4S-05

グリコゲームを題材とした素朴なナッシュ均衡導出法の評価 *

垣本健吾 † 首藤裕一 ‡

法政大学 情報科学部

1 はじめに

本稿では、対称な二人零和不完全情報ゲームであるグリコゲームを対象として、二人零和ゲームにおいてナッシュ均衡の近似解を求める素朴な方法の有効性を評価する。対称な二人零和ゲームにおいてナッシュ均衡を構成する戦略は、いかなる戦略に対しても不利にならないという重要な意味を持ち、ナッシュ均衡を計算する手法は広く研究されている。例えば、Zinkevich ら [1] は、完全記憶な二人零和不完全情報ゲームのナッシュ均衡を計算するアルゴリズムを提案し、強いポーカー AI の設計に応用した。よく知られたミニマックス定理によれば、一般の二人零和ゲームにおいてマックスミニ戦略とミニマックス戦略がナッシュ均衡を構成する。すなわち、プレイヤ 1,2 の混合戦略の集合をそれぞれ Σ_1, Σ_2 とし、プレイヤ 1,2 がそれぞれ混合戦略 $\sigma_1 \in \Sigma_1, \sigma_2 \in \Sigma_2$ を選択した場合のプレイヤ 1 の期待利得を $F(\sigma_1, \sigma_2)$ とすると、$\min_{\sigma_2 \in \Sigma_2} F(\sigma_1^*, \sigma_2) = \max_{\sigma_1 \in \Sigma_1} \min_{\sigma_2 \in \Sigma_2} F(\sigma_1, \sigma_2)$ および $\max_{\sigma_1 \in \Sigma_1} F(\sigma_1, \sigma_2^*) = \min_{\sigma_2 \in \Sigma_2} \max_{\sigma_1 \in \Sigma_1} F(\sigma_1, \sigma_2)$ を満たす戦略の組 (σ_1^*, σ_2^*) がナッシュ均衡である。マックスミニ戦略 σ_1^* の計算を試みるとき、マックスミニ戦略の定義から、次のような素朴なアルゴリズムが想起される。

- 与えらた整数のパラメータ k に対し、ランダムに k 個の混合戦略 $r_1, r_2, \ldots, r_k$ を生成する。
- $\min_{i \in [1,k]} F(\sigma, r_i)$ を最大化する戦略 $\sigma^* \in \Sigma_1$ を計算し、σ^* をマックスミニ戦略の近似解とする。

$F(\sigma_1, \sigma_2)$ の形状に仮定を置かないとき、特定の σ_2 の近傍で $F(\sigma^*, \sigma_2)$ の値が極端に小さくなることがあるかもしれない。このとき、生成した $r_1, r_2, \ldots, r_k$ が σ_2 の近傍に含まれないとき、得られる戦略 σ^* はマックスミニ戦略と大きく乖離する。したがって、一般の二人零和ゲームに対し、σ^* は何らの精度保証を与えないと考えられる。一方で、単純な構成の二人零和ゲームであれば、このアルゴリズムでマックスミニ戦略に十分近い混合戦略が得られるかもしれない。

本研究では、この問いを明らかにするための端緒として、単純な構成の二人零和不完全情報ゲームであるグリコゲーム（じゃんけんグリコ）にこの素朴なアルゴリズムを適用する。後述するように、渡邊 [2] によってこのゲームのナッシュ均衡は解析的に求まることが明らかにされている。本研究では、シミュレーション実験により、素朴なアルゴリズムで得たグリコゲームのナッシュ均衡の精度を評価する。

2 グリコゲーム

プレイヤー 1,2 が $N \geq 1$ 段の階段を先に登ることを目指す。各時刻 $t = 1, 2, \ldots$ でプレイヤ 1,2 はじゃんけんを行う、すなわち、両者は同時にグー、チョキ、パーのいずれかの手を出す。じゃんけんの勝敗（あいこを含む）は標準的なルールに従って決定するものとする。（パーがグーに勝つ等。）いずれか一方がじゃんけんに勝利した場合、勝者は自らが出した「手」に応じて階段を登る。具体的には、グーで勝った場合には 1 段上り、それ以外の手で勝った場合には 2 段登る。*1 ただし、いずれかのプレイヤーが途中で N 段目に達した場合（すなわち合計 N 段を登った場合）には、その時点でゲームは終了となり、そのプレイヤーがこのゲームの勝者となる。$N-1$ 段目に位置しているプレイヤーがチョキやパーで勝った場合も、そのプレイヤーは勝者となる。あいこの場合にはいずれのプレイヤーも階段を登らない。ゲームの勝者

* Evaluation of Naive Method for Computing Nash Equilibrium with the Glico Game

† Kengo Kakimoto, Hosei University

‡ Yuichi Sudo, Hosei University

*1 グーで勝った場合には 3 段、それ以外の手で勝った場合には 6 段登るのが標準的なルールであるが、それぞれ 1 段、2 段に変更しても比率は変わらないため、本質を失わずにゲームを簡単化できる。

は利得 1 を獲得し、敗者は利得 -1 を得る。

グリコゲームの状態は、プレイヤ 1,2 がそれぞれあと何段登れば勝利するかを表す n, m の組 $(n, m) \in [1, N] \times [1, N]$ で表す。状態集合 $[1, N] \times [1, N]$ を S とおく。プレイヤ 1,2 の戦略集合 Σ は同一であり、任意の戦略 $\sigma = (p_G(n, m), p_C(n, m))_{(n,m)\in S} \in \Sigma$ は各状態 $(n, m) \in S$ においてそれぞれグー、チョキを出す確率 $p_G(n, m), p_C(n, m)$ を与える。ただし、任意の n, m に対して $p_G(n, m) + p_C(n, m) \leq 1$ であり、$1 - p_G(n, m) - p_C(n, m)$ が状態 (n, m) においてパーを出す確率に相当する。

渡邊 [2] はグリコゲームのナッシュ均衡の解析的な計算方法を与えた。証明は割愛するが、以下の方程式を解くことによってプレイヤ 1 のマックスミニ戦略 $\sigma^* = (q_G(n, m), q_C(n, m))_{(n,m)\in S}$ を再帰的に求めることができる。

$$v_{n,m} = q_G \cdot v_{n,m} + q_C \cdot v_{n,m-1} + (1 - q_G - q_C)v_{n-2,m}$$
$$v_{n,m} = q_G \cdot v_{n-1,m} + q_C \cdot v_{n,m} + (1 - q_G - q_C)v_{n,m-2}$$
$$v_{n,m} = q_G \cdot v_{n,m-2} + q_C \cdot v_{n-2,m} + (1 - q_G - q_C)v_{n,m}$$

ただし、方程式内では q_G, q_C, q_P の引数 (n, m) は割愛している。また、任意の $i \geq 0$ について $v_{0,i} = v_{-1,i} = 1$ および $v_{i,0} = v_{i,-1} = -1$ とする。ここで、変数 $v_{n,m}$ は状態 (n, m) からゲームを開始したときにナッシュ均衡戦略でプレイヤ 1 が得る利得の期待値に相当する。

3 素朴なアルゴリズムの実装

本節では、次節の評価実験で用いる、グリコゲームに対する素朴なアルゴリズムの実装を説明する。グリコゲームは展開型の不完全情報ゲームであるが、前節の方程式で見たように、各 $(n, m) \in S$ に対応した N^2 個の標準型（戦略型）ゲーム $J(n, m)$ を連結したゲームとみることもできる。具体的には、$J(n, m)$ は（あいこが続く限りやり直す）一回のじゃんけんで構成され、グーで勝った（負けた）場合には利得 $v_{n-1,m}$（$v_{n,m-1}$）を獲得し、チョキまたはパーで勝った（負けた）場合には利得 $v_{n-2,m}$（$v_{n,m-2}$）を獲得するゲームと考える。したがって、グリコゲームそのものに対するランダムな戦略を k 個生成するのではなく、n, m について昇順に、$J(n, m)$ に対するランダム戦略 $r_i(n, m) = (p^i_G(n, m), p^i_C(n, m))$ を k 個生成し、$\min_{i\in[1,k]} F_{n,m}(\sigma, r_i)$ を最大化する $\sigma^*(n, m)$ を求める。最終的に $\sigma^* = (\sigma^*(n, m))_{(n,m)\in S}$ をグリコゲーム全体のマックスミニ戦略として出力する。

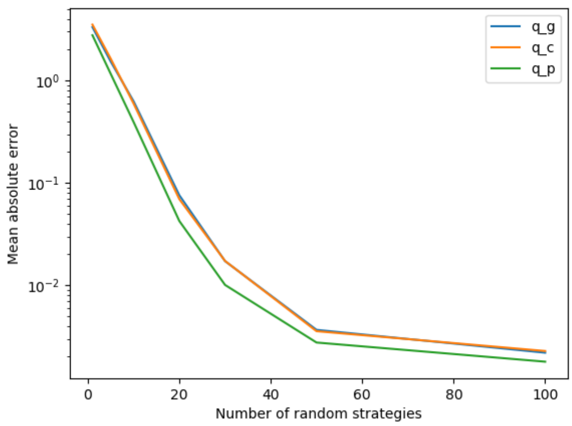

図 1 階段数が 10 のときの平均絶対誤差。縦軸は対数スケール。

$\min_{i\in[1,k]} F_{n,m}(\sigma^*(n, m), r_i(n, m))$ を最大化する $\sigma^*(n, m)$ を求めるためにはまず、任意の混合戦略 s_1, s_2 に対する期待利得 $F_{n,m}(s_1, s_2)$ の計算方法が必要である。この計算は、$v_{n-2,m}, v_{n-1,m}, v_{n,m-2}, v_{n,m-1}$ の値が与えられれば、単純な非線形式を計算するだけで求まる。$v_{n-2,m}, v_{n-1,m}, v_{n,m-2}, v_{n,m-1}$ については、正確な値は使えないが、$v_{x,y}$ として $\min_{i\in[1,k]} F_{x,y}(\sigma^*(x, y), r_i(x, y))$ を代用することで近似的に計算する。詳細は割愛するが、この期待利得計算を用いて、決定的に選んだ 200 個の戦略から $\min_{i\in[1,k]} F_{n,m}(\sigma, r_i)$ を最大化する $\sigma^*(n, m)$ を決定する。

4 実験

$n = 10, k = 1$，5，10，20，30，50，100 の設定で前節のアルゴリズムを実行し、得られた戦略 σ^* と渡辺 [2] の解析的手法で得たマックスミニ戦略との絶対誤差を評価した。表 1 に、$n = 10$ に対し 100 回の試行を行って評価した平均絶対誤差を示す。ランダム戦略の数 k に応じて高速に誤差が減少することが確認できた。

参考文献

[1] M. Zinkevich, M. Johanson, M. Bowling, and C. Piccione. Regret minimization in games with incomplete information. *Advances in neural information processing systems*, 20:1729–1736, 2007.

[2] 渡辺隆裕. グリコ・チョコレート・パイナップルゲームのゲーム理論による解. http://nabenavi.net/gcp/, 2020.