

Theorem:

let $n \geq 4$ given a n-2-g iterative social agreement game. placing no constraints on the initial actions of the players assume the players use HCR, the following holds:

1. for every $\epsilon > 0$ there exists $M = M(\epsilon)$, which is bounded such that if the system runs for $i \geq M(\epsilon)$ iterations that we reach a convention with probability $p > 1 - \epsilon$
2. once a convention is reached it will never be left
3. if a social convention is reached it will not lead to payoff lower than the maximin value.
4. if there exists a social convention which is rational with respect to maximin then such rational convention will be reached

Proof of part 4 (for coordination game):

then next process will happen with probability $p = \frac{1}{f(n)}$ and lead to social convention in

$g(n)$ iterations. The process: 2 agents will meet until all others forget their history. Denote the 2 agents i and j. Now for every agent k (other than i and j), i will meet with k and afterwards again with j. Eventually every agent will accept the strategy of agent i. If after $g(n)$ iterations the social convention was not reached, the aforementioned process will happen again. Thus if the system will run for $M = k \cdot f(n) \cdot g(n)$ the social convention will be reached with probability e^{-k} ($k > -\log(\epsilon)$).

Definition

Let g be a social agreement game. Consider the corresponding n-2-g iterative game and the n(n-1) games that might be played in iteration t. Let $X_n(t)$ be a random variable that contains the number of games that might be played in iteration t and get a payoff lower than in a rational social convention.

Let T(n) be a function that associates with n a number of iterations. Given selection rule R and some distribution on the initial actions we say that R guarantees a rational social convention after T(n) iterations, if $E(X_n(T(n)))$ converges to 0.

Theorem:

Let g be as before assume that any action is placed with some constant probability in the initial phase. If R guarantees social convention in the corresponding n-2-g game in T(n) iterations than $T(n) = \Omega(n \log n)$

Proof

Let $Y_n(i)$ denote a random variable that contains the number of players that weren't chosen yet in any iteration of the n-2-g game. Than $E(X_n(i)) \geq k \cdot E(Y_n(i))$ (k some constant), in particular

$E(X_n(T(n))) \geq k \cdot E(Y_n(T(n)))$. It is suffice to show that if $E(Y_n(T(n)))$ converges to 0 as a function of n, than $T(n) = \Omega(n \log n)$. The probability that some agent i won't be chosen for $n \cdot f(n)$

iterations is $(1 - \frac{2}{n})^{n \cdot f(n)}$, that converges to $e^{-2 \cdot f(n)}$. Thus if $e^{-2 \cdot f(n)} > \frac{1}{n}$ than

$E(Y_n(T(n))) > 1$. Since the inequality $e^{-2 \cdot f(n)} \leq \frac{1}{n}$ must hold we get that $f(n) \geq 0.3 \log n$.

Partial Control

חלק מהסוכנים בשליטתנו וחלקם אינם:
במסגרת של התפתחות קונבנציות יש סוכנים רציונלים וסוכנים צייתנים הרעיון הוא להפוך מספר סוכנים ל"שוטרים".

מודל מורה תלמיד:
משחקים משחק חוזר סוכן אחד בשליטתנו (מורה) וסוכן אחר לא (תלמיד). התלמיד ישתמש באלגוריתם למידה.

מודלי התלמיד הם בדר"כ מודלים של למידה על ידי חיזוקים.

המקרה הפשוט מצריך שתי פעולות הן למורה והן לתלמיד. ניתן לתיאור על ידי משחק הבא כאשר שחקן השורות הוא התלמיד:

	A	B
1	a	b
2	c	d

רוצים ללמד לשחק אותו את האסטרטגיה 1.
אם $a > c, b > d$ אזי טריוויאלי, כנ"ל אם $a < c, b < d$.
משפחה אחרת של מקרים: אם $d > b, a > c$ לדוגמא:

	A	B
1	5	1
2	2	7

אזי על ידי משחק של A השחקן הלומד ילמד לשחק 1.

במקרה $c > a > d > b$

	A	B
1	(a,a)	(b,-b)
2	(c,-c)	(d,d)

למרות שהמקרה נראה טריוויאלי הוא לא בהכרח כזה. בגלל העובדה שלפנינו משחק חוזר שבו לתלמיד יש אלגוריתם למידה.

אסטרטגיית לימוד אופטימלית:

אסטרטגיית לימוד מקסימלית היא פונקציה שמחזירה פעולה בכל איטרציה (בהתבסס בפעולות הקודמות שהתבצעו).

$M(a) =$ הערך שהמורה נותן לתלמיד עבור פעולה a.

כלומר ברקע יש אסטרטגיה או אלגוריתם למידה לתלמיד ואנו מסתכלים באלגוריתם II

של המורה. Π מסוים משרה התפלגות $Pr_{\Pi,k}$ על פעולות האפשריות בזמן k .
 לכן ערך Π שזכה נתון על ידי: $Val(\Pi) = \sum_{k=0}^{\infty} \gamma^k \cdot E_k(U)$ כאשר
 $E_k(U) = \sum_{a \in A_t} Pr_{\Pi,k}(a) \cdot U(a)$

איפיון אסטרטגיות אופטימליות לפי תהליך החלטה מרקובי (markov decision process):

S – State

A – Possible actions

$P: S \times S \times A \rightarrow [0,1]$

$\gamma: S \rightarrow R$

בהינתן מצב $s \in S$ ואיזושהי מדיניות $\Pi: S \rightarrow A$ מוגדרת פונקציית ההתפלגות $P_{s,\Pi,k}$ על S .
 כאשר $P_{s,\Pi,k}(s')$ היא ההסתברות להגיע בדיוק תוך k שלבים מ s ל s' כאשר המדיניות היא Π .