1M2-2

不完全私的観測付き繰り返しゲームにおける均衡発見プログラム

Equilibrium Search Program of Repeated Games with Private Monitoring

山本 駿 *1 Shun Yamamoto　岩崎 敦 *2 Atsushi Iwasaki　趙 登吉 *1 Dengji Zhao　横尾 真 *1 Makoto Yokoo

*1九州大学 システム情報科学府
Graduate School of Information Science and Electrical Engineering at Kyushu University

*2電気通信大学 大学院情報システム学研究科
Graduate School of Information Systems, University of Electro-Communications

In an infinite repeated game, players repeatedly play the same stage game over an infinite horizon. In a private monitoring case, a player cannot directly observe opponent's action; she observes a noisy private signal. In this paper, by utilizing an equilibrium analysis program, we developed an evolutionary search program that finds an equilibrium. We apply our program to repeated prisoner's dilemma, and find a new equilibrium in this game.

1. 序論

無限回繰り返しゲームは，長期的関係にあるプレイヤ間の暗黙の協調を説明するためのモデルであり，主に経済分野で企業間の談合といった協調行動を分析するために研究が行われてきた [岡田 11]．繰り返しゲームにおいては，プレイヤが他のプレイヤの行動をどの程度正確に観測できるかによって問題の性質が大きく変化する．他のプレイヤの行動が完全に観測できる場合は完全観測 (perfect monitoring) と呼ばれ，既に多くの研究成果が得られている．一方，他のプレイヤの行動を，ノイズを含むシグナルを通して不完全にしか観測できない場合は，不完全観測 (imperfect monitoring) と呼ばれる．特に，プレイヤが観測するシグナルが私的なもので，そのシグナルを他のプレイヤが観測できない場合は，不完全私的観測 (imperfect private monitoring) と呼ばれている．不完全私的観測は自然な仮定であり，様々な応用事例が存在するが，均衡の判定自体が難しい問題となっている．

不完全私的観測付き繰り返しゲームにおいて，有限状態プレオートマトン (finite state preautomaton, pre-FSA) と初期相関装置 (initial correlation device) で示される戦略が均衡を構成するか否かを，部分観測可能マルコフ過程 (Partially Observable Markov Decision Process, POMDP) における最適プランを求めるプログラム (POMDP solver) を用いて検証する手法が既に開発されている [ジョ 13]．しかし POMDP solver は一般には有限時間で終了することが保証されない．本研究ではまず，有限時間で終了することが保証される，上記手法の改良方法を示す．さらに，上記手法を部品として用いて，進化的アルゴリズムにより均衡を発見するプログラムを開発する．さらに，代表的なゲームである囚人のジレンマに関して，開発したプログラムを用いて発見した均衡戦略を示す．

2. 私的観測付き繰り返しゲーム

本章では，本研究で扱う私的観測付き繰り返しゲームに関する基本的な用語を定義する．

連絡先: 山本駿，九州大学大学院システム情報科学府，812-0395 福岡県福岡市西区元岡 744 番地，(092)802-3576，syamamoto@agent.inf.kyushu-u.ac.jp

2.1 繰り返しゲーム

本節では文献 [Kandori 10] に基づき，2 人ゲームにおける私的観測付き無限回繰り返しゲームをモデル化する．ただし本論文で扱う手法は n 人プレイヤ，非対称ゲームにまで容易に拡張できる．

私的観測付き無限回繰り返しゲームでは，プレイヤ $i \in \{1,2\}$ は同じステージゲーム (stage game) を無限回 $t = 1, 2, \ldots$ に渡って繰り返す．各ステージゲームにおいて，プレイヤ i はまず，有限集合 A_i から行動 a_i を選択する．次にプレイヤ i は行動の組合せ $\boldsymbol{a} = (a_1, a_2)$ に関する私的なシグナル $\omega_i \in \Omega_i$ を観測する．プレイヤが行動の組合せ $\boldsymbol{a}$ を選択したとき，生起するシグナルの組合せが $\boldsymbol{\omega}$ である確率を $o(\boldsymbol{\omega} \mid \boldsymbol{a})$ で与える．このとき，$o_i(\omega_i \mid \boldsymbol{a})$ を Ω_i の限界分布 (marginal distribution) と呼ぶ．さらに，ステージゲームにおけるプレイヤ i の利得を利得関数 $g_i(\boldsymbol{a})$ で与える．

例として代表的なゲームである囚人のジレンマを説明する．囚人のジレンマにおいては，プレイヤは $N = \{1,2\}$ の 2 人が存在し，行動の集合は $A_1 = A_2 = \{C, D\}$ であり，シグナルの集合は $\Omega_1 = \Omega_2 = \{g, b\}$ である．ここで C は協力を，D は非協力を意味する．また行動 C/D に対する “正しい” シグナルを g/b とする．つまり，プレイヤ 2 が C を選んだとき，プレイヤ 1 にとって，g を受け取る確率が十分高いが，b を受け取る可能性もあるとする．また利得関数は以下で与えられるとする．

	$a_2 = C$	$a_2 = D$
$a_1 = C$	$1, 1$	$-1, 1.5$
$a_1 = D$	$1.5, -1$	$0, 0$

プレイヤ i は自分の行動 a_i とシグナル ω_i から “認識利得” (recognized payoff) $\pi_i(a_i, \omega_i)$ を受け取る．このため，プレイヤ i の期待利得は $\sum_{\boldsymbol{\omega} \in \prod_{i \in N} \Omega_i} \pi_i(a_i, \omega_i) o(\boldsymbol{\omega} \mid \boldsymbol{a})$ で与えられる．本論文では，期待利得とステージゲームの利得が一致するように認識利得が選ばれていることを仮定する．ステージゲームは無限回繰り返し行われるので，t 回目の行動の組合せを $\boldsymbol{a}^t$ で表すと，プレイヤ i の割引利得は割引率 $\delta \in (0,1)$ により $\sum_{t=1}^{\infty} \delta^t g_i(\boldsymbol{a}^t)$ となる．

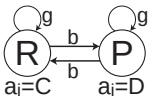

図 1: 1-MP

2.2 プレイヤの計画

有限状態オートマトン (finite state automaton, FSA) は無限の長さを持つプレイヤの計画を簡潔に表現する方法としてよく用いられる．FSA M_i を $\langle \Theta_i, \hat{\theta}_i, f_i, T_i \rangle$ で表す．ここで，Θ_i は状態の集合を表し，$\{\theta_i^1, \ldots, \theta_i^{k_i}\}$ を表す．なお，k_i はプレイヤ i の FSA の状態数を表す．また，$\hat{\theta}_i \in \Theta_i$ は初期状態を表し，$f_i : \Theta_i \to A_i$ は各状態における選択行動を表し，$T_i : \Theta_i \times \Omega_i \to \Theta_i$ は決定的な状態遷移を表す．

有限状態プレオートマトン (finite state preautomaton, pre-FSA) とは初期状態を指定しない FSA のことである．つまり pre-FSA m_i は $\langle \Theta_i, f_i, T_i \rangle$ で定義される．また pre-FSA m_i に初期状態 $\hat{\theta}_i$ を指定して得られる FSA を $(m_i, \hat{\theta}_i)$ で表す．この記述法からも明らかなように pre-FSA は FSA（計画）の集合を表し，1 つの状態に対し 1 つの FSA が対応する．$\boldsymbol{m}$ で pre-FSA の組合せ $(m_1, \ldots, m_n)$ を示す．以下の 例 1 に示す pre-FSA は，後述する均衡判定プログラムにより，あるパラメータの範囲で均衡を構成することが判明している．

例 1 (1-MP) 繰り返し囚人のジレンマにおいて，それぞれのプレイヤは図 1 で表される pre-FSA に従うと仮定する．この pre-FSA を本論文では一回相互罰則 (1-period Mutual Punishment, 1-MP) と呼ぶ．2 人のプレイヤの初期状態が共に R である場合，g を見続ける限りは互いに C を行い続けることになる．さらに，そのときプレイヤが間違って b を観測し，P に状態遷移したとしても，近い将来にプレイヤは両方 P に状態遷移し互いに罰則を行い，また R に戻ることが可能である．

3. 信念に基づく戦略

ゲーム理論における戦略は，起こりうるすべての状況でプレイヤがとる計画を示すものである．つまり，予定された計画だけでなく，そこから誤って逸脱した後の計画も，戦略は同様に記述しなければならない．たとえば，プレイヤの初期状態が示す行動が協力である場合に，誤って非協力を選択した状況においても，戦略はこの後プレイヤがとる計画を指定する必要がある．このような計画をオフパスの計画とよび，予定される行動の後の計画をオンパスの計画と呼ぶ．従来の繰り返しゲームの研究においては，プレイヤの戦略は，プレイヤの私的な行動と観測の履歴から現在の行動への写像として記述されていた．しかしながら，文献 [Kandori 10] で述べられているように，私的観測においてはこの方法ではオフパスの計画の記述が膨大になる．そこで本論文では，文献 [Kandori 10] で示されている，信念分割と pre-FSA と初期相関装置を使ってプレイヤの戦略を簡潔に記述する方法を用いる．

定義 1 (初期相関装置) 初期相関装置 $r \in \Delta(\prod_{i \in N} \Theta_i)$ は各プレイヤの初期状態に関する同時分布を示す．ここで，$\Delta(\prod_{i \in N} \Theta_i)$ は $\prod_{i \in N} \Theta_i$ 上の任意の確率分布の集合である．初期相関装置はプレイヤの状態の組合せ $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n) \in \prod_{i \in N} \Theta_i$ を確率 $r(\boldsymbol{\theta})$ で選び，任意のプレイヤ i に θ_i を初期状態として推薦する．プレイヤ i が θ_i を推薦される確率を $r_i(\theta_i)$ とする．

プレイヤ i が初期相関装置から初期状態 θ_i を推薦されたとする．このとき，プレイヤ i は他のプレイヤがどの初期状態を推薦されたかを知ることができない．しかしながら，事前に定められた初期相関装置の同時分布と，自分に推薦された初期状態から，ベイズの定理を用いて他のプレイヤの状態に関する初期信念 $r_{-i}(\cdot|\theta_i)$ を計算可能である．またプレイヤ i が持ちうる初期信念の集合を $B_i(r) = \{r_{-i}(\cdot|\theta_i) | \theta_i \in \Theta_i, r_i(\theta_i) > 0\}$ とする．

さらにプレイヤ i の現在の信念が b_i で，選んだ行動が a_i，得られたシグナルが ω_i であった場合の事後信念を $\chi_i[(a_i, \omega_i), b_i]$ で表す．また，履歴 $h_i^t \in H_i^t := (A_i \times \omega_i)^t$ に対し，b_i を事前信念とする履歴 h_i^t の後の事後信念を $\chi_i[h_i^t, b_i]$ で表す．

この初期信念と事後信念を一般に信念と呼び，$b_i(\boldsymbol{\theta}_{-i})$ で示す．ここで，$\boldsymbol{\theta}_{-i}$ はプレイヤ i 以外の状態の組合せを示す．また，$(\theta_i, \boldsymbol{\theta}_{-i})$ でプレイヤ i の状態が θ_i でプレイヤ i 以外のプレイヤの状態の組合せが $\boldsymbol{\theta}_{-i}$ であることを示す．

定義 2 (信念分割) プレイヤ i の信念分割 D_i を $(D_i^1, \ldots, D_i^{k_i})$ で表す．ここで，$\forall D_i^l \in D_i, D_i^l \subseteq \Delta(\prod_{j \neq i} \Theta_j)$ を満たす．

2 つの信念分割 D_i と $\acute{D}_i$ に対して，$D_i \subseteq \acute{D}_i$ は任意の l に対して，$D_i^l \subseteq \acute{D}_i^l$ を意味する．同様に，2 人のプレイヤの信念分割の組合せ $\boldsymbol{D}$ と $\acute{\boldsymbol{D}}$ に対して，$\boldsymbol{D} \subseteq \acute{\boldsymbol{D}}$ は任意の i に対して，$D_i \subseteq \acute{D}_i$ を意味する．

後で示すように，信念分割の各要素には，pre-FSA の一つの状態が対応しており，プレイヤの信念がある信念分割に含まれているときは，対応する pre-FSA の状態から始める計画を用いるというように戦略を定義する．つまり，$b_i \in D_i^l$ のときは θ_i^l から始める計画を用いる．たとえば，2 人のプレイヤが 1-MP に従っているときに，$D_1 = (D_1^R, D_1^P) = ([0.7, 0.9], [0.1, 0.7])$ という分割を構成するとする．ここで，$[0.7, 0.9]$ は 0.7 から 0.9 までのすべての信念を表す．この信念分割と pre-FSA の組で，信念が $[0.7, 0.9]$ に含まれるときは，プレイヤ 1 は計画を状態 R から始め，信念が $[0.1, 0.7]$ に含まれるときは，プレイヤ 1 は計画を状態 P から始めるという戦略を表現する．さらに信念が $[0.1, 0.9]$ に含まれないとき，つまり信念分割に含まれないときは，どの計画を用いるかは未定義となっている．

定義 3 (適応性) ある信念分割 D_i が初期相関装置 r に適応するとは次の条件を満たすことである．(i) $\forall \theta_i^l$ に対して，$r_i(\theta_i^l) > 0$ ならば，$r_{-i}(\cdot|\theta_i^l) \in D_i^l$ が成り立ち，(ii) $\forall t \geq 1, \forall b_i^0 \in B_i(r), \forall h_i^t \in H_i^t := (A_i \times \Omega_i)^t, \chi_i[h_i^t, b_i^0] \in D_i$ が成り立つ．また $\boldsymbol{D}$ が r に適応するとは，任意の i に対し $D_i \in \boldsymbol{D}$ が r に適応することを意味する．

定義 3 の 1 つ目の条件は，プレイヤがある状態を初期相関装置から推薦されるならば，その状態に対応する信念分割にそのときの初期信念が含まれることを意味する．2 つ目の条件は，初期相関装置から得られる初期信念を事前信念として，任意の履歴の後の事後信念を計算すると，その事後信念が，必ずいずれかの信念分割に含まれることを意味する．

定義 4 (信念に基づく戦略) プレイヤ i の信念に基づく戦略を (m_i, D_i, r) によって定義する．なお，D_i は r に適応することを仮定する．まず，プレイヤ i の均衡上の振舞いは (m_i, θ_i) で示される．ここで，θ_i は r によって推薦される状態を指す．次

に逸脱した後の計画は再帰的に以下のようにする．まずプレイヤ i が逸脱したとき持つ信念を b_i とする．そして $b_i \in D_i^l$ となる D_i^l を見つけ，この後のプレイヤ i の計画を (m_i, θ_i^l) で定める．D_i は r に適応することから，$b_i \in D_i^l$ を満たす D_i^l は必ず存在する．さらにこの後，プレイヤ i が逸脱することがあったら同様の方法で計画を定義する．

定義 5 (RFSE) $(\boldsymbol{m}, \boldsymbol{D}, r)$ が再起有限状態均衡 (Resilient Finite State Equilibrium, RFSE) を構成するとは以下の 2 条件を満たすことを意味する：(1) $\boldsymbol{D}$ が r に適応する．(2) 任意のプレイヤ $i \in N$ の任意の初期信念 $b_i^0 \in B_i(r)$ を事前信念とする，あらゆる履歴の後の事後信念 b_i に対して，プレイヤの割引期待利得を最大とする計画は (m_i, θ_i^l) となる．ここで，θ_i^l は $b_i \in D_i^l$ が成立するように選ばれる．

この定義は，RFSE が初期相関装置のもとでの逐次均衡であることを意味する．注意すべき点として，本論文で示す均衡判定アルゴリズムは，pre-FSA で記述可能な，比較的簡単な戦略しか扱うことはできないが，プレイヤが逸脱可能な戦略の空間に関しては，一切制限を加えていないということがある．つまり，他のプレイヤが RFSE を構成する pre-FSA に従う限り，均衡となる pre-FSA 以上に大きな割引期待利得をプレイヤ i に与える戦略は，pre-FSA で記述できないような複雑な戦略も含めて一切存在しない．

4. 均衡判定プログラム

本章では，均衡判定プログラムについて説明する．文献 [ジョ 13] のプログラムとの違いは，POMDP-solver を利用せず，有限のステップで均衡のチェックを可能としている点である．まず，ゲームに参加するすべてのプレイヤの pre-FSA の積である joint pre-FSA を定義する．

定義 6 (Joint pre-FSA) Pre-FSA の組合せ $\boldsymbol{m} = (m_1, \ldots, m_n)$ に対して，joint pre-FSA $(\boldsymbol{\Theta}, f, T)$ を定義する．各プレイヤの pre-FSA を $m_i = \langle \boldsymbol{\Theta}_i, f_i, T_i \rangle$ とすると $\boldsymbol{\Theta} = \prod_{i \in N} \Theta_i$，$f : \boldsymbol{\Theta} \to \prod_{i \in N} A_i$ は $f(\boldsymbol{\Theta}) = (f_1(\theta_1), \ldots, f_n(\theta_n))$，$T : (\boldsymbol{\Theta} \times \prod_{i \in N} \Omega_i) \to \boldsymbol{\Theta}$ は $(T_1(\theta_1, \omega_1), \ldots, T_n(\theta_n, \omega_n))$ として与えられる．

Joint pre-FSA は初期相関装置による初期状態の分布とともに，結合状態の集合 $\boldsymbol{\Theta}$ 上のマルコフ連鎖を定義する．そのマルコフ連鎖の遷移確率行列を $Q_{\boldsymbol{m}}$ とする．

定義 7 (正則性) Joint pre-FSA の遷移確率行列である $Q_{\boldsymbol{m}}$ が正則であるとは，ある $t \geq 1$ に対して，$Q_{\boldsymbol{m}}^t$ のすべての要素が厳密に正となることを意味する．

$v_i^{\boldsymbol{\theta}}$ を joint pre-FSA の結合状態が $\boldsymbol{\theta}$ のときのプレイヤ i の割引期待利得とする．この値は joint pre-FSA に基づき，$v_i^{\boldsymbol{\theta}} = g_i(f(\boldsymbol{\theta})) + \delta \sum_{\boldsymbol{\omega} \in \prod_{j \in N} \Omega_j} v_i^{\boldsymbol{\theta}'} \cdot o(\boldsymbol{\omega} | f(\boldsymbol{\theta}))$ を解くことで求められる．ここで，$\boldsymbol{\theta}' = T(\boldsymbol{\theta}, \boldsymbol{\omega})$ である．

定義 8 (割引期待利得) プレイヤ i の信念が b_i で与えられたときの割引期待利得を $V_i^{M_i}(b_i)$ で表す．特に $V_i^{(m_i, \theta_i)}(b_i)$ は $\sum_{\boldsymbol{\theta}_{-i} \in \prod_{j \neq i} \Theta_j} v^{(\theta_i, \boldsymbol{\theta}_{-i})} b_i(\boldsymbol{\theta}_{-i})$ で計算される．

定義 9 (1 回限りの拡張) pre-FSA m_i に対して，1 回限りの拡張という特殊な FSA を以下のように作る．(i) $a_i \in A_i$ を行う新しい状態 θ_i' を初期状態とし，(ii) さらに ω_i を観測した後は，FSA $(m_i, \theta_i^{\omega_i})$ をプレイする．なお，$\theta_i^{\omega_i} \in \Theta_i$ である．

プレイヤ i の pre-FSA m_i について，1 回限りの拡張で作ることのできるすべての FSA の集合を $Ex(m_i)$ で表す．この集合は，定義より $|A_i| \cdot k_i^{|\Omega|}$ 個の FSA を要素に持つ．また，1 回限りの拡張の概念を用いて，目標信念分割と呼ばれる信念分割を導入する．これは，1 回限りの拡張によって期待利得が向上しない信念を集めた信念分割である．

定義 10 (目標信念分割) プレイヤ i の目標信念分割 $\hat{D}_i$ は $\hat{D}_i^l$ のそれぞれに含まれる信念 $\forall b_i \in \hat{D}_i^l$ を $V_i^{(m_i, \theta_i^l)}(b_i) \geq V_i^{M_i}(b_i), \forall M_i \in Ex(m_i)$ を満たすように選んだものである．

目標信念分割の組合せを $\hat{\boldsymbol{D}}$ で表す．$V_i^{(m_i, \theta_i)}(b_i)$ は信念 b_i に対して線形である．さらに線形方程式を解くことで $\hat{D}_i$ を得られることから，$\hat{D}_i^l$ は凸多面体である．次に，この目標信念分割と RFSE の関係を示す．

定理 1 $\hat{\boldsymbol{D}}$ は r に適応する場合，かつその場合に限り，pre-FSA の組合せ $\boldsymbol{m}$ と初期相関装置 r は RFSE を構成する．

証明は紙面の都合上省略する．

次に，$\hat{\boldsymbol{D}}$ が r に適応するか判定する方法を説明する．以下の性質を用いる．

補助定理 1 ある t^* に対して，$t < t^*$ を満たす任意の履歴 h_i^t に対して，初期信念から到達可能なすべての信念が目標信念分割に含まれ，$t = t^*$ を満たす任意の h_i^t に対して，初期信念のみならず，すべての可能な信念から到達可能なすべての信念が目標信念分割に含まるなら，$\hat{D}_i$ は r に適応する．

補助定理 1 の性質，および文献 [Phelan 12] に示されている，Q_m が正則であるとき，十分長い履歴の後に到達可能な信念集合が収束するという性質により，D_i が r に適応することを有限ステップで判定するアルゴリズムが構築できる．詳細は紙面の都合上省略する．

5. 均衡発見プログラム

本章では，遺伝的アルゴリズムをもとに開発した均衡発見プログラムを紹介する．遺伝的アルゴリズムは，個体に遺伝子組み換えの操作を加え，評価値の高い個体の持つよい性質を残しながら世代交代をすることで解を探索する [Whitley 94]．本論文では pre-FSA を個体／解の候補とし，優れた解を探索する．評価値を与える関数（評価関数）を定義するため，まず必須到達点の定義を示す．

定義 11 (必須到達点) Q_m が正則であるとき，任意の初期信念から，ある行動 a_i とあるシグナル ω_i の観測を繰り返した後の事後信念は 1 点に収束する．このときの信念を必須到達点とよび $\hat{b}_i^{(a_i, \omega_i)}$ で表す．また，必須到達点の集合を $\hat{B}_i = \{\hat{b}_i^{(a_i, \omega_i)} | a_i \in A_i, \omega_i \in \Omega_i\}$ で表す．

必須到達点は，どのような初期信念を選んでも必ず到達可能な点であり，必須到達点が目標信念分割に含まれていなければ均衡にはなり得ない．よって，必須到達点が目標信念分割に含ま

れているか否かを評価関数に反映することは自然である．一方，目標信念分割は必ずしも存在しないため，目標信念分割が存在しない場合は別の評価関数を用いる必要がある．信念分割 D_i と信念 b_i に対して，$b_i \in D_i$ は $b_i \in D_i^l$ を満たす $D_i^l \in D_i$ が存在することを表す．

定義 12 (目標信念分割が存在するときの評価関数) 2 つの信念 b_i, b_i' のマンハッタン距離を $MDistance(b_i, b_i') = \sum_{\boldsymbol{\theta}_{-i} \in \Pi_{j \neq i} \Theta_j} |b_i(\boldsymbol{\theta}_{-i}) - b_i'(\boldsymbol{\theta}_{-i})|$ で定義する．また，これを用いて信念分割 D_i と信念 b_i のマンハッタン距離を $MDistance(D_i, b_i) = \min_{b_i' \in D_i} MDistance(b_i', b_i)$ で定義する．このとき，目標信念分割が存在するときの評価関数 ψ を必須到達点と目標信念分割のマンハッタン距離の平均値として以下のように定義する．

$$\psi(m_i) = \frac{\sum_{\hat{b}_i \in \hat{B}_i} MDistance(\hat{D}_i, \hat{b}_i)}{|\hat{B}_i|} \quad (1)$$

定義 13 (目標信念分割が存在しないときの評価関数) pre-FSA m_i と信念 b_i が与えられたとき，その pre-FSA が定める計画の割引期待利得の最大値を $\hat{V}(m_i, b_i) = \max_{\theta_i \in \Theta_i} V_i^{(m_i, \theta_i)}(b_i)$ で定める．同様に，pre-FSA m_i に対する 1 回限りの拡張の集合 $Ex(m_i)$ の割引期待利得の最大値を $\hat{V}(Ex(m_i), b_i) = \max_{M_i \in Ex(m_i)} V_i^{M_i}(b_i)$ で定める．事前に定めておいた代表的な信念の集合 B_i^{pre} に対して，目標信念分割の存在しないときの評価関数 ψ を以下のように定める．

$$\psi(m_i) = \frac{\sum_{b_i \in B_i^{pre}} (\hat{V}(Ex(m_i), b_i) - \hat{V}(m_i, b_i))}{|B_i^{pre}|}$$

次に世代交代の方法を説明する．まず，中間世代の個体を選ぶ．中間世代の個体とは現世代から選ばれ，その後次世代となる個体のことで，その個体数を交叉率に応じて定め，評価の高いものから選ぶ．次に交叉の準備をする．評価値の下での順位に対して，事前に定めた選択確率に応じて，母親と父親を中間世代の個体から選ぶ．次に交叉を行う．pre-FSA の状態には 1 から順に識別子が与えられている．交叉ではランダムに選ばれた値に対して，子はその値までの識別子を持つ状態を母親から受け継ぎ，残りを父親から受け継ぐ．このとき，その状態における選択行動と状態遷移も一緒に受け継いで，新しい pre-FSA を構成する．最後に，突然変異では事前に定められた突然変異確率に応じて，受け継いだ選択行動と状態遷移をランダムに変更する．

この親の選択方法はランキング方式と呼ばれる．別の一般的な交叉方法として，ルーレット方式というものがある．ランキング方式が評価値の下での順位に応じて，選択確率が決まるのに対し，ルーレット方式は評価値の大きさに応じて，選択確率が決まる．一般にランキング方式よりルーレット方式のほうが，評価値の高い個体が選ばれやすいため，収束が早いという長所がある．しかし，本論文の問題設定では，pre-FSA の性質に応じて 2 つの異質な評価関数が排他的に用いられ，評価値に大きな差がつく可能性がある．よって，ランキング方式を用いることとする．

上記の評価関数の値は，小さい方が望ましいと考えることは自然であるが，均衡を構成する pre-FSA との間の距離と厳密に対応するものではない．ここでの距離とは，ある pre-FSA から均衡を構成する pre-FSA を得るために必要な行動選択と状態遷移の変更回数を意味する．このため，我々は各世代の人口を多く，また交叉のときに選ばれる選択確率を下位のものでも比較的高く，さらに突然変異の確率を高く設定することで，各世代で多様性が保証されるようにしている．

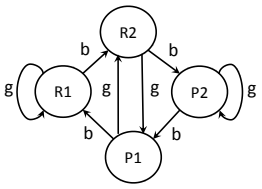

図 2: 1-MP1-Delay

この均衡発見プログラムを用いて，ステージゲームが囚人のジレンマであるゲームにおいて，図 2 で示される新しい均衡を発見した．この pre-FSA で双方の初期状態が $R1$ のとき，最初は C を行い続ける．ノイズによって観測エラーが生じた場合，すなわち b を観測した場合には，D を行うが，互いに罰し合うとすぐにまた協力しあう．この均衡の挙動は初期状態が R の 1-MP のものと似ている．しかし，罰則を与えるタイミングを $R2$ で 1 回協力することで遅らせている．このことにより，この均衡では $R1$ を初期状態とするとき，いくつかのパラメータ設定においては，1-MP 以上の割引期待利得を与えることが示されている．

6. 結論

本論文では不完全私的観測付き繰り返しゲームにおける均衡判定プログラムを改良し，この均衡判定プログラムを部品として用いる遺伝的アルゴリズムに基づく均衡発見プログラムを開発した．このプログラムを用いて，代表的なゲームの 1 つである囚人のジレンマにおいて，状態数が 4 つの新しい均衡を発見した．発見した新しい均衡は，今まで見つけ出された均衡のなかで，最大の割引期待利得を与える．状態数が 4 つの pre-FSA で正則なものをすべて数え上げることは困難であり，本論文で示したような均衡探索アルゴリズムは，より状態数の多い pre-FSA によって構成される均衡を発見するために有用であると考えられる．今後の課題として，囚人のジレンマ以外の不完全私的観測の応用事例に関して，本プログラムを適用して新しい均衡を発見することが挙げられる．

参考文献

[ジョ 13] ジョヨンジュン, 岩崎 敦, 神取 道宏, 小原 一郎, 横尾 真: 部分観測可能マルコフ決定過程を用いた私的観測付き繰返しゲームにおける均衡分析プログラム, 情報処理学会論文誌, 2445–2456 (2012)

[Kandori 10] Kandori, M. and Obara, I.: Towards a Belief-Based Theory of Repeated Games with Private Monitoring: An Application of POMDP (2010). mimeo (http://mkandori.web.fc2.com/papers/KObb10June4.pdf).

[岡田 11] 岡田 章: ゲーム理論（新版）, 有斐閣 (2011)

[Phelan 12] Phelan, C. and Skrzypacz, A.: Beliefs and Private Monitoring. *Review of Economic Studies* 79(4):1637-1660, (2012).

[Whitley 94] Whitley, D.: A Genetic Algorithm Tutorial., *Statistics and Computing*, 4:65-85, (1994)