日常生活におけるゲーム理論

2003MM044 勝野幾子
指導教員: 佐々木克巳

1　はじめに

本研究では、鈴木[1]の第1部で述べられている、非協力ゲーム理論の理解を深めることを目的としている。ゲームをゼロ和と非ゼロ和の二通りに分け、ゼロ和ではミニマックス原理について述べ、非ゼロ和では囚人のジレンマを中心として考察を行う。ここでは、囚人のジレンマについて簡単に取り上げた後、無限繰り返し囚人のジレンマにおける有効的と思われる3つの戦略の証明を主とし、ノイズが発生したときに最適と思える戦略を考察する。

2　囚人のジレンマ

一般に、囚人のジレンマ型ゲームは次の表1のように定義することができる。各アルファベットの意味は次の通りである。R（Reward：報酬），S（Sucker：お人よし），T（Temptation：誘惑），P（Punishment：懲罰)。ただし、$S<P<R<T$，$S+T<2R$とする。

表 1: 囚人のジレンマ型ゲームの利得行列

	C(協力)	D(裏切)
C(協力)	(R, R)	(S, T)
D(裏切)	(T, S)	(P, P)

両プレイヤー共に自己の利得を最大にする最適反応戦略に従って戦略を選ぶと、囚人のジレンマにおける均衡点は戦略の組（D,D)であり、均衡利得は（P,P)である。これはどのプレイヤーも戦略を変更する誘因を持たない状態、つまりナッシュ均衡点となるが、より高い利得（R,R)が存在する。このとき、他のプレイヤーの利益を下げることなく、自分の利益を上げることができない状態なのでパレート最適ではない。

3　無限繰り返しゲーム

下の表は括弧内の左側の数値をAの利得とし、右側の数値をBの利得とする。

表 2: 囚人のジレンマの利得行列

A＼B	協力	裏切
協力	(5,5)	(0,8)
裏切	(8,0)	(2,2)

有限繰り返しのときには（裏切，裏切）のみが均衡経路となっていたものを、無限回繰り返されるものとする。なお、各プレイヤーのゲーム全体としての利得は、毎回得られる利得の合計とする。各回に得られる利得を$r_1, r_2, \ldots$とするとき、無限回繰り返しにおける利得を、

$$r_1 + r_2\,\delta + r_3\,\delta^2 + \ldots$$

と定義することにする。ここで、δは$0<\delta<1$を満たす実数であり、これを割引因子と呼ぶ。将来のゲームの利得を現在価値として計算するためδが掛けられ、それまでの利得に足されているのである。

このとき、有限繰り返しの場合と同じく、戦略Dを取り続ける支配戦略固定型戦略の組は均衡点になっている。

以下では、プレイヤーが取る戦略の選択肢を固定し、割引因子が十分に大きいときにナッシュ均衡となる戦略の組を示すものとする。ただし、3.1～3.3で考える戦略は、1つの戦略とそれにノイズと呼ばれる誤動作を入れた戦略のみとする。例えば、以下に述べる戦略aを両者ともに用いている状態から、一方が戦略を変えても利得が増えないことを示す。3.4では、3.1～3.3の戦略とそれらにノイズを入れた戦略の組を考える。

3.1　戦略a トリガー戦略

最初にCを選択し、以後は相手がCを取る限りは自身もCを取り、もし相手がDを取ればそれ以後何が起きても常にDを取る戦略のことをトリガー戦略という。

両プレイヤー共にトリガー戦略を用いている状態から、プレイヤー1がある回tでトリガー戦略を取ることをやめ、戦略Cから戦略Dに変え、またトリガー戦略に戻るとする。$(t-1)$回目までは両プレイヤーとも戦略Cを取っていたので、各回で利得4を得ている。t回目でプレイヤー1が戦略Cから戦略Dに変更するので、プレイヤー1の利得は8、プレイヤー2の利得は0となる。$(t+1)$回以降は、プレイヤー2は戦略Dを取り続けるのでプレイヤー1の利得は$(t+1)$回目では0、それ以降は2になる。以上を表にまとめると表3のようになる。表3はプレイヤー1の利得のみを示しているが、プレイヤー2もt回目と$(t+1)$回目の利得が入れ替わるだけなのでたいした違いはない。

表 3: 戦略aの繰り返し

	1	⋯	t-1	t	t+1	t+2	⋯
戦略aを続	5	⋯	5	5	5	5	⋯
t回目で逸脱	5	⋯	5	8	0	2	⋯

表3の上段と下段の利得の差を計算する。

$$(5+\ldots)-(5+\ldots+5\,\delta^{t-2}+8\,\delta^{t-1}+0\,\delta^{t}+2\,\delta^{t+1}+\ldots)$$
$$= -3\,\delta^{t-1}+5\,\delta+3\,\delta^{t+1}+\ldots$$
$$= \delta^{t-1}(-3+5\,\delta+3\,\delta^{2}+\ldots)$$
$$= \frac{\delta^{t-1}}{1-\delta}(-2\,\delta^{2}+8\,\delta-3)$$

これより、$(-2\,\delta^2+8\,\delta-3)>0$ならば、$t$回目に協調行動から逸脱しても利得は増加しないことが分かる。よって、この場合には戦略の組（a,a）でゲームは均衡する。これより、戦略の組（a,a）は均衡点となるのである。

3.2 戦略b 反射戦略

最初にCを選択し、以降は前回相手の選んだ選択をそのまま選ぶ戦略のことを反射戦略という。

戦略aと同様、両プレイヤー共に反射戦略を用いている状態から、プレイヤー1がある回tで戦略Dに変え、またそれ以降はそれまでどおりに反射戦略を続けたとする。すると、$(t+1)$回目までは利得8を得ていたが、t回目で以降は利得は8と0を交互に繰り返していくことになる。

$(5+...)-(5+...+5\,\delta^{\,t-2}+8\,\delta^{\,t-1}+0\,\delta^{\,t}+...)$
$=\ \delta^{\,t-1}(\frac{-3+5\,\delta^{\,2}}{1-\delta^{\,2}})$

これより、$(-3+5\,\delta^{\,2})$>0ならば、t回目に協調行動から逸脱しても利得は増加しないことが分かる。よって、この場合には戦略の組（b,b）は均衡点となるのである。

3.3 戦略c 仏の顔も3度まで戦略

最初はCを取り、それ以後相手がDを取っても3回までは許し、3回を超えると1回前の相手の行動の反射行動を取る戦略のことを仏の顔も3度まで戦略という。

先に挙げた2つの戦略のときと同様に、両プレイヤー共に仏の顔も3度まで戦略を用いている状態から戦略を変えて、という実験をすると、結果は戦略aとほぼ同じになる。よって、戦略の組（c,c）は均衡点となる。

3.4 その他

他にも、(a,b）,（a,c）,（b,c）が均衡点となる。 ここでは、(a,b）の証明だけをするものとする。プレイヤー1が戦略a、プレイヤー2が戦略bを取るものとし、プレイヤー1がある回tでCからDに変更し、また戦略aに戻るとする。プレイヤー1の利得を表4 に示す。

表 4: 戦略aとbの繰り返し

	1	···	t-1	t	t+1	t+2	t+3	···
戦略a	5	···	5	5	5	5	5	···
逸脱	5	···	5	8	0	8	2	···

これは戦略の組（a,a）と似たような結果を得ることができるので、この場合も戦略の組（a,b）は均衡点となる。他の戦略の組（a,c）は（a,b）に、(b,c）は（b,b）と似たような結果を得ることができるので、この2つもナッシュ均衡となるのである。戦略の組（a,a）,（b,b）,（c,c）,（a,b）,（a,c）,（b,c）は均衡点となる。

3.5 結果

以上より、割引因子が十分に大きければそれぞれの戦略の組がナッシュ均衡となることが分かった。これは、お互いの戦略が最適反応戦略となって均衡しているからである。

4 錯誤と寛容

Cを取っているにも関わらずDを取ったと判断される誤認や誤動作のことをノイズと言う。

鈴木[3]が行った実験を例に挙げる。表2の利得表で表現される囚人のジレンマを200回繰り返すものとし、逆戻り推論は行われないことを前提とする。また、各プレイヤーがトリガー戦略や反射戦略などの戦略を取っているときに、戦略が指示する行動とは違う行動を誤って取る確率を考え、相手にDを取ったとしても即座にそれを処罰する行動を取らずに、それを許容することを考えて、ゲームの状況を次のように定めることにする。

- 各プレイヤーは、自分の手番のときに、そのときとった戦略の指示とは異なる行動を、0.03の確率で取る。これは、各プレイヤーの意図とは無関係に誤った行動を取ることを意味する。
- どんな戦略を取った場合でも、その戦略の指示では戦略Dを取るべき回に、確率pで戦略Cを取る。
- 戦略a〜c以外に、よく定義された10個の戦略を総当たり戦で実験を行うものとする。
- 繰り返す回数は200回であり、この実験を10回行う。この繰り返しゲームにおける各プレイヤーの利得は、実験を10回行った平均値とする。

許容する確率が$p\ =\ 0$の場合には、仏の顔も3度まで戦略の組（c,c）が均衡点となる。その際の均衡利得は（4.12,4.12）で、ノイズが発生しない場合の（5,5）より小さくなる。ノイズの発生しない場合に均衡点となっていた（a,a）などその他の戦略の組は均衡点とならない。戦略cは元々相手がDをある程度許容する戦略であるから、多少のノイズが混じったとしても良い結果が得られたのである。

許容する確率が$p=0.15$〜0.45の場合には、反射戦略の組（b,b）が均衡点となる。均衡利得は許容度が大きくなるほど大きくなり、$p=0.45$の場合の均衡利得は（4.79,4.97）であり、$p=0$の場合での均衡点だった戦略の組（c,c）は均衡点とはならない。

許容する確率が$p=0.50$の場合には、トリガー戦略の組（a,a）が均衡点となる。このときの均衡利得は（4.18,4.18）となり、前の二つの場合での均衡点だった戦略の組は均衡点とはならない。このことは寛容の限度が50%以下であることを示している。

5 終わりに

本研究において、自分の利得を最大化するために行動することが必ずしもパレート最適にならないことを、囚人のジレンマ型ゲームを通して理解した。そして、一度きりの思考ではなく無限に思考し続ける無限繰り返しの場合、それぞれの戦略の組がナッシュ均衡となり、お互いの戦略が最適反応戦略となって均衡していることを、ノイズを入れた戦略を混ぜることにより証明した。今回、無限繰り返しゲームにおける有効的と思われる戦略を取り上げたのは3つであったが、もちろん他にもいくつかの戦略が考えられる。

参考文献

[1] 鈴木光男：新ゲーム理論, 勁草書房(1994).

[2] 鈴木光男：ゲーム理論入門, 共立出版株式会社(1981).

[3] 鈴木雄一:錯誤と寛容のある無限繰り返しゲームの均衡戦略, 東京理科大学工学部経営工学科卒業論文(1991).