11

Θεωρία παιγνίων 8-6-18

Επαναλαμβανόμενα Παίγνια

Έστω ένα στατικό παίγνιο πλήρους πληροφόρησης, G. 1 επαναλαμβανόμενο παίγνιο G(T) αποτελείται από έναν αριθμό επαναλήψεων, T, του G το οποίο στο συγκεκριμένο πλαίσιο λέγεται παίγνιο σταδίου (stage game) με δεδομένο ότι αποτελεί το παίγνιο το οποίο επαναλαμβάνεται κάθε φορά (ή στάδιο)

$T \in N$, $1 < T < +\infty$:

Με δεδομένο ότι το G επαναλαμβάνεται $1 < T < +\infty$ ($T \in N$) φορές, τότε η κατάλληλη ιδέα επίλυσης του G(T) είναι η SPNE, δηλ. η λύση η οποία χρησιμοποιείται στα δυναμικά παίγνια.
Ισχύει το ακόλουθο θεώρημα: Έστω G(T) ένα επαναλαμβανόμενο παίγνιο, όπου το παίγνιο σταδίου G έχει μοναδική ισορροπία Nash σε καθαρές στρατηγικές του S^*. Αν $1 < T < +\infty$ ($T \in N$), τότε το G(T) έχει μοναδική SPNE την $S_t^* = S^*$, $\forall\, t = 1, 2, \ldots, T$

Άρα στο Δίλημμα των Φυλακ. (στατικό παίγνιο πλήρους πληροφόρησης), το οποίο έχει μοναδική ισορ. Nash σε καθαρές στρατηγικές $S^* = (0,0)$, για $T \in N$, $1 < T < +\infty$ επαναλήψεις η SPNE του G(T) (επαναλ. παιγνίου): Δίλ. των Φυλακισμένων είναι η $S_t^*(0,0)$, $\forall t$. Άρα σε κάθε επανάληψη οι 2 φυλακισμένοι θα παίζουν τη μοναδική ισορροπία Nash σε καθαρές στρατηγικές.
Ο λόγος που συμβαίνει αυτό είναι η έλλειψη αξιοπιστίας: κανένας παίκτης δεν εμπιστεύεται τον άλλον, ώστε να παίζει την Μ.Ο. μέρος της στρατηγικής του. Τονίζουμε ότι, σε πειράματα τα οποία έχουν γίνει το θεώρημα αυτό μπορεί να παραβιαστεί, δηλ. οι 1 κ' 2 (παίκτες) να μην παίζουν τη μοναδική ισορροπία Nash (0,0).
Όσον αφορά τη θεωρία, το αποτέλεσμα του παραπάνω θεωρήματος, μπορεί να αρθεί σε 2 περιπτώσεις: 1 το παίγνιο παίζεται επ' αόριστο, δηλ. έχει τέλος, αλλά είναι άγνωστο ποιά φορά είναι η τελευταία. 2 Το παίγνιο παίζεται επ' άπειρον.

Στις περιπτώσεις αυτές, με δεδομένο ότι το τέλος είναι άγνωστο ή δεν υπάρχει, δεν μπορούμε να χρησιμοποιήσουμε την SPNE πηγαίνοντας από το τέλος προς τα πίσω.

Δ.Φ.

1 \ 2	MO	O
MO	1,1	-1,2
O	2,-1	0,0

$S_1 = S_2 = \{MO, O\}$ (χώροι στρατηγικής)

Έστω ότι το δίλημμα του φυλακισμένου επαναλαμβάνεται επ' άπειρο και συγκεκριμένα σε κάθε στάδιο η πιθανότητα επανάληψης είναι $p>0$. Για ποιο p οι 1 και 2 έχουν κίνητρο να επιλέξουν από κοινού την Pareto άριστη στρατηγική (MO, MO). Αν οι 1 κ 2 επιλέξουν (MO, MO), τότε η απόδοση του καθενός είναι:

$$1 + p + p^2 + \dots = \frac{1}{1-p}$$

(Θυμίζουμε ότι: $a + aw + aw^2 + \dots = \frac{a}{1-w}$, $|w|<1$) φθίνουσα γεωμετρική σειρά

Αν οι 1 κ 2 επιλέξουν μια στρατηγική "μια σου και μια μου" ("tit for tat") και παίξουν πάντα O σε περίπτωση που ο άλλος παίκτης τους "καρφώσει", τότε η απόδοση του παίκτη που θα το κάνει αυτό θα είναι 2. Άρα οι παίκτες θα συνεργαστούν παίζοντας MO, MO αν και μόνο αν:

$$\frac{1}{1-p} > 2 \Leftrightarrow 1 > 2-2p \Leftrightarrow 2p > 1$$

$$\Leftrightarrow \boxed{p > \frac{1}{2}}$$

, δηλ. ανν η πιθανότητα επανάληψης είναι (γνησίως) μεγαλύτερη από $\frac{1}{2}$.

11

		2	
		MO	O
1	MO	a,a	b,c
	O	c,b	d,d

$c > a > d > b$

Γενική μορφή (Δ.Φ.)

Απόδοση από (MO, MO): $a + a\cdot\rho + a\cdot\rho^2 + \dots = \frac{a}{1-\rho}$

Απόδοση από στρατηγική "tit for tat":

$$c + d\rho + d\rho^2 + \dots = c + (d + d\rho + \dots)\rho =$$
$$= c + \rho\cdot\frac{d}{1-\rho}$$

Άρα η (MO, MO) συμφέρει για ρ τ.ω. $\frac{a}{1-\rho} > c + \frac{d\rho}{1-\rho} \Leftrightarrow$

$\Leftrightarrow a > c - c\rho + d\rho \Leftrightarrow (c-d)\rho > c-a \Leftrightarrow \boxed{\rho > \frac{c-a}{c-d}}$

Προηγουμένως: $a=1$, $c=2$, $d=0 \Leftrightarrow$

$\Leftrightarrow \rho > \frac{c-a}{c-d} = \frac{2-1}{2-0} = \frac{1}{2}$

Έστω G ένα στατικό παίγνιο πλήρους πληροφόρησης το οποίο έχει μοναδική ισορροπία Nash σε καθαρές στρατηγικές και $G(+\infty, \delta)$ το επαναλαμβανόμενο παίγνιο στο οποίο ο αριθμός των επαναλήψεων είναι άπειρος και $\delta < 1$ είναι ένας συντελεστής προεξόφλησης των αποδόσεων

SOS

Θεώρημα (Λαϊκό Θεώρημα / Folk theorem): Έστω $G(+\infty, \delta)$, $\delta < 1$, ένα επαναλαμβανόμενο παίγνιο, όπου το παίγνιο σταδίου G έχει ισορροπία Nash με απόδοση των παικτών (n - το πλήθος) $(v_1^*, v_2^*, \dots, v_n^*)$. Αν $(v_1, v_2, \dots, v_n)$ είναι ένας άλλος συνδυασμός αποδόσεων τ.ω. $v_i > v_i^*$ $\forall i$ και δ επαρκώς κοντά στο 1, τότε υπάρχει ισορροπία Nash του $G(+\infty, \delta)$ στην οποία η μέση απόδοση κάθε παίκτη $1, 2, \dots, n$ είναι αυθαίρετα κοντά στο $(v_1, v_2, \dots, v_n)$ διάνυσμα.

Θα εξηγήσουμε το παραπάνω θεώρημα με αναφορά στο $G(+\infty, \delta)$, G = δίλημμα του φυλακισμένου!

2

Στο G (Δ.Φ.) είχαμε $< v_1^*, v_2^* > = < -6, -6 >$.
Το δημώδες Θεώρημα λέει ότι υπάρχει δ ($\simeq 1$) τ.ω. στο $G(+\infty, \delta)$ οι παίκτες να μπορούν να πετύχουν μια μέση απόδοση αυθαίρετα κοντά στη $< v_1, v_2 >$, όπου $v_1 > -6$, $v_2 > -6$. Άρα οι 1 κ' 2 στο $G(+\infty, \delta)$ για δ επαρκώς κοντά στο 1 μπορούν να καταλήξουν σε ~~αφ~~ μια απόδοση από κείνη της $(0,0)$ (μοναδική Ν.Ε σε καθαρές στρατηγικές)

$\boxed{\text{SOS}}$ (κυρίως το γράφημα)

Αριθμητικό παράδειγμα (δίλημμα του φυλ.)

	ΜΟ	Ο
ΜΟ	4,4	−1,5
Ο	5,−1	1,1

Ορίζουμε ως μέση απόδοση τη σειρά
$$(1-\delta) \sum_{t=1}^{\infty} \delta^{t-1} v_t^i, \quad i \in \{1,2\}$$ ("κανονικοποιημένη" Π.Α. στην παρούσα)

Εν προκειμένω $< v_1^*, v_2^* > = < 1, 1 >$.
Το δημώδες Θεώρημα λέει ότι ~~οι 1 και 2 μπορούν να παίξουν στρατηγική με μέση απόδοση $< v_1, v_2 >$, $v_1 > v_1^*$, $v_2 > v_2^*$~~
υπάρχει δ επαρκώς κοντά στο 1 τ.ω. όταν το συγκεκριμένο παίγνιο επαναληφθεί ∞ φορές οι ~~και~~ 1 κ' 2 να μπορούν να πετύχουν μέση απόδοση αυθαίρετα κοντά στην $< v_1, v_2 >$, $v_1 > v_1^*$, $v_2 > v_2^*$

Συγκεκριμένα έστω ότι ο παίκτης 1 παίζει με τέτοιο τρόπο ώστε σε κάθε στάδιο του παιγνίου να εναλλάσσει την απόδοσή του από το 5 στο −1, ενώ ο 2 παίζει έτσι ώστε η απόδοσή του να εναλλάσσεται από το −1 στο 5. Τότε:
Μέση απόδοση παίκτη 1 από τη συγκεκριμένη στρατηγική $(5,-1)$ =
$$= (1-\delta)\left[5 + (-1)\delta + 5\delta^2 + (-1)\delta^3 + \dots\right] =$$
$$= (1-\delta)\left[5(1+\delta^2+\dots) + (-1)(\delta+\delta^3+\dots)\right] =$$
$$= (1-\delta)\left(\frac{5}{1-\delta^2} - \frac{\delta}{1-\delta^2}\right) = (1-\delta)\frac{5-\delta}{1-\delta^2} = \frac{5-\delta}{1+\delta}$$

(2)

Μέση απόδοση παίκτη 2 από τη συγκεκριμένη στρατηγική $(-1,5)=$

$$=(1-\delta)\left(-1+5\delta+(-1)\delta^2+5\delta^3+\dots\right)=$$

$$=(1-\delta)\left[(-1)(1+\delta^2+\dots)+5(\delta+\delta^3+\dots)\right]=$$

$$=(1-\delta)\left(-\frac{1}{1-\delta^2}+\frac{5\delta}{1-\delta^2}\right)=(1-\delta)\frac{5\delta-1}{1-\delta^2}=\frac{5\delta-1}{1+\delta}$$

Καθώς το $\delta \to 1$ ("αυθαίρετα κοντά" στο 1) τότε

$$\lim_{\delta\to 1} \text{Μέση απόδοση } 1=\lim_{\delta\to 1}\frac{5-\delta}{1+\delta}=\frac{5-1}{1+1}=\frac{4}{2}=2=\lim_{\delta\to 1} \text{Μέση απόδοση } 2=$$

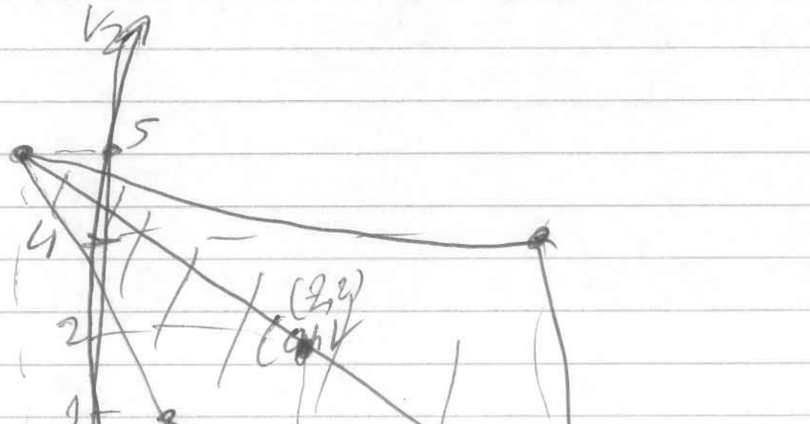
$$=\lim_{\delta\to 1}\frac{5\delta-1}{1+\delta}=\frac{5\cdot 1-1}{1+1}=\frac{4}{2}=2$$

Άρα επιλέγοντας δ αυθαίρετα κοντά στο 1, οι 1 και 2 παίρνουν μια απόδοση αυθαίρετα κοντά στο $\langle v_1,v_2\rangle=\langle 2,2\rangle$ με $2>1$ $\forall$ κάθε παίκτη.

Σημείωση: Το διάνυσμα $\langle v_1,v_2\rangle=\langle 2,2\rangle$ είναι στοιχείο του συνόλου $conv\{(4,4),(5,-1),(-1,5),(1,1)\}$ και συγκεκριμένα

είναι ο ακόλουθος κυρτός συνδυασμός:

$$0{,}5\cdot\langle 5,-1\rangle+0{,}5\cdot\langle -1,5\rangle=\langle 2,2\rangle \in conv\{(4,4),(-1,5),(5,-1),(1,1)\}$$



(Θυμίζουμε: κυρτό κέλυφος συνόλου $C \subseteq \mathbb{R}^n$ = σύνολο όλων των κυρτών συνδυασμών σημείων του C:

$$\operatorname{conv} C = \{ \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \dots + \theta_k x_k : x_i \in C,\ \theta_i \geq 0,\ i = 1, 2, \dots, k,\ \theta_1 + \theta_2 + \dots + \theta_k = 1 \}$$