数学特論 B4 のレジュメ

12 月 5 日 (木曜 2 限)

数学特論 B4 の講義 (稲浜担当分) について

これから 6–7 回にわたり，確率論に関するささやかなトピックを講義します．体系だった本格的な数学講義というわけではなく，どちらかというと「お話」といった雰囲気が強くなると思います (つまり厳密な証明はあまりしない，ということです).

教科書　教科書は特に指定しません．私のウェブサイトに講義レジュメを貼り付ける予定です.

授業内容　マルコフ連鎖（チェーン）についての入門的な講義をします．現代確率論はルベーグ積分論に基づいているのですが，これを勉強するのは 3 年前期なので，現時点では本格的なことはできません．そこで高校確率論と線形代数との知識があればある程度は楽しめるトピックとしてマルコフ連鎖を選びました．マルコフ性というのは現代確率論の中で中心的な概念なので，これを知っておくことは数学的にも有用だと思います．またマルコフ連鎖の数学以外の分野への応用も膨大にあります (例えば Wikipedia を参照せよ).

この講義の趣旨にのっとり，厳密な議論をするという形はとらず，「お話」をすることになります．今回はダイアコニスによる「トランプのシャッフル 7 回 (8 回) 問題」を「オチ」にして 1 月中旬ぐらいまで話す予定です．（なおここでいう「シャッフル」は「リフル・シャッフル (riffle shuffle)」の意味で，トランプの山をまず二つに分けて左右の手に持ち，そのあと左右の手から下に向けてパラパラと落とす，というややカッコイイ切り方のことです)．ダイアコニス先生は

トランプは 7 回 (8 回?) のシャッフルで十分よく混ざる

ことを証明しました．もちろんこの定理を厳密に証明するのは授業時間の関係で無理なのですが,

- この問題をどう数学的に定式化するのか，マルコフ連鎖がなぜ出てくるのか,
- 特に「よく切れる」という漠然とした感じをどう定式化するのか,
- 定式化した後，どういう道筋で議論すると証明できるか,

などが「鑑賞のポイント」になると思います．この問題を巨大な有限群である 52 次対称群 (置換群) S_{52} の上のマルコフ連鎖として定式化するのが一番大きな話の流れです.

成績　基本的に宿題レポートで決まります．レポート問題は比較的簡単な問題を丁寧に説明するような感じにして，出題したいと思います (指定教科書がないため).

なお出席はとりません。よって授業中の私語と食事は禁止（ただしフタの付いた容器にいれた飲料のみ例外的に認めます）だから欠席しようと思えばできますが、あまりさぼっていると当然おちこぼれやすくなります。

担当者について

稲浜　譲（数理学研究院、W1-D-733 室)
E-mail: `inahama@math.kyushu-u.ac.jp`
Website: http://www2.math.kyushu-u.ac.jp/ inahama/

準備（とくに条件付き確率，確率変数，確率過程について）

• この講義を通じて, $\mathbf{N} = \{0, 1, 2, \ldots\}$ とします．またマルコフ連鎖が動き回る集合 S は有限集合だとします．S を**状態空間 (state space)** と呼びます．

• $\mu = \{\mu(i)\}_{i \in S}$ が S 上の (確率) 分布であるとは $\mu(i) \geq 0$ $(\forall i \in S)$ で $\sum_{i \in S} \mu(i) = 1$ ということ．S 上の (確率) 分布の全体を $\mathrm{Prob}(S)$ と書く．特に 1 点集合 $\{j\}$ の上に全ての確率 (重み) が集中している分布を δ_j と書く．この場合 $\delta_j(i)$ はクロネッカーのデルタになる．なおマルコフ連鎖論では，分布 μ は「横ベクトル」とみなす．

• まず高校数学流の確率論の枠組みで議論します．現代確率論はコルモゴロフ流にルベーグ積分の意味で確率を基礎づけるのが普通ですが，２年生の段階ではまだルベーグ積分を習ってないので仕方ありません．

• 条件付き確率も高校生風に定義します．すなわち事象 B が与えられたときの事象 A の条件付き確率を

$$\mathbf{P}(A \mid B) := \frac{\mathbf{P}(B \cap A)}{\mathbf{P}(B)}$$

と定めます．もちろん $\mathbf{P}(B) = 0$ のときは定義できません．ところがマルコフ連鎖の文献には独特の習慣があります．$\mathbf{P}(A \mid B)$ という記号を書くときに，$\mathbf{P}(B) = 0$ だったらどうするかを説明しないのが普通です．(しかし以下のどちらかの理由で大丈夫です．(i) そもそも $\mathbf{P}(B) = 0$ となる場合を考える必要がない，または (ii) $\mathbf{P}(B)\mathbf{P}(A \mid B)$ のように「セット」で出てくる.)

• S 値の確率変数 X も高校風に理解する．これは $\mu(i) := \mathbf{P}(X = i)$ と定めることにより自然に S 上の分布を誘導する．確率変数 X の「正体」はむしろ誘導された分布である．

• S 値の確率過程 $(X_k)_{k \in \mathbf{N}} = (X_0, X_1, X_2, \ldots)$ とは $\mathbf{N}$ で添え字付けられた確率変数の列のことである．このような確率過程の分布とは，

$$\mathbf{P}(X_0 = i_0, X_1 = i_1, \ldots, X_N = i_N), \qquad (N \in \mathbf{N},\ i_0, \ldots, i_N \in S)$$

のことである．同値だが，任意の $N \in \mathbf{N}$ を固定するごとに $(X_0, \ldots, X_N)$ を $\mathrm{Map}(\{0, \ldots, N\} \to S)$ 値の確率変数とみなした場合にこの集合に誘導される分布 (の族) のことである．確率過程の「正体」もむしろその分布であり，確率論の習慣では同じ分布を持つ確率過程は同じとみなす．(注意：時間の集合を $\mathbf{N}$ ではなく，その有限部分集合にとりかえた場合も確率過程の定義はまったく同様にできる．)

• 推移確率行列の導入．行列 $P = (P(i,j))_{i,j \in S}$ が**推移確率行列** であるとは，

$$P(i,j) \geq 0 \quad (\forall (i,j) \in S \times S) \qquad \text{かつ} \qquad \sum_{j \in S} P(i,j) = 1 \quad (\forall i \in S)$$

を満たすことである．これはつまり各行ベクトル (横ベクトル) が S 上の分布だというのと同値である．

マルコフ連鎖の定義 マルコフ連鎖というのは確率過程のうち次のような非常に良い性質を持つものをさす．なおこの講義では斉時的 (時間的に一様な) マルコフ連鎖のみを扱う．

次の時刻 $n+1$ における状態は現在の時刻 n における状態のみで決まり過去の履歴 (時刻 $0, 1, \ldots, n-1$ における状態) にはよらない

直感的な説明だが，これだけでは数学的に何を言っているかはっきりしないので，正式な定義を与える．この定義内の確率が時刻 N について整合性があることに注目せよ．(すなわち，最初から番号 $N-1$ を考えた場合の確率と，まず番号 N に関する確率を考えのちに i_N について和を取ったものとが一致している．)

定義 1. *μ を有限集合 S 上の分布とし，P を推移確率行列だとする．S 値の確率過程 $(X_k)_{k\in\mathbf{N}} = (X_0, X_1, X_2, \ldots)$ が初期分布 μ，推移確率行列 P を持つ (斉時的) マルコフ連鎖であるとは，*

$$\mathbf{P}(X_0 = i_0, X_1 = i_1, \ldots, X_N = i_N) = \mu(i_0)P(i_0, i_1) \times \cdots \times P(i_{N-1}, i_N) \qquad (N \in \mathbf{N},\ i_0, \ldots, i_N \in S)$$

を満たすことである．

以下ではこれを単に (P, μ) マルコフ連鎖とよぶ．もし $\mathbf{P}(X_n = i) \neq 0$ ならば、$\mathbf{P}(X_{n+1} = j \mid X_n = i) = P(i, j)$ に注意せよ．この左辺の条件付き確率が時刻 n に無関係なので「定常」とよぶ．

定理 2. *S 値の確率過程 $(X_k)_{k\in\mathbf{N}}$ が (P, μ) マルコフ連鎖であることは次の条件と同値である．X_0 の誘導する分布が μ であり，かつ任意の $N, i_0, \ldots, i_{N-1}, i, j \in S$ s.t. $\mathbf{P}(X_0 = i_0, X_1 = i_1, \ldots, X_{N-1} = i_{N-1}, X_N = i) \neq 0$ に対して*

$$\mathbf{P}(X_{N+1} = j \mid \underbrace{X_0 = i_0, X_1 = i_1, \ldots, X_{N-1} = i_{N-1}}_{\text{過去の履歴}}, X_N = i) = P(i, j)$$

以下では <u>(P, μ) マルコフ連鎖の分布を $\mathbf{P}_\mu$ と書き，特に $\mu = \delta_j$ のときは単に $\mathbf{P}_j$ と書く．</u> したがって例えば

$$\mathbf{P}_\mu\Big((X_k)_{k\in\mathbf{N}}\text{に関する条件}\Big)$$

と書いたらもう自動的に (X_k) は (P, μ) マルコフ連鎖だと思っているわけである．

マルコフ連鎖は「重み (確率) の運び屋」だと自然に思える．時刻 n に単位質量が $i \in S$ にあった場合に，時刻 $n+1$ においてはその重みが S 上の別の点 j にどうばらまかれるかを指定するのが推移確率行列 $P = (P(i, j))$ だというわけだ．

またマルコフ連鎖を定義した際には，「時刻 n まで行ってから時刻 $n+1$ でどうなるか」を指定したわけだが，実はもっと一般に「時刻 n まで行ってから時刻 $n+m$ でどうなるか」も自動的にわかる．(これをチャップマン・コルモゴロフ方程式とよぶことが多い)

定理 3. *$n, m \in \mathbf{N}$ と $i, j \in S$, $\mu \in \mathrm{Prob}(S)$ は任意だとする．*

- *$\mathbf{P}_\mu(X_n = j) = (\mu P^n)_j$．すなわち，$(X_k)_{k\in\mathbf{N}}$ が (P, μ) マルコフ連鎖ならば，X_n が S 上に誘導する分布は μP^n (横ベクトル！) である．*
- *$\mathbf{P}_i(X_n = j) = \mathbf{P}_\mu(X_{n+m} = j \mid X_m = i) = P^n(i, j) \qquad$ (P の n 乗の (i, j) 成分)*

証明は定義から出発してごく簡単にできる．さらにもっと一般に次が示せる (しかもそんなに難しくない)．
“$\mathbf{P}_\mu(X_m = i) \neq 0$ ならば，時刻 m を時刻の原点にとり直した過程 $(X_{m+k})_{k=0,1,2,\ldots}$ は条件付き確率 $\mathbf{P}_\mu(\bullet \mid X_m = i)$ の下で (P, δ_i) マルコフ連鎖になる”．