数理解析研究所講究録
299巻 1977年 114-123

連続時間の非協力 n 人マルコフゲームについて

新大理　田中謙輔

§1. 問題の定式化

ここでは連続時間の非協力 n 人マルコフゲームを $2(n+1)+1$ 個の組 $(S, A_1, A_2, \cdots, A_n, g, r^{(1)}, \cdots r^{(n)}, \alpha)$ で決定する，ただし (1) $S=\{1,2,3,\cdots,m\}$ はシステムの有限状態空間，(2) A_i, $i=1,2,\cdots,n$, は Polish 空間の空でないボレル部分集合でプレーヤー i の行動空間，(3) $g=g_{s\cdot}(a_1,a_2,\cdots\cdots a_n)$ は各 $(s,a_1,a_2,\cdots,a_n)\in S\times\prod_{i=1}^{n}A_i$ に対して，S 上の有限値関数で状態変化の推移率，(4) $r^{(i)}$, $i=1,2,\cdots,n$, は $S\times\prod_{i=1}^{n}A_i$ の上の有界ボレル可測関数でプレーヤー i の利得，(5) α は正の実数で割引因子．このゲームでは各々のプレーヤー i がシステムの状態を連続的に観測し，互に相談又は協力することなく現在の状態 $s\in S$ のみによって行動 $a_i\in A_i$ を選択する．この結果各プレーヤー i は利得 $r_s^{(i)}(a_1,a_2,\cdots\cdots,a_n)$ を得てシステムの状態 s は g に従って新しい状態 s'

に移る。そこで問題はこのようなゲームを無限に続けるとき各プレーヤーiが合計割引き期待利得を最大にすることである。

各プレーヤーiの戦略$\pi^{(i)}=\pi^{(i)}(t)$ はシステムの過去の歴史とは独立で現在の状態のみに依存していると仮定し$\{\mu_t^{(i)}\}$で定める，$\mu_t^{(i)}$ は各々のsと $t\in[0,\infty)$に対して(A_i, $\mathcal{B}(A_i)$)の上の確率測度$\mu_t^{(i)}(\cdot|s)$であり，各々の$M\in\mathcal{B}(A_i)$ とsに対して$[0,\infty)$の上のルベーク可測関数$\mu_\cdot^{(i)}(M|s)$であるとする．このような各プレーヤーiの戦略$\pi^{(i)}=\pi^{(i)}(t)$をマルコフ戦略と呼び，特に各々の$t\in[0,\infty)$に対して$\mu_t^{(i)}$がSからP_{A_i}への写像μで時間に独立であるとき$\pi^{(i)}=\pi^{(i)}(t)$を定常戦略と呼ぶことにする，ただしP_{A_i}は$(A_i,\ \mathcal{B}(A_i))$の確率測度の全体とする。

この話を通して各$(a_1,a_2,\cdots,a_n)\in\prod_{i=1}^{n}A_i$に対応する推移率行列 $Q(a_1,a_2,\cdots,a_n)=\{q_{s,s'}(a_1,a_2,\cdots,a_n);\ s,s'\in S\}$に次のような仮定をおく．

仮定1．各々の$s,s'\in S$に対して，$q_{s,s'}(a_1,a_2,\cdots,a_n)$は$\prod_{i=1}^{n}A_i$の上の連続関数であり，すべての$(a_1,a_2,\cdots,a_n)\in\prod_{i=1}^{n}A_i$に対して$q_{s,s'}(a_1,a_2,\cdots,a_n)\neq 0$ $\forall s,s'\in S$，$q_{s,s'}(a_1,a_2,\cdots,a_n)\geq 0$ $s\neq s'$，$\sum_{s'}q_{s,s'}(a_1,a_2,\cdots,a_n)=0$とする。したがって各プレーヤーのマルコフ戦略の組$(\pi^{(1)},\pi^{(2)},\cdots,\pi^{(n)})$

に対応する推移率は各々の $t \geqq 0$ に対して

$$g_{s,s'}(t, \pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)}) = \int \cdots \int g_{s,s'}(a_1, a_2, \cdots, a_n) \prod_{i=1}^{n} d\mu_t^{(i)}(a_i \mid s)$$

と定義できて,仮定1より $g_{s,s'}(t, \pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)}) \geqq 0 \quad s \neq s'$, $\sum_{s'} g_{s,s'}(t, \pi^{(1)}, \pi^{(2)}, \cdots \pi^{(n)}) = 0$ を満足している。このとき上の条件より次のようなコロモゴロフの微分方程式を満している $Q(t, \pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)})$ に対応する推移確率行列 $F(t, t', \pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)}) = \{f_{s,s'}(t, t', \pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)})\,;\, s, s' \in S\}$ が一意に決まる,

$$\frac{\partial}{\partial t'} F(t, t', \pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)}) = F(t, t', \pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)})\, Q(t', \pi^{(1)}, \cdots, \pi^{(n)}),$$

ただし $F(t, t', \pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)}) = I \quad$ a,e $t' \in [t, \infty)$.

ゲームは常に時刻0から出発するものとし, $F(0, t, \pi^{(1)}, \cdots, \pi^{(n)})$ のかわりに $F(t, \pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)})$ と書くことにする.

各プレーヤーがマルコフ戦略の組 $(\pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)})$ を用いたときのプレーヤー i の状態 s と時刻 t における期待利得率は

$$r_s^{(i)}(t, \pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)}) = \int \cdots \int r_s^{(i)}(a_1, a_2, \cdots, a_n) \prod_{i=1}^{n} d\mu_t^{(i)}(a_i \mid s)$$

で与えられるので,合計割引期待利得を次のように定義する;

$$\psi_s^{(i)}(\pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)}) = \int_0^{\infty} e^{-\alpha t} \sum_{s'} f_{s,s'}(t, \pi^{(1)}, \cdots, \pi^{(n)})\, r_{s'}^{(i)}(t, \pi^{(1)}, \cdots, \pi^{(n)})\, dt.$$

ここでマルコフ戦略の組 $\{\pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)}\}$ とすべての i と s に対して

$$\psi_s^{(i)}(\pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)}) = \sup_{\sigma^{(i)} \in \Pi^{(i)}} \psi_s^{(i)}(\pi^{(1)}, \cdots, \sigma^{(i)}, \cdots, \pi^{(n)}),$$

ただし$\pi^{(i)}$はプレーヤーiのマルコフ戦略の全体とする，
が成立するとき$\{\pi^{(1)},\pi^{(2)},\cdots,\pi^{(n)}\}$を平衡点と呼び，各マルコフ戦略$\pi^{(i)}$はプレーヤー$i$の平衡戦略と呼ぶことにする．

§2．平衡定常戦略の存在について

まず最初に次のような記号と仮定を述べることにする．

$(\prod_{i=1}^{n} P_{A_i})^S$：$S$から$\prod_{i=1}^{n} P_{A_i}$への写像の全体

$\bar{\mu} \equiv (\mu^{(1)},\mu^{(2)},\cdots,\mu^{(n)}) \in (\prod_{i=1}^{n} P_{A_i})^S$,

$\hat{\mu}^{(i)} \equiv (\mu^{(1)},\cdots,\mu^{(i-1)},\mu^{(i+1)},\cdots,\mu^{(n)}) \in (\prod_{j=1, j\neq i}^{n} P_{A_j})^S$,

$(\bar{\mu};\sigma^{(i)}) \equiv (\mu^{(1)},\mu^{(2)},\cdots,\mu^{(i-1)},\sigma^{(i)},\mu^{(i+1)},\cdots\mu^{(n)})$, $\sigma^{(i)} \in (P_{A_i})^S$,

$r_s^{(i)}(\bar{\mu}) \equiv \int\cdots\int r_s^{(i)}(a_1,a_2,\cdots,a_n)\prod_{i=1}^{n} d\mu^{(i)}(a_i|s)$,

$g_{s,s'}(\bar{\mu}) \equiv \int\cdots\int g_{s,s'}(a_1,\cdots,a_n)\prod_{i=1}^{n} d\mu^{(i)}(a_i|s)$.

仮定2．各A_i，$i=1,2,\cdots,n$，はコンパクトな距離空間で，各々のiとsに対して$r_s^{(i)}(a_1,a_2,\cdots,a_n)$は$\prod_{i=1}^{n} A_i$の上で連続とする。

したがって仮定1と仮定2から$\prod_{i=1}^{n} P_{A_i}$は弱位相でコンパクトな距離空間となり，各々のs, s', iに対して$r_s^{(i)}(\bar{\mu})$と$g_{s,s'}(\bar{\mu})$は$\prod_{i=1}^{n} P_{A_i}$の上で有限な連続関数となる。

このとき仮定1と仮定2のもとで次の定理が証明できる．

定理2.1　各プレーヤーの定常戦略の組$\bar{\mu} \equiv (\mu^{(1)},\mu^{(2)},\cdots\cdots,\mu^{(n)})$に対して，$v_s^{(i)}(\bar{\mu}) = \psi_s^{(i)}(\bar{\mu})$は

$$\alpha v_s^{(i)}(\bar{\mu}) = r_s^{(i)}(\bar{\mu}) + \sum_{s'} g_{s,s'}(\bar{\mu}) v_{s'}^{(i)}(\bar{\mu}) \qquad (2.1)$$

の唯一つの解である．

<u>証明</u>　まず（2.1）の解が$\psi_s^{(i)}(\bar{\mu})$であることを示す．
（2.1）の両辺に$e^{-\alpha t}f_{\ell,s}(t,\bar{\mu})$を掛け，$s$について加えてコロモゴロフの微分方程式を用いれば，各ℓに対して次式を得る．

$$\alpha e^{-\alpha t}\sum_s f_{\ell,s}(t,\bar{\mu})v_s^{(i)}(\bar{\mu}) = e^{-\alpha t}\sum_s f_{\ell,s}(t,\bar{\mu})r_s^{(i)}(\bar{\mu}) + \\ + e^{-\alpha t}\sum_s \frac{\partial}{\partial t}f_{\ell,s}(t,\bar{\mu})v_s^{(i)}(\bar{\mu}). \qquad (2.2)$$

このとき（2.2）の両辺を$t\in[0,\infty)$について積分して

$$v_\ell^{(i)}(\bar{\mu}) = \int_0^\infty e^{-\alpha t}\sum_s f_{\ell,s}(t,\bar{\mu})r_s^{(i)}(\bar{\mu})\,dt$$

が得られる。

次に$\psi_s^{(i)}(\bar{\mu})$が（2.1）の解であることを示す．
各々のsと$t\geqq 0$に対して次式が成立する．

$$\psi_s^{(i)}(\bar{\mu}) = \int_0^t e^{-\alpha u}\sum_{s'} f_{s,s'}(u,\bar{\mu})r_{s'}^{(i)}(\bar{\mu})\,du + \\ + e^{-\alpha t}\sum_{s'} f_{s,s'}(t,\bar{\mu})\psi_{s'}^{(i)}(\bar{\mu}) \qquad (2.3)$$

（2.3）の両辺をtで微分し，$t\to 0$として

$$\lim_{t\to 0^+} f_{s,s'}(t,\bar{\mu}) = \delta_{s,s'}$$

と

$$\lim_{t\to 0^+}\frac{\partial}{\partial t}f_{s,s'}(t,\bar{\mu}) = g_{s,s'}(\bar{\mu})$$

を用いることによって

$$r_s^{(i)}(\bar{\mu}) - \alpha\psi_s^{(i)}(\bar{\mu}) + \sum_{s'} g_{s,s'}(\bar{\mu})\psi_{s'}^{(i)}(\bar{\mu}) = 0$$

が得られ証明は終る．

ここで $C(S)$ を S の上の有限値関数の全体とし，$C(S)$ に含まれる v に対して $\|v\| = \max_s |v(s)|$ とおくと，$C(S)$ は完備な距離空間となる。次に定常戦略の組 $\bar{\mu} = (\mu^{(1)}, \mu^{(2)}, \cdots\cdots, \mu^{(n)})$ に対して $Q(\bar{\mu})$ に対応する新しい確率推移行列 $P(\bar{\mu})$ を次のように定義する；

$$P(\bar{\mu}) \equiv I + \frac{1}{M} Q(\bar{\mu}),$$

ただし M は $|q_{s,s'}(\bar{\mu})|$ の最大値とする．

即ち (s, s') 要素は $p_{s,s'}(\bar{\mu}) = \delta_{s,s'} + \frac{1}{M} q_{s,s'}(\bar{\mu})$ によって与える．更に各 i に対して写像 $T^{(i)}: C(S) \to C(S)$ を次のように定義する； $v \in C(S)$ と $\bar{\mu} \in (\prod_{i=1}^{n} P_{A_i})^S$ に対して

$$(T^{(i)}v)(s) \equiv \sup_{\alpha^{(i)}(s) \in P_{A_i}} \left\{ \frac{r_s^{(i)}(\bar{\mu}; \alpha^{(i)})}{\alpha + M} + \right.$$

$$\left. + \frac{M}{\alpha + M} \sum_{s'} p_{s,s'}(\bar{\mu}; \alpha^{(i)}) v(s') \right\},$$

ただし α は割引因子である．

このとき右辺は $\prod_{i=1}^{n} P_{A_i}$ の上で連続であるので sup を max と書き換えることができる．更に $T^{(i)}$ が縮少写像であることを利用して次の補助定理を証明できる．

補助定理 I．各々の i と s に対して

$$\alpha v_s^{(i)}(\hat{\mu}^{(i)}) = \max_{\alpha^{(i)}(s) \in P_{A_i}} \left\{ r_s^{(i)}(\bar{\mu}; \alpha^{(i)}) + \right.$$

$$\left. + \sum_{s'} q_{s,s'}(\bar{\mu}; \alpha^{(i)}) v_{s'}^{(i)}(\hat{\mu}^{(i)}) \right\} \qquad (2.4)$$

を満足する $v_s^{(i)}(\hat{\mu}^{(i)})$ が存在する．

証明　$0 < M/(\alpha + M) < 1$ であるから，各 $T^{(i)}$ は縮少写像

となり唯一つの不動点が存在する．よってこの不動点を $v_s^{(i)}(\hat{\mu}^{(i)})$ とすれば，次式が成立する．

$$v_s^{(i)}(\hat{\mu}^{(i)}) = \max_{a^{(i)}(s)\in P_{A_i}} \left\{ \frac{r_s^{(i)}(\bar{\mu}; a^{(i)})}{\alpha + M} + \right.$$
$$\left. + \frac{M}{\alpha+M}\sum_{s'} p_{s,s'}(\bar{\mu}; a^{(i)}) v_{s'}^{(i)}(\hat{\mu}^{(i)}) \right\} \quad (2.5)$$

(2.5) で $p_{s,s'}(\bar{\mu}; a^{(i)})$ を $q_{s,s'}(\bar{\mu}; a^{(i)})$ で置き換えることによって (2.4) が成立し，証明は終る．

次に補助定理 1 における $v_\cdot^{(i)}(\hat{\mu}^{(i)}) \in C(S)$ に対して

$$K_s^{(i)}(\bar{\mu}; a^{(i)}) \equiv r_s^{(i)}(\bar{\mu}; a^{(i)}) + \sum_{s'} q_{s,s'}(\bar{\mu}; a^{(i)}) v_{s'}^{(i)}(\hat{\mu}^{(i)})$$

を定義し，更に $G^{(i)}(\hat{\mu}^{(i)}) \equiv \{\lambda^{(i)};\ ^\forall s \in S,\ K_s^{(i)}(\bar{\mu}; \lambda^{(i)}) = \max_{a^{(i)}(s)\in P_{A_i}} K_s^{(i)}(\bar{\mu}; a^{(i)})\}$ を定義する．このとき $G^{(i)}(\hat{\mu}^{(i)})$ は空でなく，閉で凸な部分集合となる．ここで写像 $G: (\prod_{i=1}^n P_{A_i})^S \longrightarrow (\prod_{i=1}^n P_{A_i})^S$ を次のように定義する；各々の $\bar{\mu}$ に対して

$$G(\bar{\mu}) \equiv \{\bar{\lambda} = (\lambda^{(1)}, \lambda^{(2)}, \cdots, \lambda^{(n)});\ \lambda^{(i)} \in G^{(i)}(\hat{\mu}^{(i)}),\ i=1,2,\cdots,n\}.$$

このとき写像 G は点から閉凸集合への写像となる．したがって拡張された不動点定理を用いるために次の補助定理が重要である．

補助定理 2．写像 G は上半連続である．

証明　すべての m に対して，$\bar{\lambda}_m \in G(\bar{\mu}_m)$ で，$m \to \infty$ のとき $\bar{\lambda}_m = (\lambda_m^{(1)}, \lambda_m^{(2)}, \cdots, \lambda_m^{(n)}) \Rightarrow \lambda_0 = (\lambda_0^{(1)}, \lambda_0^{(2)}, \cdots, \lambda_0^{(n)})$，$\bar{\mu}_m = (\mu_m^{(1)}, \mu_m^{(2)}, \cdots, \mu_m^{(n)}) \Rightarrow \bar{\mu}_0 = (\mu_0^{(1)}, \mu_0^{(2)}, \cdots, \mu_0^{(n)})$ ならば $\bar{\lambda}_0 \in G(\bar{\mu}_0)$ が成立することを示せば充分である，た

だし $\Rightarrow$ は $\prod_{i=1}^{n} P_{A_i}$ における弱収束を意味している.

補助定理1を用いると, $\overline{\lambda_m} \in G(\overline{\mu_m})$ と各々の i に対して

$$\alpha V_s^{(i)}(\hat{\mu}_m^{(i)}) = \max_{\alpha^{(i)}(s) \in P_{A_i}} K_s^{(i)}(\overline{\mu_m}; \alpha^{(i)})$$
$$= K_s^{(i)}(\overline{\mu_m}; \lambda_m^{(i)}) \qquad (2.6)$$

で任意の $\alpha^{(i)} \in (P_{A_i})^s$ に対して

$$\alpha V_s^{(i)}(\hat{\mu}_m^{(i)}) \geqq K_s^{(i)}(\overline{\mu_m}; \alpha^{(i)}) \qquad (2.7)$$

が成立する.

更に $m \to \infty$ のとき各 s について $V_s^{(i)}(\hat{\mu}_m^{(i)}) \longrightarrow V_s^{(i)}(\hat{\mu}_0^{(i)})$ が成立する.

よって (2.6) と (2.7) で $m \to \infty$ とするとき

$$\alpha V_s^{(i)}(\hat{\mu}_0^{(i)}) = K_s^{(i)}(\overline{\mu_0}; \lambda_0^{(i)})$$

と任意の $\alpha^{(i)} \in (P_{A_i})^s$ に対して

$$\alpha V_s^{(i)}(\hat{\mu}_0^{(i)}) \geqq K_s^{(i)}(\overline{\mu_0}; \alpha^{(i)})$$

が成立し証明は終る.

よって拡張された不動点定理より, 各々の i, s に対して $\overline{\mu_*} \in G(\overline{\mu_*})$ となる $\overline{\mu_*} = (\mu_*^{(1)}, \mu_*^{(2)}, \cdots, \mu_*^{(n)}) \in (\prod_{i=1}^{n} P_{A_i})^s$ が存在する, 即ち

$$\alpha V_s^{(i)}(\hat{\mu}_*^{(i)}) = \max_{\alpha^{(i)}(s) \in P_{A_i}} \{ r_s^{(i)}(\overline{\mu_*}; \alpha^{(i)}) +$$
$$+ \sum_{s'} g_{s,s'}(\overline{\mu_*}; \alpha^{(i)}) V_{s'}^{(i)}(\hat{\mu}_*^{(i)}) \} \qquad (2.8)$$
$$= r_s^{(i)}(\overline{\mu_*}) + \sum_{s'} g_{s,s'}(\overline{\mu_*}) V_{s'}^{(i)}(\hat{\mu}_*^{(i)}),$$

で, 任意の $\alpha^{(i)} \in (P_{A_i})^s$ に対して

$$\alpha v_s^{(i)}(\hat{\mu}_*^{(i)}) \geqq r_s^{(i)}(\bar{\mu}_*; \alpha^{(i)}) + \\ + \sum_{s'} g_{s,s'}(\bar{\mu}_*; \alpha^{(i)}) v_{s'}^{(i)}(\hat{\mu}_*^{(i)}) \quad (2.9)$$

が成立する。

このとき補助定理1と2を用いて、次の定理を証明することができる。

定理2.2　このゲームには平衡点が存在して、各プレーヤーは平衡定常戦略をもっている。

証明　定理2.1を用いて，(2.8) より各々の i，s に対して

$$v_s^{(i)}(\hat{\mu}_*^{(i)}) = \psi_s^{(i)}(\bar{\mu}_*) \quad (2.10)$$

が成立する．

定常戦略の組 $\hat{\mu}_*^{(i)}$ とプレーヤー i の任意のマルコフ戦略 $\alpha^{(i)}$ に対して，任意の時刻 $t \geqq 0$ で次式が得られる．

$$\alpha v_s^{(i)}(\hat{\mu}_*^{(i)}) \geqq r_s^{(i)}(t, \bar{\mu}_*; \alpha^{(i)}) + \\ + \sum_{s'} g_{s,s'}(t, \bar{\mu}_*; \alpha^{(i)}) v_{s'}^{(i)}(\hat{\mu}_*^{(i)}). \quad (2.11)$$

(2.11) の両辺に $e^{-\alpha t} f_{l,s}(t, \bar{\mu}_*; \alpha^{(i)})$ を掛け，s について加えてコロモゴロフの微分方程式を用いると

$$\alpha e^{-\alpha t} \sum_s f_{l,s}(t, \bar{\mu}_*; \alpha^{(i)}) v_s^{(i)}(\hat{\mu}_*^{(i)}) \geqq \alpha e^{-\alpha t} \sum_s f_{l,s}(t, \bar{\mu}_*; \alpha^{(i)}) \cdot \\ \cdot r_s^{(i)}(t, \bar{\mu}_*; \alpha^{(i)}) + e^{-\alpha t} \sum_s \frac{\partial}{\partial t} f_{l,s}(t, \bar{\mu}_*; \alpha^{(i)}) v_s^{(i)}(\hat{\mu}_*^{(i)}) \quad (2.12)$$

が得られる．

更に (2.12) の両辺を $t\in[0,\infty)$ で積分して，すべての s に対して

$$\mathcal{V}_s^{(i)}(\hat{\mu}_*^{(i)}) \geq \mathcal{V}_s^{(i)}(\bar{\mu}_*;\alpha^{(i)}) \tag{2.13}$$

を得る。

よって (2.10) と (2.13) より定常戦略の組 $\bar{\mu}_*=(\mu_*^{(1)},\cdots,\mu_*^{(n)})$ は平衡点で，この各要素 $\mu_*^{(i)}$ は各プレーヤー i の平衡定常戦略になっている。

参 考 文 献

[1] K.Fan, Fixed-point and minimax theorems in locally convex topological linear spaces. Proc. Nat. Acad. Sci., U.S.A., 88 (1952), 121-126.

[2] P.Kakumanu, Continuous time Markov decision models with applications to optimization problems, Tech. Rep. 63, Dept. O.R., Cornell University.

[3] P.Kakumanu, Continuously discounted Markov decision model with countable state and action space, Ann. Math. Statist., 42 (1971), 919-926.

[4] T.Parthasarathy, Discounted, positive, and noncooperative stochastic games, Intern. J. Game Theory, 2 (1973), 25-37.

[5] K.Tanaka and K.Wakuta, On continuous time Markov games with countable state space, to appear.

[6] M.Takahashi, Equilibrium points of stochastic noncooperative n-person games, J. Sci. Hiroshima Univ. Ser. A, 28 (1964), 95-99.