

PONAVLJAJOČE IGRE IN ZAPORNIKOVA DILEMA

MIHA NOVOSELEC

Fakulteta za matematiko in fiziko
Univerza v Ljubljani

V članku so obravnavane ponavljajoče igre, ki so ločene na končne in neskončne igre. Predstavljen je pojem strategije kot stroja. Pri iskanju Nashevih ravnovesij v neskončnih ponavljajočih igrah je izpostavljen pomemben razred ljudskih izrekov. Predstavljen je primer zapornikove dileme in vpliv ponavljanja igre na oblikovanje Nashevega ravnovesja.

REPEATED GAMES AND PRISONER'S DILEMMA

The article explores repeated games, distinguishing between finite and infinite games. It introduces the concept of strategy as a machine. In examining Nash equilibria within infinite repeated games, special attention is given to a significant class of folk theorems. The prisoner's dilemma is used as an example to illustrate how repetition influences the formation of Nash equilibrium.

1. Uvod

Ponavljajoče igre so razred ekstenzivnih iger, kjer igralci ponavljajo eno in isto strateško igro. Glavna ideja je, da si igralec ne želi vedno igrati strategije, ki bi mu prinesla kratkoročno prednost, saj bi mu lahko to pri končnem izkupičku škodovalo. Igralec se zaveda, da trenutna odločitev vpliva na prihodnje poteze ostalih igralcev ter lahko izzove sodelovanje, maščevanje, kaznovanje in grožnje.

Glede na trajanje bomo igre ločili na končne in neskončne ponavljajoče igre. Pri slednjih se bomo v veliki meri oprli na ljudske izreke, ki nam bodo pomagali pri iskanju Nashevega ravnovesja. Tako ime so dobili, ker jih ne moremo pripisati enemu samemu znanemu avtorju (v angleški literaturi so znani kot 'folk theorems'). Izbira, kdaj obravnavati neskončne ponovljive igre in kdaj končne ponovljive, temelji na tem, ali igralci vedo, kdaj se bo igra zaključila. V kolikor je igra sicer končna, vendar pa igralci ne vedo, kdaj se konča, še vedno igro obravnavamo kot neskončno ponavljajočo igro. Informacija o trajanju igre namreč vpliva na njihovo odločanje.

V članku se bomo v veliki meri zanašali na knjigo 'M.J. Osborne, A. Rubinstein: A Course in Game Theory'[1], iz katere bomo vzeli ključne izreke in dokaze. Knjiga [2] nam bo v pomoč pri bolj konkretnih zgledih, delo [3] pa za bolj praktičen aspekt teme.

Za začetek definirajmo strateške igre s funkcijami preferenc in pojem Nashevega ravnovesja.

Definicija 1 (Strateška igra s funkcijami preferenc). Strateška igra s funkcijami preferenc je trojica $(N, (A_i)_{i \in N}, (u_i)_{i \in N})$, pri čemer:

- N je neprazna množica igralcev
- Za vsakega igralca $i \in N$ je A_i neprazna množica akcij, med katerimi lahko izbira igralec i . Množico profilov akcij označimo z $\mathcal{A} = \prod_{i \in N} A_i$.
- Za vsakega igralca $i \in N$ je $u_i : \mathcal{A} \rightarrow \mathbb{R}$ funkcija preferenc na $\mathcal{A}$.

Definicija 2 (Nashevo ravnovesje). Naj bo $\Gamma = (N, (A_i)_{i \in N}, (u_i)_{i \in N})$ strateška igra s funkcijami preferenc. Naj bo $a^* = (a_i^*)_{i \in N}$ eden od profilov akcij. Profil akcij a^* je Nashevo ravnovesje, če za vse $i \in N$ in vse $b_i \in A_i$ velja $u_i(a^*) \geq u_i(a^*|b_i)$.

Prikažimo zapornikovo dilemo v matriki:

Zgled 3 (Zapornikova dilema). Zapornikova dilema je strateška igra za dva igralca, kjer se dva obtoženca odločata, ali zločin priznati ali ne. Tako imata na razpolago dve akciji, torej $A_1 = A_2 = \{\text{Prizna}, \text{Ne prizna}\} = \{P, N\}$. Števila znotraj matrike predstavljajo koristi za oba igralca v nekem stanju, prvo število za prvega, drugo število pa za drugega.

	P	N
P	2,2	0,3
N	3,0	1,1

Vemo, da bo Nashevo ravnovesje v profilu akcij (N, N) , saj se v tistem trenutku nobenemu od igralcev ne izplača spremeniti strategije. Opazimo namreč, da v kolikor prvi igralec spremeni svojo akcijo iz N na P , bo na slabšem, saj je $0 < 1$. Simetrično velja tudi za drugega igralca. Lahko razmislimo, da noben drug profil strategij ne more biti Nashevo ravnovesje. Obravnavamo profil (P, P) , kar ni Nashevo ravnovesje, saj prvi igralec raje izbere akcijo N , namreč $3 > 2$. Sedaj obravnavajmo še (N, P) , kar prav tako ni Nashevo ravnovesje, saj bi v tem primeru drugi igralec raje izbral akcijo N .

V zgledu 29 bomo predstavili dvoobdobno zapornikovo dilemo kot ekstenzivno igro.

2. Neskončne ponavljajoče igre

Želimo modelirati igro, pri kateri igralci neskončnokrat ponovijo strateško igro G . Obravnavali bomo igre, pri katerih je množica akcij za vse igralce kompaktna množica, preferenčne relacije pa bodo zvezne funkcije. Pomagamo si z ekstenzivnimi igrami s popolno informacijo, igralci si torej zapomnijo predhodne poteze vseh igralcev. Dodatno želimo še to, da igralci sočasno opravljajo poteze.

Za začetek definirajmo še ekstenzivno igro s popolno informacijo.

Definicija 4 (Ekstenzivna igra s popolno informacijo). Ekstenzivna igra s popolno informacijo je štirica $(N, T, P, (u_i)_{i \in N})$, pri čemer je:

- N neprazna množica igralcev
- T drevo s korenem, ki nima neskončne poti, njegova množica listov je $L(T)$
- $P : V(T) \setminus L(T) \rightarrow N$ funkcija, ki pove, kdo je na vrsti v vozlišču v
- $u_i : L(T) \rightarrow \mathbb{R}$ funkcija preferenc za igralca i

Definicija 5 (Neskončna ponavljajoča igra). Naj bo $G = (N, (A_i)_{i \in N}, (\geq_i)_{i \in N})$ strateška igra, kjer je N neprazna množica igralcev. Akcije igralcev označimo z $(A_i)_{i \in N}$ in velja $A = \times_{i \in N} A_i$. Neskončna ponavljajoča igra G je ekstenzivna igra s popolno informacijo $(N, H, P, (\geq_i^*)_{i \in N})$, pri čemer velja:

- $H = \emptyset \cup (\cup_{t=1}^{\infty} A^t) \cup A^{\infty}$, kjer je $\emptyset$ prvotna zgodovina in A^{∞} množica neskončnih zaporedij $(a^t)_{t=1}^{\infty}$ profilov akcij v G .

- $P(h) = N$ za vsako nekončno zgodovino $h \in H$, funkcija P pa določa, kateri igralec je na vrsti v nekem obdobju
- $\geq_i^*$ je preferenčna relacija na A^∞ ki razširi preferenčno relacijo $\geq_i$ tako, da zadošča naslednjim pogojem šibke separabilnosti: Če $(a^t) \in A^\infty, a \in A, a' \in A$ in $a \geq a'$, potem

$$(a^1, \dots, a^{j-1}, a, a^{j+1}, \dots) \geq_i^* (a^1, \dots, a^{j-1}, a', a^{j+1}, \dots) \quad \text{za vse } j.$$

V definiciji z $\geq_i^*$ označujemo preferenčno relacijo v ponavljajoči igri, z $\geq_i$ pa preferenčno relacijo v prvotni strateški igri G . V primeru stroge separabilnosti uporabljamo oznako $>_i^*$. Igralčeva preferenčna relacija v ponavljajoči igri bo odvisna od funkcije izplačil u_i , ki predstavlja njegovo preferenčno relacijo v prvotni igri G .

Po vsaki nekončni zgodovini si vsak igralec $i \in N$ izbere akcijo iz A_i . Akcije za vse igralce so določene s pomočjo strategije; to je funkcija igralca i , ki vsakemu končnemu zaporedju izidov igre G priredi akcijo iz igralčevega nabora akcij A_i .

Preferenčno relacijo $\geq^*$ lahko definiramo na več načinov, za vsako definicijo bomo prilagodili potrebne izreke za iskanje Nashevih ravnovesij. Definiramo torej naslednje tri:

Kriterij diskontiranja Imamo dve zaporedji realnih števil (v_i^t) in (w_i^t) . Kriterij pravi, da obstaja tako število $\delta \in (0, 1)$, da bo zaporedje (v_i^t) realnih števil vsaj tako dobro kot zaporedje (w_i^t) natanko tedaj, ko $\sum_{t=1}^{\infty} \delta^{t-1} (v_i^t - w_i^t) \geq 0$. Številu δ pravimo diskontni faktor. Igralec zaporedje (v_i^t) oceni s pomočjo vsote $\sum_{t=1}^{\infty} \delta^{t-1} v_i^t$. Igralčev profil izplačil za zaporedje profilov izplačil iz prvotne igre $(v^t)_{t=1}^{\infty}$ je kar $\left((1 - \delta) \sum_{t=1}^T \delta^{t-1} v_i^t \right)_{i \in N}$.

Opomba 6. Za izplačila $(v^t)_{t=1}^{\infty}$ in $\delta \in (0, 1)$ imenujemo $(1 - \delta) \sum_{t=1}^T \delta^{t-1} v^t$ diskontirano povprečje izplačil za igralca.

Kriterij limite povprečij Zaporedje realnih števil (v_i^t) preferiramo pred zaporedjem realnih števil (w_i^t) natanko tedaj, ko je $\liminf_{T \rightarrow \infty} \sum_{t=1}^T (v_i^t - w_i^t)/T > 0$. Ekvivalentno, ko obstaja $\epsilon > 0$, za katerega je $\sum_{t=1}^T (v_i^t - w_i^t)/T > \epsilon$ za vse T . Igralčev profil izplačil je $\lim_{T \rightarrow \infty} \sum_{t=1}^T (v_i^t/T)_{i \in N}$.

Kriterija se razlikujeta po tem, da je pri kriteriju diskontiranja lahko pomembna že sprememba v enem samem obdobju, pri kriteriju limite povprečij pa je pomembna sprememba v povprečju. Spremembe v nekem končnem številu obdobj za kriterij limite povprečij niso pomembe. Igralca, ki se po njem odloča, zanima dolgoročno izplačilo in je pripravljen utrpeti nekaj izgube v končno mnogo začetnih obdobjih.

Kriterij prehitavanja Zaporedje (v_i^t) preferiramo pred zaporedjem (w_i^t) natanko tedaj, ko je $\liminf_{T \rightarrow \infty} \sum_{t=1}^T (v_i^t - w_i^t) > 0$.

Zgled 7 (Primerjava kriterijev). Z zgledom želimo pokazati, kako so različna zaporedja ovrednotena s temi tremi kriteriji, in katera zaporedja ti kriteriji preferirajo. Vzamimo nekaj zaporedij kot testne primere:

- Zaporedje $(0, 0, \dots, 2, 2, \dots)$ je po kriteriju limite povprečij boljše kot zaporedje $(1, 1, \dots)$.
- Zaporedje $(1, -1, 0, 0, 0, \dots)$ je po kriteriju diskontiranja (in za vse $\delta \in (0, 1)$) boljše kot zaporedje $(0, 0, \dots)$, glede na ostala dva kriterija pa sta zaporedji enako dobri.

- Zaporedje $(-1, 2, 0, 0, 0, \dots)$ je po kriteriju prehitevanja boljše kot $(0, 0, \dots)$, po kriteriju limite povprečij pa sta zaporedji enako dobri.

Naj bo $G = (N, (A_i), (\geq_i))$ strateška igra in u_i za vse $i \in N$ funkcija izplačil, ki predstavlja preferenčno relacijo $\geq_i$. Z $u(a)$ označimo profil $(u_i(a))_{i \in N}$.

Definicija 8 (δ -diskontirana neskončna ponavljajoča igra). δ -diskontirana neskončna ponavljajoča igra od G je neskončna ponavljajoča igra od G , kjer za vse $i \in N$ funkcijo izplačil u_i zamenjamo z δu_i .

Na podoben način definiramo igri za ostala dva kriterija in govorimo o neskončni ponavljajoči igri od G s kriterijem limite povprečij in o neskončni ponavljajoči igri od G s kriterijem prehitevanja.

Definicija 9 (Profil izplačil). Vektor $v \in \mathbb{R}^N$ imenujemo profil izplačil igre $(N, (A_i), (u_i))$, če obstaja izid $a \in A$, za katerega je $u(a) = v$.

Definicija 10 (Dopustni profil izplačil). Vektor $v \in \mathbb{R}^N$ je dopustni profil izplačil igre $(N, (A_i), (u_i))$, če je konveksna kombinacija profilov izplačil, ki leži v A . Pišemo $v = \sum_{a \in A} \alpha_a u(a)$, kjer so α_a nenegativna racionalna števila in $\sum_{a \in A} \alpha_a = 1$.

Opomba 11. Dopustni profil izplačil igre $(N, (A_i), (u_i))$ ni nujno profil izplačil igre $(N, (A_i), (u_i))$.

2.1 Strategije kot stroji

Vsak igralec se pri igri odloči, kako bo igral. Zato definiramo stroj, ki bo implementiral igralčevo strategijo v ponavljajočo igro. S pomočjo stroja namreč najlažje opišemo igralčevo strategijo v različnih stanjih igre. V vsakem obdobju se nahajamo v novem stanju.

Definicija 12 (Stroj za igralca i). Stroj za igralca i v neskončni ponavljajoči igri od $(N, (A_i), (\geq_i))$ je četverica $(Q_i, q_i^0, f_i, \tau_i)$, pri čemer:

- Q_i je množica stanj
- $q_i^0 \in Q_i$ je prvotno stanje
- $f_i : Q_i \rightarrow A_i$ je funkcija, ki vsakemu stanju pripiše akcijo (pravimo ji tudi izhodna funkcija)
- $\tau_i : Q_i \times A \rightarrow Q_i$ je funkcija, ki vsakemu paru, sestavljenemu iz stanja in profila akcij, pripiše stanje (pravimo ji tudi prehodna funkcija)

Oglejmo si nekaj zgledov strojev, pri katerih igramo ponavljajočo igro zapornikove dileme z akcijama $\{P, N\}$ in izplačili, ki so bila predstavljena v Zgledu 1.

Zgled 13 (Stroj za t.i. ‘sprožilno’ strategijo). V tem zgledu bomo spoznali, kako se obnašajo posamezne komponente stroja kot strategije. Stroj $(Q_i, q_i^0, f_i, \tau_i)$, ki izvaja to strategijo, izbira P , dokler sta oba igralca v preteklosti izbirala P , sicer pa presedla na N . Če podrobno opišemo posamezne komponente:

- $Q_i = \{\mathcal{P}, \mathcal{N}\}$,
- $q_i^0 = \mathcal{P}$
- $f_i(\mathcal{P}) = P$ in $f_i(\mathcal{N}) = N$
- $\tau_i(\mathcal{P}, (P, P)) = \mathcal{P}$ in $\tau_i(\mathcal{N}, (Y, Z)) = \mathcal{N}$, če $(\mathcal{X}, (Y, Z)) \neq (\mathcal{P}, (P, P))$

Pri tem s pisanimi črkami označujemo stanja, s tiskanimi pa akcije.

Zgled 14. Oglejmo si še en primer strategij kot strojev ter zabeležimo izide igralcev. Določimo najprej stroja za oba igralca, označimo ju s S_1 in S_2 . Stroj 1 igra takole: Igra P , dokler igralec 2 igra P , sledi igranje N v treh obdobjih, nato pa se vrne na P , v kolikor igralec 2 igra N na mestu, kjer bi moral igrati P . Stroj torej kaznuje nasprotnika tri kroge in mu potem odpusti. Opazimo, da je smiselno gledati vsaj 4 obdobja. Stroj 2 pa igra takole: Začne s P in s tem nadaljuje dokler nasprotnik igra N . Če pa nasprotnik igra P , v naslednjem obdobju tudi sam spremeni na P in to igra, dokler nasprotnik spet ne igra N . V naslednjem obdobju po tem stroj 2 spet izbere P .

Zabeležimo sedaj izide igralcev v tabeli:

Obdobje	Stanje S_1	Stanje S_2	Izid	Izplačili
1	P_0	R_0	(P,P)	2,2
2	P_0	R_1	(P,N)	0,3
3	P_1	R_1	(N,N)	1,1
4	P_2	R_0	(N,P)	3,0
5	P_3	R_0	(N,P)	3,0
6	P_0	R_0	(P,P)	2,2

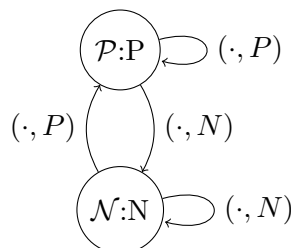
Stroja začneta v stanjih P_0 in R_0 . Izhodni funkciji strojev S_1 in S_2 tema stanjema pripiše akciji (P, P) . Prehodna funkcija za S_1 ohrani stroj v stanju P_0 , prehodna funkcija za S_2 pa zaradi izida v prešnji igri stanje spremeni na R_1 . Par stanj v novem obdobju je sedaj (P_0, R_1) . Na podoben način se premikamo do šestega obdobja, kjer opazimo, da se stanja in izidi igre venomer ponavljajo. Spet se namreč nahajamo v prvotnem stanju.

Zgled 15 (Strategija Tit-for-tat). Poglejmo si še eno strategijo, t.i. Tit-for-tat strategijo. Pokažemo, da lahko strategijo kot stroj prikažemo z diagramom. Značilnost te strategije je, da je dolžina kaznovanja nasprotnika odvisna od tega, koliko časa nas on kaznuje. Če nasprotnik igra P , potem tudi mi igramo P , če pa presedla na N , pa tudi mi zamenjamo svojo akcijo v naslednjem krogu.

Zapišimo strategijo kot stroj po definiciji:

- $Q_i = \{\mathcal{P}, \mathcal{N}\}$
- $q_i^0 = \mathcal{P}$
- $f_i(\mathcal{P}) = P$ in $f_i(\mathcal{N}) = N$
- $\tau_i(\mathcal{P}, (\cdot, P)) = \mathcal{P}$, $\tau_i(\mathcal{P}, (\cdot, N)) = \mathcal{N}$, $\tau_i(\mathcal{N}, (\cdot, P)) = \mathcal{P}$, $\tau_i(\mathcal{N}, (\cdot, N)) = \mathcal{N}$

Oglejmo si diagram, s katerim predstavimo stroj:



2.2 Nasheva ravnovesja

V tem razdelku se bomo ukvarjali z Nashevimi ravnovesji v neskončnih ponavljajočih igrah. Opazimo namreč, da množica Nashevih ravnovesij ne vsebuje zgolj izidov, ki so ponovitve Nashevega ravnovesja prvotne igre, saj se lahko igralci poslužujejo različnih odstopanj, kot so npr. kaznovanje.

Definicija 16 (Izvršljiv profil izplačil). Imamo dva profila izplačil v in w . Pravimo, da je profil izplačil w izvršljiv, če velja $w_i \geq v_i$ za vse $i \in N$, kjer je v_i dopustni profil izplačil. Profil izplačil w je strogo izvršljiv, če velja $w_i > v_i$ za vse $i \in N$.

Definicija 17 (Minmax izplačilo v strateški igri). Minmax izplačilo igralca i v strateški igri je

$$\min_{a_{-i} \in A_{-i}} \left(\max_{a_i \in A_i} u_i(a_i, a_{-i}) \right),$$

pri čemer je A_j množica akcij in u_j funkcija izplačil za vse $j \in N$.

Naj bo G strateška igra in profil $u(a)$ izvršljiv v igri G . Če je $a \in A$ izid v G , pravimo, da je a izvršljiv izid igre G . Sedaj iščemo rešitve minimizacijskega problema, ki smo ga navedli v definiciji 17. Naj bo $p_{-i} \in A_{-i}$ ena od teh rešitev. Za vsak profil $a \in A$ definiramo akcijo $b_i(a_{-i}) \in A_i$ igralca i , ki je najboljši odgovor na a_{-i} in seveda leži v množici najboljših odgovorov B_i (torej $b_i(a_{-i}) \in B_i(a_{-i})$). S takim naborom akcij p_i lahko najbolj kaznujemo igralca i v igri G .

2.3 Ljudski izreki za različne oblike preferenčnih funkcij

Sedaj bomo navedli nekaj t.i. ljudskih izrekov. Preučili bomo nekatere možne strategije in kako se igralci odzovejo. Ideja je ta, da želijo igralci slediti nekemu ravnovesju, če pa se eden od igralcev odloči odstopati od ravnovesja, ga preostali igralci kaznujejo. Glede na različne kriterije tvorimo različne strategije, s katerimi se povrnemo v ravnovesje. Govorili bomo tudi o popolnem ravnovesju podigre. Par strategij v ekstenzivni igri je popolno ravnovesje podigre, če par strategij predstavlja Nashevo ravnovesje v vsaki podigri prvotne igre.

Opomba 18. O ravnovesni poti govorimo tedaj, ko igralci v vseh obdobjih sledijo tistim profilom akcij, ki jih bo pripeljalo v ravnovesje. V tem poglavju bomo govorili o odmikih iz ravnovesne poti, torej profil akcij igralca i v nekem obdobju ne bo več ustrazal ravnovesju, zato ga npr. preostali kaznujejo. O podigrah govorimo tedaj, ko začnemo opazovati igro šele po neki zgodovini h . Na podoben način kot v ‘celih’ igrah tudi pri podigrah iščemo ravnovesja.

Načelo enega odklona Noben igralec ne mora povišati svojih izplačil z odstopanjem v enem obdobju.

Navedemo posledico 144.3 iz [1] s strani 144:

Trditev 19. Vsak profil izkupičkov v Nashevem ravnovesju neskončne ponavljajoče igre od $G = (N, (A_i), (u_i))$ po kriteriju limite povprečij je izvršljiv profil izplačil igre G . Enako velja za vse $\delta \in (0, 1)$ in profil izplačil v Nashevem ravnovesju δ -diskontirane neskončne ponavljajoče igre od G .

Dokaz. Denimo, da je w profil izplačil v neskončni ponavljajoči igri od G po kriteriju limite povprečij, ki pa ni izvršljiv. Torej je $w_i < v_i$. Potem w ni profil izplačil v Nashevem ravnovesju ponavljajoče igre, saj igralcu i strategija $s'_i(h) = b_i(s_{-i}(h))$ za vsako zgodovino h in v vsakem obdobju da vsaj izplačilo v višini v_i . To velja za vsak profil strategij s in torej w_i ni profil izplačil v Nashevem ravnovesju. Enak dokaz priteče tudi δ -diskontirani neskončni ponavljajoči igri od G . ■

Navedemo posledico 144.3 iz [1] s strani 144:

Izrek 20 (Nashev ljudski izrek za kriterij limite povprečij). *Imamo kriterij limite povprečij in neskončno ponavljajočo igro. Vsak dopusten izvršljiv profil izplačil strateške igre $G = (N, (A_i), (u_i))$ je profil izplačil (v Nashevem ravnovesju v neskončni ponavljajoči igri od G po kriteriju limite povprečij).*

Dokaz. Naj bo $w = \sum_{a \in A} (\beta_a / \delta) u(a)$ dopustni izvršljiv profil izplačil strateške igre G , kjer je $\beta_a \in \mathbb{Z}$ za vse $a \in A$ in $\gamma = \sum_{a \in A} \beta_a$. Naj bo (a^t) ciklično zaporedje profilov akcij dolžine γ . V zaporedju se profil a ponovi β_a krat. Oglejmo si sedaj strategijo s_i za ponavljajočo igro. V tej strategiji igralec i v vsakem obdobju t izbere a_i^t , razen če v enem od prejšnjih obdobjih t' nek drug igralec ni izbral $a_i^{t'}$. V tem primeru igralec i izbere $(p_{-j})_i$, kjer j označuje prvo obdobje, v katerem se ta odmik pojavi. Torej je s_i strategija, ki igralcu i zagotavlja izplačilo w_i v vsakem obdobju in w je profil izplačil Nashevega ravnovesja v neskončni ponavljajoči igri od G . ■

Navedemo posledico 145.2 iz [1] s strani 145:

Izrek 21 (Nashev ljudski izrek za kriterij diskontranja). *Naj bo w strogo izvršljiv dopustni profil izplačil strateške igre $G = (N, (A_i), (u_i))$. Za vsak $\epsilon > 0$ obstaja tak $\underline{\delta} < 1$, da v kolikor je $\delta > \underline{\delta}$, ima δ -diskontirana neskončna ponavljajoča igra od G Nashevo ravnovesje, v katerem je profil izplačil w' , za katerega velja $|w' - w| < \epsilon$.*

Imamo različne strategije, recimo, da kaznujemo igralca, ki nekako odstopa od ravnovesja. Igralec, ki odstopa, bo kaznovan (njegovo izplačilo bo zgolj minmax) natanko tokrat, da ne bo več v prednosti zaradi odstopanja. Po kriteriju limit povprečij taka strategija vrne enak izkupiček kot če bi se samo držali le poti ravnovesja.

Navedemo posledico 146.2 iz [1] s strani 146:

Izrek 22 (Popoln ljudski izrek za kriterij limite povprečij). *Imamo kriterij limite povprečij in neskončno ponavljajočo igro. Vsak dopusten strogo izvršljiv profil izplačil igre $G = (N, (A_i), (u_i))$ je popolno ravnovesje profila izplačil podigre (v neskončni ponavljajoči igri od G po kriteriju limite povprečij).*

Pred dokazom navedimo še naslednje: Naj bo $w = \sum_{a \in A} \alpha_a u(a)$ dopustni profil in denimo, da je $\alpha_a = \beta_a / \gamma$ za vse $a \in A$, kjer $\beta_a \in \mathbb{Z}$ in $\gamma = \sum_{a \in A} \beta_a$. Osredotočimo se na zaporedje izidov v ponavljajoči igri, ki je sestavljeno iz ponovitev cikla dolžine δ , v kateri se vsak $a \in A$ ponovi natanko β_a -krat. Tak cikel izidov prinese povprečno izplačilo w .

Skica dokaza. Naj bo $w = \sum_{a \in A} (\beta_a / \delta) u(a)$ dopusten strogo izvršljiv profil izplačil strateške igre G in naj bo $(a^k)_{k=1}^\delta$ zaporedje profilov akcij, v katerem se profil a pojavi β_a -krat za vse $a \in A$. Želimo skonstruirati strategijo, ki generira zaporedja profilov akcij v G in je sestavljena iz ponovitev cikla $(a^k)_{k=1}^\delta$. Če eden od igralcev, i , odstopa od poti ravnovesja v obdobju (v končno mnogo obdobjih), v katerem naj ne bi bil nihče kaznovan, potem ostali igralci igralca i kaznujejo v naslednjih obdobjih. To storijo tako, da igrajo p_{-i} , tako dolgo, dokler se prednost igralca i ne izniči. Kaznovani igralec se posledično vrne v ravnovesje, in spet začne na začetku cikla, zaradi kriterija limite povprečij pa je profil izplačil v vsakem obdobju enak w . Skonstruiramo strategijo, ki si jo bralec lahko pogleda v [1]. ■

Naslednji izrek nam da analogen rezultat kot izrek 4, vendar pa tu upoštevamo kriterij prehitevanja.

Navedemo posledico 149.1 iz [1] s strani 149:

Izrek 23 (Popoln ljudski izrek za kriterij prehitevanja). *Imamo kriterij prehitevanja in neskončno ponavljajočo igro. Za vsak strogo izvršljiv izid a^* igre $G = (N, (A_i), (u_i))$ obstaja popolno ravnovesje podigre (v neskončni ponavljajoči igri od G po kriteriju prehitevanja), ki generira pot (a^t) , v kateri je $a^t = a^*$ za vsak t .*

Dokaz. Naj bo $M = \max_{i \in N, a \in A} u_i(a)$. Imejmo sedaj strategijo, pri kateri vsak igralec i uporablja naslednji stroj:

- Igralec se lahko nahaja v stanju, ki ustreza stanju v ravnovesni poti (označimo ga z Norm) ali pa v stanju, ko mora biti nek drug igralec kaznovan. To označimo kot $\{\text{Norm}\} \cup \{P(j, t)\}$, kjer je $j \in N$ igralec, ki je odstopil od ravnovesne poti, $t \in \mathbb{Z}_+$ pa število obdobj, v katerem mora biti kaznovan.
- Prvotno stanje je Norm.
- Izhodna funkcija je določena za obe stanji. Če je igralec v stanju Norm, potem igra a_i^* . V stanju $P(j, t)$ pa igra (p_{-j}) , če $i \neq j$ in $b_i(p_{-j})$, če $i = j$.
- Definiramo še prehode glede na izid $a \in A$: Iz stanja Norm gremo v Norm, razen, če za nekega igralca j velja $a_{-j} = a_{-j}^*$ in $a_j \neq a_j^*$. V tem primeru se premaknemo v stanje $P(j, t)$, kjer je $t \in \mathbb{Z}$ najmanjše tako število, da velja $M + tv_j < (t + 1)u_j(a^*)$. Obravnajamo še situacijo, ko smo začeli v stanju $P(j, t)$. Če za vsaj dva igralca $i \in N$ velja $a_{-j} = p_{-j}$ ali $a_l \neq (p_{-j})_l$, potem gremo iz stanja $P(j, t)$ v $P(j, t - 1)$, v kolikor $t \geq 2$, in v stanje Norm, v kolikor $t = 1$. Če pa veljata pogoja $a_{-j} \neq p_{-j}$ in $a_l = (p_{-j})_l$, v kolikor je $l \neq j^*$, se premaknemo v $P(j^*, T(j, t))$. Pri tem je $T(j, t)$ tako veliko število, da vsota izplačil igralca j , ki jih dobi iz stanj $P(t, j)$ in obdobj $T(t, j)$, v katerih bo kaznovan, večja kot vsota izplačil, ki bi jih dobil s tem, ko se je odmaknil od ravnovesja ter $T(j, t)v_j^*$.

Ta strategija torej zatre vsak poskus igralca, da bi povečal svoj izkupiček, saj je tako odstopanje srečano s kaznovanjem s strani ostalih igralcev. Ta profil strategij je prav tako popolno ravnovesje podigre. ■

Naslednji izrek nam da analogen rezultat kot izrek 4, vendar pa tu upoštevamo kriterij diskontiranja.

Navedemo posledico 151.1 iz [1] s strani 151:

Izrek 24 (Popoln ljudski izrek za kriterij diskontiranja). *Naj bo a^* strogo izvršljiv izid igre $G = (N, (A_i), (u_i))$. Denimo, da obstaja tak nabor $(a(i))_{i \in N}$ strogo izvršljivih izidov igre G , da za vse $i \in N$ velja $a^* >_i a(i)$ in $a(j) >_i a(i)$ za vse $j \in N$. Potem obstaja tak $\underline{\delta} < 1$, da za vse $\delta > \underline{\delta}$ obstaja popolno ravnovesje podigre v δ -diskontirani neskončni ponavljajoči igri od G , ki generira pot (a^t) , v kateri je $a^t = a^*$ za vsak t .*

Dokaz izpustimo, bralec si ga lahko ogleda v [1]. Ponovno skonstruiramo strategijo, tokrat tako, pri kateri imamo zopet igralca, ki odstopi od ravnovesne poti. Vendar pa so tukaj igralci, ki kaznujejo igralca, željnega profitirati, nagrajeni.

Zgled 25. Poglejmo si sedaj na primeru ene od strategij, ki smo jih prej spoznali. Odločali se bomo po kriteriju diskontiranja. Prvi igralec bo uporabljal sprožilno strategijo iz zgleda 13. Če se bo tudi drugi igralec ravnal po sprožilni strategiji, bo izid v vsakem obdobju kar (P, P) in bodo izplačila za oba igralca kar $(2, 2, \dots)$. Če pa se drugi igralec odloči, da bo ubral drugo strategijo, bo v vsaj enem obdobju igral N , zaradi česar bo prvi igralec v vseh prihodnjih obdobjih igral N . Zatorej bo za drugega igralca kar najbolje, da v vsakem naslednjem obdobju igra N , saj je to najboljši odgovor

na N . Do trenutka, ko oba igrata P , bo v vsakem obdobju izplačilo kar 2 za oba igralca, ko pa drugi spremeni akcijo, pa bodo izidi kar $(3, 1, 1, 1, \dots)$. Diskontirano povprečje v tem primeru bo

$$(1 - \delta)(3 + \delta + \delta^2 + \dots) = (1 - \delta) \left(3 + \frac{\delta}{1 - \delta} \right) = 3(1 - \delta) + \delta.$$

Torej drugi igralec ne more povišati svojega izkupička natanko tedaj, ko

$$3(1 - \delta) + \delta \leq 2,$$

oziroma $\delta \geq \frac{1}{2}$. Prav tako lahko zaključimo, da v kolikor je $\delta \geq \frac{1}{2}$, je to popolno Nashevo ravnovesje v δ -diskontirani neskončni ponavljajoči igri zapornikove dileme.

3. Končne ponavljajoče igre

Sedaj se osredotočimo na končne ponavljajoče igre. Definiramo jih podobno kot neskončne ponavljajoče igre. Spet namreč opisujemo ekstenzivno igro s popolno informacijo, igralci pa poteze izvajajo sočasno.

Definicija 26 (Končna ponavljajoča igra). Naj bo $G = (N, (A_i)_{i \in N}, (\geq_i)_{i \in N})$ strateška igra, kjer je N neprazna množica igralcev. $Z(A_i)_{i \in N}$ označimo akcije igralcev. Naj bo $A = \times_{i \in N} A_i$. Neskončna ponavljajoča igra G je ekstenzivna igra s popolno informacijo $(N, H, P, (\geq_i^*)_{i \in N})$, pri čemer velja:

- $H = \emptyset \cup (\cup_{t=1}^T A^t) \cup A^T$, kjer je $\emptyset$ prvotna zgodovina in A^T množica končnih zaporedij $(a^t)_{t=1}^T$ profilov akcij v G .
- $P(h) = N$ za vsako nekončno zgodovino $h \in H$
- $\geq_i^*$ je preferenčna relacija na A^T , ki razširi preferenčno relacijo $\geq_i$ tako, da zadošča naslednjim pogojem šibke separabilnosti: Če $(a^t) \in A^T, a \in A, a' \in A$, in $a \geq a'$, potem

$$(a^1, \dots, a^{j-1}, a, a^{j+1}, \dots) \geq_i^* (a^1, \dots, a^{j-1}, a', a^{j+1}, \dots) \quad \text{za vse } j.$$

Preferenčno relacijo $\geq_i^*$ vsakega igralca i predstavimo s funkcijo $\sum_{t=1}^T u_i(a^t)/T$. Pri tem je u_i funkcija izplačil in predstavlja preference igralca i v osnovni strateški igri. Na to igro se nanašamo kot na T -obdobno ponavljajočo igro od $(N, (A_i), (u_i))$.

Intuicija za izid v Nashevem ravnovesju končnih ponavljajočih iger je ta, da mora biti Nashevo ravnovesje v zadnjem obdobju ponavljajoče igre kar Nashevo ravnovesje v osnovni strateški igri G , saj igralcem v zadnjem obdobju ne preti več nevarnost, da bi bili kasneje kaznovani. To vpliva na celotno igro, še posebej v primeru kot je zapornikova dilema, kjer je Nashevo ravnovesje enako minmax izplačilu vsakega igralca, zaradi česar bo moral biti vsak izid v ponavljajoči igri kar Nashevo ravnovesje osnovne igre. Zapišimo posledice, ki nam pomagajo to bolj formalno utemeljiti. Pri tem bo v končnih ponavljajočih igrah veljalo, da lahko popolno ravnovesje podigre nastopi le tam, kjer je izpolnjeno načelo enega odklona.

Navedemo posledico 155.1 iz [1] s strani 155:

Trditev 27. Naj bo profil izplačil v vsakem Nashevem ravnovesju strateške igre G kar profil (v_i) minmax izplačil v G . Za vse T in izid $(a^1, \dots, a^T)$ vsakega Nashevega ravnovesja v T -obdobni ponavljajoči igri od G velja, da je a^t Nashevo ravnovesje v G za vsak $t = 1, \dots, T$.

Dokaz. Imamo strateško igro $G = (N, (A_i), (u_i))$. Imamo Nashevo ravnovesje s in $a = (a^1, \dots, a^T)$ izid Nashevega ravnovesja s v T -obdobni ponavljajoči igri od G . Dokazujemo s protislovjem. Denimo, da za neko obdobje t , a^t ni Nashevo ravnovesje igre G . Naj bo $t \geq 1$ zadnje obdobje t , za katerega a^t ni več Nashevo ravnovesje. Torej tedaj velja $u_i(a_{-1}^t, a_i) > u_i(a^t)$. Imamo še eno strategijo $\hat{s}_i$ igralca i , ki se od s_i razlikuje le po tem, da po zgodovini $(a^1, \dots, a^{t-1})$ vzame a_i . Po vsakršni daljši zgodovini h pa vzame akcijo, ki nam da igralcu i vsaj minmax izplačilo. Pri tem upoštevamo profil akcij $s_i(h)$, ki so jih nameravali izvršiti ostali igralci po zgodovini h . Izid profila $(s_{-i}, \hat{s}_i)$ je kar nekončna zgodovina $\hat{a}$, ki je enak nekončni zgodovini a v obdobju $t - 1$. To lahko zapišemo kot: $u_i(\hat{a}^t) > u_i(a^t)$ in $u_i(\hat{a}^r) \geq u_i(a^r)$ za $r \geq t + 1$. Ugotovili smo, da igralec i preferira $\hat{a}$ pred a , kar pa je jasno v protislovju s predpostavko, da je s Nashevo ravnovesje ponavljajoče igre. ■

Navedemo posledico 157.2 iz [1] s strani 157:

Izrek 28. *Naj bo G strateška igra, ki ima natanko en profil izplačil v Nashevem ravnovesju. Potem za vse T velja, po vsaki zgodovini, da izberemo profil akcij, ki je popolno ravnovesje podigre. Velja, da je izbran profil akcij za popolno ravnovesje podigre v T -ponavljajoči igri G kar Nashevo ravnovesje igre G .*

Dokaz. Izid v vsaki podigri, ki se začne v obdobju T (vseh popolnih ravnovesij podiger ponavljajoče igre), je Nashevo ravnovesje igre G . Izid vseh igralcev je neodvisen od zgodovine. Enako razmišljanje apliciramo na obdobje $T - 1$. Zopet velja, da je profil akcij v tej podigri Nashevo ravnovesje igre G . Z indukcijo lahko to apliciramo na vsa dotedanja obdobja, kar zaključí dokaz. ■

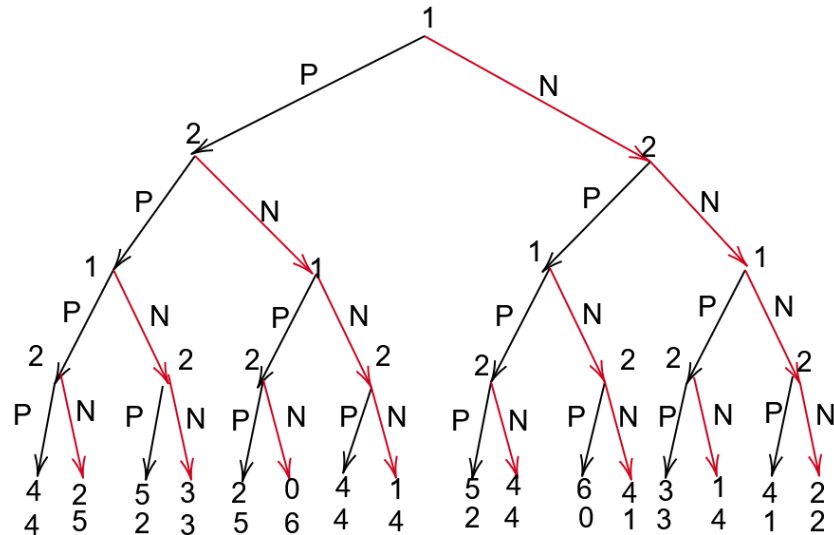
Prvotna enoobdobna igra ima lahko tudi več Nashevih ravnovesij, kar terja nove izreke, vendar pa se s tem v tem članku ne bomo bolj podrobno seznanili. V kolikor bralca zanima več, si lahko za začetek pogleda snov v delu [1].

Prišli smo torej do sklepa, da je v zgledu zapornikove dileme Nashevo ravnovesje profil akcij (N, N) . Tudi v vsakem obdobju ponavljajoče igre bo venomer tak izid, torej bo v T -obdobni zapornikovi dilemi izid kar T -kratna ponovitev profila akcij (N, N) .

4. Ponavljajoča zapornikova dilema

Zgled 29 (Predstavitev ponavljajoče zapornikove dileme z drevesom). Predstavimo ponavljajočo zapornikovo dilemo na način, kot predstavljamo ekstenzivne igre. V zgledu se ukvarjamo z dvoobdobno ponavljajočo zapornikovo dilemo. Primer rešujemo kot običajno ekstenzivno igro, od spodaj navzgor. Vrednosti v listih predstavljajo koristi obeh igralcev, zgornje število je korist prvega, spodnje pa korist drugega. V vozliščih števili 1 in 2 označujeta, kateri igralec izbira, na povezavah pa sta označeni akciji $\{P, N\}$ za vsakega igralca. Igralec, ki je na potezi, torej primerja obe vrednosti, in seveda izbere potezo z večjo koristnostjo zanj. Izbrane povezave označimo z rdečo barvo.

Opazimo seveda, da je odločitev igralca v obdobju T neodvisna od odločitev v prejšnjih obdobjih. Profil akcij v Nashevem ravnovesju bo (N, N) .



Slika 1. Dvoobobna zapornikova dilema

5. Zanimivost

V 80ih letih prejšnjega stoletja je Robert Axelrod naprosil nekaj strokovnjakov s področij matematike, ekonomije, psihologije in sociologije, ki so bili seznanjeni s problemom zapornikove dileme, da oddajo svoje predloge za strategijo v končni ponavljajoči igri zapornikove dileme. Prejel je 14 odgovorov in jih pomeril proti strategiji, ki naključno (in z verjetnostjo $1/2$) izbira akciji N in P . Igra je imela 200 rund, prav tako je vsaka strategija igrala proti vsaki petkrat. Najbolje se je odrezala strategija tit-for-tat psihologa Anatola Rapoport. Kasneje je Axelrod organiziral še en turnir, na katerega je povabil vse udeležence iz prejšnjega, dodatne tekmovalce pa pridobil z oglaševanjem v računalniški reviji. Tokrat je tekmovalo 62 strategij, dolžina igre je bila določena verjetnostno, na koncu pa se je še vedno za najboljšo izkazala strategija tit-for-tat.

6. Zaključek

Predstavili smo osnovno idejo ponavljajočih iger in formulirali ljudske izreke za neskončne ponavljajoče igre, preko katerih iščemo Nasheva ravnovesja. Pokazali smo, da je v končni ponavljajoči igri zapornikove dileme Nashevo ravnovesje kar profil akcij, kjer igralca v vsakem obdobju igrata (N, N) . Lahko bi temo nadaljevali še v smeri, ko bi bilo več Nashevih ravnovesij. Dokazov smo se predvsem pri različnih kriterijih v neskončnih ponavljajočih igrah izogibali, vendar pa njihov ogled toplo priporočam, saj so v njih ponazorjene zanimive strategije. Bralec si jih lahko ogleda v [1]. V kolikor bralca zanimajo bolj praktični rezultati v povezavi z zapornikovo dilemo, naj si ogleda vir [2]. V viru [3] je predstavljen še morda nekoliko drugačen pogled na tematiko, tam je zapornikova dilema predstavljena kot drevo in je morda bolj celovito predstavljen ta aspekt.

LITERATURA

- [1] M.J. Osborne, A. Rubinstein, *A Course in Game Theory*, (1994), 133–160.
- [2] M.J. Osborne, *An introduction to game theory*, (2004), 419–464.
- [3] MIT OpenCourseWare, *Repeated games, chapter 12*, (2012), Dostopno na povezavi: https://ocw.mit.edu/courses/14-12-economic-applications-of-game-theory-fall-2012/54f153ef7ed7800f367ce2b4021a18db_MIT14_12F12_chapter12.pdf, dostop 15.2.2025.