F-036

繰り返しゲームにおける高エラー率に対する代表的な戦略の性質
Nature of representative strategies against high error rates in repeated games

間宮 安曇 [1]　　一ノ瀬 元喜 [2]
Azumi Mamiya　　Genki Ichinose

1. はじめに

繰り返しゲームは、長期の個体間の関係を探求するための代表的な例である。また、それは利己的個体間でどのように協力や競争が個人間で生まれるかを理解するのを助ける。現実社会では、エラーは常に発生するため、繰り返しゲームにおいて、エラーの影響を考えることは、重要である。本研究では、繰り返しゲームにおける代表的なゼロ行列式戦略や無条件戦略に対する知覚エラーの影響を調査する。

2. モデル

2 人プレイヤー$i \in \{\mathrm{X,Y}\}$の繰り返し囚人のジレンマゲームを考える。それぞれのプレイヤーは、いずれかの行動$\mathrm{a}_i \in \{\mathrm{C,D}\}$を選択する。相手の選択した行動を直接見ることはできないが、相手からの信号$\omega_i \in \{\mathrm{g,b}\}$ を受け取ることができる。g と b は good と bad を意味する。これらの信号は、相手の選択した行動だけでなく、エラーの影響を受ける。2 人のプレイヤーが選択した行動がそれぞれa_X, a_Yのとき、2 人のプレイヤーがそれぞれ受け取る信号がω_X, ω_Yである確率は、$\pi(\omega_X, \omega_Y | a_X, a_Y)$のように表す。もし、プレイヤーY が C（または、D）の行動を選択したとして、プレイヤーXがb（または、g）の信号を受け取った場合、エラーが起きたということになる。ε をどちらか 1 人のプレイヤーにエラーが起きる確率とし、ξ を両方のプレイヤーに起きる確率とする。すると、$1-2\varepsilon-\xi$ は、どちらもエラーしない確率となる。1 回のゲームの利得は、自分の行動と受け取った信号によって決まる。実際の利得は、プレイヤーが C の行動を選択し、g の信号を受け取った場合は$\mathrm{u_i(C,g)} = 1$、b の信号を受け取った場合は$\mathrm{u_i(C,b)} = -\mathrm{L}$とし、プレイヤーが D の行動を選択し、$g$ の信号を受け取った場合は$\mathrm{u_i(D,g)} = 1+\mathrm{G}$、$b$ の信号を受け取った場合は$\mathrm{u_i(D,b)} = 0$とする。また、L と G は正の値とする。1 回のゲームでプレイヤーがそれぞれa_X, a_Yの行動を選択したときのプレイヤーi の得られる期待利得は、

$$f_i(a_X, a_Y) = \sum_{\omega} u_i(a_i, \omega_i)\pi(\omega_X, \omega_Y | a_X, a_Y) \qquad (1)$$

と表される[1, 2]。$\{a_X, a_Y\}$が{C, C},{C,D},{D,C},{D,D}のときのゲーム 1 回あたりの期待利得をそれぞれ $\mathrm{R_E}$ $\mathrm{S_E}$, $\mathrm{T_E}$, $\mathrm{P_E}$,とする。式(1)より $R_E = 1-(L+1)(\varepsilon+\xi)$, $S_E = -L+(1+L)(\varepsilon+\xi)$, $T_E = (1+G)(1-\varepsilon-\xi)$, $P_E = (1+G)(\varepsilon+\xi)$のように計算される。プレイヤーX のゲームの利得行列を$\boldsymbol{U_X} = (R_E, S_E, T_E, P_E)$,プレイヤーY は、$\boldsymbol{U_Y} = (R_E, T_E, S_E, P_E)$となる。プレイヤーの行動は、前回のゲームの結果のみを見て決める。前回のゲーム結果が Cg,Cb,Dg,Db（1,2,3,4 とラベルを付ける）のとき、次のゲームでプレイヤーX が C の行動を選択する確率をそれぞれp_1, p_2, p_3, p_4とすると、戦略は$\boldsymbol{p} = (p_1, p_2, p_3, p_4)$と書ける。同様に、プレイヤーY についても$\boldsymbol{q} = (q_1, q_2, q_3, q_4)$とする。$p, q$の下付き文字は前回のゲームの結果を意味する。

この繰り返しゲームをマルコフ過程とみて、遷移行列を M とする。また、$\boldsymbol{v}$を M の定常分布、$\boldsymbol{f}$を任意の列ベクトルとすると、$\boldsymbol{v}$と$\boldsymbol{f}$の内積は下記の式(2)の行列式の形で表すことができる[3]。

$$\boldsymbol{v} \cdot \boldsymbol{f} \equiv D(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{f}) =$$
$$\begin{vmatrix} \tau p_1 q_1 + \epsilon p_1 q_2 + \epsilon p_2 q_1 + \xi p_2 q_2 - 1 & \mu p_1 + \eta p_2 - 1 & \mu q_1 + \eta q_2 - 1 & f_1 \\ \epsilon p_1 q_3 + \xi p_1 q_4 + \tau p_2 q_3 + \epsilon p_2 q_4 & \eta p_1 + \mu p_2 - 1 & \mu q_3 + \eta q_4 & f_2 \\ \epsilon p_3 q_1 + \tau p_3 q_2 + \xi p_4 q_1 + \epsilon p_4 q_2 & \mu p_3 + \eta p_4 & \eta q_1 + \mu q_2 - 1 & f_3 \\ \xi p_3 q_3 + \epsilon p_3 q_4 + \epsilon p_4 q_3 + \tau p_4 q_4 & \eta p_3 + \mu p_4 & \eta q_3 + \mu q_4 & f_4 \end{vmatrix} \qquad (2)$$

式(2)では、$\tau = 1-2\epsilon-\xi, \mu = 1-\epsilon-\xi, \eta = \epsilon+\xi$とする。無限回ゲームを繰り返したときプレイヤーX の期待利得は、

$$\mathrm{s_X} = \frac{\boldsymbol{v} \cdot \boldsymbol{U_X}}{\boldsymbol{v} \cdot \boldsymbol{1}} = \frac{D(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{U_X})}{D(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{1})} \qquad (3)$$

となり、同様に、プレイヤーY の期待利得は、

$$\mathrm{s_Y} = \frac{\boldsymbol{v} \cdot \boldsymbol{U_Y}}{\boldsymbol{v} \cdot \boldsymbol{1}} = \frac{D(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{U_Y})}{D(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{1})} \qquad (4)$$

のように式(2)を用いて表すことができる。ここで$\boldsymbol{1} =$ (1,1,1,1)とする。

α, β, γを任意の定数とすると、$\mathrm{s_x}$と$\mathrm{s_Y}$を線形結合したものが式(2)を用いると式(5)のように書ける。

$$\alpha \mathrm{s_X} + \beta s_Y + \gamma = \frac{D(\boldsymbol{p}, \boldsymbol{q}, \alpha\boldsymbol{U_X} + \beta\boldsymbol{U_Y} + \gamma\boldsymbol{1})}{D(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{1})} \qquad (5)$$

ここで、式(5)の右辺の分子がゼロになれば、2 人のプレイヤー間の期待利得は直線関係になる。行列式の性質から式(6)を満たせば直線関係になり、それはゼロ行列式戦略と呼ばれている。このゼロ行列式戦略は、式(6)にプレイヤーY の戦略$\boldsymbol{q}$の要素の項が含まれていないので、プレイヤーY 戦略に関係なく、利得の直線関係を一方的に設定できる。

$$\begin{gathered} \mu p_1 + \eta p_2 - 1 = \alpha R_E + \beta R_E + \gamma, \\ \eta p_1 + \mu p_2 - 1 = \alpha S_E + \beta T_E + \gamma, \\ \mu p_3 + \eta p_4 = \alpha T_E + \beta S_E + \gamma, \\ \eta p_3 + \mu p_4 = \alpha P_E + \beta P_E + \gamma \end{gathered} \qquad (6)$$

3. 結果

本研究では、エラー率ϵ, ξ を変化させて、エラーの影響によってプレイヤーの利得の関係がどのように変化するか、ゼロ行列式戦略の特に Equalizer 戦略と Extortioner 戦略、無条件戦略の ALLD 戦略と ALLC 戦略について、それぞれシミュレーションと数理解析を行った。ここでは、$\mathrm{L = G = 0.5}$と設定する。

3.1 ゼロ行列式戦略に対する知覚エラーの影響

プレイヤーX が Equalizer 戦略を採るとプレイヤーY がどんな戦略を採ろうとプレイヤーY の利得$\mathrm{s_Y}$は固定される。本研究で扱う Equalizer戦略はエラーなしでの$\mathrm{s_Y}$ =0.5とする戦略とし、式(6)で$\alpha = 0, \mathrm{p_1} = 2/3, p_4 = 1/3$とすると$\boldsymbol{p} =$

1 静岡大学 工学部 数理システム工学科
2 静岡大学学術院工学領域数理システム工学系列

(2/3,1/3,2/3,1/3) と決まる。プレイヤーX が Extortioner 戦略を採るとプレイヤーY がどんな戦略を採ろうとs_xがs_Yを上回る。本研究で Extortioner 戦略は、エラーなしの場合での直線の傾きを $\chi = 2$ にする戦略とし、式(6)で$\alpha = 2/7, \beta = -4/7, \gamma = 0$とすると$\boldsymbol{p} = (5/7,0,5/7,0)$と決まる。

図 1 はプレイヤーX の戦略を Equalizer 戦略として、プレイヤーY の戦略$\boldsymbol{q}$を乱数で 1,000 戦略分ランダムに決定して、それぞれについて無限回ゲームを行った時の期待利得の関係を示したものである。同様に、図 2 はプレイヤーX の戦略を Extortioner 戦略とした場合である。図 1 からエラーなしの場合、Equalizer 戦略なので 1,000 個のすべての点がs_Y =0.5 の直線上にあるのは自明だが、エラーありの場合では、1,000 個の点がs_Y =0.5 の直線上に沿っていないことがわかる。また、プレイヤーの期待利得はエラーありの場合でも直線関係を保っているように見える。実際、エラーありの場合でも、3.3 で示すようにプレイヤーの期待利得は直線関係となる。さらに、エラー率がある程度までは、エラー率が大きくなるほど、直線の傾きは負の方向に大きくなるという性質があることが分かった。よって、利得の直線関係は保てるが、s_Yが 0.5 に固定されないため、エラーがあると Equalizer 戦略ではなくなる。エラー率を最大$(2\epsilon + \xi = 1)$にした場合でもやはり利得の直線関係が成り立ち、特に$\epsilon = 0, \xi = 1$ のときの直線は、エラーなしの場合の直線s_Y =0.5と同じ直線であることが分かった。図 2 から、Extortioner 戦略で、エラーなしの場合では、直線の傾き$\chi = 2$となっているのは自明だが、エラー率が大きくなるほど、傾きが正の方向に大きくなっていき、さらにエラー率が大きくなると、直線が負の傾きになる性質があることが分かった。よって、エラーがある場合、s_Yがs_xを上回ることがあるので、直線関係は保てるが Extortioner 戦略ではなくなる。また、Extortioner 戦略でもエラーがある場合で Equalizer 戦略と同じように直線関係が保たれることが 3.3 の方法で示せる。

3.2 無条件戦略に対する知覚エラーの影響

常に裏切る ALLD 戦略$\boldsymbol{p} = (0,0,0,0)$、常に協力する ALLC 戦略$\boldsymbol{p} = (1,1,1,1)$についてもゼロ行列式戦略と同様にシミュレーションと数理解析を行った。ALLD 戦略について、エラーなしの場合と小さいエラー率の場合では、s_Yがs_Xを上回ることはなかったが、エラー率を最大にした場合で、特に$\epsilon = 0, \xi = 1$の場合と$\epsilon = 0.25, \xi = 0.5$の場合は$s_y = s_X$となる点を除いて常に$s_y$が$s_X$を上回ることが分かった。ALLC 戦略については、ALLD 戦略の反対で、エラーなしの場合と小さいエラー率の場合では、s_Xがs_Yを上回ることはなかったが、エラー率を最大にした場合で、特に$\epsilon = 0, \xi = 1$の場合と$\epsilon = 0.25, \xi = 0.5$の場合は$s_y = s_X$となる点を除いて常に$s_X$が$s_Y$を上回ることが分かった。また、これら２つの戦略は、エラーなしの場合、プレイヤーの利得関係は直線になるが、エラーありの場合でも直線関係になることが分かった。そのことは 3.3 から示せる。

3.3 利得関係が直線になることの確認

プレイヤーの利得関係が直線になることを示すには、式(6)を満たすことを確認すれば良い。例えば、エラー率が$\epsilon = 0.2, \xi = 0.2$のときに Equalizer 戦略が直線関係になるかを示す。式(6)にα, β, γ以外の変数を代入して、α, β, γについて解くと、$\alpha = 3, \beta = 5/3, \gamma = -7/3$と求めることができるので直線であることが示せる。同様に、$\epsilon = 0.2, \xi = 0.2$以外のエラー率でも直線になることが確認できた。さらに、Extortioner 戦略、ALLD 戦略、ALLD 戦略でも今回設定したエラー率では直線になることが確認できた。

4. おわりに

繰り返しゲームにおける代表的なゼロ行列式戦略や無条件戦略に対する知覚エラーの影響を数理解析、シミュレーションによって調査した。その結果、エラーを加えても無条件戦略は直線の利得関係を保てることが分かった。また、ゼロ行列式戦略の Equalizer 戦略と Extortioner 戦略は、エラーが入るとそれぞれ Equalizer 戦略と Extortioner 戦略ではなくなってしまうが、直線関係を保っているのでゼロ行列式戦略の性質は満たす。今後の展開として、これらの数学的性質が一般に成り立つかを調べる予定である。

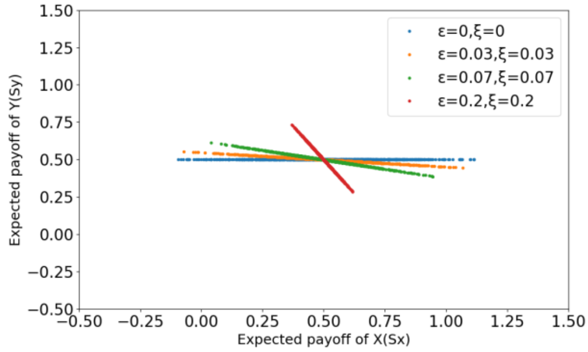

図 1Equalizer に対するエラーの影響

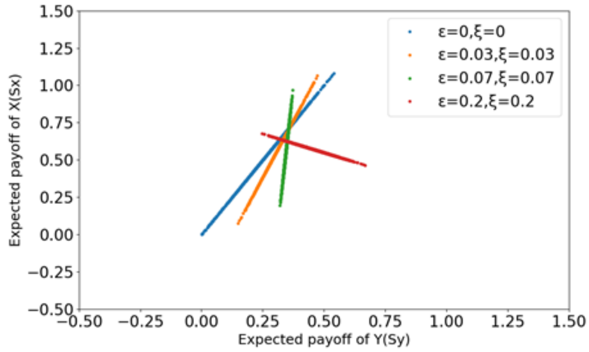

図 2 Extortioner に対するエラーの影響

参考文献

[1] Dong Hao, Zhihai Rong, Tao Zhou, “Extortion under uncertainty : Zero-determinant strategies in noisy games”, *Phys. Rev. E* 91, 052803 (2015).

[2] Tadashi Sekiguchi, “Efficiency in repeated prisoner's dilemma with private monitoring, *J. Econ. Theor*. 76, 345-361 (1997).

[3] William H. Press, Freeman J. Dyson, “Iterated Prisoner’s Dilemma contains strategies that dominate any evolutionary opponent”, *PNAS* 08540 (2012).