# 持続的相互協力状態発現の判定基準

塩村　尊[*1]　大田　圭佑[*2]

**要　旨**

本稿の目的は，繰り返し囚人のジレンマゲームにおける持続的相互協力状態の発現を判定するために３つの基準を設けてその判定を自動化するとともに，初期戦略分布の偏りを前提としたゲームを行い，環境に依存しない戦略評価の下，通常のエリート保存と強エリート保存，２ついずれかの意味でのエリート保存を学習過程としての遺伝的アルゴリズムと併用するならば持続的相互協力状態が発現する可能性が大きくなることを数値的に確認することにある．特に，強エリート保存を併用するとき，初期戦略分布に１人でもGRIM戦略者が存在するならば持続的相互協力状態が必然の結果となることを確認する．合わせて，環境に依存する戦略評価の下では持続的相互協力状態への進化は困難であることも確認する．もとより相互協力状態発現の定義に関して研究者間で共有されたものがなく，我々が与えた判定基準についても若干の曖昧性は残っている．しかしながら，判定を自動化したことは数値実験結果の解釈の時間短縮と数値データに基づく客観的な分析に寄与することになるであろう．

キーワード：繰り返し囚人のジレンマゲーム，遺伝的アルゴリズム，マルチエージェントモデル

# Three Criteria on Judging Sustainable Cooperation

Takashi SHIOMURA, Keisuke OHTA

**Abstract**

This paper presents three criteria that can automatically judge whether sustainable cooperation finally emerges in an iterative Prisoner’s Dilemma game. Using the three criteria, we numerically confirm that the potential for developing a sustainably cooperative state increases, if we use simple genetic algorithm in conjunction with elite preservation in either of the two meanings as a learning process: One is usual elite preservation and the other referred to elite preservation *in the strong sense*. The game has a bias in the initial distribution of strategies and evaluates the player’s payoff independently of the environment. We emphasize that sustainable cooperation becomes inevitable if we use elite preservation in the strong sense, when there is at least one grim trigger strategy in the initial distribution. Moreover, we

[*1] 関西大学総合情報学部　[*2] 関西大学総合情報学部生（2017年３月卒業）

demonstrate the difficulty of sustainable cooperation under the environment-dependent evaluation. Since there is no common definition of sustainable cooperation among researchers, our criteria might indicate some ambiguity. The criteria, however, seem to contribute to shortening the hours for interpreting the results of numerical calculations and analyzing them objectively.

Key words: iterative Prisoner's Dilemma game, genetic algorithm, multi-agent model

## 1．はじめに

文献［1］と［2］の出版をきっかけとして，これまでさまざまな相互協力関係の自発的発現可能性に関する研究が行われてきた．特に，Axelrod[3]－[5]が得た結論，すなわち繰り返し囚人のジレンマ（以下 IPD）ゲームにおける TFT 戦略（初回は協調的な戦略をとり，2 回目以降は前回，対戦相手のとった戦略を採用する行動計画）の平均的優位性と互恵的協力関係の発現可能性をめぐる議論は今現在においても精力的になされている．こうした一連の研究の結果，Axelrod の IPD ゲームに関するシミュレーション結果は様々な条件に依存しており，長期記憶と先見能力を前提としない限定合理性の観点に立つ進化論的アプローチをとる限りにおいて，集団内に全面協力と全面裏切りの永続的振動が起こる可能性があることはよく知られている．たとえば，Hoffmann[6]，Lindgren[7]，および Nowak and Sigmund[8]を参照されたい．そこで問題になるのが，1 度構築された全面協力状態が Pareto 劣位な全面裏切り状態に移ってしまうことを回避する策の検討である．

塩村，藤本[9]は Axelrod[10]や Hoffmann[6]と同様に遺伝的アルゴリズム（以下 GA）を 1 つの学習過程とみなし，IPD ゲームにおいて 1）エージェント（プレイヤー）は過去最高のパフォーマンスを示した戦略を記憶にとどめ続け，かつ 2）今現在観察できる環境に適応するだけではなく，考えられるすべての可能性を考慮した上で戦略を評価するという環境非依存型戦略評価の下，ほぼ確実に環境変化と突然変異に対して頑強な全面協力状態へと集団を導くことが可能であることを示した．また，彼らの分析はプレイヤー全員が環境非依存型戦略評価を行うという前提の下，持続的相互協力状態が発現するためには戦略分布の均一性は必ずしも必要ではなく，初期戦略分布において総当たり戦における互恵的最優秀戦略，すなわち GRIM 戦略がごくわずかでも存在すれば十分であることを示唆している．ここで GRIM 戦略とは，相手に裏切られないかぎり協調的行動をとるが，一度でも裏切られると以後は相手を裏切り続ける行動計画のことであり，GRIM TRIGGER 戦略とも呼ばれる．

本稿の目的は，繰り返し囚人のジレンマゲームにおける持続的相互協力状態の発現を判定するための基準を設けてその判定を自動化するとともに，初期戦略分布の偏りを前提としたゲームを行い，環境に依存しない戦略評価の下，遺伝的アルゴリズムとともに補完アルゴリズムとしてのエリート保存を併用するならば持続的相互協力状態が発現する可能性が大きくなること

を数値的に確認することにある．我々は通常のエリート保存に加えて強エリート保存と呼ばれる２タイプのエリート保存を検討するが，特に，後者を併用するならば，初期戦略分布に１人でもGRIM戦略者が存在すると進化の過程において100％持続的相互協力状態が発現することを確認する．合わせて，環境に依存する戦略評価の下では持続的相互協力状態への進化はきわめて困難であることも確認する．もとより相互協力状態発現の定義に関して研究者間で共有されたものがない．それゆえ，我々が与えた判定基準に関しても曖昧性は存在するかもしれないが，判定を自動化したことは数値実験結果の解釈の時間短縮と数値データに基づく客観的な議論に寄与することになるであろう．

## ２．数値実験の設定

本稿では次のようにしてGAを実装している．集団中の各個体は以下に述べる26の戦略評価のための対戦相手（試験体）と囚人のジレンマゲームを行うものとする．このゲームにおいては対戦する２人のプレイヤーそれぞれが強調と裏切りの２戦略を持っている．以下では協調と裏切りを各々，$C$ と $D$，または１と０で表す．

各個体は前回プレイの結果，すなわち前回実現した２人の戦略の組を記憶しており，その結果に応じて今回とるべき戦略が遺伝子に組み込まれている．ただし，初期状態においては前回プレイの結果がランダムに与えられている．このとき個体は記憶部を含めて遺伝子長７のビットストリングにより表現することができる．図１はこれを表したものであり，各遺伝子座の対立遺伝子は｛0，1｝とする．個々の遺伝子座の意味については表１を参照されたい．先頭２つの遺伝子座からなる記憶部を除く５ビットで表現可能な戦略の数は$32=2^5$あるが，このうち６戦略は他の戦略と実質的に同じものである（文献[10]，p.409，表A.1参照）．

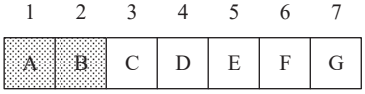



図１　戦略の遺伝子表現：先頭２ビットは記憶部

表１　遺伝子座の意味

| 遺伝子座 | 意味 |
|---|---|
| 1 | A：前回プレイの自身の手 |
| 2 | B：前回プレイの相手の手 |
| 3 | C：初回の自身の手 |
| 4 | D：記憶部が（$C$, $C$）＝（1, 1）の場合の自身の手 |
| 5 | E：記憶部が（$C$, $D$）＝（1, 0）の場合の自身の手 |
| 6 | F：記憶部が（$D$, $C$）＝（0, 1）の場合の自身の手 |
| 7 | G：記憶部が（$D$, $D$）＝（0, 0）の場合の自身の手 |

各個体（戦略），あるいはエージェントは，あらかじめ定められた回数試験体と囚人のジレンマゲームを繰り返すものとして，彼が獲得した利得の平均値をもって戦略の評価値，すなわち適応度とする．

適応度を評価する方法はいくつかあるが，本稿では次の２つの方法を採用した．１つは，各エージェントが集団からランダムに選ばれた26個体を試験体として，これら試験体とあらかじめ定められた回数だけ「実際に」対戦して，その成績をもって優秀な戦略を探し出す方法である．今ひとつは，各エージェントが表現型の異なる26個体を試験体としてあらかじめ定められた回数だけ「頭の中で」対戦して試行錯誤により優秀な戦略を探るものである．この場合，利得の平均値はエージェントが考える戦略の評価値にすぎないことに注意されたい．塩村，藤本[9]は前者と後者を各々，環境依存型，環境非依存型評価と呼んでいるが，これら２つの評価方法の決定的な違いは，戦略評価がその時々の集団中の戦略分布に依存するか否かである．

各エージェントの戦略評価は観察，またはコミュニケーションを通じて集団中の全エージェントに伝わるものとして，続けてルーレット選択，すなわち適応度に比例した確率で次世代の個体を生み出す父親と母親を決定し，これら両親の遺伝子を受け継ぐ子供を一様交差により作り出す．さらに個々の子供に対して，ランダムに選ばれた１つの遺伝子座のビットをあらかじめ定められた確率（突然変異確率）で反転し，突然変異体を作り出すものとする．これらの選択と遺伝的操作はあらかじめ定められた数の子供が誕生するまで繰り返される．以上の操作が完了するまでの時間（ステップ）を１世代と定義する．

### 2.1 GA の設定

上に述べたように，本稿で実装した GA は文献［９］とほぼ同じであるが，次の１点において異なる．すなわち，初期状態において GRIM 戦略をとる子供の数があらかじめ指定されており，残りすべての子供の戦略は GRIM 戦略とは表現型の異なる25戦略からランダムに決定される．

表２は我々が用いた成分ゲームの利得表であるが，これは文献［６］と同じものである．割引因子と試験体との対戦回数は各々，1.0，100とし，交差確率と突然変異確率はそれぞれ，0.5と0.001で与えられている．戦略（子供）の数は104＝26× 4 であり，この数は世代を通じて維持される．

表２　成分ゲームの利得表：プレイヤーＡとＢは各々，子供と試験体

| A＼B | *C* | *D* |
|---|---|---|
| *C* | (0.6, 0.6) | (0, 1) |
| *D* | (1, 0) | (0.2, 0.2) |

数値実験は戦略評価方法が環境依存型と非依存型の各々について，ルーレット選択のみを用いた場合，これに通常のエリート保存を併用した場合，および強エリート保存を併用した場合

を検討した．ここで一般に，エリート保存とはGAの補完アルゴリズムであり，過去最高の評価値を持つ戦略，すなわちエリートを次世代に強制的に残す方法である．通常のエリート保存はエリートを遺伝的操作の結果生まれた子供の中に挿入するが，強エリート保存は次世代の子供を生み出す親の1人に設定する．これらを人の学習過程とみなすならば，前者は過去最高評価の案を強制的に残すことを意味し，後者は過去最高の案と直近の優れた案から新しい案が生み出されることを意味する．強エリート保存の詳細については文献[9]を参照されたい．

### 2.2 持続的相互協力状態発現の判定

今，$t$世代における集団の平均利得を$x(t)$で表し，過去最高の集団の平均利得を$x_{max}(t)$で表す．便宜上，集団を構成するエージェントの多くが採用する戦略を代表的戦略と呼び，過去最高の集団の平均利得をもたらす代表的戦略を最優秀戦略と呼ぶが，これは必ずしも我々が期待する互恵的戦略ではないことに注意されたい．

まず，集団が最優秀戦略に達した回数を$s$として$s=0$で初期化し，次の条件1をチェックする．この条件が満たされたとき，$t$世代において集団を構成するエージェントの多くが最優秀戦略を採用していると判断する．

**条件1**：$\varepsilon > 0$を十分小さな数として

$$\left| \frac{x(t) - x_{max}(t)}{x_{max}(t)} \right| \leq \varepsilon$$

読者は，条件1において我々が過去最高の個人の利得と比較しなかったことを疑問に思われるかもしれない．容易に想像できるように，過去最高の個人の利得は全面裏切り戦略が全面協調戦略と対戦したときに獲得できる．それゆえ，もし過去最高の個人の利得との相対誤差をチェックすると，集団が「ほぼ」全面裏切り状態にあるときにのみに条件が満たされることになり，集団が全面協力状態にあるか否かを判定しようとする我々の意図に反することになる．条件1が過去最高の集団の平均利得との相対誤差をチェックしている理由はこれである．

さて，議論の単純化のため，すべてのエージェントが同じ最優秀戦略を採用していると仮定すると，各回のゲームで実現する戦略の組は（$C$，$C$）か（$D$，$D$）のいずれかになる．それゆえ，初回の手が$C$であったとするならば，次回は両プレイヤーともに遺伝子座4の戦略を採用する（表1参照）．これが$C$であったとするならば，以後は相互協力状態が続くことになる．逆に遺伝子座4の戦略が$D$であったとするならば，両プレイヤーともに3回目のゲームで遺伝子座7の戦略を採用するはずであるが，これがもし$C$であったとするならば相互裏切りと相互協力の循環が起こる．他方，$D$であったとするならば相互裏切り状態が続くことになる．

一方，初回の手が$D$であったと仮定すると，次回は両プレイヤーともに遺伝子座7の戦略を採用するはずである．これがもし$D$であったとするならば，以後は相互裏切り状態が続くこと

になる．他方，$C$であったとするならば，３回目のゲームで遺伝子座４の戦略を採用するが，これが$C$であったとするならば，以後は相互協力状態が続く．もし$D$であったとするならば，相互裏切りと相互協力の循環が起こる．

結局，最終的に持続的相互協力状態へと導く戦略は遺伝子型で，（11***）であるか，（01**1）である．ただし，先頭２つの遺伝子座（記憶部）は除いている．また，＊は０か１のいずれかを表している．そこで，最優秀戦略が互恵的なものであるか否かを判定するために次の条件２を考える．

**条件２**：$t$世代における集団の各遺伝子座の協調戦略出現頻度を格納したベクトルを$\Xi(t) \equiv (\xi_i(t))$，$1 \leq i \leq 7$で表す．また，$\delta > 0$を十分小さな数とする．このとき

$$\left|\xi_3(t) - 1\right| + \left|\xi_4(t) - 1\right| < \delta$$

または，

$$\left|\xi_3(t) - 0\right| + \left|\xi_4(t) - 1\right| + \left|\xi_7(t) - 1\right| < \delta$$

が成立する．ここで，協調戦略出現頻度とは遺伝子が１になっている子供の比率である．

$T_{max}$を最大世代数，すなわち最大計算回数として，十分大きな世代，$T_{test}$（$T_{max} > T_{test} > 0$）以降の各$t$世代で条件１と２が「同時に」満たされるならば上に定義した変数，$s$を１加算することとしよう．$t > T_{test}$で２条件をチェックする理由は我々の関心が最終的な戦略の進化にあるからである．このとき，変数$s$は集団が互恵的最優秀戦略に達した世代数を記録することになる．

最後に下の条件３をチェックし，もしこれが満たされるならば最終的に集団は持続的相互協力状態に達したと判定する．

**条件３**：予め定めた，１以下ではあるが十分１に近い数，$p > 0$について

$$\frac{s}{T_{max} - T_{test}} \geq p$$

図２は上の持続的相互協力状態発現判定のアルゴリズムをフローチャートで表したものである．Trueで初期化されたフラグ，flagは互恵的最優秀戦略への進化を判定するためのものであり，条件３が満たされなかったときにFalseになる．もし，flagが初期値，Trueのままであるならば，我々は集団が持続的相互協力状態に進化したと判定する．

### 2.3 パラメータ設定

以上の手続きにおいて，予め指定しておかなければならないパラメータは$T_{test}$，$T_{max}$，$\varepsilon$，$\delta$，および$p$の５つであるが，これらの値をどのように決めるかについて明確な基準はない．この点

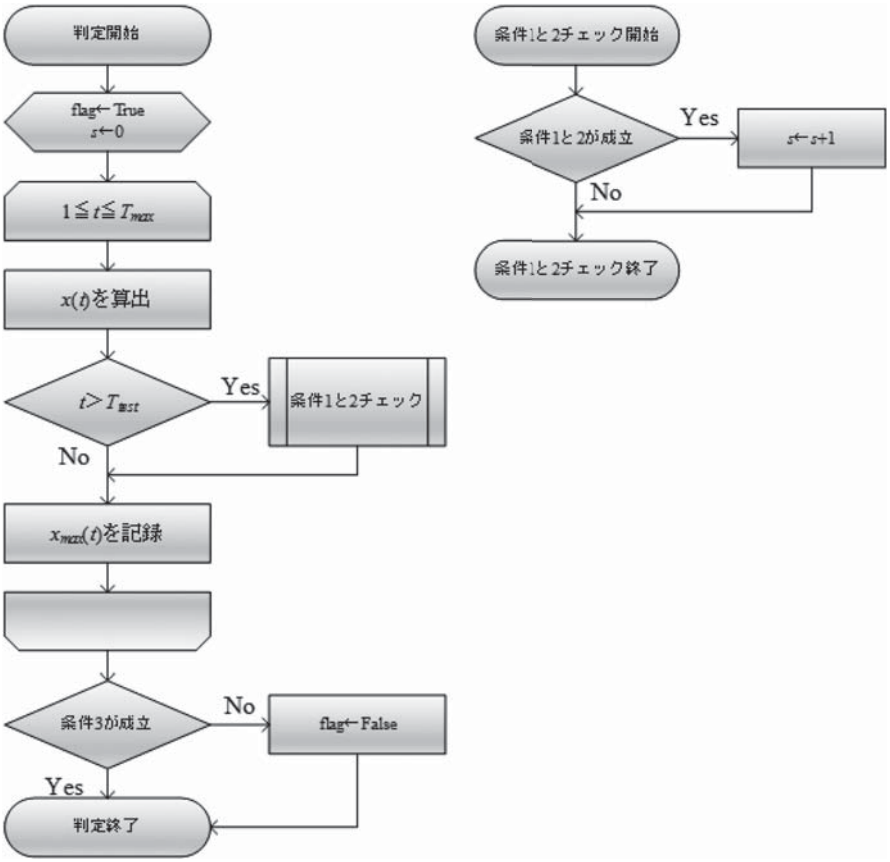



図２　持続的相互協力状態発現判定アルゴリズム

については経験的に定める他はないように思われる．

いうまでもなく，数値実験は繰り返し行う必要がある．加えて，我々の関心は最終的に相互協力状態が発現するか否かにあるので，最大世代数も十分大きくとる必要がある．他方，文献［９］と同じ設定の下，現時点でのPCの処理速度を考えるならば，最大世代数，$T_{max}$は５万が限界であるように思われる[1)]．それゆえ，本稿では$T_{max}$＝５万に固定する．

我々のこれまでの数値実験結果によれば，戦略評価方法が環境依存型，非依存型にかかわらず，もし最終的に相互協力状態に進化するとするならば，それはGRIM戦略であることが多い．この場合の最大平均利得は環境依存型評価の場合は0.6，非依存型評価の場合は約0.641になるが，条件１の許容相対誤差を決定するに当たり，子供104人中の５％程度のGRIM戦略からのズレを許容することにしよう．

---

1)　Intel Core i7−3770CPU 3.40GHz，クアッドコア，メモリ32GB，Windows 10 64bit環境の下でも１回の数値実験に約７時間かかる．なお，プログラムはVisual Basic 2015で記述されている．

表3は環境依存型と非依存型の場合における計算結果の1例であるが，遺伝子座列の数値は協調戦略出現頻度，すなわち対応する遺伝子座の遺伝子が協調戦略になっている子供の104人中の比率である．ここで，GRIM戦略の記憶部を除く遺伝子型は（11000）であることに注意されたい．表中第1行の環境依存型に関するシミュレーション結果は遺伝子座4がGRIM戦略と異なるものが集団中，3.8%存在していることを示している．他方，第2行の非依存型に関するシミュレーション結果は遺伝子座7がGRIM戦略と異なるものが集団中，3.8%存在していることを示している．それゆえ，最後の列はGRIMではない戦略を採用している子供が3.8%含まれる場合の集団の平均利得を表している．

表3　協調戦略出現頻度と集団の平均利得の例

| 評価方法 | 遺伝子座3 | 遺伝子座4 | 遺伝子座5 | 遺伝子座6 | 遺伝子座7 | 平均利得 |
|---|---|---|---|---|---|---|
| 依存型 | 1.000 | 0.962 | 0.000 | 0.000 | 0.000 | 0.501 |
| 非依存型 | 1.000 | 1.000 | 0.000 | 0.000 | 0.038 | 0.630 |

GRIM戦略を採用していない子供がどのような戦略を採用しているかによって集団の平均利得は変わるが，この数値例を目安として許容相対誤差を決定する．そこで，環境依存型の場合の許容相対誤差，$\varepsilon$を$0.165 = (0.6 - 0.501)/0.6$より若干大きい，17%とし，環境非依存型評価の場合は$0.017 = (0.641 - 0.630)/0.641$よりも若干大きく，2%とする．

我々の主たる関心は環境非依存型評価にある．この場合，経験的に初期個体中に1人でもGRIM戦略者が存在するならば，通常のGAに2ついずれかの意味でのエリート保存を併用すると100世代以内にほぼ確実に，少なくとも1世代は相互協力状態が発現する．これに対してGAのみの場合は数万世代を経ないと相互協力が発現しない場合が多い（5万世代が経過しても発現しない場合もある）．しかしながら，概ね2万世代以内には相互協力状態が発現する．そこで，条件1と2のチェックを開始する世代，$T_{test}$を2万とする（正確には2万＋1世代からチェックが始まる）．したがって，条件1と2は$T_{test} < t \leq T_{max}$なる$t$，すなわち3万世代にわたってチェックされることになる．

条件2の許容誤差，$\delta$に関しては統計的に5%の誤差を許容する場合が多いので，5%を与える．他方，最終判定基準である条件3は2万世代から5万世代までの3万世代中の相互協力発現比率を計算することを意味するが，緩やかな基準としてこれが90%以上の場合と，それよりも厳しい基準として95%以上の場合の2つを検討する．すなわち，$p$=0.90，あるいは$p$=0.95である．

これまでの議論から，持続的相互協力状態への進化を判定するためのパラメータ設定値は表4のとおりになる．

表４　パラメータ設定値：$\varepsilon^*$は環境非依存型評価の場合は２％，依存型評価の場合は17％とする

| パラメータ | 意味 | 設定値 |
|---|---|---|
| $T_{test}$ | 条件１と２のチェックを開始する世代 | ２万 |
| $T_{max}$ | 最大世代数（最大計算回数） | ５万 |
| $\varepsilon$ | 条件１の許容相対誤差 | $\varepsilon^*$ |
| $\delta$ | 条件２の許容誤差 | 0.05 |
| $p$ | 条件３の最小値 | 0.90，または0.95 |

## ３．数値実験結果

以下の議論においては環境依存型評価と非依存型評価のそれぞれに関して，ルーレット選択のみ，これにエリート保存を併用した場合，および強エリート保存を併用した場合の３ケースについて初期GRIM戦略者の数を０から５まで変化させつつ，100個の異なる乱数の種を用いた数値実験が行われている．したがって，総実験回数は3,600＝２×３×６×100回である．初期GRIM戦略者の最大数を５までに限定した理由は，経験的にこれ以上GRIM戦略者の数を増やしても結論には影響しないと判断できることと，時間的制約からである．

### 3.1　環境依存型評価

まず，エージェントが近視眼的な行動をとる環境依存型評価から検討を始めよう．あらかじめ他のエージェントとの対戦回数を定めて戦略評価を行っているので，正確にいえば我々のゲームは文献[11]や[12]の$N$・ストライク・アウト・モデルに近いものであるが，対戦回数は十分に大きい（$N=100$）ので，近似的な意味ではあるが無限回繰り返しゲームであるといってよいであろう．実際，仮に割引因子が１に近い0.9で与えられたとするならば，$0.9^{100}\approx 2.67\times 10^{-5}$となるので101回目以降の対戦結果は戦略評価にほとんど反映されない．以下では，逐一このことを断らない．

塩村，藤本[9]の数値実験結果から環境依存型評価の場合，持続的相互協力状態への進化は起こりにくいことが予想される．表５は条件３，

$$\frac{s}{T_{max}-T_{test}}\geq p$$

の右辺，$p$を0.95と0.90に設定した場合に持続的相互協力状態が発現したと判定された回数の比率（100回中）である．他方，図３は条件３の左辺，すなわち相互協力発現率の期間，$T_{max}-T_{test}$を通じた平均値を算出した結果であるが，図に示しているのはこの平均値の100回の数値実験を通した平均値である．凡例のルーレット，エリート，および強エリートはそれぞれルーレット選択のみを使用した場合（通常のGA），これに通常のエリート保存を併用した場合，およ

び強エリート保存を併用した場合を意味する．

表５　環境依存型評価における持続的相互協力発現率（100回中）

| | 95%判定 | | | 90%判定 | | |
|---|---|---|---|---|---|---|
| GRIM 数 | ルーレット | エリート | 強エリート | ルーレット | エリート | 強エリート |
| 0 | 0% | 0% | 17% | 0% | 0% | 17% |
| 1 | 0% | 1% | 22% | 0% | 2% | 22% |
| 2 | 0% | 0% | 24% | 0% | 2% | 24% |
| 3 | 0% | 0% | 33% | 0% | 1% | 33% |
| 4 | 0% | 0% | 42% | 0% | 2% | 42% |
| 5 | 0% | 0% | 48% | 0% | 3% | 48% |

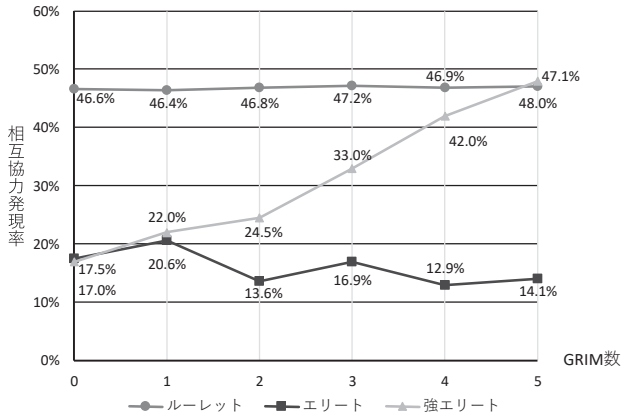



図３　環境依存型評価における相互協力発現率平均値の推移

これらの表と図から，我々の推論が妥当なものであることが確認されるが，特にルーレット選択のみを用いたGAの下では持続的相互協力状態に進化しないことが確認できる．これは環境依存型評価を用いた場合，集団は全面協力と全面裏切りを繰り返すために相互協力状態を持続できないことによる．また，持続的相互協力発現の判定のために用いた$p$の値を0.95から0.90へと緩めても結果はほとんど変わらないことが確認できる．これは世代を通じた相互協力発現率がルーレット選択のみの場合，全面裏切りと全面協力状態の永続的振動を反映して0.9を超えることがほとんどないこと，および２ついずれかの意味でのエリート保存を併用した場合は比較的初期段階で全面裏切りか，全面協力状態のいずれかに固定されてしまうことを反映してゼロに近いか，１に近いかが比較的明確に分かれるためである（表６参照）．

大変興味深いことに，図３からルーレット選択のみの場合，初期GRIM戦略者数にかかわらず相互協力発現率平均値がほぼ一定であるであることがわかる．この理由は進化の過程，あるいは歴史を決定づける乱数の種が同じであるならば，初期GRIM戦略者数にかかわらず同じような進化をたどることによる．

表 6　環境依存型評価における相互協力発現率平均値に関する基本統計：各ケース100回の数値実験から得られた結果であり，統計値は小数点以下 3 桁に丸めた結果

| | ルーレット | | | | エリート | | | | 強エリート | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GRIM 数 | 平均 | 分散 | 最大 | 最小 | 平均 | 分散 | 最大 | 最小 | 平均 | 分散 | 最大 | 最小 |
| 0 | .466 | .024 | .891 | .116 | .175 | .087 | .886 | .000 | .170 | .114 | 1.000 | .000 |
| 1 | .464 | .024 | .891 | .116 | .206 | .103 | .997 | .000 | .220 | .171 | 1.000 | .000 |
| 2 | .468 | .025 | .891 | .116 | .136 | .072 | .925 | .000 | .245 | .182 | 1.000 | .000 |
| 3 | .472 | .025 | .891 | .116 | .169 | .092 | .925 | .000 | .330 | .221 | 1.000 | .000 |
| 4 | .469 | .023 | .891 | .116 | .129 | .072 | .926 | .000 | .420 | .243 | 1.000 | .000 |
| 5 | .471 | .025 | .891 | .116 | .141 | .086 | .926 | .000 | .480 | .249 | 1.000 | .000 |

1 例として，乱数の種を 1 に固定した場合の進化の過程をグラフに表したものが図 4 である．これは進化の過程を表す指標として集団の平均利得を選び，その100世代ごとの推移をグラフに表したものである．凡例中の GRIM 0 等は初期戦略分布における GRIM 数をゼロから 5 まで変化させたことを意味している．なお，集団の平均利得が約0.6と0.2になっている場合が，それぞれ全面協力状態と全面裏切り状態に対応している．

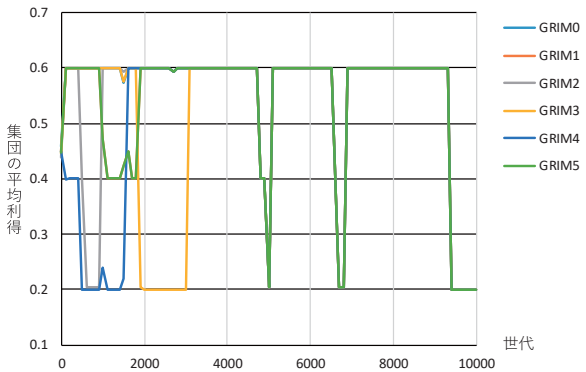



図 4　環境依存型ルーレット選択における集団の平均利得の推移：乱数の種は 1 で固定

図から初期 GRIM 戦略者数の違いを反映して初期段階では平均利得の推移に違いが現れるものの，約 4 千世代以降では完全に一致することが確認されよう．他の乱数の種を選んだ場合も同様である．このことから通常の GA の下では初期戦略分布それ自身は十分に時間が経過した後の進化にあまり影響を与えないことが示唆される．

一方，ルーレット選択に通常のエリート保存を併用した場合，併用しない場合と比較して持続的相互協力状態の発現率が小さくなる．この理由は，多くの場合において全面裏切り戦略が最優秀戦略になり，この戦略が最終的に集団に広がるためである．また，図 3 から GA に通常のエリート保存を併用した場合，初期 GRIM 戦略者数と相互協力発現率平均値がほぼ無相関になっていることが確認できるが，これはルーレット選択のみを使用した場合とは理由が異なる．

実際，同じ乱数の種を用いても進化の過程は異なる．

図5は同じ乱数の種，58を用いているが，初期GRIM戦略者数を1と2に変えた場合の進化の過程をグラフに表したものである．図より一方は全面協力状態に進化しているのに対して，他方は全面裏切り状態に進化していることが確認できる．図4と比較されたい．

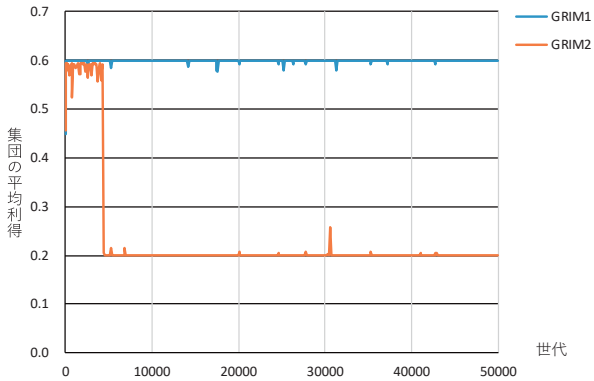



図5　環境依存型エリート保存併用における集団の平均利得の推移：乱数の種は58で固定

通常のエリート保存を併用した場合，初期GRIM戦略者数と相互協力発現率平均値がほぼ無相関になる理由は，相互協力発現率平均値が小さいことからうかがえるように，初期GRIM戦略者数にかかわらず期間の大部分，特にチェックを開始する期間，$T_{test}$ 以降，全面裏切り状態になる場合が多く，この結果100回の数値実験を通した平均値に初期GRIM戦略者数の差の影響が現れにくくなっているのである．

ルーレット選択に強エリート保存を併用した場合も，併用しない場合と比較して持続的相互協力状態が小さくなる理由は，初期段階において必ずしも協力的ではない最優秀戦略がエリート保存よりもさらに急速に集団中に広がるためであるが，強エリート保存の下，初期段階における協力的な戦略で平衡状態に落ち着く場合も少なからずあることを強調しておく．また，他の2ケースと異なり，初期GRIM戦略者数と持続的相互協力発現率，あるいは相互協力発現率平均値との間に明確な正相関が見られることにも注意されたい．これは初期GRIM戦略者数が多ければ多いほどGRIM戦略者同士が対戦する可能性が大きくなり，この互恵的な戦略が最優秀戦略として選ばれ，かつ他の戦略を完全に淘汰してしまうほどに集団内で急速に勢力を拡大するためである．実際，持続的相互協力状態が発現するほとんどのケースはGRIM戦略によってそれが実現している．たとえば，初期GRIM戦略数が5の場合，持続的相互協力が発現した48回中，42回が最終的にGRIM的な戦略へと進化し，5回が2回目以降，Pavlov的行動をとる戦略（記憶部を除く遺伝子型が01001），Pavlov戦略（記憶部を除く遺伝子型が11001）に1回進化している．ここで，Pavlov戦略とは初回は協調的な戦略をとり，2回目以降は前回の両プレイヤーの戦略が同じ場合は協調的戦略を，そうではない場合は裏切り戦略をとる行動計画である．

結局のところ，ルーレット選択のみ，ルーレット選択と通常のエリート保存の併用，および強エリート保存を併用した場合の進化の過程の違いは，最優秀戦略が集団中で広がるスピードの差にあるといえる．

### 3.2 環境非依存型評価

これまでの議論から，目先の利益に捕らわれて戦略を評価する環境依存型評価の下では持続的相互協力状態の発現は極めて困難であることが確認された．そこでつぎに，本稿の主たる目的である環境非依存型評価を行うという前提の下，強エリート保存を併用するならば初期戦略分布に 1 人でも GRIM 戦略者が存在すると，100%持続的相互協力状態が発現することを示そう．環境依存型評価と異なり，この場合のエージェントは長期的な観点に立ち，極めて慎重に戦略評価を行っていると想定されている．

先と同様に，表 7 は条件 3 の右辺，$p$ を0.95と0.90に設定した場合に持続的相互協力状態が発現したと判定された回数の比率である．他方，図 6 は図 3 と同様に相互協力発現率の期間を通じた平均値を算出した結果である．

表 7　環境非依存型評価における持続的相互協力発現率（100回中）

| | 95%判定 | | | 90%判定 | | |
|---|---|---|---|---|---|---|
| GRIM 数 | ルーレット | エリート | 強エリート | ルーレット | エリート | 強エリート |
| 0 | 3% | 49% | 15% | 39% | 98% | 15% |
| 1 | 3% | 51% | 100% | 43% | 100% | 100% |
| 2 | 3% | 51% | 100% | 39% | 100% | 100% |
| 3 | 1% | 51% | 100% | 38% | 100% | 100% |
| 4 | 1% | 49% | 100% | 39% | 100% | 100% |
| 5 | 1% | 49% | 100% | 40% | 100% | 100% |

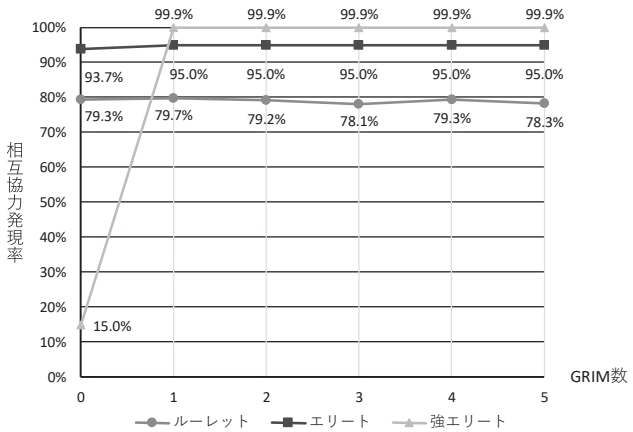



図 6　環境非依存型評価における相互協力発現率平均値の推移

環境依存型評価の場合と異なり，ルーレット選択のみの場合でも持続的相互協力状態が発現していると判定されたケースがあるが，これは戦略評価方法を変えたことにより全面協力状態と全面裏切り状態の永続的振動が部分的に抑えられることによる．２ついずれかの意味でのエリート保存を併用した場合における，この効果は著しい．実際，表８において，相互協力発現率平均値の分散が著しく小さくなっている．特に強エリート保存を併用した場合には初期GRIM戦略者が集団中に１人でもいると，100％持続的相互協力状態が発現する．期間を通じた相互協力発現率の平均値がわずかに100％に至らない（差は0.1％程度）理由は突然変異による撹乱によるものと考えられる．

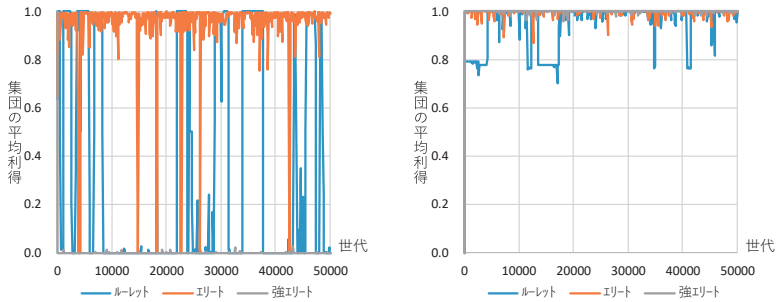
表８　環境非依存型評価における相互協力発現率平均値に関する基本統計：E-$x$ は$10^{-x}$程度の大きさであることを意味する

| | ルーレット | | | | エリート | | | | 強エリート | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GRIM 数 | 平均 | 分散 | 最大 | 最小 | 平均 | 分散 | 最大 | 最小 | 平均 | 分散 | 最大 | 最小 |
| 0 | .793 | .033 | .953 | .186 | .937 | .007 | .958 | .221 | .150 | .127 | .999 | .000 |
| 1 | .797 | .039 | .953 | .186 | .950 | E-5 | .960 | .933 | .999 | E-8 | 1.000 | .999 |
| 2 | .792 | .031 | .953 | .238 | .950 | E-5 | .960 | .933 | .999 | E-8 | 1.000 | .999 |
| 3 | .781 | .041 | .953 | .000 | .950 | E-5 | .960 | .933 | .999 | E-8 | 1.000 | .999 |
| 4 | .793 | .037 | .953 | .000 | .950 | E-5 | .960 | .933 | .999 | E-8 | 1.000 | .999 |
| 5 | .783 | .041 | .953 | .000 | .950 | E-5 | .960 | .933 | .999 | E-8 | 1.000 | .999 |

典型例として，図７は変換式，

$$\text{変換データ} = \frac{\text{元データ} - \text{最小値}}{\text{最大値} - \text{最小値}}$$

により集団の平均利得が０以上，１以下の値を持つようにスケール変換したときの推移を表したものであり，図 $a$ と $b$ は各々，環境依存型評価の場合と非依存型評価の場合である．ともに，乱数の種を27に固定しており，初期GRIM戦略者数を１に設定している．




（$a$）　環境依存型評価　　（$b$）　環境非依存型評価

図７　環境非依存型評価の振動抑制効果

図より，戦略評価方法を変えたことにより振動が抑えられることが確認されるであろう．ここで環境依存型評価の場合，平均利得はプレイヤーの実際の対戦結果から算出されるのに対して，非依存型評価の場合はプレイヤーの頭の中での対戦結果から算出されることを思い出されたい．この結果，集団の平均利得が取り得る最大値と最小値は２つの評価方法では異なる．スケール変換を行って２つを比較している理由はこのためである．

興味深いことに，ルーレット選択に通常のエリート保存を併用した場合，初期GRIM戦略者の数にかかわらず比較的高確率で持続的相互協力状態が発現するのに対して，強エリート保存を併用した場合，初期GRIM戦略者がゼロのときには他２つのケースよりも持続的相互協力率が小さくなっている．特に，判定基準を90％に緩めたときは，前者はほぼ100％持続的相互協力状態が発現しているのに対して後者は15％に過ぎない．

この理由は，最優秀戦略が集団中で広がるスピードと互恵的突然変異戦略の勢力拡大のしやすさの違いにあると考えられる．初期GRIM戦略者数がゼロの場合，強エリート保存を併用すると互恵的ではない戦略が最優秀戦略となり，次世代の子孫を生み出す「片親」に固定されてしまうためにこれが集団中に急速に広がり，たとえ互恵的突然変異戦略が出現したとしても，彼らはその勢力を拡大しにくくなっている．これに対して通常のエリート保存を併用した場合は互恵的ではない戦略が最優秀戦略になったとしても，強エリート保存を併用する場合に比較すると集団中で広がるスピードが緩やかであり，ひとたびGRIM戦略のような互恵的戦略が出現すると，これが次世代を生み出す「両親」として選び出され，したがって集団中に互恵的戦略が広がる可能性があるからである．

## ４．結　語

以上，我々は持続的相互協力状態発現を判定するための３つの基準を設けてその判定を自動化するとともに初期戦略分布の偏りを前提とした２タイプのIPDゲームを行ってきた．我々はこれら２タイプの一方を環境依存型と呼び，他方を非依存型と呼んできたが，前者はAxelrod[10]やHoffmann[6]が行った数値実験と同種の戦略評価方法であり，後者は塩村，藤本[9]により導入されたものである．

Sugdenは“長期にわたる一連のゲームにおいて，人びとは最も成功したと認識した戦略を用いる傾向がある”と述べているが（文献[13]，p.18参照），これをGAに直接反映させたものがエリート保存である．人間社会のシミュレーションとしてモデルにこれを組み込むことは自然に思えるが，明示的にGAにエリート保存を併用したモデルは我々が知る限り，本稿と塩村，藤本[9]を除けば梅原，小林[14]しかない．ただし，最後のものは研究報告というよりも，GAの解説を意図したものである．著者達はエリート保存と呼んではいないが，Ohdaira and Terano[15]では本稿における強エリート保存に近いアルゴリズムが組み込まれている．

持続的相互協力状態発現判定のために本稿で用いた基準は，十分大きな世代以降で 1）集団

の平均利得と過去最高の集団の平均利得との相対誤差が十分小さい．2）集団の遺伝子が互恵的な相互協力状態へと導く戦略になっている．これら2条件が同時に満たされるならば，集団は互恵的最優秀戦略を採用していると判定し，さらにそれが持続的なものであることを判定するために 3）チェック期間の大部分で集団が相互協力状態に達している．以上の3つからなる．

環境依存型評価，すなわち極めて場当たり的，あるいは近視眼的な戦略評価を行ったときのIPDゲームの結果は次のように要約される．1）2ついずれかの意味でのエリート保存を併用しない通常のGAの場合，初期戦略分布にかかわらず全面協力と全面裏切りの振動を回避することは不可能である．この事実は十分な世代を経ると進化の過程が初期戦略分布の違いにかかわらず同じものになってしまうことによる．ただし，我々が考慮しなかった極端な初期戦略分布の偏りがある場合を除く．2）2ついずれかの意味でのエリート保存を併用したとしても，換言するならば過去を通して優秀な戦略を考慮したとしても，持続的相互協力状態が発現することは極めて困難であり，特に通常のエリート保存を併用すると，全面裏切りに進化してしまう危険性が大きい．一方，強エリート保存を併用した場合は初期GRIM戦略者数と持続的相互協力発現率の明確な正の相関関係が現れる．

他方，非依存型評価，すなわちあらゆる可能性を考慮した極めて慎重な戦略評価を行ったときのIPDゲームに関しては以下のとおりである．1）通常のGAにおいても持続的相互協力状態への進化が可能である．これは全面協力と全面裏切りの振動を抑えることによる．2）この効果はエリート保存を併用した場合に顕著に現れ，特に強エリート保存を併用するという条件の下では，初期戦略分布に1人でもGRIM戦略者が存在するならば100％相互協力状態が発現する．注目に値するのは，通常のエリート保存を併用するという条件の下，たとえ初期戦略分布にGRIM戦略を欠いていたとしても100％近く相互協力状態が発現するという点である．一方，強エリート保存を併用するという条件の下では，相互協力発現率は3つのケースの中では最低であった．この理由は2つのエリート保存では（必ずしも互恵的ではない）最優秀戦略が集団中で広がるスピードと互恵的突然変異戦略の勢力拡大のしやすさに違いがあることによる．

以上の結果は，本稿で考察した持続的相互協力状態発現のための判定基準により得られたものである．上述したとおり，相互協力状態発現の定義に関して研究者間で共有されたものはなく，それゆえに我々の判定基準に関しても曖昧性は残ってはいるが，判定を自動化したことにより数値実験結果の解釈の時間短縮と数値データに基づく「ある程度」客観的な議論が可能になったことを最後に強調しておく．

**参考文献**

［1］Axelrod, R.: *The Evolution of Cooperation*, Basic Book, New York（1984）.
［2］Maynard Smith, J.: *Evolution and the Theory of Games*, Cambridge University Press, New York（1982）.
［3］Axelrod, R.: Effective choice of Prisoner’s Dilemma, *Journal of Conflict Resolution*, Vol.24, No.1, pp.3–25（1980）.

［4］Axelrod, R.: More effective choice of Prisoner's Dilemma, *Journal of Conflict Resolution*, Vol.24, No.3, pp.379-403（1980）.
［5］Axelrod, R.: The emergence of cooperation among egoists, *American Political Science Review*, Vol.75, pp.306-318（1981）.
［6］Hoffmann, R.: The ecology of cooperation, *Theory and Decision*, Vol.50, pp.101-118（2001）.
［7］Lindgren, K.: Evolutionary phenomena in simple dynamics, in C. G. Langton, J. D. Farmer, S. Rasmussen and C. Taylor eds.: *Artificial Life II*, pp.295-312, Addison-Wesley, Redwood City CA（1991）.
［8］Nowak, M. A. and Sigmund, K.: A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game, *Nature*, Vol.364, pp.56-58（1993）.
［9］塩村尊，藤本貴士：遺伝的アルゴリズムによる永続的互恵主義の可能性，情報処理学会論文誌，Vol.56，No.1，pp.401-410（2015）.
［10］Axelrod, R.: The Evolution of strategies in the iterated Prisoner's Dilemma, in L. Davis ed., *Genetic algorithms and Simulated Annealing*, pp.32-41, Pitman, London（1987）.
［11］太田勝造：法律　社会科学の理論とモデル７，東京大学出版会（2000）.
［12］細江守紀，太田勝造編著：法の経済分析　契約，企業，政策，勁草書房（2001）.
［13］Sugden, R.: *The Economics of Rights, Co-operation and Welfare*, 2nd edition, Palgrave Macmillan, UK（2005）.
［14］梅原嘉介，小林敬治：進化ゲーム理論と遺伝的アルゴリズム，工学社（2007）.
［15］Ohdaira, T. and Terano, T: Cooperation in the Prisoner's Dilemma game based on the second-best decision, *Journal of Artificial Societies and Social Simulation*, Vol.12, No.4（2009）.