5C-02

公的観測下の繰り返しプロジェクトゲームにおける協力のダイナミクス

五十嵐 瞭平 *　　岩崎 敦 *
Ryohei Igarashi　　Atsushi Iwasaki

1　はじめに

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 [2] であり，主に経済学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発展してきた. 繰り返しゲームにおいて，見間違え (不完全観測) は現実的で重要な仮定である．2 人がまったく見間違えない「完全」観測下では，常に裏切り（ALLD）や一度でも裏切られたら許さない（Grin Trigger，GRIM）といった非協力的な戦略しか生き残らないことが知られている [3]．不完全観測下の繰り返しゲームでは，自分の行動と利得から相手の行動を推測できないようにするため，お互いの行動で決まるステージゲーム利得を，自分の行動とシグナルから決まる実現利得の見間違えの確率に関する期待値として定義する．本論文では，囚人のジレンマと同じ構造をもつプロジェクトゲームを扱い，プロジェクトの成功報酬とプロジェクトに対する努力コストが分離して利得を定義するケース（分離実現利得）と分離せずに定義するケース（非分離実現利得）を分析する．

繰り返しゲームの戦略は，昨日までの行動と観測の履歴から今日の行動への写像で定義する．ゲームを無限回繰り返すとき，戦略空間は無限になるので，すべての均衡戦略を具体的に特定することは現実的ではない．そこで，プレイヤが取りうる戦略を状態数 2 以下の有限状態機械 (Finite State Automaton, FSA) に限定する．戦略を FSA に限定したときの期待利得をマルコフ決定過程に基づいて計算した利得表から，突然変異付きレプリケータ方程式を構成し，その帰結を吟味する．

その結果，報酬とコストを分離するとき，実現利得が存在するための，利得とシグナル分布のパラメータの条件を明らかにした．さらに，この条件の下では，不寛容な戦略である GRIM や ALLD しか生き残らないことを明らかにした．

2　モデル

本章では文献 [2] に基づいて，2 人公的観測付き無限回繰り返しプロジェクトゲームをモデル化する．ここでプレイヤ $i \in \{1,2\}$ はステージゲームを無限期間 $t = 0,1,2,\cdots$ に渡って繰り返す．各期においてプレイヤ i は有限集合 A から行動 a_i を選択し，その行動の組を $\boldsymbol{a} = (a_1, a_2) \in A^2$ とする．次に，プレイヤ i は $\boldsymbol{a}$ に関する公的なシグナル $\omega \in \Omega$ を観測する．また， プレイヤが $\boldsymbol{a}$ を選択したとき ω が生起する同時確率を $p(\omega \mid \boldsymbol{a})$ とし， この同時確率を与える分布のことをシグナル分布と呼ぶ. ステージゲームは無限回繰り返し行われるので，プレイヤ i の割引利得和は割引因子 $\delta \in (0,1)$ により $\sum_{t=1}^{\infty} \delta^t g_i(\boldsymbol{a}^t)$ となる. ただし，$g_i(\cdot)$ の値は表 1 に示す囚人のジレンマの利得表に従う．

表 1: 囚人のジレンマ ($g > 0$，$l > 0$，および $|g - l| < 1$)

	$a_2 = C$	$a_2 = D$
$a_1 = C$	$1, 1$	$-l, 1+g$
$a_1 = D$	$1+g, -l$	$0, 0$

表 2: 実現利得の利得表 (行動とシグナルが独立)

	$\omega = G$	$\omega = B$
$a_1 = C$	$V(G) - P(C)$	$V(B) - P(C)$
$a_1 = D$	$V(G) - P(D)$	$V(B) - P(D)$

次にプレイヤ 2 の行動に関するプレイヤ 1 のノイズを含む観測をプレイヤ 1 のシグナルとし，$\omega \in \{G, B\}$ (*GOOD*, *BAD*) とする．G はプレイヤ 2 の C，B はプレイヤ 2 の D に関するシグナルである．プレイヤ 2 についても同様である．公的観測とはお互いに同じシグナルを観測する観測構造である．ここで，P_{a_1,a_2} はプレイヤ 1 と 2 の行動プロファイルが (a_1, a_2) のとき，公的シグナルが $GOOD$ となる確率である．

不完全観測において重要な実現利得を定義する．まず，分離実現利得では，実現利得を $V(\omega)$ と $P(a_i)$ と定義し，利得表は表 2 となる．次に，非分離実現利得では，実現利得を $V(\omega \mid a_i)$ と定義する．また，実現利得をシグナル分布に関して期待値をとると表 1 に一致する．

戦略は，そのプレイヤの過去の行動と受け取ったシグナルから現在の行動への写像で表現され，先に述べた通り，本研究では状態数 2 以下の非同相な 26 個の FSA に限定する．代表的な戦略としては，まず ALLD と GRIM があり，この 2 戦略はほぼすべてのステージゲーム利得に対して均衡を形成する．また，この 2 戦略よりも範囲は狭いが均衡を形成する戦略として，はじめ協力し，相手の行動を真似する戦略である “しっぺ返し” (Tit-For-Tat，TFT)，裏切られたら一度だけ相手を処罰し協力へ戻る戦略である “Forgiver”（FGV），“勝ち残り，負け逃げ” (Win-Stay，Lose-Shift, WSLS) の 3 つの戦略が存在する．

* 電気通信大学大学院情報理工学研究科

このような数ある戦略の中から有効な戦略を発見する方法の1つとして，突然変異付きレプリケータダイナミクス [1] がある．本論文では，その方程式を

$$\dot{x_i} = x_i\left[f_i(\vec{x}) - \phi(\vec{x})\right] + u\left(\frac{1}{n} - x_i\right),\ i = 1, ..., n$$

と定義する．$\phi(\cdot)$ を全ての戦略の利得の平均 $\sum_j x_j f_j(\vec{x})$，$f_j(\cdot)$ を $\sum_m x_j a_{jm}$ とする．ただし，a_{jm} は戦略 j をとるプレイヤが戦略 m を取るプレイヤと無限回プレイしたときの割引利得和である．

3 実現利得の存在条件

本章では，実現利得によって表現可能なステージゲーム利得とシグナル分布のパラメータ範囲を示す．ステージゲーム利得は実現利得のシグナル分布に関する期待値である．例えば，分離実現利得のケースでは式 1 が成り立つ．

$$\begin{bmatrix} P_{CC} & 1-P_{CC} & -1 & 0 \\ P_{CD} & 1-P_{CD} & -1 & 0 \\ P_{DC} & 1-P_{DC} & 0 & -1 \\ P_{DD} & 1-P_{DD} & 0 & -1 \end{bmatrix} \begin{bmatrix} V(G) \\ V(B) \\ P(C) \\ P(D) \end{bmatrix} = \begin{bmatrix} 1 \\ -l \\ 1+g \\ 0 \end{bmatrix} \tag{1}$$

式 1 が解を持つ条件より定理 1 が成り立つ．

Theorem 1. 分離実現利得のケースでは，式 2 を満たすときのみステージゲーム利得が存在する．

$$(1+g)(P_{CC} - P_{CD}) - (1+l)(P_{DC} - P_{DD}) = 0 \tag{2}$$

次に，非分離実現利得では，行列式が 0 でないため定理 2 が成り立つ．

Theorem 2. 非分離実現利得のケースでは，ステージゲーム利得が常に存在する．

4 実験結果

図 1 に $P_{CC} = 0.90, P_{CD} = P_{DC} = 0.40, P_{DD} = 0.30$ の公的観測下における最大多数戦略を示す．最大多数戦略とは，収束時に最も多くの人口を獲得した戦略を意味する．図の横軸は自分の裏切りによる利得の増分 g，縦軸は相手の裏切りによる損失 l に対応し，0.01 刻みで $[0.01, 1.00]$ をプロットした．なお，$\delta = 0.90, u = 0.01$ とした．$P_{CC} = 0.90, P_{CD} = P_{DC} = 0.40, P_{DD} = 0.30$ のとき，分離実現利得で表現可能なステージゲーム利得は，囚人のジレンマで要求される $|g - l| < 1$ を満たさないため，存在しない．一方で，非分離実現利得では常にステージゲーム利得が存在する．図 1 が示すように，g と l がある程度より大きい領域では GRIM や ALLD といった不寛容な戦略が最大多数となるが，g と l が小さい領域では FGV や TFT が最大多数となり，協力的な戦略が広い範囲で生き残る．

次に，$P_{CC} = 0.90, P_{CD} = P_{DC} = 0.70, P_{DD} = 0.50$ の公的観測を考える．このシグナルパラメータでは式 2 より，分離実現利得で表現可能なステージゲーム利得範囲として $g = l$ を得る．そこで，図 2 に $g = l$ において，g を変化させ最大多数戦略の推移を示した．g が 0.01 では FGV，0.02 から 0.05 では WSLS が最大多数となるが，それより大きいときは GRIM もしくは ALLD が最大多数となる．

以上より，分離実現利得では厳しい条件の下でのみステージゲーム利得が存在することを示した．また，非分離実現利得では広いパラメータ範囲において協力的な戦略が生き残るのに対し，分離実現利得でほとんどの利得範囲において不寛容な戦略しか生き残らない．

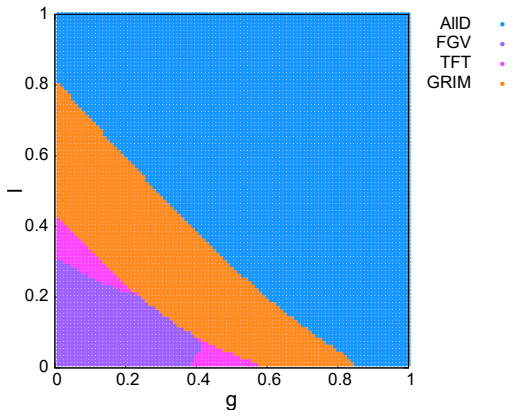

図 1: 公的観測下のダイナミクス ($P_{CC} = 0.90, P_{CD} = P_{DC} = 0.40, P_{DD} = 0.30$)

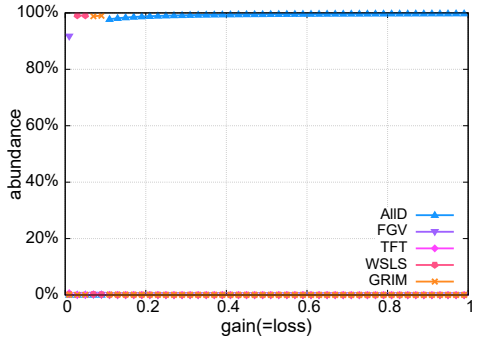

図 2: gain(=loss) による最大多数戦略の変化 ($P_{CC} = 0.90, P_{CD} = P_{DC} = 0.70, P_{DD} = 0.50$)

参考文献

[1] B. M. Zagorsky, J. G. Reiter, K. Chatterjee, and M. A. Nowak. Forgiver triumphs in alternating prisoner's dilemma. *PLOS ONE*, pp. 1–8, 2013.

[2] 神取. 人はなぜ協調するのか –くり返しゲーム理論入門–. 三菱経済研究所, 2015.

[3] 西野上, 五十嵐, 岩崎. 私的観測下の繰り返し囚人のジレンマにおける協力のダイナミクス. 第 19 回情報科学技術フォーラム, 2020.