連鎖

2 つめの不等式の両辺に $v_i > 0$ をかけて i に関して和をとると v が $\rho(A)$ に対応する固有ベクトルだったので

$$|\lambda| \sum_i v_i |x_i| \leq \sum_{i,j} v_i a_{i,j} |x_j| = \rho(A) \sum_j v_j |x_j|$$

もし $|\lambda| = \rho(A)$ が成り立つのであればこの不等式でも、1 つ前の不等式でも等号が成立している。すなわち

$$|\sum_i a_{i,j} x_j| = \sum_i a_{i,j} |x_j|$$

が成り立つ。A の全ての成分が正なのでベクトル x は定数倍 (複素数) することで正のベクトルになる。
先の命題より $|\lambda| = \rho(A)$ であれば $\lambda = \rho(A)$ になる。 $\square$

系

A を非負の正方行列とし、$\exists k$ s.t. A^k は正の正方行列とする。このとき A の固有値 λ のうち $\lambda \neq \rho(A)$ となるものは $|\lambda| < \rho(A)$ である。

注意：実際周期 (d とする) のある確率行列 P に対しては $\rho(P) = 1$ となるが $\omega = \exp(i2\pi/d)$ として $\omega, \omega^2, \ldots, \omega^{d-1}$ が出てくる。

問題

A の固有値を用いて A^k の固有値を表すことで系を証明せよ。

命題 (確率行列の極限)

P を確率行列とし既約、周期 1 (非周期) とする。このとき $\lim_{n\to\infty} P^n$ が存在し

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix}$$

と書け、$\mu > 0$ は唯一の定常分布になる。

<u>証明</u>
確率行列なので $u = {}^t(1, \ldots, 1)$ とすれば

$$Pu = u$$

を満たしている。Perron-Frobenius の定理を適用することで $\rho(P) = 1$ であることが分かる。

Markov 連鎖

一方既約で周期 1 なので十分大きな k をとれば P^k は正の正方行列になる。従って命題より 1 以外の固有値 λ は $|\lambda| < \rho(P) = 1$ を満たしている。
また当然

$$vP = v$$

を満たすベクトル v も存在するがこちらも Perron-Frobenius の定理から $v > 0$ となる。これを規格化して

$$\mu = \frac{1}{\sum_i v_i} v$$

とおくことで唯一の定常分布を得られる。(当然 $\mu > 0$)

Markov 連鎖

$Q^{-1}PQ = \Lambda$ と対角化 (必要ならばジョルダン標準形に) するが $\lambda_1 = 1$ とする。このとき $Q = (q_1, q_2, \ldots, q_n)$, $Q^{-1} = {}^t(q'_1, q'_2, \ldots, q'_n)$ と書いたとき q_1, q'_1 は固有値 1 に対する固有ベクトルであり ${}^tq'_1 q_1 = 1$ を満たしているので特に

$$q_1 = {}^t(1, 1, \ldots, 1), \qquad {}^tq'_1 = (\mu_1, \mu_2, \ldots, \mu_n) = \mu$$

と取ってよい。ここで 1 以外の固有値 λ が $|\lambda| < 1$ であることから

$$\begin{aligned}\lim_{n\to\infty} P^n &= \lim_{n\to\infty} Q\Lambda^n Q^{-1} \\ &= Q \begin{pmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} Q^{-1} = \begin{pmatrix} 1 & 0 \cdots 0 \\ 1 & 0 \cdots 0 \\ \vdots & \vdots \\ 1 & 0 \cdots 0 \end{pmatrix} Q^{-1} = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix}\end{aligned}$$

$\square$

Markov 連鎖

平均再帰時間を考えるためには再帰確率が 1 でなければ ∞ になる。まず再帰性、非再帰性について準備する。
T 到達時刻または再帰時刻で

$$T_x := \min\{n \geq 1; X_n = x\}$$

とする。ただし $\min \emptyset = \infty$ である。
最終的に考えたいのは $E[T_x | X_0 = x]$ であるが
$P(T_x < \infty | X_0 = x) < 1$ ではこの値が ∞ になる。

定義 (再帰、非再帰)

状態 $x \in S$ が再帰的であるとは

$$P(T_x < \infty | X_0 = x) = 1$$

であること。また状態 x が再帰的でないとき非再帰的であると呼ぶ。

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/3 & 2/3 \end{pmatrix}$$

の場合に実際に計算してみると (場所と時間が混同するので場所に関しては上から順に a, b, c とする)

$$\begin{aligned} &P(T_a = 1 | X_0 = a) = P(X_1 = a | X_0 = a) = \frac{1}{2} \\ &P(T_a = 2 | X_0 = a) = P(X_2 = a, X_1 \neq a | X_0 = a) \\ &= P(X_2 = a | X_1 = b) P(X_1 = b | X_0 = a) \\ &\quad + P(X_2 = a | X_1 = c) P(X_1 = c | X_0 = a) = 0 \end{aligned}$$

以下同様に考えると

$$P(2 \leq T_a < \infty | X_0 = a) = 0$$

は分かる。

Markov 連鎖

同様に

$$
\begin{aligned}
&P(T_b = 1 | X_0 = b) = P(X_1 = b | X_0 = b) = \frac{1}{2}\\
&P(T_b = 2 | X_0 = b) = P(X_2 = b, X_1 \neq b | X_0 = b)\\
&= P(X_2 = b | X_1 = a) P(X_1 = a | X_0 = b)\\
&\quad + P(X_2 = b | X_1 = c) P(X_1 = c | X_0 = b) = \frac{1}{6}
\end{aligned}
$$

以下同様に考えると $k \geq 2$ で

$$
P(T_b = k | X_0 = b) = \frac{1}{6} \left(\frac{2}{3} \right)^{k-2}
$$

と計算できるが実際には本質的に 2×2 行列しか見ていないので計算できる。もう少し複雑になると直接的に確率を計算することはほぼ不可能になる。

定理 (再帰的)

状態 x が再帰的である必要十分条件は

$$\sum_{n=0}^{\infty} P(X_n = x | X_0 = x) = \infty$$

である。

<u>証明</u>

$$P_n = P(X_n = x | X_0 = x), \quad f_n = P(T_x = n | X_0 = x)$$

とおく。

Markov 連鎖

$X_n = x$ という事象をはじめて x に戻ってきた時間で分けると

$$\{X_n = x\} = \bigcup_{k=1}^{n} \{X_n = x, T_x = k\}$$

と排反事象に分けられ、

$$\{T_x = k\} = \{X_k = x, X_{k-1} \neq x, X_{k-2} \neq x, \ldots, X_1 \neq x\}$$

と書けるので

Markov 連鎖

これらの式に Markov 性を加えれば

$$
\begin{aligned}
P_n &= P(X_n = x | X_0 = x) \\
&= P(\bigcup_{k=1}^{n} P(\{X_n = x, T_x = k\} | X_0 = x) \\
&= \sum_{k=1}^{n} P(X_n = x, X_k = x, X_{k-1} \neq x, X_{k-2} \neq x, \ldots, X_1 \neq x | X_0 = x) \\
&= \sum_{k=1}^{n} P(X_n = x | X_k = x) \\
&\qquad \times P(X_k = x, X_{k-1} \neq x, X_{k-2} \neq x, \ldots, X_1 \neq x | X_0 = x) \\
&= \sum_{k=1}^{n} P(X_n = x | X_k = x) P(T_x = k | X_0 = x) = \sum_{k=1}^{n} P_{n-k} f_k
\end{aligned}
$$

$P_0 = 1$ を加えて両辺 n に関して和をとり、和の順序交換、変数変換をすることで

$$\sum_{n=0}^{\infty} P_n = 1 + \sum_{n=1}^{\infty} \sum_{k=1}^{n} P_{n-k} f_k$$
$$= 1 + \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} P_{n-k} f_k = 1 + \sum_{k=1}^{\infty} f_k \sum_{n=0}^{\infty} P_n$$

と計算できるので少なくとも形式的には

$$\sum_{k=1}^{\infty} f_k = 1 - \frac{1}{\sum_{n=0}^{\infty} P_n}$$

と計算でき定理の主張と一致する。

Markov 連鎖

もちろん $\sum P_n < \infty$ のときに非再帰的という事実に関してはこれで証明になる。しかし $\sum P_n = \infty$ のときには $\sum f_n \neq 0$ くらいしか分からない。
母関数の方法 収束因子をつけて極限を取る。前ページの計算に $|s| < 1$ として s^n を加えてみる。

$$\sum_{n=0}^{\infty} P_n s^n = 1 + \sum_{n=1}^{\infty} (\sum_{k=1}^{n} P_{n-k} f_k) s^n = 1 + \sum_{n=1}^{\infty} \sum_{k=1}^{n} P_{n-k} s^{n-k} f_k s^k$$

$$= 1 + \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} P_{n-k} s^{n-k} f_k s^k = 1 + \sum_{k=1}^{\infty} f_k s^k \sum_{n=0}^{\infty} P_n s^n$$

Markov 連鎖

$|s| < 1$ なので

$$\sum_{k=1}^{\infty} f_k s^k = 1 - \frac{1}{\sum_{n=0}^{\infty} P_n s^n}$$

は正しく $s \to 1$ の極限をとれば $\sum P_n = \infty$ の時にも定理の主張が正しいことが分かる。 $\square$

問題

この極限操作が正しいことを示せ。（正しい理由を説明せよ）

定理 (再帰性と同値類)

$x \leftrightarrow y$ かつ x が再帰的であれば y も再帰的である。

証明

$x \leftrightarrow y$ なので

$$\exists n_1, n_2 \text{ s.t. } P(X_{n_1} = y | X_0 = x) > 0,\ P(X_{n_2} = x | X_0 = y) > 0$$

である。Markov 性から

$$\begin{aligned}
&P(X_{n_2+n+n_1} = y | X_0 = y) \\
&= \sum_{z,w} P(X_{n_2+n+n_1} = y, X_{n+n_1} = z, X_{n_1} = w | X_0 = y) \\
&\geq P(X_{n_2+n+n_1} = y, X_{n+n_1} = x, X_{n_1} = x | X_0 = y) \\
&= P(X_{n_2} = y | X_0 = x) P(X_n = x | X_0 = x) P(X_{n_1} = x | X_0 = y)
\end{aligned}$$

先の定理 (再帰性) を適用する

$$\begin{aligned}
&\sum_{k=0}^{\infty} P(X_k = y | X_0 = y) \\
&\geq \sum_{n=0}^{\infty} P(X_{n_2+n_1+n} = y | X_0 = y) \\
&\geq \sum_{n=0}^{\infty} P(X_{n_2} = y | X_0 = x) P(X_n = x | X_0 = x) P(X_{n_1} = x | X_0 = y) \\
&= \infty
\end{aligned}$$

より y も再帰的である。 $\square$

Markov 連鎖

系 (再帰性と同値類)

$x \leftrightarrow y$ かつ x が非再帰的であれば y も非再帰的である。

問題

定理の主張の対偶をとることでこの系を示せ。

命題 (再帰性と同値類)

x が再帰的でかつ $x \to y$ であれば $y \to x$ である。

証明

$x \to y$ より $\{n; P(X_n = y | X_0 = x) > 0\} \neq \emptyset$ であるから

$$n = \min\{n; P(X_n = y | X_0 = x) > 0\}$$

とする。すなわち $\{x_1, x_2, \ldots, x_{n-1}\} \cap \{x, y\} = \emptyset$ かつ

$$P(X_n = y, X_{n-1} = x_{n-1}, \ldots, X_1 = x_1 | X_0 = x) > 0$$

とできる。

Markov 連鎖

$y \not\to x$ とすると非再帰的であることを示す。
今 $y \not\to x$ とすると $\forall m \geq 1$,

$$P(X_m \neq x, X_{m-1} \neq x, \dots X_1 \neq 1 | X_0 = y) = 1$$

なので $\forall m \geq 1$ に対して

$$\begin{aligned}
&P(X_{n+m} \neq x, X_{n+m-1} \neq x, \dots, X_{n+1} \neq x, \\
&\qquad X_n = y, X_{n-1} = x_{n-1}, \dots, X_1 = x_1 | X_0 = x) \\
&= P(X_{n+m} \neq x, X_{n+m-1} \neq x, \dots, X_{n+1} \neq x | X_n = y) \\
&\qquad P(X_n = y, X_{n-1} = x_{n-1}, \dots, X_1 = x_1 | X_0 = x) \\
&= P(X_n = y, X_{n-1} = x_{n-1}, \dots, X_1 = x_1 | X_0 = x) > 0
\end{aligned}$$

(m によらずに正の値で下から評価できる。)

$F_N = \bigcap_{n=1}^{N}\{X_n \neq x\}$ とおくと F_N は単調減少列で

$$\{T_x = \infty\} = (\bigcup_{n\geq 1}\{X_n = x\})^c = \bigcap_{n\geq 1}\{X_n = x\}^c = \lim F_N$$

従って

$$\begin{aligned}
&P(T_x = \infty, X_n = y | X_0 = x) \\
&= P(\lim_{m\to\infty}\{X_{n+m} \neq x, \ldots, X_{n+1} \neq x, X_n = y\} | X_0 = x) \\
&= \lim_{m\to\infty} P(X_{n+m} \neq x, \ldots, X_{n+1} \neq x, X_n = y | X_0 = x) \\
&\geq P(X_n = y, X_{n-1} = x_{n-1}, \ldots, X_1 = x_1 | X_0 = x) > 0
\end{aligned}$$

となり x が非再帰的になる。 □

Markov 連鎖

系 (再帰性と同値類)

$\exists y$ s.t. $x \to y$ かつ $y \not\to x$ であれば x は非再帰的である。

問題

命題の主張の対偶をとることでこの系を示せ。

Markov 連鎖

まとめ

- x が再帰的と $\sum_n P(X_n = x | X_0 = x) = \infty$ は同値である。
- $x \leftrightarrow y$ であれば x, y 共に再帰的か、共に非再帰的である。
- $x \to y$ かつ x が再帰的であれば $y \to x$ である。
- $\exists y$ s.t. $x \to y$ かつ $y \not\to x$ であれば x は非再帰的である。

実際に適用してみる。
Perron-Frobenius の定理から結論付けられた命題 (確率行列の極限) を使えば P を既約、周期 1 の確率行列とすると

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix}$$

と書け、$\mu > 0$ は唯一の定常分布になる。
すなわち $\exists n_0$ s.t. $\forall n \geq n_0$ ならば $P^n_{x,x} \geq \mu_x/2 > 0$ ととれるため明らかに

$$\sum_n P(X_n = x | X_0 = x) = \sum_n P^n_{x,x} = \infty$$

より全ての x で再帰的になる。

Markov 連鎖

周期が $d \geq 2$ の場合も系として $\exists n$ s.t.

$$P^n = \begin{pmatrix} P_1 & & & \\ & P_2 & & \\ & & \ddots & \\ & & & P_d \end{pmatrix}$$

とブロック行列に分解できて $P_i > 0$ であることを用いれば先の議論と同じように

$$\sum_m P(X_m = x | X_0 = x) = \sum_m P^m_{x,x} \geq \sum_k P^{kn}_{x,x} = \infty$$

より全ての x で再帰的になる。

再帰性に関して

実際に Markov 過程の再帰性、非再帰性の問題で面白い話が出てくるためには状態空間 S を可算集合にする必要がある。
例えば d 次元のランダムウォークは平均が 0 の場合 $d = 1, 2$ で再帰的、$d \geq 3$ で非再帰的になる。

Markov 連鎖

平均再帰時間の問題を考えるが次にあげるランダムナイトの問題 (random knight problem) を取り上げる。
チェスは 8×8 の盤上で行なわれる競技であるが、その一つの駒であるナイトは次のような 8 点に移動可能である (黒丸から白丸へ移動可能)。

	○		○	
○				○
		●		
○				○
	○		○	

このナイトが酔っぱらってランダムに動き出した時の平均再帰時間 $E[T_x|X_0 = x]$ を求めたい。但しナイトの動き方は Markov 過程でその推移法則は盤内の移動可能な点に等確率で移動するものとする。

Markov 連鎖

$E[T_x|X_0=x]$ を定義どおりに $P(T_x=k|X_0=x)$ を求めることで計算をすることはかなり大変でやる気がおきないが、実は次のようになる。

168	112	84	84	
112	84	56	56	
84	56	42	42	
84	56	42	42	

ただし対称性があるため盤の左上の半分だけを記載している。

同様にキング、クイーンの場合

キング (周り 8 箇所へ移動可能)

○	○	○
○	●	○
○	○	○

,

140	84	84	84	
84	$105/2$	$105/2$	$105/26$	
84	$105/2$	$105/2$	$105/2$	
84	$105/2$	$105/2$	$105/2$	

クイーン (縦横斜めどこまででも移動可能)

○		○		○
	○	○	○	
○	○	●	○	○
	○	○	○	
○		○		○

,

$208/3$	$208/3$	$208/3$	$208/3$	
$208/3$	$1456/23$	$1456/23$	$1456/23$	
$208/3$	$1456/23$	$1456/23$	$1456/23$	
$208/3$	$1456/23$	$1456/23$	$1456/23$	

Markov 連鎖

x を固定する (毎に)

$$\mu_y = \mu_y^x = \sum_{n=0}^{\infty} P(X_n = y, T_x > n | X_0 = x)$$

と定義する。
注意：(1) $x \not\to y$ であれば $\mu_y = 0$ である。
(2) $X_n = x$ ならば $T_x \leq n$ なので $n \geq 1$ では
$P(X_n = x, T_x > n | X_0 = x) = 0$ である。従って

$$\mu_x = P(X_0 = x, T_x > 0 | X_0 = x) = 1$$

である。
(3) x を再帰的とすると $x \to y$ であれば $y \to x$ にもなるので x を再帰的とすると μ は $x \leftrightarrow y$ となる同値類の中でだけ正になり得る。

定理 (定常測度)

x が再帰的であればこの μ は定常測度である。

証明

$y \neq x$ とする。Markov 性を用いて

$$\begin{aligned}
&P(X_{n+1} = z, X_n = y, T_x > n | X_0 = x) \\
&= P(X_{n+1} = z, X_n = y, X_{n-1} \neq x, \ldots, X_1 \neq x | X_0 = x) \\
&= P(X_{n+1} = z | X_n = y) P(X_n = y, X_{n-1} \neq x, \ldots, X_1 \neq x | X_0 = x) \\
&= P_{y,z} P(X_n = y, T_x > n | X_0 = x)
\end{aligned}$$

と書ける。

Markov 連鎖

これを $y(\neq x)$ に関して和を取る。ただし $y = x$ とすると $P(X_n = y, T_x > n | X_0 = x) = 0$ になるのでこの項も右辺に加えて

$$
\begin{aligned}
&\sum_{y \neq x} P(X_{n+1} = z, X_n = y, T_x > n | X_0 = x) \\
&= \sum_{y} P_{y,z} P(X_n = y, T_x > n | X_0 = x)
\end{aligned}
$$

この左辺は（$n = 0$, $n \geq 1$ で別々に見る必要があるが）

$$
\begin{aligned}
&P(X_{n+1} = z, T_x > n+1 | X_0 = x), \quad z \neq x \text{ のとき} \\
&P(T_x = n+1 | X_0 = x), \quad z = x \text{ のとき}
\end{aligned}
$$

である。

Markov 連鎖

さらにこれを n に関して和をとる。
$z \neq x$ のとき $P(X_0 = z | X_0 = x) = 0$ なのでこれを加えて左辺の和は

$$\sum_{n=0}^{\infty} P(X_{n+1} = z, T_x > n+1 | X_0 = x)$$
$$= \sum_{n=0}^{\infty} P(X_n = z, T_x > n | X_0 = x) = \mu_z$$

$z = x$ のときは再帰的なので

$$\sum_{n=0}^{\infty} P(T_x = n+1 | X_0 = x) = P(T_x < \infty | X_0 = x) = 1 = \mu_x$$

Markov 連鎖

一方右辺は和の順序交換を行なうことで z によらず

$$\sum_{n=0}^{\infty}\sum_{y} P_{y,z}P(X_n = y, T_x > n|X_0 = x)$$
$$= \sum_{y}\sum_{n=0}^{\infty} P(X_n = y, T_x > n|X_0 = x)P_{y,z} = \sum_{y} \mu_y P_{y,z}$$

従って $\mu = \mu P$ である。 □

定理 (一意性)

P が既約で再帰的であれば定常測度は定数倍を除いてただ一つである。さらに全ての点で正の値をとる。

注意：(1) この定理の再帰的の意味は既約を仮定しているのである点 x で再帰的であれば全ての点で再帰的になるし、ある点 x で非再帰的であればすべての点で非再帰的になる。

Markov 連鎖

注意：(2) 1 次元のランダムウォークの場合定常測度 μ は各 $x \in \mathbb{Z}$ に対して

$$\mu_x = q\mu_{x+1} + p\mu_{x-1}$$

を満たすがこれを 2 項間漸化式として解くとその基底は

$$\mu_x^1 = 1, \quad \mu_x^2 = \left(\frac{p}{q}\right)^x, \quad p \neq q \text{ のとき}$$
$$\mu_x^1 = 1, \quad \mu_x^2 = x, \quad p = q \text{ のとき}$$

となるので $p \neq q$ ではこの 2 つの基底の重ね合わせで $\nu = c_1\mu^1 + c_2\mu^2$, ただし $c_1, c_2 \geq 0$, $(c_1, c_2) \neq 0$ ととれば全て定常測度になるが $p = q$ の場合は $\nu = c_1\mu^1$ ただし $c_1 > 0$ だけが定常測度になる。実際 1 次元のランダムウォークは 1 歩の平均が 0 のとき再帰的でそれ以外では非再帰的になる。

証明

まず x を 1 点固定して作った $\mu = \mu^x$ は P が再帰的であれば定常測度になっている。また $x \to y$ より $\{n; P(X_n = y | X_0 = x) > 0\} \neq \emptyset$ であるから

$$n = n_y = \min\{n; P(X_n = y | X_0 = x) > 0\}$$

と定義すると

$$0 < P(X_n = y | X_0 = x) = P(X_n = y, T_x > n | X_0 = x) \leq \mu_y$$

を満たしている。

定常測度 $\tilde{\nu}$ が与えられた時には μ と等しいことをいうが、実際には $\exists a \in S$, $C > 0$ s.t. $\tilde{\nu} = C\mu^a$ を示す。この議論を $\tilde{\nu} = \mu^y$ として使うと a は任意にとれ、特に $a = x$ としてもよく、結局全ての定常測度は $\mu = \mu^x$ の定数倍であることが分かる。
定常測度 $\tilde{\nu}$ が与えられたとき少なくとも 1 点 $a \in S$ で $\tilde{\nu}_a > 0$ となるので

$$\nu_x = \frac{\tilde{\nu}_x}{\tilde{\nu}_a}$$

とする。当然 ν も定常測度である。この ν が μ^a と等しいことを示す。

Markov 連鎖

ν も定常測度なので $\nu P = \nu$ をみたすので次のように書けるが、特に a だけを特別視して書くと

$$
\begin{aligned}
\nu_z &= \sum_y \nu_y P_{y,z} \\
&= \nu_a P_{a,z} + \sum_{y \neq a} \nu_y P_{y,z}
\end{aligned}
$$

ν の定義から $\nu_a = 1$ なので

$$\nu_z \geq P_{a,z}$$

である。この不等式を上の等式の和の中の ν に代入することで

$$
\begin{aligned}
\nu_z &\geq P_{a,z} + \sum_{y \neq a} P_{a,y} P_{y,z} \\
&= P_{a,z} + P(X_2 = z, X_1 \neq a | X_0 = a) \\
&= P_{a,z} + P(X_2 = z, T_a > 1 | X_0 = a)
\end{aligned}
$$

Markov 連鎖

同様にこの不等式を上の等式の和の中の ν に代入することで

$$
\begin{aligned}
\nu_z &\geq P_{a,z} + \sum_{y \neq a} P_{a,y} P_{y,z} + \sum_{y \neq a} P_{y,z} P(X_2 = y, T_a > 1 | X_0 = a) \\
&= P_{a,z} + P(X_2 = z, X_1 \neq a | X_0 = a) \\
&\qquad + P(X_3 = z, X_2 \neq a, T_a > 1 | X_0 = a) \\
&= P_{a,z} + P(X_2 = z, T_a > 1 | X_0 = a) \\
&\qquad + P(X_3 = z, T_a > 2 | X_0 = a)
\end{aligned}
$$

この代入操作を繰り返すことで、以下を得る。

$$
\nu_z \geq P_{a,z} + \sum_{n=2}^{N} P(X_n = z, T_a > n-1 | X_0 = a), \quad \forall N
$$

特に $z \neq a$ であれば右辺を書き換えることができ

$$\nu_z \geq \sum_{n=0}^{N} P(X_n = z, T_a > n | X_0 = a) \to \mu_z^a, \quad (N \to \infty)$$

である。また $z = a$ では $\nu_a = 1$ であったので

$$\nu_z \geq \mu_z^a, \quad \forall z$$

である。さらに ν が定常測度であるのでこの不等式とあわせて

$$1 = \nu_a = \sum_y \nu_y P_{y,a}^n \geq \sum_y \mu_y^a P_{y,a}^n = \mu_a^a = 1$$

すなわち $P_{y,a}^n > 0$ となる y では $\nu_y = \mu_y^a$ にならなければいけない。既約性よりすべての y に対して $P_{y,a}^n > 0$ となる n があるので $\nu = \mu^a$ である。 $\square$

Markov 連鎖

定理 (定常分布と再帰時間)

P が既約で定常分布 π を持つならば全ての点は再帰的で

$$\pi_x = \frac{1}{E[T_x|X_0 = x]} = \frac{1}{\sum_{n=0}^{\infty} P(T_x > n|X_0 = x)}$$

注意： $\sum_x \mu_x < \infty$ となる定常測度があれば

$$\pi_x = \frac{\mu_x}{\sum_y \mu_y}$$

とすることで定常分布を持つ。

証明
定理の最終式の 2 つめの等式は次のような一般的な事実による。

問題

以下を示せ
Y を非負整数値を取る確率変数とする時

$$E[Y] = \sum_{n=0}^{\infty} P(Y > n)$$

($\infty = \infty$ も許す。)

Markov 連鎖

再帰性について
π は少なくとも 1 点 x で $\pi_x > 0$ になる。また π は定常分布なので

$$\pi_z = (\pi P^n)_z = \sum_y \pi_y P(X_n = z | X_0 = y) \geq \pi_x P(X_n = z | X_0 = x)$$

と書けるが既約性より $P(X_n = z | X_0 = x) > 0$ となる n が存在する。従って $\pi > 0$ である。
$\pi_z = \sum_y \pi_y P^n_{y,z}$ の両辺を n に関して和を取ると $\pi > 0$ より

$$\infty = \sum_{n=0}^{\infty} \pi_z = \sum_{n=0}^{\infty} \sum_y \pi_y P^n_{y,z} = \sum_y \pi_y \sum_{n=0}^{\infty} P^n_{y,z}$$

Markov 連鎖

ここで $P^n_{y,z} = P(X_n = z | X_0 = y)$ を Markov 過程がはじめて z に到達した時間で分解することで

$$P^n_{y,z} = P(X_n = z | X_0 = y) = \sum_{k=1}^{n} P(X_n = z, T_z = k | X_0 = y)$$

$$= \sum_{k=1}^{n} P(X_n = z, X_k = z, T_z = k | X_0 = y)$$

$$= \sum_{k=1}^{n} P(X_n = z | X_k = z) P(T_z = k | X_0 = y)$$

$$= \sum_{k=1}^{n} P^{n-k}_{z,z} P(T_z = k | X_0 = y)$$

と書けるためこれを使う。

和の順序交換を行なうことで

$$\sum_{n=0}^{\infty} P_{y,z}^{n} = \sum_{n=0}^{\infty}\sum_{k=1}^{n} P_{z,z}^{n-k} P(T_z = k | X_0 = y)$$
$$= \sum_{l=0}^{\infty} P_{z,z}^{l} \sum_{k=1}^{\infty} P(T_z = k | X_0 = y)$$
$$= \sum_{l=0}^{\infty} P_{z,z}^{l} P(T_z < \infty | X_0 = y) \leq \sum_{l=0}^{\infty} P_{z,z}^{l}$$

$\sum_y \pi_y = 1$ とあわせて

$$\infty = \sum_{y} \pi_y \sum_{n=0}^{\infty} P_{y,z}^{n} \leq \sum_{y} \pi_y \sum_{l=0}^{\infty} P_{z,z}^{l} \leq \sum_{l=0}^{\infty} P_{z,z}^{l}$$

より z の再帰性が分かる。

最後に

$$\pi_x = \frac{1}{\sum_{n=0}^{\infty} P(T_x > n | X_0 = x)}$$

を示す。
先の定理 (定常測度)、(一意性) を用いれば各 x に対して定めた

$$\mu_y^x = \sum_{n=0}^{\infty} P(X_n = y, T_x > n | X_0 = x)$$

と、π を含めても定数倍を除いて等しいので

$$\pi = \frac{1}{\sum_y \mu_y^x} \mu^x, \quad \forall x$$

と書ける。

Markov 連鎖

$\mu_x^x = 1$ であったことと和の交換により

$$\sum_y \mu_y^x = \sum_y \sum_{n=0}^{\infty} P(X_n = y, T_x > n | X_0 = x)$$
$$= \sum_{n=0}^{\infty} \sum_y P(X_n = y, T_x > n | X_0 = x)$$
$$= \sum_{n=0}^{\infty} P(T_x > n | X_0 = x)$$

より

$$\pi_x = \frac{1}{\sum_{n=0}^{\infty} P(T_x > n | X_0 = x)}$$

を得る。 □

Markov 連鎖

この定理 (定常分布と再帰時間) を適用すれば定常分布を求めれば その逆数として平均再帰時間が求められる。
Perron-Frobenius の定理を適用すれば (状態空間が有限集合で) 考えている推移確率行列 P が既約であれば定常分布はただ 1 つしかないことが分かり、$\mu P = \mu$ の自明でない解 μ を求めればよいことは分かる。
一方で現実的には、例えばランダムナイトの問題では P は (かなりの数の成分は 0 であるが) 64×64 の行列になり (対称性をうまく使えば 10×10 の行列に減らせるがそれでも)、このサイズの行列の計算を行なう気にはなかなかなれない。(P のかなりの数の成分は 0 になると書いたが、基本変形を繰り返して行くうちに 0 でない成分だらけになることは簡単に予想できる。)
ランダムナイトの例が簡単に計算できるのは次のような理由による。

Markov 連鎖

定義 (グラフ上のランダムウォーク)

Markov 過程の推移確率行列 P が以下を満たすときグラフ上のランダムウォークと呼ぶ。

$P_{x,x} = 0$

$P_{x,y} > 0$ ならば $P_{y,x} > 0$

$x, y, z \in S$ に対して $P_{x,y} > 0$ かつ $P_{x,z} > 0$ ならば $P_{x,y} = P_{x,z}$

注意：これは次のように考える。
P が与えられたとき、状態空間 S を頂点集合とする。(向き、重みをつけない) 辺の集合 E を $P_{x,y} > 0$ (このとき $P_{y,x} > 0$ にもなる) となる (x, y) の組全体とすると自然にグラフ $G = (S, E)$ が定義できる。
この Markov 過程はこのグラフ上を辺にそってランダムに移動するがどの辺を選択するかは現在いる頂点から出ている辺を等確率で選択する。

Markov 連鎖

グラフ上のランダムウォークの推移確率行列 P が与えられた時に $\mu = (\mu_1, \ldots, \mu_n)$ を $\mu_x = \#\{y; P_{x,y} > 0\}$ で与える。
注意：各成分 μ_x はグラフの言葉で書くと x の次数である。

定理 (グラフ上のランダムウォーク)

P はこの μ に対して対称である。
従って μ は定常測度になる。

注意：定常測度が分かれば

$$\pi = \frac{1}{\sum_x \mu_x}\mu$$

と規格化することで定常分布を得られるが、この分母に表れる $\sum_x \mu_x$ はグラフの言葉で書くと (数え方によるが向き無しで考えているので) 辺の総数の 2 倍になる。

証明

x, y によらず $\mu_x P_{x,y} = 1$ である。 □

Markov 連鎖

ランダムナイト、ランダムキング、ランダムクイーンはすべてこのグラフ上のランダムウォークになるため計算は容易で、先にあげたように計算できた。
一方ランダム金、ランダム銀は $P_{x,y} > 0$ ならば $P_{y,x} > 0$ を満たさないため簡単には計算できない。
次のページは数値計算の結果 9×9 で行なっているが右半分が入っていない（対称性がある)。
（81×81 行列の計算を行なっている。）

金 (周り 8 箇所中斜め後ろ以外へ移動可能)

○	○	○
○	●	○
	○	

30.791328	17.815554	17.111911	16.881724	16.8248
35.324925	24.915491	25.076442	25.071163	25.0627
85.940838	67.42103	71.810424	73.470653	73.8932
215.529198	179.831729	201.712943	212.137365	215.112
550.125457	476.83005	558.232434	603.699589	617.821
1419.270985	1261.152818	1527.8363	1696.096853	1752.161
3681.111795	3328.221102	4149.011752	4722.384145	4924.695
9499.054514	8719.120537	11269.92047	13163.67467	13854.67
29920.38808	28154.66878	38005.94435	45254.73054	47919.13

銀 (周り 8 箇所中、横、後ろ以外へ移動可能)

○	○	○
	●	
○		○

67.813349	45.504115	44.201301	45.805548	45.524978
49.629111	24.907000	27.663759	27.158334	27.633778
72.368415	43.904072	40.570712	42.020332	41.348125
121.449120	65.709962	65.736386	63.129401	64.097269
185.413674	106.341529	99.792808	99.383662	97.215698
296.189600	163.363762	157.864619	151.766057	152.140169
461.498716	258.665768	242.750465	236.878209	232.707452
742.552272	405.113035	376.585647	361.925390	359.450153
2025.565173	1020.418715	955.758462	919.546091	904.813474

Markov 連鎖

スペクトルギャップは名前は仰々しいが実際は確率行列の最大固有値と次に大きい固有値の差である。
これが意味を持つためには例えば全ての固有値が実数である必要がある。
この講義では P はある μ に対して対称であることを仮定する。
注意：対称でない場合も P にたいして定常分布 μ を出してこの μ に対して $P^* = P^*_\mu$, $P_s := \frac{1}{2}(P + P^*)$ と定義することで P_s は μ 対称でかつ確率行列になるのでこの P_s に関するスペクトルギャップを考えることは可能。
スペクトルギャップの評価を行うことが多いが、基本的にはその後で使いたいこと、例えば流体力学極限の証明など、が控えているために行なっているが、もっと単純に次のようなことを考える。

Markov 連鎖

Markov 過程 X_n は物理現象に対応しているものとすると、その物理現象に対応した各時刻での物理量 (実数値になるものとする) は関数 $f : S \to \mathbb{R}$ を用いて $f(X_n)$ と表せるが、実際は時刻 0 での X_0 の値が決まっているとは考えにくくある分布 ν (定常分布ではないものを想定する) に従っていると考えることが自然で、その平均量 $E[f(X_n)]$ を観測していると思う方が自然である。
いままでの話から $E[f(X_n)] \to E[f], (n \to \infty)$ と収束していくことはほぼ明らかで収束値の $E[f]$ と書いたものは定常分布 μ による平均であることもほぼ明らかである。
ここで実際に見たいものは極限へ収束していく速さである。

Markov 連鎖

推移確率行列 P は μ 対称な確率行列なので全ての固有値は実数になる。従って

$$Q^{-1}PQ = \Lambda, \qquad P = Q\Lambda Q^{-1}$$

$$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}, \quad Q^{-1} = \begin{pmatrix} \mu^1 \\ \vdots \\ \mu^n \end{pmatrix}, \quad Q = \left(q^1 \cdots q^n\right)$$

と対角化可能であるが、$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ と順番になっていると思ってよい。また既約性を仮定すれば Perron-Frobenius の定理より

$$\lambda_1 = 1 > \lambda_2 \geq \ldots \geq \lambda_n \geq -1$$

であり ${}^t q^1 = {}^t 1 = (1, \ldots, 1)$, μ^1 を定常分布ととれる。

Markov 連鎖

状態空間を $S = \{x_1, x_2, \ldots, x_n\}$ とし、S 上の関数 f に対してベクトル ${}^t f = (f_1, \ldots, f_n)$ を $f_i = f(x_i)$ とし $\nu = (\nu_1, \ldots, \nu_n)$ に対して

$$E[f(X_m)] := \sum_{x,y} \nu_x P(X_m = y | X_0 = x) f(y)$$

は前ページの設定に加えて $n \times n$ 行列で i, i 成分のみ 1, 他は 0 となる行列を E_i とすると、以下のように書き換えることが可能である

$$\begin{aligned} E[f(X_m)] &= \sum_{x,y} \nu_x P(X_m = y | X_0 = x) f(y) \\ &= \nu\{Q\Lambda^m Q^{-1}\} f = \sum_{i=1}^{n} \nu\{Q\lambda_i^m E_i Q^{-1}\} f \\ &= \sum_{i=1}^{n} \lambda^m (\nu q^i)(\mu^i f) \end{aligned}$$

Markov 連鎖

特に μ^1 は定常分布なので $(\mu^1 f) = E_{\mu^1}[f]$、$q^1 = 1$ で ν は確率測度であるので $(\nu q^i) = 1$ になることから

$$E[f(X_m)] = E_{\mu^1}[f] + \sum_{i=2}^{n} \lambda_i^m (\nu q^i)(\mu^i f)$$

と書け $(\nu q^i)(\mu^i f)$ は m に依存しないので ($\lambda_n \neq -1$ ならば) $E[f(X_m)]$ は $E_{\mu^1}[f]$ へ収束し、その収束の速さは $\max\{|\lambda_2|, |\lambda_n|\}$ で決まりこの値が 1 にならなければ指数的に速く減衰する。

P をそのまま見ると ($\lambda_n \neq -1$ でも) 収束の速さを知るには λ_2, λ_n の 2 つの量を知る必要がある。
これは不便であるので lazy Markov 過程もしくは連続時間 Markov 過程を次のように考える。

定義 (lazy, 連続 Markov 過程)

P を確率行列とする。P に対応する lazy Markov 過程とは推移確率行列が $\frac{1}{2}(P+I)$ で与えられるものとし、対応する連続時間 Markov 過程とは $P^t_{x,y} := (\exp\{t(P-I)\})_{x,y}$ と与えた時に $t_i > 0$ に対して $T_i = \sum_{j=1}^{i} t_i$ と書く時に

$$
\begin{aligned}
&P(X_{T_n} = x_n, X_{T_{n-1}} = x_{n-1}, \dots, X_{T_1} = x_1 | X_0 = x_0) \\
&= P^{t_1}_{x_0,x_1} P^{t_2}_{x_1,x_2}, \cdots, P^{t_n}_{x_{n-1},x_n}
\end{aligned}
$$

と与えられるものとする。

Markov 連鎖

lazy Markov 過程は元々の Markov 過程が与えられたときに各時間ステップで、さらにコイン投げをして表がでたら元々の Markov 過程に従って動き、裏が出たらその場に留まると思えばよい。このため lazy 怠け者という名前になっている。

Markov 連鎖

問題

A は対角化可能で

$$Q^{-1}AQ = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

と書けたとする。
このとき $(c_1A + c_2I)$, は同じ Q を用いて対角化可能で

$$Q^{-1}(c_1A + c_2I)Q = \begin{pmatrix} c_1\lambda_1 + c_1 & & \\ & \ddots & \\ & & c_1\lambda_n + c_2 \end{pmatrix}$$

であることを示せ。

Markov 連鎖

問題

A は対角化可能で

$$Q^{-1}AQ = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

と書けたとする。
このとき $\exp tA$, は同じ Q を用いて対角化可能で

$$Q^{-1}\exp\{tA\}Q = \begin{pmatrix} \exp(t\lambda_1) & & \\ & \ddots & \\ & & \exp(t\lambda_n) \end{pmatrix}$$

であることを示せ。

Markov 連鎖

これらの事実から lazy Markov 過程、もしくは連続時間 Markov 過程を考えることにより λ_2、もしくは $\lambda_1 - \lambda_2 = 1 - \lambda_2$ を考えればよくなる。
この講義では (連続時間 Markov 過程に対応させて) $L := P - I$ とおく。このとき問題から L の最大固有値が 0 であることが分かり、Perron-Frobenius の定理からこれ以外の固有値は真に負であることが分かる。

定義 (スペクトルギャップ)

確率行列 P が与えられたとき $L = P - I$ の第 2(最大) 固有値の絶対値をスペクトルギャップと呼ぶ。また L, P をあまり区別せず、同じものになるので L のスペクトルギャップ、P のスペクトルギャップ等と呼ぶ。

注意：L の固有値は真に負なので次のように呼んでもよい
確率行列 P が与えられたとき $-L = -(P - I)$ の第 2(最小) 固有値をスペクトルギャップと呼ぶ。

Markov 連鎖

通常行列の設定で連続時間化をする場合 L ではなく記号 Q を用いる。またここで考えている行列 $P-I$ は Q-行列と呼ばれる。
一方で次のように思い直すことにする。
状態空間 $S=\{x_1,x_2,\dots,x_n\}$ から実数への関数全体の空間は $n=\#S$ 次元の線形空間になるがこの基底 $f_1,f_2,\dots,f_n$ として

$$f_i(x)=\begin{cases}1 & x=x_i\\ 0 & x\neq x_i\end{cases}$$

ととることができる。
また一方推移確率行列 P を用いてこの関数空間上の線形作用素 L を

$$Lf(x)=\sum_y P_{x,y}(f(y)-f(x))=\sum_y P_{x,y}f(y)-f(x)$$

と定めるとこの L の 基底 $f_1,f_2,\dots,f_n$ による表現行列を考えると実は $P-I$ になっている。

Markov 連鎖

さらに同じ空間での線形作用素 L のうち

$$Lf(x) = \sum_y C_{x,y}(f(y) - f(x))$$

の形で書け、$C_{x,y} \geq 0$ と与えられるものを考える。ただし $C_{x,x} = 0, \forall x$ とする。

問題

$a = \max_x \sum_y C_{x,y}$ としたときにこの $\frac{1}{a}L$ の 基底 $f_1, f_2, \ldots, f_n$ による表現行列を $P - I$ と書くと

$$P_{x,y} = \begin{cases} \frac{1}{a} C_{x,y} & x \neq y \\ 1 - \frac{1}{a}\sum_z C_{x,z} & x = y \end{cases}$$

であることを示せ

Markov 連鎖

(この線形作用素 L と表現行列 $L = a(P - I)$ を同一視して同じ記号で書くが) この L に対して $\exp(tL)$ を考えるとこれから連続時間の Markov 過程が考えられ a はスピードアップ (スピードダウンもあり得る) に対応するパラメータと思えばここまで拡張して考えた方が自然である。
また解析的には行列よりも線形作用素の方がより取り扱いやすいのでこの講義では線形作用素を用いて記述することにする。
ただしいつでも表現行列を通して行列に戻れることは注意する。
またなるべく行列での表記も残すことにする。

Markov 連鎖

確率行列 P に関しては既約を定義したが、自然に拡張して

定義 (既約)

$Lf(x) = \sum_y C_{x,y}(f(y) - f(x))$ で与えられる L が既約であるとは $\forall x, y,\ \exists n, x = x_0, x_1, \ldots, x_n$ s.t. $C_{x_{i-1},x_i} > 0,\ 1 \leq \forall y \leq n$ である。

以後 L は既約であると仮定する。

問題

先に与えた自然な基底 $f_1, \ldots, f_n$ に対して表現行列を考えて $L = a(P - I)$ と書いたとき L が既約であることと P が既約であることは同値であることを示せ。

問題

P を確率行列として対応する定常分布を $\mu = (\mu_1, \ldots, \mu_n)$ とする。このとき $Lf(x) = \sum_y P_{x,y}(f(y) - f(x))$ と定義した L に対して $E_\mu[f(X)] = \sum_x \mu_x f(x)$ と定義すると

$$E_\mu[Lf(X)] = 0, \quad \forall f$$

を示せ。
逆に $Lf(x) = \sum_y C_{x,y}(f(y) - f(x))$ で与えられる L から自然な基底 $f_1, \ldots, f_n$ に対して表現行列を考えて $L = a(P - I)$ と書く。分布 $\mu = (\mu_1, \ldots, \mu_n)$ が

$$E_\mu[Lf(X)] = 0, \quad \forall f$$

を満たせばこの μ は P の定常分布になることを示せ。

Markov 連鎖

定義 (μ 対称)

$Lf(x) = \sum_y C_{x,y}(f(y) - f(x))$ で与えられる L が μ に関して対称であるとは $\forall x, y,\ \mu_x C_{x,y} = \mu_y C_{y,x}$ を満たすことである。

問題

L が μ 対称ならば任意の関数 f, g に対して

$$E_\mu[fLg] = E_\mu[gLf]$$

を示せ。

Markov 連鎖

問題

P を確率行列として μ 対称とする。このとき $Lf(x)=\sum_y P_{x,y}(f(y)-f(x))$ と定義した L に対して L も μ 対称になることを示せ。
逆に $Lf(x)=\sum_y C_{x,y}(f(y)-f(x))$ で与えられる L から自然な基底 $f_1,\dots,f_n$ に対して表現行列を考えて $L=a(P-I)$ と書く。L が μ 対称であれば P も μ 対称であることを示せ。

Markov 連鎖

問題

P を確率行列として μ を定常分布とする。(対称ではないものとする。) これに対応する L ($Lf(x) = \sum_y P_{x,y}(f(y) - f(x))$) と、$\mu$ による重み付き内積 $\langle \cdot, \cdot \rangle_\mu$ を用いて P^* を定義したがこの P^* に対応する L^* ($L^* f(x) = \sum_y P^*_{x,y}(f(y) - f(x))$) は実際に共役の関係

$$E_\mu[fLg] = E_\mu[gL^* f], \quad \forall f, g$$

を満たすことを示せ。

問題

先の問題の逆で $Lf(x) = \sum_y C_{x,y}(f(y) - f(x))$ で与えられる L から確率行列 P を出し、この定常分布を μ とする。
(先にあげた問題から $E\mu[f] = 0, \forall f$ を満たす分布 μ としてもよい。) (対称ではないものとする。) これに対して L^* を $E_\mu[fLg] = E_\mu[gL^*f], \forall f, g$ を満たすように定義する。このとき L^* も $L^*f(x) = \sum_y C^*_{x,y}(f(y) - f(x))$ の形で与えられ $C^*_{x,y} \geq 0$ であることを示せ。
またこの L^* から与えられる確率行列 $\tilde{P}^*$ は P^* と一致するか？

自然に f の平均、分散を

$$E[f] = E_\mu[f] = \sum_x f(x)\mu_x, \quad V[f] = V_\mu[f] = E_\mu[(f - E_\mu[f])^2]$$

と定義する。さらに μ に対称な L に対して Dirichlet 形式 D を 2 次形式として

$$D[f] = D_\mu[f] = E_\mu[f(-L)f]$$

と定義する。

問題

$Lf(x) = \sum_y C_{x,y}(f(y) - f(x))$ で与えられる L が μ に関して対称であるときに

$$D[f] = \frac{1}{2}\sum_{x,y} \mu_x[C_{x,y}(f(y) - f(x))^2]$$

と書けることを示せ。

Markov 連鎖

これらを行列の言葉で書くには重み付き内積 $\langle \cdot, \cdot \rangle_\mu$ を使い、関数 f から自然にベクトル $f = {}^t(f_1, \ldots, f_n)$ を $f_1 = f(x_1)$ で書けば

$$E[f] = \langle 1, f \rangle_\mu = \langle f, 1 \rangle_\mu, \quad E[f^2] = \langle f, f \rangle_\mu,$$
$$V[f] = \langle f - E[f], f - E[f] \rangle_\mu, \quad D[f] = \langle f, (-L) f \rangle_\mu$$

と書ける。ただし $1 = {}^t(1, \ldots, 1)$ であり L も表現行列を通して行列になっているものとしている。

命題 (スペクトルギャップの変分表現)

L のスペクトルギャップ λ は

$$\lambda = \inf \left\{ \frac{D[f]}{V[f]} \middle| E[f] = 0 \right\}$$

で与えられる。

証明

行列表現を用いると、

$$E[f] = \langle 1, f \rangle_{\mu} = \mu f$$

と書ける。ベクトル μ は P の最大固有値 1 の左固有ベクトルであり $-L$ の 最小固有値 0 に対する左固有ベクトルでもある。すなわち $\mu f = E[f] = 0$ を満たすということは f は $-L$ の最小固有値に対する固有ベクトルの直交補空間に属していることを意味する。従ってこの inf を使って書き表される λ は第 2 固有値を表している。 □

Markov 連鎖

スペクトルギャップに関していろいろと書いてきたが、実際に L が与えられた時にスペクトルギャップを正確に求めることは非常に困難である。
確率過程のスケール極限を考える上では (状態空間も変化するが) 推移確率行列の族 $\{P_n\}_n$ が与えられている、もしくは線形作用素の族 $\{L_n\}_n$ が与えられている時にこれらのスペクトルギャップのオーダーの評価が必要になる。
本講義ではこの路線に沿った方向の解析を行なう。

Markov 連鎖

命題 (ポアンカレの不等式)

$B(r)$ を $\mathbb{R}^d$ の半径 r の球とする。このとき次元 d だけに依存する定数 $C = C_d$ が存在して任意の $u \in C^1(B(r))$ に対して

$$\int_{B(r)} \left| u(x) - \frac{1}{|B(r)|} \int_{B(r)} u(y)dy \right|^2 dx \leq Cr^2 \int_{B(r)} |Du(x)|^2 dx$$

が成立する。

問題

この命題を証明せよ。($d = 1$ だけでもよい)

Markov 連鎖

簡単のため $d = 1$ とし、$B(r) = [-r, r]$ とする。この命題 (ポアンカレの不等式) において $u \in C^2$ とすると部分積分を用いることで

$$\int_{-r}^{r} |u'(x)|^2 dx = [u(x)u'(x)]_{x=-r}^{r} - \int_{-r}^{r} u(x)u''(x)dx$$

と書けるので端点 $r, -r$ で $u(x)$ もしくは $u'(x)$ が無視できる状況 (ディリクレ、ノイマン境界条件) の下ではポアンカレの不等式は

$$\int_{B(r)} \left| u(x) - \frac{1}{|B(r)|} \int_{B(r)} u(y)dy \right|^2 dx \le Cr^2 \int_{B(r)} u(x)(-\Delta u)(x)dx$$

と書いても構わない。

Markov 連鎖

確率論では ($B(r)$ 上の) ブラウン運動は重要であるが、元の Markov 過程の話と対応させると定常分布が $B(r)$ 上の一様分布、$L = \Delta$ (もしくは $L = \frac{1}{2}\Delta$) を生成作用素とすることが知られている。
先のページの最後に与えた不等式を両辺 $|B(r)|$ で割ると左辺側が $V[u]$、右辺側が $Cr^2 D[u]$ に対応しているすなわち

$$V[u] \leq Cr^2 D[u]$$

をいつでも満たしていることが分かる。これをスペクトルギャップの変分表現に代入することで

$$\lambda = \inf\{\frac{D[f]}{V[f]} | E[f] = 0\} \geq \frac{1}{Cr^2}$$

すなわちスペクトルギャップの下からの評価を与えたことになる。

Markov 連鎖

一方再び変分表現で $f(x) = x^3$ を代入することで

$$\lambda \leq \frac{D[x^3]}{V[x^3]} \leq \frac{\frac{12}{5}r^5}{\frac{2}{7}r^7} \leq \frac{C'}{r^2}$$

と同じ $1/r^2$ のオーダーで与えられることが分かる。
またブラウン運動のスケーリング法則として

命題 (ブラウン運動のスケーリング)

$B = (B_t)_{t\geq 0}$ を $B_0 = 0$ とするブラウン運動とすると、有限次元分布の意味で

$$(B_t)_t = (\alpha B_{\alpha^2 t})_t$$

が良く知られているがこの時間と空間のスケーリング (α^2, α) はスペクトルギャップのオーダーから決まっていると思える。(他の確率過程を時間と空間のスケーリングをした時にブラウン運動へ収束することが多いが、このスケーリングの関係がスペクトルギャップのオーダーから決まっていると思える。)

Markov 連鎖

命題 (スペクトルギャップの変分表現) があるので、スペクトルギャップの上からの評価は比較的容易にできる。(多くの場合は f の候補がすぐにあがる。)
下からの評価は一般的には難しいが、ポアンカレの不等式の形で

$$V[f] \leq CD[f]$$

と書いた時に全ての f に対してこの不等式を満たす最小の C こそがスペクトルギャップの逆数になり、$-\Delta$ に対するポアンカレの不等式の証明のように Cauchy Schwarz の不等式と Dirichlet 形式の比較をうまく利用することで (定数はうまく出ないが) 証明できることが多い。

問題

X, Y を独立同分布の確率変数とする。このとき

$$V[f] = V[f(X)] = \frac{1}{2}E[(f(X) - f(Y))^2]$$

を示せ。

完全グラフ上のランダムウォーク
$S=\{1,2,\dots,n\}$ として推移確率行列 P を

$$P_{i,j}=\frac{1}{n},\quad \forall i,j$$

とおくと $\mu=(\mu_1,\dots,\mu_n)$ を $\mu_i=1/n, \forall i$ としたものが定常分布になり明らかに P は μ 対称である。なおこの推移確率行列で与えられる Markov 連鎖は完全グラフ上のランダムウォークと考えられる。

命題 (完全グラフ上の RW のスペクトルギャップ)

上で与えた P のスペクトルギャップは 1 である。

証明

先の問題を使えば

$$V[f] = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(f(i) - f(j))^2\frac{1}{n^2}$$

$$= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{n}(f(i) - f(j))^2\frac{1}{n} = D[f]$$

よりスペクトルギャップは 1 になる。 $\square$

注意：もっと単純に P の固有値は 1 と 0 で 0 が $(n-1)$ 重根になっていることがすぐに分かる。

Markov 連鎖

ランダムウォーク
S を変えずに推移確率行列を

$$P_{i,j} = \begin{cases} \frac{1}{2} & |i-j| = 1, i = j = 1 \text{ または } i = j = n \\ 0 & \text{その他} \end{cases}$$

とおくと先ほどと同様に $\mu_i = 1/n, \forall i$ としたものが定常分布になり明らかに P は μ 対称である。なおこの推移確率行列は (通常の)1 次元ランダムウォークと考えられる。

命題 (完全グラフ上の RW のスペクトルギャップ)

上で与えた P のスペクトルギャップは $O(1/n^2)$ である。

注意：実際は第 2 固有関数は三角関数 $\cos(\frac{\pi}{n-1}(x-1))$ と分かっているので定数まで込めて計算可能

Markov 連鎖

証明

上からの評価

$f(x) = x$ とすれば

$$E[f] = \frac{n+1}{2}, \quad E[f^2] = \frac{(n+1)(2n+1)}{6},$$

$$D[f] = \sum_{i,j} P_{i,j}(f(j) - f(i))^2 \mu_i = \sum_{i=2}^{n} 1^2 \frac{1}{n} = \frac{n-1}{n}$$

従って

$$\lambda \le \frac{D[f]}{V[f]} = \frac{12}{n^2} + O(1/n^3)$$

下からの評価
完全グラフの時と同じ式からはじめる。すなわち (少し和の書き方を変えてあるが)

$$V[f] = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(f(i)-f(j))^2\frac{1}{n^2}$$

$$= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{1}{n}(f(j)-f(i))^2\frac{1}{n}$$

であるが $j \geq i$ なので

$$f(j)-f(i) = \sum_{k=i+1}^{j} f(k)-f(k-1)$$

と書ける。

Markov 連鎖

さらにこれに Cauchy Schwarz の不等式を適用することで

$$\{f(j)-f(i)\}^2 \le (j-i)\sum_{k=i+1}^{j}\{f(k)-f(k-1)\}^2$$

を代入し、和の順序交換をすることで

$$V[f] \le \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{1}{n}(j-i)\sum_{k=i+1}^{j}\{f(k)-f(k-1)\}^2\frac{1}{n}$$

$$=\sum_{k=2}^{n}\{f(k)-f(k-1)\}^2\sum_{i=1}^{k}\sum_{j=k+1}^{n}(j-i)\frac{1}{n}\frac{1}{n}$$

$$\le\sum_{k=2}^{n}\{f(k)-f(k-1)\}^2\sum_{i=1}^{n}\sum_{j=1}^{n}(n)\frac{1}{n}\frac{1}{n}=n^2D[f]$$

Markov 連鎖

全ての関数に対して成り立っているので

$$\lambda \geq \frac{1}{n^2}$$

が分かる。 $\square$

Markov 連鎖

下からの評価に関して

下からの評価に関しては、分散 $V[f]$ をうまく式変形して行き Dirichlet 形式 $D[f]$ で評価したと見えるが、次のように 2 つの部分に分けて考える。
(1) 完全グラフのスペクトルギャップの (下からの) 評価
(2) 完全グラフのランダムウォークに対応する Dirichlet 形式 $D_m[f]$ と本来考えているランダムウォークに対応する Dirichlet 形式 $D[f]$ の比較、評価

このように考えると (ある程度) 見通しはよい。一方で今回の場合は (1) が等号で示せたのでよかったが、多くの場合 (1) (2) 共に Cauchy Schwarz の不等式を (本質的に) 適用していることになり、等号成立条件の関数が (1) (2) で異なってしまい、結果的にオーダーの意味でもよい評価が出せないことはあり得る。

出生死亡過程
ランダムウォークを一般化する。S を変えずに L を

$$Lf(i) = b_i\{f(i+1) - f(i)\} + d_i\{f(i-1) - f(i)\}$$

とする。但し $d_1 = 0, b_n = 0$ でそれ以外は $b_i > 0, d_i > 0$ とする。名前から想像されるように、(ある生物種の) 人口 (個体数) だけを観察する。このモデル化として、(1) 増減は 1 単位づつ (2) 増減のおこりやすさは現在の人口に依存するこの 2 つの性質を持つ Markov 過程はこの生成作用素 L を用いて書ける。
なお今回は $S = \{1, 2, \ldots, n\}$ としたが $S = \{0, 1, 2, \ldots\}$ のように絶滅状態 0 を加えると同時に人口はいくらでも大きくなれ、絶滅状態からは人口は増えない、すなわち $b_0 = 0$ を設定にするのがもともとのモデル。（このため既約にはならない。）

Markov 連鎖

$\tilde{\mu} = (\tilde{\mu}_1, \ldots, \tilde{\mu}_n)$ を漸化的に

$$\tilde{\mu}_1 = 1, \qquad \tilde{\mu}_{x+1} = \frac{b_x}{d_{x+1}} \mu_x$$

とおくと明らかに L は $\tilde{\mu}$ に関して対称になる。$Z = \sum_i \tilde{\mu}_i$ とおき $\mu = \frac{1}{Z}\tilde{\mu}$ と規格化することで確率分布 μ とする。

この L のスペクトルギャップの下からの評価を行うが、まず、見通しがよくなるように完全グラフ (平均場近似) に対応するような Dirichlet 形式 D_m を導入して二段階で評価をしてみる。先ほどのランダムウォークの場合と同じように分散 $V[f]$ を式変形すると、

$$V[f] = \frac{1}{2}\sum_{i,j}\{f(j) - f(i)\}^2 \mu_i \mu_j$$

この右辺を $D_m[f]$ と思うことにすると対応する L_m は

$$L_m f(i) = \sum_j \mu_j \{f(j) - f(i)\}$$

と書ける。

この L_m に対応する D_m ともともともとの D との比較をするが次のように式変形をする。

$$D_m[f] = \sum_{i<j} \{f(j) - f(i)\}^2 \mu_i \mu_j$$
$$= \sum_{i<j} \{ \sum_{k=i+1}^{j} (f(k) - f(k-1)) \sqrt{a_k} \frac{1}{\sqrt{a_k}} \}^2 \mu_i \mu_j$$

ここで $\{a_k; k\}$ を導入したがどのようにとるかは後にしてここで意味を持つように $a_k > 0$ だけを仮定する。
Cauchy Schwarz の不等式を適用すると次のように書ける。

$$D_m[f] \leq \sum_{i<j} \mu_i \mu_j \{ \sum_{k=i+1}^{j} (f(k) - f(k-1))^2 a_k \} \{ \sum_{k=i+1}^{j} \frac{1}{a_k} \}$$

ここで $c_{i,j} = \sum_{k=i+1}^{j} \frac{1}{a_k}$ とおき、和の順序交換を行なうことで

$$D_m[f] \le \sum_{i<j} \mu_i \mu_j \{ \sum_{k=i+1}^{j} (f(k) - f(k-1))^2 a_k \} c_{i,j}$$

$$= \sum_{k=2}^{n} (f(k) - f(k-1))^2 a_k \sum_{i=1}^{k-1} \sum_{j=k}^{n} c_{i,j} \mu_i \mu_j$$

と書ける。この式の最後に表れる和の部分は k だけに依存するので A_k と書くことにし、さらにその k に関する最大値を A と書くことにする。特に $a_k = d_k \mu_k$ とおくと

$$A_k := \sum_{i=1}^{k-1} \sum_{j=k}^{n} c_{i,j} \mu_i \mu_j = \sum_{i=1}^{k-1} \sum_{j=k}^{n} \sum_{l=i+1}^{j} \frac{1}{d_l \mu_l} \mu_i \mu_j, \quad A = \max_k A_k$$

と書ける。

Markov 連鎖

上の $D_m[f]$ の右辺に $a_k = d_k \mu_k$ を代入することで

$$
\begin{aligned}
D_m[f] &\le \sum_{k=2}^{n} (f(k) - f(k-1))^2 d_k \mu_k A_k \\
&\le A \sum_{k=2}^{n} (f(k) - f(k-1))^2 d_k \mu_k = A D[f]
\end{aligned}
$$

と書けるので全てをあわせるとこの $1/A$ は L のスペクトルギャップの下からの評価になっている。

一方で上からの評価はスペクトルギャップの変分表現から関数 f を持ってくることにするが、次のような方針で関数を求める。
下からの評価を行った式変形では本質的に 2 回しか不等式を用いていない。順番に
(1) Cauchy Schwarz の不等式
(2) A_k をその最大値 A で置き換える
であるが、(1) が本質的で (2) は (少なくとも n に関するオーダーを評価するにあたっては) あまり効いていないと予想して f の候補を探す。(実際には (2) をうまく使う方法を出せるかは不明)
Cauchy Schwarz の不等式を用いているので等号成立条件を考えれば (定数倍を除いて)

$$(f(k) - f(k-1))\sqrt{a_k} = \frac{1}{\sqrt{a_k}}$$

を満たすことである。

定数項は意味を持たないので簡単のため $f(1)=0$ となるものとすると

$$f(k)=f(k-1)+\frac{1}{a_k}=\sum_{l=2}^{k}\frac{1}{d_k\mu_k}$$

と書くことができる。この関数に対して $V[f],D[f]$ を計算してみると

$$E[f]=\sum_{k=1}^{n}\mu_k\sum_{l=2}^{k}\frac{1}{d_k\mu_k},\quad E[f^2]=\sum_{k=1}^{n}\mu_k\Big(\sum_{l=2}^{k}\frac{1}{d_k\mu_k}\Big)^2,$$

$$V[f]=E[f^2]-(E[f])^2,\quad D[f]=\sum_{k=2}^{n}\frac{1}{d_k\mu_k}$$

であるので、モデルにあわせて計算してみるしかないようである。

Markov 連鎖

もう少し一般化したものが Saloff-Coste [Sal] にあるが (記号などはこちらの書き方に変えるが) さらに少し拡張したものをここにあげる。
記号の準備
状態空間 $S=\{x_1,\ldots,x_n\}$ とし、生成作用素は
$Lf(x)=\sum_y C_{x,y}(f(y)-f(x))$ で与えられ、$C_{x,y}$ は既約、さらにある確率分布 $\mu=(\mu_{x_1},\ldots,\mu_{x_n})$ に関して対称とする。
$\gamma=\gamma(x,y)=\{x_0,x_1,\ldots,x_k\}$ が道 (path) であるとは
$x_0=x, x_k=y$ かつ $C_{x_{i-1},x_i}>0\ \forall 1\le i\le k$ を満たすこととする。
(この k を道 γ の長さと呼ぶ。)
$\delta=\delta(x,y)=(\gamma_1,\ldots,\gamma_l,a_1,\ldots,a_l)$ が重み付き道であるとは
$\gamma_1,\ldots,\gamma_l$ は全て道でかつ $\sum_i a_i=1,\ a_i>0,\forall 1\le i\le l$ を満たすこととする。(この道 γ_i のそれぞれの長さは違っても構わない。)
同じ名前重み付きの道、同じ記号 δ を用いて各 x,y を止める毎に重み付きの道 $\delta(x,y)$ を固定したものとし、これらの δ を全て集めた集合を Γ とする。

Markov 連鎖

辺の重み w は $\forall x, y \in S$ に対して $w_{x,y} = w_{y,x} > 0$ と定義されているものとする。(実際は $C_{x,y} > 0$ となる x, y の組だけ定義されていれば十分。) この重み w と道 $\gamma = (x_1, \ldots, x_k)$ に対して

$$|\gamma|_w = \sum_{j=1}^{k} \frac{1}{w_{x_{j-1},x_j}}$$

と定義する。

Markov 連鎖

定理 (Saloff-Coste の拡張)

重み付き道 δ と辺の重み w を決める毎に

$$A = A(\delta, w) := \max\{\frac{w_{u,v}}{C_{u,v}\mu_v}\sum_{x,y,i}\mu_x\mu_y a_i|\gamma_i(x,y)|_w; u, v, C_{u,v} > 0\}$$

が定まる。但し $\sum_{x,y,i}$ は $\{x, y, i; \exists j \text{ s.t. } x^i_{j-1} = u, x^i_j = v\}$ の上での和を取っている。このとき $\lambda \geq 1/A$ である。

系

$B = \min\{A(\delta, w); \delta, w\}$ とおくと $\lambda \geq 1/B$ である。

証明
先ほどの証明と同様に δ を一つ固定すると各 x, y に対して (x, y と y, x で表記上は同じように見えるが違うものであっても構わない。)

$$f(x) - f(y) = \sum_{i=1}^{l} \sqrt{a_i}\sqrt{a_i}\{\sum_{j=1}^{m}(f(x^i_j) - f(x^i_{j-1}))\sqrt{w_j}\frac{1}{\sqrt{w_j}}\}$$

但し $\delta = (\gamma_1, \ldots, \gamma_l, a_1, \ldots, a_l)$ に対して $\gamma_i = \{x^i_1, x^i_2, \ldots, x^i_m\}$, $m = m_i$, $w_j = w_{x^i_{j-1}, x^i_j}$ である。Cauchy Schwarz の不等式を適用することで以下のように評価できる。

$$\begin{aligned} \{f(x) - f(y)\}^2 &\leq \sum_i a_i \sum_i a_i \{\sum_{j=1}^{m}(f(x^i_j) - f(x^i_{j-1}))\sqrt{w_j}\frac{1}{\sqrt{w_j}}\}^2 \\ &\leq \sum_i a_i \sum_{j=1}^{m}(f(x^i_j) - f(x^i_{j-1}))^2 w_j |\gamma_i(x, y)|_w \end{aligned}$$

Markov 連鎖

前回と同じような計算をすると

$$
\begin{aligned}
V[f] &= \frac{1}{2}\sum_{x,y}(f(y)-f(x))^2\mu_x\mu_y \\
&\le \frac{1}{2}\sum_{x,y}\mu_x\mu_y\sum_i a_i\sum_{j=1}^{m}(f(x^i_j)-f(x^i_{j-1}))^2 w_j|\gamma_i(x,y)|_w \\
&= \frac{1}{2}\sum_{u,v}(f(u)-f(v))^2 w_{u,v}\sum_{x,y,i}\mu_x\mu_y a_i|\gamma_i(x,y)|_w \\
&= \frac{1}{2}\sum_{u,v}(f(u)-f(v))^2 C_{u,v}\mu_v\frac{w_{u,v}}{C_{u,v}\mu_v}\sum_{x,y,i}\mu_x\mu_y a_i|\gamma_i(x,y)|_w \\
&\le \max\{\frac{w_{u,v}}{C_{u,v}\mu_v}\sum_{x,y,i}\mu_x\mu_y a_i|\gamma_i(x,y)|_w; u,v\}D[f]
\end{aligned}
$$

但し $\sum_{x,y,i}$ は $\{x,y,i;\exists j \text{ s.t. } x^i_{j-1}=u, x^i_j=v\}$ の上での和を取っている。 $\square$

Markov 連鎖

実際に出生死亡過程で (特殊な) レート b_k, d_k を代入してみた場合のスペクトルギャップを評価してみる。今までの結果をまとめると次のような計算をすればよいことが分かるが、和ではなく積分で近似して (できるものとして) 計算する。また基本的にはサイズ n に関するオーダーが出れば (上からと下からの評価のオーダーが一致すれば) 十分と思うことにする。

Markov 連鎖

$$\tilde{\mu}_1 = 1, \quad \tilde{\mu}_{x+1} = \frac{b_x}{d_{x+1}}\mu_x, \quad Z = \sum_i \tilde{\mu}_i, \quad \mu = \frac{1}{Z}\tilde{\mu}$$

$$A_k = \sum_{i=1}^{k-1}\sum_{j=k}^{n}\sum_{l=i+1}^{j}\frac{1}{d_l\mu_l}\mu_i\mu_j, \quad A = \max_k A_k$$

$$E[f] = \sum_{k=1}^{n}\mu_k\sum_{l=2}^{k}\frac{1}{d_k\mu_k}, \quad E[f^2] = \sum_{k=1}^{n}\mu_k(\sum_{l=2}^{k}\frac{1}{d_k\mu_k})^2,$$

$$V[f] = E[f^2] - (E[f])^2, \quad D[f] = \sum_{k=2}^{n}\frac{1}{d_k\mu_k}, \quad C = \frac{D[f]}{V[f]}$$

$$\frac{1}{A} \le \lambda \le C$$

Markov 連鎖

[1] $b_k = d_{k+1}$ のケース。
このとき μ が一様分布になることに注意する。
(1) $d \equiv 1$

$$A_k \cong \int_0^k (\int_k^n (\int_x^y \frac{1}{n} du) dy) dx = \frac{1}{2}(nk - k^2) \leq \frac{1}{8} n^2$$

$$E[f] \cong \int_0^n \frac{1}{n} \int_0^x n du dx = \frac{1}{2} n^2,$$

$$E[f^2] \cong \int_0^n \frac{1}{n} (\int_0^x n du)^2 dx = \frac{1}{3} n^4$$

$$V[f] \cong \frac{1}{12} n^4, \quad D[f] \cong \int_0^n n dx = n^2$$

$$\frac{8}{n^2} \leq \lambda \leq \frac{12}{n^2}$$

実際は $f(x) = \cos(\pi x/n)$ がほぼ固有関数になるため $\lambda \cong \frac{\pi^2}{n^2}$

(2) $d_k = k^\alpha$ $(\alpha < 0)$ のケース。

$$A_k \cong \int_0^k (\int_k^n (\int_x^y \frac{1}{n}\frac{1}{u^\alpha} du) dy) dx = O(n^{-\alpha+2})$$

$$E[f] \cong \int_0^n \frac{1}{n} \int_0^x n \frac{1}{u^\alpha} du dx = O(n^{-\alpha+2}),$$

$$E[f^2] \cong \int_0^n \frac{1}{n} (\int_0^x n \frac{1}{u^\alpha} du)^2 dx = O(n^{-2\alpha+4})$$

$$V[f] \cong O(n^{-2\alpha+4}), \quad D[f] \cong \int_0^n n \frac{1}{u^\alpha} dx = O(n^{-\alpha+2})$$

$$\lambda \cong O(n^{\alpha-2})$$

Markov 連鎖

粒子系への応用
簡単のため排他過程で話をするが、多くのモデルで適用可能な方法である。
実際多くの論文でこの方法を適用していると考えられるが、大抵通常の $\mathbb{Z}^d$ 格子 (を有限系に制限したもの) を使い、直接的に証明してしまっている。

Markov 連鎖

S (もしくは $\{S_n\}$) を Markov 過程の状態空間とし、対応する生成作用素は $Mf(x) = C_{x,y}(f(y) - f(x))$, とし、$C_{x,y}$ は既約で対応する生成作用素は $Mf(x) = C_{x,y}(f(y) - f(x))$, とし、$C_{x,y}$ は既約で $C_{x,y} = C_{y,x}$ とする。
この S に対して粒子系の状態空間 Λ_k (もしくは $\{\Lambda_{n,k}\}$) を

$$\Lambda_k := \{\eta \in \{0,1\}^S; \sum_x \eta_x = k\}$$

とする。排他過程 (川崎モデル) の生成作用素 L を

$$Lf(\eta) = \sum_{x,y} C_{x,y}\eta_x(1-\eta_y)(f(\eta^{x,y}) - f(\eta))$$

と定義する。ただし $\eta \in \Lambda_k$ に対して

$$(\eta^{x,y})_z = \begin{cases} \eta_y & \text{if } z = x \\ \eta_x & \text{if } z = y \\ \eta_z & \text{otherwise} \end{cases}$$

Markov 連鎖

このモデルでは $\eta_x = 1$ は点 x に粒子あり、$\eta_x = 0$ ならば点 x に粒子はなしと解釈し、$\sum_x \eta_x = k$ に制限しているので総粒子数が k 個である。この個々の粒子は M を生成作用素とするランダムウォークをするがジャンプしたい行き先に既に他の粒子がある場合はジャンプをすることをとりやめる (排他的) 規則でランダムに時間発展して行く。

命題 (対称測度)

π を Λ_k 上の一様分布とすると L は π に関して対称

証明

$\eta_x = 1$, $\eta_y = 0$ の場合に $\pi(\eta)C_{x,y} = \pi(\eta^{x,y})C_{y,x}$ を示せばよいが仮定より明らか。 □

Markov 連鎖

定理 (平均場近似の場合のスペクトルギャップ)

$\#S = n$ として $C_{x,y} = 1/n, \forall x, y$ とすると L のスペクトルギャップは

$$\lambda = 1$$

である。

Markov 連鎖

平均場近似の生成作用素 L^m を

$$L^m f(\eta) = \sum_{x,y} \frac{1}{n} \eta_x (1-\eta_y)(f(\eta^{x,y} - f(\eta))$$

とし、対応する Dirichlet 形式を $D^m[f]$ と書くことにする

定理 (平均場近似との比較定理)

重み付き道 δ と辺の重み w を決める毎に定理 (Salof-Coste の拡張) で与えられる $A = A(\delta, w)$ を用いて

$$D^m[f] \leq AD[f]$$

系

L のスペクトルギャップ λ は

$$\lambda \geq 1/B$$

Markov 連鎖

証明

$\gamma(x,y) = \{x_0 = x, x_1, \dots, x_n = y\}$ と $\eta_x = 1, \eta_y = 0$ を満たす η が与えられた時に $\{\eta^i\}_{i=0}^{2n-1}$ を以下のように帰納的に定める。

$$\eta^0 = \eta, \quad \eta^i = \begin{cases} (\eta^{i-1})^{x_{i-1}, x_i} & 1 \le i \le n \\ (\eta^{i-1})^{x_{2n-i-1}, x_{2n-i}} & n+1 \le i \le 2n-1 \end{cases}$$

この $\{\eta^i\}_{i=0}^{2n-1}$ は次の命題を満たすように作ったものである。

命題 (η^i の性質)

$\eta^{2n-1} = \eta^{x,y}$,
$1 \le i \le n$ ならば $(\eta^i)_{x_{i-1}} = 1$
$n+1 \le i \le 2n-1$ ならば $(\eta^i)_{x_{2n-i}} = 0$

問題

命題の性質を満たしていることを確かめよ。

Markov 連鎖

これは x にあった粒子を (粒子のない) y へ移すことを意味するが、Markov 過程でジャンプできる道のうち特に γ に沿って一歩づつ動かしたものであるがこの命題の性質から次のことが分かる。

系

(η^i) を使って粒子を移動した時、γ から自然に表れる各辺 (x_{i-1}, x_i) $(1 \le i \le n)$ に対して粒子は 1 回づつしかジャンプしていない。

命題

点 u, v 道 $\gamma = \gamma(x, y) = \{x_i\}_{i=1}^{n}$ と配置 ξ に対して $\exists j$ s.t. $u = x_{j-1}, v = x_j$, $\xi_u = 1$, $\xi_v = 0$ とする。このとき $\exists k$ s.t. $\xi = \eta^k$ となる η が唯 1 つ決まる。

問題

命題の性質を満たしていることを確かめよ。

Markov 連鎖

先の定理をほぼなぞることになるが先の定理の証明では
$(f(y) - f(x))^2$ を評価していたが $\eta_x = 1, \eta_y = 0$ の下で
$(f(\eta^{x,y}) - f(\eta))^2$ を同じように評価する。
先ほどと同じように $w_j = w_{x_{j-1},x_j}$, $(1 \leq j \leq n)$,
$w_j = w_{x_{2n-i-1},x_{2n-i}}$, $(n+1 \leq j \leq 2n-1)$ として

$$f(\eta^{x,y}) - f(\eta) = \sum_{i=1}^{l} \sqrt{a_i}\sqrt{a_i}\{\sum_{j=1}^{2n-1}(f(\eta^j) - f(\eta^{j-1}))\sqrt{w_j}\frac{1}{\sqrt{w_j}}\}$$

と書けるのでこれに Cauchy Shwarz の不等式を適用する。

$$\begin{aligned}
&\{f(\eta^{x,y}) - f(\eta)\}^2 \\
&\leq \sum_i a_i\{\sum_{j=1}^{2n-1}(f(\eta^j) - f(\eta^{j-1}))^2 w_j|\gamma_i(x,y)|_w
\end{aligned}$$

Markov 連鎖

これに η^i の性質と π が一様分布であることを加えると
$u_j = x_{j-1}$, $(1 \le j \le n)$, $u_j = x_{2n-i-1}$, $(n+1 \le j \le 2n-1)$ として

$$
\begin{aligned}
&\eta_x(1-\eta_y)\{f(\eta^{x,y}) - f(\eta)\}^2\pi(\eta) \\
&\le \sum_i a_i \{ \sum_{j=1}^{2n-1} (f(\eta^j) - f(\eta^{j-1}))^2 \pi(\eta^{j-1}) w_j |\gamma_i(x,y)|_w \} \\
&= \sum_i a_i \{ \sum_{j=1}^{2n-1} \eta^j_{u_j}(1-\eta^j_{u_j+1})(f(\eta^j) - f(\eta^{j-1}))^2 \pi(\eta^{j-1}) \\
&\qquad\qquad \times w_j |\gamma_i(x,y)|_w \}
\end{aligned}
$$

と書ける。
また M を生成作用素とする Markov 過程の定常分布は $\mu_x = 1/n$ であることに注意する。

Markov 連鎖

$D^m[f]$ の定義へ今までの結果を代入して計算すると

$$D^m[f] = \frac{1}{2}\sum_{\eta}\sum_{x,y}\frac{1}{n}\eta_x(1-\eta_x)(f(\eta^{x,y})-f(\eta))^2\pi(\eta)$$

$$\le \frac{1}{2}\sum_{\eta}\sum_{x,y}\mu_y\sum_i a_i\Bigg\{\sum_{j=1}^{2n-1}$$

$$\eta^j_{u_j}(1-\eta^j_{u_{j+1}})(f(\eta^j)-f(\eta^{j-1}))^2 w_j|\gamma_j(x,y)|_w\Bigg\}\pi(\eta)$$

$$= \frac{1}{2}\sum_{\eta}\sum_{\xi}\sum_{x,y}\mu_y\sum_i a_i\Bigg\{\sum_{j=1}^{2n-1}\mathbf{1}_{\{\xi=\eta^j\}}$$

$$\xi_{u_j}(1-\xi_{u_{j+1}})(f(\xi^{u_j,u_{j+1}})-f(\xi))^2 w_j|\gamma_j(x,y)|_w\Bigg\}\pi(\eta)$$

Markov 連鎖

$D^m[f]$ の定義へ今までの結果を代入して計算すると和の交換が可能で先のページの最後から続けて

$$= \frac{1}{2} \sum_{\xi} \sum_{u,v} (f(\xi^{u,v}) - f(\xi))^2 \pi(\xi) w_{u,v}$$

$$\times \sum_{\eta} \sum_{x,y,i} \mathbf{1}_{\{\xi = \eta^i\}} \mu_y a_i |\gamma_i(x,y)|_w$$

$$= \frac{1}{2} \sum_{\eta} \sum_{u,v} (f(\eta^{u,v}) - f(\eta))^2 C_{u,v} \pi(\eta)$$

$$\frac{w_{u,v}}{C_{u,v}\mu_v} \sum_{x,y,i} \mu_x \mu_y a_i |\gamma_i(x,y)|_w = AD[f] \quad \square$$

注意：先のページで書いたがこの証明では π が一様分布であることすなわち $\forall \eta, \xi \in \Lambda_k$ に対して $\pi(\eta) = \pi(\xi)$ であることを全面的に使っている。これを仮定しないと $\pi(\eta^{i-1})/\pi(\eta)$ のような量が出てくるがこれを評価するのは以外と厳しい。
π が一様分布であることは C の対称性 $C_{x,y} = C_{y,x}$ から来ており、これをある μ に関して対称に変更すると π が一様分布ではなくなり、μ に関して技術的な仮定をすることで (この仮定だけに依存した) 定数 C を使って $D^m[f] \leq CAD[f]$ と書けるが、証明はかなり煩雑でこの定数もかなり大きくなる。

参考文献

少なくとも以下の教科書などは参考にしている。但し、ずいぶん前に読んで勉強した教科書の内容をそれと認識せずに参考にしていることもありえるのでその点に関してはご容赦頂きたい。
Markov 過程の基本的なところでは [K] の教科書の該当部分を多く参考にしている。
ランダムナイトに関しては [AF] の教科書からアイデアを得ているが、証明は [K] の教科書に沿っている。
Perron-Frobenius の定理は [V] の教科書の証明に沿っているが、確率論的には [Sai][Sal][Se] の方がよいかもしれない。なお [Sai] は非負の行列でなく正の行列で証明をしている。
スペクトルギャップに関しては [Sal] の結果を中心に、粒子系（格子気体）への応用を述べた。粒子系への応用に関して明示的に書かれたもの（教科書、論文）はないと思う。

参考文献

[AF] Aldous D.; Fill J.A.
Reversible Markov Chains and Random Walks on Graphs,
(未刊の教科書 現在 (2015/11) も 「Aldous Fill Markov chain」で検索をすれば pdf、 html バージョンを発見できる)
[K] 小谷真一, 測度と確率, 岩波書店
[LPW] Levin D.A.; Peres Y,; Wilmer E.L.
Markov Chains and Mixing Times, AMS
[Sai] 斎藤正彦, 線形代数入門, 東京大学出版会
[Sal] Saloff-Coste L.
Lectures on Finite Markov Chains, Lecture on Probability 1665
[Se] Seneta E.
Non-negative Matrices and Markov Chains, Springer
[V] Varga R.S.
Matrix Iterative Analysis, Springer