# 完全情報ゲームの拡張と繰り返しゲーム Repeated Game

## -- 合理的思考の技術 7 --

小林憲正

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

May 27, 2013

# Outline

# 完全情報ゲームの定式化

## Definition (完全情報ゲーム)

$\Gamma = \langle \mathcal{K}, P, u \rangle$:

- $\mathcal{K} = \langle H, A \rangle$ 根付き木 rooted tree:
  - $H$ ノード node（履歴 history ）の集合
    最終履歴 terminal history の集合を $Z(\subset H)$ と書く
  - $A$ 枝 arc（行動 actions ）の集合
- $P : H \setminus Z \to N$ 各決定ノードに、そこで決定するプレーヤーを対応付けるプレーヤー関数
  - $N$ プレーヤーの集合
- $u_i : Z \to \Re$ プレーヤー$i \in N$ の効用関数.

# ノード・枝と履歴・行動

完全情報ゲームの木では、決定ノードと履歴、枝と行動が丁度一対一対応する。

- 木の根に対応して、空の履歴 null history （初期履歴 initial history）$\varnothing \in H$ という記号を用いる
- 履歴を空の履歴からの行動プロファイル（行動の組）の列で表すことができる e.g.) $h = (a^1, ..., a^T)$
  (右上の添字は、プレーヤーの添字ではないことに注意。)
- 同様に、履歴の合成を考えることができる e.g.) $h = (h', h'')$

# 不確実性を含む拡張

## Definition (不確実性を含む完全情報ゲーム)

$\Gamma = \langle \mathcal{K}, P, f_c, u \rangle$: (拡張部分のみ記述)

- $P : H \setminus Z \to N \oplus \{c\}$ がプレーヤー関数 (プレーヤー$c$ は不確実性 chance または自然 nature と呼ばれる),
- $f_c(\cdot|h) \in \Delta(A(h))$ は $h \in P^{-1}(\{c\})$ に対する $A(h)$ 上の確率分布

後ろ向き帰納法は、決定の木と同様に行う。

# 同時決定を含む拡張 [7]

## Definition (同時決定を含む完全情報ゲーム)

$\Gamma = \langle \mathcal{K}, P, u \rangle$: (拡張部分のみ記述)

- $P : H \setminus Z \to \mathcal{P}(N)$ 各履歴 $h \in H$ に、その直後に行動するプレーヤーの部分集合を対応させる関数,
- $A(h) = \times_{i \in P(h)} A_i(h)$ 各履歴 $h$ 直後の行動の集合上のプレーヤー分割.

## Definition (戦略 Strategy)

同時決定を含む完全情報ゲーム $\Gamma = \langle \mathcal{K}, P, u \rangle$ におけるプレーヤー $i \in N$ の (純粋) 戦略は、全ての $i \in P(h)$ を満たす終端でない履歴 $h \in H \setminus Z$ に $A_i(h)$ 上の行動を対応付ける関数。

# 部分ゲーム

## Definition (部分ゲーム Subgame)

同時決定を含む完全情報ゲーム $\Gamma = \langle \mathcal{K}, P, u \rangle$ の履歴 $h \in H$ 後の部分ゲームとは、$\Gamma$ の構造を全て保存したゲーム $\Gamma(h) = \langle N, \mathcal{K}|_h, P|_h, u|_h \rangle$:

- $\forall h' \in H|_h, (h, h') \in H$
- $\forall h' \in H|_h, P|_h(h') = P(h, h')$
- $\forall h' \in Z|_h, u_i|_h(h') = u_i(h, h')$

## Definition (戦略の制限 Restriction)

$\Gamma = \langle \mathcal{K}, P, u \rangle$ におけるプレーヤー$i \in N$ の戦略 $s_i$ の部分ゲーム $\Gamma(h)$ 上への制限 $s_i|_h$ は以下を満たす

$$\forall h' \in H|_h, s_i|_h(h') = s_i(h, h')$$

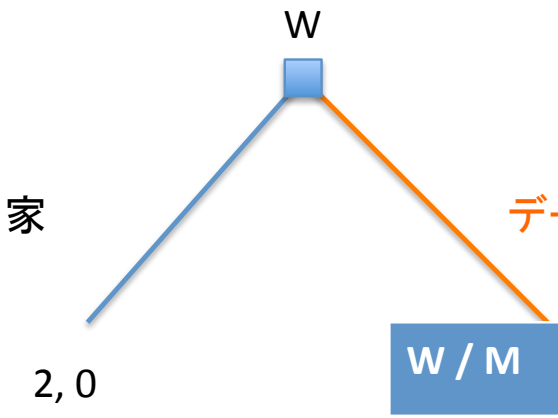
# 部分ゲーム完全ナッシュ均衡 Subgame-Perfect Nash Equilibrium (SPNE)

## Definition (部分ゲーム完全ナッシュ均衡 SPNE)

同時決定を含む完全情報ゲーム $\Gamma = \langle N, \mathcal{K}, P, u \rangle$ における戦略プロファイル $s^*$ が部分ゲーム完全均衡であるとは、$\forall h \in H$ の部分ゲーム $\Gamma(h)$ において、$s^*$ の $\Gamma(h)$ 上への制限がナッシュ均衡となっていることである。

- 部分ゲーム完全ナッシュ均衡は完全情報ゲームにおける後ろ向き帰納法による解の拡張である。
- 有限完全情報ゲームには（純粋戦略の）部分ゲーム完全ナッシュ均衡が存在する（存在定理）。これは後ろ向き帰納法を用いて数学的帰納法で証明できる。

# 同時決定を含む展開型ゲームの例




| W / M | 美術館 | スポーツ |
|---|---|---|
| 美術館 | 3, 1 | 0, 0 |
| スポーツ | 0, 0 | 1, 3 |

W の主張「デートでスポーツ鑑賞するくらいなら、家にいる」
$\rightarrow$ ２段階目のデートゲームでは W は断固美術館を選ぶ (!?)
$\rightarrow (s_W, s_M)$ = ( (家, スポーツ), スポーツ) は SPNE ではあるが、妥当 reasonable な均衡でないとみなされる。

# 精緻化 Refinement 基準の例 -- 前向き帰納法 Forward Induction[5]

- W の主張に基づいて妥当でない均衡を除去する考え方を前向き帰納法 forward induction という。
  - 一般には、前向き帰納法により除去されない解は安定 stable であるとも言われ、標準形表現をした際に、弱支配戦略の除去によって除去されない。
  - Q. 先のオプション付きデートゲームの唯一の前向き帰納法で残る SPNE ((デート, 美術館), 美術館) がこの性質を満たすことを確かめよ。
- 後ろ向き帰納法や前向き帰納法のように、納得のいく基準を満たすナッシュ均衡のみを解とみなし、行動の予測精度を高めようとする方法を均衡の精緻化 refinement という。
  - 精緻化基準は状況によらず普遍的に異論なく適用可能というわけではない。
  - Q. 前向き帰納法の突っ込みどころを探してみよう！

# Outline

# 無限繰り返しゲーム

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 $h = (a^1, ..., a^t) \in A^t \subset H$ 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。
- 後に見るように、$T$ 回繰り返しゲームの $T$ を限りなく大きくした有限回繰り返しゲームと無限回繰り返しゲームでは、均衡の様相が大きく異なる。

# 割引利得 Discounted Payoff

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。[1]

## Definition (割引利得 Discounted Payoff)

1, ..., 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、

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

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 $\delta_i$ は一般には $i \in N$ により異なる。

[1]割引利得のモデルの候補として双曲線割引 hyperbolic discounting も提案されているが、本講義の理論分析では両者の間に大きな違いはない。

# 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、その期に属する任意の履歴 $h \in A^t$ でプレーするゲームの物理的環境は割引利得モデルではほぼ不変。それぞれの $h$ 以降のプレーで得られる利得分布は、$\sum_{t'=1}^{t} \delta_i u_i(a^{t'})$ だけ平行移動しただけ。
- しかし、(後述するが) 繰り返しゲームでは、過去の他のプレーヤーのプレーに依存して制裁 punishment や報酬 reward を施すことが、プレーの本質となる。
- 上記は、我々にとって、サンクコストの誤謬と反対に、とりわけゲーム状況において過去の記憶が重要な一つの大きな要因。
- ゲーム理論と標準的な自然科学理論との大きな違いの一つは、前者では一般に履歴に大きく依存せずに状態を記述することが難しいことがある。(非マルコフ性)

# 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、すなわち

$$\{\sum_{a \in A} \alpha_a u(a) | (\forall a \in A, \alpha_a \geq 0) \wedge \sum_{a \in A} \alpha_a = 1\}$$

ここでは、効用に四則演算を施していることに注意。繰り返しゲームの分析では、純粋戦略の分析の範囲内でも基数効用としての性質が重要である。

# 個人合理性

## Definition (ミニマックス Minmax 利得)

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

プレーヤー$i$ に対する最も厳しい制裁から得られる $i$ の利得。

## Definition (個人合理性 Individual Rationality)

$u \in \Re^N$

- $u \geq \underline{v}$ 個人合理的 individually rational
- $u \gg \underline{v}$ 狭義個人合理的 strictly individually rational

達成可能かつ個人合理的な利得プロファイルの集合を $V$ と表すことがある。

# トリガー戦略 Trigger Strategy としっぺ返し Tit for Tat

## Definition (トリガー戦略)

$v \in V$ をプレーヤー間で達成を目指す利得の合意内容とする。以下の戦略をトリガー戦略と呼ぶ。

- 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 $j$ に対して制裁を課すプレーをとり続ける。

似たようなプレーの仕方に「しっぺ返し（目には目を、歯には歯を）」がある。これは、相手の戦略を一期遅れで模倣する戦略。

# フォーク定理 Folk Theorem

**Theorem (ナッシュ・フォーク定理 [4])**

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 $w \in V$ に対し、$v$ 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 戦略を制裁とするトリガー戦略の組はナッシュ均衡。 □

一般の展開型ゲーム同様、制裁に信憑性があるとは限らない。

# 実例 -- チキン

## Example (チキン)

| | ハト | タカ |
|---|---|---|
| ハト | $3,3$ | $1,4$ |
| タカ | $4,1$ | $0,0$ |

(ハト, ハト) を繰り返すことを合意とし、両者の制裁ともタカ戦略とするトリガー戦略の組は、このゲームの無限回繰り返しゲームのナッシュ均衡。

制裁には大きな痛みが伴い、片方が妥協してしまいたい誘惑にかられる。

# 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 $a \in A$ が、$G$ のあるナッシュ均衡 $a^* \in A$ を強パレート支配しているとする（すなわち $u(a) > u(a^*)$ 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.**

$a^* \in A$ を制裁にしたトリガー戦略の組は SPNE。 □

より一般には、特に３人以上での場合、制裁したプレーヤーに報いることによって、ナッシュフォーク定理と同様の広範囲の利得の組が SPNE としても達成可能。

# 実例 -- 囚人のジレンマ

## Example (囚人のジレンマ)

| | C | D |
|---|---|---|
| C | $3,3$ | $1,6$ |
| D | $6,1$ | $2,2$ |

(C, D) と (D, C) を適当な割合でプレーすることを繰り返すことを合意とし、制裁は (D, D) の繰り返しとするトリガー戦略の組は SPNE。

(C, C) は一回交代で交互に (C, D) と (D, C) をプレーする繰り返しにパレート支配される。この分析により、この例のような利得双行列を囚人のジレンマと呼ばないこともある。(C を繰り返すのが協力的と解釈する立場)

# 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。

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。[3]

# 望ましい小さな非合理性としての $\epsilon$-均衡

## Definition ($\epsilon$-均衡)

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

$$u_i(s_i, s_i^*) - u_i(s^*) \leq \epsilon$$

- 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、$\epsilon$ 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。[6]

# 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

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

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

[5] Elon Kohlberg and Jean-Francois Mertens.
On the strategic stability of equilibria.
Econometrica, 54(5):1003--1037, 1986.

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

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