強化学習の TD(0) 法:数学的な証明について

川端諒

概要

機械学習のうち，教師なし学習の一つである強化学習というものがある．他の機械学習と同様に，強化学習にも様々な学習アルゴリズムが存在する．その中の一つに TD(0) 法と呼ばれるものがある．強化学習が解説されている多くの教科書や技術書では，TD(0) 法のアルゴリズムだけが載っていて，数学的な証明は無いかあっても正確な証明が載っていることはほぼ無い．そこで，この論文ではごく簡単なモデルにおいて TD(0) 法が数学的に正しいアルゴリズムであることを証明する．

目次

1 研究背景

機械学習がしばしば様々な場面で利用されている．様々な教科書や技術書が世に出回っている．それらの本の中には具体的な学習のアルゴリズムの解説や，そのアルゴリズムを具体的な問題に適用したときの例などが描かれている．しかし概要に述べたように，その様々な教科書や技術書において数学的な根拠は省かれている．たとえば，「アルゴリズムに従うとある極限値が得られる」といったことが教科書に書かれていた場合を考える．教科書にはそのアルゴリズムの具体的な手順が書かれていて，その次には実際そのアルゴリズムをパソコンで実装するとどうなるかという話に進む．アルゴリズムに従うことにより望ましい結果が得られるという数学的な理由や証明は書かれていないことが多い．また，書かれていたとしても高度に数学的だからという理由で省かれたり，数式を用いて証明のスケッチだけを書いている場合がほとんどである．

その一方で，筆者は大学四年生および修士一年生でマルコフ連鎖を学んだ．マルコフ連鎖が利用されている機械学習の一つとして強化学習と呼ばれるものがある．修士一年の後半ではその強化学習について理論を学んでいたが，先ほど述べたように数学的根拠が省かれて説明されることが多くあった．用いていた教科書 (Csabà Szepesvàri [1]) では TD(0) 法というアルゴリズムが紹介されていた．

価値関数の式中にはマルコフ連鎖 (X_n) の n に関する無限級数がある．すなわち，コンピュータ上でこの価値関数を計算しようとしても真の値を算出することは難しい．なぜなら，時刻 n を無限大まで大きくしてシミュレーションを行わなければならないからだ．したがって，実際に計算機を用いて学習を行う場合は価値関数の値は推定しなければならない．TD(0) 法はその推定方法の一つであった．TD(0) 法のアルゴリズムに関連して，この論文では次のことを行った．すなわち，最適な方策を見つけて価値関数の最大化をするという，強化学習の目標を達成する数理的なアルゴリズムを与えその証明をした．さらに具体的に述べると，簡単なモデルを数学的に定義し，そのモデル上で価値関数を最大化する方策をみつけるアルゴリズムを与え，それが数学的に正しいアルゴリズムであることの証明を与えた．

また，強化学習の数学モデルである MRP の具体的な構成の一環として，状態空間が $[0, 1]$ であるマルコフ連鎖の構成を行った．

2　強化学習と TD(0) 法の概説

まずは強化学習についてその概要を述べる．

強化学習のモデルには大きく「状態」と「行動」と「報酬」の三つの要素がある．ある状態にいるとき，ある行動を取ることにより別の状態に遷移し，遷移した状態に対応して「報酬」を受け取る．そしてまた，もう一度行動をとることにより，別の状態に遷移して，報酬を受け取る．このプロセスを繰り返すことで，様々な状態を移動しながらいくらかの報酬を得ることができる．強化学習の目的は，「行動」の選び方を最適なものにすることで最終的に得られる報酬の総和の期待値を最大化することである．身近な例としては，AlphaGo と呼ばれる囲碁をするマシーンや Atari2600 というゲームがある．(久保隆宏 [2])

なお，この節の内容はその多くを Csabà Szepesvàri [1]，森村哲郎 [3] を参考にした．

2.1　MDP/MRP

強化学習を数式を用いて扱うための基礎的なモデルが MDP である．

以下，$(\Omega, \mathcal{F})$ を可測空間とする．この確率測度や確率変数を定める場合，全てのこの測度空間上で定める．

強化学習の議論する際のモデルとして最も基本的なものが MDP である．

定義 2.1. 以下に定める記号の五つ組 $(\mathcal{X}, \mathcal{A}, \mathcal{P}_0)$ を MDP (Markov Decision Process) という．

- $\mathcal{X}$:状態の集合，高々可算
- $\mathcal{A}$:行動の集合，高々可算
- 写像 $\mathcal{P}_0$:遷移確立カーネル．$(x,a) \in \mathcal{X} \times \mathcal{A}$ に対して，$\mathcal{X} \times \mathbb{R}$ 上で定義された関数

$$\mathcal{P}_0(\cdot \mid x, a)$$

を対応付ける．この関数 $\mathcal{P}_0(\cdot \mid x, a)$ は次を満たす．

$$0 \leq \mathcal{P}_0(\cdot \mid x, a) \leq 1, \quad \sum_U \mathcal{P}_0(U \mid x, a) = 1.$$

$U \subset \mathcal{X} \times \mathbb{R}$ に対して

$$\mathcal{P}_0(U \mid x, a)$$

は，状態が x において行動 a をとったとき，次の状態と得られる「報酬」が U に入っている確率を表す．

定義 2.1 において「報酬」という言葉があるが，これはのちに定義を述べる．また，$(\mathcal{X}, 2^{\mathcal{X}}), (\mathcal{A}, 2^{\mathcal{A}})$ で可測空間を定める．

定義 2.2. 状態遷移確率カーネル $\mathcal{P} : \mathcal{X} \times \mathcal{A} \times \mathcal{X} \to [0,1]$ は

$$\mathcal{P}(x,a,y) = \mathcal{P}_0(\{y\} \times \mathbb{R} \mid x,a)$$

である．

定義 2.3. MDP において行動の集合の要素の個数が 1 つのみのとき，その MDP を MRP(Markov Reward Process) という．

MDP は強化学習におけるモデルで，$\mathcal{X}$ が問題の環境を表している．

定義 2.4. γ は割引率と呼ばれるハイパーパラメータである．$0 \leq \gamma \leq 1$ をみたす．

2.2 MDP によって定まる状態, 行動, 報酬の系列

定義 2.5. MDP を一つ固定する．報酬関数 $R : \mathcal{X} \times \mathcal{A} \to \mathbb{R}$ はその分布が $\mathcal{P}_0$ に従うような関数である．すなわち，状態が x, 行動が a であるときに，遷移する状態が y で，報酬が $U \subset \mathcal{B}(\mathbb{R})$ に含まれている確率は

$$\mathcal{P}_0(\{y\} \times U \mid x,a)$$

である．また，報酬関数 R は有界関数であるとする．

簡単のために，報酬関数を $R : \mathcal{X} \ni y \mapsto u \in \mathbb{R}$ とすることもある．この場合，報酬関数は現在の状態にのみに依存した分布にしたがっている．

定義 2.6. 即時報酬関数 $r : \mathcal{X} \times \mathcal{A} \to \mathbb{R}$ は状態が x, 行動を a を取った時の報酬関数の期待値である．つまり，

$$r(x,a,y) = \mathbb{E}[R(x,a,y)].$$

ただし，ここのでの期待値は確率測度 $\mathcal{P}_0(\{y\} \times \cdot \mid x,a)$ によるものである．

R が有界関数なので r も有界関数である．

定義 2.7. π が定常方策とは，$\pi : \mathcal{X} \times \mathcal{A} \to [0,1]$ という写像で次の条件を満たすもの．

$$0 \leq \pi(x,a) \leq 1, \quad \sum_{a \in \mathcal{A}} \pi(x,\cdot) = 1 \quad \forall x \in \mathcal{X}$$

という写像であることをいう．これは状態 x において行動 a が選択される確率を表す．条件付き確率の意味を込めて $\pi(a \mid x)$ とも書く．定常方策 π 全体の集合を Π で表すことにする．

定常方策 $\pi \in \Pi$ を任意に一つ固定する．$t \in \mathbb{N}$ は現在の時刻とする．X_0^π を時刻 0 における状態を表す Ω 上の確率変数とする．このとき，A_0^π を定常方策 π に従うように定める．すなわち，$A_0^\pi : \Omega \to \mathcal{A}$ であり，

$$A_0^\pi \sim \pi(\cdot \mid X_0^\pi).$$

状態遷移確立カーネルから，$X_1^\pi : \Omega \to \mathcal{X}$ が定まる．つまり，

$$X_1^\pi \sim \mathcal{P}(X_0^\pi, A_0^\pi, \cdot).$$

X_1^π を観測すれば，報酬 $R_1^\pi : \Omega \to \mathbb{R}$ が定まる．すなわち，

$$R_1^\pi \sim \mathcal{P}_0(X_1^\pi \times \{\cdot\} \mid X_0^\pi, A_0^\pi).$$

こうして得られた X_1^π に対して同様の操作をおこなうことで $(A_1^\pi, X_2^\pi, R_2^\pi)$ を得る．これを繰り返して，確率過程 $((X_t^\pi, A_t^\pi, R_{t+1}^\pi); t \geq 0)$ が決まる．

3 節において方策や報酬などの設定を簡単にした場合の確率過程 $((X_t^\pi, A_t^\pi, R_{t+1}^\pi); t \geq 0)$ の具体的な構成を述べる．そしてその場合において (X_t^π) がマルコフ連鎖であることを示す．

2.3　価値関数

MDP と定常方策を一つ固定し，その MDP によって定まる確率過程 $((X_t^\pi, A_t^\pi, R_{t+1}^\pi); t \geq 0)$ を考える．

定義 2.8. 収益 $\mathcal{R}$ は次で定義される．

$$\mathcal{R} = \sum_{t=0}^{\infty} \gamma^t R_{t+1}^\pi.$$

価値関数は強化学習において重要な値である．

定義 2.9. 価値関数 $V : \Pi \times \mathcal{X} \to \mathbb{R}$ とは，

$$V(\pi, x) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R_{t+1} \;\middle|\; X_0 = x\right].$$

なお価値関数を考える際は，X_0 は任意の $x \in \mathcal{X}$ について，$\mathbb{P}(X_0 = x) > 0$ を満たしているとする．

$V(\pi, x)$ は，定常方策 π に従う MDP が状態 x からスタートしたとき，得られる収益の期待値を表している．

2.4　ベルマン作用素

ベルマン作用素は強化学習において基本的な作用素である．マルコフ性と条件付き平均の性質を用いて価値関数を捉えなおしたものである．補題 4.3 も参照．

定義 2.10. MDP を一つ固定する．$\pi \in \Pi$ を定常方策とする．π に従うベルマン作用素 $T^\pi : \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}}$ を次の式により定義する．

$$(T^\pi V)(x) := r(x, \pi(x)) + \gamma \sum_{y \in \mathcal{X}} \mathcal{P}(x, \pi(x), y) V(y).$$

$0<\gamma<1$ のとき，連立方程式

$$(TV)(x)=V(x)\quad (x\in\mathcal{X})$$

は唯一の解 V をもつ．この方程式をベルマン方程式という．これは補題 4.1 と深くかかわりを持っている．

2.5 TD(0) 法

価値関数には以下のように無限級数が含まれている．

$$V(\pi,x)=\mathbb{E}\left[\sum_{t=0}^{\infty}\gamma^t R_{t+1}\ \middle|\ X_0=x\right].$$

したがって，コンピュータで価値関数の計算をするときは無限大の時刻まで X_n のシミュレーションをしなければならない．それは非現実的なので，価値関数の値は推定したい．そこで TD(0) 法についてそのアルゴリズムを述べる．そのまえに，一つ定義しなければならないことがある．

定義 2.11. 非負の数列 $\{\alpha_t\}_{t\geq 0}$ が RM 条件を満たすとは

1. $\sum_{t=0}^{\infty}\alpha_t=\infty$.
2. $\sum_{t=0}^{\infty}\alpha_t^2<\infty$.

を満たすことである．

さて，以下に TD(0) 法のアルゴリズムを述べる．

TD(0) 法のアルゴリズム

MDP と定常方策 π がある．このときの，価値関数 $V(\pi,x)$ の値を推定する．数列 $\{\alpha_t\}$ は RM 条件を満たすとする．以下で $t\geq 0$ とする．以下の手順に則って関数列 $V_t(\pi,\cdot)$ を考える．ただし，全ての t について $V_t(\pi,\cdot):\mathcal{X}\to\mathbb{R}$ である．$V_0(\pi,\cdot):\mathcal{X}\to\mathbb{R}$ は恒等的に 0 に等しいとする．

手順 1. ステップ幅と呼ばれる δ_{t+1} を次のように定める．

$$\delta_{t+1}=R_{t+1}+\gamma V_t(\pi,X_{t+1}^{\pi})-V_t(\pi,X_t^{\pi}).$$

手順 2. $t+1$ 番目の価値関数の推定 V_{t+1} を次のように定める．

$$V_{t+1}(\pi,x)=V_t(\pi,x)+\alpha_t\delta_{t+1}1_{\{X_t=x\}}\quad (\forall x\in\mathcal{X}).$$

状態 X_t をコンピュータで観測するたびに関数 V_t を更新していく．X_t での価値と X_{t+1} での価値を割引率で重みを付けたものの差と，報酬の和を関数 V_t の更新幅として用いている．この点が勾配降下法によく似ている．

このアルゴリズムが収束するかどうかは大きな問題である．状態空間が有限であり，(X_t) がエルゴード的なマルコフ連鎖の場合，この TD(0) のアルゴリズムは収束するといわれている．研究背景でも述べたとおり，この文書では TD(0) 法のこの勾配降下法のような手法に注目して，価値関数の最大化を測るアルゴリズムを提示し，その正当性を示している．

3　MDP/MRP の具体例の構成

この節ではマルコフ連鎖の構成を行う．もっと詳しく言うと，MDP/MRP の具体的な構成を行う．ここでは，簡単のために行動は状態によって確実に決まるとする．すなわち，方策 π は決定的な方策であるとする．別の言い方をすると，$\pi : \mathcal{X} \to \mathcal{A}$ は写像であるとする．また，簡単のために報酬関数 R のランダム性も省く．すなわち，状態の集合 $\mathcal{X}$ が与えられた時，報酬関数 R は $R : \mathcal{X} \to \mathbb{R}$ なる有界関数であるとする．

状態空間が離散集合の場合，適当な確率空間上にマルコフ連鎖を構成できることが知られている．(舟木 [5,pp.96, 命題 7.7]) そこで，状態空間が連続の場合にマルコフ連鎖を構成することを考える．この構成が示されれば MDP/MRP の数学的な実例を一つ提示できたことになる．

3.1　確率変数の定義

モデルを設定するために記号の定義をする．

定義 3.1. 以下のようにそれぞれの記号を定める．

- $\mathcal{X} = [0,1]$ を状態空間とする．
- $\mathcal{A}$ は行動の空でない高々可算集合とする．
- $(\mathcal{X}, \mathcal{B}[0,1])$ で可測空間を定める
- $(\mathcal{A}, 2^{\mathcal{A}})$ で可測空間を定める．
- $R : \mathcal{X} \to \mathbb{R}$ を有界かつ連続な関数とする．
- $\pi\colon \mathcal{X} \longrightarrow \mathcal{A}$ は写像である
- π は $\mathcal{X}/\mathcal{A}$-可測とする．

次に，この節で構成するマルコフ連鎖の遷移確率関数に相当する関数 f の定義を行う．

定義 3.2. $f^{\pi} : \mathcal{X} \times \mathcal{A} \times \mathcal{X} \longrightarrow [0,1]$ を次のような写像とする．

$$f^{\pi}(x,a,y) = \begin{cases} 0 & (\text{if} \quad a \neq \pi(x)) \\ f(x,a,y) & (\text{if} \quad a = \pi(x)) \end{cases}.$$

ただし，$x \in \mathcal{X}$ を任意に固定し，$a = \pi(x)$ のとき，次の式を満たすとする．

$$\int_0^1 f^{\pi}(y \mid x,a)dy = 1.$$

この f は後に構成するマルコフ連鎖の遷移確率に相当するものである．したがって，条件付き確率であるということを明示的にするために $f(x,a,y)$ を

$$f(y \mid x,a)$$

と書くことにする．

確率空間の定義を行う．

定義 3.3. 次のようにそれぞれの記号を定める．ただし，以下で i は自然数である．

- m_i は $\mathbb{R}$ 上のルベーグ測度である．
- 確率空間 $(\Omega_i, \mathcal{F}_i, \mu_i)$ を全て $([0,1], \mathcal{B}[0,1], m_i)$ とする．
- 直積確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ を

$$\Omega = \prod_{i=1}^{\infty} \Omega_i, \quad \mathcal{F} = \prod_{i=1}^{\infty} \mathcal{F}_i, \quad \mathbb{P} = \prod_{i=1}^{\infty} m_i$$

  と定義する．

定義 3.4. i 番目の確率空間上の確率変数 $U_i : \Omega_i \longrightarrow \mathbb{R}$ を次を満たすとする．

- 確率変数の族 $\{U_i\}_i$ は確率測度 $\mathbb{P}$ に関して独立．
- 任意の $\omega_i \in \Omega_i$ について $U_i(\omega_i) = \omega_i$ である．

補題 3.5. 任意の $i \in \mathbb{N}$ に対して

$$U_i \sim Unif[0,1]$$

である．ただし，$Unif[0,1]$ は $[0,1]$ 上の一様分布を表す．つまり，U_i の密度関数 f_{U_i} は次のように表される．

$$f_{U_i}(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

.

証明. $i \in \mathbb{N}$ を任意に選ぶ．$\Omega_i = [0,1]$ なので実数 a に対して次のようになる．

$$\mathbb{P}(U_i \leq a) = \begin{cases} 0 & (a \leq 0) \\ m_i[0,a] & (0 \leq a \leq 1) \\ 1 & (1 < a) \end{cases}.$$

一方，関数 $f : \mathbb{R} \longrightarrow \mathbb{R}$ を

$$f(x) = \begin{cases} 1 & (0 \leq x \leq 1) \\ 0 & (x < 0, 1 < x) \end{cases}$$

とすると，

$$\int_{-\infty}^{a} f(x)dx = \begin{cases} 0 & (a \leq 0) \\ m_i[0,a] & (0 \leq a \leq 1) \\ 1 & (1 < a) \end{cases}$$

である．したがって，任意の実数 a について

$$\mathbb{P}(U_i \leq a) = \int_{-\infty}^{a} f(x)dx$$

が成り立つ．よって U_i は $[0,1]$ 上の一様分布に従う． □

次に分布に相当する記号を定義する．

定義 3.6. 関数 $F : \mathcal{X} \times \mathcal{A} \times \mathbb{R} \longrightarrow [0,1]$ を次のように定める．

$$F(y \mid x, a) = \begin{cases} 0 & (\text{if} \quad a \neq \pi(x)) \\ 1 & (\text{if} \quad y > 1) \\ \int_0^y f(z \mid x, a) dz & (\text{if} \quad 0 \leq y \leq 1) \\ 0 & (\text{if} \quad y < 0) \end{cases}.$$

定義 3.6 の F を分布としてもつ確率変数の定義をする．

定義 3.7. $X_0, A_0 (= \pi(X_0))$ は与えられた定数とする．$i \geq 1$ に対して，$\rho_i \colon \prod_{n=1}^{\infty} \Omega_n \longrightarrow \Omega_i$ を $\omega \in \Omega$ の i 番目の成分への射影とする．すなわち，$\Omega \ni \omega = (\omega_1, \omega_2, \cdots)$ としたとき，$\rho_i(\omega) = \omega_i$ である．$(\Omega, \mathcal{F}, \mathbb{P})$ 上の確率変数族 $\{X_i\}_{i \in \mathbb{N}}$ を次で定義する．$\Omega \ni \omega = (\omega_1, \omega_2, \cdots)$ について

$$\begin{array}{ccccc} X_1(\omega) & = & \inf\{y \in [0,1] \,;\, U_1(\rho_1(\omega)) \leq F(y \mid X_0, A_0)\} & \equiv & F^{-1}(U_1 \mid X_0, A_0), \\ X_2(\omega) & = & \inf\{y \in [0,1] \,;\, U_2(\rho_2(\omega)) \leq F(y \mid X_1(\omega), A_1(\omega))\} & \equiv & F^{-1}(U_2 \mid X_1, A_1), \\ \vdots & & \vdots & & \vdots, \\ X_i(\omega) & = & \inf\{y \in [0,1] \,;\, U_i(\rho_i(\omega)) \leq F(y \mid X_{i-1}(\omega), A_{i-1}(\omega))\} & \equiv & F^{-1}(U_i \mid X_{i-1}, A_{i-1}), \\ \vdots & & \vdots & & \vdots \end{array}$$

と定める．なお，各 X_i について inf をとる集合が空集合の場合，0 を割り当てる．

補題 3.8. 定義 3.7 で定めた写像は well-defined である．

証明. $i \in \mathbb{N}$, $\omega \in \Omega$ を固定する．定義 3.7 より，閉区間 $[0,1]$ の部分集合の inf を取っているので，$X_i(\omega)$ の値は一意に定まる．再び $X_i(\omega)$ の定義より，$X_i(\omega)$ の確率変数がとる値は $0 \leq X_i(\omega) \leq 1$ である． □

X_i 達は全て $\Omega = \prod_i \Omega_i$ 上で定義された関数であるが，n を一つ固定したとき，任意の $\omega \in \Omega$ に対して $X_n(\omega)$ の値は ω の最初の n 個の成分だけに依存している．

補題 3.9. 任意の $n (\geq 1)$ について，$s, t \in \Omega$ が

$$\rho_i(s) = \rho_i(t) \quad (1 \leq i \leq n)$$

をみたしているならば

$$X_n(s) = X_n(t)$$

が成り立つ．

証明. 帰納法で証明する．$n = 1$ とする．$s, t \in \Omega$ は $\rho_1(s) = \rho_1(t)$ を満たしているとすると，

$$\begin{aligned} X_1(s) &= \inf\{y \in [0,1] \,;\, U_1(\rho_1(s) \leq F(y \mid X_0, A_0)\} \\ &= \inf\{y \in [0,1] \,;\, U_1(\rho_1(t)) \leq F(y \mid X_0, A_0)\} \\ &= X_1(t). \end{aligned}$$

$k \geq 1$ とする．$n = k$ のときに $X_k(s) = X_k(t)$ が成立とする．$s, t \in \Omega$ は $\rho_i(s) = \rho_i(t) \quad (1 \leq i \leq k+1)$ を満たしているとすると，

$$X_{k+1}(s) = \inf\{y \in [0,1]\ ;\ U_{k+1}(\rho_{k+1}(s)) \leq F(y \mid X_k(s), A_k(s))\}$$

帰納法の仮定より

$$\begin{aligned}
&= \inf\{y \in [0,1]\ ;\ U_{k+1}(\rho_{k+1}(s)) \leq F(y \mid X_k(t), A_k(t)\} \\
&= \inf\{y \in [0,1]\ ;\ U_{k+1}(\rho_{k+1}(t)) \leq F(y \mid X_k(t), A_k(t)\} \\
&= X_{k+1}(t).
\end{aligned}$$

□

補題 3.9 より，$X_n(\omega)$ の値を知るためには ω のはじめの n 個の成分だけを見ればよいということが分かる．そこで，定義域をはじめの n 個の成分にだけ狭めた確率変数を定義しておく．

定義 3.10. $n \in \mathbb{N}$ とする．X_n の定義域を $\Omega_1 \times \cdots \times \Omega_n$ とみなした確率変数を Y_n とする．また，$B_n = \pi(Y_n)$ と定める．すなわち，$Y_0 := X_0, B_0 := A_0$ と定義し，$n \geq 1$ については，以下のように定義する．$\omega \in \Omega$ に対して，

$$\begin{array}{ccc}
Y_1(\rho_1(\omega)) & = & \inf\{y \in [0,1]\ ;\ U_1(\rho_1(\omega)) \leq F(y \mid X_0, A_0)\}, \\
Y_2(\rho_1(\omega), \rho_2(\omega)) & = & \inf\{y \in [0,1]\ ;\ U_2(\rho_2(\omega)) \leq F(y \mid Y_1(\rho_1(\omega)), B_1(\rho_1(\omega)))\}, \\
\vdots & & \vdots, \\
Y_{i+1}(\tilde{\omega}) & = & \inf\Big\{y \in [0,1]\ ;\ U_{i+1}(\rho_{i+1}(\omega)) \leq F\Big(y \mid Y_i(\hat{\omega}), B_i(\hat{\omega})\Big)\Big\}, \\
\vdots & & \vdots
\end{array}$$

ただし，$\tilde{\omega} = (\rho_1(\omega), \cdots, \rho_{i+1}(\omega)), \hat{\omega} = (\rho_1(\omega), \cdots, \rho_i(\omega))$ である．

系 3.11. 定義 3.10 で定めた確率変数は well-defined である．さらに，$\omega \in \Omega$ の最初の n 個の成分が $\omega_1, \cdots, \omega_n$ とし，$\tilde{\omega} = (\omega_1, \cdots, \omega_n)$ とすれば，次が成立する．

$$X_n(\omega) = Y_n(\tilde{\omega}).$$

3.2 (X_n) がマルコフ連鎖であること

3.2.1 X_n の分布関数

定義 3.6 の F を分布として持つように定義 3.7 で確率変数 X_n たちを定めていた．そこで，実際に分布が F であることを確かめる．まずは，X_1 の分布を計算する．

補題 3.12. $b \in [0,1]$ とする．このとき，

$$\{X_1 \leq b\} = \left\{\omega \in \Omega \;\middle|\; \rho_1(\omega) \leq \int_0^b f(z|X_0, A_0)dz\right\}$$

である．

証明. $\omega \in \left\{\omega \in \Omega \,\middle|\, \rho_1(\omega) \leq \int_0^b f(z|X_0, A_0)dz\right\}$ を任意にとる．すると，

$$\rho_1(\omega_1) \leq \int_0^b f(z|X_0, A_0)dz$$

なので，

$$b \in \left\{y \in [0,1] \,\middle|\, \omega_1 \leq \int_0^y f(z|X_0, A_0)dz\right\}$$

であることが分かる．したがって，

$$\inf \left\{y \in [0,1] \,\middle|\, \omega_1 \leq \int_0^y f(z|X_0, A_0)dz\right\} \leq b$$

である．X_1 の定義より

$$X_1(\omega) \leq b$$

となる．したがって，$\omega \in \{X_1 \leq b\}$ である．

$\omega \in \{X_1 \leq b\}$ を取る．$X_1(\omega) \leq b$ が成立しているが，$X_1(\omega) < b$ および，$X_1(\omega_1) = b$ の二つの場合が考えられる．$X_1(\omega) < b$ の場合は明らかに

$$\rho_1(\omega) \leq \int_0^b f(z|X_0, A_0)dz$$

が成立している．$b = X_1(\omega)$ の場合を考えよう．inf の定義より任意の $\varepsilon > 0$ に対してある $y \in \left\{y \in [0,1] \,\middle|\, \omega_1 \leq \int_0^y f(z|X_0, A_0)dz\right\}$ が存在し，

$$\rho_1(\omega) \leq \int_0^y f(z|X_0, A_0)dz \leq \int_0^{b+\varepsilon} f(z|X_0, A_0)dz$$

となる．よって，$\varepsilon \downarrow 0$ とすると，

$$\rho_1(\omega) \leq \int_0^b f(z|X_0, A_0)dz$$

となる．したがって，$\omega \in \left\{\omega \in \Omega \,\middle|\, \rho_1(\omega) \leq \int_0^b f(z|X_0, A_0)dz\right\}$. □

命題 3.13. $b \in \mathbb{R}$ とする．$\mathbb{P}(X_1 \leq b) = F(b \mid X_0, A_0)$ が成立する．すなわち，X_1 の分布 F_{X_1} は $F(\cdot \mid X_0, A_0)$ である．

証明. $b < 0$ のとき，$0 \leq X_1 \leq 1$ より，$\mathbb{P}(X_1 \leq b) = \mathbb{P}(\emptyset) = 0$ である．また，F の定義より，$F(b \mid X_0, A_0) = 0$ である．ゆえに，$\mathbb{P}(X_1 \leq b) = F(b \mid X_0, A_0)$.

$b > 1$ のとき，同様に，$\mathbb{P}(X_1 \leq b) = 1$ であり，$F(b \mid X_0, A_0) = 1$ である．

$b \in [0,1]$ の場合について考える．補題 3.12 より，

$$\begin{aligned}
\mathbb{P}(X_1 \leq b) &= \mathbb{P}\left(\left\{\omega \in \Omega \;\middle|\; \rho_1(\omega) \leq \int_0^b f(z|X_0, A_0)dz\right\}\right) \\
&= m_1\left(\left[0,\ \int_0^b f(z|X_0, A_0)dz\right]\right) \\
&= \int_0^b f(z|X_0, A_0)dz
\end{aligned}$$

である．一方，$F(\cdot \mid X_0, A_0)$ の定義より，

$$F(b \mid X_0, A_0) = \int_0^b f(z|X_0, A_0)dz$$

であるから，

$$\mathbb{P}(X_1 \leq b) = F(b \mid X_0, A_0)$$

である． □

次に，X_2 の分布を考える．X_1 の時とは異なり，「条件付きの部分」にランダム性が現れる．X_2 の定義から，X_1 について知っていなければならない．X_1, A_1 の条件つきの，X_2 の分布を考えねばならない．一方，X_1 の分布を考える際に X_0 も書いていたが，これは予め与えられるものなのでランダム性が無かった．

補題 3.14.

$$\{X_2 \leq b\} = \left\{\omega \in \Omega \;\middle|\; \rho_2(\omega) \leq \int_0^b f(z|X_1(\omega), A_1(\omega))dz\right\}.$$

証明. $\omega \in \left\{\omega \in \Omega \;\middle|\; \rho_2(\omega) \leq \int_0^b f(z|X_1(\omega), A_1(\omega))dz\right\}$ を取る．すると，

$$\rho_2(\omega) \leq \int_0^b f(z|X_1(\omega), A_1(\omega))dz$$

なので，

$$b \in \left\{y \in [0,1] \;\middle|\; \rho_2(\omega) \leq \int_0^y f(z|X_1(\omega), A_1(\omega))dz\right\}$$

であることが分かる．したがって，

$$\inf\left\{y \in [0,1] \;\middle|\; \rho_2(\omega) \leq \int_0^y f(z|X_1(\omega), A_1(\omega))dz\right\} \leq b$$

である．X_2 の定義より

$$X_2(\omega) \leq b$$

となる．したがって，$\omega \in \{X_2 \leq b\}$ である．

逆に，$\omega \in \{X_2 \leq b\}$ を取る．$X_2(\omega) \leq b$ が成立しているが，$X_2(\omega) < b$ および，$X_2(\omega) = b$ の二つの場合が考えられる．$X_2(\omega) < b$ の場合は明らかに

$$\omega \leq \int_0^b f(z|X_1(\omega), A_1(\omega))dz$$

が成立している．$b = X_2(\omega)$ の場合を考えよう．inf の定義より任意の $\varepsilon > 0$ に対してある $y \in \left\{y \in [0,1] \;\middle|\; \rho_2(\omega) \leq \int_0^y f(z|X_1(\omega), A_1(\omega))dz\right\}$ が存在し，

$$\omega \leq \int_0^y f(z|X_1(\omega), A_1(\omega))dz \leq \int_0^{b+\varepsilon} f(z|X_1(\omega), A_1(\omega))dz$$

となる．よって，$\varepsilon \downarrow 0$ とすると，

$$\omega \leq \int_0^b f(z|X_1(\omega), A_1(\omega))dz$$

となる．したがって，$\omega \in \left\{\omega \in \Omega \;\middle|\; \rho_2(\omega) \leq \int_0^b f(z|X_1(\omega), A_1(\omega))dz\right\}$. □

命題 3.15. X_2 の条件付き分布関数 $F_{X_2}(\cdot \mid X_1, A_1)$ は $F(\cdot \mid X_1, A_1)$ である．すなわち，任意の $b \in \mathbb{R}$ について，

$$\mathbb{P}(X_2 \leq b \mid X_1, A_1) = F(b \mid X_1, A_1)$$

が成立する．

証明. $\mathbb{P}(X_2 \leq b \mid X_1, A_1)$ とは $\mathbb{E}(I_{\{X_2 \leq b\}} \mid X_1, A_1)$ のことなので，任意の $G \in \sigma(X_1, A_1)$ について次が成立することを示せばよい.

$$\int_G I_{\{X_2 \leq b\}} d\mathbb{P} = \int_G \int_0^b f(z \mid X_1, A_1) dz d\mathbb{P}.$$

表記の簡単さのために

$$F = \left\{\omega \in \Omega \;\middle|\; \rho_2(\omega) \leq \int_0^b f(z|X_1(\omega), A_1(\omega))dz\right\}$$

とおく．

補題 3.14 より

$$
\begin{aligned}
&\int_G I_{\{X_2 \leq b\}}(\omega) d\mathbb{P} \\
&= \int_G I_F(\omega) d\mathbb{P} \\
&= \int I_{G\cap F}(\omega) d\mathbb{P} \\
&= \int I_{G\cap F}(\omega)\ \mathbb{P}_1 \otimes \mathbb{P}_2(d(\omega_1, \omega_2)) \\
&= \int I_{G\cap F}(\omega)\ \mathbb{P}_2(d\omega_2)\mathbb{P}_1(d\omega_1) \\
&= \int_G \int_0^b f(z \mid X_1(\omega), A_1(\omega)) dz d\mathbb{P}_1.
\end{aligned}
$$

□

同様にすると X_n の条件付き分布も計算できる．

補題 3.16.

$$\{X_n \leq b\} = \left\{\omega \in \Omega \;\middle|\; \rho_n(\omega) \leq \int_0^b f(z|X_{n-1}(\omega), A_{n-1}(\omega))dz\right\}.$$

証明. $\omega \in \left\{\omega \in \Omega \;\middle|\; \rho_n(\omega) \leq \int_0^b f(z|X_{n-1}(\omega), A_{n-1}(\omega))dz\right\}$ を取る．すると，

$$\rho_n(\omega) \leq \int_0^b f(z|X_{n-1}(\omega), A_{n-1}(\omega))dz$$

なので，

$$b \in \left\{y \in [0,1] \;\middle|\; \rho_n(\omega) \leq \int_0^y f(z|X_{n-1}(\omega), A_{n-1}(\omega))dz\right\}$$

であることが分かる．したがって，

$$\inf\left\{y \in [0,1] \;\middle|\; \rho_n(\omega) \leq \int_0^y f(z|X_{n-1}(\omega), A_{n-1}(\omega))dz\right\} \leq b$$

である．X_n の定義より

$$X_n(\omega) \leq b$$

となる．したがって，$\omega \in \{X_n \leq b\}$ である．

逆に，$\omega \in \{X_n \leq b\}$ を取る．$X_n(\omega) \leq b$ が成立しているが，$X_n(\omega) < b$ および，$X_n(\omega) = b$ の二つの場合が考えられる．$X_n(\omega) < b$ の場合は明らかに

$$\omega_n \leq \int_0^b f(z|X_{n-1}(\omega), A_{n-1}(\omega))dz$$

が成立している．$b = X_n(\omega)$ の場合を考えよう．inf の定義より任意の $\varepsilon > 0$ に対してある $y \in A$ が存在し，

$$\omega_n \leq \int_0^y f(z|X_{n-1}(\omega), A_{n-1}(\omega))dz \leq \int_0^{b+\varepsilon} f(z|X_{n-1}(\omega), A_{n-1}(\omega))dz$$

となる．よって，$\varepsilon \downarrow 0$ とすると，

$$\omega_n \leq \int_0^b f(z|X_{n-1}(\omega), A_{n-1}(\omega))dz$$

となる．したがって，$\omega_n \in \left\{\omega \in \Omega \;\middle|\; \rho_n(\omega) \leq \int_0^b f(z|X_{n-1}(\omega), A_{n-1}(\omega))dz\right\}$. □

命題 3.17. $n \geq 1$ とする．X_n の条件付き分布関数 $F_{X_n}(\cdot \mid X_{n-1}, A_{n-1})$ は $F(\cdot \mid X_{n-1}, A_{n-1})$ である．すなわち，任意の $b \in \mathbb{R}$ について，

$$\mathbb{P}(X_n \leq b \mid X_{n-1}, A_{n-1}) = F(b \mid X_{n-1}, A_{n-1})$$

が成立する．

証明. $\mathbb{P}(X_n \leq b \mid X_{n-1}, A_{n-1})$ とは $\mathbb{E}(I_{\{X_n \leq b\}} \mid X_{n-1}, A_{n-1})$ のことなので，任意の $G \in \sigma(X_1, A_1, \cdots, X_{n-1}, A_{n-1})$ について次が成立することを示せばよい．

$$\int_G I_{\{X_n \leq b\}}(\omega)d\mathbb{P} = \int_G \int_0^b f(z \mid X_{n-1}(\omega), A_{n-1}(\omega))dzd\mathbb{P}.$$

表記の簡単さのために

$$F = \left\{\omega \in \Omega \;\middle|\; \rho_n(\omega) \leq \int_0^b f(z|X_{n-1}(\omega), A_{n-1}(\omega))dz\right\}$$

とおく．

補題 3.16 より，次のような計算ができる．

$$\begin{aligned}
&\int_G I_{\{X_n \leq b\}}(\omega)d\mathbb{P} \\
=&\int_G I_F(\omega)d\mathbb{P} \\
=&\int I_{G\cap F}(\omega)d\mathbb{P} \\
=&\int I_{G\cap F}(\omega)d\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_n \\
=&\int I_{G\cap F}(\omega)\ \mathbb{P}_n(d\omega_n)\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_{n-1}(d(\omega_1, \cdots, \omega_{n-1})) \\
=&\int_G \int_0^b f(z \mid X_{n-1}(\omega), A_{n-1}(\omega))dzd\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_{n-1}.
\end{aligned}$$

□

3.2.2 f が遷移確率になっていること

(X_n) がマルコフ連鎖になっていることを示そう．

命題 3.18. $i \in \mathbb{N}$ を任意にひとつ固定する．全ての $A \in \mathcal{B}[0,1]$ に対して

$$\mathbb{P}(X_i \in A \mid X_{i-1}, A_{i-1}) = \int_A f(y \mid X_{i-1}, A_{i-1}) dy$$

が成立している．

証明. $i = 1$ のとき．すべての $G \in \mathcal{B}[0,1]$ について

$$\mathbb{P}(X_1 \in G \mid X_0, A_0) = \int_G f(y \mid X_0, A_0) dy$$

を示す．つまり，任意の $F \in \sigma(X_0, A_0)$ について

$$\int_F I_{\{X_1 \in G\}}(\omega) \mathbb{P}(d\omega) = \int_F \int_G f(y \mid X_0, A_0) dy d\mathbb{P}$$

を示せばよい．だがここで，X_0, A_0 は与えられた定数関数なので，$\sigma(X_0, A_0)$ は最も簡単な σ-alg.，すなわち，$\sigma(X_0, A_0) = \{\emptyset, [0,1]\}$ である．$F = \emptyset$ の場合は両辺とも 0 になるので正しい．$F = [0,1]$ の場合は次にように計算できる．

$$\begin{aligned}
&\int_F I_{\{X_1 \in G\}}(\omega) \mathbb{P}(d\omega) \\
&= \mathbb{P}(\{X_1 \in G\}) \\
&= \mathbb{P}_1(\{X_1 \in G\}).
\end{aligned}$$

補題 3.12 より，任意の $b \in [0,1]$ に対して

$$\mathbb{P}_1(\{X_1 \leq b\}) = \int_0^b f(y \mid X_0, A_0) dy$$

が成立する．$\mathcal{I} := \{[0,a] \mid a \in [0,1]\}$ は π 系であり，$\sigma(\mathcal{I}) = \mathcal{B}[0,1]$ だから，

$$\mathbb{P}_1(\{X_1 \in G\}) = \int_G f(y \mid X_0, A_0) dy$$

が成り立つ．

$i = 2$ のときすべての $G \in \mathcal{B}[0,1]$ について

$$\mathbb{P}(X_2 \in G \mid X_1, A_1) = \int_G f(y \mid X_1, A_1) dy$$

を示す．したがって，任意の $F \in \sigma(X_1, A_1)$ について

$$\int_F I_{\{X_2 \in G\}}(\omega) \mathbb{P}(d\omega) = \int_F \int_G f(y \mid X_1(\omega), A_1(\omega)) dy \mathbb{P}(d\omega)$$

を示せばよい．$F=[0,a], G=[0,b]$ $(a,b\in[0,1])$ として，

$$\begin{aligned}\int_F I_{\{X_2\in G\}}(\omega)\mathbb{P}(d\omega) &= \int_0^a\int_0^b f(y\mid X_1(\omega),A_1(\omega))dy\mathbb{P}(d\omega)\\ &= \int_0^a F(b\mid X_1(\omega),A_1(\omega))\mathbb{P}(d\omega)\\ &= \int_0^a \mathbb{P}(X_2\le b\mid X_1(\omega),A_1(\omega))\mathbb{P}(d\omega)\\ &= \int_0^a I_{\{X_2\le b\}}(\omega)\mathbb{P}(d\omega).\end{aligned}$$

なお，2 行目から 3 行目の変形は命題 3.15 を用いた．また，3 行目から 4 行目の変形は $[0,a]\in\sigma(X_1,A_1)$ であることより条件付き平均の定義から従う．$\sigma(\mathcal{I}\times\mathcal{I})=\mathcal{B}[0,1]\times\mathcal{B}[0,1]$ であるから[*1]，$\mathcal{B}[0,1]\times\mathcal{B}[0,1]$ 上で

$$\int_F I_{\{X_2\in G\}}(\omega)\mathbb{P}(d\omega)=\int_F\int_G f(y\mid X_1(\omega),A_1(\omega))dy\mathbb{P}(d\omega)$$

が言えた．$\sigma(X_1,A_1)\subset\mathcal{B}[0,1]$ なので，任意の $F\in\sigma(X_1,A_1)$ について

$$\int_F I_{\{X_2\in G\}}(\omega)\mathbb{P}(d\omega)=\int_F\int_G f(y\mid X_1(\omega),A_1(\omega))dy\mathbb{P}(d\omega)$$

もいえたことになる． □

3.2.3 X_2 のマルコフ性

ここから，X_n がマルコフ性を持つことを示す．まずは X_2 についてマルコフ性を見る．

命題 3.19. 全ての $C\in\mathcal{B}[0,1]$ について

$$\mathbb{P}(X_2\in C\mid X_0,X_1)=\mathbb{P}(X_2\in C\mid X_1)$$

が成立する．

証明. X_0 は与えられた定数関数なので $\sigma(X_0,X_1)=\sigma(X_1)$ だから，

$$\mathbb{P}(X_2\in C\mid\sigma(X_0,X_1))=\mathbb{P}(X_2\in C\mid\sigma(X_1)).$$

□

[*1] 証明してない

3.2.4 X_3 のマルコフ性

次に，X_3 についてマルコフ性を持つことをしめすが，その前にいくつかの補題を証明する．

定義 3.20. $a, b \in [0,1],\ C \in \mathcal{B}[0,1]$ とする．集合 $G_{1a}, G_{2b}, G_{3C}, H_{1a}, H_{2b}, H_{3C}, K_{1a}, K_{2b}$ をそれぞれ次のように定義する．

$$
\begin{aligned}
G_{1a} &:= \left\{ t \in \Omega \;\middle|\; U_1(\rho_1(t)) \leq \int_0^a f(y \mid X_0, A_0) dy \right\}, \\
G_{2b} &:= \left\{ t \in \Omega \;\middle|\; U_2(\rho_2(t)) \leq \int_0^b f(y \mid X_1(t), A_1(t)) dy \right\}, \\
G_{3C} &:= \{ X_3 \in C \}, \\
H_{1a} &:= \left\{ (\omega_1, \omega_2, \omega_3) \in \Omega_1 \times \Omega_2 \times \Omega_3 \;\middle|\; U_1(\omega_1) \leq \int_0^a f(y \mid X_0, A_0) dy \right\}, \\
H_{2b} &:= \left\{ (\omega_1, \omega_2, \omega_3) \in \Omega_1 \times \Omega_2 \times \Omega_3 \;\middle|\; U_2(\omega_2) \leq \int_0^b f(y \mid Y_1(\omega_1), B_1(\omega_1)) dy \right\}, \\
H_{3C} &:= \{ Y_3 \in C \}, \\
K_{1a} &:= \left\{ (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 \;\middle|\; U_1(\omega_1) \leq \int_0^a f(y \mid X_0, A_0) dy \right\}, \\
K_{2b} &:= \left\{ (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 \;\middle|\; U_2(\omega_2) \leq \int_0^b f(y \mid Y_1(\omega_1), B_1(\omega_1)) dy \right\}.
\end{aligned}
$$

補題 3.21.

$$G_{1a} \cap G_{2b} \cap G_{3C} = (H_{1a} \cap H_{2b} \cap H_{3C}) \times \prod_{i=4}^{\infty} \Omega_i.$$

証明. $t \in G_{1a} \cap G_{2b} \cap G_{3C}$ を任意に固定する．$(\rho_1(t), \rho_2(t), \rho_3(t))$ について調べる．$t \in G_{1a}$ より，

$$U_1(\rho_1(t)) \leq \int_0^a f(y \mid X_0, A_0) dy$$

を満たす．したがって，$(\rho_1(t), \rho_2(t), \rho_3(t)) \in H_{1a}$ である．また，$t \in G_{2b}$ であることから

$$U_2(\rho_2(t)) \leq \int_0^b f(y \mid X_1(t), A_1(t)) dy$$

が成立する．系 3.10 より，$X_1(t) = Y_1(\rho_1(t)), A_1(t) = B_1(\rho_1(t))$ なので，

$$U_2(\rho_2(t)) \leq \int_0^b f(y \mid Y_1(\rho_1(t)), B_1(\rho_1(t))) dy$$

が成立する．したがって，$(\rho_1(t), \rho_2(t), \rho_3(t)) \in H_{2b}$ である．再び，系 3.10 より，$(\rho_1(t), \rho_2(t), \rho_3(t)) \in H_{3C}$ がわかる．したがって，

$$(\rho_1(t), \rho_2(t), \rho_3(t)) \in H_{1a} \cap H_{2b} \cap H_{3C}$$

である．よって，

$$t \in (H_{1a} \cap H_{2b} \cap H_{3C}) \times \prod_{i=4}^{\infty} \Omega_i$$

逆に，$t \in (H_{1a} \cap H_{2b} \cap H_{3C}) \times \prod_{i=4}^{\infty} \Omega_i$ を任意に取る．$(\rho_1(t), \rho_2(t), \rho_3(t)) \in H_{1a}$ より，

$$U_1(\rho_1(t)) \le \int_0^a f(y \mid X_0, A_0) dy$$

が成り立つので，$t \in G_{1a}$ である．$(\rho_1(t), \rho_2(t), \rho_3(t)) \in H_{2b}$ より，

$$U_2(\rho_2(t)) \le \int_0^b f(y \mid Y_1(\rho_1(t)), B_1(\rho_1(t))) dy$$

であり，系 3.10 より，$X_1(t) = Y_1(\rho_1(t)), A_1(t) = B_1(\rho_1(t))$ なので，

$$U_2(\rho_2(t)) \le \int_0^b f(y \mid X_1(t), A_1(t)) dy$$

が分かる．したがって，$t \in G_{2b}$ である．系 3.10 より，$t \in G_3$ がわかる．したがって，

$$t \in G_{1a} \cap G_{2b} \cap G_{3C}$$

である． □

補題 3.22. $\omega_1 \in \Omega_1, \omega_2 \in \Omega_2$ を固定する．全ての $\omega_3 \in \Omega_3$ について

$$I_{K_{1a} \cap K_{2b}}(\omega_1, \omega_2) = I_{H_{1a} \cap H_{2b}}(\omega_1, \omega_2, \omega_3).$$

証明. $I_{K_{1a} \cap K_{2b}}(\omega_1, \omega_2) = 1$ のとき, K_{1a}, K_{2b} の定義から, 全ての $\omega_3 \in \Omega_3$ に対して $(\omega_1, \omega_2, \omega_3) \in H_{1a} \cap H_{2b}$ である．したがって，$I_{H_{1a} \cap H_{2b}}(\omega_1, \omega_2, \omega_3) = 1$ である．

$I_{K_{1a} \cap K_{2b}}(\omega_1, \omega_2) = 0$ のときは，(ω_1, ω_2) が K_{1a}, K_{2b} の少なくとも一方に属していないことになるので，$(\omega_1, \omega_2, \omega_3)$ も H_{1a}, H_{2b} の少なくとも一方には属していない．したがって，$I_{H_{1a} \cap H_{2b}}(\omega_1, \omega_2, \omega_3) = 0$ □

補題 3.23. $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$ が固定されている．$C \in \mathcal{B}[0,1]$ とする．このとき，

$$\int_{r \in \Omega_3} I_{H_{3C}}(\omega_1, \omega_2, r) \mathbb{P}_3(dr) = \int_C f(y \mid Y_2(\omega_1, \omega_2), B_2(\omega_1, \omega_2)) dy.$$

証明. 系 3.11 と命題 3.18 とフビニの定理より従う．

□

次とその次の二つの補題は，補題 3.26 の証明のために必要な補題である．

補題 3.24. X, Y を空でない集合，$f\colon X \to Y$ を写像とする．$\mathcal{G}$ は Y 上の σ-alg. であるとする．このとき，$f^{-1}(\mathcal{G}) := \{f^{-1}(G) \mid G \in \mathcal{G}\}$ は X 上の σ-alg. である．

証明． $X = f^{-1}(Y)$ であり，$Y \in \mathcal{G}$ より，$X \in f^{-1}(\mathcal{G})$ である．

$F \in f^{-1}(\mathcal{G})$ とする．定義より，ある $G \in \mathcal{G}$ が存在し，

$$F = f^{-1}(G)$$

と書ける．このとき，

$$F^c = (f^{-1}(G))^c = f^{-1}(G^c)$$

であり，$G^c \in \mathcal{G}$ なので，$F^c \in f^{-1}(\mathcal{G})$ である．

$(F_i)_{i\in\mathbb{N}} \subset f^{-1}(\mathcal{G})$ を任意に選ぶ．定義より，各 i について，ある $G_i \in \mathcal{G}_i$ が存在し，

$$F_i = f^{-1}(G_i)$$

と書ける．このとき，

$$\bigcup_i F_i = \bigcup_i f^{-1}(G_i) = f^{-1}\left(\bigcup_i G_i\right)$$

であり，$\bigcup_i G_i \in f^{-1}(\mathcal{G})$ なので，$\bigcup_i F_i \in f^{-1}(\mathcal{G})$ である． □

補題 3.25. X, Y をそれぞれ空でない集合とする．$\mathcal{C}$ を Y の部分集合の族とする．$f\colon X \longrightarrow Y$ という写像が与えられている．$\sigma(\mathcal{C})$ は $\mathcal{C}$ により生成される σ-alg.，つまり，$\mathcal{C}$ を含む Y 上の最小の σ-alg. を表すとする．このとき，

$$f^{-1}(\sigma(\mathcal{C})) = \sigma(f^{-1}(\mathcal{C}))$$

が成立する．

証明． 補題 3.24 より，$f^{-1}(\sigma(\mathcal{C}))$ は σ-alg. である．定義より，$f^{-1}(\sigma(\mathcal{C})) \supset f^{-1}(\mathcal{C})$ である．生成された σ-alg. の最小性より $f^{-1}(\sigma(\mathcal{C})) \supset \sigma(f^{-1}(\mathcal{C}))$ である．

Y の部分集合族 $\mathcal{D}$ を次のように定義する；

$$\mathcal{D} := \{A \subset Y \mid f^{-1}(A) \in \sigma(f^{-1}(\mathcal{C}))\}$$

とする．ここで，$\mathcal{D}$ が $\mathcal{C}$ を含む σ-alg. であることを示す．$C \in \mathcal{C}$ とする．当然，$C \in \sigma(\mathcal{C})$ であり，$f^{-1}(C) \in f^{-1}(\mathcal{C})$ より，

$$f^{-1}(C) \in \sigma(f^{-1}(\mathcal{C}))$$

である．したがって，

$$C \in \mathcal{D}$$

である．よって，$\mathcal{C} \subset \mathcal{D}$ が成り立つ．

次に，$\mathcal{D}$ が σ-alg. であることを示そう．まず，$Y \in \sigma(\mathcal{C})$ である．$f^{-1}(Y) = X$ であり，$\sigma(f^{-1}(\mathcal{C}))$ は X 上の σ-alg. なので，

$$X \in \sigma(f^{-1}(\mathcal{C}))$$

である．したがって，$f^{-1}(Y) \in \sigma(f^{-1}(\mathcal{C}))$ がなりたつ．よって，

$$Y \in \mathcal{D}$$

である．次に，$A \in \mathcal{D}$ とする．$\mathcal{D}$ の定義より，

$$f^{-1}(A) \in \sigma(f^{-1}(\mathcal{C}))$$

である．よって，$(f^{-1}(A))^c = X - f^{-1}(A) = f^{-1}(Y - A) = f^{-1}(A^c)$ も，$\sigma(f^{-1}(\mathcal{C}))$ に属する．したがって，

$$A^c \in \mathcal{D}$$

である．最後に，$(A_i)_{i\in\mathbb{N}} \subset \mathcal{D}$ とする．各 i について，$\mathcal{D}$ の定義から，

$$f^{-1}(A_i) \in \sigma(f^{-1}(\mathcal{C}))$$

である．したがって，$\bigcup_i f^{-1}(A_i)$ も $\sigma(f^{-1}(\mathcal{C}))$ に属する．$\bigcup_i f^{-1}(A_i) = f^{-1}(\bigcup_i A_i)$ であるから，

$$\bigcup_i A_i \in \mathcal{D}$$

である．以上により，$\mathcal{D}$ は σ-alg. であることが証明できた．したがって，

$$\sigma(\mathcal{C}) \subset \mathcal{D}$$

である．よって，

$$f^{-1}(\sigma(\mathcal{C})) \subset f^{-1}(\mathcal{D})$$

である．いっぽう，$\mathcal{D}$ の定義より

$$f^{-1}(\mathcal{D}) \subset \sigma(f^{-1}(\mathcal{C}))$$

ゆえに，

$$f^{-1}(\sigma(\mathcal{C})) \subset \sigma(f^{-1}(\mathcal{C})).$$

□

補題 3.26. $Z\colon \Omega \to \mathbb{R}^2$ で，$Z(\omega) = (X_1(\omega), X_2(\omega))$，$\mathcal{G} := \{A \times B \mid A, B \in \mathcal{B}[0,1]\}$ とする．このとき，

$$\sigma(X_0, X_1, X_2) = \sigma(Z^{-1}(\mathcal{G})) = Z^{-1}(\sigma(\mathcal{G}))$$

が成立する．

証明. 補題 3.25 より，$\sigma(Z^{-1}(\mathcal{G})) = Z^{-1}(\sigma(\mathcal{G}))$ である．

$\sigma(Z^{-1}(\mathcal{G}))$ 上で X_1, X_2 が可測であることを示そう．$B \in \mathcal{B}[0,1]$ を任意に一つ固定する．

$$X_1^{-1}(B) = Z^{-1}(B \times [0,1]),$$
$$X_2^{-1}(B) = Z^{-1}([0,1] \times B),$$

より，

$$X_1^{-1}(B), X_2^{-1}(B) \in \sigma(Z^{-1}(\mathcal{G}))$$

である．したがってこれらの確率変数は $\sigma(Z^{-1}(\mathcal{G}))$ 上で可測．

逆に，$\sigma(X_0, X_1, X_2) \supset Z^{-1}(\mathcal{G})$ であることを示そう．$F \in Z^{-1}(\mathcal{G})$ を取る．すると，定義よりある $A, B \in \mathcal{B}[0,1]$ が存在して $F = Z^{-1}(A \times B)$ と書ける．よって，

$$Z^{-1}(A \times B) = X_1^{-1}(A) \cap X_2^{-1}(B)$$

である．$\sigma(X_0, X_1, X_2)$ 上では X_1, X_2 は可測なので，

$$X_1^{-1}(A), X_2^{-1}(B) \in \sigma(X_0, X_1, X_2)$$

である．よって，

$$F = X_1^{-1}(A) \cap X_2^{-1}(B) \in \sigma(X_0, X_1, X_2).$$

したがって，

$$Z^{-1}(\mathcal{G}) \subset \sigma(X_0, X_1, X_2)$$

である．

以上により

$$\sigma(X_0, X_1, X_2) = \sigma(Z^{-1}(\mathcal{G}))$$

が示せた． □

補題 3.27.

$$\begin{aligned}
\pi([0,1]) &:= \{[0,x] \mid 0 \le x \le 1\}, \\
(\pi([0,1]))^2 &:= \{[0,a] \times [0,b] \mid 0 \le a, b \le 1\}, \\
\mathcal{I} &:= \{X_1^{-1}([0,a]) \cap X_2^{-1}([0,b]) \mid a, b \in [0,1]\},
\end{aligned}$$

とする．このとき

$$\sigma(X_0, X_1, X_2) = \sigma(\mathcal{I}).$$

が成立する．

証明. $\mathcal{B}([0,1]) \otimes \mathcal{B}([0,1])$ を $\bigotimes_{i=1}^2 \mathcal{B}([0,1])$ と書くことにする．

補題 3.26 で定義した Z に対して，$\mathcal{I} = Z^{-1}\big((\pi([0,1]))^2\big)$ である．したがって，$\sigma(\mathcal{I}) = \sigma(Z^{-1}\big((\pi([0,1]))^2\big))$ が成立する．補題 3.25 より，$\sigma(\mathcal{I}) = Z^{-1}(\sigma\big((\pi([0,1]))^2\big))$ である．補題 3.26 で定義した $\mathcal{G}$ に対して，$\sigma(\mathcal{G}) = \bigotimes_{i=1}^2 \mathcal{B}([0,1])$ である．したがって，$\sigma\big((\pi([0,1]))^2\big) =$

$\bigotimes_{i=1}^{2} \mathcal{B}([0,1])$ を示せばよい．すると補題 3.26 より，

$$\begin{aligned}
\sigma(\mathcal{I}) &= \sigma(Z^{-1}\big((\pi([0,1]))^2\big) \\
&= Z^{-1}(\sigma\big((\pi([0,1]))^2\big)) \\
&= Z^{-1}\left(\bigotimes_{i=1}^{2} \mathcal{B}([0,1])\right) \\
&= Z^{-1}(\sigma(\mathcal{G})) \\
&= \sigma(Z^{-1}(\mathcal{G})) \\
&= \sigma(X_0, X_1, X_2)
\end{aligned}$$

が分かる．

明らかに $(\pi([0,1]))^2 \subset \bigotimes_{i=1}^{2} \mathcal{B}[0,1]$ である．

任意に $A \times B \in \mathcal{G}$ を取ると，

$$A \times B = (A \times [0,1]) \cap ([0,1] \times B)$$

である．$A \in \mathcal{B}[0,1] = \sigma(\pi([0,1]))$ であるから，$A \times [0,1] \in \sigma\big((\pi([0,1]))^2\big)$ である．これは B に関しても同様である．ゆえに，

$$A \times B \in \sigma\big((\pi([0,1]))^2\big)$$

である．よって，

$$\mathcal{G} \subset \sigma\big((\pi([0,1]))^2\big)$$

つまり，

$$\bigotimes_{i=1}^{2} \mathcal{B}[0,1] = \sigma(\mathcal{G}) \subset \sigma\big((\pi([0,1]))^2\big)$$

である． □

次の命題で，X_3 についてもマルコフ性を満たすことを証明する．

命題 3.28. 全ての $C \in \mathcal{B}[0,1]$ について

$$\mathbb{P}(X_3 \in C \mid X_0, X_1, X_2) = \mathbb{P}(X_3 \in C \mid X_2).$$

証明． $C \in \mathcal{B}[0,1]$ を任意に選ぶ．命題 3.18 より，

$$\mathbb{P}(X_3 \in C \mid X_2) = \int_C f(y \mid X_2, A_2) dy$$

である．よって，

$$\mathbb{P}(X_3 \in C \mid X_0, X_1, X_2) = \int_C f(y \mid X_2, A_2) dy$$

を示せばよい．条件付き期待値の定義から，すべての $F \in \sigma(X_0, X_1, X_2)$ について

$$\int_F I_{\{X_3 \in C\}} d\mathbb{P} = \int_F \int_C f(y \mid X_2, A_2) dy d\mathbb{P}$$

を示せばよい．

$\mathcal{I} := \{X_1^{-1}([0,a]) \cap X_2^{-1}([0,b]) \mid a, b \in [0,1]\}$ とおく．$F \in \mathcal{I}$ を任意にえらび，$F = \{X_1 \leq a\} \cap \{X_2 \leq b\}$ とする．ただし，$a, b, c \in [0,1]$ なる定数である．

$$\begin{aligned}\int_F I_{\{X_3 \in C\}}(t)\mathbb{P}(dt) &= \int_\Omega I_{\{X_1 \leq a\} \cap \{X_2 \leq b\} \cap \{X_3 \in C\}}(t)\mathbb{P}(dt) \\ &= \int_\Omega I_{G_{1a} \cap G_{2b} \cap G_{3C}}(t)\mathbb{P}(dt) \\ &= \mathbb{P}(G_{1a} \cap G_{2b} \cap G_{3C}).\end{aligned}$$

補題 3.21 より，

$$\begin{aligned}\mathbb{P}(G_{1a} \cap G_{2b} \cap G_{3C}) &= \mathbb{P}\left((H_{1a} \cap H_{2b} \cap H_{3C}) \times \prod_{i=4}^{\infty} \Omega_i\right) \\ &= \Big(\mathbb{P}_1 \otimes \mathbb{P}_2 \otimes \mathbb{P}_3(H_{1a} \cap H_{2b} \cap H_{3C})\Big) \times 1 \\ &= \int_{x \in \Omega_1 \times \Omega_2 \times \Omega_3} I_{H_{1a} \cap H_{2b} \cap H_{3C}}(x)\mathbb{P}_1 \otimes \mathbb{P}_2 \otimes \mathbb{P}_3(dx) \\ &= \int_{x \in \Omega_1 \times \Omega_2 \times \Omega_3} I_{H_{1a} \cap H_{2b}}(x) I_{H_{3C}}(x)\mathbb{P}_1 \otimes \mathbb{P}_2 \otimes \mathbb{P}_3(dx) \\ &= \int_{s \in \Omega_1 \times \Omega_2} \left(\int_{r \in \Omega_3} I_{H_{1a} \cap H_{2b}}(s,r) I_{H_{3C}}(s,r)\mathbb{P}_3(dr)\right) \mathbb{P}_1 \otimes \mathbb{P}_2(ds).\end{aligned}$$

なお，最後の行の変形にはフビニの定理を用いた．補題 3.22 より，

$$\begin{aligned}&\int_{s \in \Omega_1 \times \Omega_2} \left(\int_{r \in \Omega_3} I_{H_{1a} \cap H_{2b}}(s,r) I_{H_{3C}}(s,r)\mathbb{P}_3(dr)\right) \mathbb{P}_1 \otimes \mathbb{P}_2(ds) \\ =& \int_{s \in \Omega_1 \times \Omega_2} \left(\int_{r \in \Omega_3} I_{K_{1a} \cap K_{2b}}(s,r) I_{H_{3C}}(s,r)\mathbb{P}_3(dr)\right) \mathbb{P}_1 \otimes \mathbb{P}_2(ds) \\ =& \int_{s \in \Omega_1 \times \Omega_2} I_{K_{1a} \cap K_{2b}}(s) \left(\int_{r \in \Omega_3} I_{H_{3C}}(s,r)\mathbb{P}_3(dr)\right) \mathbb{P}_1 \otimes \mathbb{P}_2(ds).\end{aligned}$$

補題 3.23 より，

$$\begin{aligned}&\int_{s \in \Omega_1 \times \Omega_2} I_{K_{1a} \cap K_{2b}}(s) \left(\int_{r \in \Omega_3} I_{H_{3C}}(s,r)\mathbb{P}_3(dr)\right) \mathbb{P}_1 \otimes \mathbb{P}_2(ds) \\ =& \int_{s \in \Omega_1 \times \Omega_2} I_{K_{1a} \cap K_{2b}}(s) \left(\int_{r \in \Omega_3} I_{\{Y_3 \in C\}}(s,r)\mathbb{P}_3(dr)\right) \mathbb{P}_1 \otimes \mathbb{P}_2(ds) \\ =& \int_{s \in \Omega_1 \times \Omega_2} I_{K_{1a} \cap K_{2b}}(s) \left(\int_C f(y \mid Y_2(s), B_2(s))dy\right) \mathbb{P}_1 \otimes \mathbb{P}_2(ds).\end{aligned}$$

さて，$\Omega \ni \omega = (\omega_1, \omega_2, \omega_3, \cdots)$ を固定したとき，

$$X_2(\omega) = Y_2(\omega_1, \omega_2), A_2(\omega) = B_2(\omega_1, \omega_2)$$

である．だから，この固定した ω に対して

$$\int_C f(y \mid X_2(\omega), A_2(\omega))dy = \int_C f(y \mid Y_2(\omega_1, \omega_2), B_2(\omega_1, \omega_2))dy$$

である．また，$F \in \sigma(X_0, X_1, X_2)$ より，

$$I_F(\omega) = I_{K_{1a} \cap K_{2b}}(\omega_1, \omega_2)$$

も成立する．よって，

$$\begin{aligned}
&\int_F \int_C f(y \mid X_2(\omega), A_2(\omega)) dy d\mathbb{P} \\
&= \int_{\omega \in \Omega} I_F(\omega) \left(\int_C f(y \mid X_2(\omega), A_2(\omega)) dy \right) \mathbb{P}(d\omega) \\
&= \int_{\omega \in \Omega} I_{K_{1a} \cap K_{2b}}(\omega_1, \omega_2) \left(\int_C f(y \mid Y_2(\omega_1, \omega_2), B_2(\omega_1, \omega_2)) dy \right) \mathbb{P}(d\omega) \\
&= \int_{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2} I_{K_{1a} \cap K_{2b}}(\omega_1, \omega_2) \left(\int_C f(y \mid Y_2(\omega_1, \omega_2), B_2(\omega_1, \omega_2)) dy \right) \mathbb{P}_1 \otimes \mathbb{P}_2(d(\omega_1, \omega_2)) \\
&= \int_{s \in \Omega_1 \times \Omega_2} I_{K_{1a} \cap K_{2b}}(s) \left(\int_C f(y \mid Y_2(s), B_2(s)) dy \right) \mathbb{P}_1 \otimes \mathbb{P}_2(ds)
\end{aligned}$$

となる．

したがって任意の $F \in \mathcal{I}$ に対して

$$\int_F I_{\{X_3 \in C\}} d\mathbb{P} = \int_F \int_C f(y \mid X_2, A_2) dy d\mathbb{P}$$

が成立する．

$\mathcal{I}$ は π 系であり，補題 3.26 より，$\sigma(\mathcal{I}) = \sigma(X_0, X_1, X_2)$ である．よって，任意の $F \in \sigma(X_0, X_1, X_2)$ に対して

$$\int_F I_{\{X_3 \in C\}} d\mathbb{P} = \int_F \int_C f(y \mid X_2, A_2) dy d\mathbb{P}$$

が成立する． $\square$

3.2.5 X_n のマルコフ性

$i = n(\geq 3)$ のときは $i = 3$ のときと同様にできる．基本的に 3.2.5 内の X_3 を X_n に，X_2 を X_{n-1} に置き換えるとよい．

X_n のマルコフ性を調べるためにいくつかの補題を提示するが，その証明は X_3 のマルコフ性を調べたときと同じなので省略する．

定義 3.29. n を自然数とする．$a_1, \cdots, a_{n-1}, b_1, \cdots, b_{n-1}, c_1, \cdots, c_{n-1} \in [0, 1]$ とする．$C, D \in \mathcal{B}[0, 1]$ とする．このとき，Ω の部分集合 $G_1, \cdots, G_{n-1}, G_n$ と $\Omega_1 \times \cdots \times \Omega_n$ の部分集合 $H_1, \cdots, H_{n-1}, H_n$，および $\Omega_1 \times \cdots \times \Omega_{n-1}$ の部分集合 $K_1, \cdots, K_{n-1}$ を以下のように定義

する．

$$
\begin{aligned}
G_1 &:= \{t \in \Omega \mid U_1(\rho_1(t)) \le \int_0^a f(y \mid X_0, A_0)dy\},\\
&\vdots\\
G_{n-1} &:= \{t \in \Omega \mid U_{n-1}(\rho_{n-1}(t)) \le \int_0^{a_{n-1}} f(y \mid X_{n-2}(t), A_{n-2}(t))dy\},\\
G_n &:= \{X_n \in C\},\\
H_1 &:= \{t \mid U_1(t_1) \le \int_0^{b_1} f(y \mid Y_0, B_0)dy\},\\
&\vdots\\
H_{n-1} &:= \{t \mid s = (t_1, \cdots, t_{n-2}),\ U_{n-1}(t_{n-1}) \le \int_0^{b_{n-1}} f(y \mid Y_{n-2}(s), Y_{n-2}(s))dy\}\\
H_n &:= \{Y_n \in D\},\\
K_1 &:= \{r \mid U_1(r_1) \le \int_0^{c_1} f(y \mid X_0, A_0)dy\},\\
&\vdots\\
K_{n-1} &:= \{r \mid u = (r_1, \cdots, r_{n-2}),\ U_{n-1}(r_{n-1}) \le \int_0^{c_{n-1}} f(y \mid Y_{n-2}(u), B_{n-2}(u))dy\}.
\end{aligned}
$$

補題 3.30.

$$G_1 \cap \cdots \cap G_n = (H_1 \cap \cdots \cap H_n) \times \prod_{i=n+1}^{\infty} \Omega_i.$$

補題 3.31. $\omega_1 \in \Omega_1, \cdots, \omega_{n-1} \in \Omega_{n-1}$ を固定する．全ての $\omega_n \in \Omega_n$ について

$$I_{K_1 \cap \cdots \cap K_{n-1}}(\omega_1, \cdots, \omega_{n-1}) = I_{H_1 \cap \cdots \cap H_{n-1}}(\omega_1, \cdots, \omega_n).$$

補題 3.32. $(\omega_1, \cdots, \omega_{n-1}) \in \Omega_1 \times \cdots \Omega_{n-1}$ が固定されている．$C \in \mathcal{B}[0,1]$ とする．このとき，

$$\int_{r \in \Omega_n} I_{H_n}(\omega_1, \cdots, \omega_{n-1}, r)\mathbb{P}_n(dr) = \int_C f(y \mid Y_{n-1}(\omega_1, \cdots, \omega_{n-1}), B_{n-1}(\omega_1, \cdots, \omega_{n-1}))dy..$$

補題 3.33.

$$\mathcal{I} := \{X_1^{-1}([0, a_1]) \cap \cdots \cap X_{n-1}^{-1}([0, a_{n-1}]) \mid a_1, \cdots, a_{n-1} \in [0,1]\}$$

とする．このとき，

$$\sigma(X_0, X_1, \cdots, X_{n-1}) = \sigma(\mathcal{I})$$

が成立する．

命題 3.34. 全ての $C \in \mathcal{B}[0,1]$ について

$$\mathbb{P}(X_n \in C \mid X_0, X_1, \cdots, X_{n-1}) = \mathbb{P}(X_n \in C \mid X_{n-1}).$$

証明． $C \in \mathcal{B}[0,1]$ を任意に選ぶ．命題 3.18 より，

$$\mathbb{P}(X_n \in C \mid X_{n-1}) = \int_C f(y \mid X_{n-1}, A_{n-1})dy$$

である．よって，

$$\mathbb{P}(X_n \in C \mid X_0, X_1, \cdots, X_{n-1}) = \int_C f(y \mid X_{n-1}, A_{n-1})dy$$

を示せばよい．条件付き期待値の定義から，すべての $F \in \sigma(X_0, X_1, \cdots, X_{n-1})$ について

$$\int_F I_{\{X_n \in C\}} d\mathbb{P} = \int_F \int_C f(y \mid X_{n-1}, A_{n-1})dyd\mathbb{P}$$

を示せばよい．集合 $\mathcal{I} := \{X_1^{-1}([0,a_1]) \cap \cdots \cap X_{n-1}^{-1}([0,a_{n-1}]) \mid a_1, \cdots, a_{n-1} \in [0,1]\}$ とおく．$F \in \mathcal{I}$ を任意にえらび，$F = \{X_1 \leq a_1\} \cap \cdots \cap \{X_{n-1} \leq a_{n-1}\}$ とする．ただし，$a_1, \cdots, a_{n-1} \in [0,1]$ なる定数である．

$$\begin{aligned}
\int_F I_{\{X_n \in C\}}(t)\mathbb{P}(dt) &= \int_\Omega I_{\{X_1 \leq a\} \cap \cdots \cap \{X_{n-1} \leq a_{n-1}\} \cap \{X_n \in C\}}(t)\mathbb{P}(dt) \\
&= \int_\Omega I_{G_1 \cap \cdots \cap G_n}(t)\mathbb{P}(dt) \\
&= \mathbb{P}(G_1 \cap \cdots \cap G_n).
\end{aligned}$$

補題 3.30 より，

$$\begin{aligned}
\mathbb{P}(G_1 \cap \cdots \cap G_n) &= \mathbb{P}\left((H_1 \cap \cdots \cap H_n) \times \prod_{i=n+1}^{\infty} \Omega_i\right) \\
&= \mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_n(H_1 \cap \cdots \cap H_n) \times 1 \\
&= \int_{x \in \Omega_1 \times \cdots \times \Omega_n} I_{H_1 \cap \cdots \cap H_n}(x)\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_n(dx) \\
&= \int_{x \in \Omega_1 \times \cdots \times \Omega_n} I_{H_1 \cap \cdots \cap H_{n-1}}(x) I_{H_n}(x)\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_n(dx) \\
&= \int_{s \in \Omega_1 \times \cdots \times \Omega_{n-1}} \left(\int_{r \in \Omega_n} I_{H_1 \cap \cdots \cap H_{n-1}}(s,r) I_{H_n}(s,r)\mathbb{P}_n(dr)\right) \mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_{n-1}(ds).
\end{aligned}$$

なお，最後の行の変形にはフビニの定理を用いた．補題 3.31 より，

$$\begin{aligned}
&\int_{s \in \Omega_1 \times \cdots \times \Omega_{n-1}} \left(\int_{r \in \Omega_n} I_{H_1 \cap \cdots \cap H_{n-1}}(s,r) I_{H_n}(s,r)\mathbb{P}_n(dr)\right) \mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_{n-1}(ds) \\
=& \int_{s \in \Omega_1 \times \cdots \times \Omega_{n-1}} \left(\int_{r \in \Omega_n} I_{K_1 \cap \cdots \cap K_{n-1}}(s) I_{H_n}(s,r)\mathbb{P}_n(dr)\right) \mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_{n-1}(ds) \\
=& \int_{s \in \Omega_1 \times \cdots \times \Omega_{n-1}} I_{K_1 \cap \cdots \cap K_{n-1}}(s) \left(\int_{r \in \Omega_n} I_{H_n}(s,r)\mathbb{P}_n(dr)\right) \mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_{n-1}(ds).
\end{aligned}$$

補題 3.32 より，

$$\int_{s\in\Omega_1\times\cdots\times\Omega_{n-1}} I_{K_1\cap\cdots\cap K_{n-1}}(s)\left(\int_{r\in\Omega_n} I_{H_n}(s,r)\mathbb{P}_n(dr)\right)\mathbb{P}_1\otimes\cdots\otimes\mathbb{P}_{n-1}(ds)$$
$$=\int_{s\in\Omega_1\times\cdots\times\Omega_{n-1}} I_{K_1\cap\cdots\cap K_{n-1}}(s)\left(\int_0^{a_n} f(y\mid Y_{n-1}(s),B_{n-1}(s))dy\right)\mathbb{P}_1\otimes\cdots\otimes\mathbb{P}_{n-1}(ds).$$

さて，$\Omega\ni\omega=(\omega_1,\cdots,\omega_{n-1},\cdots)$ を固定したとき，

$$X_{n-1}(\omega)=Y_{n-1}(\omega_1,\cdots,\omega_{n-1}),A_{n-1}(\omega)=B_{n-1}(\omega_1,\cdots,\omega_{n-1})$$

だから，

$$\int_0^{a_n} f(y\mid X_{n-1}(\omega),A_{n-1}(\omega))dy=\int_0^{a_n} f(y\mid Y_{n-1}(\omega_1,\cdots,\omega_{n-1}),B_{n-1}(\omega_1,\cdots,\omega_{n-1}))dy$$

である．また，$F\in\sigma(X_0,X_1,\cdots,X_{n-1})$ より，

$$I_F(\omega)=I_{K_1\cap\cdots\cap K_{n-1}}(\omega_1,\cdots,\omega_{n-1})$$

も成立する．よって，

$$\int_F\int_0^{a_n} f(y\mid X_{n-1}(\omega),A_{n-1}(\omega))dyd\mathbb{P}$$
$$=\int_{\omega\in\Omega} I_F(\omega)\left(\int_0^{a_n} f(y\mid Y_{n-1}(\omega_1,\cdots,\omega_{n-1}),B_{n-1}(\omega_1,\cdots,\omega_{n-1}))dy)\right)\mathbb{P}(d\omega)$$
$$=\int_{\omega\in\Omega} I_{K_1\cap\cdots\cap K_{n-1}}(\omega_1,\cdots,\omega_{n-1})\left(\int_0^{a_n} f(y\mid Y_{n-1}(\omega_1,\cdots,\omega_{n-1}),B_{n-1}(\omega_1,\cdots,\omega_{n-1}))dy\right)\mathbb{P}(d\omega)$$
$$=\int_{s\in\Omega_1\times\cdots\times\Omega_{n-1}} I_{K_1\cap\cdots\cap K_{n-1}}(s)\left(\int_0^{a_n} f(y\mid Y_{n-1}(s),B_{n-1}(s))dy\right)\mathbb{P}_1\otimes\cdots\otimes\mathbb{P}_{n-1}(ds)$$

となる．

したがって，$\mathcal{I}$ 上で

$$\int_F I_{\{X_n\in C\}}d\mathbb{P}=\int_F\int_0^{a_n} f(y\mid X_{n-1},A_{n-1})dyd\mathbb{P}$$

が成立する．

$\mathcal{I}$ は π 系であり，補題 3.33 より，$\sigma(\mathcal{I})=\sigma(X_0,X_1,\cdots,X_{n-1})$ である．よって，任意の $F\in\sigma((X_0,X_1,\cdots,X_{n-1})$ に対して

$$\int_F I_{\{X_n\in C\}}d\mathbb{P}=\int_F\int_0^{a_n} f(y\mid X_{n-1},A_{n-1})dyd\mathbb{P}$$

が成立する． □

以上により，(X_n) がマルコフ連鎖であることが証明された．

命題 3.35. $(\Omega,\mathcal{F},\mathbb{P})$ 上の確率変数列 $((X_n,A_n);n\in\mathbb{N})$ はマルコフ連鎖になる．このマルコフ連鎖の遷移確立は $f(y\mid x,a)$ である．

4　主定理とその証明

簡単な強化学習のモデルを定義し，そのモデルの上で価値関数を最大にする方策を見つけるためのアルゴリズムを提示する．

4.1　モデルの定義

4×4 のマス目の迷路がある．左上の角がスタート地点で，右下の角がゴール地点とする．この迷路に挑戦する人は時間が一つ進むたびに，ランダムに上下左右の一つ隣のマスに移動する．[*2]各マス目を左上から順に $x_1, x_2, \cdots, x_{16}$ と名づけ，その集合を $\mathcal{X}$ とする．すなわち，

$$\mathcal{X} := \{(i, j) \mid i, j \in \{1, 2, 3, 4\}\}$$

と定義し，

$$x_1 = (1, 1), x_2 = (1, 2), \cdots, x_5 = (2, 1), \cdots, x_{16} = (4, 4)$$

と定義するのである．任意のマス x に対して $R(x)$ という実数値の報酬が与えられる．$\max_{x \in \mathcal{X}} |R(x)| < \infty$ であるとする．

$$\Pi := \{(q_1, q_2, q_3) \in [0, 1]^3 \mid \forall i, 0 \leq q_i \leq 1, \quad \text{and} \quad q_1 + q_2 + q_3 \leq 1\}$$

とおく．π は行動と遷移確立を一つにまとめた考えたものと見ることができる．

状態空間を $\mathcal{X}$ とするマルコフ連鎖 (X_t) を考えたい．$\pi \in \Pi$ を任意にひとつ固定する．このとき (X_t) の遷移関数は次で与える．また，以下で $e_1 = (1, 0), e_2 = (0, 1)$ である．$q_4 = 1 - (q_1 + q_2 + q_3)$ である．

$$f^\pi(x, y) = \begin{cases} 0 & \text{if} \quad y - x \neq \pm e_1, \pm e_2 \quad \text{or} \quad x = (4, 4), x \neq y \\ 1 & \text{if} \quad x = y = (4, 4) \\ q_1 & \text{if} \quad y - x = e_1 \quad \text{or} \quad x = y, x = (4, 2), (4, 3) \\ q_2 & \text{if} \quad y - x = e_2 \quad \text{or} \quad x = y, x = (2, 4), (3, 4) \\ q_3 & \text{if} \quad y - x = -e_1 \quad \text{or} \quad x = y, x = (1, 2), (1, 3) \\ q_4 & \text{if} \quad y - x = -e_2 \quad \text{or} \quad x = y, x = (2, 1), (3, 1) \\ q_3 + q_4 & \text{if} \quad x = y = (1, 1) \\ q_1 + q_4 & \text{if} \quad x = y = (1, 4) \\ q_2 + q_3 & \text{if} \quad x = y = (4, 1) \end{cases}.$$

遷移関数を遷移行列としてとらえた 16×16 の行列を f^π と表すことにする．

この遷移確率を持つマルコフ連鎖を (X^π) で表すことにする．

後の計算のために次の補題を述べておく．

[*2] 「迷路の左端にいるため左に進めない」など，ランダムに選ばれた進行方向に進むことができない場合はその場にとどまることにする．

補題 4.1. $I - \gamma f^\pi$ には逆行列が存在し，それは次の行列の等比級数である．

$$(I - \gamma f^\pi)^{-1} = \sum_{n \geq 0} (\gamma f^\pi)^n$$

証明． Appendix の 6.2 節を参照. □

4.2 価値関数

$\gamma \in [0, 1]$ を割引率とする．価値関数 $V(\pi, x)\colon \Pi \times \mathcal{X} \to \mathbb{R}$ は次のように定義される．

$$V(\pi, x) = \mathbb{E}\left[\sum_{n=0}^{\infty} \gamma^n R(X_{n+1}^\pi) \mid X_0^\pi = x\right] \quad (\forall x \in \mathcal{X}).$$

補題 4.2. $V(\pi, x)$ は存在する．すなわち，$\mathbb{E}\left[\sum_{n=0}^{\infty} \gamma^n R(X_{n+1}^\pi) \mid X_0^\pi = x\right]$ は値を持つ．

証明．

$$\begin{aligned}
&\mathbb{E}\left[\sum_{n=0}^{\infty} \gamma^n R(X_{n+1}^\pi) \mid X_0^\pi = x\right] \\
\leq &\mathbb{E}\left[M \sum_{n=0}^{\infty} \gamma^n \mid X_0^\pi = x\right] \\
= &\frac{M}{1 - \gamma}.
\end{aligned}$$

よって値を持つ. □

4.3 価値関数の偏微分可能性と偏微分の平均による表現

補題 4.3. 全ての $x \in \mathcal{X}, \pi \in \Pi$ について次が成り立つ．

$$V(\pi, x) = \mathbb{E}\left[R(X_1^\pi) + \gamma V(\pi, X_1^\pi) \mid X_0^\pi = x\right].$$

証明.

$$
\begin{aligned}
V(\pi, x) &= \mathbb{E}\left[\sum_{n=0}^{\infty} \gamma^n R(X_{n+1}^{\pi}) \mid X_0^{\pi} = x\right] \\
&= \mathbb{E}\left[R(X_1^{\pi}) + \sum_{n=1}^{\infty} \gamma^n R(X_{n+1}^{\pi}) \mid X_)^{\pi} = x\right] \\
&= \mathbb{E}\left[R(X_1^{\pi}) \mid X_0^{\pi} = x\right] + \mathbb{E}\left[\mathbb{E}\left[\gamma \sum_{n=1}^{\infty} \gamma^{n-1} R(X_{n+1}^{\pi} \mid X_0^{\pi} = x, X_1^{\pi}\right] \mid X_0^{\pi} = x\right] \\
&= \mathbb{E}\left[R(X_1^{\pi}) \mid X_0^{\pi} = x\right] + \mathbb{E}\left[\mathbb{E}\left[\gamma \sum_{n=1}^{\infty} \gamma^{n-1} R(X_{n+1}^{\pi} \mid X_1^{\pi}\right] \mid X_0^{\pi} = x\right] \\
&= \mathbb{E}\left[R(X_1^{\pi}) \mid X_0^{\pi} = x\right] + \mathbb{E}\left[\gamma V(X_1^{\pi}) \mid X_0^{\pi} = x\right] \\
&= \mathbb{E}\left[R(X_1^{\pi}) + \gamma V(\pi, X_1^{\pi}) \mid X_0^{\pi} = x\right].
\end{aligned}
$$

□

価値関数は π に関して微分可能である．

補題 4.4. $V(\pi, x)$ は π に関して勾配が存在する．すなわち，$\frac{\partial}{\partial q_i} V(\pi, x) \quad (i = 1, 2, 3)$ が存在する．

証明. $\pi = (q_1, q_2, q_3), \Delta = (h_1, 0, 0)$ とする．ただし，h_1 は十分小さい．すなわち $q_1 + h_1 + q_2 + q_3 < 1$ である．

$$
\lim_{h_1 \to 0} \frac{V(\pi, x) - V(\pi + \Delta, x)}{h_1}
$$

が存在することを言えば十分である．価値関数の差から計算する．

$$
\begin{aligned}
&V(\pi, x) - V(\pi + \delta, x) \\
=&E^{\pi}\left[R(X_1^{\pi}) + \gamma V(\pi, X_1^{\pi}) \mid X_0 = x\right] - E^{\pi+\Delta}\left[R(X_1^{\pi+\Delta}) + \gamma V(\pi + \Delta, X_1^{\pi+\Delta}) \mid X_0 = x\right] \\
=&\sum_{y \in \mathcal{X}} (R(y) + \gamma V(\pi, y)) f^{\pi}(x, y) - \sum_{y \in \mathcal{X}} (R(y) + \gamma V(\pi + \Delta, y)) f^{\pi+\Delta}(x, y) \\
=&\sum_{y \in \mathcal{X}} R(y) \left(f^{\pi}(x, y) - f^{\pi+\Delta}(x, y)\right) + \gamma \sum_{y \in \mathcal{X}} \left(V(\pi, y) f^{\pi}(x, y) - V(\pi + \Delta, y) f^{\pi+\Delta}(x, y)\right) \\
=&\sum_{y \in \mathcal{X}} R(y) \left(f^{\pi}(x, y) - f^{\pi+\Delta}(x, y)\right) \\
&\qquad + \gamma \sum_{y \in \mathcal{X}} \left(V(\pi, y) \left(f^{\pi}(x, y) - f^{\pi+\Delta}(x, y)\right) + \left(V(\pi, y) - V(\pi + \Delta, y)\right) f^{\pi+\Delta}(x, y)\right).
\end{aligned}
$$

したがって，

$$
\begin{aligned}
&(V(\pi, x) - V(\pi + \Delta, x)) - \gamma \sum_{y \in \mathcal{X}} (V(\pi, y) - V(\pi + \Delta, y)) f^{\pi+\Delta}(x, y) \\
=& \sum_{y \in \mathcal{X}} \left\{R(y) \left(f^{\pi}(x, y) - f^{\pi+\Delta}(x, y)\right) + \gamma V(\pi, y) \left(f^{\pi}(x, y) - f^{\pi+\Delta}(x, y)\right)\right\}.
\end{aligned}
$$

行列を用いると次のように表すことができる．

$$(I-\gamma f^{\pi+\Delta})(V(\pi,\cdot)-V(\pi+\Delta,\cdot))^\top=\left(f^\pi-f^{\pi+\Delta}\right)(R+\gamma V(\pi,\cdot))^\top.$$

補題 4.1 より $(I-\gamma f^{\pi+\Delta}$ には逆行列が存在する．それを両辺に掛けると次のように計算が進む．

$$\begin{aligned}
&(V(\pi,\cdot)-V(\pi+\Delta,\cdot))^\top\\
=&(I-\gamma f^{\pi+\Delta})^{-1}\left(f^\pi-f^{\pi+\Delta}\right)(R+\gamma V(\pi,\cdot))^\top\\
=&\left(\sum_{n=0}^{\infty}\left(\gamma f^{\pi+\Delta}\right)^n\right)\left(f^\pi-f^{\pi+\Delta}\right)(R+\gamma V(\pi,\cdot))^\top\\
=&\left(\sum_{n=0}^{\infty}\gamma^n\left(f^{\pi+\Delta}\right)^n\right)\left(f^\pi-f^{\pi+\Delta}\right)(R+\gamma V(\pi,\cdot))^\top\\
=&\sum_{n=0}^{\infty}\gamma^n\left\{\left(f^{\pi+\Delta}\right)^n\left(f^\pi-f^{\pi+\Delta}\right)(R+\gamma V(\pi,\cdot))^\top\right\}.
\end{aligned}$$

したがって，これを再びベクトルの成分で見直すと次のようになる．

$$\begin{aligned}
&V(\pi,x)-V(\pi+\Delta,x)\\
=&\sum_{n=0}^{\infty}\gamma^n\left\{\sum_{z\in\mathcal{X}}\left(\sum_{y\in\mathcal{X}}\left(f^{\pi+\Delta}\right)^n(x,y)\left(f^\pi-f^{\pi+\Delta}\right)(y,z)\right)(R(z)+\gamma V(\pi,z))\right\}.
\end{aligned}$$

である．両辺を h_1 で割って $h_1\to 0$ とすると

$$\begin{aligned}
&\lim_{h_1\to 0}\frac{V(\pi,x)-V(\pi+\Delta,x)}{h_1}\\
=&\lim_{h_1\to 0}\frac{1}{h_1}\sum_{n=0}^{\infty}\gamma^n\left\{\sum_{z\in\mathcal{X}}\left(\sum_{y\in\mathcal{X}}\left(f^{\pi+\Delta}\right)^n(x,y)\left(f^\pi-f^{\pi+\Delta}\right)(y,z)\right)(R(z)+\gamma V(\pi,z))\right\}.
\end{aligned}$$

右辺の極限について考える．

$$\begin{aligned}
&\frac{\gamma^n}{h_1}\left\{\sum_{z\in\mathcal{X}}\left(\sum_{y\in\mathcal{X}}\left(f^{\pi+\Delta}\right)^n(x,y)\left(f^\pi-f^{\pi+\Delta}\right)(y,z)\right)(R(z)+\gamma V(\pi,z))\right\}\\
=&\gamma^n\left\{\sum_{z\in\mathcal{X}}\left(\sum_{y\in\mathcal{X}}\left(f^{\pi+\Delta}\right)^n(x,y)\frac{\left(f^\pi-f^{\pi+\Delta}\right)(y,z)}{h_1}\right)(R(z)+\gamma V(\pi,z))\right\}\\
\leq&\gamma^n\sum_{z\in\mathcal{X}}|R(z)+\gamma V(\pi,z)|\left(\sum_{y\in\mathcal{X}}\left|\left(f^{\pi+\Delta}\right)^n(x,y)\right|\left|\frac{\left(f^\pi-f^{\pi+\Delta}\right)(y,z)}{h_1}\right|\right).
\end{aligned}$$

ここで，R,V は有界であり，$\left(f^{\pi+\Delta}\right)^n$ は n-step 遷移関数なのでその絶対値は 1 以下である．また，$\dfrac{\left(f^\pi-f^{\pi+\Delta}\right)(y,z)}{h_1}$ に関しては平均値の定理よりある $0\leq h\leq h_1$ が存在して $\Delta'=(h,0,0)$ としたとき，

$$\frac{\left(f^\pi-f^{\pi+\Delta}\right)(y,z)}{h_1}=\frac{\partial}{\partial q_1}f^{\pi+\Delta'}(y,z)$$

となる．しかし，

$$\left|\frac{\partial}{\partial q_1} f^{\pi+\Delta'}(y,z)\right| \leq 1$$

であるから，これも有界．したがって，

$$\begin{aligned}
&\frac{\gamma^n}{h_1}\left\{\sum_{z\in\mathcal{X}}\left(\sum_{y\in\mathcal{X}}\left(f^{\pi+\Delta}\right)^n(x,y)\left(f^{\pi}-f^{\pi+\Delta}\right)(y,z)\right)(R(z)+\gamma V(\pi,z))\right\}\\
\leq&\gamma^n\sum_{z\in\mathcal{X}}M(\sum_{y\in\mathcal{X}}1\times 1)\\
=&\gamma^n M\left|\mathcal{X}\right|
\end{aligned}$$

となる．ただし，M は $|R(z)+\gamma V(\pi,z)| \leq M$ なる定数で，$|\mathcal{X}|$ は $\mathcal{X}$ の要素の個数．よって，有界収束定理より

$$\begin{aligned}
&\lim_{h_1\to 0}\frac{V(\pi,x)-V(\pi+\Delta,x)}{h_1}\\
=&\sum_{n=0}^{\infty}\gamma^n\left\{\sum_{z\in\mathcal{X}}\left(\sum_{y\in\mathcal{X}}\left(f^{\pi+\Delta}\right)^n(x,y)\lim_{h_1\to 0}\frac{\left(f^{\pi}-f^{\pi+\Delta}\right)(y,z)}{h_1}\right)(R(z)+\gamma V(\pi,z))\right\}\\
=&\sum_{n=0}^{\infty}\gamma^n\left\{\sum_{z\in\mathcal{X}}\left(\sum_{y\in\mathcal{X}}\left(f^{\pi+\Delta}\right)^n(x,y)\frac{\partial}{\partial q_1}f^{\pi}(y,z)\right)(R(z)+\gamma V(\pi,z))\right\}.
\end{aligned}$$

したがって，微分可能である． □

補題 4.5.

$$\frac{\partial}{\partial q_i}V(\pi,x)=\sum_{y\in\mathcal{X}}\left\{(R(y)+\gamma V(\pi,y))\frac{\partial}{\partial q_i}f^{\pi}(x,y)+\gamma f^{\pi}(x,y)\frac{\partial}{\partial q_i}V(\pi,y)\right\}.$$

証明.

$$\begin{aligned}
&\frac{\partial}{\partial q_i}V(\pi,x)\\
=&\frac{\partial}{\partial q_i}\sum_{y\in\mathcal{X}}(R(y)+\gamma V(\pi,y))f^{\pi}(x,y)\\
=&\sum_{y\in\mathcal{X}}\left\{(R(y)+\gamma V(\pi,y))\frac{\partial}{\partial q_i}f^{\pi}(x,y)+\gamma f^{\pi}(x,y)\frac{\partial}{\partial q_i}V(\pi,y)\right\}.
\end{aligned}$$

□

補題 4.5 において，$f^{\pi}(x,\cdot)$ を確率測度と思うと次が得られる．

補題 4.6.

$$\frac{\partial}{\partial q_i}V(\pi,X_0^{\pi})=\mathbb{E}^{f^{\pi}(X_0^{\pi},\cdot)}\left[\frac{\{R(X_1^{\pi})+\gamma V(\pi,X_1^{\pi})\}\frac{\partial}{\partial q_i}f^{\pi}(X_0^{\pi},X_1^{\pi})}{f^{\pi}(X_0^{\pi},X_1^{\pi})}+\gamma\frac{\partial}{\partial q_i}V(\pi,X_1^{\pi})\right]. \quad (1)$$

ただし，$\mathbb{E}^{f^\pi(X_0^\pi,\cdot)}$ は $f^\pi(X_0^\pi,\cdot)$ を確率測度として見たときの平均を表す．

補題 4.6 の期待値の中身を F とおくことにする．

定義 4.7. $F\colon \Pi\times\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ を次で定義する．

$$F(\pi,X_0^\pi,X_1^\pi)=\frac{\{R(X_1^\pi)+\gamma V(\pi,X_1^\pi)\}\nabla_\pi f^\pi(X_0^\pi,X_1^\pi)}{f^\pi(X_0^\pi,X_1^\pi)}+\gamma\nabla_\pi V(\pi,X_1^\pi). \tag{2}$$

この表現により，V の勾配を次のように表すことができる．

$$\nabla_\pi V(\pi,X_0^\pi)=\mathbb{E}^{f^\pi(X_0^\pi,\cdot)}\left[F(\pi,X_0^\pi,X_1^\pi)\right].$$

4.4 価値関数の解析

価値関数の解析のためにノルムを導入する．今回のモデルの場合，最終的に価値関数は π に関してリプシッツ連続 (命題 4.17) であることが分かる．有限次元空間の場合，どのノルムを用いても良いので計算しやすいノルムを採用する．

$\|\cdot\|_\infty$ を ∞-ノルムとする．すなわち，$\mathbf{a}\in\mathbb{R}^{16}$ に対し，

$$\|\mathbf{a}\|_\infty=\sum_{i=1}^{16}|a_i|$$

と定める．また，行列ノルムを ∞-ノルムから誘導されるノルムとする．すなわち，$A\in\mathbb{R}^{16\times16}$ に対して

$$\|A\|_\infty=\max_{1\le i\le 16}\sum_{j=1}^{16}|a_{ij}|$$

とする．A のノルムは行ごとに成分の絶対値の和を計算したものの最大である．

Appendix の節で，$\|\mathbf{a}\|_\infty$ から $\|A\|_\infty$ が誘導されること，$\|A\|_\infty,\|\mathbf{a}\|_\infty$ は両立するノルムであることを証明する．

まず，今回の状況において価値関数は有界であることを示す．

補題 4.8. すべての $\pi\in\Pi, x\in\mathcal{X}$ について $V(\pi,x)$ は一様有界．

証明.

$$V(\pi,x)=\mathbb{E}\left[\sum\gamma^n R(X_{n+1})\mid X_0=x\right]$$

であるが，ここで今回のモデルの仮定より R は有界で，$0<\gamma<1$ である．したがってある定数 M が存在して

$$V(\pi,x)\le\frac{1}{1-\gamma}\sup_{y\in\mathcal{X}}R(y)\le\frac{M}{1-\gamma}$$

となる．よって $V(\pi,x)$ は一様有界． □

価値関数の偏微分の差を評価する．その計算のためにこの補題を示す．

補題 4.9. $i = 1, 2, 3$ とする．$x, y \in \mathcal{X}, \quad \pi, \pi' \in \Pi$ とする．このとき，

$$\frac{\partial}{\partial q_i} f^{\pi}(x, y) = \frac{\partial}{\partial q_i} f^{\pi'}(x, y)$$

である．

証明. $\pi = (q_1, q_2, q_3), \pi' = (q_1', q_2', q_3')$ とする．f^{π} の各 q_i を q_i' に変えたものが $f^{\pi'}$ であるから，$\frac{\partial}{\partial q_i} f^{\pi}(x, y) = \frac{\partial}{\partial q_i} f^{\pi'}(x, y)$ がなりたつ． □

補題 4.10.

$$\begin{aligned}
&\left(\frac{\partial}{\partial q_i} V(\pi, x) - \frac{\partial}{\partial q_i} V(\pi', x)\right)(1 - \gamma f^{\pi}(x, x)) - \sum_{y \neq x} \left\{\gamma f^{\pi}(x, y) \left(\frac{\partial}{\partial q_i} V(\pi, x) - \frac{\partial}{\partial q_i} V(\pi', x)\right)\right\} \\
&\quad = \sum_{y \in \mathcal{X}} \left\{\gamma \left(V(\pi, y) - V(\pi', y)\right) \frac{\partial}{\partial q_i} f^{\pi}(x, y) + \gamma \left(f^{\pi}(x, y) - f^{\pi'}(x, y)\right) \frac{\partial}{\partial q_i} V(\pi, y)\right\}.
\end{aligned}$$

証明. $\pi, \pi' \in \Pi$ を任意に固定する．補題 4.5 より，

$$\frac{\partial}{\partial q_i} V(\pi, x) = \sum_{y \in \mathcal{X}} \left\{(R(y) + \gamma V(\pi, y)) \frac{\partial}{\partial q_i} f^{\pi}(x, y) + \gamma f^{\pi}(x, y) \frac{\partial}{\partial q_i} V(\pi, y)\right\}$$

である．したがって，

$$\begin{aligned}
&\frac{\partial}{\partial q_i} V(\pi, x) - \frac{\partial}{\partial q_i} V(\pi', x) \\
&= \sum_{y \in \mathcal{X}} \left\{(R(y) + \gamma V(\pi, y)) \frac{\partial}{\partial q_i} f^{\pi}(x, y) + \gamma f^{\pi}(x, y) \frac{\partial}{\partial q_i} V(\pi, y)\right\} \\
&\quad - \sum_{y \in \mathcal{X}} \left\{(R(y) + \gamma V(\pi', y)) \frac{\partial}{\partial q_i} f^{\pi'}(x, y) + \gamma f^{\pi'}(x, y) \frac{\partial}{\partial q_i} V(\pi', y)\right\} \\
&= \sum_{y \in \mathcal{X}} \left\{ R(y) \left(\frac{\partial}{\partial q_i} f^{\pi}(x, y) - \frac{\partial}{\partial q_i} f^{\pi'}(x, y)\right) + \gamma \left(V(\pi, y) \frac{\partial}{\partial q_i} f^{\pi}(x, y) - V(\pi', y) \frac{\partial}{\partial q_i} f^{\pi'}(x, y)\right) \right. \\
&\qquad \left. + \gamma \left(f^{\pi}(x, y) \frac{\partial}{\partial q_i} V(\pi, y) - f^{\pi'}(x, y) \frac{\partial}{\partial q_i} V(\pi', y)\right)\right\}.
\end{aligned}$$

ここで補題 4.9 より，

$$R(y) \left(\frac{\partial}{\partial q_i} f^{\pi}(x, y) - \frac{\partial}{\partial q_i} f^{\pi'}(x, y)\right) = 0$$

である．したがって，

$$
\begin{aligned}
&\frac{\partial}{\partial q_i}V(\pi,x)-\frac{\partial}{\partial q_i}V(\pi',x)\\
&=\sum_{y\in\mathcal{X}}\Biggl\{\gamma\left(V(\pi,y)\frac{\partial}{\partial q_i}f^{\pi}(x,y)-V(\pi',y)\frac{\partial}{\partial q_i}f^{\pi'}(x,y)\right)\\
&\qquad\qquad+\gamma\left(f^{\pi}(x,y)\frac{\partial}{\partial q_i}V(\pi,y)-f^{\pi'}(x,y)\frac{\partial}{\partial q_i}V(\pi',y)\right)\Biggr\}\\
&=\sum_{y\in\mathcal{X}}\Biggl\{\gamma\left(V(\pi,y)-V(\pi',y)\right)\frac{\partial}{\partial q_i}f^{\pi}(x,y)\\
&\qquad\qquad+\gamma\left(f^{\pi}(x,y)\frac{\partial}{\partial q_i}V(\pi,y)-f^{\pi'}(x,y)\frac{\partial}{\partial q_i}V(\pi,y)+f^{\pi'}(x,y)\frac{\partial}{\partial q_i}V(\pi,y)-f^{\pi'}(x,y)\frac{\partial}{\partial q_i}V(\pi',y)\right)\Biggr\}\\
&=\sum_{y\in\mathcal{X}}\Biggl\{\gamma\left(V(\pi,y)-V(\pi',y)\right)\frac{\partial}{\partial q_i}f^{\pi}(x,y)+\gamma\left(f^{\pi}(x,y)-f^{\pi'}(x,y)\right)\frac{\partial}{\partial q_i}V(\pi,y)\\
&\qquad\qquad+\gamma f^{\pi'}(x,y)\left(\frac{\partial}{\partial q_i}V(\pi,y)-\frac{\partial}{\partial q_i}V(\pi',y)\right)\Biggr\}.
\end{aligned}
$$

よって，

$$
\begin{aligned}
&\left(\frac{\partial}{\partial q_i}V(\pi,x)-\frac{\partial}{\partial q_i}V(\pi',x)\right)-\sum_{y\in\mathcal{X}}\gamma f^{\pi'}(x,y)\left(\frac{\partial}{\partial q_i}V(\pi,x)-\frac{\partial}{\partial q_i}V(\pi',x)\right)\\
&=\sum_{y\in\mathcal{X}}\left\{\gamma\left(V(\pi,y)-V(\pi',y)\right)\frac{\partial}{\partial q_i}f^{\pi}(x,y)+\gamma\left(f^{\pi}(x,y)-f^{\pi'}(x,y)\right)\frac{\partial}{\partial q_i}V(\pi,y)\right\}.
\end{aligned}
$$

したがって，

$$
\begin{aligned}
&\left(\frac{\partial}{\partial q_i}V(\pi,x)-\frac{\partial}{\partial q_i}V(\pi',x)\right)(1-\gamma f^{\pi}(x,x))-\sum_{y\neq x}\gamma f^{\pi'}(x,y)\left(\frac{\partial}{\partial q_i}V(\pi,x)-\frac{\partial}{\partial q_i}V(\pi',x)\right)\\
&=\sum_{y\in\mathcal{X}}\left\{\gamma\left(V(\pi,y)-V(\pi',y)\right)\frac{\partial}{\partial q_i}f^{\pi}(x,y)+\gamma\left(f^{\pi}(x,y)-f^{\pi'}(x,y)\right)\frac{\partial}{\partial q_i}V(\pi,y)\right\}.
\end{aligned}
$$

□

補題 4.10 を行列を用いて書くと評価がしやすくなる．

$\pi\in\Pi$ を一つ固定する．$V(\pi,\cdot)$ は価値関数の取る値を並べたベクトルとする．f^{π} は (X^{π}) の遷移確率行列を表すとする．$\partial q_i V(\pi,\cdot),\partial q_i f^{\pi}$ はそれぞれのベクトル・行列の要素を q_i 成分で偏微分したベクトル・行列を表す．I は単位行列を表すとする．

系 4.11. 補題 4.10 より，

$$
(I-\gamma f^{\pi})\left(\partial q_i V(\pi,\cdot)-\partial q_i V(\pi',\cdot)\right)^{\top}=\left(\partial q_i f^{\pi}\right)\gamma\left(V(\pi,\cdot)-V(\pi',\cdot)\right)^{\top}+\gamma\left(f^{\pi}-f^{\pi'}\right)\left(\partial q_i V(\pi,\cdot)\right)^{\top}
$$

である．

補題 4.1 より，$I-\gamma f^\pi$ の逆行列は存在するので，系 4.11 の両辺の左側からかける．

$$
\begin{aligned}
&(\partial q_i V(\pi,\cdot)-\partial q_i V(\pi',\cdot))\\
&=(I-\gamma f^\pi)^{-1}\left\{(\partial q_i f^\pi)\,\gamma\left(V(\pi,\cdot)-V(\pi',\cdot)\right)^\top+\gamma\left(f^\pi-f^{\pi'}\right)(\partial q_i V(\pi,\cdot))^\top\right\}.
\end{aligned}
$$

この両辺にノルムを適用する．現在，ノルムは両立しており，三角不等式を用いることで次の評価が得られる．

系 4.12.

$$
\begin{aligned}
&\|\partial q_i V(\pi,\cdot)-\partial q_i V(\pi',\cdot)\|\\
\le&\left\|(I-\gamma f^\pi)^{-1}\right\|\|\partial q_i f^\pi\|\,\gamma\,\|V(\pi,\cdot)-V(\pi',\cdot)\|+\left\|(I-\gamma f^\pi)^{-1}\right\|\gamma\left\|f^\pi-f^{\pi'}\right\|\|\partial q_i V(\pi,\cdot)\|.
\end{aligned}
\tag{3}
$$

上式の右辺のノルムを評価する．すなわち以下のノルムについて考える．

- $\|V(\pi,\cdot)-V(\pi',\cdot)\|$,
- $\|\partial q_i V(\pi,\cdot)\|$,
- $\left\|(I-\gamma f^\pi)^{-1}\right\|$,
- $\left\|f^\pi-f^{\pi'}\right\|$.

補題 4.13. $\left\|f^\pi-f^{\pi'}\right\|\le 6\|\pi-\pi'\|$.

証明. $q_4=1-q_1-q_2-q_3$ であることに注意すれば．三角不等式を用いる．

$$
\begin{aligned}
\frac{\left\|f^\pi-f^{\pi'}\right\|}{\|\pi-\pi'\|}&=\frac{\max_{i=1,\cdots,16}\sum_{j=1}^{16}\left|(f^\pi)_{ij}-\left(f^{\pi'}\right)_{ij}\right|}{\max_{i=1,2,3}|q_i-q_i'|}\\
&\le\frac{2|q_1-q_1'|+2|q_2-q_2'|+2|q_3-q_3'|}{\max_{i=1,2,3}|q_i-q_i'|}\\
&\le 2+2+2=6.
\end{aligned}
$$

□

補題 4.14. ある定数 M が存在して $\|V(\pi,\cdot)-V(\pi',\cdot)\|\le\frac{6M}{1-\gamma}\|\pi-\pi'\|$.

証明.

$$
\begin{aligned}
&V(\pi, x) - V(\pi', x) \\
=&\mathbb{E}^{\pi}\left[R(X_1^{\pi}) + \gamma V(\pi, X_1^{\pi}) \mid X_0 = x\right] - \mathbb{E}^{\pi'}\left[R(X_1^{\pi'}) + \gamma V(\pi', X_1^{\pi'}) \mid X_0 = x\right] \\
=&\sum_{y \in \mathcal{X}} \left(R(y) + \gamma V(\pi, y)\right) f^{\pi}(x, y) - \sum_{y \in \mathcal{X}} \left(R(y) + \gamma V(\pi', y)\right) f^{\pi'}(x, y) \\
=&\sum_{y \in \mathcal{X}} \left\{R(y) + \gamma V(\pi, y) - \gamma V(\pi', y) f^{\pi}(x, y) + \gamma V(\pi', y) f^{\pi}(x, y)\right\} f^{\pi}(x, y) - \sum_{y \in \mathcal{X}} \left(R(y) + \gamma V(\pi', y)\right) f^{\pi'}(x, y) \\
=&\sum_{y \in \mathcal{X}} \left\{R(y) f^{\pi}(x, y) + \gamma\left(V(\pi, y) - V(\pi', y)\right) f^{\pi}(x, y)\right\} - \sum_{y \in \mathcal{X}} \Big\{R(y) f^{\pi'}(x, y) + \gamma\left(f^{\pi}(x, y) - f^{\pi'}(x, y)\right) V(\pi', y \\
=&\sum_{y \in \mathcal{X}} \left\{\left(f^{\pi}(x, y) - f^{\pi'}(x, y)\right)\left(R(y) - \gamma V(\pi'), y\right) + \gamma\left(V(\pi, y) - V(\pi', y)\right) f^{\pi}(x, y)\right\}.
\end{aligned}
$$

したがって，

$$
\left(V(\pi, x) - V(\pi', x)\right) - \sum_{y \in \mathcal{X}} \gamma f^{\pi}(x, y)\left(V(\pi, y) - V(\pi', y)\right) = \sum_{y \in \mathcal{X}} \left(f^{\pi}(x, y) - f^{\pi'}(x, y)\right)\left(R(y) - \gamma V(\pi', y)\right).
$$

これを行列を使って記述すると次のようになる．

$$
\left(I - \gamma f^{\pi}\right)\left(V(\pi, \cdot) - V(\pi', \cdot)\right)^{\top} = \left(f^{\pi} - f^{\pi'}\right)\left(R + \gamma V(\pi', \cdot)\right).
$$

両辺に $(I - \gamma f^{\pi})$ の逆行列をかける．

$$
\left(V(\pi, \cdot) - V(\pi', \cdot)\right)^{\top} = \left(I - \gamma f^{\pi}\right)^{-1}\left(f^{\pi} - f^{\pi'}\right)\left(R + \gamma V(\pi', \cdot)\right).
$$

両辺にノルムを適用する．

$$
\begin{aligned}
&\left\|\left(V(\pi, \cdot) - V(\pi', \cdot)\right)\right\| \\
=&\left\|\left(I - \gamma f^{\pi}\right)^{-1}\left(f^{\pi} - f^{\pi'}\right)\left(R + \gamma V(\pi', \cdot)\right)\right\| \\
\leq&\left\|\left(I - \gamma f^{\pi}\right)^{-1}\right\|\left\|f^{\pi} - f^{\pi'}\right\|\left\|R + \gamma V(\pi', \cdot)\right\| \\
\leq&\left\|\left(I - \gamma f^{\pi}\right)^{-1}\right\| 6\left\|\pi - \pi'\right\|\left\|R + \gamma V(\pi', \cdot)\right\|.
\end{aligned}
$$

ここで，$\left\|\left(I - \gamma f^{\pi}\right)^{-1}\right\|$ について評価する．

$$
\begin{aligned}
\left\|\left(I - \gamma f^{\pi}\right)^{-1}\right\| &= \left\|\sum (\gamma f^{\pi})^{n}\right\| \leq \sum \left\|(\gamma f^{\pi})^{n}\right\| \\
&\leq \sum \left\|\gamma f^{\pi}\right\|^{n} = \sum \gamma^{n} = \frac{1}{1 - \gamma}.
\end{aligned}
$$

一方，R と V は有界であることが分かっているので，ある定数 M が存在して

$$
\left\|R + \gamma V(\pi', \cdot)\right\| \leq M
$$

とできる．以上により，

$$\begin{aligned}
&\|(V(\pi,\cdot)-V(\pi',\cdot))\| \\
\leq& \left\|(I-\gamma f^{\pi})^{-1}\right\| 6\|\pi-\pi'\|\|R+\gamma V(\pi',\cdot)\| \\
\leq& \frac{6M}{1-\gamma}\|\pi-\pi'\|.
\end{aligned}$$

□

補題 4.15. $\|\partial q_i V(\pi,\cdot)\|$ は一様に有界.

証明. 4.5 より，

$$\frac{\partial}{\partial q_i}V(\pi,x)=\sum_{y\in\mathcal{X}}\left\{(R(y)+\gamma V(\pi,y))\frac{\partial}{\partial q_i}f^{\pi}(x,y)+\gamma f^{\pi}(x,y)\frac{\partial}{\partial q_i}V(\pi,y)\right\}$$

である．したがって，

$$\frac{\partial}{\partial q_i}V(\pi,x)-\sum_{y\in\mathcal{X}}\gamma f^{\pi}(x,y)\frac{\partial}{\partial q_i}V(\pi,y)=\sum_{y\in\mathcal{X}}(R(y)+\gamma V(\pi,y))\frac{\partial}{\partial q_i}f^{\pi}(x,y).$$

これを行列を用いて表すと次のようになる．

$$(I-\gamma f^{\pi})(\partial q_i V(\pi,\cdot))^{\top}=(\partial q_i f^{\pi})(R+\gamma V(\pi,\cdot))^{\top}.$$

先ほどの計算と同様にすると次を得る．

$$(\partial q_i V(\pi,\cdot))^{\top}=(I-\gamma f^{\pi})^{-1}(\partial q_i f^{\pi})(R+\gamma V(\pi,\cdot))^{\top}.$$

両辺にノルムを適用する．

$$\begin{aligned}
\left\|(\partial q_i V(\pi,\cdot))^{\top}\right\| &= \left\|(I-\gamma f^{\pi})^{-1}(\partial q_i f^{\pi})(R+\gamma V(\pi,\cdot))^{\top}\right\| \\
&\leq \left\|(I-\gamma f^{\pi})^{-1}\right\|\|(\partial q_i f^{\pi})\|\|(R+\gamma V(\pi,\cdot))\| \\
&\leq \frac{1}{1-\gamma}16M \quad (M \text{ は定数}).
\end{aligned}$$

□

これらのノルムの評価を用いることで系 4.12 をさらに評価できる．

系 4.16.

$$\|\partial q_i V(\pi,\cdot)-\partial q_i V(\pi',\cdot)\|\leq\frac{192\gamma M}{1-\gamma}\|\pi-\pi'\|.$$

命題 4.17. $\nabla_{\pi}V(\pi,x)$ で V の π に関する勾配を表すとする．すなわち，

$$\nabla_{\pi}V(\pi,x)=\left(\frac{\partial}{\partial q_1}V(\pi,x),\quad\frac{\partial}{\partial q_2}V(\pi,x),\quad\frac{\partial}{\partial q_3}V(\pi,x)\right).$$

とする．$\nabla_\pi V(\pi, x)$ は π に関してリプシッツ連続である．即ち，任意の $\pi, \pi' \in \Pi$ に対してある定数 L が存在して

$$\|\nabla_\pi V(\pi, x) - \nabla_\pi V(\pi', x)\| \leq L \|\pi - \pi'\|$$

が成立する．

証明． 系 4.16. □

補題 4.18.

$$\|\nabla_\pi V(\pi, \cdot)\|^2 \leq C (1 + \|V(\pi, \cdot)\|)$$

となる定数 C が存在する．

証明． 補題 4.15 より $\|\nabla_\pi V(\pi, \cdot)\|$ は有界なので，主張の不等式を成立させるような十分大きな定数 C を取ることができる． □

4.5 目的

方策 π をパラメータとして，価値関数を最大化することを考えたい．つまり，

$$\min_{\pi \in \Pi} (-V(\pi, \cdot))$$

を調べたい．

まずは議論のために必要な Robbins Monro 条件 について記す．

定義 4.19. 実数列 $\{\alpha_n\}$ が RM 条件を満たすとは，次を満たすこと．

$$\sum \alpha_n = +\infty, \quad \sum \alpha_n^2 < \infty$$

例えば，数列 $\{\frac{1}{n}\}$ はこの条件を満たす．

以下，数列 $\{\alpha_n\}$ を RM 条件を満たす実数列とする．

4.6 π の探索法:勾配降下法

目的のために用いる手法は勾配降下法の計算を利用する．すなわち，大雑把に言えば価値関数の微分を計算し極値を求めることで，価値関数を最大にする方策 π を求める．この節ではその具体的な計算法について，書き記す．数学としての理論的な方法と，実際に計算機を用いてシミュレーションする際の方法を述べていく．

4.6.1 アルゴリズム:数学的な記述

最初に任意の方策 π を選ぶことで帰納的に更新されていく．以下に手順を記すが，その中にはいくつか突飛と思えるような記号の定義が登場する．その根拠については手順を述べたあとに記述する．

π_n の更新アルゴリズム

$n \geq 0$ とする．$\pi_0 \in \Pi$ を任意に一つ選ぶ．ただし，以下ではである．

手順 1.　$\pi_n \in \Pi$ に対し，遷移関数を f^{π_n} に持ち，初期状態 $X_n^{\pi_n} := X_n^{\pi_{n-1}}$ とするマルコフ連鎖 $(X_t^{\pi_n})_{t \geq n}$ を考えることができる．ただし，$n = 0$ の場合，初期状態は $X_0^{\pi_0}$ である．

手順 2.　選んだ π_n に対して次のような遷移関数 g^{π_n} と，それに対応するマルコフ連鎖 $(Y_t^{\pi_n})_{t \geq n}$ を考える．

$$
\begin{aligned}
&0 \leq g^{\pi_n}(x, y) \leq 1 \quad (\forall x, y \in \mathcal{X}), \\
&\sum_{y \in \mathcal{X}} g^{\pi_n}(x, y) = 1 \quad (\forall x \in \mathcal{X}), \\
&Y_n^{\pi_n} := X_n^{\pi_{n-1}} \quad (\text{ただし}, Y_0^{\pi_0} = X_0^{\pi_0}), \\
&\sum_{y \in \mathcal{X}} g^{\pi_n}(Y_n^{\pi_n}, y) \Big\{ -V(\pi_n, y) + V(\pi_n, Y_n^{\pi_n}) \Big\} \leq 0.
\end{aligned}
$$

手順 3.　$G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_n})$ を次のように定義する．

$$
\begin{aligned}
&G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_n}) \\
&= \frac{\left(R(Y_1^{\pi_n}) + \gamma V(\pi_n, Y_{n+1}^{\pi_n})\right) \nabla_\pi f^{\pi_n}(Y_n^{\pi_n}, Y_{n+1}^{\pi_n}) + \gamma f^{\pi_n}(Y_n^{\pi_n}, Y_{n+1}^{\pi_n}) \nabla_\pi V(\pi_n, Y_{n+1}^{\pi_n})}{g^{\pi_n}(Y_n^{\pi_n}, Y_{n+1}^{\pi_n})}.
\end{aligned}
$$

ただし，$g^{\pi_n}(Y_n^{\pi_n}, Y_{n+1}^{\pi_n}) = 0$ のとき，

$$
G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_n}) = 0
$$

と定義する．

手順 4.　次の式にしたがって，π_{n+1} を得る．

$$
\pi_{n+1} = \pi_n - \alpha_{n+1} G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_n}).
$$

手順 5.　π_{n+1} に対して手順 1. から適用する．

さて，ここからは先ほど述べた手順の中で登場した，新たな記号の定義の正当性や根拠を説明する．

手順 2. について　g^{π_n} は具体的には次のような例が存在する．

$$
g^{\pi_n}(Y_n^{\pi_n}, y) = \begin{cases} 0 & (\text{if} \quad y \in \{z \mid -V(\pi_n, z) + V(\pi_n, Y_n^{\pi_n}) > 0\}) \\ p_y (> 0) & (\text{if} \quad y \in \{z \mid -V(\pi_n, z) + V(\pi_n, Y_n^{\pi_n}) \leq 0\} =: A_n). \end{cases}
$$

ただし，$\sum_{y \in \mathcal{X}} g^{\pi_n}(x, y) = \sum_{y \in A_n} p_y = 1$ であるとする．

手順 3. について　補題 4.5 より次のように計算する．

$$
\begin{aligned}
&\nabla_\pi V(\pi_n, Y_n^{\pi_n}) \\
=&\nabla_\pi \mathbb{E}\left[R(X_1^{\pi_n}) + \gamma V(\pi_n, X_1^{\pi_n}) \mid X_0^{\pi_n} = Y_n^{\pi_n}\right] \\
=&\sum_{y \in \mathcal{X}} \left\{ \left(R(y) + \gamma V(\pi_n, y)\right) \nabla_\pi f^{\pi_n}(Y_n^{\pi_n}, y) + \gamma f^{\pi_n}(Y_n^{\pi_n}, y) \nabla_\pi V(\pi_n, y) \right\} \\
=&\sum_{y \in \mathcal{X}} \frac{g^{\pi_n}(Y_n^{\pi_n}, y)}{g^{\pi_n}(Y_n^{\pi_n}, y)} \left\{ \left(R(y) + \gamma V(\pi_n, y)\right) \nabla_\pi f^{\pi_n}(Y_n^{\pi_n}, y) + \gamma f^{\pi_n}(Y_n^{\pi_n}, y) \nabla_\pi V(\pi_n, y) \right\} \\
=&\mathbb{E}^{g^{\pi_n}}\left[\frac{\left(R(Y_{n+1}^{\pi_n}) + \gamma V(\pi_n, Y_{n+1}^{\pi_n})\right) \nabla_\pi f^{\pi_n}(Y_n^{\pi_n}, Y_{n+1}^{\pi_n}) + \gamma f^{\pi_n}(Y_n^{\pi_n}, Y_{n+1}^{\pi_n}) \nabla_\pi V(\pi_n, Y_{n+1}^{\pi_n})}{g^{\pi_n}(Y_n^{\pi_n}, Y_{n+1}^{\pi_n})}. \right]
\end{aligned}
$$

ここで，$\mathbb{E}^{g^{\pi_n}}$ は確率測度 $g^{\pi_n}(Y_n^{\pi_n}, \cdot)$ よる期待値を表すとする．さて，この期待値の中身を G と置くことにする．

$$
\begin{aligned}
&G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_n}) \\
&= \frac{\left(R(Y_{n+1}^{\pi_n}) + \gamma V(\pi_n, Y_{n+1}^{\pi_n})\right) \nabla_\pi f^{\pi_n}(Y_n^{\pi_n}, Y_{n+1}^{\pi_n}) + \gamma f^{\pi_n}(Y_n^{\pi_n}, Y_{n+1}^{\pi_n}) \nabla_\pi V(\pi_n, Y_{n+1}^{\pi_n})}{g^{\pi_n}(Y_n^{\pi_n}, Y_{n+1}^{\pi_n})}.
\end{aligned}
$$

手順 4. について　ここが勾配降下法の計算を行っている部分．

方策の更新のアルゴリズムの時に新しくマルコフ連鎖を定義したが，図に表すと次のような状況になっている．

$$
\begin{tikzcd}
X_0^{\pi_0} \arrow[r, "f^{\pi_0}"] \arrow[dr, "f^{\pi_0}"'] & X_1^{\pi_0} \arrow[r, "f^{\pi_0}"] \arrow[d, dash] & X_2^{\pi_0} \arrow[r, "f^{\pi_0}"] & X_3^{\pi_0} \arrow[r, "f^{\pi_0}"] & \cdots \\
& X_1^{\pi_1} \arrow[r, "f^{\pi_1}"] \arrow[dr, "f^{\pi_1}"'] & X_2^{\pi_1} \arrow[r, "f^{\pi_1}"] \arrow[d, dash] & X_3^{\pi_1} \arrow[r, "f^{\pi_1}"] & \cdots \\
& & X_2^{\pi_2} \arrow[r, "f^{\pi_2}"] \arrow[dr, "f^{\pi_2}"'] & X_3^{\pi_2} \arrow[r, "f^{\pi_2}"] \arrow[d, dash] & \cdots \\
& & & X_3^{\pi_3} \arrow[r, "f^{\pi_3}"] \arrow[dr, "f^{\pi_3}"'] & \cdots \\
& & & & \ddots
\end{tikzcd}
$$

$$
\begin{tikzcd}
Y_0^{\pi_0} \arrow[r, "g^{\pi_0}"] \arrow[dr, "g^{\pi_0}"'] & Y_1^{\pi_0} \arrow[r, "g^{\pi_0}"] \arrow[d, dash] & Y_2^{\pi_0} \arrow[r, "g^{\pi_0}"] & Y_3^{\pi_0} \arrow[r, "g^{\pi_0}"] & \cdots \\
& Y_1^{\pi_1} \arrow[r, "g^{\pi_1}"] \arrow[dr, "g^{\pi_1}"'] & Y_2^{\pi_1} \arrow[r, "g^{\pi_1}"] \arrow[d, dash] & Y_3^{\pi_1} \arrow[r, "g^{\pi_1}"] & \cdots \\
& & Y_2^{\pi_2} \arrow[r, "g^{\pi_2}"] \arrow[dr, "g^{\pi_2}"'] & Y_3^{\pi_2} \arrow[r, "g^{\pi_2}"] \arrow[d, dash] & \cdots \\
& & & Y_3^{\pi_3} \arrow[r, "g^{\pi_3}"] \arrow[dr, "g^{\pi_3}"'] & \cdots \\
& & & & \ddots
\end{tikzcd}
$$

4.6.2　アルゴリズム:計算機を用いる場合

基本的には数学的な手順と同様である．異なる部分は g の設定である．モンテカルロ法などの技法を使い，価値関数の値をシミュレーションを行う．その標本平均を V の値として用いる．すなわち，$V(\pi, X_0^{\pi_0})$ の値を知るために，M 回のシミュレーションを行う．この標本平均を $\bar{V}(\pi, X_0^{\pi_0})$ とする．この標本平均を用いて g^{π} を定義する．

4.7 アルゴリズムが収束することの証明

前節で方策の更新方法を設定したが，この更新式を十分な回数行うと収束することを示せばよい．即ち，今回のモデルに対しては次を証明すればよいことになる．この定理は Gilles Pagés [4, pp.180-181] の結果を自分のモデルに適用したものである．

定理 4.20. $\pi_0 \in \Pi$ を任意に固定する．マルコフ連鎖 $(X_t^{\pi_0})_{t\geq 0}$ がある．このマルコフ連鎖の遷移確率関数は f^{π_0} であるとする．
任意の $\pi \in \Pi$ について次のような条件を満たす遷移確率関数 g^π とマルコフ連鎖 $(Y_t^\pi)_{t\geq 0}$ が存在すると仮定する．

$$
\begin{aligned}
&0 \leq g^\pi(x,y) \leq 1 \quad (\forall x, y \in \mathcal{X}),\\
&\sum_{y\in\mathcal{X}} g^\pi(x,y) = 1 \quad (\forall x \in \mathcal{X}),\\
&\sum_{y\in\mathcal{X}} g^\pi(Y_n^\pi, y)\Big\{-V(\pi,y) + V(\pi, Y_n^\pi)\Big\} \leq 0 \quad (\forall n \in \mathbb{N}).
\end{aligned}
$$

4.6 節のアルゴリズムに従って列 (π_n) を得る．

$$\pi_{n+1} = \pi_n - \alpha_{n+1} G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_n}).$$

さらに，以下の三つの条件が満たされているとする．

1. $\nabla_\pi V$ はリプシッツ連続
2. $\|\nabla_\pi V(\pi,\cdot)\|^2 \leq C\left(1 + \|V(\pi,\cdot)\|\right)$
3. $\sum \alpha_n = \infty, \quad \sum \alpha_n^2 < \infty$

このとき，次が成り立つ．

1. $\pi_{n+1} - \pi_n$ は概収束し，かつ，1 次平均収束する．
2. 関数列 $\left(V(\pi_n,\cdot)\right)_n$ は L^1-有界で，概収束する．

定理 4.20 の証明. $(\mathcal{G}_n)$ をマルコフ連鎖 $(Y_n^{\pi_n}, g^{\pi_n})$ に関する natural filtration とする．条件付き平均 $\mathbb{E}\left[V(\pi_{n+1}, Y_{n+1}^{\pi_{n+1}}) - V(\pi_n, Y)_n^{\pi_n} \mid \mathcal{G}_n\right]$ について考える．

$$
\begin{aligned}
&\mathbb{E}\left[V(\pi_{n+1}, Y_{n+1}^{\pi_{n+1}}) - V(\pi_n, Y_n^{\pi_n}) \mid \mathcal{G}_n\right]\\
=&\mathbb{E}\left[V(\pi_{n+1}, Y_{n+1}^{\pi_{n+1}}) - V(\pi_{n+1}, Y_n^{\pi_n}) + V(\pi_{n+1}, Y_n^{\pi_n}) - V(\pi_n, Y_n^{\pi_n}) \mid \mathcal{G}_n\right]\\
=&\mathbb{E}\left[V(\pi_{n+1}, Y_{n+1}^{\pi_{n+1}}) - V(\pi_{n+1}, Y_n^{\pi_n}) \mid \mathcal{G}_n\right] + \mathbb{E}\left[V(\pi_{n+1}, Y_n^{\pi_n}) - V(\pi_n, Y_n^{\pi_n}) \mid \mathcal{G}_n\right].
\end{aligned}
$$

そこで，二つの条件付き平均

1. $\mathbb{E}\left[V(\pi_{n+1}, Y_{n+1}^{\pi_{n+1}}) - V(\pi_{n+1}, Y_n^{\pi_n}) \mid \mathcal{G}_n\right]$

2. $\mathbb{E}\left[V(\pi_{n+1}, Y_n^{\pi_n}) - V(\pi_n, Y_n^{\pi_n}) \mid \mathcal{G}_n\right]$

を評価する．

1. について． $Y_{n+1}^{\pi_{n+1}} = Y_{n+1}^{\pi_n}$ であることに注意する．

$$
\begin{aligned}
&\mathbb{E}\left[V(\pi_{n+1}, Y_{n+1}^{\pi_{n+1}}) - V(\pi_{n+1}, Y_n^{\pi_n}) \mid \mathcal{G}_n\right] \\
=&\mathbb{E}\left[V\left(\pi_n - \alpha_{n+1}G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_n}), Y_{n+1}^{\pi_n}\right) - V\left(\pi_n - \alpha_{n+1}G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_n}), Y_n^{\pi_n}\right) \mid \mathcal{G}_n\right] \\
=&\sum_{y\in\mathcal{X}} \left\{V\left(\pi_n - \alpha_{n+1}G(\pi_n, Y_n^{\pi_n}, y), y\right) - V\left(\pi_n - \alpha_{n+1}G(\pi_n, Y_n^{\pi_n}, y), Y_n^{\pi_n}\right)\right\} g^{\pi_n}(Y_n^{\pi_n}, y) \\
\leq&0.
\end{aligned}
$$

2. について．

$$
\begin{aligned}
&\mathbb{E}\left[V(\pi_{n+1}, Y_n^{\pi_n}) - V(\pi_n, Y_n^{\pi_n}) \mid \mathcal{G}_n\right] \\
=&\sum_{y\in\mathcal{X}} \left\{V\left(\pi_n - \alpha_{n+1}G(\pi_n, Y_n^{\pi_n}, y), Y_n^{\pi_n}\right) - V(\pi_n, Y_n^{\pi_n})\right\} g^{\pi_n}(Y_n^{\pi_n}, y)
\end{aligned}
$$

である．そこで，中かっこ $\{\quad\}$ の中を評価する．$V\left(\pi_n - \alpha_{n+1}G(\pi_n, Y_n^{\pi_n}, y), Y_n^{\pi_n}\right) = V\left(\pi_{n+1}, Y_n^{\pi_n}\right)$ であるから，これをテイラーの定理を用いて展開する．

$$
\begin{aligned}
&V\left(\pi_{n+1}, Y_n^{\pi_n}\right) \\
=&V\left(\pi_n, Y_n^{\pi_n}\right) + \langle \nabla_\pi V\left(\pi, Y_n^{\pi_n}\right), \pi_{n+1} - \pi_n\rangle \quad \text{for some} \quad \pi \in (\pi_n, \pi_{n+1}) \\
=&V\left(\pi_n, Y_n^{\pi_n}\right) + \langle \nabla_\pi V\left(\pi, Y_n^{\pi_n}\right) + \nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right) - \nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right), \pi_{n+1} - \pi_n\rangle \\
=&V\left(\pi_n, Y_n^{\pi_n}\right) + \langle \nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right), \pi_{n+1} - \pi_n\rangle \\
&\qquad + \langle \nabla_\pi V\left(\pi, Y_n^{\pi_n}\right) - \nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right), \pi_{n+1} - \pi_n\rangle \\
=&V\left(\pi_n, Y_n^{\pi_n}\right) - \alpha_{n+1}\langle \nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right), G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_n+1})\rangle \\
&\qquad + \langle \nabla_\pi V\left(\pi, Y_n^{\pi_n}\right) - \nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right), \pi_{n+1} - \pi_n\rangle
\end{aligned}
$$

命題 4.17 より，

$$
\begin{aligned}
\leq&V\left(\pi_n, Y_n^{\pi_n}\right) - \alpha_{n+1}\langle \nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right), G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_n+1})\rangle + L\left\|\pi_{n+1} - \pi_n\right\|^2 \\
=&V\left(\pi_n, Y_n^{\pi_n}\right) - \alpha_{n+1}\langle \nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right), G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_n+1})\rangle \\
&\qquad + L\alpha_{n+1}^2\left\|G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_n+1})\right\|^2.
\end{aligned}
$$

以上により，次の評価を得る．

$$
\begin{aligned}
&\mathbb{E}\left[V(\pi_{n+1}, Y_{n+1}^{\pi_{n+1}}) - V(\pi_n, Y_n^{\pi_n}) \mid \mathcal{G}_n\right] \leq 0 + \mathbb{E}\left[V(\pi_{n+1}, Y_n^{\pi_n}) - V(\pi_n, Y_n^{\pi_n}) \mid \mathcal{G}_n\right] \\
\Leftrightarrow& \\
&\mathbb{E}\left[V(\pi_{n+1}, Y_{n+1}^{\pi_{n+1}}) \mid \mathcal{G}_n\right] \\
\leq&V(\pi_n, Y_n^{\pi_n}) + \mathbb{E}\left[V(\pi_{n+1}, Y_n^{\pi_n}) - V(\pi_n, Y_n^{\pi_n}) \mid \mathcal{G}_n\right] \\
\leq&V(\pi_n, Y_n^{\pi_n}) + \mathbb{E}\Big[-\alpha_{n+1}\langle \nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right), G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_n+1})\rangle + L\alpha_{n+1}^2\left\|G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_n+1})\right\|^2 \mid \mathcal{G}_n\Big].
\end{aligned}
$$

さらなる評価のために次の二つを評価しよう．

- $\mathbb{E}\left[-\alpha_{n+1}\langle\nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right), G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_n+1})\rangle \mid \mathcal{G}_n\right]$
- $\mathbb{E}\left[L\alpha_{n+1}^2\left\|G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_{n+1}})\right\|^2 \mid \mathcal{G}_n\right]$

1. について

$$\begin{aligned}
&\mathbb{E}\left[-\alpha_{n+1}\langle\nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right), G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_n+1})\rangle \mid \mathcal{G}_n\right] \\
=&\sum_{y\in\mathcal{X}} -\alpha_{n+1}\langle\nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right), G(\pi_n, Y_n^{\pi_n}, y)\rangle g^{\pi_n}(Y_n^{\pi_n}, y) \\
=&-\alpha_{n+1}\langle\nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right), \sum_{y\in\mathcal{X}} G(\pi_n, Y_n^{\pi_n}, y) g^{\pi_n}(Y_n^{\pi_n}, y)\rangle \\
=&-\alpha_{n+1}\langle\nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right), \nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right)\rangle \\
\leq&-\alpha_{n+1}\left\|\nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right)\right\|^2.
\end{aligned}$$

2. について　補題 4.18 より

$$\mathbb{E}\left[L\alpha_{n+1}^2\left\|G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_{n+1}})\right\|^2 \mid \mathcal{G}_n\right] \leq C(1+V(\pi_n, Y_n^{\pi_n}))L\alpha_{n+1}^2.$$

以上により，最終的に次を得る．

$$\begin{aligned}
\mathbb{E}\left[V(\pi_{n+1}, Y_{n+1}^{\pi_{n+1}}) \mid \mathcal{G}_n\right] &\leq V(\pi_n, Y_n^{\pi_n}) - \alpha_{n+1}\left\|\nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right)\right\|^2 + \tilde{C}(1+V(\pi_n, Y_n^{\pi_n}))\alpha_{n+1}^2 \\
&= V(\pi_n, Y_n^{\pi_n})\left(1+\tilde{C}\alpha_{n+1}^2\right) - \alpha_{n+1}\left\|\nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right)\right\|^2 + \tilde{C}\alpha_{n+1}^2.
\end{aligned}$$

ただし，ここで $\tilde{C} := CL$ とおいている．そこで

$$\begin{cases} S_n := V(\pi_n, Y_n^{\pi_n}) + \sum_{i=i}^n \alpha_i \left\|\nabla_\pi V\left(\pi_{i-1}, Y_{i-1}^{\pi_{i-1}}\right)\right\| + \tilde{C}\sum_{k\geq n+1}\alpha_k^2 \\ \beta_n := \prod_{k=1}^n\left(1+\tilde{C}\alpha_k^2\right) \end{cases}.$$

とおくと，$\left(\frac{S_n}{\beta_n}\right)$ が非負の優マルチンゲールであることが分かる．実際，$\mathbb{E}\left[S_n \mid \mathcal{G}_n\right]$ を次のように

評価できる．

$$
\begin{aligned}
&\mathbb{E}\left[S_n \mid \mathcal{G}_n\right] \\
=&\mathbb{E}\left[V(\pi_{n+1}, Y_n^{\pi_{n+1}}) + \sum_{i=i}^{n+1} \alpha_i \left\|\nabla_\pi V\left(\pi_{i-1}, Y_{i-1}^{\pi_{i-1}}\right)\right\| + \tilde{C} \sum_{k \geq n+2} \alpha_k^2 \mid \mathcal{G}_n\right] \\
=&\mathbb{E}\left[V(\pi_{n+1}, Y_n^{\pi_{n+1}}) \mid \mathcal{G}_n\right] + \sum_{i=i}^{n+1} \alpha_i \left\|\nabla_\pi V\left(\pi_{i-1}, Y_{i-1}^{\pi_{i-1}}\right)\right\| + \tilde{C} \sum_{k \geq n+2} \alpha_k^2 \\
\leq& V(\pi_n, Y_n^{\pi_n}) - \alpha_{n+1} \left\|\nabla_\pi V\left(\pi_n, Y_n^{\pi_n}\right)\right\|^2 + \tilde{C}(1 + V(\pi_n, Y_n^{\pi_n}))\alpha_{n+1}^2 \\
&\quad + \sum_{i=i}^{n+1} \alpha_i \left\|\nabla_\pi V\left(\pi_{i-1}, Y_{i-1}^{\pi_{i-1}}\right)\right\| + \tilde{C} \sum_{k \geq n+2} \alpha_k^2 \\
=& V(\pi_n, Y_n^{\pi_n})\left(1 + \tilde{C}\alpha_{n+1}^2\right) + \sum_{i=i}^{n} \alpha_i \left\|\nabla_\pi V\left(\pi_{i-1}, Y_{i-1}^{\pi_{i-1}}\right)\right\| + \tilde{C} \sum_{k \geq n+1} \alpha_k^2 \\
\leq& \left(V(\pi_n, Y_n^{\pi_n}) + \sum_{i=i}^{n} \alpha_i \left\|\nabla_\pi V\left(\pi_{i-1}, Y_{i-1}^{\pi_{i-1}}\right)\right\| + \tilde{C} \sum_{k \geq n+1} \alpha_k^2\right)\left(1 + \tilde{C}\alpha_{n+1}^2\right) \\
=& S_n\left(1 + \tilde{C}\alpha_{n+1}^2\right).
\end{aligned}
$$

ゆえに，

$$
\mathbb{E}\left[S_{n+1} \mid \mathcal{G}_n\right] \leq S_n\left(1 + \tilde{C}\alpha_{n+1}^2\right).
$$

よって，

$$
\mathbb{E}\left[\frac{1}{\left(1 + \tilde{C}\alpha_{n+1}^2\right)} \frac{S_{n+1}}{\beta_n} \mid \mathcal{G}_n\right] = \mathbb{E}\left[\frac{S_{n+1}}{\beta_{n+1}} \mid \mathcal{G}_n\right] \leq \frac{S_n}{\beta_n}.
$$

ゆえに，$\frac{S_n}{\pi_n}$ は L^1-収束し，概収束する．$\lim_{n\to\infty} \pi_n$ は存在するので，S_∞ も L^1-収束，概収束両方の意味で存在する．

$$
S_\infty = \lim_{n\to\infty}\left(V(\pi_n, Y_n^{\pi_n}) + \sum_{i=i}^{n} \alpha_i \left\|\nabla_\pi V\left(\pi_{i-1}, Y_{i-1}^{\pi_{i-1}}\right)\right\| + \tilde{C} \sum_{k \geq n+1} \alpha_k^2\right).
$$

また，$\pi_n > 1$ であるから，

$$
\sum_{i=i}^{n} \alpha_i \left\|\nabla_\pi V\left(\pi_{i-1}, Y_{i-1}^{\pi_{i-1}}\right)\right\| \leq S_n \leq S_n \pi_n
$$

である．ゆえに，

$$
\sum_{i=i}^{\infty} \alpha_i \left\|\nabla_\pi V\left(\pi_{i-1}, Y_{i-1}^{\pi_{i-1}}\right)\right\| \leq \mathbb{E}\left[S_\infty \pi_\infty\right] < \infty
$$

である．よって $\lim_{n\to\infty} V(\pi_n, Y_n^{\pi_n})$ も概収束する．

一方で，

$$\|\pi_{n+1} - \pi_n\|^2 \leq \alpha_{n+1}^2 \left\|G(\pi_n, Y_n^{\pi_n}, Y_{n+1}^{\pi_n})\right\|^2$$

なので，

$$\sum_n \mathbb{E}\left[\|\pi_{n+1} - \pi_n\|^2\right] \leq \sum_n \alpha_{n+1}^2 \tilde{C}\left(1 + V(\pi_n, Y_n^{\pi_n})\right) < \infty$$

である．したがって，

$$\lim_{n\to\infty} (\pi_{n+1} - \pi_n) = 0$$

である． □

5　結論と展望

5.1　結論

マルコフ連鎖の具体的な構成を行い，MRP の数学的な実例を一つ提示できた．また，マルコフ連鎖に勾配降下法を適用することを利用して，価値関数を最大にすると考えられる方策の入手のアルゴリズムを提示できた．

5.2　展望

価値関数を最大にすると考えられる方策の入手のアルゴリズムが実際に最適なものであるかの証明が残っている．つまり，定理 4.20 で得られる $\lim_{n\to\infty}\pi_n$ は $\max_{\pi\in\Pi} V(\pi, x)$ を実現するものであるかという証明が残っている．

6 Appendix

証明を飛ばした命題や一般化した命題などの証明.

6.1 $\mathcal{B}[0,1] = \sigma(\pi([0,1]))$ について

補題 6.1. $\mathcal{I} := \{[0,1] \cap (-\infty, a) \mid a \in \mathbb{R}\}$ とする．このとき，

$$\mathcal{B}[0,1] = \sigma(\mathcal{I})$$

が成立する．

証明. $\mathcal{I} := \{[0,1] \cap (-\infty, a) \mid a \in \mathbb{R}\}$ とする．このとき，$\mathcal{I}$ は π-system である．実際，$A, B \in \mathcal{I}$ を取ると，$\mathcal{I}$ の定義より，ある $a, b \in \mathbb{R}$ について

$$\begin{aligned} A &= [0,1] \cap (-\infty, a), \\ B &= [0,1] \cap (-\infty, b), \end{aligned}$$

と書ける．したがって，

$$A \cap B = [0,1] \cap (-\infty, \min\{a, b\}) \in \mathcal{I}.$$

次に，$\mathcal{B}[0,1] = \sigma(\mathcal{I})$ である．両辺は σ-alg. だから，

$$(\mathbb{R} \text{ の部分位相空間 } [0,1] \text{ 上の開集合全体}) \subset \sigma(\mathcal{I}),$$

および，

$$\mathcal{I} \subset \mathcal{B}[0,1],$$

だけを示せばよい．前者について : U を $[0,1]$ の開集合とすると，相対位相の定義より，ある $\mathbb{R}$ の開集合 V が存在して

$$U = [0,1] \cap V$$

と書ける．$\mathbb{R}$ の開基は開区間全体なので，ある高々可算な集合 Λ が存在して，

$$V = \bigcup_{\lambda \in \Lambda} V_\lambda$$

と書ける．ただし，V_λ は $\mathbb{R}$ の開区間である．よって，

$$U = \bigcup_{\lambda \in \Lambda} ([0,1] \cap V_\lambda)$$

である．ここで，任意の $\mathbb{R}$ 上の開区間 (a, b) について $[0,1] \cap (a, b) \in \sigma(\mathcal{I})$ である．このことは場合分けをして確認する．以下のような五つの場合に分けられる．

- $b \leq 0$ または $1 \leq a$ のとき

- $a \leq 0 < b \leq 1$ のとき
- $0 < a \leq 1 < b$ のとき
- $0 < a < b < 1$ のとき
- $a \leq 0$ かつ $1 \leq b$ のとき

($b \leq 0$ または $1 \leq a$ のとき) $[0,1] \cap (a,b) = \emptyset \in \sigma(\mathcal{I})$ である．
($a \leq 0 < b \leq 1$ のとき)

$$\begin{aligned} [0,1] \cap (a,b) &= [0,b) \\ &= [0,1] \cap (-\infty, b) \\ &= [0,1] \cap \left(\bigcup_{n \in \mathbb{N}} \left(-\infty, b - \frac{1}{n} \right] \right) \\ &= \bigcup_{n \in \mathbb{N}} \left([0,1] \cap \left(-\infty, b - \frac{1}{n} \right] \right) \in \sigma(\mathcal{I}) \end{aligned}$$

である．
($0 < a \leq 1 < b$ のとき) $[0,1] \cap (a,b) = (a,1]$ である．ところで，$[0,1] \cap (-\infty, a] \in \mathcal{I}$ なので，

$$([0,1] \cap (-\infty, a])^c \in \sigma(\mathcal{I})$$

である．今の a の値を考えると，上式は

$$(-\infty, 0) \cup (a, \infty) \in \sigma(\mathcal{I})$$

と書ける．また，$[0,1] \in \mathcal{I}$ だから[*3]，

$$[0,1] \cap \big((-\infty, 0) \cup (a, \infty) \big) \in \sigma(\mathcal{I})$$

であり，これは

$$\begin{aligned} [0,1] \cap \big((-\infty, 0) \cup (a, \infty) \big) &= \big([0,1] \cap (-\infty, 0) \big) \cup \big([0,1] \cap (a, \infty) \big) \\ &= \emptyset \cup \big([0,1] \cap (a, \infty) \big) \\ &= (a, 1] \end{aligned}$$

となるので，この場合も正しい．
($0 < a < b < 1$ のとき)

$$\begin{aligned} [0,1] \cap (a,b) &= (a,b) \\ &= (a,1] \cap [0,b) \end{aligned}$$

であるから，前二つの場合分けよりこの場合も正しい．
($a \leq 0$ かつ $1 \leq b$ のとき) $[0,1] \cap (a,b) = [0,1]$ より正しい．

---

*3 $[0,1] \cap (-\infty, 2]$ などを考える．

以上により，

$$U = \bigcup_{\lambda \in \Lambda} ([0,1] \cap V_\lambda) \in \sigma(\mathcal{I})$$

である．

後者について：$\mathcal{I}$ の元を任意に一つとる．定義よりこれは

$$[0,1] \cap (-\infty, a]$$

と書ける．a の値によってこの集合は次のように変わる．

$$[0,1] \cap (-\infty, a] = \begin{cases} \emptyset & (a < 0) \\ [0,a] & (0 \leq a < 1) \\ [0,1] & (1 \leq a) \end{cases}.$$

$\emptyset \in \mathcal{B}[0,1]$ である．また，$\varepsilon = \frac{1-a}{2}$ とおけば，

$$[0,a] = \bigcap_{n \in \mathbb{N}} \left([0,1] \cap \left(-1, a + \frac{\varepsilon}{n} \right) \right)$$

であるから，$[0,a] \in \mathcal{B}[0,1]$ である．そして，$[0,1] = [0,1] \cap (-1,2)$ なので[*4]，$[0,1] \in \mathcal{B}[0,1]$ である．

以上により

$$\mathcal{B}[0,1] = \sigma(\mathcal{I})$$

である． □

6.2 $I - \gamma f^\pi$ の逆行列について

補題を再掲する．

補題． $I - \gamma f^\pi$ には逆行列が存在し，それは次の行列の等比級数である．

$$(I - \gamma f^\pi)^{-1} = \sum_{n \geq 0} (\gamma f^\pi)^n.$$

証明は二段階に分けて行う．

まず，無限級数 $\displaystyle\sum_{n \geq 0} (\gamma f^\pi)^n$ が存在すると仮定する．これが $I - \gamma f^\pi$ の逆行列になっていることを示す．

$$\begin{aligned} (I - \gamma f^\pi) \sum_{n \geq 0} (\gamma f^\pi)^n &= \sum_{n \geq 0}^{\infty} (\gamma f^\pi)^n - \gamma f^\pi \sum_{n \geq 0}^{\infty} (\gamma f^\pi)^n \\ &= (\gamma f^\pi)^0 + \sum_{n \geq 1}^{\infty} (\gamma f^\pi)^n - \sum_{n \geq 1}^{\infty} (\gamma f^\pi)^n \\ &= I. \end{aligned}$$

*4 $[0,1]$ の開集合である．

同様に，

$$\sum_{n\geq 0}(\gamma f^{\pi})^{n}(I-\gamma f^{\pi})=I.$$

次に，無限級数 $\sum_{n\geq 0}(\gamma f^{\pi})^{n}$ が存在することを示す．そのためにいくつかの線形代数の命題を利用する．齋藤 [6,pp207, 定理 2.1,pp217, 定理 3.1] を参考にした．

定理 6.2. 実数列のべき級数 $\sum_{k=0}^{\infty}a_k x^k$ の収束半径を ρ とする．行列 X の特性根 (固有多項式の根) の絶対値が全て ρ より小さければ，行列 X のべき級数 $\sum_{k=0}^{\infty}a_k X^k$ は存在する．そうでなければ，この行列のべき級数は発散する．

実数列のべき級数 $\sum_{n\geq 0}\gamma^n x^n$ の収束半径が $|x|\leq 1$ である．したがって定理 6.2 より，$\sum_{n\geq 0}\gamma^n (f^{\pi})^n$ が収束するためには f^{π} の固有値の絶対値が 1 以下であればよい．

定理 6.3 (ペロンフロベニスの定理)**.** A を非負正方行列とする．すなわち，A の成分が全て非負であるとする．この時，以下が成立する．

1. A は非負の実固有値を持つ．
2. A の非負固有値のうち，最大のものを α とする．α に対する非負固有ベクトルが存在する．α のことを特に A のフロベニウス根という．
3. A の任意の固有値の絶対値は α 以下である
4. $A^{\top}$ のフロベニウス根は α と等しい

補題 6.4. f^{π} は 1 をフロベニウス根に持つ．したがって，f^{π} の任意の固有値の絶対値は 1 以下．

証明. $\mathbf{d}$ を成分が全て 1 の 16 次元のベクトルとする．確率行列の性質より明らかに

$$f^{\pi}\mathbf{d}=\mathbf{d}$$

である．したがって，f^{π} は 1 を固有値に持つ．

α を f^{π} の任意の固有値とする．$\mathbf{x}$ を α に対応する固有ベクトルとする．$\|\mathbf{x}\|_{\infty}=|x_k|$ とする．固有ベクトルの定義から

$$\alpha x_k=\sum_{j=1}^{16}f^{\pi}_{kj}x_j$$

であるから，この両辺の絶対値を取ると次のようになる．

$$|\alpha|\,|x_k|\leq\sum_{j=1}^{16}f^{\pi}_{kj}\,|x_j|\leq\sum_{j=1}^{16}f^{\pi}_{kj}\,|x_k|=|x_k|\,.$$

$|x_k|\neq 0$ より，両辺をこれで割れば

$$|\alpha|\leq 1$$

を得る． $\square$

以上により，等比級数 $\displaystyle\sum_{n=0}^{\infty}(\gamma f^{\pi})^n$ は存在する．

6.3 行列ノルム $\|A\|$ について

ここで示すことは以下である．

- ベクトルの ∞-ノルムから行列の ∞-ノルムが誘導されること
- ベクトルの ∞-ノルムと行列の ∞-ノルムが両立すること

補題 6.5. ベクトルの ∞-ノルムから誘導される行列ノルムを $\|A\|$ とすると，それは次である．A を n 次正方行列とするとき

$$\|A\| = \max_{1\leq i\leq n}\sum_{j=1}^{n}|a_{ij}|$$

である．

証明. 誘導されるノルムの定義から，$\|A\| = \displaystyle\max_{x\in\mathbb{R}^n, \|x\|_\infty\leq 1}\|Ax\|_\infty$ なので，

$$\max_{x\in\mathbb{R}^n, \|x\|_\infty\leq 1}\|Ax\|_\infty \leq \max_{1\leq i\leq n}\sum_{j=1}^{n}|a_{ij}|, \quad \max_{x\in\mathbb{R}^n, \|x\|_\infty\leq 1}\|Ax\|_\infty \geq \max_{1\leq i\leq n}\sum_{j=1}^{n}|a_{ij}|$$

を示せばよい．

$x\in\mathbb{R}^n$ で $\|x\|_\infty \leq 1$ なるものを任意に取る．

$$\|Ax\|_\infty = \max_{1\leq i\leq n}\left|\sum_{j=1}^{n}a_{ij}x_j\right| \leq \max_{1\leq i\leq n}\sum_{j=1}^{n}|a_{ij}|\,|x_j| \leq \max_{1\leq i\leq n}\sum_{j=1}^{n}|a_{ij}|\times 1 = \max_{1\leq i\leq n}\sum_{j=1}^{n}|a_{ij}|.$$

x は $\|x\|_\infty \leq 1$ なるもので任意に取ったから

$$\max_{x\in\mathbb{R}^n, \|x\|_\infty\leq 1}\|Ax\|_\infty \leq \max_{1\leq i\leq n}\sum_{j=1}^{n}|a_{ij}|.$$

i 行目が $\displaystyle\sum_{j=1}^{n}|a_ij|$ を最大にする行だとする (そのような行が複数ある場合は初めの行を採用する)．ベクトル $y\in\mathbb{R}^n$ を次のように定義する．

$$y_j = \begin{cases} 1 & (a_{ij}\geq 0) \\ -1 & (a_{ij}<0) \end{cases}.$$

明らかに，$\|y\|_\infty = 1 \leq 1$ である．

$$\begin{aligned}\sum_{j=1}^{n} |a_{ij}| = \sum_{j=1}^{n} |a_{ij}y_j| = \sum_{j=1}^{n} a_{ij}y_j = \left|\sum_{j=1}^{n} a_{ij}y_j\right| \\ \leq \max_{1\leq i\leq n} \left|\sum_{j=1}^{n} a_{ij}y_j\right| = \max_{1\leq i\leq n, \|y\|_\infty\leq 1} \left|\sum_{j=1}^{n} a_{ij}y_j\right| = \|A\|.\end{aligned}$$

□

補題 6.6. ベクトルの ∞-ノルムと行列の ∞-ノルムは両立する．即ち，

$$\|Ax\|_\infty \leq \|A\| \|x\|_\infty$$

が成立する．

証明. 誘導されるノルムの定義より，

$$\|A\| = \max_{x\in\mathbb{R}^n, \|x\|_\infty\neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty}$$

である．よって $y \in RR^n, \|y\|_\infty \neq 0$ を任意に一つとると

$$\|A\| \geq \frac{\|Ay\|_\infty}{\|y\|_\infty}.$$

したがって，

$$\|Ax\|_\infty \leq \|A\| \|x\|_\infty.$$

□

参考文献

[1] Szepesvári, Csaba. *Algorithms for Reinforcement Learning* Morgan and Claypool Publishers. 小山田創哲・前田新一・小山雅典 (訳) 速習強化学習―基礎理論とアルゴリズム―，共立出版，2017

[2] 久保隆宏，Phtyon で学ぶ強化学習 入門から実践まで，講談社, 2019

[3] 森村哲郎，強化学習，講談社，2019

[4] Gilles Pagés. *Numerical Probability An Introduction with Applications to Finance* Universitext, Springer. 2017.

[5] 舟木直久，確率論，朝倉書店，2004

[6] 齋藤正彦，線型代数入門，東京大学出版会，1966

謝辞

本論文の執筆にあたり，多くの方々にご支援いただきました．とりわけ，指導教官であるコハツ・ヒガ・アルトゥーロ教授には毎週議論していただき，多くのご指導をしていただきました．心から感謝いたします．