ゲーム理論

第4回 繰り返しゲーム

佐賀大学大学院 工学系研究科
知能情報システム学専攻
上田　俊
Email: sgrueda@cc.saga-u.ac.jp
https://sites.google.com/view/sgrueda/in-japanese

途中経過

- 第2回 第2章 戦略形ゲーム
- 第3回 第3章 展開形ゲーム
- 第4章 完全均衡点
  - ナッシュ均衡の精緻化
  - 部分ゲーム完全均衡点
  - 逐次均衡点
- 第5章 情報不完備ゲーム
  - ベイジアンゲーム
  - ベイジアン・ナッシュ均衡
- 第4回 第6章 繰り返しゲーム

アウトライン

- 囚人のジレンマ (おさらい)
- 繰り返しゲーム
  - 将来利得に対する割引
- 無限回繰り返し囚人のジレンマ
  - 代表的な戦略
  - フォーク定理
- 有限回繰り返し囚人のジレンマ
  - 弱い均衡概念による協力の達成

囚人のジレンマ (1/2)

- 今日の目的: 如何にして（黙秘, 黙秘）を均衡状態で達成するか.
- 1回のプレイでは（ほぼ）不可能
  - （自白, 自白）が支配戦略均衡
- 何回かゲームを繰り返す.

	黙秘	自白
黙秘	(1年, 1年)	(10年, 3ヵ月)
自白	(3ヵ月, 10年)	$\boxed{(8年, 8年)}$

支配戦略均衡

囚人のジレンマ (2/2)

- 囚人のジレンマの一般化した利得行列
- プレイヤーの行動
  - 行動C (協力, cooperation)
  - 行動D (裏切り, defection)
- 以下の条件が成立
  - $T > R > P > S$
  - $2R > (S + T)$

	C	**D**
C	(R, R)	(S, T)
D	(T, S)	(P, P)

繰り返しゲーム

- (戦略形) ゲームを何回か繰り返し行うゲーム.
  - 繰り返すゲームを成分ゲームと呼ぶ.
- 次のようなルールを持つ:
  - 毎回, プレイヤーは成分ゲームの行動を他とは独立に選択する.
  - 行動を選択するとき, プレイヤーは過去のプレイを完全に知る. (完全観測)
  - すべてのプレイヤーは割引因子 $\delta (0 < \delta < 1)$ による割引利得和を最大にする.

割引因子

- 将来利得に対する割引因子 (discount factor)
- 現在の利得と将来の利得は異なる.
  - 現在の100円 > 1ゲーム後の100円
  - $\delta = \frac{1\text{ゲーム後の}100\text{円}}{\text{現在の}100\text{円}}$
- tゲーム後に得る100円を現在の価値に換算すると $100\delta^{t-1}$
- 1ゲームごとに無限回100円を得続けると, 割引利得和は…
  - $100 + 100\delta + 100\delta^2 + \cdots = \frac{100}{1-\delta}$

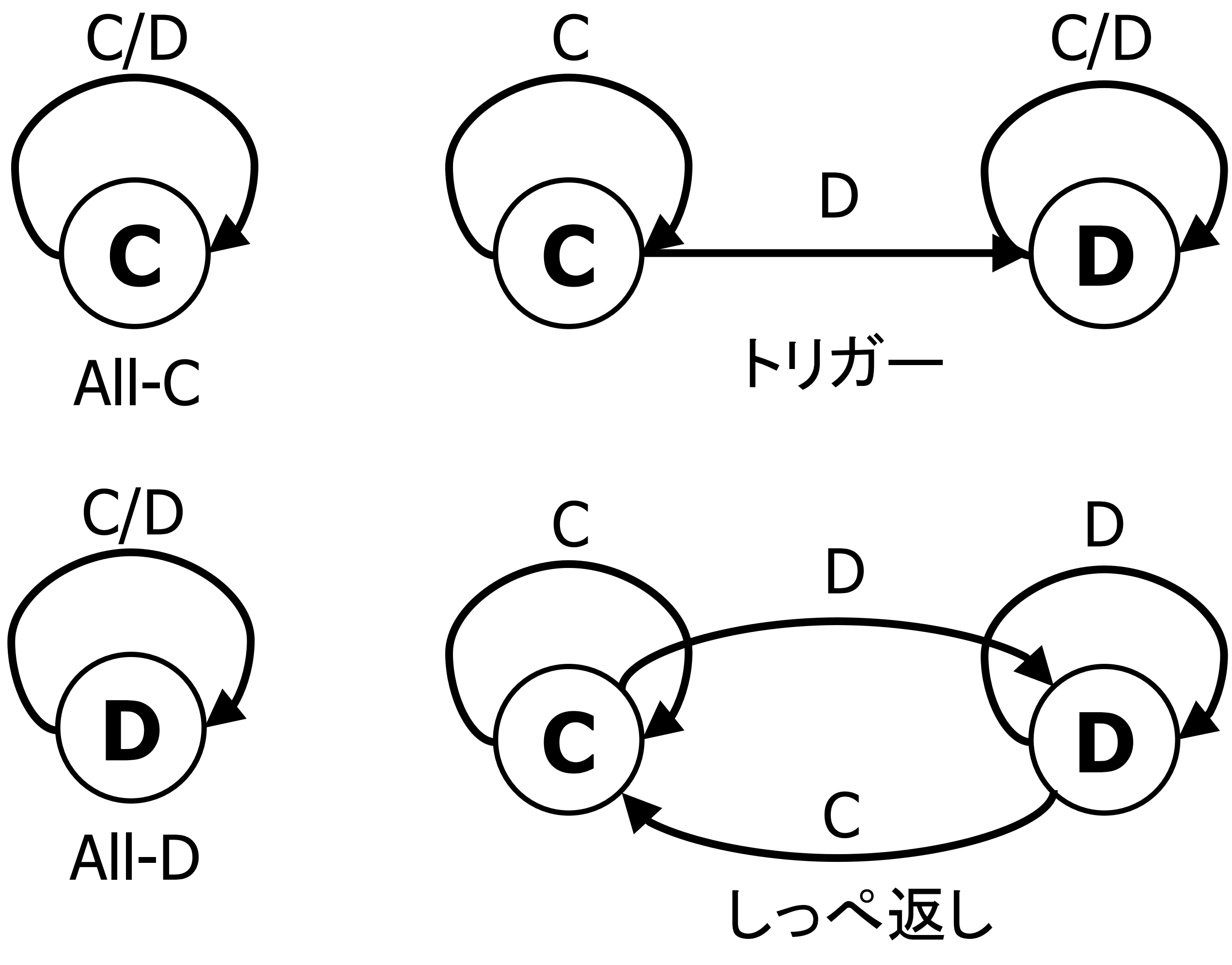
アウトライン

- 囚人のジレンマ (おさらい)
- 繰り返しゲーム
  - 将来利得に対する割引
- 無限回繰り返し囚人のジレンマ
  - 代表的な戦略
  - フォーク定理
- 有限回繰り返し囚人のジレンマ
  - 弱い均衡概念による協力の達成

繰り返し囚人のジレンマの戦略

- All-C: 過去のプレイ結果によらず，常に C をとる.
- All-D: 過去のプレイ結果によらず，常に D をとる.
- トリガー (trigger): 最初は C をとる．以後，双方が C をとる限り C をとる．しかし，1回でも一方が D をとれば，その後，D をとり続ける.
- しっぺ返し (tit for tat): 最初は C をとる．以後，相手の前回の行動と同じものをとる.

戦略のオートマトンを用いた表現



ナッシュ均衡のチェック

	All-C	All-D	トリガー	しっぺ返し
All-C	×	×	×	×
All-D		○	×	×
トリガー			○	?
しっぺ返し				○

トリガー戦略によるナッシュ均衡

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

- 定理 繰り返し囚人のジレンマ･ゲームにおいて, プレイヤーの将来利得に対する割引因子 δ が,
$$\delta \geq \frac{T-R}{T-P}$$
であるならば, トリガー戦略の組は繰り返しゲームのナッシュ均衡点である.

トリガー戦略によるナッシュ均衡

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

- 2人がともにトリガー戦略を用いるときの割引利得和
  - $R + \delta R + \delta^2 R + \cdots = \frac{R}{1-\delta}$
- P_1が t 回目で行動を D に変更すると…
  - P_2は $t+1$ 回目以降，Dを取り続ける．
  - t 回目以後 (t 回目を含む) のP_1の割引総利得和

$$T + \delta P + \delta^2 P + \cdots = T + \frac{\delta}{1-\delta} P$$

トリガー戦略によるナッシュ均衡

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

- 先ほどの割引総利得和を比較し，逸脱しない方が利得が高くなれば良いので，

$$\frac{R}{1-\delta} > T + \frac{\delta}{1-\delta}P$$

- プレイヤー2の方も同様．
- したがって，δ が条件を満たせば，トリガー戦略の組がナッシュ均衡になる．■
- 例えば，$P = -3, R = 5, S = -4, T = 6$とすると，$\delta \geq {}^{1}/_{8}$ ならば，条件を満たす．

しっぺ返しによるナッシュ均衡

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

- 定理 繰り返し囚人のジレンマ･ゲームにおいて, プレイヤーの将来利得に対する割引因子 δ が,
$$\delta \geq \max\left(\frac{T-R}{T-P}, \frac{T-R}{R-S}\right)$$
であるならば, しっぺ返し戦略の組は繰り返しゲームのナッシュ均衡点である.
- 証明は省略.

無限回繰り返しゲームでのナッシュ均衡

- 戦略形ゲームでは，混合戦略まで含めれば，ナッシュ均衡となる戦略の組が少なくともひとつ存在する.
- 無限回繰り返すと？
  - 成分ゲームでナッシュ均衡であれば，無限回繰り返しゲームでもナッシュ均衡
  - 無限回繰り返すことでナッシュ均衡となる戦略の組は増える (でないと繰り返す意味がない)
  - ではどのくらい増える？

ミニマックス利得と個人合理性

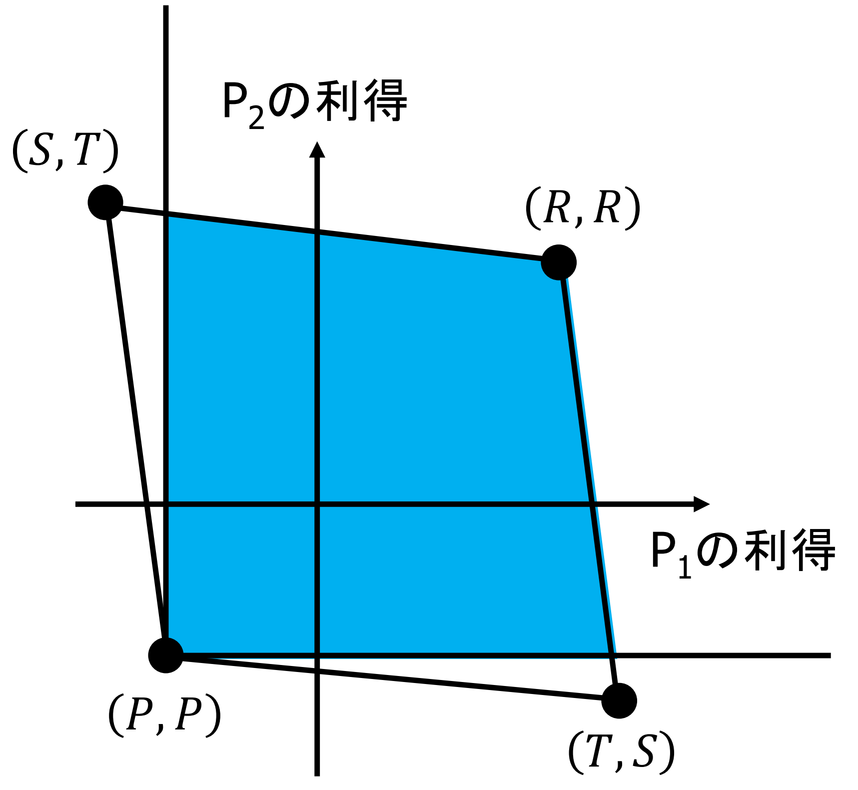
- 以下で定義される利得 v_i を**ミニマックス利得**と呼ぶ:
  - $v_i = \min_{a_{-i}} \max_{a_i} f_i(a_i, a_{-i})$
  - f_i, a はそれぞれ成分ゲームの利得関数，行動の組
- 他のプレイヤーによる処罰を受けているときに得られる最低限の利得
- すべてのプレイヤー i に対して，$f_i(a) \geq v_i$ が成立する行動の組 a を**個人合理的**であるという.
  - 不等号が厳密に成立するときは，**強く個人合理的**

フォーク定理 (1/2)

- 成分ゲームの強く個人合理的な任意の行動の組 a に対して，将来利得の割引因子 δ が，
$$\delta \geq \frac{\max_{b_i} f_i(b_i, a_{-i}) - f_i(a)}{\max_{b_i} f_i(b_i, a_{-i}) - v_i}, \forall i \in N$$
を満たすならば，それぞれの割引総利得和が $f_i(a)$ と等しくなる戦略の組が存在して，かつそれはナッシュ均衡点である．

フォーク定理 (2/2)

- つまり，割引因子 δ が1に十分近いとき，任意の強く個人合理的な利得を達成するナッシュ均衡が存在する．
- ゲームを繰り返すことで，ゲームの任意の状態に到達できる．



個人合理的利得ベクトルの集合

アウトライン

- 囚人のジレンマ (おさらい)
- 繰り返しゲーム
  - 将来利得に対する割引
- 無限回繰り返し囚人のジレンマ
  - 代表的な戦略
  - フォーク定理
- 有限回繰り返し囚人のジレンマ
  - 弱い均衡概念による協力の達成

有限回繰り返し囚人のジレンマ

- 最後のゲームが存在するので，帰納法で均衡点を求めることができる.
  - 最後のゲームでは，裏切っても何の影響もないので，行動 D を選択する.
  - 最後のひとつ前のゲームでは，最後のゲームは裏切るということがわかっているので，行動 D 選択する.
  - ...以下繰り返し.
- 常に行動 D を選択することが支配戦略均衡となる.
  - 繰り返し回数は，有限回では不十分.

限定合理性

- 前のスライドの議論は正しいのか.
  - 1000回繰り返しの囚人のジレンマをプレイする実験を行うと，最初の何回かは協力行動をするプレイヤーが確認される.
- 現実では，繰り返し回数が大きいときに，ゲームの最終回から後向き帰納法で行動を推論しているのか？
- 人間は必ずしも完全に合理的ではない ⇒ 限定合理性 (bounded rationality)

ε-均衡点と有限トリガー戦略

- ε-均衡点
  - 現在の戦略と最適反応との利得の差が ε 以内である戦略の組
  - $\varepsilon = 0$ だと，普通のナッシュ均衡
- 有限トリガー戦略
  - t 回目までは，行動 C を選択する．ただし，相手が逸脱した場合，直ちに行動 D に移行する．
  - $t+1$ 回目以後は，以前の行動に関わらず行動 D を選択する．
- 条件 (繰り返し回数，t, εの値等) 次第では，トリガー戦略の組が ε-均衡点になる．

まとめ

- 繰り返しゲーム
  - 特に繰り返し囚人のジレンマ
  - 無限回繰り返すことで協力が達成可能
- フォーク定理
  - ゲームを無限回繰り返すことで，成分ゲームのいかなる状態にも均衡状態で到達できる.
- 最新の研究テーマ
  - 私的観測下の繰り返し囚人のジレンマ
  - 相手の行動が正しく観測できない.