マルコフ連鎖入門　第 7 回

永幡幸生
新潟大学
nagahata@eng.niigata-u.ac.jp

2021 第 1 学期

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

$$
\begin{aligned}
&Q^{-1}PQ = \Lambda, \qquad P = Q\Lambda Q^{-1} \\
&\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}, \quad Q^{-1} = \begin{pmatrix} \mu^1 \\ \vdots \\ \mu^n \end{pmatrix}, \quad Q = \begin{pmatrix} q^1 \cdots q^n \end{pmatrix}
\end{aligned}
$$

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

ベクトル ${}^t f = (f_1, \dots, f_n)$ を $f_i = f(x)$ とし $\nu = (\nu_1, \dots, \nu_n)$ 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。

Markov 連鎖

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

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

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 $(c_1 A + c_2 I)$, は同じ Q を用いて対角化可能で

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

であることを示せ。

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 $S=\{x_1,x_2,\ldots,x_n\}$ 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 $f_1,f_2,\ldots,f_n$ 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 $f_1,f_2,\ldots,f_n$ 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

(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。

Markov 連鎖

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 $Lf(x)=(P-I)_{x,y}(f(y)-f(x))$ 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
$Lf(x) = (P - I)_{x,y}(f(y) - f(x)$ 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 $f_1, \ldots, f_n$ 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。

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 $(Lf(x) = (P - I)_{x,y}(f(y) - f(x))$ と、μ 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^* $(L^*f(x) = (P^* - I)_{x,y}(f(y) - f(x))$ は実際に共役の関係

$$E_\mu[fLg] = E_\mu[gLf], \quad \forall f, g$$

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？