

不完全観測下における損失回避性が均衡の安定性に及ぼす影響

The Effect of Loss Aversion on Equilibrium Stability in Incomplete Observations

四辻 隆¹澤 亮治¹Yutaka Yotsutsuji¹Ryoji Sawa¹

概要

本論文では、私的不完全観測繰り返しゲームにおいて、進化的安定戦略のシミュレーション研究を行った。私的観測とはプレーヤーが相手の行動を見ることができない代わりに、相手の行動に依存して決定するシグナルを受け取る状況である。本稿ではプレーヤーが損失回避性という行動経済学的性質を持つことを仮定した。損失回避性(loss aversion)とはプレーヤーが利益よりも損失を評価する性質のことである。そのような選好のもと、修正進化的安定性の概念を用いて、どのような戦略が安定になるかを考察した。その結果、Pavlov-C、Pavlov-D戦略が、幅広いエラー率と利得の範囲で修正進化的安定戦略になることが分かった。また私的観測下の有限オートマトンゲームを Pairwise-Fermi comparison モデルに拡張し、ネットワーク上の繰り返し囚人のジレンマゲームを行い、対戦する相手が限られた集団におけるシミュレーションの結果、協調的な戦略が生き残り、幅広い利得範囲で協力が維持されることが分かった。

1 はじめに

ゲーム理論においてはプレーヤーが完全合理的であると仮定されてきたが、現実の人間は完全合理的であるとは言い難い。[1]は、プロスペクト理論において参照点依存型効用関数を定義し、人々は効用の最大化を行うのではなく、損得の感情的なバイアスの影響を受けることを考慮した意思決定モデルを示した。参照点依存とは、効用は最終的な状態によって決定されるのではなく、基準となる参照点からの移動の方向により評価されることである。プロスペクト理論では同じ額の利得と損失では損失のほうが心理的な影響が大きいことが分かっている。そのため、実際の人間は利益の最大化のみに動機づけられるのではなく、損失の最小化によっても動機づけられる、限定合理的なプレーヤーと言える。

繰り返しゲームとは、短期的な相互関係では成しえなかった協力関係が、長期的関係になると協力が達成できることを説明するモデルである。繰り返しゲームの研究において、相手の行動を観測できないが、プレーヤーたちの行動を不完全に表す、自分にしか見ることができないシグナルを観測する場合を、不完全私的観測付き繰り返しゲームと呼ぶ[2]。このゲームの均衡を部分観測可能マルコフ決定過程(Partially Observable Markov Decision Process, POMDP)を用いて安定な均衡を見つける手法が[3]によって示されている。均衡における振る舞いを、有限オートマトンを用いることで複雑な戦略を限定して表現でき、現実の人間の行動に近い分析が可能になる。しかし、有限オートマトン研究においてプレーヤーの社会的選好を考慮した研究は存在していない。[1]により人間は

合理的でなく、限定合理的であることも示されており、ゲーム理論の前提が覆されている。また、社会的選好がプレーヤーの協力行動を促進することや[4]、公平性を繰り返し囚人のジレンマゲームに導入し、公平性が高いほど協力が保たれることも示されている[5]。[6]では損失回避性を持つプレーヤーを繰り返し囚人のジレンマゲームに導入することで、協力が保たれることが示されている。このように多くの文献が、人々の長期的関係において社会的選好の協りに与える影響が大きいことを示している。これらの先行研究は囚人のジレンマに社会的選好を導入したが、完全観測を前提に理論研究を行っている。私的観測では[7]などの理論研究が行われているが、社会的選好などの意思決定バイアスを考慮した研究は著者らが調べた範囲では見られない。そのため、今回は社会的選好(損失回避性)を持ったプレーヤーを前提に、私的観測で安定なマシンを、シミュレーションを通して見つけていく。

無限回繰り返しゲームでは均衡が無数存在することが分かっているが、その均衡を絞るアプローチとして進化的安定性の分析が用いられてきた。[8]は進化的安定性の定義を用いて有限オートマトンでの分析を行った。しかしこの論文ではプレーヤーは自分たちの過去の行動を完全に観測できることを前提にしてきた。実際には自分が協力したのに相手に裏切られるといったようなこと、つまり不完全観測であることが多い。ここで、不完全観測下での繰り返しゲームは他人が過去にどのような行動をとったか直接観察できない状況を指す。今回は、[8]の定義を私的観測に拡張してシミュレーションを行う。しかし、私的観測下では進化的安定戦略は状態数よりも利得で決まることが大半を占めていた。利得を変化させ、レプリケータダイナミクスを用いた私的観測下の繰り返し囚人のジレンマでは、利得の大きさに応じて4つの社会が現れることが分かっている[9]。そのため、進化的安定戦略の定義を用いたシミュレーションにおいても利得の範囲に応じて戦略の安定性が異なってくると考えられる。そのため、今回は利得を一定の範囲で変化させ、利得の大きさが均衡の安定性に及ぼす影響も分析した。

また本稿ではネットワーク互恵モデルに Pairwise-Fermi Comparison の戦略更新モデルを取り入れて、追加シミュレーションを行った。ネットワーク互恵とは主体間の相互作用に偏りがある場合を言い、このモデルでは高い利得を獲得している隣人の戦略を確率的に模倣する。[8]での進化的安定性の定義は、全ての戦略に対する利得の大きさと状態数が一定の条件を満たすことが求められていた。しかし、現実社会においては社会全員と関わるのではなく、自身の近くにいる人とのみ関わり、相互作用を行うことが多い。これは新型コロナウイルス禍で顕著だが、地理的な距離などにより制約を受けるためだ。[10]は、一回限りの囚人のジレンマにおいて All-C 戦略と All-D 戦略の2戦略のみで格子状のグラフに配置された隣人

¹ 筑波大学大学院システム情報工学研究群

Degree Program in Systems and Information Engineering, University of Tsukuba

とゲームを行うシミュレーションを行い、その結果 All-C 戦略を取る協力者が一定割合社会に生き残ることを示した。つまり、ネットワーク上のプレーヤーが利得の高い戦略をコピーすると、囚人のジレンマでも All-C 戦略が進化することが明らかにされた。そのため、有限オートマトンを用いた囚人のジレンマゲームにおいて、プレーヤーがネットワーク上に配置されていると仮定することで、社会的選好と協力率の関係についても明らかにすることができる。また今回は私的観測を前提としているが、Pairwise-Fermi comparison モデルを用いた時に、エージェントの行動エラーがネットワーク上の協力を促進することも分かっている[11]。[11]では空間囚人ジレンマゲームにおいてエージェントは過去に自分が取った戦略に影響を受けないとされてきたが、今回はエージェントの戦略を有限オートマトンで記述し、エージェントは過去の自分の戦略と相手から受けとるシグナルが分かる状況を仮定して、シミュレーションを行った。また、社会に存在するエージェントは損失回避性を持ち一期前の利得差により影響を受けるとした。

このようなシミュレーションの結果、エラー率と利得を変化させた場合、プレーヤーが損失回避性効用関数を持つ時に Pavlov 戦略が、一般的な効用関数を持つときと比較して、より幅広いパラメータの範囲で修正進化的安定戦略となることが分かった。これにより損失回避性が協調行動を促す戦略の安定性に、正の影響を与えていると推測できる。さらに、Pairwise-Fermi comparison モデルを用いたシミュレーションでは、一般的な効用関数の時と比較し、損失回避性効用関数の時は、自分が相手に裏切られた時の利得が小さいときに協調率が高くなることが分かった。加えて、修正進化的安定戦略のシミュレーションでは生き残らなかった、All-C 戦略も生き残ることを発見した。

2 モデル

本稿では二人対称プレーヤーにおける私的不完全観測の無限回繰り返しゲームを扱う。無限回繰り返しゲームでは、プレーヤー*i* は同じステージゲームを無限期間にわたり繰り返す。プレーヤー集合を $N = \{1, 2\}$ とする。 $A_i = \{C, D\}$ をプレーヤーの行動集合とする。 $\Omega_i = \{g, b\}$ はプレーヤー*i* が行動 $a_i \in A_i$ を取った時に、相手の行動を観測する代わりに受け取るプレーヤー*i* のシグナル ω_i の集合である。行動とシグナルの関係として、 $i \neq j \in N$ とする時、 $a_j = C$ をとると、 $\omega_i = g$ 、 $a_j = D$ をとると $\omega_i = b$ となる。また、 $o(\omega|a)$ をプレーヤーが行動プロファイル a を選択したときに発生するシグナルプロファイルが ω である同時確率とする。また、プレーヤー*i* が認識できる情報である行動 a_i と、観測したシグナル ω_i に依存する認識利得を $\pi_i(a_i, \omega_i)$ で表す。例として、自分が $a_i = C$ を取っている、相手のシグナルが $\omega_i = g$ の時は、

$$\pi_i(C, g) = 3$$

また、自分が $a_i = D$ を取っていて、相手のシグナルが $\omega_i = g$ の時は、

$$\pi_i(D, g) = 4$$

のように表される。よって行動プロファイル a の時のプレーヤーの期待利得は、

$$g_i(a) = \sum_{\omega \in \Omega} \pi_i(a_i, \omega_i) o(\omega|a)$$

と表す。ここで、繰り返しゲームとは上記の標準ゲームが期間 $t = \{1, \dots, T\}$ において繰り返し行われるゲームをいう。あるプレーヤー*i* の t 期までの私的履歴をそのプレーヤーの過去の行動とシグナルで表し、

$h_i^t = (a_i^1, \omega_i^1, \dots, a_i^t, \omega_i^t) \in H_i^t := (A_i \times \Omega_i^t)$ とする。 $t+1$ 期の戦略 $a_i(h_i^t) \in A_i$ の時、無限繰り返しゲームの期待利得は以下になる。

$$f_i(s_1(h_1^t), s_2(h_2^t)) = (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} g_i(s_1(h_1^t), s_2(h_2^t))$$

また、本稿ではプレーヤーの選好として一般的な効用関数に加えて、損失回避的な場合も考察する。[6]では損失回避性と参照点を繰り返し囚人のジレンマゲームに導入し、繰り返し囚人のジレンマゲームでの損失回避性の影響を分析している。

損失回避効用関数を $v_i^t(x_i^t, r_i^t)$ と表す。[6]が定義したモデルは以下のようになる。

$$v_i^t(x_i^t, r_i^t) = \begin{cases} x_i^t & \text{if } x_i^t \geq r_i^t \\ x_i^t - \lambda_i(r_i^t - x_i^t) & \text{if } x_i^t < r_i^t \end{cases}$$

ここで $x_i^t \in \mathbb{R}$ は t 期におけるプレーヤー*i* の利得を表し、 $r_i^t \in \mathbb{R}$ は t 期におけるプレーヤー*i* の参照点を表す。参照点とは損得の境目となる心理的な基準点である。今回は私的観測であり、不確実な状況を扱うため、単純化のために参照点は一期前の利得とし、 $r_i^t = x_i^{t-1}$ で表す。また、 $\lambda_i \in \mathbb{R}$ をプレーヤー*i* の損失回避性を表し、 $\lambda_i \in (0, 1)$ とする。この値が大きいほど、損失回避的であるとする。プレーヤーの効用は今期の利得 x_i^t が一期前の利得 x_i^{t-1} よりも大きい時、 $v_i^t(x_i^t, r_i^t) = x_i^t$ で与えられ、 x_i^{t-1} よりも小さい時、 $v_i^t(x_i^t, r_i^t) = x_i^t - \lambda_i(r_i^t - x_i^t)$ となり、 λ_i で利得がマイナスとなる。つまり、今期はこれくらい利得を得られるだろうという参照点よりも、実際に得た利得が低かった時には、プレーヤーは損失を感じる。本稿では損失回避性を $\lambda_i = 0.99$ と、プレーヤーが十分に損失回避的であることを想定して、シミュレーションを行った。これは $\lambda_i = 0.5$ 程度だと、一般的な効用関数の時と比較し、修正進化的安定戦略達成の範囲が大きくは異ならなかったため、 λ_i は出来る限り 1 に近い場合、 $\lambda_i = 0.99$ を想定する。また、今回は線形の効用関数を損失回避効用関数と定義しているが、既存研究[12]などで多く用いられている。

表 1: 囚人のジレンマゲーム

1,2	$a_1 = C$	$a_2 = D$
$a_1 = C$	3,3	0,4
$a_1 = D$	4,0	1,1

表 2: 利得を一般化した囚人のジレンマゲーム

$$(g > 0, \ell > 0, |g - \ell| < 1)$$

1,2	$a_2 = C$	$a_2 = D$
$a_1 = C$	1,1	$-\ell, 1+g$
$a_1 = D$	$1+g, -\ell$	0,0

表 3:シグナル分布

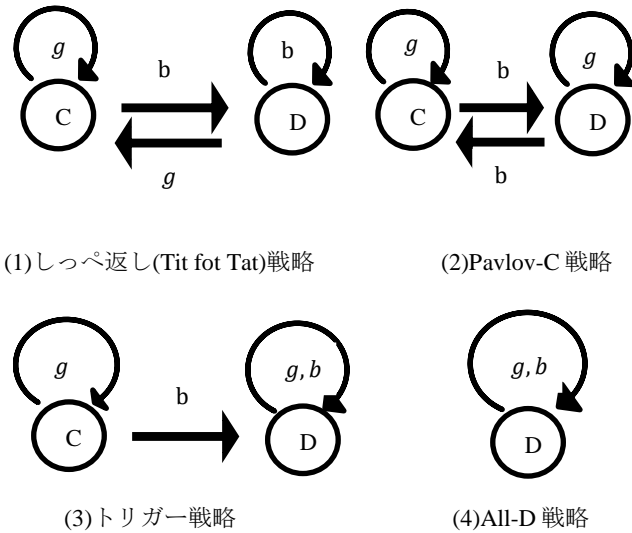
1,2	$\omega_2 = g$	$\omega_2 = b$
$\omega_1 = g$	$(1-p)^2$	$(1-p)p$
$\omega_1 = b$	$p(1-p)$	p^2

本稿では表 1 に示された四人のジレンマゲームを扱う。表中の C はプレイヤーの協力, D は裏切りを表している。今回行ったシミュレーションにおいて, エラー率を変化させたときの利得は固定して行うため, 表 1 に従う。一般化した利得表を表 2 で表す。ここで $g > 0, \ell > 0, |g - \ell| < 1$ が条件として成立する。3 つ目の条件が成り立たないとすると, 協力と裏切りを交互に出す戦略が純粋な協力よりも利得が高くなってしまい, 協力が維持できなくなる[9]。利得を変化させたときのシミュレーションは表 2 に従う。

また, 今回は私的観測なので観測にノイズが含まれる。そのため, 間違っただ信号を受け取る確率を p と表す。ここでエラー確率とは, プレーヤーに間違っただ信号が伝わってしまう独立な確率と定義する。本稿では行動プロファイル a に対して生起する私的シグナル ω の同時分布を, エラー率 p を用いて表 3 のように表す。エラー率はプレーヤーの行動に依存しない外生的なもので, プレーヤー i はこの確率を事前に認知でき, そのエラー率を考慮した期待利得を用いている。

3 有限オートマトンと戦略

図 1:有限オートマトン



(1)しっぺ返し(Tit for Tat)戦略

(2)Pavlov-C 戦略

(3)トリガー戦略

(4)All-D 戦略

本稿では以上に述べた私的観測繰り返しゲームのモデルを基に有限オートマトンを二人のプレーヤーがプレイする。ここでは各状態で取る行動と相手から受け取るシグナルで状態の遷移について定め, 毎期に取る行動を決めつける。有限オートマトンは図 1 のように表す。円の中は, オートマトンが持つ状態, 文字は自分が観測するシグナルを表している。C 状態ではプレーヤーは, $a_i = C$ を取り, D 状態では, $a_i = D$ を取る。

本稿のシミュレーションでは, 戦略として状態数 1 と状態数 2 の戦略を合わせて計 26 戦略を扱う。そのうちの

主要な戦略を 2 状態の有限オートマトンを用いて表したのが図 1 である。図 1 の(1)しっぺ返し戦略は相手が裏切った次の期は裏切りを, 相手が協力した次の期は協力を取る戦略である。また(2)の Pavlov-C 戦略は相手が協力を取り続ける限り, 自分も協力するが, 裏切りのシグナルを観測した次の期は自分も裏切る戦略である。つまり, エラーが起こり, 協力から逸脱してもすぐに協力に戻る戦略である。Pavlov-D 戦略は初期状態が D から始まる戦略であり, シグナルの構造は Pavlov-C 戦略と同様である。(3)のトリガー戦略は相手が協力を取り続ける限り自分も協力するが, 相手が裏切った時は, それ以降裏切り続ける戦略である。(4)の All-D 戦略は相手のどのようなシグナルを観測しても, 裏切り続ける戦略である。完全観測の時は, エラー率が 0 の時を言うが, しっぺ返し戦略や, All-D 戦略など多くの戦略が均衡を達成する。トリガー戦略は状態数の複雑さから均衡を達成することは無い。しかし私的観測の時はしっぺ返し戦略が均衡を達成することは無い[3]。エラー率が高い場合, 行動プロファイル(C,D)と(D,C)を繰り返し, 協力状態に復帰できず, 泥沼に陥ってしまうためだ。私的観測下で損失回避性を考慮した場合も, Pavlov-C 戦略, Pavlov-D 戦略, All-D 戦略, トリガー戦略のみが均衡を構成した。

4 修正進化的安定戦略について

有限オートマトンを用いたゲームにおいて, 進化的安定戦略の議論は, [8]が第一に利得の大きさ, 第二に有限オートマトンの複雑さを選好の優先順序として定義している。以下が修正進化的安定戦略(Modified Evolutionarily Stable Strategy: 以下 MESS)である。

プレーヤー $i \in N$ の利得を P_i , 有限オートマトンを $m \in M_i$ で表す。 $m \in M_i, i \in N$ が修正進化的安定戦略であるとは, $\forall m^* \neq m \in M_i$ において以下の不等式が成り立つことを言う。

$$\begin{cases} P_i(m, m) > P_i(m^*, m) \\ P_i(m, m) = P_i(m^*, m) \text{ and } P_i(m, m^*) > P_i(m^*, m^*) \\ P_i(m, m) = P_i(m^*, m) \text{ and } P_i(m, m^*) > P_i(m^*, m^*) \text{ and } |m| \leq |m^*| \end{cases}$$

ここで, 進化的安定戦略は修正進化的安定戦略と比較し, 定義の三つ目の不等式が $|m| < |m^*|$ となる点で違いがある。しかし, 進化的安定戦略の定義を用いた時, All-D 戦略しか進化的安定戦略にならなかったため, 定義を緩めたのが修正進化的安定戦略である。[8]では完全観測を前提としているが, 今回は定義を私的観測に拡張して繰り返し四人のジレンマゲームのシミュレーションを行った。

5 修正進化的安定戦略のシミュレーション結果

5.1 利得を変化させた時

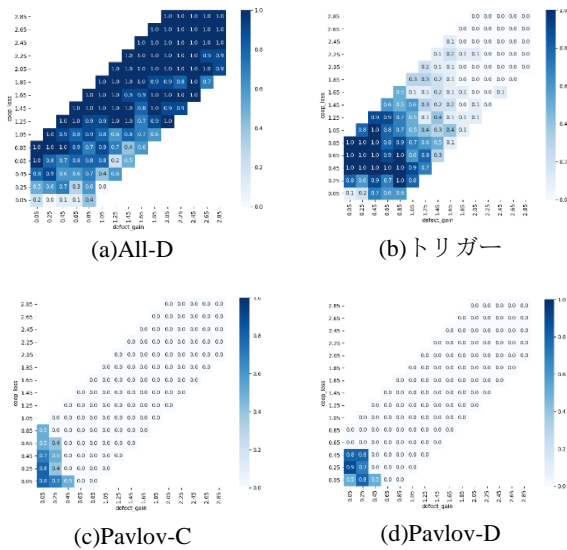
以下では利得を変化させた MESS のシミュレーション結果を, 一般的な効用関数と, 損失回避性効用関数とで分けて示す。一般的な効用関数とは損失回避性を持たない, 合理的なプレーヤーの効用関数を言う。エラー率を変化させたシミュレーションでは利得を一定とし, エラー率を変化させた場合, ほとんどの有限オートマトンの

対戦において MESS の定義から、利得の大小により MESS が決定され、状態数で MESS が決定されることは無かった。完全観測の際は、利得よりも状態数で決まることが多いが、不完全観測の際はほとんど利得の大小でマシンが MESS として選ばれた。そのため、利得を一般化し、変化させることで MESS になる範囲が変わってくると考えられる。エラー率は 0.1 で固定し、割引因子は 0.99 とする。100 期の囚人のジレンマゲームを 100 回繰り返し、それぞれのゲームの利得の期待利得を有効数字 20 桁で使用する。

以下に示す図の横軸は自分が裏切ることによる利得の増分 g を表し、図の縦軸は相手が裏切ることによる利得の損失 l を表す。利得は 0.2 刻みで [0.05, 3.00] の範囲で変化させた。各セルには同じ g と l の値で 10 回シミュレーションを行い、そのうち MESS になった割合をプロットした。各セルの色は、MESS になった割合が 1 に近くなるほど色が濃くなる。

(1) 一般的な効用関数の時

図 2: 一般的な効用関数の時のシミュレーション結果 ($\lambda_i = 0.99, \delta = 0.99, \text{エラー率 } 0.1$)



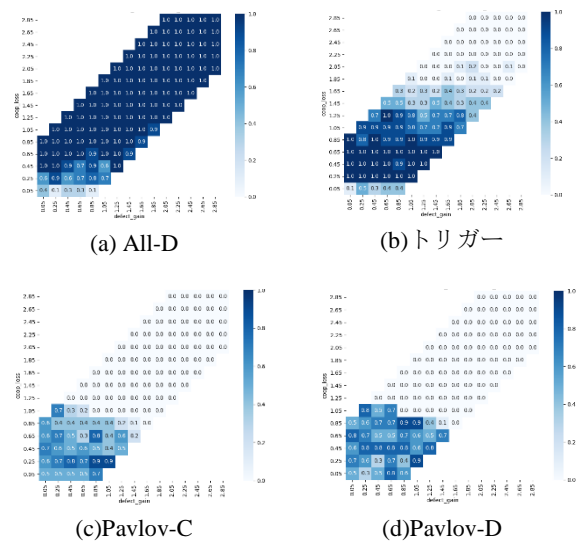
一般的な効用関数の時は Pavlov-C 戦略と Pavlov-D 戦略が g と l が小さく、 g が l よりも小さい時に MESS となる。 g が大きいと、相手を裏切った時の利得が大きくなり All-D 戦略や、トリガー戦略に有利となり、侵入されてしまう。 g が小さい時は、協力から逸脱する誘因が小さくなり、行動プロファイル $a=(C,C)$ を繰り返し高い利得を得ることで MESS を達成できる。All-D 戦略はほぼすべての g と l の値で MESS を達成できる。トリガー戦略は g と l が 1.5 程度の大きさまで MESS を達成できるが、それ以上大きくなると、All-D 戦略に侵入されてしまい MESS 不達成となる。

また All-D 戦略は g と l が大きい時、トリガー戦略は g と l が小さい時に MESS 達成度の割合が高くなる。これは g と l が大きい時は、相手を裏切ったことによる利益が大きいため、裏切る誘因が高い。そのため、常に相手に非協

力を取る、All-D 戦略が強くなる。 g と l が小さい時は相手を裏切っても利得が少なく、協力を保ったほうが高い利得を安定的に得られるため、トリガー戦略や Pavlov-C 戦略のように初期状態が状態 C から始まる戦略が MESS を達成しやすい。そのため、図 2 における All-D 戦略の MESS 達成割合は、 g と l が小さい時は低くなっているが、この時はトリガー戦略に侵入されてしまい、MESS を達成できないことが多い。[9]では、突然変異付きレプリケータダイナミクスを用いて利得を変化させ、どの戦略が生き残るかを分析している。その結果、本稿で MESS を達成した 4 戦略に加えて、All-C 戦略、OSP (one-shot punishment) 戦略、しっぺ返し戦略が、生き残っている。[9]では、Pavlov 戦略は g と l が小さい範囲でのみ生き残ったが、本稿では損失回避性効用関数の場合、より広い g と l の範囲で生き残っている。

(2) 損失回避性効用関数の時

図 3: 損失回避性効用関数の時のシミュレーション結果 ($\lambda_i = 0.99, \delta = 0.99, \text{エラー率 } 0.1$)



損失回避性効用関数の時は図 3 より、Pavlov-C 戦略と Pavlov-D 戦略が g と l が 1.5 程度までの範囲で、 g が l よりも大きい時に MESS を達成できる。一般的な効用関数の時と比較し、 g と l がより広い範囲で、 g が l よりも大きい時に MESS を達成していることが分かる。 l が大きくなると相手に裏切られた時の利得の損失が大きくなり、トリガー戦略などに侵入されてしまう。しかし、一般的な効用関数の時と異なり、Pavlov 戦略は g が 1.5 程度とある程度の大きさの時まで、MESS になる。これは例えば、自分が Pavlov 戦略で相手が All-D 戦略の時に、行動プロファイルが $a=(C,D)$ から $a=(D,D)$ に遷移し、相手の利得は $1+g \Rightarrow 0$ となり、損失回避性によりマイナスになる。 g がそこまで大きくない時はこのマイナス分が作用し、自分だけ裏切り戦略に逸脱するより Pavlov 戦略同士で行動プロファイル $a=(C,C)$ を保ったほうが、利得が高くなり Pavlov 戦略が MESS となる。All-D 戦略とトリガー戦略は一般的な効用関数の時と比較して、MESS 達成範囲に大

きな変化はないが、All-D 戦略の MESS 達成度の割合が増加していることが分かる。これは損失回避の時はプレーヤーがエラーにより何度も戦略を変更すると、損失回避性により利得がマイナスになるため、常に D を取り続ける All-D 戦略が高い利得を獲得した結果であると考えられる。

5.2 エラー率を変化させたとき

以下では修正進化的安定戦略のシミュレーション結果を示す。ここではエラー率 $p \in (0,1)$ を変化させた時のシミュレーション結果を一般的な効用関数(normal)と損失回避性効用関数(loss)の時とで分けて示す。割引因子のパラメータは $\delta = 0.99$ で固定して行い、26 戦略でシミュレーションを行った。MESS にならなかったマシンについては省略する。500 期の囚人のジレンマゲームを 100 回繰り返し行い、それぞれのゲームの利得の期待利得を有効数字 20 桁で使用し、修正進化的安定戦略の定義を用いて安定的な戦略を見つけるシミュレーションを行った。また、今回は不確実性を伴うのでシミュレーション結果に多少の誤差が出た。そのため、同様のシミュレーションを 10 回行い、MESS となる判定が 9 割以上で○、1 割より大きく 9 割未満で△、1 割以下で×という基準を設け、表にまとめた。

表 4:エラー率を変化させた時のシミュレーション結果 ($\lambda_i = 0.99, \delta = 0.99$)

	Pavlov-C		トリガー		Pavlov-D	
	loss	normal	loss	normal	loss	normal
0.01	○	○	△	△	○	○
0.02	○	○	○	△	○	○
0.03	○	○	○	○	○	○
0.04	○	○	○	○	○	○
0.05	○	△	○	○	○	○
0.06	○	△	○	○	○	○
0.07	△	×	○	○	△	×
0.08	△	×	○	○	△	×
0.09	△	×	○	○	△	×
0.1	△	×	○	○	△	×
0.11	△	×	○	○	△	×
0.12	△	×	○	○	△	×
0.13	△	×	○	○	△	×
0.14	△	×	○	○	△	×
0.15	×	×	△	○	×	×
0.16	×	×	△	○	×	×
0.17	×	×	△	△	×	×
0.18	×	×	△	△	×	×
0.19	×	×	×	△	×	×
0.2	×	×	×	×	×	×

(9 割以上 MESS : ○, 1 割より大きく 9 割未満で MESS : △, 1 割以下で MESS : ×)

表 4 から、一般的な効用関数の場合、Pavlov 戦略は $p=0.06$ 以上になると MESS を達成できないが、損失回避性効用関数の場合、Pavlov-C 戦略と Pavlov-D 戦略が、高

いエラー率の場合であっても 1 割より大きく 9 割未満の割合で MESS を達成したことが分かる。損失回避性効用関数の時に、Pavlov 戦略が高いエラー率であっても MESS を達成できた理由として、Pavlov 戦略同士から、自分だけトリガー戦略や All-D 戦略に変えたとしても、最初は行動プロファイル $a=(D,C)$ により利得 4 を得られるが、その後、Pavlov 戦略を取る相手も裏切るので行動プロファイル $a=(D,D)$ となり、利得 1 となる。一期前の利得 4 と比較して今期の利得が大きく減少したため、損失回避性により効用が大きくマイナスになってしまい、結果として、自分だけ戦略を変えても得をせず、協力から逸脱する誘因が低い。Pavlov 戦略同士で安定的に行動プロファイル $a=(C,C)$ による利得 3 を得たほうが最終的な利得の合計が大きくなる。

All-D 戦略と、トリガー戦略は損失回避性効用関数と一般的な効用関数の場合で MESS 達成に大きな違いは見られなかった。トリガー戦略はどちらの効用関数の場合でもエラー率が低い場合にしばしば MESS を達成できないことがあった。これは、トリガー戦略は、エラーが起これば、D 状態に遷移することで、All-C 戦略などから行動プロファイル $a=(D,C)$ による利得 4 を得られ、高い利得を達成できる。しかし、エラー率が低いと D 状態に遷移する確率が低く、高い利得を得られずほかのマシンに侵入されてしまう。エラー率が高くなると、All-D 戦略に侵入され、MESS を達成できない。All-D 戦略はエラー率によらずに、どちらの効用関数においても常に 9 割以上 MESS を達成したため、表では省略する。

6 私的観測下での Pairwise-Fermi Comparison モデルシミュレーション

以下では、ネットワーク上で私的観測下の繰り返し囚人のジレンマを行う時のモデルを示す。プレーヤー a はネットワーク上で結ばれた全隣人と囚人のジレンマゲームを行い、その合計利得 π_i を得る。そして Pairwise-Fermi Comparison モデル(以下 PF モデル)を用いて自分の隣人のうち、一人をランダムに選択し、選ばれた隣人が自分よりも高い利得を獲得していればいるほど、より高い確率でその相手の戦略をコピーする。以下の式がモデルを表す。

$$P_{i \leftarrow j} = \frac{1}{1 + \exp\left[\frac{\pi_i - \pi_j}{\kappa}\right]}$$

κ は温度係数と言われるパラメータを表しており、プレーヤーたちが対戦相手との得点差に対してどれくらい敏感に反応するかを表している。ここでは $\kappa = 0.1$ とする。

6.1 Pairwise-Fermi Comparison モデルでのシミュレーション

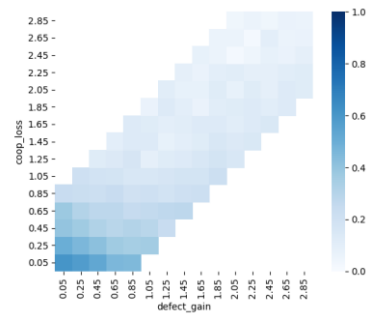
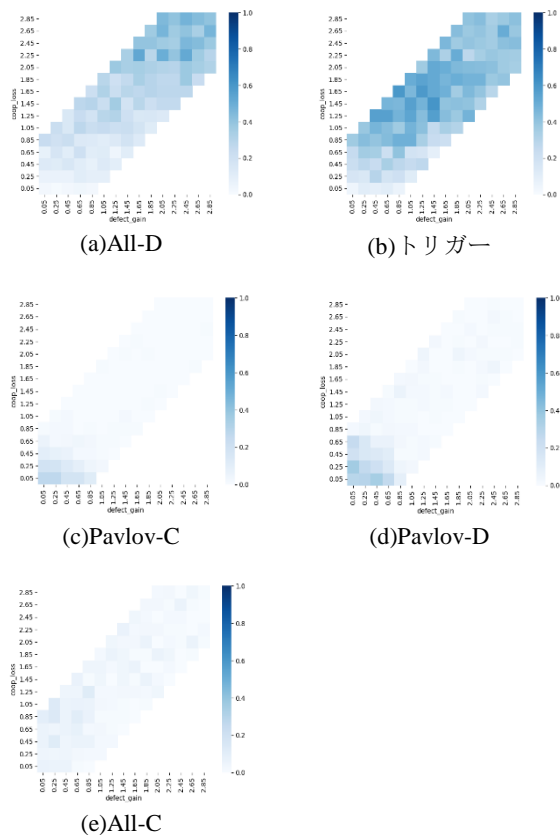
以下ではネットワーク上における私的観測下の繰り返し囚人のジレンマゲームのシミュレーション結果を示す。MESS のシミュレーションと同様に、全 2 状態以下の有限オートマトン 26 戦略でシミュレーションを行った。プレーヤーの隣人の数を表す平均次数は 8 とし、ネットワーク構造は格子グラフ(Lattice)とする。ネットワーク上に存在するプレーヤーの数は $676 (= 26^2)$ とし、初期

には26人ずつ各戦略を取るエージェントが存在している。プレイヤー達は私的不完全観測の下で、ネットワーク上で結ばれた隣人と囚人のジレンマゲームを行う。相手のシグナルを見間違えるエラー率は0.1で固定し、割引率はMESSのシミュレーションと同様に、0.99とする。利得は表2の一般化した囚人のジレンマゲームに従い、利得の範囲は0.2刻みで[0.05,3.00]の範囲で変化させた。

各セルには社会に存在する全プレイヤーのうち、該当する戦略を取る割合をプロットした。各プレイヤーは隣人と100期の繰り返し囚人のジレンマゲームを行う。このゲームを全プレイヤーが行い、これを1時間ステップとする。各時間ステップで、全プレイヤーがPFモデルに従い、戦略を更新していく。今回は50時間ステップまでシミュレーションを行うとする。今回はこのシミュレーションを10回行い、平均を取った。シミュレーションの収束判定は、協力率の変化率が 10^{-3} 以下となった時点で収束とし、また協力率が、0もしくは1となった時点でも収束と判定する。50時間ステップまでに収束と判定されなかった場合は、各期の協力率の平均値を取り、その時点で計算を終了する。

6.2 一般的な効用関数の時

図4: 一般的な効用関数の時のシミュレーション結果 ($\lambda_i = 0.99, \delta = 0.99, \text{エラー率} 0.1$)



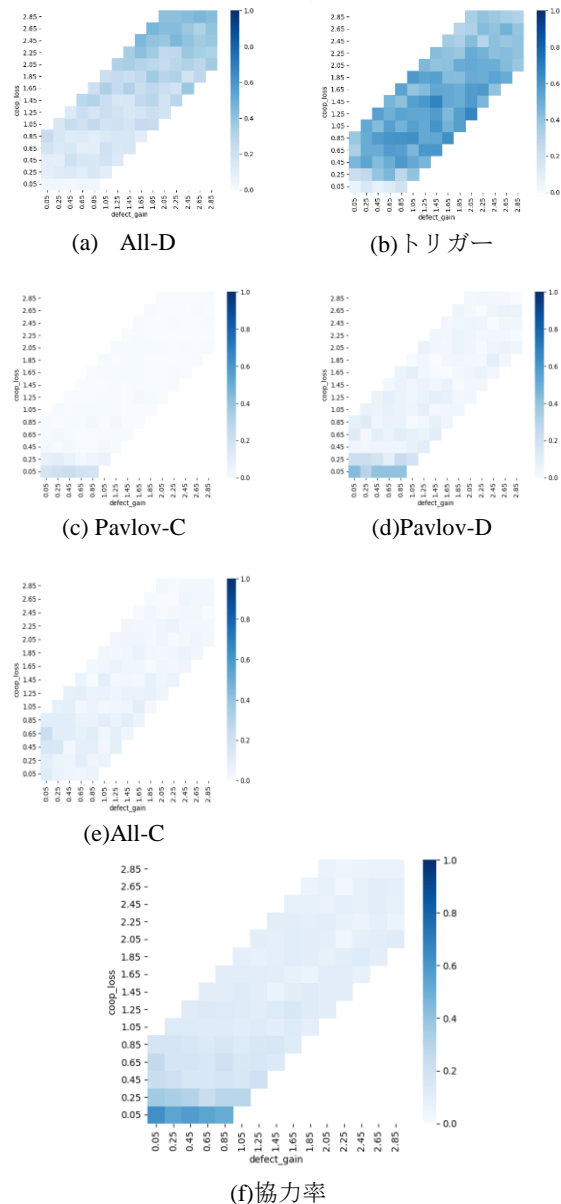
(f) 協力率

一般的な効用関数の時は、図4に示す結果となった。協力率は g と l が比較的小さいときに、高い協力率になった。戦略としてはPavlov-C, Pavlov-D戦略, トリガー戦略, All-D戦略, All-C戦略が高い割合で模倣される結果となった。MESSのシミュレーションではAll-C戦略が生き残ることはなかったが、PFモデルでのシミュレーションでは少ない割合であるが、生き残る結果となった。全有限オートマトン26戦略中、協力C状態から始まる戦略は13戦略である。All-C戦略の周囲に協力的な戦略が存在していれば、All-C戦略はエラー率によらず高い利得を獲得できるため、ほかの戦略をとるプレイヤーから戦略を真似される可能性が高い。同様に、D状態とC状態を行き来できる、Pavlov-C戦略とPavlov-D戦略も少ない割合であるが、生き残っている。つまり私的観測下においても、ネットワーク互惠の効果が働いていると考えられる。[10]のシミュレーションにおいてもネットワークの世界では、進化と学習の結果、協力プレイヤーが一定の割合で生き残ることが分かっている。私的観測下では、協力プレイヤーは相手の行動を見間違えることで裏切りプレイヤーによる搾取を防ぐことができるため、こうした観測エラーが協力者にとって有利に働く。したがって、私的観測という状況とネットワーク構造により、協力的な戦略が生き残っていると考えられる。

図4-(b)から、トリガー戦略は g と l が幅広い範囲で生き残っている。MESSシミュレーションにおいて、トリガー戦略は g と l が1.5以上になるとAll-D戦略などに侵入されて生き残ることが無かった。しかし、ネットワーク上の囚人のジレンマになると周囲にAll-D戦略が存在しないことが起こり得る。そのような状況下で、トリガー戦略は高い利得を獲得し、多くの戦略から真似されたと考えられる。

また図4-(e)より、All-D戦略は g と l がかなり大きいときに生き残っている。MESSシミュレーションでは、All-D戦略は幅広い利得の範囲で生き残っていた。ネットワーク上のシミュレーションでは、 g と l が低いと、相手を常に裏切るという厳しい態度を取るよりも、トリガー戦略のように相手が裏切るまでは自分も裏切らない戦略や、Pavlov戦略のようにD状態からC状態に復帰できる寛容な戦略が、行動プロファイル $a=(C,C)$ により高い利得を得て、周囲のプレイヤーに真似されることが多い。そのため、All-D戦略は g と l が低いと生き残ることは少ない。

6.3 損失回避性効用関数の時

図5: 損失回避性効用関数の時のシミュレーション結果 ($\lambda_i = 0.99, \delta = 0.99, \text{エラー率 } 0.1$)

損失回避性効用関数の時は、図5に示す結果となった。協力率は l が0.05と低い時に、高い協力率になった。

戦略としては一般的な効用関数の時と同様に、Pavlov-C、Pavlov-D戦略、トリガー戦略、All-D戦略、All-C戦略が高い割合で模倣される結果となった。Pavlov戦略は g が大きく l が小さい時に高い割合で生き残った。この傾向はMESSシミュレーションの時と同様であるが、生き残る範囲が図3-(c)と比較し、小さくなっている。

All-D戦略は g と l が大きいときに高い割合で生き残っている。MESSの時はAll-D戦略は1に近い割合でMESSを達成したが、PFモデルでのシミュレーションではトリガー戦略のほうが生き残る割合が高い結果となった。

MESSシミュレーションでは g と l が1.5以上になると裏切り戦略がMESSを達成し、協力的な戦略がMESSを達成することは無かったが、プレイヤーが隣人とのみ相互作用するネットワーク上での関係を考えて時は g と l が低い場合、寛容な戦略が生き残り、 g と l が大きいても一定割合の協力が保たれることが分かった。

6.4 一般的な効用関数と損失回避性効用関数の比較

一般的な効用関数の時と比較し、損失回避性効用関数の場合、協力率は l が低いときに高くなっている。この傾向は図5-(c)、図5-(d)からPavlov戦略にもみられる。これは、 l が大きくなると損失回避性を持つ場合、自分が裏切られた時のマイナス分が損失回避性の分だけ大きくなり、トリガー戦略やAll-D戦略に侵入されてしまうためである。 l が小さいと、Pavlov-C戦略やAll-C戦略など、協力的な戦略が隣人との対戦により高い利得を獲得し、生き残るため協力率も高くなると考えられる。また、損失回避性効用関数の場合、 g が高くなると協力率が高くなっている。 g が高い場合、All-D戦略やトリガー戦略などの相手を厳しく罰する戦略にとって、相手を裏切った時の利得が大きくなり g が高いと協調率が低くなるように思える。しかし、プレイヤーが損失回避性を持つ場合、 g が高いと、裏切り戦略にとっては、行動プロファイル $a=(D,C)$ から $a=(D,D)$ に遷移した時の利得のマイナス分が損失回避性によりかなり大きくなる。そのため、 g が高く l が低いと裏切り戦略をとるプレイヤーは高い利得を得られず、ほかの協力的な戦略を真似する可能性が高い。 l が0.05よりも大きくなると、裏切られた時の損失が大きくなり、徐々に協力率が下がっていくことが図4-(f)や、図5-(f)から分かる。

結果として、一般的な効用関数では、 g と l が低いとき、損失回避性効用関数では、 l が低いときに協力的なプレイヤーが多くなることが分かった。

7 おわりに

本論文は、私的観測下における修正進化的安定戦略を見つけるシミュレーションをエラー率と利得を変化させて行い、プレイヤーが損失回避性効用関数を持つ時に、Pavlov戦略が幅広い利得とエラー率の範囲で修正進化的安定戦略になることを発見した。Pairwise-Fermi comparisonモデルを用いたネットワーク上のシミュレーションでは私的観測下においてもAll-C戦略が生き残り、修正進化的安定戦略のシミュレーションと比較して、協力的な戦略が生き残ることを発見した。また、損失回避性効用関数の場合、一般的な効用関数の場合と比較し、 l が小さいときにPavlov戦略が生き残り、協力率も高くなることが分かった。今後の課題として、シミュレーションの繰り返し回数を増やすことにより、精度を上げる必要がある。またネットワークの構造を変えることや、エラー率がネットワーク上の協調率に与える影響を分析することなどが考えられる。

謝辞

小池淳平氏にはシミュレーションプログラムについて有益な助言を頂きました。心より感謝申し上げます。

参考文献

- [1] Kahneman, D. and Tversky, T., "Prospect Theory: An Analysis of Decision under Uncertainty", *Econometrica*, vol.47, pp.263-291(1979).
- [2] 神取道宏, "人はなぜ協調するのか-繰り返しゲーム理論入門-", 三菱経済研究所(2015).
- [3] 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, pp1234-1246(2012).
- [4] Fehr, E. and Schmidt, K. M., "A Theory of Fairness, Competition, and Cooperation", *Quarterly Journal of Economics*, CXIV, 817-868(1999).
- [5] Duffy, J. and Muñoz-García, F., "Patience or Fairness? Analyzing Social Preferences in Repeated Games", *Games*, 3(1): 56-77(2012).
- [6] Shalev, J., "Loss aversion equilibrium", *International Journal of Game Theory*, vol. 29(2), pp. 269-87 (2000).
- [7] Piccione, M., "The repeated prisoner's dilemma with imperfect private monitoring", *Journal of Economic Theory*, 102,(2002).
- [8] Binmore, K. and Samuelson, L., "Evolutionary Stability in Repeated Games Played by Finite Automata", *Journal of Economic Theory*, 57, 278-305 (1992).
- [9] 西野上和真, 五十嵐瞭平, 岩崎敦, "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講演論文集, 第2分冊, p27-p34 (2020).
- [10] May, R. M. and Nowak, M. A., "Evolutionary games and spatial chaos", *Nature*, Vol.359, pp.826-829(1992).
- [11] Alam, M., Nagashima, K., and Tanimoto, J., "Various error settings bring different noise-driven effects on network reciprocity in spatial prisoner's dilemma", *Chaos, Soliton and Fractals*, 114, pp. 338-346(2018).
- [12] Köszegi, B. and Rabin, M., "Reference-Dependent Risk Attitudes", *American Economic Review*, 91 (A): 1047-7(2007).