マルコフ連鎖の基本

@litharge3141

2020 年 5 月 2 日

1　マルコフ連鎖の基本

有限ないし高々可算の状態空間を持ち，かつ離散的という最もシンプルな場合を通して，マルコフ過程の概念を整理する．大数の強法則を証明したことがある程度の知識を仮定する．本文全体を通して I を高々可算集合とし，その σ-代数として I の部分集合全体 $\mathcal{P}(I)$ を取る．また，自然数の全体 $\mathbb{N}$ は 0 を含むものとする．

1.1　マルコフ連鎖の定義

時間変化する確率変数は，各時刻で I 上の確率分布を与える．それが時間無限大でどうなるかとか，そういう問題を考えたい．I は高々可算だから，I 上の分布はうまく言い換えられる．

Definition 1.1. 写像 $\nu\colon I \to [0,1]$ が確率ベクトルであるとは，$\sum_{i\in I}\nu(i) = 1$ が成立することをいう．ν を $(\nu_i)_{i\in I}$ ともかき，$\nu(i)$ を単に ν_i とかく．写像 $A\colon I \times I \to [0,1]$ が確率行列であるとは，任意の $i \in I$ に対して，$\sum_{j\in I} A(i,j) = 1$ が成立することをいう．A を $(A_{ij})_{i,j\in I}$ ともかき，$A(i,j)$ を単に A_{ij} とかく．

確率ベクトル ν に対して $P\colon \mathcal{P}(I) \to [0,1]$ を $P(E) \coloneqq \sum_{i\in E}\nu_i$ により定めると，P は I 上の確率測度となる．逆に I 上の確率測度 P に対して $\nu_i \coloneqq P(\{i\})$ と定めると，ν は確率ベクトルになる．したがって，I 上の確率分布を定めることは，確率ベクトルを定めることと同じである．

Example 1.1. $i \in I$ に対して，確率ベクトル δ_i を

$$\delta_i(j) = \begin{cases} 1 & (j = i) \\ 0 & (j \neq i) \end{cases}$$

によって定めることができる．この記号 δ_i は後で用いる．

Theorem 1.1. 確率ベクトル ν と確率行列 A の積 $\nu A\colon I \to [0,1]$ を $j \in I$ に対し $\nu A(j) \coloneqq \sum_{i\in I}\nu_i A_{ij}$ によって定めることができ，νA は確率ベクトルになる．確率行列 A, B の積 $AB\colon I \times I \to [0,1]$ を $AB(i,j) = \sum_{k\in I} A_{ik}B_{kj}$ によって定めることができ，AB は確率行列になる．

Proof. 非負項の二重級数はいつでも和をとる順序を交換できるので，定理が従う．　□

確率行列は確率ベクトルの変換を定めるが，これは分布の変換を定めているのと同じことである．

Example 1.2 (破産問題)**.** $I = \mathbb{Z}$ とし, X_n を n 回目の試行の後の所持金とする．確率 $1/2$ で $X_{n+1} = X_n + 1$ とし，確率 $1/2$ で $X_{n+1} = X_n - 1$ とするような賭けを考える．最初の所持金を i とすると，初期分布は δ_i で

与えられる．この試行の確率行列は

$$A_{ij} = \begin{cases} 1/2 & (j = i \pm 1) \\ 0 & (\text{それ以外}) \end{cases}$$

によって与えられる．

初期分布と時間によらない分布の変換が与えられているとき，それにしたがって発展するような分布を持つ確率変数列のことをマルコフ連鎖という．素直に数式で表現すると次のようになる．

Definition 1.2 (Markov 連鎖っぽいなにか)**.** 確率ベクトル ν と確率行列 A が与えられたとする．$(\Omega, \mathcal{F}, P)$ を確率空間として，Ω から I への確率変数列 $(X_n)_{n=0}^{\infty}$ が遷移行列 A，初期分布 ν をもつマルコフ連鎖であるとは，任意の $E \subset I$ に対して $P(X_0 \in E) = \sum_{i \in E} \nu_i$ が成立し，さらに任意の $n \in \mathbb{N}$ と任意の $E \subset I$ に対して $P(X_{n+1} \in E) = \sum_{i \in E} \sum_{j \in I} P(X_n = j) A_{ji}$ が成立することをいう．

この定義は任意の $i \in I$ に対して $P(X_{n+1} = i) = \sum_{j \in I} P(X_n = j) A_{ji}$ が成立すること，と書き換えてもよい．この定義は次の定義を採用すればそれから導かれる．

Definition 1.3 (Markov 連鎖)**.** 確率ベクトル ν と確率行列 A が与えられたとする．$(\Omega, \mathcal{F}, P)$ を確率空間として，Ω から I への確率変数列 $(X_n)_{n=0}^{\infty}$ が遷移行列 A，初期分布 ν をもつマルコフ連鎖であるとは，任意の $n \in \mathbb{N}$ と任意の $i_0, \ldots, i_n \in I$ に対して，$P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \nu_{i_0} A_{i_0 i_1} A_{i_1 i_2} \cdots A_{i_{n-1} i_n}$ が成立することをいう．マルコフ連鎖を初期分布と遷移行列，確率測度との組にして $((X_n)_{n=0}^{\infty}, A, \nu, P)$ と書き表すこともある．

この定義は任意の $i, j \in I$ に対して $P(X_0 = i) = \nu_i$ および $P(X_{n+1} = i \mid X_n = j) = A_{ji}$ が成立すること，と書き換えてもよい．っぽいなにかのほうだと後で証明が回らなくなるようなので，このノートではこちらの定義を採用する．実は同値だったとか，具体的にどこがまずいのかとか分かったら追記する．マルコフ連鎖が常に存在するかは自明ではないので，存在を示す．

Theorem 1.2. 確率ベクトル ν と確率行列 A が与えられたとする．このとき，ある確率空間 $(\Omega, \mathcal{F}, P)$ と Ω から I への確率変数列 $(X_n)_{n=0}^{\infty}$ が存在して，遷移行列 A，初期分布 ν をもつマルコフ連鎖となる．

Proof. コルモゴロフの拡張定理を使うために，便宜的に $\mathbb{R}^n$ 上の確率測度を構成する．I は高々可算集合，したがって単射 $\phi\colon I \to \mathbb{N}$ が存在する．$n = 1, 2, \ldots$ に対して写像 $\mu_n\colon \mathcal{B}(\mathbb{R}^n) \to \mathbb{R}$ を $\mu_n(E) \coloneqq \sum_{(\phi(i_0), \ldots, \phi(i_{n-1})) \in E} \nu_{i_0} A_{i_0 i_1} A_{i_1 i_2} \cdots A_{i_{n-2} i_{n-1}}$ と定めると，これは $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ 上の確率測度になる．さらに拡張定理の仮定である整合条件 $\mu_{n+1}(E \times \mathbb{R}) = \mu_n(E)$ が満たされることも示せるので，コルモゴロフの拡張定理から $\mathbb{R}^{\mathbb{N}}$ 上の確率測度 μ で任意の $n = 1, 2, \ldots$ と任意の $A \in \mathcal{B}(\mathbb{R}^n)$ に対して $\mu(A \times \mathbb{R}^{\mathbb{N}}) = \mu_n(A)$ を満たすものが一意的に存在する．$n = 1, 2, \ldots$ に対して $Z_n\colon \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ を n 成分への射影 $Z_n((x_1, x_2, \ldots)) = x_n$ によって定めて，$n \in \mathbb{N}$ に対して $Y_n \coloneqq Z_{n+1}$ とする．$(Y_n)_{n=0}^{\infty}$ はほとんどいたるところ $\phi(I)$ に値を取る確率変数列で，任意の $n \in \mathbb{N}$ と任意の $i_0, \ldots, i_n \in I$ に対して，$P(Y_0 = \phi(i_0), Y_1 = \phi(i_1), \ldots, Y_n = \phi(i_n)) = \nu_{i_0} A_{i_0 i_1} A_{i_1 i_2} \cdots A_{i_{n-1} i_n}$ を満たす．零集合上の値を修正して I 値にすれば，求めるマルコフ連鎖 $(X_n)_{n=0}^{\infty}$ が得られる． $\square$

しばしば現れてきた $\{\omega \mid X_0(\omega) = i_0, \ldots, X_n(\omega) = i_n\}$ のような形の事象全体で作られる，自然な増大情報系という概念を導入しておくと便利である．

Definition 1.4. $(\Omega,\mathcal{F},P)$ を確率空間として，Ω から I への確率変数列 $X=(X_n)_{n=0}^{\infty}$ が与えられたとする．$n\in\mathbb{N}$ に対して $\mathcal{F}$ の部分 σ-代数 $\mathcal{F}_n$ を $\mathcal{F}_n\coloneqq\sigma(X_0,\dots,X_n)$ と定める．$(\mathcal{F}_n)_{n=0}^{\infty}$ を X に関する自然な増大情報系という．

I は高々可算集合なので，$F\subset I^n$ に対して $E=(X_0,\dots,X_n)^{-1}(F)$ という形でかける E は $E=\bigcup_{(i_0,\dots,i_n)\in F}\{X_0=i_0,\dots,X_n=i_n\}$ という交わらない可算和で書き直せる．このような元全体で生成されるのが $\mathcal{F}_n$ である．各時間 n において，X に関連する事象で確率を計算し得るものは全てここに属する．そのような中で最小のもの，というのが自然という言葉の意味である．

マルコフ連鎖は時刻 $n+1$ での確率分布が時刻 n での分布にのみ依存するという性質を持つ．これをマルコフ性という．それを示そう．

Theorem 1.3 (マルコフ性 1)**.** $(\Omega,\mathcal{F},P)$ を確率空間として，Ω から I への確率変数列 $(X_n)_{n=0}^{\infty}$ は遷移行列 A，初期分布 ν をもつマルコフ連鎖であるとする．このとき，$P(X_n=i_n,\dots,X_0=i_0)>0$ が成り立つような任意の $n\in\mathbb{N}$ と任意の $i_0,\dots,i_{n+1}\in I$ に対して，$P(X_{n+1}\mid X_n=i_n,\dots,X_0=i_0)=A_{i_ni_{n+1}}$ が成立する．

Proof. $n\in\mathbb{N}$ と $i_0,\dots,i_{n+1}\in I$ が任意に与えられたとすると，$P(X_{n+1}=i_{n+1}\mid X_n=i_n,\dots,X_0=i_0)P(X_n=i_n,\dots,X_0=i_0)=P(X_{n+1},X_n=i_n,\dots,X_0=i_0)$ が成り立つ．$(X_n)_{n=0}^{\infty}$ がマルコフ連鎖であるから，$P(X_n=i_n,\dots,X_0=i_0)=\nu_{i_0}A_{i_0i_1}\cdots A_{i_{n-1}i_n}$ および $P(X_{n+1},X_n=i_n,\dots,X_0=i_0)=\nu_{i_0}A_{i_0i_1}\cdots A_{i_ni_{n+1}}$ となる．これを代入して計算をすれば定理を得る． □

マルコフ性は別の定式化をすることもできる．次の性質も成り立つ．

Theorem 1.4 (マルコフ性 2)**.** $(\Omega,\mathcal{F},P)$ を確率空間として，Ω から I への確率変数列 $(X_n)_{n=0}^{\infty}$ は遷移行列 A，初期分布 ν をもつマルコフ連鎖であるとする．$P(X_m=i)>0$ が成り立つような任意の $i\in I$ と任意の $m\in\mathbb{N}$ に対して，$Y_n\coloneqq X_{n+m}$ として $(Y_n)_{n=0}^{\infty}$ を定めると，$((Y_n)_{n=0}^{\infty},A,\delta_i,P(\cdot\mid X_m=i))$ はマルコフ連鎖となる．

Proof. $i\in\mathbb{N}$ と $i_m,i_{m+1},\dots,i_{m+n}\in I$ が任意に与えられたとする．

$$
\begin{aligned}
&P(Y_0=i_m,\dots,Y_n=i_{m+n}\mid X_m=i)\\
&=P(X_m=i,X_m=i_m,\dots,X_{m+n}=i_{m+n})/P(X_m=i)\\
&=\frac{\sum_{i_0,\dots,i_{m-1}\in I}P(X_0=i_0,\dots,X_{m-1}=i_{m-1},X_m=i,X_m=i_m,\dots,X_{m+n}=i_{m+n})}{\sum_{j_0,\dots,j_{m-1}\in I}P(X_0=j_0,\dots,X_{m-1}=j_{m-1},X_m=i)}\\
&=\frac{\sum_{i_0,\dots,i_{m-1}\in I}\delta_i(i_m)\nu_{i_0}A_{i_0i_1}\cdots A_{i_{m+n-1}i_{m+n}}}{\sum_{i_0,\dots,i_{m-1}\in I}\nu_{j_0}A_{j_0j_1}\cdots A_{j_{m-1}i}}\\
&=\delta_i(i_m)A_{i_mi_{m+1}}\cdots A_{i_{m+n-1}i_{m+n}}\frac{\sum_{i_0,\dots,i_{m-1}\in I}\nu_{i_0}A_{i_0i_1}\cdots A_{i_{m-1}i}}{\sum_{i_0,\dots,i_{m-1}\in I}\nu_{j_0}A_{j_0j_1}\cdots A_{j_{m-1}i}}\\
&=\delta_i(i_m)A_{i_mi_{m+1}}\cdots A_{i_{m+n-1}i_{m+n}}
\end{aligned}
$$

したがって，示された． □

Theorem1.4 を使うとより強い次の結果を示すことができる．

Theorem 1.5 ($\mathcal{F}_m$ との独立性)**.** $(\Omega,\mathcal{F},P)$ を確率空間として，Ω から I への確率変数列 $(X_n)_{n=0}^{\infty}$ は遷移行

列 A，初期分布 ν をもつマルコフ連鎖であるとする．$P(X_m = i) > 0$ が成り立つような任意の $i \in I$ と任意の $m \in \mathbb{N}$ に対して，$Y_n \coloneqq X_{n+m}$ として $(Y_n)_{n=0}^{\infty}$ を定める．このとき，確率空間 $(\Omega, \mathcal{F}, P(\cdot \mid X_m = i))$ の下で $(Y_n)_{n=0}^{\infty}$ と $\mathcal{F}_m$ は独立である．

Proof. $n \in \mathbb{N}$ と $j_0, j_1, \ldots, j_m, i_m, i_{m+1}, \ldots, i_{m+n} \in I$ が任意に与えられたとする．$E \coloneqq \{X_0 = j_0, \ldots, X_m = j_m\}$ とおく．$E \in \mathcal{F}_m$ は定義から明らかである．この E に対して

$$\begin{aligned} &P((Y_0 = i_m, \ldots, Y_n = i_{m+n}) \cap E \mid X_m = i) \\ &= P(Y_0 = i_m, \ldots, Y_n = i_{m+n} \mid X_m = i) P(E \mid X_m = i) \end{aligned} \tag{1}$$

を示す．$(X_n)_{n=0}^{\infty}$ のマルコフ性から

$$\begin{aligned} &P((Y_0 = i_m, \ldots, Y_n = i_{m+n}) \cap E \mid X_m = i) \\ &= \delta_i(i_m) P(X_0 = j_0, \ldots, X_m = j_m, X_m = i_m, \ldots, X_{m+n} = i_{m+n}) / P(X_m = i) \\ &= \delta_{j_m i_m} \delta_i(i_m) \nu_{j_0} A_{j_0 j_1} \cdots A_{i_{m+n-1} i_{m+n}} / P(X_m = i) \end{aligned}$$

および

$$\begin{aligned} P(E \mid X_m = i) &= \delta_i(j_m) P(X_0 = j_0, \ldots, X_{m-1} = j_{m-1}, X_m = j_m) / P(X_m = i) \\ &= \delta_i(j_m) \nu_{j_0} A_{j_0 j_1} \cdots A_{j_{m-1} j_m} / P(X_m = i) \end{aligned}$$

が得られる．また Theorem1.4 により，

$$P(Y_0 = i_m, \ldots, Y_n = i_{m+n} \mid X_m = i) = \delta_i(i_m) A_{i_m i_{m+1}} \cdots A_{i_{m+n-1} i_{m+n}}$$

となる．これらの式から (1) が得られた．一般の $E \in \mathcal{F}_m$ に対して証明するため，上の結果を利用する．

$$\begin{aligned} \mathcal{C} \coloneqq \{E \mid & \forall n \in \mathbb{N}, \forall i_m, \ldots, i_{m+n} \in I, \\ & P((Y_0 = i_m, \ldots, Y_n = i_{m+n}) \cap E \mid X_m = i) \\ & = P(Y_0 = i_m, \ldots, Y_n = i_{m+n} \mid X_m = i) P(E \mid X_m = i)\} \end{aligned}$$

とおく．$E = \{X_0 = j_0, \ldots, X_m = j_m\}$ の形で表される E の全体は $\mathcal{C}$ に含まれるから，$\mathcal{C}$ が σ-代数であることを示せばよいが，それは $P(\cdot \mid X_m = i)$ が確率測度であることから直ちにしたがう．よって示された． □

1.2 到達確率と差分作用素

マルコフ性の応用として到達確率と差分作用素を扱う．