繰り返しゲーム Repeated Game

– 合理的思考の技術 8 –

小林憲正

Department of Value and Decision Science (VALDES)
Tokyo Institute of Technology

May 26, 2014

繰り返しゲーム

同じゲーム $G = \langle N, A, u \rangle$ （これを成分ゲーム constituent game (stage game) という）のプレーを繰り返すゲーム。

- 繰り返しゲームでは、履歴は $h = (a^1, \ldots, a^t) \in A^t \subset H$ となる。a^t は t 期目の行動の組。
- 一般に無限の展開形ゲームを考えることができる。無限回繰り返しゲームもその一種。
- 各期に得られる利得 u をなんらかの方法で集計した利得 U が繰り返しゲームの利得となる。一般に、weak separability という性質を満たす。$\forall (a^t)_{t=1}^T \in A^T \forall a' \in A$

$$u_i(a^t) \geq u_i(a') \Rightarrow U_i((a^s)_{s=1}^T) \geq U_i(a^1, \ldots, a^{t-1}, a', a^{t+1}, \ldots, a^T)$$

割引利得 Discounted Payoff

無限繰り返しゲームの定式化としては、各期に利得を得て、それを割引因子 discount factor で割り引いて足しあわせた割引利得を全体ゲームの利得とするものが最も代表的である。

Definition (割引利得 Discounted Payoff)

$1, \ldots, T$ 期に得られる利得の列 $(u_i(a^t))_{t=1}^T$ の割引因子 $\delta_i \in (0,1)$ のもとでの割引利得は、

$$U_i((a^t)_{t=1}^T) := (1-\delta_i) \sum_{t=1}^{T} \delta_i^{t-1} u_i(a^t)$$

割引利得に関する注意

- 右辺冒頭の

$$1 - \delta_i = \frac{1}{\sum_{t=1}^{\infty} \delta_i^{t-1}} \qquad (\delta_i \in (0,1))$$

は、重み付き平均のための規格化定数。後述の成分ゲームと関係付けた理論分析の表現を簡単にする便宜上導入される。

- 割引利得は、ファイナンスにおいて一定利子率で割り引く割引現在価値 net present value (NPV) の一般化。
割引因子 δ_i は一般には $i \in N$ により異なる。
- 行動経済学では、割引因子のべき乗による割引利得がしばしば批判され、より現実の意思決定主体の感覚を表現する候補として双曲線割引 hyperbolic discounting も提案されている [10] が、本講義の理論分析では両者の間に大きな違いはない。

平均利得 Means Payoff

有限回繰り返しゲームの分析で通常用いられる効用関数としては、単純に各期の利得の和をとったものがある。

Definition (平均利得 Means Payoff)

$$U_i((a^t)_{t=1}^T) := \sum_{t=1}^{T} \frac{1}{T} u_i(a^t)$$

- 割引利得同様に、$1/T$ で規格化し、数理分析的には等価な平均値で表現する。
- $T \to \infty$ の極限をとっても効用関数が発散しないようにする効果もある。

プレーヤー間の協力により達成可能な利得

無限回繰り返しゲームでは、適当な回数の割合で A 上の結果を混ぜてプレーすることで合意する（公的混合 public randomizing）ことにより、以下の集合上の任意の利得の組を近似的に達成できる。

Definition (達成可能 Feasible な利得プロファイル)

$G = \langle N, A, u \rangle$ 上で達成可能な利得プロファイル $v \in \Re^N$ の集合は、A 上の結果で得られる利得プロファイルの凸結合

$$\Delta(\{u(a) | a \in A\})$$

混合戦略との比較

達成可能な利得の組は混合戦略の組による効用可能集合 UPS よりも広い。

Q. 相関均衡点を思い出しながら、なぜ、達成可能な利得プロファイルの集合のほうが混合戦略の組の UPS より広いか、考えてみよう。

個人合理性

Definition (ミニマックス Minmax 利得)

$$\underline{v}_i = \min_{a_{-i}\in A_{-i}} \max_{a_i \in A_i} u_i(a)$$

上記の解 $p_{-i} \in A_{-i}$ は、プレーヤー i に対する最も厳しい制裁 punishmentと解釈できる。

Example (囚人のジレンマ)

$$\underline{v} = u(D, D), \qquad p_1 = D, p_2 = D$$

Definition (個人合理性 Individual Rationality)

$u \in \Re^N$

- $u \geq \underline{v}$ 個人合理的 individually rational (enforceable[9])
- $u \gg \underline{v}$ 狭義個人合理的 strictly individually rational (strictly enforceable[9])

Example (囚人のジレンマ)

1 \ 2	C	D
C	$3, 3$	$1, 7$
D	$7, 1$	$2, 2$

Q. この囚人のジレンマの達成可能、かつ狭義個人合理的な利得プロファイルの集合を $u_1 u_2$ 空間上に描き、(C, C) がこの集合上でパレート効率的でないことを確かめよ。

このことより、

囚人のジレンマの定義

通常の序数効用の満たす性質に加えて、

$$u(C, C) > \frac{1}{2}(u(D, C) + u(C, D))$$

を満たす利得双行列のみを囚人のジレンマと呼ぶこともある。

機械 Machine (オートマトン Automaton) 的戦略 [9]

後に見るように、繰り返しゲームの解は単純な状態遷移的戦略の組によって特徴付けが可能であることが知られている。このことが、単純な戦略を遺伝子で表現する進化ゲーム理論の発端ともなる。

Definition (オートマトン的戦略)

$\langle N, A, u\rangle$ の繰り返しゲームにおけるプレーヤー i のオートマトン的戦略は：

- Q_i 状態 state の集合
- $q_0 \in Q_i$ 初期状態 initial state
- $f_i : Q_i \to A_i$ 出力関数 output function
- $\tau_i : Q_i \times A \to Q_i$ 状態遷移関数 transition function

オートマトン的戦略の入力は今期の行動の組のみという意味でマルコフ的 Markovian である [7]。

オートマトン的戦略の例

Definition (トリガー戦略 trigger strategy)

- 他のプレーヤーが合意内容に従ってプレーしている間は、自分も合意内容に従う
- 他のプレーヤー誰かひとり j が逸脱した場合、その次の期からプレーヤー $i \neq j$ は j に対して制裁を課すプレーをとり続ける
- これ以外の詳細（例えば、逸脱したプレーヤーがその後どうするか、とか 3 人以上のゲームで複数の他のプレーヤーが同時に逸脱した場合への対処）は任意

Example (囚人のジレンマ)

合意内容 (C, C)、逸脱 D、制裁 D

フォーク定理 Folk Theorem

Theorem (ナッシュ· フォーク定理 [6])

狭義個人合理的かつ達成可能な成分ゲーム G の任意の利得の組 w に対し、w と無限に近い利得の組を結果とする G の無限回繰り返しゲームのナッシュ均衡が存在する。

Proof.

minmax 戦略 p_{-i} を制裁とするトリガー戦略の組はナッシュ均衡。 □

Example (囚人のジレンマ)

1 \ 2	C	D
C	$3, 3$	$1, 7$
D	$7, 1$	$2, 2$

合意 – (C,D), (D, C) を交互に繰り返す, 制裁 – D のトリガー戦略の組はナッシュ均衡

愛と憎しみは表裏一体!? – フォーク定理と感情

フォーク定理では、長期関係における信用を達成する装置としては制裁が基本的。しかし、一般の展開型ゲーム同様、制裁にはしばしばコストがかかるため、信憑性があるとは限らない。
これに関連して、感情や行動の遺伝形質に関する研究がたくさんある：

実験経済学 意地悪 spite な感情がむしろ協力を促進しやすい [3, 4]

進化ゲーム理論 “punish or perish”[11] – 協力と報復 retaliation は共進化 coevolve

完全フォーク定理 Perfect Folk Theorem

Theorem

成分ゲーム G における行動の組 $a \in A$ において、各プレーヤーj に対して G のあるナッシュ均衡 $a^{j*} \in A$ が存在して、$u_j(a) > u_j(a^{j*})$ とすると、*SPNE* が存在して、その結果は a のみを無限回繰り返すようにすることができる。

Proof.

i のみが逸脱した場合、$j \neq i$ が a_j^{i*} を制裁とするプレーを繰り返し、i が a_{-i}^{i*} をこの制裁に対する最適応答としてプレーする戦略の組は SPNE。 □

より一般には、制裁したプレーヤーに報いることによって、より広範囲の利得の組が SPNE としても達成可能 [6] 。

Example (囚人のジレンマ)

囚人のジレンマでのトリガー戦略の制裁 D に対して、逸脱者も D をとると、トリガー戦略の組は SPNE にもなる。

Example (チキン)

1 \ 2	Dove	Hawk
D	$3, 3$	$1, 4$
H	$4, 1$	$0, 0$

- (D, D) を繰り返すことを合意とし、単一プレーヤーが逸脱するまでは、これをプレーする
- 相手が逸脱して H をとると、報復措置として、その次の期より、H をとり続け、相手はビビって最適応答として D をとり続ける

という戦略の組は、このチキンゲームの無限回繰り返しゲームの SPNE。

有限回繰り返しゲームと部分ゲーム完全均衡

Theorem (有限回繰り返しゲームと部分ゲーム完全均衡)

有限回繰り返しゲームでは、成分ゲームのナッシュ均衡が一つしか存在しない場合、その均衡をすべての期にプレーする戦略プロファイルのみが部分ゲーム完全均衡となる。

Example (囚人のジレンマ)

有限回繰り返し囚人のジレンマゲームの唯一のナッシュ均衡は、全てのステージで、D をプレーする組。

成分ゲームに均衡が二つ以上存在する場合は、パレート支配される均衡のプレーを脅しに用いることにより、最終期のみ裏切るような、フォーク定理と似たようなプレーが均衡となり得る。[5]

Example (囚人のジレンマ + 鹿狩 Stag Hunt)

囚人のジレンマと鹿狩を合わせた以下のゲームについて考える。

	Coop	Defect	Punish
C	10, 10	5, 11	0, 8
D	11, 5	9, 9	0, 8
P	8, 0	8, 0	7, 7

概要 – ゲームのナッシュ均衡は (D, D), (P, P) の 2 つ。最終期 T は (D, D) をプレーすることを既定路線とし、それ以前に相手が逸脱した場合に (P, P) のプレーを信憑性のある制裁としてちらつかせる。

- (C, C) を $T-1$ 期まで繰り返しプレーし、最終期 T は (D, D) をプレーすることを合意とする。この合意内容に従う場合、最終期の部分ゲームはナッシュ均衡。
- 最終期前に、もし相手が逸脱した場合は、制裁措置として、その次の期より自分は P をプレーするトリガー戦略をとり、相手もそれに対して最適応答 P をとるとすると、この部分ゲームのプレーも成分ゲームのナッシュ均衡の繰り返しなので部分ゲーム完全。
- 相手の逸脱による利益 1 は、制裁による損失の最低値 2 を下回る。

望ましい小さな非合理性としての ϵ-均衡

Definition (ϵ-均衡)

$\epsilon > 0$ に対して、戦略プロファイル s^* が ϵ-均衡であるとは、$\forall s_i \in S_i$,

$$u_i(s_i, s^*_{-i}) - u_i(s^*) \leq \epsilon$$

Example (囚人のジレンマ)

1 \ 2	C	D
C	$3,3$	$1,7$
D	$7,1$	$2,2$

の 100 回繰り返しゲームを考える。

- SPNE では全ステージで (D, D) をプレーし平均利得 (2,2)
- (C,D), (D, C) の交互プレーを合意、D を制裁とするトリガー戦略の組は $\epsilon \geq 1$ で ϵ-均衡であり、平均利得はほぼ (4, 4)

ムカデゲームのプレー同様に、微小な非合理性が劇的な改善をもたらし得るかもしれない。

有限回繰り返しと無限回繰り返しの接続

繰り返しゲームの理論では、完全フォーク定理と有限回繰り返しゲームの SPNE に大きな違いがある。モデルの頑健性 robustness の観点からすると望ましくなく見える。この乖離を埋める試みがいくつかある。

- 割引因子が十分に大きなプレーヤー同士が十分長期にわたってプレーする場合、ϵ-均衡の ϵ をゼロに近くしてフォーク定理と似たような結果が得られる。[2, 1]
  ($\epsilon - \delta$ 論法に類似)
- 似たような分析として、最終期 T に関する微小な共有知識の破れを仮定しただけでも、フォーク定理と似た結果が得られることが知られている。[8]

繰り返しゲームとヒステリシス Hysterisis

- 繰り返しゲームでは、履歴は過去のプレーのみによって特徴付けられる。つまり、同じ t 期であれば、その期以降の各部分ゲームの物理的環境はほぼ不変。それぞれの h 以降のプレーで得られる利得分布は、$\sum_{t'=1}^{t} \delta_i u_i(a^{t'}) \quad (\sum_{t'=1}^{t}(1/t)u_i(a^{t'}))$ だけ平行移動しただけ。
- また、無限回繰り返しゲームでは、有限オートマトンのように、過去のプレーのみ考慮してプレーを決定する単純なアルゴリズムで SPNE をプレー可能。

Q. 以上を考慮して、小論テーマ「我々はなぜ歴史を学ぶのか？」について再考せよ。

まとめ – 繰り返しゲームの面白さ

- 協力関係の直感のモデル化
- 均衡の著しい多様性
- 調整の必要性
- 無限と有限の著しい違い
- 非合理性の効果

References

[1] Jean-Pierre Benoit and Vijay Krishna.
Finitely repeated games.
Econometrica, 53(4):905–922, 1985.

[2] Jean-Pierre Benoit and Vijay Krishna.
Nash equilibria of finitely repeated games.
International Journal of Game Theory, 16(3):197–204, 1987.

[3] Timothy N. Cason, Tatsuyoshi Saijo, Takehiko Yamato, and Konomu Yokotani.
Non-excludable public good experiments.
Games and Economic Behavior, 49(1):81 – 102, 2004.

[4] TimothyN. Cason, Tatsuyoshi Saijo, and Takehiko Yamato.
Voluntary participation and spite in public good provision experiments: An international comparison.
Experimental Economics, 5(2):133–153, 2002.

[5] James W. Friedman.
Cooperative equilibria in finite horizon noncooperative supergames.
Journal of Economic Theory, 35(2):390 – 398, 1985.

[6] Drew Fudenberg and Eric Maskin.
The folk theorem in repeated games with discounting or with incomplete information.
Econometrica, 54(3):533–554, 1986.

[7] Eric Maskin and Jean Tirole.
Markov perfect equilibrium: I. observable actions.
Journal of Economic Theory, 100(2):191 – 219, 2001.

[8] Abraham Neyman.
Cooperation in repeated games when the number of stages is not commonly known.
Econometrica, 67(1):45–64, 1999.

[9] Martin J. Osborne and Ariel Rubinstein.
A Course in Game Theory.
MIT Press, Cambridge, 1994.

[10] Ariel Rubinstein.
“ economics and psychology”? the case of hyperbolic discounting.
International Economic Review, 44(4):1207–1216, 2003.

[11] Karl Sigmund.
Punish or perish? retaliation and collaboration among humans.
Trends in Ecology Evolution, 22(11):593 – 600, 2007.