日本オペレーションズ・リサーチ学会
2021年 秋季研究発表会

ほぼ公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

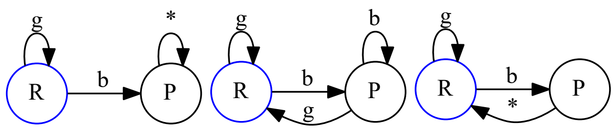
電気通信大学 *五十嵐瞭平 IGARASHI Ryohei
05000493 電気通信大学 岩崎敦 IWASAKI Atsushi

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 [4] 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．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（GRIM, 図 1a）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 [5]．しかし意外なことに有名なしっぺ返し（Tit-For-Tat，TFT, 図 1b）が生き残ることはない．一方で，2 人がお互いの見間違えをほぼ共有する「ほぼ公的」観測下でなら TFT が均衡になりうる [1]．そこで，本論文では「ほぼ公的」観測下の繰り返し囚人のジレンマを突然変異付きレプリケータダイナミクスを用いて分析した．

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，ほぼ公的観測下では AllD や GRIM に加えて，相手を 1 度だけ処罰したら協力に戻る戦略である Forgiver (FGV, 図 1c) [2] が広い利得パラメータの領域で生き残ることがわかった．一方で，TFT は FGV と比べてかなり狭い領域でしか生き残らなかった．



(a) GRIM (b) TFT (c) FGV

図 1: 主要な FSA

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: (a_1, a_2) のときのシグナル分布

	$w_2 = g$	$w_2 = b$
$w_1 = g$	$P_{a_1,a_2}(1-e)^2 + (1 - Pa_1, a_2)e^2$	$e(1-e)$
$w_1 = b$	$e(1-e)$	$P_{a_1,a_2}e^2 + (1 - Pa_1, a_2)(1-e)^2$

2. モデル

本章では文献 [3] に基づいて，2 人私的観測付き無限回繰り返しゲームをモデル化する．ここでプレイヤ $i \in \{1, 2\}$ はゲームを無限期間 $t = 0, 1, 2, \cdots$ 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 $\boldsymbol{a}$ に関する私的なシグナル $\omega_i \in \Omega$ を観測する．$\boldsymbol{w}$ をシグナルの組 $(\omega_1, \omega_2) \in \Omega^2$ とする．また， プレイヤが $\boldsymbol{a}$ を選択したとき $\boldsymbol{w}$ が生起する同時確率を $o(\boldsymbol{w} \mid \boldsymbol{a})$ 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．成分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．

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 の私的シグナルとし，$\omega \in \{g, b\}$ ($good$，bad) とする．正しい観測ではプレイヤ 2 が C を選択した際のプレイヤ 1 の私的シグナルは g， D を選択した際の私的シグナルは b となる．プレイヤ 2 についても同様である． 表 2 にほぼ公的観測のシグナル分布を示す． 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，e は公的シグナルを見間違える確率（独立誤差）である，

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

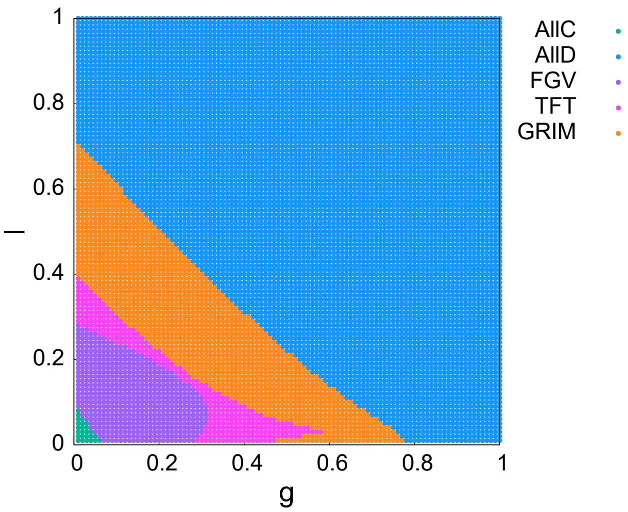


図 2: ほぼ公的観測下のダイナミクス

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．

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 [2] がある．本論文では，その方程式を

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

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. ほぼ公的観測下のダイナミクス

図 2 にほぼ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．ここで，シグナル分布のパラメータは $P_{cc} = 0.90, P_{cd} = P_{dc} = 0.40, P_{DD} = 0.30, e = 0.01$，割引因子は $\delta = 0.9$，初期戦略分布は一様分布，突然変異率 u を 0.01 とした．

図 2 が示すように，どんな戦略が生き残るかは利得構造に依存する．まず，g と l が大きいとき，裏切りによる損益が大きいため，他のどの戦略も協力を維持するに十分な将来利得を獲得できない．そのため，AllD が最大多数戦略となり，単独で他戦略を淘汰する．次に，g および l が中程度のとき，GRIM が最大多数戦略となる．このとき，プレイヤは最初はお互いに協力するが，一度でも裏切りが発生すると永遠に裏切り続ける．

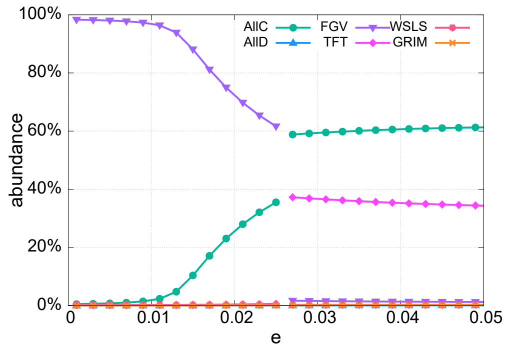


図 3: 独立誤差 e による戦略比率の変化

g や l がやや小さくなることによりプレイヤは初め協力を取るようになるが，裏切りの後に協力に戻ろうとはしない．

最後に g と l が小さいとき，FGV もしくは TFT が最大多数となるが，その領域は FGV の方が TFT がよりかなり広く，FGV が TFT を淘汰していると言える．これは，TFT より FGV の方が見間違えが起こった後の処罰状態から協力状態に戻りやすいためである．

ここで、独立誤差 e がダイナミクスの帰結に与える影響を吟味する．e はほぼ公的観測の特徴である見間違えの共有度合いを決めるパラメータである．図 3 に $g = 0.10, l = 0.10$ に e を変化させながら，収束時の戦略の比率を示した．見間違えが共有されやすい (e が比較的小さい) ときは，FGV が他の戦略をほぼ淘汰する．一方で，e が増加して，見間違いが共有されにくくなる (e が比較的大きい) と FGV はほぼ生き残らなくなり，代わりに TFT と AllC が共存するようになった．

参考文献

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

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

[3] ヨンジュン, 岩崎, 神取, 小原, 横尾. 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り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, pp. 1234–1246, 2012.

[4] 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.

[5] 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.