UNIVERSITAT DE BARCELONA

Facultat de Matemàtiques
i Informàtica

GRAU DE MATEMÀTIQUES

Treball final de grau

JOCS ESTOCÀSTICS

Autor: Daniel Reverter Condal

Director: Dr. Carles Rovira Escofet
Realitzat a: Departament de Matemàtiques i Informàtica

Barcelona, 21 de juny de 2020

Abstract

Stochastic games are a generalization of Markov's decision processes and normal games. We assume that the game has a finite number of states and players. These play stage by stage and the actions they take determine the transition probabilities from one state to another. The number of stages played is not bounded in length. In this work we will study this type of game and all the prerequisites for a person with no experience in the field of game theory to understand each of the results obtained.

Resum

Els jocs estocàstics són una generalització dels processos de decisió de Markov i dels jocs en forma normal. Suposem que el joc té un nombre finit d'estats i de jugadors. Aquests juguen etapa per etapa i les accions que prenen determinen la probabilitat de transició d'un estat a un altre. El nombre d'etapes jugades no està acotat. En aquest treball estudiarem aquest tipus de joc i tots els requisits previs perquè una persona sense experiència en l'àmbit de la teoria de jocs entengui cada un dels resultats obtinguts.

2010 Mathematics Subject Classification. 91A05, 91A06, 91A10, 91A15

Agraïments

M'agradaria agraïr al meu tutor Dr. Carles Rovira per l'ajut i el suport en la realització d'aquest treball, especialment durant aquest període de temps difícil per a tothom. També vull agraïr a la meva família, especialment als meus pares, i als meus amics per estar al meu costat al llarg de tot aquest camí.

Índex

Introducció

Motivació

Al llarg de la carrera de Matemàtiques, el que més m'ha motivat a aprendre el temari que se m'ha ensenyat ha estat la cerca d'aplicacions útils d'aquest temari a la vida quotidiana. És cert que al cap dels anys també he trobat el gust a les matemàtiques que s'han desenvolupat més per la bellesa de crear-les que per la seva utilitat immediata, però mai tant com les matemàtiques aplicades. És per això que vaig triar fer la menció en Economia, i entre les assignatures que vaig cursar la que més em va cridar l'atenció, amb diferència, és la de teoria de jocs. Pot ser que em cridés l'atenció pel contingut tan interessant i fàcilment aplicable a situacions mundanes, o per l'excel·lent didàctica i dedicació del professor que impartia l'assignatura, en Mikel Álvarez, o, segurament, per la combinació de totes dues. Quan es va acabar el quadrimestre i, amb ell, l'assignatura, em vaig quedar amb ganes de continuar aprenent més sobre el tema. És per això que he decidit fer el treball de fi de grau sobre teoria de jocs, aprofundint sobre un apartat d'aquesta teoria que requereix, d'entrada, una bona base matemàtica: els jocs estocàstics.

Objectius

L'objectiu principal del treball és l'estudi general dels jocs estocàstics de dos jugadors, posant èmfasi en els resultats sobre les estratègies estacionàries, amb l'objectiu final de demostrar l'existència d'equilibris en jocs estocàstics de suma general. Gran part del treball, però, consisteix en construïr les bases de la teoria de joc des de zero, amb l'objectiu que una persona que no tingui experiència en l'àmbit de la teoria de jocs pugui seguir la major part dels raonaments que s'hi presenten.

Estructura de la Memòria

El treball està separat en tres capítols.

En el primer capítol s'introdueixen els conceptes més elementals de la teoria de jocs, com els dels agents o les seves motivacions. A continuació s'estableix la base de la teoria d'utilitat amb els axiomes de von-Neumann i la seva conclusió immediata: el teorema de von-Neumann, per a demostrar l'existència de funcions d'utilitat que a partir del capítol 2 suposarem que els agents intentaran maximitzar.

El segon capítol tracta principalment sobre el tipus de joc més elemental que s'estudia en la teoria de jocs: els jocs en forma normal (o jocs normals). Ens és imprescindible entendre aquest tipus de jocs ja que sense ells no podriem arribar a definir jocs estocàstics. A part d'analitzar aquests jocs es dona una demostració del teorema de Nash a partir del teorema del punt fix de Brower.

A l'últim capítol s'introdueix el concepte de joc estocàstic de dos jugadors, es defineixen els conceptes clau relacionats i es donen resultats sobre els jocs estocàstics de suma zero i de suma general jugats amb estratègies estacionàries.

1 Teoria de la utilitat

L'objectiu d'aquest primer capítol és introduïr els conceptes i supòsits sobre els quals es construeix la base de la teoria de jocs.

El primer concepte és els dels agents. La teoria de jocs és l'estudi matemàtic de les interaccions entre agents independents i autointeressats. Aquests agents són els jugadors, i podrien ser, per exemple, persones o col·lectius. Que els agents siguin autointeressats no vol dir necessàriament que només es preocupin per ells mateixos, vol dir que tenen certa preferència per uns possibles estats del món respecte uns altres i actuen per a portar el món cap a aquests estats. Per a modelitzar aquest comportament s'ha desenvolupat el que es diu la teoria de la utilitat, que permet quantificar aquesta preferència d'un agent respecte un estat del món o un altre. A continuació introduiré de manera matemàtica aquest concepte d'utilitat i arribaré, gràcies al teorema de von Neumann-Morgenster, a definir el concepte de funció d'utilitat.

Per a fer-ho definirem la relació de preferència $\succeq$. Sigui O un conjunt finit de resultats (d'una possible acció). Intuïtivament aquests resultats són el canvi en l'estat del món que un agent pot aconseguir mitjançant accions. Per a qualsevol parella $o_1, o_2 \in O$, notarem amb $o_1 \succeq o_2$ si l'agent prefereix el resultat o_1 al resultat o_2. Notarem amb $o_1 \sim o_2$ si $o_1 \succeq o_2$ i $o_2 \succeq o_1$, i direm que o_1 i o_2 són equivalents. Notarem amb $o_1 \succ o_2$ si $o_1 \succeq o_2$ i no són equivalents, i direm que l'agent prefereix o_1 a o_2 de manera forta.

Una *loteria* és una distribució de probabilitat sobre possibles resultats escrita

$$[p_1 : o_1, p_2 : o_2, ..., p_k : o_k]$$

amb $o_i \in O$, $p_i > 0$ $\forall i \in \{1, 2, ..., k\} \subset \mathbb{N}$, i $\sum_{i=1}^{k} p_i = 1$. p_i denota la probabilitat que la loteria esculli el resultat situat a la posició i.

Sigui Γ el conjunt format per totes les loteries. Ampliarem el concepte de preferència $\succeq$ als elements de Γ, fent que puguem considerar loteries sobre resultats com si fossin resultats de per si, i per tant podem considerar també loteries de loteries. Per exemple,

$$l = [p_1 : o_1, p_2 : o_2, p_3 : [p_4 : o_1, p_5 : o_2]], \text{ amb } p_1 + p_2 + p_3 = p_4 + p_5 = 1$$

és una loteria. En general, anomenem $P_l(o_i)$ a la probabilitat que o_i sigui escollit per la loteria l. En l'exemple anterior, $P_l(o_1) = p_1 + p_3 \cdot p_4$.

1.1 Axiomes de von Neumann-Morgenstern

Basarem la teoria de la utilitat en uns axiomes que s'adeqüen a la idea intuïtiva que tenim sobre les preferències. El primer axioma ens diu que, donats dos resultats, un agent sempre preferirà un, o l'altre, o li seràn indiferents.

Axioma 1.1. $\forall o_i, o_j \in O$ $o_i \succeq o_j$ o bé $o_j \succeq o_i$

El segon axioma aporta transitivitat a la relació $\succeq$. Si no l'agafèssim com a axioma es donarien casos que no consideraríem racionals.

Axioma 1.2. Si $o_i \geq o_j$ i $o_j \geq o_k$ aleshores $o_i \geq o_k$

El tercer axioma serveix per a relacionar loteries a partir de relacions conegudes entre resultats, i diu que si a un agent li són indiferents dos resultats, també li són indiferents dues loteries que només varien en quin d'aquests resultats li és ofert.

Axioma 1.3. Si $o_1 \sim o_2$, aleshores per a tota sèrie de resultats $o_3, \ldots, o_k$ i tota sèrie de probabilitats $p, p_3, \ldots, p_k$ que compleixin $p + \sum_{i=3}^{k} p_i = 1$, tenim

$$[p : o_1, p_3 : o_3, \ldots, p_k : o_k] \sim [p : o_2, p_3 : o_3, \ldots, p_k : o_k]$$

El quart axioma ens diu que a un agent li són indiferents les loteries que indueixen les mateixes probabilitats sobre resultats, sigui quina sigui la forma de les loteries presentades.

Axioma 1.4. Si $\forall o_i \in O$, $P_{l_1}(o_i) = P_{l_2}(o_i)$ aleshores $l_1 \sim l_2$

El cinqué axioma ens diu que si un agent prefereix un resultat que un altre, també preferirà una loteria que assigna més probabilitat al resultat que prefereix.

Axioma 1.5. Si $o_1 \succ o_2$ i $p > q$ aleshores $[p : o_1, 1-p : o_2] \succ [q : o_1, 1-q : o_2]$

Lema 1.6. *Sigui $\geq$ una relació de preferència. Si $\geq$ satisfà els axiomes 1.1, 1.2, 1.3, 1.4 i 1.5, i si $o_1 \succ o_2$ i $o_2 \succ o_3$, aleshores existeix $p \in [0,1]$ tal que*

- $\forall p' < p$, $o_2 \succ [p' : o_1; (1-p') : o_3]$
- $\forall p'' > p$, $[p'' : o_1; (1-p'') : o_3] \succ o_2$

Demostració. Considerem la següent loteria:

$$l(p) := [p : o_1, (1-p) : o_3]$$

Triem un $p_- \in [0,1]$ tal que $o_2 \succ l(p_-)$. Aquest p_- existeix ja que $o_2 \succ o_3$. Per exemple, $p_- = 0$ satisfà la condició, ja que

$$l(p_-) = l(0) = [0 : o_1, 1 : o_3] \sim o_3$$

per l'axioma 1.4. A més,

$$l(p_-) \succ l(p') \;\; \forall p' \in [0, p_-)$$

per l'axioma 1.5, i

$$\forall p' \leq p_-, \;\; o_2 \succ l(p')$$

per l'axioma 1.2. Considerem un $p_+ \in \Gamma$ tal que $l(p_+) \succ o_2$. De la mateixa manera, per 1.5, tenim que

$$l(p') \succ l(p_+) \;\; \forall p' \in (p_+, 1]$$

i per 1.2 tenim que

$$\forall p' \geq p_+, \;\; l(p') \succ o_2$$

De moment hem relacionat $l(p)$ i o_2 per a tots els valors de p excepte aquells entre p_- i p_+. Considerem $p^* = (p_- + p_+)/2$. Per l'axioma 1.1, $o_2 \succ l(p^*)$, o bé $l(p^*) \succ o_2$ o bé $o_2 \sim l(p^*)$.

Considerem primer el cas $o_2 \sim l(p^*)$. No pot passar que existeixi un $p' \neq p^*$ amb $o_2 \sim l(p')$, ja que si fos així, per l'axioma 1.2 tindriem que $l(p^*) \sim l(p')$, i com que $o_1 \succ o_3$, això contradiria l'axioma 1.5. Per tant, $\forall p' \neq p^*$, tenim que $o_2 \succ l(p')$, o bé $l(p') \succ o_2$. Si hi hagués un $p' > p^*$ amb $o_2 \succ l(p')$, aleshores $\forall p'' < p', o_2 \succ l(p'')$, entrant en contradicció amb $o_2 \sim l(p^*)$. Per tant $\forall p' > p^*$ tenim que $l(p') \succ o_2$ De la mateixa manera, no pot haver-hi un $p' < p^*$ amb $l(p') \succ o_2$, i per tant $\forall p' < p^*$ tenim que $o_2 \succ l(p')$. Per tant, en aquest cas tindriem relacionats $l(p)$ i o_2 per a tots els valors de p tal i com ens diu el lema.

Considerem ara el cas $o_2 \succ l(p^*)$. Tenim que $\forall p' \leq p^*$, $o_2 \succ l(p')$. Podem per tant redefinir $p_- := p^*$. Si considerem el cas $l(p^*) \succ o_2$, tindrem que $l(p') \succ o_2$ $\forall p' \geq p^*$, i en aquest cas redefiniriem $p_+ := p^*$. En qualsevol dels dos casos, la longitud de l'interval (p_-, p_+) és reduïda a la meitat. Podem iterar el mateix argument i, o bé eventualment trobarem un p^* amb $o_2 \sim l(p^*)$, o bé en el límit, p_- s'aproparà a un p per l'esquerra i p_+ s'aproparà al mateix p per la dreta. □

El que aquest lema no ens indica és la relació de preferència entre o_2 i $[p : o_1; (1-p) : o_3]$. L'últim axioma ho fa, i ens diu que si hi ha 3 resultats no equivalents dos a dos, existeix una loteria entre el resultat més preferit i el resultat menys preferit que per a l'agent és equivalent al resultat del mig.

Axioma 1.7. Si $o_1 \succ o_2$ i $o_2 \succ o_3$, aleshores $\exists p \in [0,1]$ tal que $o_2 \sim [p : o_1; (1-p) : o_3]$

Exemple 1.8. Imaginem que un agent ha guanyat un premi de 100 euros (que notem amb o_2), i se li ofereix tornar a jugar i tenir la possibilitat de doblar els seus diners amb una probabilitat de p (notat amb o_1) o perdre'ls tots (notat amb o_3). A algunes persones els hi seria indiferent guanyar el premi o jugar-se-la si la probabilitat de doblar fos de $\frac{1}{2}$. Algunes persones averses al risc necessitarien que la probabilitat de doblar els diners fos més elevada, i algunes amants del risc jugarien la loteria fins i tot amb una probabilitat menor. Sigui com sigui, aquest axioma ens garantitza que per a la persona que estigui jugant, existeix una probabilitat (única per el lema 1.6) amb la qual li és indiferent quedar-se els 100 euros guanyats o jugar a "doble o res".

1.2 Funcions d'utilitat

Gràcies als anteriors axiomes podem arribar a demostrar un teorema d'extrema importància en la teoria de jocs, el teorema de von Neumann-Morgenstern. Aquest teorema ens garanteix l'existència de funcions reals 1-dimensionals que els agents preferiran maximitzar. Aquestes funcions les anomenarem *funcions d'utilitat*. Aquest resultat ens permet suposar que es pot quantificar el concepte abstracte de preferències per treballar la teoria de jocs d'una manera quantitativa, i arribar a resultats molt més útils a la pràctica que si treballèssim de manera qualitativa.

Teorema 1.9. **(Teorema de von Neumann-Morgenstern)** *Si una relació de preferència $\succeq$ compleix els axiomes 1.1, 1.2, 1.3, 1.4, 1.5 i 1.7, aleshores existeix una funció $u : \Gamma \to [0,1]$ tal que:*

1. $u(o_1) \geq u(o_2) \Leftrightarrow o_1 \succeq o_2$

2. $u([p_1 : o_1, \ldots, p_k : o_k]) = \sum_{i=1}^{k} p_i u(o_i)$

Demostració. Primer considerem el cas que a l'agent li siguin indiferents tots els possibles resultats. En aquest cas $u(o_i) = 0\ \forall o_i \in O$, i $u(l) = 0\ \forall l \in \Gamma$. I per tant 1 i 2 són trivialment certs.

Ara considerem el cas que a l'agent no li siguin indiferents tots els possibles resultats. Aleshores existeix un subconjunt d'O d'un o més resultats més preferits per l'agent, que anomenarem $\overline{O}$. Formalment, $\exists \overline{O} \subset O$ tal que

$$\forall \overline{o} \in \overline{O} \quad \forall o \in O \setminus \overline{O}, \quad \overline{o} \succ o$$

i a més

$$\forall \overline{o}_1, \overline{o}_2 \in \overline{O}, \quad \overline{o}_1 \sim \overline{o}_2$$

De la mateixa manera existeix un subconjunt d'O, que anomenarem $\underline{O}$ d'un o més elements menys preferits per l'agent. Agafem $\overline{o}$ un element de $\overline{O}$, i $\underline{o}$ un element de $\underline{O}$. Per a qualsevol $o_i \in O$, definim $u(o_i) := p_i$, on p_i és el nombre per al qual es compleix $o_i \sim [p_i : \overline{o}, (1-p_i) : \underline{o}]$. Si $o_i \in O \setminus (\overline{O} \cup \underline{O})$, per 1.7 aquest nombre existeix, i per 1.6 és únic. Si $o_i \in \overline{O}$, aquest nombre és 1, i si $o_i \in \underline{O}$, aquest nombre és 0.

Primer demostrarem que $\forall o_1, o_2 \in O$, $u(o_1) \geq u(o_2) \Leftrightarrow o_1 \succeq o_2$. Per la tria de $u(o_1)$, tenim que

$$o_1 \sim [u(o_1) : \overline{o}; (1 - u(o_1)) : \underline{o}]$$

Notem amb l_1 a aquesta loteria. De la mateixa manera

$$o_2 \sim [u(o_2) : \overline{o}; (1 - u(o_2)) : \underline{o}]$$

i notem amb l_2 a aquesta loteria.

Veiem primer que $u(o_1) \geq u(o_2) \Rightarrow o_1 \succeq o_2$. Si $u(o_1) > u(o_2)$, com que $\overline{o} \succ \underline{o}$ tenim, per l'axioma 1.5 que $l_1 \succ l_2$. Per tant tenim que $o_1 \sim l_1 \succ l_2 \sim o_2$, i per els axiomes 1.1 i 1.2 que $o_1 \succ o_2$. Si $u(o_1) = u(o_2)$, aleshores l_1 i l_2 són la mateixa loteria, i per tant tindriem $o_1 \sim l_1 \equiv l_2 \sim o_2$, i per l'axioma 1.2 tenim que $o_1 \sim o_2$.

Anem a veure ara que $o_1 \succeq o_2 \Rightarrow u(o_1) \geq u(o_2)$. Per contrarecíproc, és suficient veure $u(o_1) \ngeq u(o_2) \Rightarrow o_1 \nsucceq o_2$, que per l'axioma 1.1 és el mateix que veure que $u(o_2) > u(o_1) \Rightarrow o_2 \succ o_1$, i això ja ho hem vist abans (amb o_1 i o_2 intercanviats). Per tant ja hem demostrat 1.

Ara anem a demostrar que $u([p_1 : o_1, ..., p_k : o_k]) = \sum_{i=1}^{k} p_i u(o_i)$. Sigui

$$u^* = u([p_1 : o_1, ..., p_k : o_k])$$

Per la construcció de $u(o_i)$ tenim que

$$o_i \sim [u(o_i) : \overline{o}, (1 - u(o_i)) : \underline{o}]$$

Per l'axioma 1.3 podem substituïr cada o_i de la definició de u^* per la loteria $[u(o_i) : \overline{o}, (1.u(o_i))) : \underline{o}]$, arribant a

$$u^* = u([p_1 : [u(o_1) : \overline{o}, (1 - u(o_1)) : \underline{o}], ..., p_k : [u(o_k) : \overline{o}, (1 - u(o_k)) : \underline{o}]])$$

Fent servir l'axioma 1.4 arribem a

$$u^* = u([\sum_{i=1}^{k} p_i u(o_i)) : \overline{o}, 1 - (\sum_{i=1}^{k} p_i u(o_i)) : \underline{o}])$$

Per la definició de u,

$$u^* = \sum_{i=1}^{k} p_i u(o_i)$$

que és el que volíem. □

A la pràctica voldrem tenir funcions d'utilitat amb valors reals no necessàriament continguts en $[0,1]$. Sigui $u(o)$ una funció d'utilitat descrita com en el teorema. Aleshores per a qualssevol constants reals a>0 i b, tenim que $u^*(o) := au(o) + b$ és també una funció d'utilitat, en el sentit que compleix les propietats 1. i 2. del teorema. Per tant podem tenir funcions d'utilitat amb imatges contingudes a qualsevol interval $[M_-, M_+]$, amb $M_-, M_+ \in \mathbb{R}$.

2 Jocs en forma normal

A partir d'ara podrem suposar que els agents, en cas de trobar-se davant d'una decisió, triaran si és possible realitzar l'acció que els porti al resultat que maximitzi la seva funció d'utilitat. Però imaginem-nos que tenim a la vegada a dos o més agents que intenten maximitzar les seves funcions d'utilitat mitjançant accions, i que aquestes accions no només afecten la seva funció d'utilitat, sinó la de tots els altres agents. Els jocs en forma normal són la manera més comuna de representar i analitzar aquest tipus de situacions. En aquest capítol els definirem, juntament amb altres conceptes bàsics de teoria de jocs per a tenir els fonaments necessaris per a poder demostrar el teorema de Nash i, eventualment, desenvolupar la teoria sobre els jocs estocàstics.

2.1 Introducció als jocs en forma normal

Definició 2.1. *Un joc en forma normal és una 3-tupla (N,A,u), on*

- *N és un conjunt finit d'n jugadors*
- *$A = A_1 \times \ldots \times A_n$, on $A_i = \{a_{i_1}, \ldots, a_{i_{m_i}}\}$, $m_i \in \mathbb{N}$, és el conjunt finit d'accions disponibles per al jugador i. Cada vector $a = (a_1, \ldots, a_n) \in A$ es diu un perfil d'accions*
- *$u = (u_1, \ldots, u_n)$, on $u_i : A \to \mathbb{R}$ és una funció d'utilitat per al jugador i*

Cal notar que ara el domini de les funcions d'utilitat és un conjunt d'accions A, i no un conjunt de resultats O. Estem assumint implícitament que $A = O$, ja que les accions conjuntes dels jugadors canvien l'estat del món, i els possibles resultats estan determinats per aquestes accions.

Els jocs en forma normal (també anomenats jocs normals) ens serveixen per a modelitzar situacions en les que un nombre finit d'agents actuen de forma simultània, i el conjunt d'aquestes accions canvia l'estat del món i, conseqüentment, aporten una utilitat a cada un dels jugadors que podem quantificar gràcies a les funcions d'utilitat que, dins dels supòsits del capítol 1, sabem que existeixen.

Durant aquest capítol suposarem que els agents tenen tota l'informació del joc que estan jugant, és a dir que cada agent sap quines possibles accions pot jugar cada un dels jugadors, i totes les funcions d'utilitat. Es poden estudiar jocs en els que els jugadors no tenen accés a tota l'informació, anomenats *jocs d'informació imperfecta*, però no forma part dels objectius d'aquest treball.

A part de modelitzar una gran amplitud de situacions aplicables a la realitat (des de tàctiques de mercat d'empreses multinacionals fins a debats familiars, passant per ecosistemes, congestions de trànsit, i altres), formen la base per a estudiar jocs més complexes. A més, alguns dels jocs més complexes es poden separar, i fins i tot a vegades reduïr a jocs en forma normal.

Exemple 2.2. Per fer-nos una idea de què poden representar els jocs normals, podem mirar el següent exemple bàsic. Suposem que en un racó d'un bosc hi viuen dos animals, i en aquest racó hi ha un sol taronger. Els dos animals tenen altres fonts de menjar, però prefereixen les taronges a qualsevol altra font. Els animals poden adoptar dues actituds diferents: poden cooperar per compartir les taronges (ho notarem amb C) o poden actuar de manera agressiva per intentar quedar-se-les totes (ho notarem amb A). Si tots dos cooperen, compartiran les taronges i direm que ambdós tindran una utilitat de $\frac{1}{2}$. Si un dels dos es comporta de manera agressiva i l'altre no, el primer es quedarà totes les taronges i direm que tindrà una utilitat d'1, mentres que el segon tindrà una utilitat de 0. Si tots dos es comporten de manera agressiva, lluitaran i en sortiran malparats, aconseguint tots dos una utilitat negativa arbitràriament gran, que en aquest exemple direm que és -100. Aquest és un joc en forma normal on:

- $N = \{1, 2\}$
- $A = A_1 \times A_2$, on $A_1 = (C, A)$ i $A_2 = (C, A)$
- $u = (u_1, u_2)$, amb $u_1(C,C) = \frac{1}{2}$, $u_1(C,A) = 0$, $u_1(A,C) = 1$, $u_1(A,A) = -100$, $u_2(C,C) = \frac{1}{2}$, $u_2(C,A) = 1$, $u_2(A,C) = 0$, $u_2(A,A) = -100$

2.2 Representació matricial

Una manera simple i molt usada de representar aquesta informació és mitjançant matrius. En jocs de dos jugadors tindríem una matriu on cada fila denota una acció del perfil d'accions del jugador 1 i cada columna denota una acció del perfil d'accions del jugador 2, i dins de la matriu cada element és el vector d'utilitat u que s'obtindria jugant les accions corresponents. Així, en el nostre exemple, l'element de la primera fila i la primera columna seria el vector $u(C,C) = (u_1(C,C), u_2(C,C)) = (\frac{1}{2}, \frac{1}{2})$, i tota la matriu seria:

	C	A
C	$(\frac{1}{2}, \frac{1}{2})$	(0,1)
A	(1,0)	(-100,-100)

En cas que cada jugador tingués més accions disponibles, simplement la matriu tindria més files o columnes. Si hi hagués un tercer jugador, cada acció d'aquest jugador la representaríem amb una matriu, i els vectors d'utilitat tindrien 3 elements. Si hi hagués 4 jugadors, hauríem de representar cada acció del jugador 4 amb un grup de matrius (per exemple, si hi haguessin 4 jugadors amb 3 accions cada un ho representaríem 3 grups de 3 matrius de 3 files i 3 columnes, en total 9 matrius 3x3). El nombre de matrius necessàries per a representar un joc augmenta de forma exponencial en relació al nombre de jugadors. Per tant aquest mètode de representar jocs resulta útil en jocs amb un nombre petit de jugadors, tot i que és possible fer-lo servir per a n jugadors amb n arbitràriament gran.

Per exemple, un joc de 3 jugadors on cada un tingués dues accions disponibles (A_1 o A_2 pel jugador 1, B_1 o B_2 pel jugador 2 i C_1 o C_2 pel jugador 3) es podria representar amb dues matrius de la següent manera:

C_1

	B_1	B_2
A_1	(0,2,0)	(1,2,-4)
A_2	(3,2,1)	(-1,5,-4)

C_2

	B_1	B_2
A_1	(2,3,5)	(3,2,1)
A_2	(2,1,3)	(-1,3,0)

On el tercer nombre de cada vector d'utilitat és ara la utilitat que rebria el jugador 3, i la tria d'acció del jugador 3 determinaria si estariem jugant la primera matriu (si triés jugar C_1) o la segona matriu (si triés jugar C_2).

2.3 Conceptes bàsics

Definició 2.3. *Un joc en forma normal és un joc de suma constant si existeix una constant c tal que per a cada perfil d'accions* $a \in A$, $\sum_{i=1}^{n} u_i(a) = c$. *Si* $c = 0$, *diem que el joc és un joc de suma zero.*

Exemple 2.4. Si juguem al joc de pedra, paper, tisores i suposem que guanyar aporta una utilitat d'1, empatar aporta una utilitat de $\frac{1}{2}$ i perdre aporta una utilitat de 0, el joc seria un joc normal de suma constant amb $c = 1$. En forma de matriu,

	Pe	Pa	T
Pe	$(\frac{1}{2},\frac{1}{2})$	(0,1)	(1,0)
Pa	(1,0)	$(\frac{1}{2},\frac{1}{2})$	(0,1)
T	(0,1)	(1,0)	$(\frac{1}{2},\frac{1}{2})$

Quan un jugador ha de triar quina acció jugar pot, simplement, escollir-ne una i jugar-la. En aquest cas direm que el jugador segueix una *estratègia pura*. Una altra estratègia que podria seguir seria assignar una probabilitat a cada una de les seves accions disponibles, triant una d'aquestes accions de manera aleatòria. En aquest cas direm que el jugador segueix una *estratègia mixta*. Més formalment,

Definició 2.5. *Sigui (N,A,u) un joc en forma normal. El conjunt d'estratègies mixtes per al jugador i és* $S_i = \Pi(A_i)$*, on* $\Pi(A_i)$ *és el conjunt de totes les distribucions de probabilitat sobre* A_i*. El conjunt de perfils d'estratègies mixtes és* $S = S_1 \times ... \times S_n$

Notarem amb s_i l'estratègia mixta usada pel jugador i. Notarem amb $s_i(a_{i_k})$ la probabilitat que l'acció a_{i_k} sigui jugada en l'estratègia mixta s_i, i en cas que $s_i(a_{i_k}) = 1$, a vegades escriurem a_{i_k} en lloc de s_i.

Definició 2.6. *El suport d'una estratègia mixta* s_i *per al jugador i és el subconjunt d'estratègies pures* $\{a_{i_k} | s_i(a_{i_k}) > 0\}$.

Necessitem estendre la definició de funció d'utilitat per tal que abarqui perfils d'estratègies mixtes. Això ho fem amb el concepte d'utilitat esperada.

Definició 2.7. *Sigui (N,A,u) un joc en forma normal. La utilitat esperada* u_i *per al jugador i per al perfil d'estratègies mixtes* $s = (s_1, ..., s_n)$ *es defineix per*

$$u_i(s) = \sum_{a \in A} u_i(a) \prod_{j=1}^{n} s_j(a_j)$$

Cal notar l'abús de notació en fer servir u_i tant per utilitat com per a utilitat esperada.

Exemple 2.8. Suposem que tenim el següent joc de dos jugadors:

	R	S
C	(1,2)	(-1,2)
D	(3,-3)	(-6,4)

i suposem que els jugadors juguen una estratègia mixta s on el primer jugador juga C amb una probabilitat de $\frac{1}{2}$ i D amb una probabilitat de $\frac{1}{2}$, i el segon jugador juga R amb una probabilitat de $\frac{1}{4}$ i juga S amb una probabilitat de $\frac{3}{4}$. Aleshores, la utilitat esperada per al jugador 1 seria:

$$\begin{aligned} u_1(s) &= u_1(C,R)\cdot s_1(C)\cdot s_2(R) + u_1(C,S)\cdot s_1(C)\cdot s_2(S) + \\ &+ u_1(D,R)\cdot s_1(D)\cdot s_2(R) + u_1(D,S)\cdot s_1(D)\cdot s_2(S) = \\ &= 1\cdot\frac{1}{2}\cdot\frac{1}{4} + (-1)\cdot\frac{1}{2}\cdot\frac{3}{4} + 3\cdot\frac{1}{2}\cdot\frac{1}{4} + (-6)\cdot\frac{1}{2}\cdot\frac{3}{4} = \frac{-17}{8} \end{aligned}$$

2.4 Estratègies òptimes

El següent pas per a acabar de construïr la base de la teoria de jocs és parlar d'estratègies òptimes, és a dir, estratègies que maximitzin la utilitat esperada del jugador. En jocs d'N jugadors, amb $N > 1$, cada jugador intenta maximitzar la seva utilitat, fet que fa difícil per a un jugador concret trobar una estratègia que la maximitzi, ja que aquesta estratègia depén de la tria d'estratègies de tots els altres jugadors. El concepte més usat per a estudiar aquestes situacions és l'equilibri de Nash.

Notem que, si un agent sabés quines accions jugarien els altres agents, el problema de trobar una estratègia que maximitzés la seva utilitat esperada seria molt més senzill. Donat un perfil d'estratègies $s = (s_1, ..., s_n)$, definim $s_{-i} := (s_1, ..., s_{i-1}, s_{i+1}, ..., s_n)$, i podem escriure s com $s = (s_i, s_{-i})$.

Definició 2.9. *Una millor resposta del jugador i al perfil d'estratègies s_{-i} és una estratègia mixta $s_i^* \in S_i$ tal que $u_i(s_i^*, s_{-i}) \geq u_i(s_i, s_{-i})$ per a tot $s_i \in S_i$*

Cal notar que una millor resposta no és sempre única, ara ho veurem.

Exemple 2.10. Anem a mirar el joc de tres jugadors que hem vist anteriorment, a la segona part de l'exemple 2.2, que té com a matriu:

C_1

	B_1	B_2
A_1	(0,2,0)	(1,2,-4)
A_2	(3,2,1)	(-1,5,-4)

C_2

	B_1	B_2
A_1	(2,3,5)	(3,2,1)
A_2	(2,1,3)	(-1,3,0)

En el cas que el jugador 1 jugués A_1 (amb probabilitat 1) i el jugador 2 jugués B_1 (amb probabilitat 1), la millor resposta del jugador 3 a aquest perfil d'estratègies seria jugar C_2 amb probabilitat 1, i aquesta millor resposta seria única, ja que $u_3(A_1, B_1, C_1) = 0$ i $u_3(A_1, B_1, C_2) = 5$, i qualsevol estratègia mixta on el jugador 3 li dongui una probabilitat $p > 0$ a l'acció C_1 li donarà una utilitat esperada més petita que la que guanyaria jugant l'estratègia pura C_2.

De la mateixa manera, la millor resposta del jugador 2 al perfil d'estratègies on el jugador 1 juga A_1 de forma pura (i.e. amb probabilitat 1) i el jugador 3 juga C_2 de forma pura és jugar B_1 de forma pura, ja que així la seva utilitat seria 3 enlloc de 2 (si jugués B_2 de forma pura) o d'un nombre estrictament entre 3 i 2 (si jugués una estratègia mixta).

Ara anem a veure quina és la millor estratègia del jugador 1 al perfil d'estratègies on el jugador 2 juga B_1 amb probabilitat 1 i el jugador 3 juga C_2 amb probabilitat 1. Com veiem, tant si juga A_1 com si juga A_2 com si juga una estratègia mixta on jugués A_1 amb probabilitat $p > 0$ i A_2 amb probabilitat $1 - p$, tindria en tots els casos una utilitat esperada de 2, i per tant qualsevol d'aquestes estratègies seria una millor resposta al perfil d'estratègies dels jugadors 2 i 3.

Més generalment, excepte en el cas en que només hi ha una millor resposta que és jugar una estratègia pura, el nombre de millors respostes és infinit, ja que quan el suport d'una millor resposta inclou dues o més accions, al jugador li són indiferents aquestes accions, ja que si no ho fossin, preferiria reduir la probabilitat d'una o més d'aquestes a 0. Per tant qualsevol estratègia mixta que contingui aquest suport també és una millor resposta.

2.5 Equilibri de Nash

Definició 2.11. *Un perfil d'estratègies $s = (s_1, ..., s_n)$ és un equilibri de Nash si per a tot i, s_i és una millor resposta a s_{-i}.*

Direm que un equilibri de Nash $s = (s_1, ..., s_n)$ és fort si per a tot jugador i, i per a tota estratègia $s_i' \neq s_i$ tenim que $u_i(s_i, s_{-i}) > u_i(s_i', s_{-i})$, i direm que és dèbil si $u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i})$ i no és fort.

En l'exemple anterior hem demostrat que el perfil d'estratègies $s = (A_1, B_1, C_2)$ és un equilibri de Nash, més concretament un equilibri de Nash dèbil (ja que el jugador 2 podria canviar la seva estratègia guanyant exactament la mateixa utilitat que si no la canviés).

Intuitivament, els equilibris de Nash són molt estables, ja que cap jugador tindria cap incentiu per a canviar la seva estratègia si sabés que tots els jugadors triaran estratègies que portin a un equilibri de Nash. La pregunta que ens fem ara és sobre l'existència d'equilibris de Nash. El famós teorema de Nash ens diu que tots els jocs en forma normal tenen com a mínim un equilibri de Nash. Per a demostrar-lo necessitem uns resultats previs.

Definició 2.12. *Donats vectors $v_0, \ldots, v_n$ d'un espai euclidi i escalars $\lambda_0, \ldots, \lambda_n \geq 0$, tals que $\sum_{i=0}^{n} \lambda_i = 1$, anomenarem combinació convexa de $v_0, \ldots, v_n$ al vector $\sum_{i=0}^{n} \lambda_i v_i$*

Definició 2.13. *Direm que un conjunt de vectors $\{v_0, \ldots, v_n\}$ d'un espai euclidi són afinament independents si*

$$\sum_{i=0}^{n} \lambda_i v_i = 0 \ i \ \sum_{i=0}^{n} \lambda_i = 0 \Rightarrow \lambda_0 = \ldots = \lambda_n = 0$$

Definició 2.14. *Un n-símplex $v_0 \ldots v_n$ és el conjunt de combinacions convexes dels vectors afinament independents $v_0, \ldots, v_n$, és a dir,*

$$\left\{ \sum_{i=0}^{n} \lambda_i v_i : \forall i \in \{0, \ldots, n\}, \lambda_i \geq 0 \ i \ \sum_{i=0}^{n} \lambda_i = 1 \right\}$$

Notem que el conjunt de perfils d'estratègies mixtes $S = S_1 \times \ldots \times S_n$ és un producte cartesià de símplexs, ja que cada estratègia mixta $s_i \in S_i$ es pot entendre com un punt d'un símplex.

Teorema 2.15. **(Teorema del punt fix de Brower per a productes cartesians de símplexs)** *Sigui $S = S_1 \times \ldots \times S_n$ un producte cartesià de símplexs, i sigui $f : S \to S$ una funció contínua. Aleshores f té un funt fix, és a dir, $\exists s \in S$ tal que $f(s) = s$*

La demostració d'aquest teorema no és difícil però és llarga i no forma part dels objectius d'aquest treball. Es basa en el teorema del punt fix de Brower per a símplexs i en el fet que els productes cartesians de símplexs són homeomorfs a símplexs. Podeu trobar la demostració, per exemple, al capítol 6 del llibre *Fixed Point Theorems with Applications to Economics and Game Theory*[1].

Una altra manera de demostrar el teorema 2.15 és usant el teorema del punt fix de Kakutani, que es pot trobar a l'article *A generalization of Brouwer's fixed point theorem*, del volum número 8 del *Duke Mathematical Journal*, a les pàgines 457-459.

Amb aquest teorema ja podem, per fi, demostrar el teorema de Nash, que ens garanteix l'existència d'equilibris de Nash en jocs normals.

Teorema 2.16. **(Teorema de Nash)** *Cada joc amb un nombre finit de jugadors i de perfils d'accions té un equilibri de Nash.*

Demostració. Donat un perfil d'estratègies $s \in S$, per a tot $i \in N$ i per a tot $a_{i_k} \in A_i$, definim

$$\varphi_{i,a_{i_k}}(s) = \max\{0, u_i(a_{i_k}, s_{-i}) - u_i(s)\}$$

Definim també la funció $f : S \to S$, fent servir la notació $f(s) = s'$, com:

$$s'_i(a_{i_k}) = \frac{s_i(a_{i_k}) + \varphi_{i,a_{i_k}}(s)}{\sum_{a_{i_j} \in A_i} s_i(a_{i_j}) + \varphi_{i,a_{i_j}}(s)} = \frac{s_i(a_{i_k}) + \varphi_{i,a_{i_k}}(s)}{1 + \sum_{a_{i_j} \in A_i} \varphi_{i,a_{i_j}}(s)}$$

La funció f és contínua ja que es pot comprovar que cada $\varphi_{i,a_{i_k}}$ és contínua, i el denominador és positiu ja que, de fet, és més gran o igual que 1. Pel teorema del punt fix de Brower per a productes cartesians de símplexs, f té un punt fix. Anem a veure que els punts fixos de f són equilibris de Nash.

Primer de tot, si s és un equilibri de Nash, aleshores totes les φ són 0, i per tant s és un punt fix de f.

Anem a veure el recíproc. Considerem s un punt fix de f. Triem $i \in \{1, .., n\}$. Ha d'existir com a mínim una acció a_{i_l} que pertany al suport de s_i tal que $u_i(a_{i_l}, s_{-i}) \leq u_i(s)$, ja que si s_i és millor resposta a s_{-i}, aleshores qualsevol a_{i_l} del suport de s_i complirà que $u_i(a_{i_l}, s_{-i}) = u_i(s)$, i si s_i no és millor resposta a s_{-i}, aleshores agafem un a_{i_l} del suport de s_i que no estigui al suport d'una millor resposta a s_{-i}, i tindrem que $u_i(a_{i_l}, s_{-i}) < u_i(s)$. Per la definició de φ, $\varphi_{i,a_{i_l}}(s) = 0$. Com que s és un punt fix de f, $s'_i(a_{i_l}) = s_i(a_{i_l})$. Considerem la segona igualtat de la definició de $s'_i(a_{i_l})$. El numerador es simplifica a $s_i(a_{i_l})$, i és estrictament positiu ja que a_{i_l} és del suport de i. Per tant el denominador ha de ser 1, i per tant per a cada $j \in \{1, ..., m_i\}$, $\varphi_{i,a_{i_j}}(s) = 0$. Per la definició de φ, això només pot passar quan el jugador i no pot millorar la seva utilitat esperada canviant a una estratègia pura. Repetint el mateix raonament per a cada $i \in \{1, .., n\}$, tenim que s és un equilibri de Nash.$\square$

2.6 Estratègies maxmin i minmax

Els equilibris de Nash són uns subconjunts dels perfils d'estratègia que compleixen unes propietats que els fa interessants, però això no treu que puguem buscar altres perfils d'estratègia amb altres propietats que els facin interessants a la seva manera. A continuació descriuré alguns d'aquests subconjunts àmpliament usats en teoria de jocs.

Definició 2.17. *Una estratègia maxmin del jugador i és* $\arg\max_{s_i} \min_{s_{-i}} u_i(s_i, s_{-i})$*, i el valor maxmin del jugador i és* $\max_{s_i} \min_{s_{-i}} u_i(s_i, s_{-i})$*, on* $\arg\max$ *denota l'argument del màxim, que són els punts del domini d'una funció en els que el valor d'aquesta funció assoleix el seu màxim.*

Exemple 2.18. Agafem el joc de l'exemple 2.8, que tè com a matriu:

	R	S
C	(1,2)	(-1,2)
D	(3,-3)	(-6,4)

Si el jugador 1 tria jugar l'acció C, obtindrà una utilitat d'1 o de -1 depenent del que trii el jugador 2. Si tria l'acció D, obtindrà una utilitat de 3 o de -6. Per tant l'estratègia maxmin pel jugador 1 és jugar C, i el seu valor maxmin és -1.

L'estratègia maxmin és una estratègia (no necessàriament única) que maximitza la utilitat del jugador i en el pitjor dels casos, és a dir en la situació en que tots els altres jugadors juguen les estratègies que fan que el jugador i guanyi el mínim possible. El valor maxmin és el valor mínim d'utilitat que guanyaria el jugador i fent servir una estratègia maxmin. En qualsevol cas, si els altres jugadors juguessin estratègies que no fessin que el jugador i guanyés el mínim possible, si el jugador fa servir una estratègia maxmin aconseguiria una utilitat de, com a mínim, el valor maxmin.

Un concepte molt lligat a l'estratègia maxmin és l'estratègia minmax. En un joc de dos jugadors, l'estratègia minmax per al jugador i és aquella estratègia que minimitza l'utilitat màxima que pot aconseguir l'altre jugador, que li direm jugador $-i$. Més formalment,

Definició 2.19. *En un joc de dos jugadors, una estratègia minmax per al jugador i contra el jugador $-i$ és* $\arg\min_{s_i} \max_{s_{-i}} u_{-i}(s_i, s_{-i})$, *i el valor minmax del jugador $-i$ és* $\min_{s_i} \max_{s_{-i}} u_{-i}(s_i, s_{-i})$.

Les estratègies minmax per a jocs normals de suma zero també s'anomenen *estratègies òptimes*.

En jocs d'n jugadors, definir l'estratègia minmax del jugador i contra el jugador j és més complicat, ja que el jugador i no pot assegurar per si sol que el jugador j aconsegueixi la mínima utilitat. Seria possible si suposèssim que tots els jugadors diferents de j es posessin d'acord per a aconseguir aquest propòsit. Més formalment,

Definició 2.20. *En un joc d'n jugadors, una estratègia minmax del jugador i contra el jugador $j \neq i$ és el component i del perfil d'estratègies mixtes s_{-j} de* $\arg\min_{s_{-j}} \max_{s_j} u_j(s_j, s_{-j})$, *i el valor maxmin del jugador j és* $\min_{s_{-j}} \max_{s_j} u_j(s_j, s_{-j})$.

En jocs de 2 jugadors, donat un jugador, el seu valor minmax i el seu valor maxmin coincideixen, mentre que en jocs d'n jugadors el valor maxmin és sempre més petit o igual que el valor minmax. Direm que un perfil d'estratègies $s = (s_1, s_2, ..., s_n)$ és un perfil d'estratègies maxmin si s_i és una estratègia maxmin per a cada jugador $i = 1, 2, ..., n$, i en jocs de dos jugadors definim recíprocament els perfils d'estratègies minmax.

Teorema 2.21. **(Teorema Minmax)** *En un joc de dos jugadors finit i de suma zero, a cada equilibri de Nash, cada jugador rep una utilitat igual al seu valor minmax i al seu valor maxmin.*

Demostració. Pel teorema de Nash, ha d'existir com a mínim un equilibri de Nash. Sigui (s'_i, s'_{-i}) un equilibri de Nash arbitrari. Denotem amb v_i la utilitat del jugador i en aquest equilibri, amb $\overline{v}_i$ el valor maxmin i amb $\underline{v}_i$ el valor minmax d'aquest jugador. Anem a veure que $\overline{v}_i = v_i$.

Primer de tot, no és possible que $\overline{v}_i > v_i$, ja que si fos cert, aleshores en aquest equilibri de Nash el jugador i podria canviar la seva estratègia a una que no fos s'_i per aconseguir una major utilitat, i per tant (s'_i, s'_{-i}) no seria un equilibri de Nash.

Anem a veure ara que $\overline{v}_i \geq v_i$. Si ho veiem, com que sabem que $\overline{v}_i \ngtr v_i$, tindrem que $\overline{v}_i = v_i$. Per definició d'equilibri de Nash, cada jugador juga una millor resposta al que juga l'altre jugador. Per tant,

$$v_{-i} = \max_{s_{-i}} u_{-i}(s'_i, s_{-i})$$

Equivalentment, podem dir que, donada l'estratègia del jugador i, el jugador $-i$ minimitza l'oposat de la seva utilitat (ja que maximitza aquesta utilitat), és a dir,

$$-v_{-i} = \min_{s_{-i}} -u_{-i}(s'_i, s_{-i})$$

Com que el joc és de suma zero i de dos jugadors, tenim que $v_i = -v_{-i}$ i que $u_i = u_{-i}$. Per tant,

$$v_i = \min_{s_{-i}} u_i(s'_i, s_{-i})$$

Hem definit $\overline{v}_i$ com $\max_{s_i} \min_{s_{-i}} u_i(s_i, s_{-i})$. Per la definició de màxim, tenim que

$$\max_{s_i} \min_{s_{-i}} u_i(s_i, s_{-i}) \geq \min_{s_{-i}} u_i(s'_i, s_{-i})$$

Per tant tenim que $\overline{v}_i \geq v_i$, i per tant $\overline{v}_i = v_i$. De la mateixa manera, fent servir la definició de mínim i girant algunes desigualtats demostraríem que $\underline{v}_i = v_i$. □

Aquest teorema ens diu, en particular, que en jocs finits de dos jugadors de suma zero, donat un jugador, el seu valor maxmin i el seu valor minmax és el mateix. Per conveni, el valor maxmin del jugador 1 es diu el *valor del joc*.

2.7 ϵ-equilibris

Els últims subconjunts de perfils d'estratègia que mencionaré en aquest capítol són els ϵ-equilibris de Nash. Formalment,

Definició 2.22. *Fixat un $\epsilon > 0$, direm que un perfil d'estratègies $s = (s_1, ..., s_n)$ és un ϵ-equilibri de Nash si per a tot jugador i, i per a tota estratègia $s'_i \neq s_i$ tenim que $u_i(s_i, s_{-i}) \geq u_i(s'_i, s_{-i}) - \epsilon$.*

Aquests equilibris són útils per tenir presents els casos en què un jugador no té incentius per canviar la seva estratègia si la diferència d'utilitats en canviar-la és massa petita (més petita que ϵ). Un exemple senzill seria un joc d'un jugador on aquest està caminant pel carrer i es troba una moneda d'1 cèntim al terra i pot triar entre agafar-la o seguir caminant. Podem dir que agafar-la li otorga una utilitat d'1 i seguir caminant li otorga una utilitat de 0, però l'utilitat d'1 és prou petita com perquè no es molesti en canviar la seva estratègia de seguir caminant. En aquest cas, seguir caminant és un ϵ- equilibri de Nash per qualsevol $\epsilon \geq 1$.

En jocs normals tenim garantida l'existència d'ϵ-equilibris de Nash, ja que qualsevol equilibri de Nash és també un ϵ-equilibri de Nash per a qualsevol $\epsilon > 0$.

3 Jocs estocàstics

Ara que tenim definides les bases de la teoria de jocs podem començar a analitzar l'objecte principal d'aquest treball, els jocs estocàstics. Resumidament, els jocs estocàstics són un tipus de jocs en què els jugadors comencen jugant un joc de forma normal com els que hem vist al capítol anterior: juguen les accions que han triat i reben la utilitat que els hi toqui depenent de les accions triades per tots els jugadors, però la diferència està en que seguidament jugaran a un altre joc en forma normal, jugaràn les seves accions, rebran utilitat i després jugaran a un altre joc de forma normal, repetint aquest procès indefinidament. A més, exceptuant el primer joc en forma normal, no està determinat quin joc normal es jugarà després, ja que la tria d'aquest següent joc és aleatòria, i la funció de probabilitat de transició depén del joc actual i de les accions jugades.

Es poden estudiar jocs estocàstics d'n jugadors, amb $n \in \mathbb{N}$, però primer ens centrarem en jocs estocàstics de dos jugadors.

3.1 Introducció als jocs estocàstics

Definició 3.1. *Un joc estocàstic de dos jugadors és una 6-tupla* $(S, \{A^s, s \in S\}, \{B^s, s \in S\}, u^1, u^2, p)$*, on*

- $S = \{1, 2, ..., z\}, z \in \mathbb{N}$, *és un conjunt finit de jocs normals, també anomenats estats*
- $A_s = \{1, 2, ..., m_s\}, m_s \in \mathbb{N}$, *és el conjunt finit d'accions disponibles per al jugador 1 al joc* $s \in S$
- $B_s = \{1, 2, n_s\}, n_s \in \mathbb{N}$, *és el conjunt finit d'accions disponibles per al jugador 2 al joc* $s \in S$
- $u^1 : \bigcup_{s \in S} (s \times A_s \times B_s) \to \mathbb{R}$, *és la funció d'utilitat pel jugador 1*
- $u^2 : \bigcup_{s \in S} (s \times A_s \times B_s) \to \mathbb{R}$, *és la funció d'utilitat pel jugador 2*
- $p : \bigcup_{s \in S} (s \times A_s \times B_s) \to \Delta^z$ *és la funció de probabilitat de transició, amb*

  $$p(s, i, j) = (p(1|s, i, j), p(2|s, i, j), ..., p(z|s, i, j))$$

  i amb

  $$\Delta^z = \{(v_1, ..., v_z) \in \mathbb{R}^z | v_i \geq 0 \;\; \forall i \in (1, ..., z), \sum_{i=1}^{z} v_i = 1\}$$

Un lector que estigui familiaritzat en el concepte de cadenes de Markov i, més generalment, en els processos de decisió de Markov pot haver notat que els jocs estocàstics són una generalització d'aquests. En efecte, un procès de decisió de Markov és equivalent a un joc estocàstic amb un únic jugador. És per això que l'estudi de tots dos temes està estretament lligat.

Els jocs estocàstics es poden representar de manera similar als jocs normals, amb matrius. Simplement ara s'ha d'afegir la probabilitat de transició a cada possible acció comuna. Això normalment es fa dividint cada casella de la matriu en dues caselles, i en una casella escrivim el que teniem en els jocs normals, és a dir el vector d'utilitats, i en l'altra casella escrivim el vector de probabilitats de transició (o si volem ser rigorosos podriem dir que estem escrivint dues matrius en l'espai que ocupa una, una matriu per els vectors d'utilitat i una altra pels vectors de probabilitat de transició).

Exemple 3.2. Suposem que tenim els següents dos jocs normals, que els anomenarem J_1 i J_2.
J_1 té com a matriu:

	d	e	f
a	(3,5)	(2,-4)	$(\frac{1}{2},0)$
b	(2,3)	(-3,4)	(0,0)
c	$(0,\frac{5}{2})$	(0,1)	(100,100)

J_2 té com a matriu:

	i	j
g	(-50,-50)	(10,-20)
h	(-20,20)	(-200,-200)

Suposem ara que si estem jugant al joc J_1, les probabilitats de passar al joc J_2 són 0 en tots els casos excepte en el que es juga (c,f), en el qual hi ha una probabilitat de $\frac{1}{4}$ de passar al joc J_2 (i per tant una probabilitat de $\frac{3}{4}$ de continuar jugant J_1). A més, si s'està jugant el joc J_2, hi ha una probabilitat de $\frac{1}{2}$ de canviar al joc J_1 si es juga (g,i), una probabilitat d'1 de canviar al joc J_1 si es juga (h,j) i una probabilitat de 0 de canviar al joc J_1 en la resta de casos. Podem expressar aquesta informació en les mateixes matrius, escrivint ara els vectors d'utilitat a dalt a l'esquerra de la casella on estaven i els vectors de probabilitat de transició a baix a la dreta de la casella que els hi correspon, subdividint cada casella en 2 amb una línia diagonal per evitar confusions. En el nostre exemple, les matrius quedarien així:

J_1:

	d	e	f
a	(3,5) (1,0)	(2,-4) (1,0)	$(\frac{1}{2},0)$ (1,0)
b	(2,3) (1,0)	(-3,4) (1,0)	(0,0) (1,0)
c	$(0,\frac{5}{2})$ (1,0)	(0,1) (1,0)	(100,100) $(\frac{3}{4},\frac{1}{4})$

J_2:

	i	j
g	(-50,-50) $(\frac{1}{2},\frac{1}{2})$	(10,-20) (0,1)
h	(-20,20) (0,1)	(-200,-200) (1,0)

Definició 3.3. *Donat un joc estocàstic de 2 jugadors i $t \in \mathbb{N}$, suposem que s'han jugat t etapes del joc estocàstic i que s'han jugat els jocs $(s_1, s_2, ..., s_t)$ en aquest ordre. Suposem també que el jugador 1 ha jugat $(i_1, i_2, ..., i_t)$ (en aquest ordre) i el jugador 2 ha jugat $(j_1, j_2, ..., j_t)$ (en aquest ordre). Definim la història del joc estocàstic en el moment t com $h_t = (s_1, i_1, j_1, s_2, i_2, j_2, ..., s_t, i_t, j_t)$.*

3.2 Conjunts d'estratègies

En els jocs estocàstics definim estratègia mixta de la mateixa manera que en els jocs normals. Durant cada etapa del joc els jugadors saben la història del joc estocàstic però no saben quines estratègies mixtes ha fet servir l'altre jugador.

A l'hora de decidir quina acció jugar, els jugadors poden seguir qualsevol tipus d'estratègia. Per exemple, poden tenir present la història del joc o poden ignorar-la i prendre la seva decisió només en funció de l'estat en el què està el joc en el moment de prendre-la.

Dins de totes les estratègies possibles, podem considerar 3 subconjunts que són de gran importància a l'hora d'estudiar els jocs estocàstics. El primer d'ells és el d'estratègies estacionàries. Diem que un jugador segueix una estratègia estacionària si tè fixada una estratègia mixta per a cada estat, i la fa servir sempre que el joc estocàstic estigui en aquell estat, independentment de la història. Formalment,

Definició 3.4. *Una estratègia estacionària pel jugador 1 és un element* $x \in X = \prod_{s=1}^{z} \Delta^{m_s}$*, on cada element de* Δ^{m_s} *indica la probabilitat otorgada per l'estratègia mixta del jugador 1 de jugar la seva respectiva acció d'*A_s*. Recíprocament, una estratègia estacionària pel jugador 2 és un element* $y \in Y = \prod_{s=1}^{z} \Delta^{n_s}$*.*

És a dir, per al jugador 1 una estratègia estacionària x és un vector $x = (x_1, \ldots, x_z)$ on x_s denota l'estratègia que jugarà el jugador 1 en l'estat s, i aquesta estratègia és un vector $x_s = (x_{s_1}, \ldots, x_{s_{m_s}})$, on cada x_{s_i} denota la probabilitat del jugador 1 de jugar l'acció $i \in A_s$ en el joc s. Per tant $\sum_{i \in S} x_{s_i} = 1$ per a cada $s \in S$.

El següent subconjunt d'estratègies és una generalització de les estratègies estacionàries anomenades estratègies de Markov, en les que el jugador que les segueix té fixada una estratègia mixta per a cada estat del joc, però ara diferenciant en quin moment $t \in \mathbb{N}$ del joc estem, és a dir que l'estratègia mixta que tè fixada per al joc s_1 al moment $t = 1$ pot diferir de la que té fixada per al joc s_1 al moment $t = 2$. Formalment,

Definició 3.5. *Una estratègia de Markov per al jugador 1 és una funció* $f : \mathbb{N} \to X$*. Recíprocament, una estratègia de Markov per al jugador 2 és una funció* $g : \mathbb{N} \to Y$*. El conjunt d'estratègies de Markov del jugador 1 es denota amb F i la del jugador 2 es denota amb G.*

Els últims conjunts d'estratègies que mencionarem són una generalització de les estratègies de Markov, les estratègies conductuals. En aquest tipus d'estratègies el jugador té present tota la història del joc per decidir quina estratègia mixta jugar en cada moment. Formalment,

Definició 3.6. *Sigui* $H_n = \{(s_1, i_1, j_1, s_2, i_2, j_2, \ldots, s_n, i_n, j_n) | s_k \in S, i_k \in A_{s_k}, j_k \in B_{s_k}\}$ *el conjunt de possibles històries fins el moment* $n \in \mathbb{N}$*. Una estratègia conductual per al jugador 1 és una funció* $\pi : \bigcup_{n=1}^{\infty} H_n \to X$*. Recíprocament, una estratègia conductual per al jugador 2 és una funció* $\sigma : \bigcup_{n=1}^{\infty} H_n \to Y$*. El conjunt d'estratègies conductuals per al jugador 1 es denota amb* Π*, i per al jugador 2 es denota amb* Σ*.*

Si volem restringir-nos al cas en que els jugadors no fan servir estratègies mixtes (només fan servir estratègies pures), notarem el conjunt d'estratègies estacionaries amb X^p i Y^p, el conjunt d'estratègies de Markov amb F^p i G^p i el conjunt d'estratègies conductuals amb Π^p i Σ^p.

3.3 Criteris d'utilitat

El fet que els jocs estocàstics siguin, en principi, de durada infinita (dic en principi ja que després veurem algun mecanisme per poder fer que siguin finits de forma artificial) ens porta al problema d'avaluar la utilitat guanyada total. No podem avaluar-la com la suma de les utilitats de cada moment del joc, ja que aquesta és infinita (i en la immensa majoria dels casos, no convergent). Necessitem, però, alguna manera d'avaluar aquesta utilitat total perquè els jugadors la puguin comparar de manera quantitativa per a poder decidir quina estratègia jugar. És per això que es fan servir principalment 3 criteris per a avaluar aquesta utilitat total, i es separa l'estudi dels jocs estocàstics en 3 blocs diferents, depenent de quin dels criteris fan servir els jugadors.
Abans de definir aquest criteris definirem primer la utilitat esperada per a un jugador al moment t.

Definició 3.7. *Sigui $(\pi, \delta) \in \Pi \times \Delta$ l'estratègia que segueixen els jugadors, i sigui $s \in S$ l'estat inicial del joc. Definim $U^k(n)$ com la variable aleatòria que representa la utilitat pel jugador k a l'etapa $n \in \mathbb{N}$. Definim la utilitat esperada a l'etapa n per al jugador k com $E_{s,\pi,\delta}[U^k(n)]$, on E denota l'esperança matemàtica, i depèn de s, π i δ.*

La utilitat esperada està ben definida: En la primera etapa del joc, π i δ determinen quina estratègia mixta jugaran els jugadors, i per tant la seva utilitat esperada. També determinen les probabilitats de transició al següent estat. Durant la segona etapa, per a qualsevol nou estat π i δ tornen a determinar quina estratègia mixta jugaran, la utilitat esperada i les següents probabilitats de transició, i seguint així està tot ben definit per a qualsevol $n \in \mathbb{N}$.

Exemple 3.8. Suposem que estem en el joc estocàstic de l'exemple 3.2, que J_1 és el joc inicial i que els jugadors juguen les següents estratègies estacionàries: Si estem en el joc J_1, el primer jugador sempre juga c i el segon jugador sempre juga f, i si estem en el joc J_2, el primer jugador juga sempre g i el segon jugador juga sempre i. Anem a calcular la utilitat esperada per a cada $n \in \mathbb{N}$ del jugador 1:

Està clar que la utilitat esperada del jugador 1 per a la primera etapa és 100, ja que es jugarà (c, f) amb probabilitat 1. La utilitat esperada per a la segona etapa és

$$100 \cdot \frac{3}{4} + (-50) \cdot \frac{1}{4} = \frac{125}{2}$$

ja que hi ha una probabilitat de $\frac{3}{4}$ de continuar jugant el joc J_1 i una probabilitat de $\frac{1}{4}$ d'estar jugant el joc J_2 a l'etapa 2. De fet, la utilitat esperada per a l'etapa n és:

$$100 \cdot p_1(n) + (-50) \cdot p_2(n)$$

on $p_i(n)$ denota la probabilitat d'estar jugant el joc J_i a l'etapa n. Com que $p_2(n) = 1 - p_1(n)$, si podem calcular $p_1(n)$ per a tot $n \in \mathbb{N}$ podrem calcular la utilitat esperada per al jugador 1 per a tot $n \in \mathbb{N}$.

Recordem que la probabilitat de transició de J_2 a J_1 si es juga (g,i) (que en el nostre cas es juga sempre que estem a J_2) és de $\frac{1}{2}$. Per tant, la probabilitat que a l'etapa n juguem J_1 és

$$p_1(n) = \frac{3}{4} p_1(n-1) + \frac{1}{2} p_2(n-1) = \frac{3}{4} p_1(n-1) + \frac{1}{2}(1 - p_1(n-1)) = \frac{p_1(n-1) + 2}{4}$$

Sabent que $p_1(1) = 1$, podem calcular $p_1(n)$ per a $n \geq 2$ de forma recursiva, i tindriem que

$$p_1(n) = \frac{1 + 2\sum_{k=0}^{n-2} 4^k}{4^{n-1}} = \frac{1 + 2(\frac{4^{n-1}-1}{4-1})}{4^{n-1}} = \frac{1 + 2(4^{n-1})}{3 \cdot 4^{n-1}} = \frac{1}{3 \cdot 4^{n-1}} + \frac{2}{3}$$

Per tant, la utilitat esperada del jugador 1 a l'etapa n és

$$100 \cdot \left(\frac{1}{3 \cdot 4^{n-1}} + \frac{2}{3}\right) - 50 \cdot \left(1 - \left(\frac{1}{3 \cdot 4^{n-1}} + \frac{2}{3}\right)\right) = 150 \cdot \left(\frac{1}{3 \cdot 4^{n-1}} + \frac{2}{3}\right) - 50 = \frac{50}{4^{n-1}} + 50$$

Com que les utilitats a cada etapa són iguals per al jugador 2 que per al jugador 1, la utilitat esperada per al jugador 2 és la mateixa.

En aquest exemple ens ha resultat fàcil calcular les utilitats esperades ja que els jugadors estan fent servir estratègies estacionàries i pures. En altres casos la complexitat dels càlculs s'incrementa, però en cap cas és impossible de determinar.

El primer dels tres mètodes que mencionaré per a definir la utilitat total d'un jugador s'anomena el criteri de la utilitat β-descomptada. És segurament el mètode sobre el qual s'ha desenvolupat més teoria de jocs estocàstics, i el primer mètode usat per Shapley quan va definir aquest tipus de jocs al seu article 'Stochastic Games'[5]. A més, moltes de les propietats i resultats de l'estudi de jocs estocàstics sobre els altres dos mètodes deriven de propietats i resultats d'aquest.

Definició 3.9. *Sigui $\beta \in (0,1)$. Sigui $(\pi, \delta) \in \Pi \times \Delta$ l'estratègia que segueixen els jugadors, i sigui $s \in S$ l'estat inicial del joc. La utilitat β-descomptada pel jugador k és*

$$\gamma_\beta^k(s, \pi, \delta) = (1 - \beta) \sum_{n=1}^{\infty} \beta^{n-1} E_{s,\pi,\delta}[U^k(n)]$$

i diem que β és el factor de descompte

En la definició, el factor $(1-\beta)$ el fem servir per a normalitzar les utilitats β- descomptades per a poder-les comparar en un futur amb un altre criteri, el criteri de la utilitat mitjana.

El factor de descompte β es pot interpretar de diverses maneres. Pot representar, per exemple, que els agents prefereixin obtenir la utilitat a curt termini, és a dir, que passa un temps entre cada etapa del joc i els jugadors devaluen la utilitat com més llunyana estigui en el temps, amb un factor de β per cada etapa. Si la utilitat fossin diners, que el factor de descompte fos β representaria un tipus d'interés compost de $\frac{1-\beta}{\beta}$ per a cada etapa, ja que diners amb valor α a la primera etapa creixerien fins a

$$\alpha\left(1 + \frac{1-\beta}{\beta}\right)^{n-1} = \alpha\left(\frac{1}{\beta}\right)^{n-1} = \alpha\beta^{-(n-1)}$$

a l'etapa n, però en estar $n-1$ etapes en el futur, els agents donarien un valor a aquests diners de $\beta^{n-1}\alpha\beta^{-(n-1)} = \alpha$, com volíem. Una altra representació que pot tenir seria que els jugadors no devaluessin la utilitat a mesura que passa el temps però que el joc acabés amb una probabilitat $(1-\beta)$ a cada etapa del joc. L'anàlisi del joc no depén de la interpretació del factor de descompte.

A vegades usarem $\gamma_\beta^k(\pi,\delta) = (\gamma_\beta^k(1,\pi,\delta), \gamma_\beta^k(2,\pi,\delta), ..., \gamma_\beta^k(z,\pi,\delta))$ per notar el vector d'utilitats β-descomptades de tots els estats $s \in S$

És interessant veure que donats k, β, s, π, δ, $\gamma_\beta^k(s,\pi,\delta)$ està acotat. Sigui

$$M = max\{|u^k(s,i,j)| : k \in \{1,2\}, i \in A_s, j \in B_s, s \in S\}$$

Aquest màxim existeix ja que A_s, B_s i S són finits. Tenim que $E_{s,\pi,\delta}[U^k(n)] \in [-M,M]$, i per tant $\gamma_\beta^k(s,\pi,\delta) \in [-M,M]$ per qualsevol $\beta \in (0,1)$, ja que

$$\gamma_\beta^k(s,\pi,\delta) \leq (1-\beta)\sum_{n=1}^{\infty}\beta^{n-1}M = (1-\beta)\frac{1}{1-\beta}M = M$$

$$\gamma_\beta^k(s,\pi,\delta) \geq (1-\beta)\sum_{n=1}^{\infty}\beta^{n-1}(-M) = (1-\beta)\frac{1}{1-\beta}(-M) = -M$$

Agafem el joc estocàstic de l'exemple 3.8 i calculem la utilitat β-descomptada per al jugador 1 (seguint amb els supòsits que la primera etapa és J_1 i que tots dos jugadors segueixen les estratègies descrites a l'exemple). Per $\beta \in (0,1)$, aquesta utilitat β-descomptada és:

$$\gamma_\beta^1(J_1,\pi,\delta) = (1-\beta)\sum_{n=1}^{\infty}\beta^{n-1}\left(\frac{50}{4^{n-1}} + 50\right) = 50\frac{8-5\beta}{4-\beta}$$

Per $\beta \in (0,1)$, aquesta utilitat β-descomptada és un nombre entre 50 i 100, tendeix a 100 quan β tendeix a 0 i tendeix a 50 quan β tendeix a 1. Això no ens ha de resultar estrany, ja que intuïtivament, si β tendeix a 0, la utilitat β-descomptada tendirà, per com està definida, a la utilitat rebuda a la primera etapa del joc, que en aquest exemple és 100. Com més aprop estigui β d'1, més pes tindran les utilitats rebudes en les altres etapes del joc, fins al límit 1, on totes les etapes serien igual de rellevants (i en haver-hi infinites etapes, el que passés a l'infinit passaria a tenir tot el pes, i les primeres etapes deixarien de tenir-ne pel fet de ser finites).

El segon criteri per a definir la utilitat total d'un jugador és el criteri de la utilitat mitjana.

Definició 3.10. *Sigui $(\pi,\delta) \in \Pi \times \Delta$ l'estratègia que segueixen els jugadors, i sigui $s \in S$ l'estat inicial del joc. La utilitat mitjana pel jugador k és*

$$\gamma^k(s,\pi,\delta) = \liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}E_{s,\pi,\delta}[U^k(n)]$$

Com abans, podem usar $\gamma^k(\pi,\delta) = (\gamma^k(1,\pi,\delta), \gamma^k(2,\pi,\delta), ..., \gamma^k(z,\pi,\delta))$ per notar el vector d'utilitats totals de tots els estats $s \in S$

La idea d'aquest criteri és senzilla: la utilitat mitjana és, tal com indica el seu nom, la mitjana de les utilitats esperades, avaluada al límit quan N tendeix a infinit. A la definició agafem $\liminf$ ja que $\lim$ pot no existir. Podríem haver definit aquesta utilitat usant $\limsup$ enlloc de $\liminf$, però per conveni s'usa aquesta. Si π, δ són estratègies estacionàries, el límit existeix ja que el límit inferior i el límit superior coincideixen. Aquest resultat és de teoria de cadenes de Markov i està demostrat al llibre 'Finite Markov Chains'[4], però no el veurem en aquest treball.

La utilitat mitjana rebuda pel jugador 1 en l'exemple 3.8 és la següent:

$$\gamma^1(J_1, \pi, \delta) = \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \frac{50}{4^{n-1}} + 50 = 50$$

que, curiosament, és igual al límit de la utilitat β-descomptada del jugador 1 d'aquest mateix joc quan β tendeix a 1, calculada anteriorment. Això no ens hauria de sorprendre, ja que hem dit que quan β tendeix a 1, totes les etapes del joc són igual de rellevants usant el criteri de la utilitat β-descomptada, com en el cas de la utilitat mitjana.

El criteri de la utilitat β-descomptada fa èmfasi en les utilitats rebudes en el futur proper, i el criteri de la utilitat mitjana fa èmfasi en les utilitats rebudes en el futur llunyà. Un altre exemple d'aquest fenòmen seria el següent: Si tenim un joc estocàstic amb dos estats (J_1 i J_2), i podem passar de l'estat J_1 a l'estat J_2 però no podem passar de l'estat J_2 a l'estat J_1, i suposem que comencem el joc a l'estat J_1, les primeres etapes no es veuran reflectides en la utilitat mitjana, només es veuran reflectides les etapes jugades a l'estat J_2. En canvi, les etapes jugades en l'estat J_1 són les primeres, i per tant es veuen clarament reflectides en la utilitat β-descomptada.

En algunes situacions és més interessant fer servir un criteri que doni la mateixa importància a utilitats rebudes en un futur proper que a les rebudes en un futur llunyà. El tercer (i últim) criteri que veurem és un d'aquests, anomenat el criteri de la utilitat total.

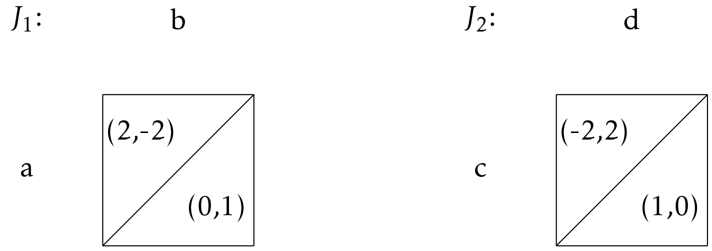
Definició 3.11. *Sigui $(\pi, \delta) \in \Pi \times \Delta$ l'estratègia que segueixen els jugadors, i sigui $s \in S$ l'estat inicial del joc. La utilitat total pel jugador k és*

$$\gamma_T^k(s, \pi, \delta) = \liminf_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} E_{s,\pi,\delta}[U^k(n)]$$

Fem servir la notació $\gamma_T^k(\pi, \delta) = (\gamma_T^k(1, \pi, \delta), \gamma_T^k(2, \pi, \delta), ..., \gamma_T^k(z, \pi, \delta))$. És important veure que si $\sum_{n=1}^{\infty} E_{s,\pi,\delta}[U^k(n)]$ existeix, aleshores és igual a $\gamma_T^k(s, \pi, \delta)$. Triem $\liminf$ a la definició pels mateixos motius que triàvem $\liminf$ a la definició d'utilitat mitjana.

Generalment la utilitat total d'un joc estocàstic serà $+\infty$ o $-\infty$, casos que no ens aporten gaire interès. Els jocs on ens resulta útil aquest criteri són els jocs en els que la utilitat mitjana és 0 per a totes o per a la majoria de les estratègies. Per exemple,

Exemple 3.12. Tenim el següent joc estocàstic de dos jugadors:

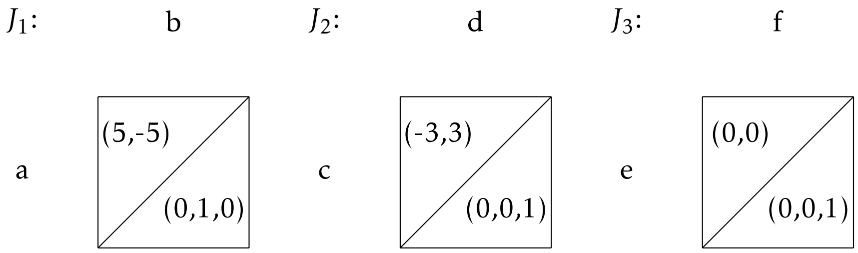


Si el joc estocàstic comença per l'estat J_1, tenim que a la primera etapa el segon jugador 'paga' 2 unitats d'utilitat al primer jugador, a la segona etapa el jugador 1 les torna, repetint-se aquesta situació infinitament. El vector de suma d'utilitats per al jugador 1 seria (2,0,2,0,2,0,...). Si, en canvi, el joc comença per l'estat J_2, aquest vector seria (-2,0,-2,0,-2,0,...). Està clar, llavors, que el primer jugador preferiria començar per J_1, però això no queda reflectit en l'utilitat mitjana, que és 0 per a tots dos jugadors tant si es comença per J_1 com si es comença per J_2. En canvi, la utilitat total per el primer jugador si es comença per J_1 és

$$\gamma_T^k(s,\pi,\delta) = \liminf_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} 2(-1)^{n-1} = 1$$

i si es comença per J_2 és -1.

La utilitat total també ens és molt útil per jocs en els que a partir d'una certa etapa els jugadors es queden atrapats en un estat on les utilitats que reben són 0. Normalment aquests estats es fan servir per a fer que el joc estocàstic sigui finit de forma artificial. Per exemple, tenim el següent joc estocàstic:



En aquest joc estocàstic, la utilitat total és, tal i com indica el nom, la suma de totes les utilitats que rebrà cada jugador. Com que $\sum_{n=1}^{\infty} E_{s,\pi,\delta}[U^k(n)]$ existeix, i és, en el cas que es comenci el joc per J_1, igual a 2 per al primer jugador, tenim que la utilitat total per al primer jugador començant per J_1 és, també, 2.

3.4 Resultats sobre estratègies estacionàries

Dels tres conjunts d'estratègies descrits els menys complexes són els de les estratègies estacionàries. A continuació donaré uns resultats útils per a l'anàlisi dels jocs estocàstics restringits a l'ús d'aquest tipus d'estratègies.

Definició 3.13. *Per a un parell d'estratègies estacionàries $(x,y) \in X \times Y$ definim:*

a) $Car^z(x) := \bigtimes_{s=1}^{z} Car(x_s)$, *on* $Car(x_s) := \{i \in A_s : x_s(i) > 0\}$, *el suport de* x *i de* x_s *respectivament.* $Car^z(y)$ *i* $Car(y_s)$ *es defineixen de forma similar.*

b) $u^k(x,y) := (u^k(1,x_1,y_1), u^k(s,x_2,y_2), ..., u^k(z,x_z,y_z))$, *on* $u^k(s,x_s,y_s) := \sum_{i=1}^{m_s}\sum_{j=1}^{n_s} x_s(i)u^k(s,i,j)y_s(j)$ *és la utilitat esperada del jugador* k *a l'estat* s*.*

c) $P(x,y)$ *és la matriu de transició de mida* $z \times z$*. L'element en la posició* (s,t) *de* $P(x,y)$ *és* $p(t|s,x_s,y_s) := \sum_{i=1}^{m_s}\sum_{j=1}^{n_s} x_s(i)p(t|s,i,j)y_s(j)$, *que és la probabilitat de transició de l'estat* s *a l'estat* t *si els jugadors juguen* x_s *i* y_s*. Notarem amb* $P(x,y)_s$ *la fila* s *de* $P(x,y)$*.*

d) $Q(x,y) := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P^n(x,y)$*. Notarem amb* $Q(x,y)_s$ *la fila* s *de* $Q(x,y)$*.*

$P(x,y)$ determina una cadena de Markov a l'espai d'estats per a cada $(x,y) \in X \times Y$.

Lema 3.14. *Sigui* $(x,y) \in X \times Y$*. Aleshores:*

a) L'element en la posició (s,t) *de la matriu* $P^{n-1}(x,y)$ *és igual a la probabilitat que a l'etapa* n *s'estigui jugant l'estat* t *si l'estat inicial és* s*.* $P^0(x,y) = Id$*.*

b) $E_{sxy}[U^k(n)] = P^{n-1}(x,y)_s u^k(x,y)$*.*

c) L'element en la posició (s,t) *de la matriu* $Q(x,y)$ *és igual al valor esperat de la mitjana de cops que s'ha jugat l'estat* t *si el joc comença en l'estat* s*.*

d) $Q(x,y)_s$ *és igual a la única distribució estacionària de la cadena de Markov relacionada amb* (x,y) *i comença a l'estat* s*.*

e) $Q(x,y)P(x,y) = Q(x,y)$*.*

f) $(Id - \beta P(x,y))$ *i* $(Id - P(x,y) + Q(x,y))$ *són matrius no singulars per a tot* $\beta \in (0,1)$*. Per tant* $(Id - \beta P(x,y) + Q(x,y))$ *no és singular per* β *aprop de 1.*

g) $Q(x,y) = \lim_{\beta\to 1}(1-\beta)(Id - \beta P(x,y))^{-1}$*.*

Els primers quatre resultats del lema són trivials a partir de la definició i els últims dos són resultats de teoria de cadenes de Markov que es poden trobar amb la seva demostració al llibre 'Finite Markov Chains'[4].

Lema 3.15. *Sigui* $(x,y) \in X \times Y$ *i* $\beta \in (0,1)$*. Aleshores:*

a) $\gamma_\beta^k(x,y) = (1-\beta)\sum_{n=1}^{\infty} \beta^{n-1}P^{n-1}(x,y)u^k(x,y)$*.*

b) $\gamma_\beta^k(x,y) = (1-\beta)(Id - \beta P(x,y))^{-1}u^k(x,y)$*.*

c) $\gamma_\beta^k(x,y)$ *és l'unic* $\alpha \in \mathbb{R}^z$ *que compleix que* $\alpha = (1-\beta)u^k(x,y) + \beta P(x,y)\alpha$*.*

Demostració 3.16. Per definició

$$\gamma_\beta^k(s,x,y) = (1-\beta)\sum_{n=1}^{\infty}\beta^{n-1}E_{s,x,y}[U^k(n)]$$

per a tot $s \in S$. Pel lema 3.14 (b) això implica

$$\gamma_\beta^k(x,y) = (1-\beta)\sum_{n=1}^{\infty}\beta^{n-1}P^{n-1}(x,y)u^k(x,y)$$

Com que

$$(1-\beta)\sum_{n=1}^{\infty}\beta^{n-1}P^{n-1}(x,y) = (1-\beta)(Id-\beta P(x,y))^{-1}$$

per a qualsevol matriu estocàstica $P(x,y)$, tenim que

$$\gamma_\beta^k(x,y) = (1-\beta)(Id-\beta P(x,y))^{-1}u^k(x,y)$$

Això implica que

$$(1-\beta)(Id-\beta P(x,y))\gamma_\beta^k(x,y) = (1-\beta)u^k(x,y)$$

i per tant $\gamma_\beta^k(s,x,y)$ és solució de

$$\alpha = (1-\beta)u^k(x,y) + \beta P(x,y)\alpha$$

Com que $(Id - \beta P(x,y)$ no és singular pel lema 3.14 (f), aquesta solució és única. □

Lema 3.17. *Sigui $(x,y) \in X \times Y$, $\beta \in (0,1)$ i $\alpha \in \mathbb{R}^z$. Aleshores:*

a) Si $\alpha \leq (1-\beta)u^k(x,y) + \beta P(x,y)\alpha$, aleshores $\alpha \leq \gamma_\beta^k(x,y)$.

b) Si $\alpha \not\leq (1-\beta)u^k(x,y) + \beta P(x,y)\alpha$, aleshores $\alpha \not\leq \gamma_\beta^k(x,y)$.

c) Passa el mateix canviant $\leq$ per $\geq$ o $\not\leq$ per $\not\geq$.

Demostració 3.18. Si

$$\alpha \not\leq (1-\beta)u^k(x,y) + \beta P(x,y)\alpha$$

aleshores

$$(Id-\beta P(x,y))\alpha \not\leq (1-\beta)u^k(x,y)$$

$(Id-\beta P(x,y))^{-1}$ és no negativa i cada columna té com a mínim un element positiu. Per tant tenim que

$$\alpha \not\leq (1-\beta)(Id-\beta P(x,y))^{-1}u^k(x,y) = \gamma_\beta^k(x,y)$$

com volíem. □

Lema 3.19. *Sigui $(x,y) \in X \times Y$. Aleshores:*

a) $\gamma^k(x,y) = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}P^{n-1}(x,y)u^k(x,y)$.

b) $\gamma^k(x,y) = Q(x,y)u^k(x,y)$.

c) $\gamma^k(x,y) = \alpha$ *per a qualsevol parella* $(\alpha,\delta) \in \mathbb{R}^z \times \mathbb{R}^z$ *que satisfà que* $\alpha = P(x,y)\alpha$ *i que* $\alpha + \delta = u^k(x,y) + P(x,y)\delta$.

d) $\gamma^k(x,y) = \lim_{\beta\to 1} \gamma^k_\beta(x,y)$.

Demostració 3.20. Per definició

$$\gamma^k(s,x,y) = \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} E_{s,x,y}[U^k(n)]$$

per a tot $s \in S$. El lema 3.14 (b) implica que

$$\gamma^k(s,x,y) = \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P^{n-1}(x,y)u^k(x,y)$$

Per resultats de cadenes de Markov de J. Kemeny i J. Snell a 'Finite Markov chains'[4] mencionat anteriorment, tenim que $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} P^{n-1}(x,y)$ existeix i és igual a $Q(x,y)$. Per tant hem demostrat (a) i (b).
Si $(\alpha,\delta) \in \mathbb{R}\times\mathbb{R}$ i $\alpha = P(x,y)\alpha$ i a més $\alpha+\delta = u^k(x,y) + P(x,y)\delta$, multiplicant la segona igualtat amb $Q(x,y)$ tenim pel lema 3.14 (e) que

$$Q(x,y)\alpha + Q(x,y)\delta = Q(x,y)u^k(x,y) + q(x,y)\delta$$

Per tant

$$Q(x,y)\alpha = q(x,y)u^k(x,y) = \gamma^k(x,y)$$

per l'apartat (b) d'aquest lema. A més, $\alpha = P(x,y)\alpha$ implica que $\alpha = Q(x,y)\alpha$, i per tant $\alpha = \gamma^k(x,y)$. Per acabar, l'apartat (d) és conseqüència directa del lema 3.14 (g) i el lema 3.15 (b).□

L'apartat (d) del lema anterior ens diu que el límit quan $\beta \to 1$ de $\gamma^k_\beta(x,y)$ coincideix amb $\gamma^k(x,y)$, com havíem intuït anteriorment.

Lema 3.21. *Sigui* $(x,y) \in X\times Y$ *i* $\alpha,\delta \in \mathbb{R}^z$. *Aleshores:*

a) Si $\alpha \leq P(x,y)\alpha$ *i* $\alpha+\delta \leq u^k(x,y) + P(x,y)\delta$ *aleshores* $\alpha \leq \gamma^k(x,y)$.

b) El mateix resultat és cert canviant $\leq$ *per* $\geq$.

Demostració 3.22.

$$\alpha \leq P(x,y)\alpha \implies \alpha \leq Q(x,y)\alpha$$

i a la vegada

$$\alpha+\delta \leq u^k(x,y) + P(x,y)\delta \implies Q(x,y)\alpha \leq Q(x,y)u^k(x,y)$$

Per tant,

$$\alpha \leq Q(x,y)u^k(x,y) = \gamma^k(x,y)$$

□

Lema 3.23. *Sigui* $(x,y) \in X\times Y$ *i suposem que* $\gamma^k(x,y) = 0$ *per a k=1,2. Aleshores:*

a) $\gamma_T^k(x,y) = \lim_{N\to\infty} \sum_{m=1}^{N} \sum_{n=1}^{m} P^{n-1}(x,y)u^k(x,y).$

b) $\gamma_T^k(x,y) = (Id - P(x,y) + Q(x,y))^{-1}u^k(x,y).$

c) $\gamma_T^k(x,y) = \alpha$ *per a qualsevol parella* $(\alpha,\delta) \in \mathbb{R}^z \times \mathbb{R}^z$ *que satisfà que* $\alpha = u^k(x,y) + P(x,y)\alpha$ *i que* $\alpha + \delta = P(x,y)\delta$.

d) $\gamma_T^k(x,y) = \lim_{\beta\to 1}(1-\beta)^{-1}\gamma_\beta^k(x,y).$

Demostració 3.24. Per definició

$$\gamma_T^k(s,x,y) = \liminf_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} E_{s,x,y}[U^k(n)]$$

per a tot $s \in S$. Per tant

$$\gamma_T^k(x,y) = \liminf_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} P^{n-1}(x,y)u^k(x,y)$$

Pel lema 3.19 (b) tenim que $0 = \gamma^k(x,y) = Q(x,y)u^k(x,y)$. Observem que per a cada $N \in \mathbb{N}$ tenim que

$$\begin{aligned}
&(Id - P(x,y) + q(x,y))\left(\frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} P^{n-1}(x,y)u^k(x,y)\right) = \\
&= \frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} [P^{n-1}(x,y)u^k(x,y) - P^n(x,y)u^k(x,y)] \\
&= \frac{1}{N} \sum_{m=1}^{N} [u^k(x,y) - P^m(x,y)u^k(x,y)] \\
&= u^k(x,y) - \frac{1}{N} \sum_{m=1}^{N} P^m(x,y)u^k(x,y)
\end{aligned}$$

Usant que $\gamma^k(x,y) = 0$ i el lema 3.14 (f) per la no singularitat de $(Id - P(x,y) + Q(x,y))$, tenim que

$$\lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \sum_{n=1}^{m} P^{n-1}(x,y)u^k(x,y)$$

existeix i que és igual a

$$(Id - P(x,y) + Q(x,y))^{-1}u^k(x,y)$$

cosa que demostra (a) i (b). Per a demostrar (c) suposem que α i γ són solució de $\alpha = u^k(x,y) + P(x,y)\alpha$ i a la vegada $\alpha + \delta = P(x,y)\delta$. Multiplicant la segona equació per $Q(x,y)$ tenim $Q(x,y)\alpha = 0$ per el lema 3.14 (e). Combinant això amb la primera equació tenim que

$$\alpha - P(x,y)\alpha + Q(x,y)\alpha = u^k(x,y)$$

Finalment, la no singularitat de $(Id - P(x,y) + Q(x,y))$ implica que

$$\alpha = (Id - P(x,y) + Q(x,y))^{-q}u^k(x,y)$$

i per tant tenim $\alpha = \gamma_T^k(x,y)$ per (b).
Per a demostrar (d) veiem que $Q(x,y)\gamma_T^k(x,y) = 0$ ja que

$$Q(x,y)\gamma_T^k(x,y) = Q(x,y)[(1-\beta)\sum_{n=1}^{\infty}\beta^{n-1}P^{n-1}(x,y)u^k(x,y)] =$$

$$= (1-\beta)\sum_{n=1}^{\infty}\beta^{n-1}Q(x,y)u^k(x,y) = 0$$

perquè $\gamma^k(x,y) = 0$. Per tant per el lema 3.15 (c) tenim que

$$\gamma_\beta^k(x,y) = (1-\beta)u^k(x,y) + \beta P(x,y)\gamma_\beta^k(x,y) - Q(x,y)\gamma_\beta^k(x,y)$$

Com que $(Id - \beta P(x,y) + Q(x,y))$ no és singular, tenim que

$$\gamma_\beta^k(x,y) = (1-\beta)(Id - \beta P(x,y) + Q(x,y))^{-1}u^k(x,y)$$

Com que $(Id - P(x,y) + Q(x,y))$ tampoc és singular, i gràcies a (b) d'aquest lema, tenim que

$$\lim_{\beta\to 1}(1-\beta)^{-1}\gamma_\beta^k(x,y) = (Id - P(x,y) + Q(x,y))^{-1}u^k(x,y) = \gamma_T^k(x,y)$$

□

Lema 3.25. *Sigui* $(x,y) \in X \times Y$ *i suposem que* $\gamma^k(x,y) = 0$ *per a k=1,2. Aleshores:*

a) Si $\alpha \leq u^k(x,y) + P(x,y)\alpha$ *i* $\alpha + \delta \leq P(x,y)\delta$*, aleshores* $\alpha \leq \gamma_T^k(x,y)$

b) El mateix resultat és cert canviant $\leq$ *per* $\geq$*.*

Demostració 3.26. De $\alpha + \delta \leq P(x,y)\delta$ tenim que $Q(x,y)\alpha + Q(x,y)\delta \leq Q(x,y)\delta$, i per tant $Q(x,y)\alpha \leq 0$. de $\alpha - P(x,y)\alpha \leq u^k(x,y)$ tenim que

$$P^{n-1}(x,y)\alpha - P^n(x,y)\alpha \leq P^{n-1}(x,y)u^k(x,y)$$

per a tot $n \in \mathbb{N}$. Això implica que per a tot $m \in \mathbb{N}$

$$\alpha - P^m(x,y)\alpha = \sum_{n=1}^{m} P^{n-1}(x,y)\alpha - P^n(x,y)\alpha) \leq \sum_{n=1}^{m} P^{n-1}(x,y)u^k(x,y)$$

I per tant

$$\alpha - \frac{1}{N}\sum_{m=1}^{N} P^m(x,y)\alpha \leq \frac{1}{N}\sum_{m=1}^{N}\sum_{n=1}^{m} P^{n-1}(x,y)u^k(x,y)$$

per a tot $n \in \mathbb{N}$. Quan N tendeix a ∞ i usant que $Q(x,y)\alpha \leq 0$, obtenim que $\alpha \leq \gamma_t^k(x,y)$.

3.5 Jocs estocàstics de suma zero

Ja definits els criteris d'utilitat, podem buscar l'existència d'estratègies que intentin maximitzar aquestes utilitats, de la mateixa manera que els equilibris de Nash dels jocs en forma normal. L'existència (o no) d'aquest tipus d'estratègies depén del criteri d'utilitat usat, del tipus d'estratègies permeses (estacionàries, de Markov o conductuals), de característiques particulars del joc estocàstic en si (com, per exemple, si és de suma zero o no) i del nombre de jugadors (que en el nostre cas sempre és 2, però totes les definicions que he donat es poden estendre a jocs estocàstics d'N jugadors, amb $N \in \mathbb{N}$).

Per començar, estudiarem els jocs estocàstics de suma zero, que en el cas de dos jugadors són aquells en els què

$$\forall s \in S,\ u^1(s,i,j) = -u^2(s,i,j)\ \ \forall i \in A_s,\ \forall j \in B_s$$

Estudiarem el cas en què el criteri d'utilitat que es fa servir és el criteri β-descomptat. En un joc de suma zero els dos jugadors tenen objectius directament oposats: el jugador 1 voldrà maximitzar la seva utilitat $\gamma_\beta^1(\pi,\delta)$, i el jugador 2 voldrà minimitzar aquesta utilitat. Assumirem, per tant, que el jugador 1 vol jugar una estratègia que maximitzi un mínim d'utilitat garantitzada (de forma similar a les estratègies maxmin del capítol 2), és a dir, que el jugador 1 busca π^* tal que $\inf_\delta \gamma_\beta^1(\pi^*,\delta) \geq \inf_\delta \gamma_\beta^1(\pi,\delta)$ per a tot $\pi \in \Pi$. És a dir que el jugador 1 està interessat en jugar l'estratègia π que li aporti la utilitat $\sup_\pi \inf_\delta \gamma_\beta^1(\pi,\delta)$. De manera similar, suposem que el jugador 2 està interessat en minimitzar la utilitat màxima que pot rebre el jugador 1, és a dir que busca jugar una estratègia δ que faci que el jugador 1 rebi una utilitat de $\inf_\delta \sup_\pi \gamma_\beta^1(\pi,\delta)$. Veiem que

$$\sup_\pi \inf_\delta \gamma_\beta^1(\pi,\delta) \leq \inf_\delta \sup_\pi \gamma_\beta^1(\pi,\delta)$$

ja que

$$\begin{gathered}\inf_\delta \gamma_\beta^1(\pi^*,\delta) \leq \gamma_\beta^1(\pi^*,\delta^*)\ \ \forall \pi^* \in \Pi,\ \ \forall \delta^* \in \Delta \Rightarrow \\ \Rightarrow \sup_\pi \inf_\delta \gamma_\beta^1(\pi,\delta) \leq \sup_\pi \gamma_\beta^1(\pi,\delta^*)\ \ \forall \delta^* \in \Delta \Rightarrow \\ \Rightarrow \sup_\pi \inf_\delta \gamma_\beta^1(\pi,\delta) \leq \inf_\delta \sup_\pi \gamma_\beta^1(\pi,\delta)\end{gathered}$$

Si $\sup_\pi \inf_\delta \gamma_\beta^1(\pi,\delta) = \inf_\delta \sup_\pi \gamma_\beta^1(\pi,\delta)$, diem que aquest vector és el *valor del joc*. Formalment,

Definició 3.27. *Sigui Γ un joc estocàstic de dos jugadors de suma zero, i sigui $\beta \in (0,1)$. Si existeix $v_\beta^1 \in \mathbb{R}^z$ tal que $\sup_\pi \inf_\delta \gamma_\beta^1(\pi,\delta) = v_\beta^1 = \inf_\delta \sup_\pi \gamma_\beta^1(\pi,\delta)$, aleshores anomenem valor del joc Γ a aquesta v_β^1.*

Definició 3.28. *Si el valor d'un joc estocàstic de dos jugadors de suma zero és v_β^1, una estratègia $\pi^* \in \Pi$ és una estratègia β-descomptada òptima per al jugador 1 si $\gamma_\beta^1(\pi^*,\delta) \geq v_\beta^1$ per a tot $\delta \in \Delta$.*

Simètricament podem definir v_β^2 i estratègia β-descomptada òptima per al jugador 2. A més, de manera similar es pot definir el valor del joc per al criteri de la utilitat mitjana (notat amb v^1) i per al criteri de la utilitat total (notat amb v_t^1), i les seves respectives estratègies òptimes. Notem que $v_\beta^2 = -v_\beta^1$ ja que $u^1(s,i,j) = -u^2(s,i,j)$ per a tot s,i,j.

Teorema 3.29. *Sigui Γ un joc estocàstic de suma zero, i sigui $\beta \in (0,1)$. Aleshores v_β^1 existeix.*

Demostració 3.30. Sigui Γ el nostre joc estocàstic. Definim, per a cada $s,l \in S$ les següents matrius:

$$A^s = \begin{bmatrix} u^1(s,1,1) & u^1(s,1,2) & \dots & u^1(s,1,n_s) \\ u^1(s,2,1) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ u^1(s,m_s,1) & \dots & \dots & u^1(s,m_s,n_s) \end{bmatrix}$$

$$P^{sl} = \begin{bmatrix} p(l|s,1,1) & p(l|s,1,2) & \dots & p(l|s,1,n_s) \\ p(l|s,2,1) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ p(l|s,m_s,1) & \dots & \dots & p(l|s,m_s,n_s) \end{bmatrix}$$

On recordem que $u^1(s,i,j)$ és la utilitat que rep el jugador 1 si en el joc s el jugador 1 juga i i el jugador 2 juga j, i $p(l|s,i,j)$ és la probabilitat de transició de l'estat s a l'estat l si el jugador 1 juga i i el jugador 2 juga j.

Notem amb a_{ij}^s cada element de A^s (és a dir $a_{ij}^s = u^1(s,i,j)$ per a cada $i \in \{1,2,...,m_s\}$, $j \in \{1,2,...,n_s\}$), i amb p_{ij}^{sl} cada element de P^{sl}.

Ja hem vist anteriorment que si fem servir el criteri β-descomptat, aleshores $|\gamma_\beta^k(s,x,y)| \leq M$, on $M = max\{|u^k(s,i,j)| : k \in \{1,2\}, i \in A_s, j \in B_s, s \in S\}$.

Donat un joc normal $s \in S$, denotem amb $val[A^s]$ el valor minmax del joc s per al primer jugador (definit a la definició 2.19), és a dir el valor minmax del joc que té A^s com a matriu d'utilitats del primer jugador.

És fàcil veure que donats dos jocs normals s,l de les mateixes dimensions (és a dir $n_s = n_l$ i $m_s = m_l$), aleshores

$$|val[A^s] - val[A^l]| \leq \max_{i,j} |a_{ij}^s - a_{ij}^l| \tag{1}$$

Definim la matriu $A^s(\vec{\alpha})$ com la matriu que tè per elements

$$(1-\beta)a_{ij}^s + \beta \sum_{l \in S} p_{ij}^{sl} \alpha_l, \;\; i \in \{1,2,...,m_s\}, j \in \{1,2,...,n_s\}$$

on $\vec{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_z)$ és un vector de nombres reals qualsevol de z components (on z és el nombre d'estats del nostre joc estocàstic).

Escollim $\vec{\alpha}_{(0)}$ arbitràriament, i definim $\vec{\alpha}_{(t)}$ recursivament de la següent manera:

$$\alpha_{s(t)} = Val[A^s(\alpha_{(t-1)})]$$

on $Val[A^s(\alpha_{(t-1)})]$ denota el valor minmax del joc normal que te com a matriu d'utilitats els elements de $A^s(\alpha_{(t-1)})$.

Veurem que el límit quan $t \to \infty$ de $\alpha_{(t)}$ existeix i és independent de l'elecció de $\alpha_{(0)}$, i que aquest vector límit (que anomenarem $\vec{\phi}$) és el valor del joc.

Considerem la transformació T:

$$T\vec{\alpha} = \vec{\gamma}, \quad \text{on } \gamma_s = Val[A^s(\vec{\alpha})]$$

Definim la norma de $\vec{\alpha}$ com

$$\|\vec{\alpha}\| = \max_s |\alpha_s|$$

Aleshores tenim

$$\begin{aligned}
\|T\vec{\gamma} - T\vec{\alpha}\| &= \max_s |Val[A^s(\vec{\gamma})] - Val[A^s(\vec{\alpha})]| \\
&\le \max_{s,i,j} |\beta \sum_{l\in S} p_{ij}^{sl}\gamma_l - \beta \sum_{l\in S} p_{ij}^{sl}\alpha_l| \\
&\le \beta \max_{s,i,j} |\sum_l p_{ij}^{sl}| \max_l |\gamma_l - \alpha_l| = \beta \max_l |\gamma_l - \alpha_l| \\
&= \beta\|\vec{\gamma} - \vec{\alpha}\|
\end{aligned}$$

on a la primera desigualtat hem fet servir (1).

En particular, $\|T^2\vec{\alpha} - T\vec{\alpha}\| \le \beta\|T\vec{\alpha} - \vec{\alpha}\|$. Com que $\beta \in (0,1)$, la successió $\vec{\alpha}_{(0)}, T\vec{\alpha}_{(0)}, T^2\vec{\alpha}_{(0)}, \ldots$ és convergent. El vector límit $\vec{\phi}$ té la propietat que $T\vec{\phi} = \vec{\phi}$. Aquest vector és únic, ja que si tinguessim un altre vector $\vec{\psi}$ tal que $T\vec{\psi} = \vec{\psi}$, aleshores

$$\|\vec{\psi} - \vec{\phi}\| = \|T\vec{\psi} - T\vec{\phi}\| \le \beta\|\vec{\psi} - \vec{\phi}\|,$$

I per tant $\|\vec{\psi} - \vec{\phi}\| = 0$. Per tant $\vec{\phi}$ és l'únic punt fix per T i és independent de $\vec{\alpha}_{(0)}$.

Anem a veure que $\vec{\phi}$ és el valor del joc estocàstic Γ. Anomenem Γ_s al joc estocàstic que té s com a estat inicial, i anomenem $\Gamma_{s(t)}$ al joc que tindríem si juguèssim t etapes del joc estocàstic Γ_s i després de t etapes deixéssim de jugar.

Nota: Aquest tipus de joc ($\Gamma_{s(t)}$) no és un joc estocàstic, ja que tè un nombre d'etapes finit. Tampoc és un joc normal, ja que tè més d'una etapa i més d'un estat. És un tipus de joc que s'anomena joc de dos jugadors de suma zero en forma extensiva, que no estudiem en aquest treball. Per a poder acabar la demostració, però, necessitem uns resultats que es poden trobar al llibre 'Theory of Games and Economic Behavior'[11], que ens diuen que aquests tipus de joc també tenen un valor (similar al valor dels jocs normals i estocàstics de dos jugadors de suma zero ja definits en aquest treball) i que si els jugadors segueixen estratègies òptimes (també similars a les definides en aquest treball), aquests garantitzen una utilitat major o igual que el valor del joc.

Tornant a la demostració, anomenem $v^1_{s(t)}$ al valor del joc en forma extensiva $\Gamma_{s(t)}$. Si el jugador 1 juga una estratègia òptima de $\Gamma_{s(t)}$ durant les primeres t etapes de Γ_s i juga arbitràriament a partir de l'etapa $t+1$ aleshores tindrà garantida (gràcies a resultats del llibre mencionat anteriorment) una utilitat β-descomptada $\gamma^1_\beta(s,x,y) \in (v^1_{s(t)} - \beta^t M, v^1_{s(t)} + \beta^t M)$. Quan $t \to \infty$, tenim que $\beta^t M \to 0$ i $v^1_{s(t)} \to \phi_s$, on ϕ_s denota la s-èssima component del vector $\vec{\phi}$. Per tant, concluim que $\vec{\phi}$ és el valor del joc Γ. $\square$

La demostració del teorema ens caracteritza v_β^1, ja que per construcció, v_β^1 és l'únic vector $\vec{\phi} = (\phi_1, ..., \phi_z) \in \mathbb{R}^z$ que és solució de l'equació

$$\phi_s = Val[A^s(\phi_s)], \ \ s \in S$$

anomenada *equació de Shapley.*

Teorema 3.31. *Sigui* Γ *un joc estocàstic de suma zero, sigui* $\beta \in (0,1)$*. Existeixen estratègies estacionàries* β*-descomptades òptimes per a cada jugador.*

Demostració 3.32. Sigui v_β^1 el valor del joc Γ. Fent servir les mateixes notacions que a la demostració 3.30, per a cada $s \in S$, agafem x_s^* i y_s^* una estratègia òptima (o estratègia min-max) per a cada jugador per al joc normal de suma zero que té com a matriu d'utilitats $A^s(v_\beta^1)$. Anem a veure que aleshores $\vec{x}^* = (x_1^*, x_2^*, ..., x_z^*)$, $\vec{y}^* = (y_1^*, y_2^*, ..., y_z^*)$ són estratègies estacionàries β-descomptades òptimes per a cada jugador.

Sigui $s \in S$. Similarment a la demostració 3.30, anomenem $\Gamma_{s(t)}$ al joc de suma zero en forma extensiva que tindríem si juguèssim t etapes del joc estocàstic Γ_s i després de t el joc acabés, però ara a cada etapa d'aquest joc el jugador 1 rep una utilitat de $a_{ij}^h + \sum_l p_{ij}^{hl} \phi_l$ en comptes d'una utilitat de a_{ij}^h. En aquest joc l'estratègia estacionària $\vec{x}^*$ garantitza al jugador 1 una utilitat de com a mínim ϕ_s, i per tant, en el joc estocàstic original, la utilitat β-descomptada per al jugador 1 si el joc s'aturés en el moment t seria, com a mínim, de

$$\phi_s - \beta^{(t-1)} \max_{h,i,j} \sum_l p_{ij}^{hl} \phi_l$$

i per tant, com aquesta utilitat serà com a mínim

$$\phi_s - \beta^{(t-1)} \max_l \phi_l$$

Com que això és cert per a valors de t arbitràriament grans, concloem que l'estratègia estacionària x_s^* és β-descomptada òptima per al jugador 1 per a l'estat inicial $s \in S$. Fent servir el mateix argument per a cada $s \in S$, i un argument simètric per al jugador 2, demostrem el teorema.□

Els resultats de Vrieze i Tijs al seu article 'Relations between game parameters, value, and optimal strategy spaces in stochastic games and construction of games with given solution'[12] al volum 31 del 'Journal of Optimization Theory and Applications' ens diuen que les estratègies estacionàries $\vec{x}^*$ i $\vec{y}^*$ descrites a la demostració 3.32 són, de fet, les úniques estratègies estacionàries que són β-descomptades òptimes per a jocs estocàstics de suma zero.

3.6 Equilibris en jocs estocàstics de suma general

Ara que ja hem vist els resultats més importants per als jocs estocàstics de suma zero ens queden per veure els resultats per a jocs estocàstics de suma general, és a dir aquells que no són de suma constant. Comencem definint equilibri en jocs estocàstics de forma anàloga a com hem definit equilibris de Nash al capítol 2.

Definició 3.33. *Sigui* Γ *un joc estocàstic. Un parell d'estratègies* $(\pi^*,\delta^*) \in \Pi \times \Delta$ *s'anomenen equilibri* β*-descomptat per a l'estat inicial s si:*

$$\gamma^1_\beta(s,\pi^*,\delta^*) \geq \gamma^1_\beta(s,\pi,\delta^*) \text{ per a tot } \pi \in \Pi, \text{ i}$$
$$\gamma^2_\beta(s,\pi^*,\delta^*) \geq \gamma^2_\beta(s,\pi^*,\delta) \text{ per a tot } \delta \in \Delta.$$

Si (π^*,δ^*) *és un equilibri* β*-descomptat per a tot* $s \in S$*, aleshores diem que* (π^*,δ^*) *és un equilibri* β*-descomptat. Es defineix equilibri mitjà i equilibri total de manera anàloga (canviant* $\gamma^i_\beta(s,\pi^*,\delta^*)$ *per* $\gamma^i(s,\pi^*,\delta^*)$ *o* $\gamma^i_T(s,\pi^*,\delta^*)$ *respectivament).*

El següent teorema és un anàleg al teorema de Nash, per a jocs estocàstics. Fink va demostrar-lo l'any 1963 per a jocs estocàstics d'n jugadors al seu article 'Equilibrium in a Stochastic n-Person Game'[9] al volum 28 de la revista 'Hiroshima Mathematical Journal'. Nosaltres demostrarem el teorema equivalent per a jocs estocàstics de dos jugadors.

Teorema 3.34. *Sigui* Γ *un joc estocàstic de 2 jugadors i* $\beta \in (0,1)$*. Suposem que els jugadors estan restringits a jugar estratègies estacionàries. Aleshores existeix un equilibri* β*-descomptat.*

Demostració 3.35. Sense pèrdua de generalitat suposarem que l'estat inicial del joc és $z \in S$. Similarment a la demostració del teorema de Nash, volem trobar una funció contínua $\tau : X \times Y \to X \times Y$ (on X, Y denoten els espais d'estratègies estacionàries com en la definició 3.4) en la què els seus punts fixos són els equilibris β-descomptats del nostre joc estocàstic, fent servir el teorema del punt fix de Brower per a demostrar la seva existència.

Recordem que un element $x \in X$ és un vector $x = (x_1,...,x_z)$, i cada element x_s d'aquest vector és en si un vector $x_s = (x_{s_1},...,x_{s_{m_s}})$, on x_{s_i} denota la probabilitat que el jugador 1 jugui l'acció $i \in A_s$ en el joc normal $s \in S$. Donat $(x,y) \in X \times Y$, per a cada $s \in S$ i per a cada $i \in A_s$, $j \in B_s$, definim:

$$\phi^1_{s_i}(z,x,y) := \max\{0, \gamma^1_\beta(z,s_i,y) - \gamma^1_\beta(z,x,y)\}$$
$$\phi^2_{s_j}(z,x,y) := \max\{0, \gamma^2_\beta(z,x,s_j) - \gamma^2_\beta(z,x,y)\}$$

on $\gamma^1_\beta(z,s_i,y)$ denota la utilitat β-descomptada per al jugador 1 si juga l'estratègia s_i, estratègia idèntica a x excepte que en el joc $s \in S$ juga l'acció $i \in A_s$ amb probabilitat 1 (és a dir $\overline{s_i}_{s_i} = 1$), i $\gamma^2_\beta(z,x,s_j)$ denota la utilitat β-descomptada per al jugador 2 si juga l'estratègia s_j, estratègia idèntica a y excepte que en el joc $s \in S$ el jugador 2 juga $j \in B_s$ amb probabilitat 1.

Alehores, definim $\tau : X \times Y \to X \times Y$ com $\tau(x,y) = (\overline{x},\overline{y})$, on $\overline{x} = (\overline{x}_1,...,\overline{x}_z)$, i $\overline{x}_s = (\overline{x}_{s_1},...,\overline{x}_{s_{m_s}})$ per a cada $s \in S$, on cada $\overline{x}_{s_i}$ l'obtenim de la següent manera:

$$\overline{x}_{s_i} = \frac{x_{s_i} + \phi^1_{s_i}(z,x,y)}{1 + \sum\limits_{j \in A_s} \phi^1_{s_j}(z,x,y)}$$

De la mateixa manera, $\overline{y} = (\overline{y}_1,...,\overline{y}_z)$, i $\overline{y}_s = (\overline{y}_{s_1},...,\overline{y}_{s_{n_s}})$ per a cada $s \in S$, on cada $\overline{y}_{s_i}$ l'obtenim:

$$\overline{y}_{s_j} = \frac{y_{s_j} + \phi^2_{s_j}(z,x,y)}{1 + \sum\limits_{k \in B_s} \phi^2_{s_k}(z,x,y)}$$

Com en el teorema de Nash, el teorema del punt fix de Brower ens garanteix l'existència de punts fixos de la funció τ. Per tant, per acabar ens queda demostrar que els punts fixos de τ coincideixen amb els equilibris β-descomptats.

Si (x^*, y^*) és un equilibri β-descomptat, aleshores $\phi^1_{s_i}(z, x^*, y^*) = 0$ per a tot $s \in S$ per a tot $i \in A_s$, ja que

$$\begin{aligned}
&\gamma^1_\beta(z, x^*, y^*) \geq \gamma^1_\beta(z, x, y^*) \text{ per a tot } x \in X \implies \\
&\gamma^1_\beta(z, s_i, y^*) - \gamma^1_\beta(z, x^*, y^*) \leq 0 \text{ per a tot } s \in S, \text{ per a tot } i \in A_s \implies \\
&\phi^1_{s_i}(z, x^*, y^*) = 0 \text{ per a tot } s \in S, \text{ per a tot } i \in A_s
\end{aligned}$$

i per tant

$$\overline{x}^*_{s_i} = \frac{x^*_{s_i} + 0}{1 + \sum_{j \in A_s} 0} = x^*_{s_i} \text{ per a tot } s \in S \text{ i per a tot } i \in A_s$$

De la mateixa manera $\phi^2_{s_j}(z, x^*, y^*) = 0$ i $\overline{y}^*_{s_j} = y^*_{s_j}$. Per tant $\tau(x^*, y^*) = (x^*, y^*)$, i per tant (x^*, y^*) és un punt fix.

Anem a veure el recíproc. Si (x^*, y^*) és un punt fix de τ, volem veure que (x^*, y^*) és un equilibri β-descomptat.

$$\tau(x^*, y^*) = (x^*, y^*) \implies \overline{x}^*_{s_i} = \frac{\overline{x}^*_{s_i} + \phi^1_{s_i}(z, x^*, y^*)}{1 + \sum_{j \in A_s} \phi^1_{s_j}(z, x^*, y^*)} \;\; \forall s \in S, \;\; \forall i \in A_s \implies$$

$$\implies \overline{x}^*_{s_i} + \overline{x}^*_{s_i} \sum_{j \in A_s} \phi^1_{s_j}(z, x^*, y^*) = \overline{x}^*_{s_i} + \phi^1_{s_i}(z, x^*, y^*) \;\; \forall s \in S, \;\; \forall i \in A_s \implies$$

$$\implies \phi^1_{s_i}(z, x^*, y^*) = \overline{x}^*_{s_i} \sum_{j \in A_s} \phi^1_{s_j}(z, x^*, y^*) \;\; \forall s \in S, \;\; \forall i \in A_s$$

Per tant, tenim que si $\overline{x}^*_{s_i} = 0$, aleshores $\phi^1_{s_i}(z, x^*, y^*) = 0$. Seguint el mateix raonament, tenim que si $\overline{y}^*_{s_j} = 0$, aleshores $\phi^2_{s_j}(z, x^*, y^*) = 0$.

Suposem que

$$\sum_{s \in S} \sum_{i \in A_s} \phi^1_{s_i}(z, x^*, y^*) > 0$$

Aleshores tindríem que

$$\overline{x}^*_{s_i} = \frac{\phi^1_{s_i}(z, x^*, y^*)}{\sum_{j \in A_s} \phi^1_{s_j}(z, x^*, y^*)}$$

per a cada $s \in S$ i per a cada $i \in A_s$ que compleixin que $\overline{x}^*_{s_i} > 0$.

A la vegada, però, existeixen una $s \in S$ i una $k \in A_s$ tal que $\overline{x}^*_{s_k} > 0$ i a la vegada $\phi^1_{s_k}(z, x^*, y^*) = 0$, ja que:

$$\gamma^1_\beta(z, x^*, y^*) \geq \min_{s \in S, i \in A_s} \{\gamma^1_\beta(z, s_i, y^*) | \overline{x}^*_{s_i} > 0\} = \gamma^1_\beta(z, s_k, y^*)$$

i per tant $\phi^1_{s_k}(z,x^*,y^*) = 0 < \overline{x}^*_{s_k}$, entrant en contradicció amb el que hem vist fa un moment. Per tant la suposició

$$\sum_{s\in S}\sum_{i\in A_s}\phi^1_{s_i}(z,x^*,y^*) > 0$$

era falsa, i per tant $\phi^1_{s_i}(z,x^*,y^*) = 0$ per a tot $s \in S$ i per a tot $i \in A_s$. Fent el mateix argument per al segon jugador arribem a que $\phi^2_{s_j}(z,x^*,y^*) = 0$ per a tot $s \in S$ i per a tot $j \in B_s$. Per tant

$$\gamma^1_\beta(z,s_i,y^*) - \gamma^1_\beta(z,x^*,y^*) \leq 0 \text{ per a tot } s \in S \text{ i per a tot } i \in A_s$$

i a la vegada

$$\gamma^2_\beta(z,x^*,s_j) - \gamma^2_\beta(z,x^*,y^*) \leq 0 \text{ per a tot } s \in S \text{ i per a tot } j \in B_s$$

Gràcies a un algoritme de R.A. Howard que es troba al seu llibre 'Dynamic Programming and Markov Processes'[3] podem concloure que

$$\gamma^1_\beta(z,x^*,y^*) \geq \gamma^1_\beta(z,x,y^*) \text{ per a tot } x \in X$$

$$\gamma^1_\beta(z,x^*,y^*) \geq \gamma^1_\beta(z,x^*,y) \text{ per a tot } y \in Y$$

i per tant (x^*,y^*) és un equilibri β-descomptat, com volíem. □

Conclusions

Al darrer capítol hem introduït tots els conceptes bàsics sobre els jocs estocàstics i hem arribat a demostrar l'existència d'equilibris en el cas que els jugadors estiguin restringits a l'ús d'estratègies estacionàries. A més, amb l'excepció d'alguns resultats imprescindibles per a arribar als nostres resultats però a la vegada massa llargs de demostrar en l'extensió d'aquest treball (com el teorema del punt fix de Brower), hem aconseguit arribar-hi construïnt els resultats de teoria de jocs que necessitàvem pràcticament des de zero, assolint així els objectius que ens plantejàvem en un principi.

El següent pas seria estendre el treball per a explorar jocs estocàstics de més de dos jugadors o eliminar la restricció de jugar estratègies estacionàries. Això, però, m'ho plantejo com a projecte personal pels anys vinents.

Referències

[1] K. Border. *Fixed Point Theorems with Applications to Economics and Game Theory.* Cambridge University Press, 1985.

[2] A. M. Fink. *Equilibrium in a stochastic n-person game*. J. Sci. Hiroshima Univ. Ser. A-I Math. 28, no. 1, 89-93, 1964.

[3] R. Howard. *Dynamic Programming and Markov Processes*. MIT Press, Cambridge, MA, 1960.

[4] J. Kemeny, J. Snell. *Finite markov chains*. Van Nostrand Princeton, NJ, 1960.

[5] L. S. Shapley. *Stochastic Games*. Proceedings of the National Academy of Sciences, 39 (10) 1095-1100, 1953.

[6] Y. Shoham, K. Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2008.

[7] M. J. Sobel. *Noncooperative Stochastic Games.* Ann. Math. Statist. 42, no. 6, 1930-1935, 1971.

[8] E. Solan, N. Vieille. *Stochastic games*. Proceedings of the National Academy of Sciences, 112 (45) 13743-13746, 2015.

[9] M. Takahashi. *Equilibrium points of stochastic non-cooperative n-person games.* J. Sci. Hiroshima Univ. Ser. A-I Math. 28, no. 1, 95-99, 1964.

[10] F. Thuijsman. *Optimality and equilibria in stochastic games*. CWI Tracts, 1992.

[11] J. Von Neumann, O. Morgenstern. *Theory of games and economic behavior.* Princeton University Press, 1944.

[12] O.J. Vrieze, S.H. Tijs. *Relations between game parameters, value, and optimal strategy spaces in stochastic games and construction of games with given solution.* J Optim Theory Appl 31, 501–513, 1980.