マルコフ連鎖入門　第2回

永幡幸生
新潟大学
nagahata@ie.niigata-u.ac.jp

2021 September

Markov 連鎖

推移確率行列もしくは、作用素が与えられたときに、スペクトルギャップを正確に計算することは難しいことが多い。
一方で確率過程のスケール極限を考える場合には、(状態空間も変化するが) 推移確率行列の列 $\{P_n\}$ もしくは作用素の列 $\{L_n\}$ が与えられたときにスペクトルギャップの n に関するオーダーの評価が必要になり、逆にオーダーの評価ができれば十分なことは多い。
この路線にそったスペクトルギャップの評価方法を与える。
基本的には Cauchy Schwartz の不等式をいかにうまく使うかという路線になる。

Markov 連鎖

この講義では、ある状態空間 $S = \{x_1, x_2, \ldots, x_r\}$ に対して、可逆確率測度 (になる) $\mu = (\mu_1, \mu_2, \ldots, \mu_r)$ があり、
$C_{x,y} \geq 0, \forall x, y \in S$, $C_{x,x} = 0$ が与えられ、μ に対して可逆、すなわち

$$\mu_x C_y = \mu_y C_{y,x}$$

を満たし、これらを用いて作用素

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

が与えられているものとする。

Markov 連鎖

自然に f の (μ による) 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。

命題

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

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

と書ける。

証明

定義通りに計算する、最後に μ に関する対称性を使うと

$$
\begin{aligned}
E[f(-L)f] &= \sum_x \left[f(x) \left\{ -\sum_y C_{x,y}(f(y)-f(x)) \right\} \right] \mu_x \\
&= \sum_{x,y} \mu_x C_{x,y} \{-f(x)\}\{f(y)-f(x)\} \\
&= \sum_{x,y} \mu_y C_{y,x} \{-f(x)\}\{f(y)-f(x)\}
\end{aligned}
$$

最後の式で和に関する変数 x, y の文字を入れ替えても変わらないので

$$
E[f(-L)f] = \sum_{x,y} \mu_x C_{x,y} \{-f(y)\}\{f(x)-f(y)\}
$$

この式と上で計算した真ん中の式は共に $E[f(-L)f]$ を表しているので

$$
\begin{aligned}
E[f(-L)f] &= \frac{1}{2}\Bigg[\sum_{x,y}\mu_x C_{x,y}\{-f(x)\}\{f(y)-f(x)\} \\
&\qquad + \sum_{x,y}\mu_x C_{x,y}\{-f(y)\}\{f(x)-f(y)\}\Bigg] \\
&= \frac{1}{2}\sum_{x,y}\mu_x C_{x,y}(f(y)-f(x))^2
\end{aligned}
$$

□

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 $f = {}^t(f_1, \ldots, f_r)$ を $f_1 = f(x_1)$ で書けば

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

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。 $\square$

Markov 連鎖

次の命題を証明なしで使う。

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。

Markov 連鎖

簡単のため $d = 1$ とすると、$B(r) = [-r, r]$ となる。この命題において $u \in C^2$ 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

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

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 $Cr^2D[u]$ に対応しているすなわち

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

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

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

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

$$V[f] \le CD[f]$$

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。

解答例
右辺側を計算する。

$$E[(f(X) - f(Y))^2] = E[(f(X))^2 - 2f(X)f(Y) + (f(Y))^2]$$

であるが X, Y は同分布なので $E[(f(X))^2] = E[(f(Y))^2]$
独立なので $E[f(X)f(Y)] = E[f(X)]E[f(Y)]$ さらに同分布なので
$E[f(X)]E[f(Y)] = E[f(X)]^2$ になる。従って

$$E[(f(X) - f(Y))^2] = 2E[(f(X))^2] - 2(E[f(X)])^2 = 2V[f(X)] \qquad \square$$

Markov 連鎖

完全グラフ上のランダムウォーク

$S = \{1, 2, \ldots, n\}$ 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 $\mu = (\mu_1, \ldots, \mu_n)$ 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 マルコフ連鎖は完全グラフ上のランダムウォークと考えられる。

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。 □
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

完全グラフ上の重み付きランダムウォーク

$S = \{1, 2, \ldots, n\}$、S 上の確率 $\mu = (\mu_1, \ldots, \mu_n) > 0$ として推移確率行列 P を

$$P_{i,j} = \mu_j, \quad \forall i$$

とおくと明らかに P は μ 対称である。この推移確率行列で与えられる マルコフ連鎖は完全グラフ上の重み付きランダムウォークと考えられる。

完全グラフ上の重み付き 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
先ほどと全く同様に

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

よりスペクトルギャップは 1 になる。 $\square$
注意：先ほどと同様に 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。

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

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

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

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

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

これを代入し、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。 $\square$

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。それは (2) の 2 つの Dirichlet 形式の比較が直接的には行えない場合も、計算や、比較のしやすい中間的な Dirichlet 形式をはさむことで結果的に (2) の比較が可能になることが多い。一方でこのような中間的な Dirichlet 形式を探すのはかなり大変である。

Markov 連鎖

出生死亡過程

ランダムウォークを一般化する。
ランダムウォークの生成作用素は

$$Lf(i) = \begin{cases} \frac{1}{2}\{f(i+1) - f(i)\} & i = 0 \\ \frac{1}{2}\{f(i+1) - f(i)\} + \frac{1}{2}\{f(i-1) - f(i)\} & i = 2, 3, \ldots n-1 \\ \frac{1}{2}\{f(i-1) - f(i)\} & i = n \end{cases}$$

と書けていたが、これを係数だけ変化させて、

$$Lf(i) = \begin{cases} b_1\{f(i+1) - f(i)\} & i = 0 \\ b_i\{f(i+1) - f(i)\} + d_i\{f(i-1) - f(i)\} & i = 2, 3, \ldots n-1 \\ d_i\{f(i-1) - f(i)\} & i = n \end{cases}$$

$b_i, d_i > 0$ とする。

Markov 連鎖

これは（実質何も変化していないが）次のように書いてもよい。

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

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。

Markov 連鎖

このモデルは、元々は生物学のモデルで、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。
生物学のモデルだと、個体数の減少はその人口に比例し、個体数の増加は人口が少ないうちは人口に比例するが、(人口が増えると、限られた食料などを奪い合うことにより) 人口が多くなるとあまり多くならない例を考えることが多いようである。
SIR モデルの近似で出てくる生成死滅過程のパラメータも同じような性質を持っていると考えられる。

Markov 連鎖

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

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

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。

Markov 連鎖

SIR モデルとの対応で、（総人口に対応する N は β の中に入っていいるものとして）

$$b_i = \beta i, \quad d_i = \gamma i$$

の場合に μ がどのようになるのかを見ておく。
定義通りに $\tilde{\mu}_x$ を計算すると

$$\tilde{\mu}_x = \left(\frac{\beta}{\gamma}\right)^x \frac{1}{x}$$

$1/x$ の項の関係で、規格化をするのは難しいが、n が十分大きければ $(\beta/\gamma)^x$ の部分がメインのパートになり、$\beta/\gamma > 1$ では x が大きいところにほぼ全ての重みがあり、$\beta/\gamma < 1$ では x が小さいところにほぼ全ての重みがある。例外的に $\beta/\gamma = 1$ では $1/x$ に比例するため $\beta/\gamma < 1$ のところと同じように x が小さいところに重みが集中する。

Markov 連鎖

SIR モデルとの関係を、考えれば十分時間がたてば、状態は重みのあるところに集中していると思えるので、$\beta/\gamma > 1$ では x が大きいところ、すなわち感染者だらけになっており、$\beta/\gamma < 1$ では x が小さいところ、すなわち許容できる状態になっていることが予想できる。
このように近似過程の生成死滅過程の可逆測度を見るだけでも、実行再生産数 $\beta/\gamma = 1$ がクリティカルであることが容易に分かる。

この生成死滅過程に対応する L のスペクトルギャップの下からの評価を行うが、ランダムウォークの時と同じように計算してみる。分散 $V[f]$ を式変形すると、

$$
\begin{aligned}
V[f] &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\{f(j)-f(i)\}^2\mu_i\mu_j \\
&= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\{f(j)-f(i)\}^2\mu_i\mu_j
\end{aligned}
$$

と書ける。$j \geq i$ なので

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

と書いて

Markov 連鎖

Cauchy Schwarz の不等式を適用することで

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

これを代入し、和の順序交換をすることで

$$\begin{aligned} V[f] &\leq \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (j - i) \sum_{k=i+1}^{j} \{f(k) - f(k-1)\}^2 \mu_i \mu_j \\ &= \sum_{k=2}^{n} \{f(k) - f(k-1)\}^2 \sum_{i=1}^{k} \sum_{j=k+1}^{n} (j - i) \mu_i \mu_j \\ &\leq \sum_{k=2}^{n} d_k \{f(k) - f(k-1)\}^2 \mu_k \sum_{i=1}^{n} \sum_{j=1}^{n} |j - i| \mu_i \mu_j \frac{1}{d_k \mu_k} \end{aligned}$$

最後の変形で Dirichlet 形式が見えるように変形を行ったが、この式から出せる最良の不等式は

$$
\begin{aligned}
&\sum_{k=2}^{n} d_k\{f(k)-f(k-1)\}^2\mu_k \sum_{i=1}^{n}\sum_{j=1}^{n}|j-i|\mu_i\mu_j\frac{1}{d_k\mu_k}\\
\leq\ &\sup_k\{\frac{1}{d_k\mu_k}\}\sum_{i=1}^{n}\sum_{j=1}^{n}|j-i|\mu_i\mu_j D[f]
\end{aligned}
$$

である。

Markov 連鎖

ランダムウォークの場合は $d_k\mu_k$ が k によらず $\frac{1}{2n}$ であったこと と (実際には係数も計算できるが、) $\sum_{i=1}^{n}\sum_{j=1}^{n}|j-i|\mu_i\mu_j = O(n)$ であることから

$$V[f] \leq O(n^2)D[f]$$

が得られ、このオーダーは実際のオーダーと一致している。一方で、SIR モデルの近似を含む生成死滅過程一般では $\inf\{d_k\mu_k\}$ が指数的に小さくなる例もあり、実際のスペクトルギャップのオーダー ($O(1)$) と合わないことの方が多い。

Markov 連鎖

Cauchy Schwarz の不等式を適用することで

$$\begin{aligned}\{f(j)-f(i)\}^2 &= \left\{\sum_{k=i+1}^{j}(f(k)-f(f-1))\right\}^2 \\ &\leq (j-i)\sum_{k=i+1}^{j}\{f(k)-f(k-1)\}^2\end{aligned}$$

としたが適用の仕方が良くなかったと思って、この部分を少しテクニカルに変えてみる。

Markov 連鎖

$$
\begin{aligned}
\{f(j)-f(i)\}^2 &= \left\{\sum_{k=i+1}^{j}(f(k)-f(k-1))\right\}^2 \\
&= \left[\sum_{k=i+1}^{j}\left\{(f(k)-f(k-1))\sqrt{a_k}\right\}\frac{1}{\sqrt{a_k}}\right]^2 \\
&\leq \left\{\sum_{k=i+1}^{j}(f(k)-f(k-1))^2 a_k\right\}\left\{\sum_{k=i+1}^{j}\frac{1}{a_k}\right\}
\end{aligned}
$$

ここで $\{a_k\}$ を導入したが、詳細は後で全体の不等式が良くなるように設定することにして、ここで意味を持つために $a_k > 0 \quad \forall k$ だけ仮定しておく。また $\displaystyle\sum_{k=i+1}^{j}\frac{1}{a_k} =: B_{ij}$ と書くことにする。

Markov 連鎖

先ほどと同様にこの不等式を代入したのち、和の順序交換を行うと

$$
\begin{aligned}
V[f] &= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\{f(j)-f(i)\}^2\mu_i\mu_j \\
&\leq \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left\{\sum_{k=i+1}^{j}(f(k)-f(f-1))^2 a_k\right\}B_{i,j}\mu_i\mu_j \\
&= \sum_{k=2}^{n}(f(k)-f(k-1))^2 a_k\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}B_{ij}\mu_i\mu_j
\end{aligned}
$$

と書けるが $\displaystyle\sum_{j=i+1}^{n}B_{ij}\mu_i\mu_j := A_k$ と書くと、

$$
\begin{aligned}
V[f] &\leq \sum_{k=2}^{n} d_k(f(k)-f(k-1))^2 \mu_k \frac{a_k A_k}{d_k \mu_k} \\
&\leq \sup_k \{\frac{a_k A_k}{d_k \mu_k}\} \sum_{k=2}^{n} d_k(f(k)-f(k-1))^2 \mu_k \\
&= \sup_k \{\frac{a_k A_k}{d_k \mu_k}\} D[f]
\end{aligned}
$$

と書ける。
$\{a_k\}$ は後で決めるとしていたが、A_k が $\{a_k\}$ により決まるため一番分かりやすい設定は

$$
a_k = d_k \mu_k
$$

で　A_k をモデルに合わせて計算してみることと思われる。
(後でいくつかの例で適用してみる。結果的には上手くいく例は多いが、SIR モデルに対応したものでは上手くいかない。)

Markov 連鎖

この計算で行っていたのは $((f(j) - f(i))^2$ のうち $V[f]$ の中にはでてくるが $D[f]$ の中にはでてこないものを Cauchy Schwarz の不等式を使って $D[f]$ のに出てくる $(f(k) - f(l))^2$ で書き換えたことと思える。
この考え方で一般化したものが Saloff-Coste 「Lectures on Finite Markov Chains」 Lecture on Probability や Levin D.A.; Peres Y, 「Markov Chains and Mixing Times」 AMS にあるが、この 2 つに書かれているものを組み合わせた定理を紹介する。

Markov 連鎖

記号の準備
状態空間 $S = \{x_1, x_2, \ldots, x_r\}$ とし生成作用素は

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

で与えられ、このマルコフ連鎖は既約であるとする。
すなわち $\forall x, y \in S$ に対して $\exists\{x_0, x_1, \ldots, x_n\} \subset S$ s.t.
$x_0 = x, x_n = y$ かつ $1 \leq \forall i \leq n$ に対して $C_{x_{i-1},x_i} > 0$ ととれる。
さらにこのマルコフ連鎖は可逆確率測度 $\mu = (\mu_{x_1}, \mu_{x_2}, \ldots, \mu_{x_r})$ を持つ。

記号の準備

$\gamma = \gamma(x, y) = \{x_0, x_1, \dots, x_n\}$, $x, y, x_0, \dots, x_n \in S$ が道 (path) であるとは
$x_0 = x, x_n = y$ かつ $C_{x_{i-1},x_i} > 0 \quad 1 \le \forall i \le n$ を満たすこと。
この n を道 γ の長さと呼ぶ。
$\delta = \delta(x, y) = (\gamma_1, \dots, \gamma_l, a_1, \dots, a_l)$ が重み付き道であるとは
$\gamma_1 = \gamma_1(x, y), \dots, \gamma_l = \gamma_l(x, y)$ が全て道であり、$\sum_i a_i = 1$,
$a_i > 0 \ 1 \le \forall i \le n$ を満たすこと。
ただし $\gamma_1, \dots, \gamma_l$ の長さは違っていても構わない。

記号の準備

道は直感的に分かるように $C_{z,w} > 0$ となるような辺を伝って x から y へ到達できる通路を示し、
重み付き道は、道を複数本準備しておいて、そこから $a_1, \ldots, a_l$ で与えられる確率で選んで通ることを想定している。
ランダムウォークを含む生成死滅過程の設定のように 1 次元の場合は、ループを排除すると、道は 1 つに定まってしまうため、意味を持たないが、2 次元のランダムウォークを考えると、多くの場合色々な道を考えることができる。

記号の準備

辺の重み w は $\forall x, y \in S$ に対して $w_{x,y} = w_{y,x} > 0$ と定義する。
($C_{x,y} > 0$ となる x, y の組にだけ定義されていれば十分)
辺の重み w と道 $\gamma = \{x_0, x_1, \ldots, x_n\}$ に対して

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

と定義する。

記号の準備

重み付き道 δ と辺の重み w を決める毎に

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

が定義できる。ただし $\sum_{x,y,i}$ は与えられた重み付きの道 $\delta = (\gamma_1, \ldots, \gamma_l, a_1, \ldots, a_l)$ の各道を $\gamma_i = \{x_0^i, x_1^i, \ldots, x_n^i\}$ としたとき、$\{x, y, i;\ \exists j \text{ s.t. } x_{j-1}^i = u, x_j^i = v\}$ での和である。

Markov 連鎖

定理

作用素 L のスペクトルギャップ λ_2 は任意の重み付き道 δ と辺の重み w に対して

$$\lambda_2 \geq \frac{1}{A(\delta, w)}$$

である。

系

作用素 L のスペクトルギャップ λ_2 は

$$\lambda_2 \geq \sup\{\frac{1}{A(\delta, w)}; \delta, w\} = \frac{1}{\inf\{A(\delta, w); \delta, w\}}$$

である。

証明

A の定義で書いたように与えられた重み付きの道 $\delta = (\gamma_1, \ldots, \gamma_l, a_1, \ldots, a_l)$ のそれぞれの道は $\gamma_i = \{x^i_0, x^i_1, \ldots, x^i_n\}$ と書く。また $w^i_j = w_{x^i_{j-1}, x^i_j}$ とする。

生成死滅過程の時の変形と同じように

$$f(x) - f(y) = \sum_{i=1}^{l} \sqrt{a_i} \left[\sqrt{a_i} \left\{ \sum_{j=1}^{n} (f(x^i_{j-1}) - f(x^i_j)) \sqrt{w^i_j} \frac{1}{\sqrt{w^i_j}} \right\} \right]$$

と書いて Cauchy Schwartz の不等式を適用すると

$$\begin{aligned} &(f(x) - f(y))^2 \\ &\quad \le \sum_{i=1}^{l} a_i \sum_{i=1}^{l} a_i \left\{ \sum_{j=1}^{n} (f(x^i_{j-1}) - f(x^i_j)) \sqrt{w^i_j} \frac{1}{\sqrt{w^i_j}} \right\}^2 \end{aligned}$$

$\sum_i a_i = 1$ であることに注意して、さらに $\{\cdots\}^2$ の部分に再度 Cauchy Schwartz の不等式を適用すると

$$
\begin{aligned}
&(f(x) - f(y))^2 \\
&\quad \le \sum_{i=1}^{l} a_i \sum_{j=1}^{n} (f(x^i_{j-1}) - f(x^i_j))^2 w^i_j \sum_{j=1}^{n} \frac{1}{w^i_j} \\
&\quad = \sum_{i=1}^{l} a_i \sum_{j=1}^{n} (f(x^i_{j-1}) - f(x^i_j))^2 w^i_j |\gamma_i(x,y)|_w
\end{aligned}
$$

になる。

Markov 連鎖

生成死滅過程の時の変形と同じように

$$
\begin{aligned}
V[f] &= \frac{1}{2}\sum_{x,y}(f(x)-f(y))^2\mu_x\mu_y \\
&\le \quad \frac{1}{2}\sum_{x,y}\sum_{i=1}^{l} a_i \sum_{j=1}^{n}(f(x^i_{j-1})-f(x^i_j))^2 w^i_{x^i_{j-1},x^i_j}|\gamma_i(x,y)|_w \\
&= \quad \frac{1}{2}\sum_{u,v}(f(u)-f(v))^2 w_{u,v}\sum_{x,y,i}\mu_x\mu_y a_i|\gamma_i(x,y)|_w
\end{aligned}
$$

と書き直せるが、この時の和の変換により $\sum_{x,y,i}$ は定理の前の A の定義のただし書きとして表れる $\{x,y,i;\ \exists j \text{ s.t. } x^i_{j-1}=u, x^i_j=v\}$ での和である。

Markov 連鎖

これより Dirichlet 形式が表れるように書き換えると

$$\begin{aligned}
&= \frac{1}{2}\sum_{u,v} C_{u,v}(f(u)-f(v))^2 \mu_v \frac{w_{u,v}}{C_{u,v}\mu_v}\sum_{x,y,i}\mu_x\mu_y a_i|\gamma_i(x,y)|_w \\
&\le \sup\{\frac{w_{u,v}}{C_{u,v}\mu_v}\sum_{x,y,i}\mu_x\mu_y a_i|\gamma_i(x,y)|_w\} D[f] \\
&= A(\delta,w)D[f]
\end{aligned}$$

すなわち全ての f に対して $V[f] \le A(\delta,w)D[f]$ を満たすので $\lambda_2 \ge \dfrac{1}{A(\delta,w)}$ である。 □

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を中心に、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

出生死亡過程での上からの評価は変分公式から、どのような関数 f を選んでも、 Dirichlet 形式 $D[f]$ と分散 $V[f]$ の比を取ればスペクトルギャップの上からの評価にはなっている。より良い評価を目指すために、下からの評価の導出を見直す。
基本的に 2 か所でしか不等式を使っていないが、1 つは

$$\left[\sum_{k=i+1}^{j}\left\{(f(k)-f(k-1))\sqrt{a_k}\right\}\frac{1}{\sqrt{a_k}}\right]^2$$

$$\leq \left\{\sum_{k=i+1}^{j}(f(k)-f(k-1))^2 a_k\right\}\left\{\sum_{k=i+1}^{j}\frac{1}{a_k}\right\}$$

とした Cauchy Schwartz の不等式、

Markov 連鎖

もう一つは

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

$$\leq \sup_k \{\frac{a_k A_k}{d_k \mu_k}\} \sum_{k=2}^{n} d_k(f(k)-f(k-1))^2 \mu_k$$

と係数を sup で前に出したことによる不等式である。
よい f を見つけようという観点から 1 つ目の Cauchy Schwartz の等号成立条件を使って、スペクトルギャップの上からの評価のための関数 f を見つける。

Markov 連鎖

Cauchy Schwartz の等号成立条件は、$\exists C$ s.t. $\forall k$

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

なので、特に $f(1) = 0,\ C = 1$ として、さらに $a_k = d_k\mu_k$ であったことから

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

を採用することにする。

Markov 連鎖

これらの設定で必要な量を計算しておくと、次のようになる。

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

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

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

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

$$\frac{1}{A} \le \lambda_2 \le C$$

Markov 連鎖

(I) $b_k = d_{k+1}$ のケース
このとき $\mu \equiv \frac{1}{n}$ つまり一様分布になることに注意する。
(1) $d \equiv 1$

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

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

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

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

$$\frac{8}{n^2} \leq \lambda_2 \leq \frac{12}{n^2}$$

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

Markov 連鎖

(2) $d_k = k^\alpha$ $(\alpha < 1)$ のケース。やはり $\mu_k \equiv \frac{1}{n}$

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

$$E[f] \cong \int_0^n \frac{1}{n} \int_0^x n \frac{1}{u^\alpha} du dx$$

$$= \frac{1}{(1-\alpha)(2-\alpha)} n^{-\alpha+2} + o(n^{-\alpha+2})$$

$$E[f^2] \cong \int_0^n \frac{1}{n} \left(\int_0^x n \frac{1}{u^\alpha} du\right)^2 dx$$

$$= \frac{1}{(1-\alpha)^2(3-2\alpha)} n^{-2\alpha+4} + o(n^{-2\alpha+4})$$

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

$$\lambda_2 \cong O(n^{\alpha-2})$$

Markov 連鎖

このように $\mu_k \equiv \frac{1}{n}$ の形の一様分布になるときは、この方法で上手くいくことが多いようである。
一方で SIR モデルの近似のように μ_k が k の値に対して指数的に変化するときには上手くいかず、別の a_k を探す必要があるようである。

[2] SIR モデルに対応する $b_k = \beta k,$, $d_k = \gamma k$ のケース。
なおこのケースでは上で挙げた A はあまり良い値を与えないが実際に計算してみる。

補題

$\alpha > 0$ として

$$\int_1^n \frac{e^{\alpha x}}{x} dx = \frac{e^{\alpha n}}{\alpha n}(1 + o(1)), \quad (n \to \infty)$$

証明
次のように評価する。
$x \in [1, n]$ であれば

$$\int_1^n \frac{e^{\alpha x}}{x} dx \geq \int_1^n \frac{e^{\alpha x}}{n} dx = \frac{1}{\alpha n}(e^{\alpha n} - e) = \frac{e^{\alpha n}}{\alpha n}(1 + o(1))$$

である。

Markov 連鎖

$1 \ll k = n - l < n$ として（後で l のオーダーを決める。）

$$\int_1^n \frac{e^{\alpha x}}{x} dx = \int_1^k \frac{e^{\alpha x}}{x} dx + \int_k^n \frac{e^{\alpha x}}{x} dx$$

であるが、先ほどと同様に $1/x$ を区間の最大値で評価することで

$$\int_1^k \frac{e^{\alpha x}}{x} dx \leq \int_1^k e^{\alpha x} dx = \frac{1}{\alpha}(e^{\alpha k} - e) \leq \frac{e^{\alpha n}}{\alpha n} \frac{n}{e^{\alpha l}}$$
$$\int_k^n \frac{e^{\alpha x}}{x} dx \leq \int_k^n \frac{e^{\alpha x}}{k} dx = \frac{1}{\alpha k}(e^{\alpha n} - e^{\alpha k}) = \frac{e^{\alpha n}}{\alpha n} \frac{n}{n-l}(1 - \frac{1}{e^{\alpha l}})$$

と評価できるが、

$$\frac{n}{e^{\alpha l}} = o(1), \quad \frac{n}{n-l} = 1 + o(1), \quad e^{-\alpha l} = o(1)$$

と取れれば上からの評価も得られる。
実際 $l = \sqrt{n}$ と取れば良い。 $\square$

Markov 連鎖

先ほどと同じように、SIR モデルの A を積分で近似して

$$\tilde{\mu}_1 = 1, \quad \tilde{\mu}_k = \frac{b_k}{d_{k+1}}\tilde{\mu}_k = \left(\frac{\beta}{\gamma}\right)^k \frac{1}{k}, \quad Z = \sum_{k=1}^{n} \tilde{\mu}_k$$

ここで $\alpha = \log\frac{\beta}{\gamma}$ すなわち $e^\alpha = \frac{\beta}{\gamma}$ とすると、

$$Z \cong \int_1^n e^{\alpha x}\frac{1}{x}dx \cong \frac{e^{\alpha n}}{\alpha n}$$

$$\begin{aligned}
A_k &\cong \int_1^k (\int_k^n (\int_x^y \frac{1}{Z}\frac{1}{\gamma}\left(\frac{\beta}{\gamma}\right)^{x+y-u}\frac{1}{xy}du)dy)dx \\
&= \frac{1}{Z}\frac{1}{\gamma}\frac{1}{\alpha}\int_1^k (\int_k^n \{e^{\alpha y} - e^{\alpha x}\}\frac{1}{xy}dy)dx \\
&= \frac{1}{Z}\frac{1}{\gamma}\frac{1}{\alpha}\{\log k \int_k^n e^{\alpha y}\frac{1}{y}dy - (\log n - \log k)\int_1^k e^{\alpha x}\frac{1}{x}dx\}
\end{aligned}$$

Markov 連鎖

この A_k を k の関数とみなして微分を取って最大値をとる k の候補を計算すると

$$A_k' = -\log n \frac{e^{\alpha k}}{k} + \frac{1}{k}\int_1^n \frac{e^{\alpha x}}{x} dx$$

より、 $\alpha > 0$ の場合は

$$\int_1^n \frac{e^{\alpha x}}{x} dx \cong \frac{e^{\alpha n}}{\alpha n}$$

を使って

$$k_0 \cong n - \frac{1}{\alpha} \log(\alpha n \log n)$$

で最大値を取っている。

この k_0 に対して

$$\log k_0 = \log\{n + O(\log(n \log n))\} = \log\{n(1 + O(\frac{\log(n \log n)}{n}))\}$$

$$= \log n + O(\frac{\log(n \log n)}{n})$$

$$n - k_0 = \frac{1}{\alpha} \log(\alpha n \log n)$$

$$e^{\alpha k_0} = \exp \alpha\{n - \frac{1}{\alpha} \log(\alpha n \log n)\} = e^{\alpha n} \frac{1}{\alpha n \log n}$$

$$e^{\alpha n} - e^{\alpha k_0} = e^{\alpha n}(1 + o(1))$$

Markov 連鎖

これらを使って k_0 を先の A_k の計算に代入することで

$$A_{k_0} \cong \frac{1}{Z}\frac{1}{\gamma}\frac{1}{\alpha}\frac{e^{\alpha n}}{\alpha n}\log n \times (1+o(1)) = \frac{1}{\gamma\alpha}\log n \times (1+o(1))$$

と $\log n$ のオーダーが出てくる。
つまりスペクトルギャップの下からの評価として次を得る。

$$\lambda_2 \geq \frac{\gamma\alpha}{\log n}$$

定理に戻って特別な道（重み）を取ることでスペクトルギャップを評価する A としてより良い値を出す。（A に関して上からの評価を与える。）(実際は a_k を取りなおしていることに対応するが、参照した論文が Saloff-Coste の結果を基に書いているため、それに従った書き方にしてある。)
この方法（道の取り方）は本質的に
[Ka] N. Kahale, Nabil; A semidefinite bound for mixing rates of Markov chains. Proceedings of the Workshop on Randomized Algorithms and Computation (Berkeley, CA, 1995). Random Structures Algorithms 11 (1997), no. 4, 299-313.
による。（[Ka] では $b_k = \beta$, $d_k = \gamma$ の非対称ランダムウォークの時の適用例として挙げられたもの [Ka] のケースの方が μ_k に $1/k$ の項がなくなるため、見通しは良い。実際に計算が近似でなくできてしまう。)

Markov 連鎖

$0 < \delta < \alpha$ として $w_{k-1,k} = e^{\delta k}$ とする。
このとき $x < y$ に対して

$$\gamma_{x,y} = \sum_{k=x}^{y} \frac{1}{w_{k-1,k}} = \sum_{k=x+1}^{y} e^{-\delta k} \cong \int_x^y e^{\delta k} dk = \frac{1}{\delta}(e^{-\delta x} - e^{-\delta y})$$

（定理中定義されていないが max を取る前のものを A_k ととること で）

Markov 連鎖

$$
\begin{aligned}
A_k &= \frac{e^{\delta k}}{\frac{1}{Z}\gamma e^{\alpha k}} \sum_{1\le x<y\le n} \frac{1}{Z} e^{\alpha x} \frac{1}{x} \frac{1}{Z} e^{\alpha y} \frac{1}{y} \gamma_{x,y} \\
&\cong \frac{1}{Z}\frac{1}{\gamma\delta} e^{(\delta-\alpha)k} \int_1^k \left(\int_k^n \frac{e^{\alpha x}}{x} \frac{e^{\alpha y}}{y} (e^{-\delta x} - e^{-\delta y}) dy \right) dx \\
&\le \frac{1}{\gamma\delta}\frac{1}{Z} e^{(\delta-\alpha)k} \int_1^k \frac{e^{(\alpha-\delta)x}}{x} dx \int_k^n \frac{e^{\alpha y}}{y} dy \\
&\cong \frac{1}{\gamma\delta}\frac{1}{Z} e^{(\delta-\alpha)k} \frac{e^{(\alpha-\delta)k}}{k} \frac{e^{\alpha n}}{n} \cong \frac{1}{\gamma\delta}\frac{1}{k}
\end{aligned}
$$

これを k に関して最大を取るので $A = \frac{1}{\gamma\delta}$ と定数のオーダーであることが分かる。

Markov 連鎖

なお逆側の不等式に関しては、

$$f(x) = \begin{cases} 1 & x = n \\ 0 & \text{その他} \end{cases}$$

のような極端な関数を使うと定数オーダーになる。

Markov 連鎖

この結果を SIR モデルに戻すと $\alpha > 0$ すなわち $\beta > \gamma$ と感染者が増える率が感染者が復帰する率に比べて大きい場合には、人口サイズによらず、感染者はすぐに爆発することを示唆している。
注意：もしスペクトルギャップが n に依存して小さくなる（A が n の関数として大きくなる）場合は、総人口が多いほど、感染者の割合が大きくなる時間が遅くなることを示唆している。