

Teorija socialne izbire

Teorija socialne izbire se ukvarja s problemom, kako individualne preference posameznikov iz neke skupine pretvoriti v kolektivno preferenco te skupine. Primer so na primer volitve, kjer želimo iz mnenj posameznikov o kandidatih določiti zmagovalca volitev, ali pa ko iz mnenj posameznih sodnikov na športnem tekmovanju (drsanje, ples, gimnastika,...) želimo določiti končni vrstni red tekmovalcev. Podobno je včasih pri odločanju v parlamentu, ko morajo poslanci izbirati med različnimi predlogi zakonov za rešitev nekega problema.

Običajno imamo neko končno množico izbir $X = \{x_1, x_2, \dots, x_n\}$, posamezniki pa te izbire individualno rangirajo glede na njihove osebne preference. Za osebne preference bomo zahtevali, da so predstavljene z binarno relacijo ne-stroge šibke urejenosti, ki jo bomo v našem primeru raje poimenovali preferenčna relacija.

DEFINICIJA. Naj bo X neprazna množica izbir. *Preferenčna relacija* na množici X je binarna relacija R na množici X , za katero velja

- stroga sovisnost: $\forall x, y \in X : xRy \vee yRx$,
- tranzitivnost: $\forall x, y, z \in X : xRy \wedge yRz \implies xRz$.

Pri preferenčni relaciji lahko posamezniki določene izbire rangirajo kot enako dobre. Iz vsake preferenčne relacije R na množici X lahko naredimo relacijo stroge šibke urejenosti (označevali jo bomo z \tilde{R}), tako da definiramo

$$x\tilde{R}y \iff xRy \wedge \neg yRx.$$

Relacija \tilde{R} je podrelacija preferenčne relacije R . Če relacijo xRy lahko preberemo kot “ x je boljši ali enako dober kot y ” ali “ x je preferiran nad ‘ y ’”, potem relacijo $x\tilde{R}y$ preberemo kot “ x je boljši kot y ” ali “ x je strogo preferiran nad ‘ y ’”.

Naj bo X neprazna množica izbir. Z $L(X)$ označimo množico vseh preferenčnih relacij na množici X , torej vseh možnih načinov, kako lahko nek posameznik rangira izbire.

DEFINICIJA. Naj bo N naravno število. *Funkcija socialne blaginje* je vsaka preslikava

$$f : L(X)^N \rightarrow L(X).$$

V zgornji definiciji število N predstavlja število posameznikov, ki se odločajo glede danih izbir iz množice X . Vsak element $\mathcal{R} = (R_1, R_2, \dots, R_N)$ iz $L(X)^N$ predstavlja nek *seznam preferenc*, ki si ga lahko predstavljamo kot vse oddane volilne lističe. Če je $\mathcal{R} = (R_1, R_2, \dots, R_N)$ seznam preferenc, definiramo $x\mathcal{R}y$, če je xR_iy za vsak $i = 1, 2, \dots, N$. To pomeni, da vsak

posameznik preferira x nad y . Podobno definiramo $x\tilde{\mathcal{R}}y$, če $x\tilde{R}_i y$ za vsak $i = 1, 2, \dots, N$, kar pomeni, da vsak posameznik strogo preferira x nad y .

Če je $\emptyset \neq Y \subset X$ podmnožica množice izbir X in R preferenčna relacija na množici X . Z $R|_Y$ označimo zožitev relacije R na množico Y . Bolj natančno, $R|_Y$ je preferenčna relacija na množici Y , definirana z $\forall x, y \in Y : xR|_Y y \iff xRy$. Za seznam preferenc $\mathcal{R} = (R_1, R_2, \dots, R_N)$ označimo $\mathcal{R}|_Y = (R_1|_Y, \dots, R_N|_Y)$.

DEFINICIJA. Naj bo N naravno število. *Funkcija socialne izbire* je vsaka preslikava

$$f : L(X)^N \rightarrow X.$$

Razlika med funkcijo socialne blaginje in funkcijo socialne izbire je, da nam prva vrne neko preferenčno relacijo na množici izbir (nam med seboj rangira vse izbire), medtem ko nam druga vrne samo “zmagovalno” izbiro. Funkcijo socialne izbire uporabljamo na primer takrat, ko volimo predsednika republike.

Izrek Arrowa o nezmožnosti

Funkcija socialne blaginje naj bi iz preferenc posameznikov določila preference skupine. Običajno si želimo, da bi posamezniki, tudi če se končen rezultat ne ujema z njihovimi željami, sam proces priznavali kot pošten. V tem razdelku bomo videli, da v primeru, ko imamo vsaj dva posameznika in vsaj tri izbire, ne obstaja funkcija socialne blaginje, ki bi vedno zadoštila nekim minimalnim pogojem pravičnosti. To seveda ne pomeni, da v nobenem primeru ne moremo pošteno sestaviti preferenčno relacijo skupine. Če imajo na primer vsi posamezniki popolnoma enake želje, potem ni težko določiti preferenčno relacijo skupine, ki se bo vsem zdela poštena. Kar ne morem doseči je, da vnaprej dogovorjen sistem določitve preferenčne relacije skupine izpade pošten ne glede na to, kako se odločajo sami posamezniki. Kaj pomeni, da je sistem (minimalno) pošten, si pogledajmo v nadaljevanju.

DEFINICIJA (Paretovo načelo). Funkcija socialne blaginje $f : L(X)^N \rightarrow L(X)$ zadošča Paretovem načelu, če za vsak par $x, y \in X$ in za vsak seznam preferenc $\mathcal{R} \in L(X)^N$, za katerega velja $x\tilde{\mathcal{R}}y$, sledi $x f(\tilde{\mathcal{R}}) y$.

Funkcija socialne blaginje zadošča Paretojemu načelu, ko bo skupina strogo preferirala izbiro x nad y vedno, če vsi posamezniki v skupino strogo preferirajo izbiro x nad y .

DEFINICIJA (Neodvisnost od irelevantnih alternativ). Funkcija socialne blaginje $f : L(X)^N \rightarrow L(X)$ je neodvisna od irelevantnih alternativ, če za vsak par $x, y \in X$ in za vsaka dva seznama preferenc $\mathcal{R} = (R_1, R_2, \dots, R_N) \in L(X)^N$ in $\mathcal{R}^* = (R_1^*, R_2^*, \dots, R_N^*) \in L(X)^N$, za katera velja $\mathcal{R}|_{\{x,y\}} = \mathcal{R}^*|_{\{x,y\}}$, sledi $f(\mathcal{R})|_{\{x,y\}} = f(\mathcal{R}^*)|_{\{x,y\}}$.

Funkcija socialne blaginje zadošča aksiomu o neodvisnosti od irelevantnih alternativ, če bo pri dveh seznamih preferenc, ki se nič ne razlikujeta, kar se tiče dveh konkretnih izbir x nad y , preferenca skupine enaka, kar se tiče teh dveh alternativ. Med drugim to pomeni, da če so vsi posamezniki že

odločeni, kar se tiče odnosa med izbirama x in y , potem ne bodo mogli vplivati na odnos skupine do tih dveh izbir s tem, da se špekulativno odločajo o ostalih izbirah.

DEFINICIJA (Obstoj diktatorja). Funkcija socialne blaginje $f : L(X)^N \rightarrow L(X)$ je diktatorska, če obstaja tak indeks i , da za vsak par $x, y \in X$ in za vsak seznam preferenc $\mathcal{R} = (R_1, R_2, \dots, R_N) \in L(X)^N$ velja $x \tilde{R}_i y \implies x f(\tilde{\mathcal{R}}) y$.

Funkcija socialne izbire je torej diktatorska, če obstaja fiksni posameznik, ki ne glede na to, kako vsi glasujejo, vedno on odloči o rangiranju izbir.

IZREK (Izrek Arrowa). *Naj bo moč množice izbir X vsaj 3. Potem je vsaka funkcija socialne blaginje $f : L(X)^N \rightarrow L(X)$, za katero velja Paretovo načelo in neodvisnost od irelevantnih alternativ, diktatorska.*

DOKAZ. Naj bodo x, y in z trije različni si elementi izmed množice izbir X . Za vsak $i = 0, 1, 2, \dots, N$ naj bo L_i množica vseh tistih seznamov preferenc $\mathcal{R} = (R_1, R_2, \dots, R_N)$, za katere velja:

- $x \tilde{R}_j y$ in $z \tilde{R}_j y$ za vse $j > i$,
- $y \tilde{R}_j x$ in $y \tilde{R}_j z$ za vse $j \leq i$.

Po Paretovem načelu za vsak $\mathcal{R} \in L_0$ velja $x f(\tilde{\mathcal{R}}) y$ in $z f(\tilde{\mathcal{R}}) y$ in za vsak seznam preferenc $\mathcal{R} \in L_N$ velja $y f(\tilde{\mathcal{R}}) x$ in $y f(\tilde{\mathcal{R}}) z$. Naj bo k najmanjše tako naravno število, da za vsak $\mathcal{R} \in L_k$ velja $y f(\tilde{\mathcal{R}}) x$. Naj bo sedaj za vsak $i = 0, 1, 2, \dots, N$ L'_i množica vseh tistih seznamov preferenc $\mathcal{R} = (R_1, R_2, \dots, R_N)$, za katere velja:

- $x \tilde{R}_j y$ za vse $j > i$,
- $y \tilde{R}_j x$ za vse $j \leq i$.

Podobno kot zgoraj, naj bo k' tisto najmanjše naravno število, da za vsak $\mathcal{R} \in L'_k$ velja $y f(\tilde{\mathcal{R}}) x$. Ker je $L_k \subset L'_k$, je seveda $k \leq k'$. Ker pa sta množici zožitev seznamov L_k in L'_k na $\{x, y\}$ enaki, je zaradi načela neodvisnosti od irelevantnih alternativ $y f(\tilde{\mathcal{R}}) x$ za vsak seznam \mathcal{R} iz L'_k in zato $k = k'$. Takemu k rečemo pivot y nad x .

Naj bo sedaj M množica vseh seznamov referenc, za katere velja:

- $y \tilde{R}_j z$ in $z \tilde{R}_j x$ za vsak $j < k$,
- $x \tilde{R}_j y$ in $y \tilde{R}_j z$ za $j \geq k$.

Ker je $M \subset L'_{k-1}$, za vsak seznam preferenc $\mathcal{R} \in M$ velja $x f(\tilde{\mathcal{R}}) y$ (tukaj uporabimo tudi načelo neodvisnosti od irelevantnih alternativ) in $y f(\tilde{\mathcal{R}}) z$ (to sledi direktno iz Paretovega načela). Označimo z M' vse tiste sezname referenc $\mathcal{R} = (R_1, R_2, \dots, R_n)$, za katere velja:

- $y \tilde{R}_k x$ in $x \tilde{R}_k z$,
- $y \tilde{R}_j x$ in $z \tilde{R}_j x$ za $j < k$,
- $x \tilde{R}_j y$ in $x \tilde{R}_j z$ za $j > k$.

Za vsak seznam referenc \mathcal{R} iz M' velja $y f(\tilde{\mathcal{R}}) x$, saj je $M' \subset L'_k$. Ker pa so sezname iz M in M' identični, če jih zožimo na $\{x, z\}$, mora veljati $x f(\tilde{\mathcal{R}}) z$

za vsak seznam iz M' , saj velja za vsak seznam iz M . Za vsak seznam \mathcal{R} iz M' torej velja $y f(\tilde{\mathcal{R}}) z$. Naj bo sedaj $\mathcal{R} = (R_1, R_2, \dots, R_N)$ poljuben seznam, za katerega velja $y \tilde{R}_k z$. Zožitev \mathcal{R} na $\{y, z\}$ je enaka zožitvi seznamov iz M' na $\{y, z\}$. Zato je po načelu neodvisnosti od irelevantnih alternativ $y f(\tilde{\mathcal{R}}) z$. Torej smo pokazali

$$y \tilde{R}_k z \implies y f(\tilde{\mathcal{R}}) z.$$

Zato rečemo, da je k diktator za y nad z .

Videli smo, da je pivot za y nad x tudi diktator za y nad z . Naj bo sedaj $k_{y/z}$ pivot za y nad z . Seveda je $k_{y/z} \leq k_{y/x}$, ker je $k_{y/x}$ diktator za y nad z . Podobno je $k_{z/y} \geq k_{y/x}$. Skupaj dobimo

$$k_{y/z} \leq k_{y/x} \leq k_{z/y}.$$

Če v vsej zgodbi zamenjamo vlogo y in z , dobimo neenakost

$$k_{z/y} \leq k_{z/x} \leq k_{y/z}.$$

Zato je

$$k_{z/y} = k_{y/z} = k_{y/x}.$$

Torej so pivotni elementi za vse pare, in posledično diktatorji za vse pare, enaki. Torej dobimo, da je tak element k diktator. \square

PRIMER. V primeru, ko razvrščamo le dve izbiri in je N vsaj 2, večinsko glasovanje zadosti vsem zahtevam poštenosti: Paretovo načelo, neodvisnost od relevantnih alternativ in odsotnost diktatorja.

PRIMER. Poglejmo si primer lepotnega tekmovanja psov. V finalno tekmovanje so prišli štirje psi, označimo jih z x, y, z, w . Vsak od treh sodnikov razvrsti pse po lepoti in svojo razvrstitev napiše na list. Vsako prvo mesto šteje 5 točk, vsako drugo mesto 4 točke, tretje 2 točki in četrto 1 točko. Vrstni red psov se nato določi tako, da se sešteje točke. Takemu glasovanju rečemo *uteženo glasovanje*.

Poglejmo, da je tak način glasovanja ne diktatorski. Če bi bil na primer prvi sodnik diktator, bi bil končni rezultat enak njegovi razvrstitvi. Recimo, da je to x, y, z, w . Vendar, če oba ostala sodnika podata na primer razvrstitev y, x, z, w , bo to tudi končna razvrstitev. Prav tako tak način zadošča Paretovem načelu: če vsi sodniki na primer izbirajo x rangirajo nad izbiro y , bo v končnem seštevku x dobil več točk kot y . Težava nastane pri neodvisnosti od irelevantnih alternativ. Recimo, da je resnično mnenje prvih dveh sodnikov razvrstitev x, z, w, y , tretjega sodnika pa razvrstitev z, w, x, y . V tem primeru bo končni rezultat z (13 točk), x (12 točk), w (8 točk), y (3 točke). Če pa prva dva sodnika spremenita svoj vrstni red v x, y, z, w , tretji sodnik pa svojo odločitev ohrani, bo končni rezultat x (12 točk), y in z (9 točk), w (6 točk). Ne glede na to, da imajo vsi sodniki pri obeh glasovanjih enak odnos o relativni razvrstitvi x in z , se njun vrstni red na koncu spremeni, kar pomeni, da končen vrstni red ni neodvisen od irelevantnih alternativ. V praksi to pomeni, da prva dva sodnika zato, da bi x -u zagotovila zmago nad z , špekulativno postavita y na drugo mesto.

Uteženo glasovanje ima še vsaj eno dodatno težavo. Lahko se zdi, da so točke, ki jih damo za $1, 2, \dots$ mesto postavljene arbitrarno. Vsekakor lahko

pri popolnoma enakem glasovanju dosežemo drugačen končni rezultat, če točkovanje spremenimo.

Gibbard-Satterthwaitov izrek

V tem razdelku bomo podoben izrek kot zgoraj pokazali tudi za funkcijo socialne izbire. Aksiomi "pravičnosti" bodo podobni.

DEFINICIJA (Paretovo načelo). Funkcija socialne izbire $f : L(X)^N \rightarrow X$ zadošča Paretovem načelu, če za vsak $x \in X$ in za vsak seznam preferenc $\mathcal{R} \in L(X)^N$, za katerega velja $x \tilde{\mathcal{R}} y$ za vsak $y \in X \setminus \{x\}$, sledi $f(\mathcal{R}) = x$.

Paretovo načelo nam v tem primeru pove, da naj funkcija socialne izbire nujno izbere x , če so vsi posamezniki izbiro x postavili strogo pred vse ostale izbire.

Načelo o neodvisnosti od irelevantnih alternativ v tem primeru zamenjamo z naslednjim

DEFINICIJA (Monotonost). Funkcija socialne izbire $f : L(X)^N \rightarrow X$ je monotona, če za vsak par seznamov preferenc $\mathcal{R} = (R_1, R_2, \dots, R_N) \in L(X)^N$ in $\mathcal{R}^* = (R_1^*, R_2^*, \dots, R_N^*) \in L(X)^N$ velja, da če je $f(\mathcal{R}) = x$ in če za vsak $i = 1, 2, \dots, N$ in za vsak $y \in X \setminus \{x\}$ velja $x \tilde{R}_i y \implies x \tilde{R}_i^* y$, potem je tudi $f(\mathcal{R}^*) = x$.

DEFINICIJA (Obstoj diktatorja). Funkcija socialne izbire $f : L(X)^N \rightarrow L(X)$ je diktatorska, če obstaja tak indeks i , da za vsak seznam preferenc $\mathcal{R} = (R_1, R_2, \dots, R_N) \in L(X)^N$ velja $f(\mathcal{R}) = x \iff \forall y \in X \setminus \{x\}, x \tilde{R}_i y$.

IZREK (Muller-Satterthwaitov izrek). *Naj bo $N \geq 2$ in moč množice X vsaj 3. Potem je vsaka funkcija socialne izbire $f : L(X)^N \rightarrow X$, za katero velja Paretovo načelo in monotonost, diktatorska.*

PRIMER. Poslanci državnega zbora imajo na mizi tri predloge novega zakona: predlog x , predlog y in predlog z , kar je tudi zaporedje, po katerem so predlogi prišli pred poslance. V takih primerih imajo protokol, da najprej večinsko glasujejo med prvo prispelima predlogoma, nato pa še večinsko glasujejo med zmagovalnim predlogom iz prvega kroga glasovanja in tretjim predlogom. Pri vsakem večinskem glasovanju neodločen izid razrešijo tako, da za zmagovalni predlog razglasijo tistega, ki je prej prispel v parlament. Izbran je tisti predlog, ki zmaga v drugem krogu glasovanja. Poglejmo si nekaj težav takšne funkcije socialne izbire.

Poglejmo najprej, da tak način izbire zagotovo ni diktatorski. Recimo, da je i -ti poslanec diktator. Potem bi vedno moral biti izbran tisti predlog, ki ga on postavi na vrh svoje preferenčne relacije. Vendar takoj vidimo, da če vsi ostali poslanci ta predlog postavijo na zadnje mesto, predlog ne more biti izbran. Poglejmo sedaj, da tak izbor tudi zadošča Paretovem načelu. Če vsi poslanci nek predlog postavijo na prvo mesto svojih preferenčnih relacij, potem bo ta predlog izbran pri večinskem glasovanju s katerim koli drugim predlogom.

Da tak način glasovanja ne zadošča monotonosti, lahko vidimo z naslednjima dvema preferenčnima seznamoma. Recimo, da 30 poslancev predloge rangira po vrstnem redu x, y, z , 30 poslancev po vrstnem redu z, x, y in 30

poslancev po vrstnem redu y, z, x . Pri večinskem glasovanju med predlogoma x in y bo zmagal x , ki ga bo potem predlog z porazil v drugem krogu. Če pa prvih 30 poslancev svoj vrstni red iz x, y, z spremenijo v y, x, z , ostalih 60 poslancev pa ohrani svoje preference, bo med x in y izglasovan y , ki bo tudi zmagovalac v drugem krogu. Tak način torej krši princip monotonosti: v prvem primeru je bil izbran z , in ker pri drugem seznamu preferenc velja, da je z pred x oziroma y vedno, ko je pred x oziroma y v prvem seznamu preferenc, bi tudi v drugem primeru moral biti izbran z .

Seveda ni potrebno, da gledamo monotonost, da opazimo pomanjkljivosti takega načina glasovanja. Poslanci bi se verjetno tudi pritožili, da je vrstni red glasovanj o predlogih precej arbitraren. Poglejmo še enkrat seznam preferenc, ko 30 poslancev predloge rangira po vrstnem redu x, y, z , 30 poslancev po vrstnem redu z, x, y in 30 poslancev po vrstnem redu y, z, x . Če bi namesto med x in y najprej glasovali med x in z , bi na koncu zmagal y , le pa bi najprej glasovali med y in z pa bi zmagal x . Dvotiren (ali več tiren, v primeru več predlogov) zaporeden večinski način glasovanja se zdi pošten takrat, ko bi bil nek predlog izbran ne glede na to, kakšen vrstni red glasovanj izberemo. To se seveda zgodi takrat, ki bi tak predlog v večinskem glasovanju premagal kateri koli drug predlog. Takemu zmagovalcu rečemo *Condorcetov zmagovalac*.

PRIMER. Na volitvah nastopajo trije kandidati: Gore, Nader in Bush. Dva sta kandidata levega političnega pola (Gore in Nader), eden pa je kandidat desnega političnega pola (Bush). Vsak volivec lahko glasuje le za enega kandidata, zmagaja pa tisti, ki dobi največ glasov (če je rezultat slučajno izenačen, izberejo kandidata po abecedi). Volivec glasuje za tistega kandidata, ki ga postavi na vrh svoje preferenčne relacije (ker pri preferenčni relaciji dovolimo, da so nekatere izbire za volivca enakovredne, lahko volivec dejansko na vrh postavi enega kandidata, za ostala dva pa mu je vseeno).

Na enak način kot v prejšnjem primeru vidimo, da tak način glasovanja zopet ni diktatorski. Podobno zadosti Paretovem načelu, saj bo zagotovo izbran nek kandidat, če bo na vrhu vseh preferenčnih relacij. Zopet pa se zatakne pri monotonosti. Recimo, da je 52% volivcev levo usmerjeni, in ne bi nikoli glasovali za kandidata iz desnega političnega pola, 48% volivcev pa je desno usmerjenih, in ne bi nikoli glasovali za kandidata iz levega političnega pola. Tako bo 52% levih volivcev Busha postavilo na zadnje mesto, 48% desno usmerjenih volivcev pa na prvo mesto. V primeru, da 5% levih volivcev na prvo mesto postave Naderja, bo volitve dobil Bush, če pa vsi (ali vsaj več kot 48% levih volivcev na prvo mesto postavijo Gora, bo na volitvah zmagal Gore. Ker v tem drugem primeru še vedno velja, da je Bush višje od katerega koli drugega kandidata na vseh tistih preferenčnih seznamih, na katerih je bil višje tudi v prvem primeru, bi moral po monotonosti zopet zmagati Bush.

Ta primer nam dobro ilustrira, kaj je težava, če funkcija socialne izbire ne zadošča monotonosti. V tem primeru se posameznikom včasih splača, da ne sporočajo svoje dejanske preference. To je seveda zelo pogost problem skoraj vseh volitev. Poglejmo si to malo bolj natančno.

Naj bo $\mathcal{R} = (R_1, R_2, \dots, R_N) \in L(X)^N$ seznam preferenc in $i \in \{1, 2, \dots, N\}$. Naj bo nadalje R'_i neka (nova) preferenca i -tega volivca. Z (R'_i, \mathcal{R}_{-i}) označimo

seznam prefenc, kjer smo v seznamu \mathcal{R} zgolj zamenjali preferenco i -tega volivca, R_i , z njegovo novo preferenco, R'_i .

DEFINICIJA. Funkcija socialne izbire $f : L(X)^N \rightarrow L(X)$ ne dopušča strategije, če za vsakega volivca i in vsak seznam preferenc \mathcal{R} ter vsako preferenco R'_i velja

$$f(R'_i, \mathcal{R}_{-i}) \neq f(\mathcal{R}) \implies f(\mathcal{R}) \tilde{R}_i f(R'_i, \mathcal{R}_{-i}).$$

IZREK (Renyjev izrek). Če je funkcija socialne izbire $f : L(X)^N \rightarrow X$ surjektivna in ne dopušča strategije, potem zadošča tako Paretoemu načelu kot tudi monotonosti.

Kot posledico dobimo Gibbard-Satterthwaitov izrek

IZREK (Gibbard-Satterthwaitov izrek). Če je $f : L(X)^N \rightarrow X$ surjektivna in ne dopušča strategije, potem je diktatorska.