Algoritmos para Juegos con Información Incompleta y No Determinismo

Rubmary Rojas

Universidad Simón Bolívar, Caracas, Venezuela

Enero 2020

Agenda

1. Teoría de juegos.
2. Juegos en forma normal:
    - Modelo y estrategias.
    - Equilibrio de Nash.
    - Regret Matching: algoritmo, experimentos y conclusiones.
3. Juegos en forma extensiva:
    - Modelo y estrategias.
    - Kuhn Poker.
    - Counterfactual Regret Minimization (CFR): algoritmo, experimentos y conclusiones.
4. Conclusiones y Recomendaciones.
5. Demo.

Teoría de Juegos

Definición

- *Estudio de* **modelos matemáticos** *de* **conflicto** *y* **cooperación***.*
- *Agentes que toman decisiones de forma* **racional** *e* **inteligente***.*

Aplicaciones

Ciencias sociales

Economía

Matemática

Computación

Juegos no deterministas con información incompleta

No determinismo

Incertidumbre probabilística:
- Lanzar dados
- Repartir cartas

Información incompleta

Información parcial sobre algunas de las acciones que fueron tomadas previamente.

Interrogantes

- ¿Qué significa que un juego sea resuelto?
- ¿Cuándo un jugador juega de forma óptima?

Objetivo General

Comprender los conceptos en el área de juegos de dos personas que involucran información incompleta y no determinismo, así como implementar los algoritmos para resolverlos, realizando experimentos sobre distintos juegos que son capturados por el modelo.

Forma Normal o Estratégica

Juegos en Forma Normal o Estratégica

Piedra, papel o tijera

	$\mathcal{R}$ (piedra)	$\mathcal{P}$ (papel)	$\mathcal{S}$ (tijera)
$\mathcal{R}$ (piedra)	$0, 0$	$-1, 1$	$1, -1$
$\mathcal{P}$ (papel)	$1, -1$	$0, 0$	$-1, 1$
$\mathcal{S}$ (tijera)	$-1, 1$	$1, -1$	$0, 0$

Juegos en Forma Normal o Estratégica

Piedra, papel o tijera

	$\mathcal{R}$ (piedra)	$\mathcal{P}$ (papel)	$\mathcal{S}$ (tijera)
$\mathcal{R}$ (piedra)	$0, 0$	$-1, 1$	$1, -1$
$\mathcal{P}$ (papel)	$1, -1$	$0, 0$	$-1, 1$
$\mathcal{S}$ (tijera)	$-1, 1$	$1, -1$	$0, 0$

jugador 1

Juegos en Forma Normal o Estratégica

Piedra, papel o tijera

	$\mathcal{R}$ (piedra)	$\mathcal{P}$ (papel)	$\mathcal{S}$ (tijera)	jugador 2
$\mathcal{R}$ (piedra)	$0, 0$	$-1, 1$	$1, -1$	
$\mathcal{P}$ (papel)	$1, -1$	$0, 0$	$-1, 1$	
$\mathcal{S}$ (tijera)	$-1, 1$	$1, -1$	$0, 0$	

Juegos en Forma Normal o Estratégica

Piedra, papel o tijera

	$\mathcal{R}$ (piedra)	$\mathcal{P}$ (papel)	$\mathcal{S}$ (tijera)
$\mathcal{R}$ (piedra)	$0,0$	$-1,1$	$1,-1$
$\mathcal{P}$ (papel)	$1,-1$	$0,0$	$-1,1$
$\mathcal{S}$ (tijera)	$-1,1$	$1,-1$	$0,0$

primer jugador **gana** 1

Juegos en Forma Normal o Estratégica

Piedra, papel o tijera

	$\mathcal{R}$ (piedra)	$\mathcal{P}$ (papel)	$\mathcal{S}$ (tijera)
$\mathcal{R}$ (piedra)	$0, 0$	$-1, 1$	$1, -1$
$\mathcal{P}$ (papel)	$1, -1$	$0, 0$	$-1, 1$
$\mathcal{S}$ (tijera)	$-1, 1$	$1, -1$	$0, 0$

segundo jugador **pierde** 1

Juegos en Forma Normal o Estratégica

Piedra, papel o tijera

	$\mathcal{R}$ (piedra)	$\mathcal{P}$ (papel)	$\mathcal{S}$ (tijera)
$\mathcal{R}$ (piedra)	$0, 0$	$-1, 1$	$1, -1$
$\mathcal{P}$ (papel)	$1, -1$	$0, 0$	$-1, 1$
$\mathcal{S}$ (tijera)	$-1, 1$	$1, -1$	$0, 0$

Elementos

1. Jugadores.
2. Acciones o estrategias puras:
   $\mathcal{R}$, $\mathcal{P}$, $\mathcal{S}$.
3. Función de pago o utilidades.

Juegos en Forma Normal o Estratégica

Piedra, papel o tijera

	$\mathcal{R}$ (piedra)	$\mathcal{P}$ (papel)	$\mathcal{S}$ (tijera)
$\mathcal{R}$ (piedra)	$0, 0$	$-1, 1$	$1, -1$
$\mathcal{P}$ (papel)	$1, -1$	$0, 0$	$-1, 1$
$\mathcal{S}$ (tijera)	$-1, 1$	$1, -1$	$0, 0$

Juego de dos jugadores de suma cero

Elementos

1. Jugadores.
2. Acciones o estrategias puras: $\mathcal{R}$, $\mathcal{P}$, $\mathcal{S}$.
3. Función de pago o utilidades.

Juegos en Forma Normal o Estratégica

Piedra, papel o tijera

	$\mathcal{R}$ (piedra)	$\mathcal{P}$ (papel)	$\mathcal{S}$ (tijera)
$\mathcal{R}$ (piedra)	$0, 0$	$-1, 1$	$1, -1$
$\mathcal{P}$ (papel)	$1, -1$	$0, 0$	$-1, 1$
$\mathcal{S}$ (tijera)	$-1, 1$	$1, -1$	$0, 0$

Juego de
dos jugadores
de suma cero

Elementos

1. Jugadores.
2. Acciones o estrategias puras: $\mathcal{R}$, $\mathcal{P}$, $\mathcal{S}$.
3. Función de pago o utilidades.

Estrategias

1. Estrategias puras: siempre se elige la misma acción.
2. Estrategias mixtas: cada acción se elige con cierta probabilidad.

Juegos en Forma Normal o Estratégica

Batalla de los sexos

		José	
		ballet	béisbol
María	ballet	$2, 1$	$0, 0$
	béisbol	$0, 0$	$1, 2$

Juegos en Forma Normal o Estratégica

Batalla de los sexos

José

María		ballet	béisbol
	ballet	$2, 1$	$0, 0$
	béisbol	$0, 0$	$1, 2$

Ninguno obtiene ganancia.

Juegos en Forma Normal o Estratégica

Batalla de los sexos

		José	
		ballet	béisbol
María	ballet	$2, 1$	$0, 0$
	béisbol	$0, 0$	$1, 2$

María obtiene una ganancia mayor que José.

Juegos en Forma Normal o Estratégica

Batalla de los sexos

		José	
		ballet	béisbol
María	ballet	$2, 1$	$0, 0$
	béisbol	$0, 0$	$1, 2$

José obtiene una ganancia mayor que María.

Juegos en Forma Normal o Estratégica

Batalla de los sexos

		José	
		ballet	béisbol
María	ballet	$2, 1$	$0, 0$
	béisbol	$0, 0$	$1, 2$

Conceptos

1. Ganancia Esperada

Valor promedio que un determinado jugador obtendría si jugara infinitas veces cuando cada jugador utiliza una estrategia dada.

Juegos en Forma Normal o Estratégica

Batalla de los sexos

		José	
		ballet	béisbol
María	ballet	$2, 1$	$0, 0$
	béisbol	$0, 0$	$1, 2$

Conceptos

1. Ganancia Esperada
2. Mejor Respuesta

La mejor forma en que puede jugar un jugador dadas las estrategias seleccionadas de sus oponentes.

Juegos en Forma Normal o Estratégica

Batalla de los sexos

		José	
		ballet	béisbol
María	ballet	$2,1$	$0,0$
	béisbol	$0,0$	$1,2$

Si María siempre elige ballet.

Conceptos

1. Ganancia Esperada
2. Mejor Respuesta

La mejor forma en que puede jugar un jugador dadas las estrategias seleccionadas de sus oponentes.

Juegos en Forma Normal o Estratégica

Batalla de los sexos

		José	
		ballet	béisbol
María	ballet	$2, 1$	$0, 0$
	béisbol	$0, 0$	$1, 2$

Lo mejor para José es siempre elegir ballet.

Conceptos

1. Ganancia Esperada
2. Mejor Respuesta

La mejor forma en que puede jugar un jugador dadas las estrategias seleccionadas de sus oponentes.

Juegos en Forma Normal o Estratégica

Batalla de los sexos

		José	
		ballet	béisbol
María	ballet	$2, 1$	$0, 0$
	béisbol	$0, 0$	$1, 2$

Conceptos

1. Ganancia Esperada
2. Mejor Respuesta
3. Equilibrio de Nash

Cada jugador utiliza una mejor respuesta frente a las estrategias de sus oponentes.

Juegos en Forma Normal o Estratégica

Batalla de los sexos

		José	
		ballet	béisbol
María	ballet	$2, 1$	$0, 0$
	béisbol	$0, 0$	$1, 2$

Conceptos

1. Ganancia Esperada
2. Mejor Respuesta
3. Equilibrio de Nash

Cada jugador utiliza una mejor respuesta frente a las estrategias de sus oponentes.

Juegos en Forma Normal o Estratégica

Batalla de los sexos

		José	
		ballet	béisbol
María	ballet	$2, 1$	$0, 0$
	béisbol	$0, 0$	$1, 2$

María no tiene motivos para cambiar su estrategia.

Conceptos

1. Ganancia Esperada
2. Mejor Respuesta
3. Equilibrio de Nash

Cada jugador utiliza una mejor respuesta frente a las estrategias de sus oponentes.

Juegos en Forma Normal o Estratégica

Batalla de los sexos

		José	
		ballet	béisbol
María	ballet	$2, 1$	$0, 0$
	béisbol	$0, 0$	$1, 2$

José no tiene motivos para cambiar su estrategia.

Conceptos

1. Ganancia Esperada
2. Mejor Respuesta
3. Equilibrio de Nash

Cada jugador utiliza una mejor respuesta frente a las estrategias de sus oponentes.

Juegos en Forma Normal o Estratégica

Batalla de los sexos

		José	
		ballet	béisbol
María	ballet	$2, 1$	$0, 0$
	béisbol	$0, 0$	$1, 2$

Conceptos

1. Ganancia Esperada
2. Mejor Respuesta
3. Equilibrio de Nash

Cada jugador utiliza una mejor respuesta frente a las estrategias de sus oponentes.

Juegos en Forma Normal o Estratégica

Batalla de los sexos

		José	
		ballet	béisbol
María	ballet	$2, 1$	$0, 0$
	béisbol	$0, 0$	$1, 2$

Conceptos

1. Ganancia Esperada
2. Mejor Respuesta
3. Equilibrio de Nash
4. Equilibrio Correlacionado

Puede haber cooperación entre los jugadores.

Juegos en Forma Normal o Estratégica

Batalla de los sexos

		José	
		ballet	béisbol
María	ballet	$2, 1$	$0, 0$
	béisbol	$0, 0$	$1, 2$

Lanzar una moneda

1. cara $\Longrightarrow$ ballet
2. sello $\Longrightarrow$ béisbol

Conceptos

1. Ganancia Esperada
2. Mejor Respuesta
3. Equilibrio de Nash
4. Equilibrio Correlacionado

Puede haber cooperación entre los jugadores.

Equilibrio de Nash

Piedra, papel o tijera

	$\mathcal{R}$ (piedra)	$\mathcal{P}$ (papel)	$\mathcal{S}$ (tijera)
$\mathcal{R}$ (piedra)	0, 0	-1, 1	1,-1
$\mathcal{P}$ (papel)	1,-1	0, 0	-1, 1
$\mathcal{S}$ (tijera)	-1, 1	1,-1	0, 0

Equilibrio de Nash

Piedra, papel o tijera

	$\mathcal{R}$ (piedra)	$\mathcal{P}$ (papel)	$\mathcal{S}$ (tijera)
$\mathcal{R}$ (piedra)	0, 0	-1, 1	1,-1
$\mathcal{P}$ (papel)	1,-1	0, 0	-1, 1
$\mathcal{S}$ (tijera)	-1, 1	1,-1	0, 0

Equilibrio de Nash

Piedra, papel o tijera

	$\mathcal{R}$ (piedra)	$\mathcal{P}$ (papel)	$\mathcal{S}$ (tijera)
$\mathcal{R}$ (piedra)	0, 0	-1, 1	1,-1
$\mathcal{P}$ (papel)	1,-1	0, 0	-1, 1
$\mathcal{S}$ (tijera)	-1, 1	1,-1	0, 0

Equilibrio de Nash

Piedra, papel o tijera

	$\mathcal{R}$ (piedra)	$\mathcal{P}$ (papel)	$\mathcal{S}$ (tijera)
$\mathcal{R}$ (piedra)	0, 0	-1, 1	1,-1
$\mathcal{P}$ (papel)	1,-1	0, 0	-1, 1
$\mathcal{S}$ (tijera)	-1, 1	1,-1	0, 0

Equilibrio de Nash

Piedra, papel o tijera

	$\mathcal{R}$ (piedra)	$\mathcal{P}$ (papel)	$\mathcal{S}$ (tijera)
$\mathcal{R}$ (piedra)	0, 0	-1, 1	1,-1
$\mathcal{P}$ (papel)	1,-1	0, 0	-1, 1
$\mathcal{S}$ (tijera)	-1, 1	1,-1	0, 0

Equilibrio de Nash

Piedra, papel o tijera

	$\mathcal{R}$ (piedra)	$\mathcal{P}$ (papel)	$\mathcal{S}$ (tijera)
$\mathcal{R}$ (piedra)	0, 0	-1, 1	1,-1
$\mathcal{P}$ (papel)	1,-1	0, 0	-1, 1
$\mathcal{S}$ (tijera)	-1, 1	1,-1	0, 0

Equilibrio de Nash

Piedra, papel o tijera

	$\mathcal{R}$ (piedra)	$\mathcal{P}$ (papel)	$\mathcal{S}$ (tijera)
$\mathcal{R}$ (piedra)	0, 0	-1, 1	1,-1
$\mathcal{P}$ (papel)	1,-1	0, 0	-1, 1
$\mathcal{S}$ (tijera)	-1, 1	1,-1	0, 0

Equilibrio de Nash

Piedra, papel o tijera

	$\mathcal{R}$ (piedra)	$\mathcal{P}$ (papel)	$\mathcal{S}$ (tijera)
$\mathcal{R}$ (piedra)	0, 0	-1, 1	1,-1
$\mathcal{P}$ (papel)	1,-1	0, 0	-1, 1
$\mathcal{S}$ (tijera)	-1, 1	1,-1	0, 0

No todos los juegos tienen un equilibrio de Nash en estrategias puras.

Equilibrio de Nash

Piedra, papel o tijera

	$\mathcal{R}$ (piedra)	$\mathcal{P}$ (papel)	$\mathcal{S}$ (tijera)
$\mathcal{R}$ (piedra)	0, 0	-1, 1	1,-1
$\mathcal{P}$ (papel)	1,-1	0, 0	-1, 1
$\mathcal{S}$ (tijera)	-1, 1	1,-1	0, 0

No todos los juegos tienen un equilibrio de Nash en estrategias puras.

Teorema de Nash

Todo juego finito tiene al menos un equilibrio de Nash (en estrategias mixtas).

Juegos de Dos Jugadores de Suma Cero

- Equilibrio de Nash como principal concepto de solución.
- Valor del juego (u): ganancia del primer jugador cuando ambos jugadores utilizan un equilibrio de Nash.
- El juego “batalla de los sexos” no cumple estas condiciones.

Regret Matching: Cálculo del Equilibrio de Nash

Esquema General:

1. Se juega de forma repetida a través del tiempo $t = 1, 2, 3, ...$.
2. A tiempo $t + 1$ cada jugador elige una acción siguiendo una estrategia mixta **determinada**.
3. La **estrategia empírica** converge a un equilibrio de Nash; en la práctica el algoritmo es detenido después de cierto número de iteraciones.

Regret Matching: Cálculo del Equilibrio de Nash

Esquema General:

1. Se juega de forma repetida a través del tiempo $t = 1, 2, 3, \ldots$.
2. A tiempo $t + 1$ cada jugador elige una acción siguiendo una estrategia mixta **determinada**.
3. La **estrategia empírica** converge a un equilibrio de Nash; en la práctica el algoritmo es detenido después de cierto número de iteraciones.

Diferentes formas de calcular la estrategia mixta conducen a diferentes algoritmos:

1. Regret condicional.
2. Regret incondicional.
3. Vector invariante de probabilidad.

Regret Matching: Procedimientos

Regret

Métrica de **arrepentimiento** *de no haber elegido una acción en particular.*

Regret Matching: Procedimientos

Regret

Métrica de **arrepentimiento** *de no haber elegido una acción en particular.*

Tres procedimientos

1. Regret condicional.

$\mathcal{R},\mathcal{S}$	$\mathcal{R},\mathcal{P}$	$\mathcal{S},\mathcal{S}$	$\bar{u}$
1	-1	0	0

Regret Matching: Procedimientos

Regret

Métrica de **arrepentimiento** *de no haber elegido una acción en particular.*

Tres procedimientos

1. Regret condicional.

$\mathcal{R}, \mathcal{S}$	$\mathcal{R}, \mathcal{P}$	$\mathcal{S}, \mathcal{S}$	$\bar{u}$
1	-1	0	0

$\mathcal{S}, \mathcal{S}$	$\mathcal{S}, \mathcal{P}$	$\mathcal{S}, \mathcal{S}$	$\bar{u}$
0	1	0	$\frac{1}{3}$

$$R_1(\mathcal{R}, \mathcal{S}) = \frac{1}{3} - 0 = \frac{1}{3}$$

Regret Matching: Procedimientos

Regret

Métrica de **arrepentimiento** *de no haber elegido una acción en particular.*

Tres procedimientos

1. Regret condicional.

$\mathcal{R},\mathcal{S}$	$\mathcal{R},\mathcal{P}$	$\mathcal{S},\mathcal{S}$	$\bar{u}$
1	-1	0	0

$\mathcal{P},\mathcal{S}$	$\mathcal{P},\mathcal{P}$	$\mathcal{S},\mathcal{S}$	$\bar{u}$
-1	0	0	-$\frac{1}{3}$

$$R_1(\mathcal{R},\mathcal{P}) = -\frac{1}{3} - 0 = -\frac{1}{3}$$

Regret Matching: Procedimientos

Regret

Métrica de **arrepentimiento** *de no haber elegido una acción en particular.*

Tres procedimientos

1. Regret condicional.
2. Regret incondicional.

$\mathcal{R}, \mathcal{S}$	$\mathcal{R}, \mathcal{P}$	$\mathcal{S}, \mathcal{S}$	$\bar{u}$
1	-1	0	0

Regret Matching: Procedimientos

Regret

Métrica de **arrepentimiento** *de no haber elegido una acción en particular.*

Tres procedimientos

1. Regret condicional.
2. Regret incondicional.

$\mathcal{R},\mathcal{S}$	$\mathcal{R},\mathcal{P}$	$\mathcal{S},\mathcal{S}$	$\bar{u}$
1	-1	0	0

$\mathcal{S},\mathcal{S}$	$\mathcal{S},\mathcal{P}$	$\mathcal{S},\mathcal{S}$	$\bar{u}$
0	1	0	$\frac{1}{3}$

$$R_1(\mathcal{S}) = \frac{1}{3} - 0 = \frac{1}{3}$$

Regret Matching: Procedimientos

Regret

Métrica de **arrepentimiento** *de no haber elegido una acción en particular.*

Tres procedimientos

1. Regret condicional.
2. Regret incondicional.

$\mathcal{R},\mathcal{S}$	$\mathcal{R},\mathcal{P}$	$\mathcal{S},\mathcal{S}$	$\bar{u}$
1	-1	0	0

$\mathcal{P},\mathcal{S}$	$\mathcal{P},\mathcal{P}$	$\mathcal{P},\mathcal{S}$	$\bar{u}$
-1	0	-1	-$\frac{2}{3}$

$$R_1(\mathcal{P}) = -\frac{2}{3} - 0 = -\frac{2}{3}$$

Regret Matching: Procedimientos

Regret

Métrica de **arrepentimiento** *de no haber elegido una acción en particular.*

Tres procedimientos

1. Regret condicional.
2. Regret incondicional.
3. Vector invariante de probabilidad.

$\mathcal{R},\mathcal{S}$	$\mathcal{R},\mathcal{P}$	$\mathcal{S},\mathcal{S}$	$\bar{u}$
1	-1	0	0

Regret Matching: Observaciones

1. Las probabilidades son elegidas proporcional a los regrets positivos.
2. El regret positivo tiende a cero cuando el número de juegos tiende a infinito.
3. Si el regret positivo es pequeño, la **estrategia empírica** es una aproximación a un equilibrio de Nash.

Regret Matching: Observaciones

1. Las probabilidades son elegidas proporcional a los regrets positivos.
2. El regret positivo tiende a cero cuando el número de juegos tiende a infinito.
3. Si el regret positivo es pequeño, la **estrategia empírica** es una aproximación a un equilibrio de Nash.

Estrategia Empírica

La probabilidad de que un determinado jugador elija una acción a es igual a:

$$p(a) = \frac{\text{número de veces que el jugador eligió } a}{\text{número total de juegos}}$$

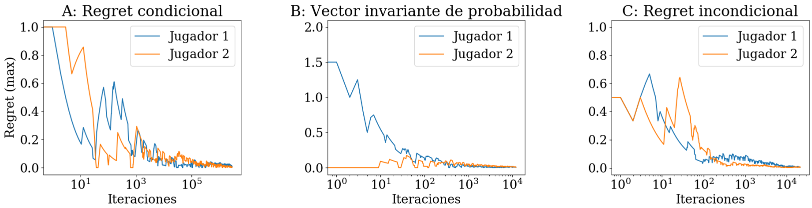
Regret Matching: Experimentos

1. 4 juegos para dos jugadores de suma cero.
2. 10 corridas por cada uno de los juegos y cada uno de los procedimientos.
3. Criterio de parada: regret incondicional menor que 0,005.
4. Gráficas del regret con respecto al número de iteraciones.
5. Verificación realizada con programación lineal.
6. Verificación: la explotabilidad mide la distancia entre la estrategia actual y un equilibrio de Nash.

Piedra, Papel o Tijera

Valor del juego (u): 0.

	A	B	C
Ganancia esperada $u(\sigma)$	$-0,000012$	$0,000004$	$0,000022$
Explotabilidad ε_σ	$0,006$	$0,010$	$0,009$
Tiempo T	$12,198$	$0,345$	$0,049$
Iteraciones I	$4.519.054,1$	$6.601,3$	$19.321,1$
T/I	$2,70\times10^{-6}$	$5,23\times10^{-5}$	$2,54\times10^{-6}$



Matching Pennies

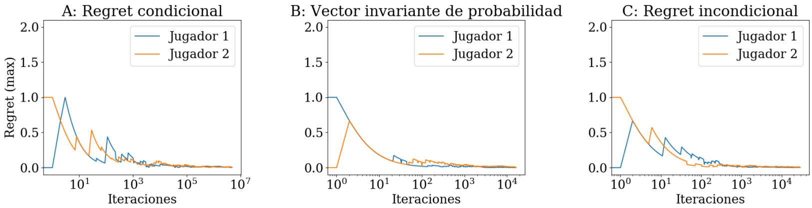
1. Cada jugador posee una moneda y elige cara o sello.
2. Misma elección: gana el primer jugador.
3. Elecciones diferentes: gana el segundo jugador.
4. Tabla de pagos.

	cara	sello
cara	1	-1
sello	-1	1

Matching Pennies

Valor del juego (u): 0.

	A	B	C
Ganancia esperada $u(\sigma)$	$0,000$	$0,000$	$0,000$
Explotabilidad ε_σ	$0,006$	$0,006$	$0,008$
Tiempo T	$10,276$	$0,777$	$0,042$
Iteraciones I	$3.892.550,4$	$25.616,6$	$16.260,5$
T/I	$2,64\times10^{-6}$	$3,03\times10^{-5}$	$2,58\times10^{-6}$



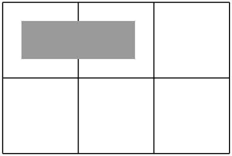
Ficha vs. Dominó

Jugador 1

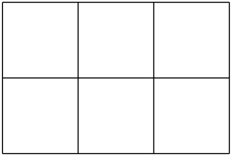
Jugador 2

Ficha vs. Dominó

Jugador 1

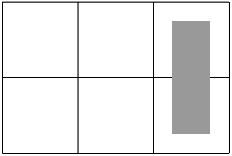


Jugador 2

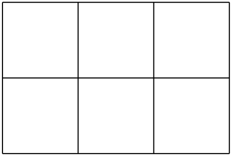


Ficha vs. Dominó

Jugador 1

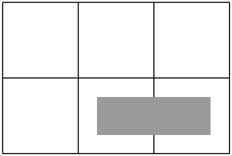


Jugador 2

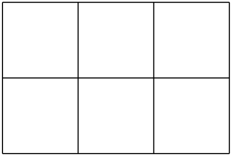


Ficha vs. Dominó

Jugador 1

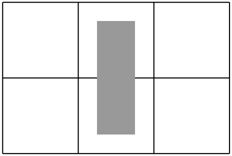


Jugador 2

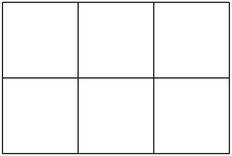


Ficha vs. Dominó

Jugador 1

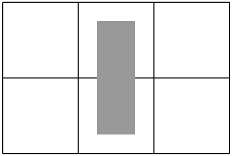


Jugador 2

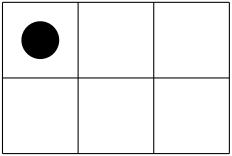


Ficha vs. Dominó

Jugador 1

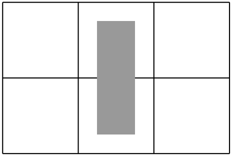


Jugador 2

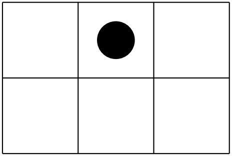


Ficha vs. Dominó

Jugador 1

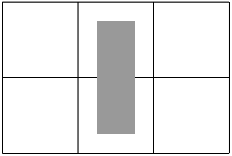


Jugador 2

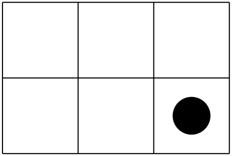


Ficha vs. Dominó

Jugador 1

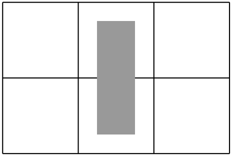


Jugador 2

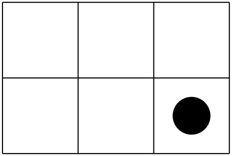


Ficha vs. Dominó

Jugador 1

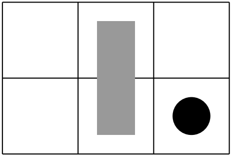


Jugador 2



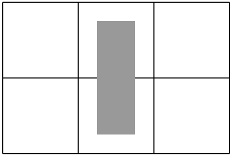
Resultado

1. La ficha y el dominó no se superponen: gana el jugador 1.

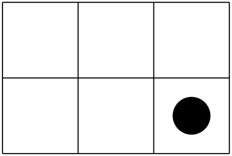


Ficha vs. Dominó

Jugador 1

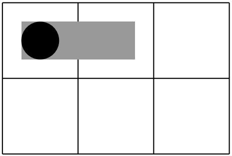


Jugador 2



Resultado

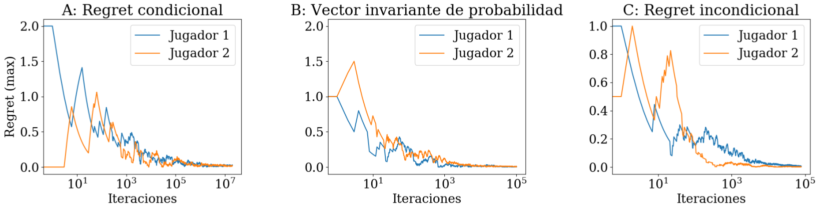
1. La ficha y el dominó no se superponen: gana el jugador 1.
2. La ficha y el dominó sí se superponen: gana el jugador 2.



Ficha vs. Dominó

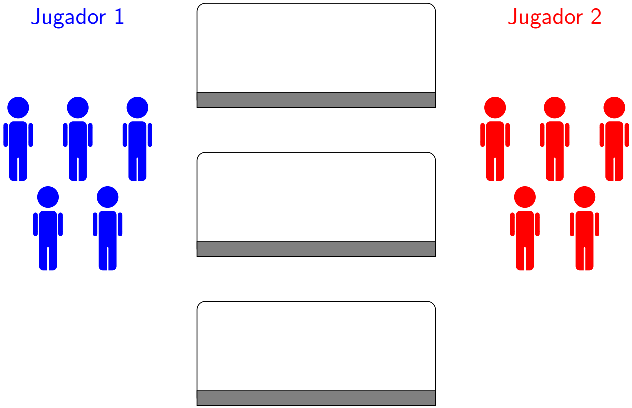
Valor del juego (u): $\frac{1}{3}$.

	A	B	C
Ganancia esperada $u(\sigma)$	$0,333$	$0,334$	$0,334$
Explotabilidad ε_σ	$0,010$	$0,007$	$0,004$
Tiempo T	$319,179$	$11,275$	$0,237$
Iteraciones I	$108.319.272,4$	$75.250,2$	$84.318,5$
T/I	$2,95\times10^{-6}$	$1,50\times10^{-4}$	$2,81\times10^{-6}$



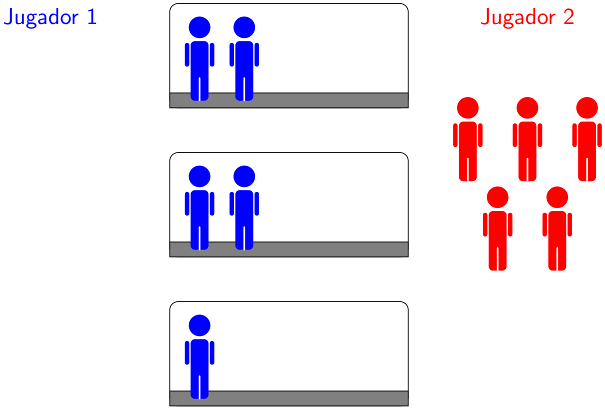
Coronel Blotto

- S soldados por jugador.
- N campos de batalla.



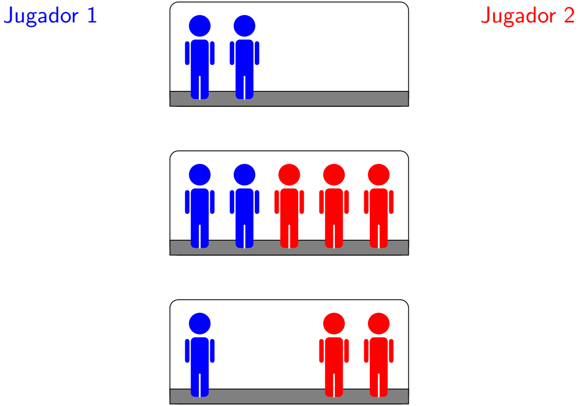
Coronel Blotto

- S soldados por jugador.
- N campos de batalla.



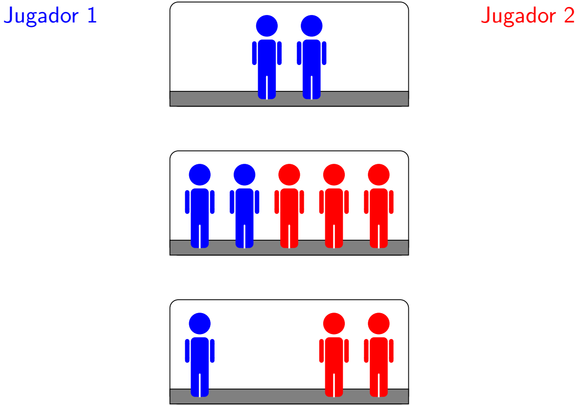
Coronel Blotto

- S soldados por jugador.
- N campos de batalla.



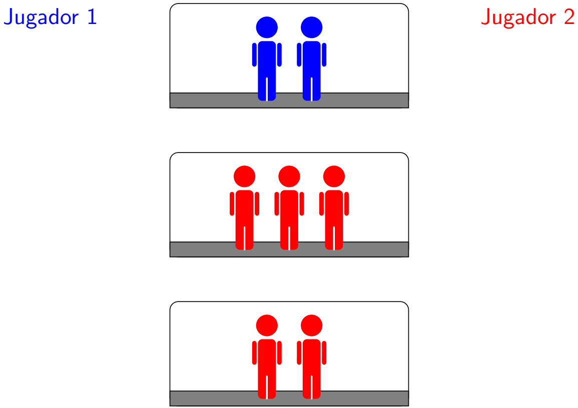
Coronel Blotto

- S soldados por jugador.
- N campos de batalla.



Coronel Blotto

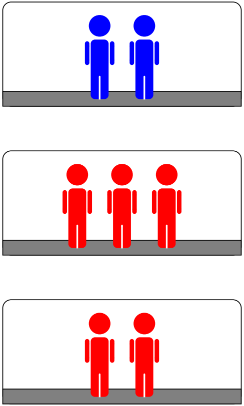
- S soldados por jugador.
- N campos de batalla.



Coronel Blotto

- S soldados por jugador.
- N campos de batalla.
- $u_1 = 1 - 2 = -1$
- $u_2 = 2 - 1 = 1.$

Jugador 1

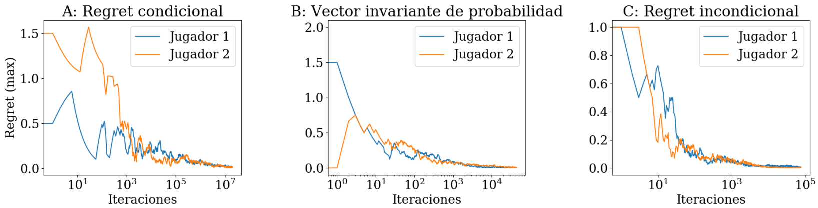


Jugador 2

Coronel Blotto

Valor del juego (u): 0.

	A	B	C
Ganancia esperada $u(\sigma)$	$0,000219$	$0,000150$	$0,000024$
Explotabilidad ε_σ	$0,010$	$0,010$	$0,009$
Tiempo T	$875,533$	$70,453$	$0,166$
Iteraciones I	$190.222.305,3$	$58.794,4$	$48.613,5$
T/I	$4,60\times10^{-6}$	$1,20\times10^{-3}$	$3,41\times10^{-6}$



Forma Normal y Regret Matching: Conclusiones

1. El equilibrio de Nash es un concepto de solución satisfactorio en juegos en forma normal de dos jugadores de suma cero.
2. Los algoritmos de Regret Matching permiten encontrar aproximaciones de un equilibrio de Nash en este tipo de juegos.
3. El procedimiento "regret condicional" necesita mayor número de iteraciones, y por lo tanto mayor tiempo, para converger.
4. El procedimiento "vector invariante de probabilidad" es el procedimiento con las iteraciones más costosas.
5. El procedimiento "regret incondicional" es el más eficiente.

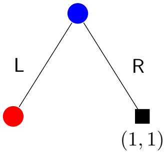
Forma Extensiva

Forma Extensiva: Modelo

Juegos secuenciales

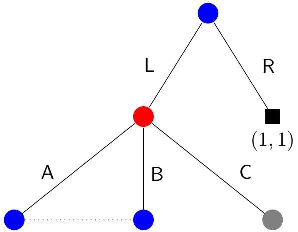
Forma Extensiva: Modelo

Juegos secuenciales



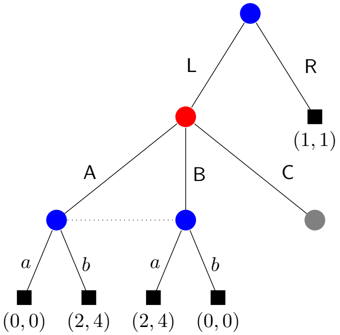
Forma Extensiva: Modelo

Juegos secuenciales



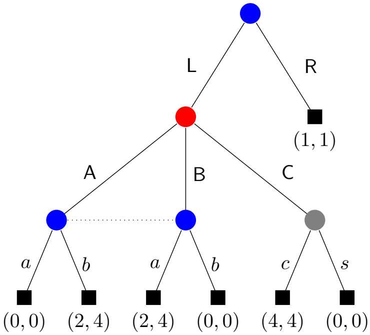
Forma Extensiva: Modelo

Juegos secuenciales



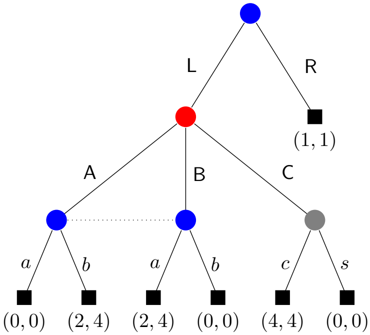
Forma Extensiva: Modelo

Juegos secuenciales



Forma Extensiva: Modelo

Juegos secuenciales

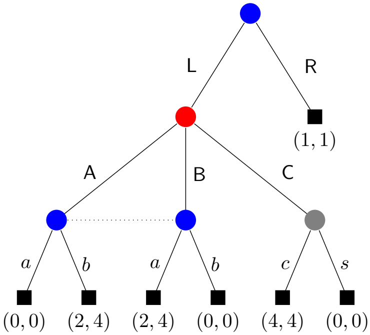


Elementos

1. Historias o nodos.
   Ej: $\emptyset$, LA, LBb, R.

Forma Extensiva: Modelo

Juegos secuenciales

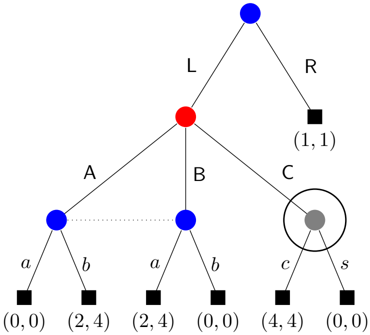


Elementos

1. Historias o nodos.
   Ej: $\emptyset$, LA, LBb, R.
2. Función que asigna a cada historia (nodo) no terminal un jugador.

Forma Extensiva: Modelo

Juegos secuenciales

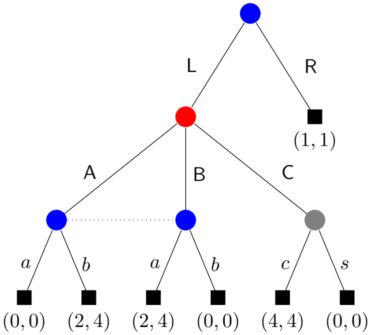


Elementos

1. Historias o nodos.
   Ej: $\emptyset$, LA, LBb, R.
2. Función que asigna a cada historia (nodo) no terminal un jugador.
   - Nodos de azar.

Forma Extensiva: Modelo

Juegos secuenciales

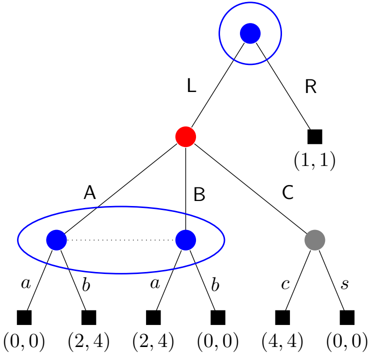


Elementos

1. Historias o nodos.
   Ej: $\emptyset$, LA, LBb, R.
2. Función que asigna a cada historia (nodo) no terminal un jugador.
   - Nodos de azar.
3. Conjuntos de información.

Forma Extensiva: Modelo

Juegos secuenciales

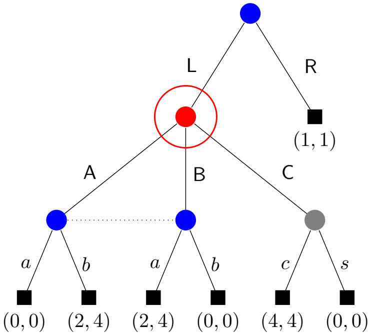


Elementos

1. Historias o nodos.
   Ej: $\emptyset$, LA, LBb, R.
2. Función que asigna a cada historia (nodo) no terminal un jugador.
   - Nodos de azar.
3. Conjuntos de información.

Forma Extensiva: Modelo

Juegos secuenciales

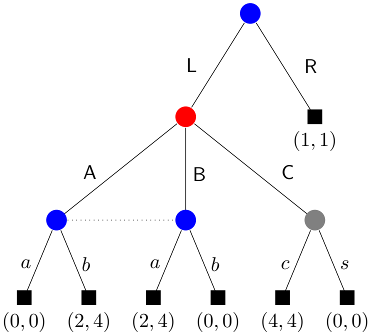


Elementos

1. Historias o nodos.
   Ej: $\emptyset$, LA, LBb, R.
2. Función que asigna a cada historia (nodo) no terminal un jugador.
   - Nodos de azar.
3. Conjuntos de información.

Forma Extensiva: Modelo

Juegos secuenciales

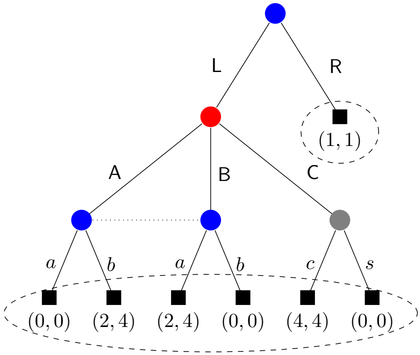


Elementos

1. Historias o nodos.
   Ej: $\emptyset$, LA, LBb, R.
2. Función que asigna a cada historia (nodo) no terminal un jugador.
   - Nodos de azar.
3. Conjuntos de información.
4. Función que asigna por cada historia (nodo) terminal y cada jugador una utilidad.

Forma Extensiva: Modelo

Juegos secuenciales

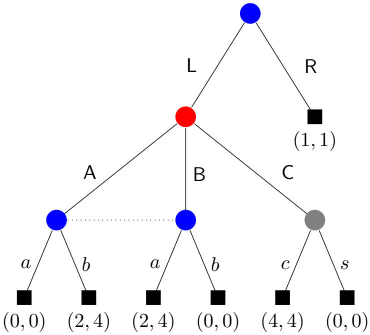


Elementos

1. Historias o nodos.
   Ej: $\emptyset$, LA, LBb, R.
2. Función que asigna a cada historia (nodo) no terminal un jugador.
   - Nodos de azar.
3. Conjuntos de información.
4. Función que asigna por cada historia (nodo) terminal y cada jugador una utilidad.

Forma Extensiva: Modelo

Juegos secuenciales

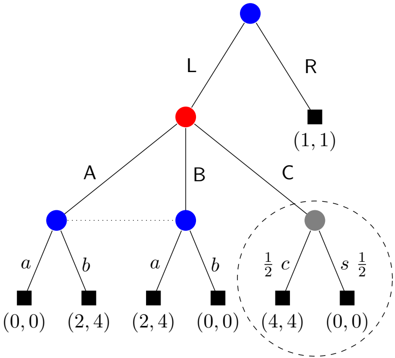


Elementos

1. Historias o nodos.
   Ej: $\emptyset$, LA, LBb, R.
2. Función que asigna a cada historia (nodo) no terminal un jugador.
   - Nodos de azar.
3. Conjuntos de información.
4. Función que asigna por cada historia (nodo) terminal y cada jugador una utilidad.
5. Distribución de probabilidad sobre el conjunto de acciones en cada nodo de azar.

Forma Extensiva: Modelo

Juegos secuenciales

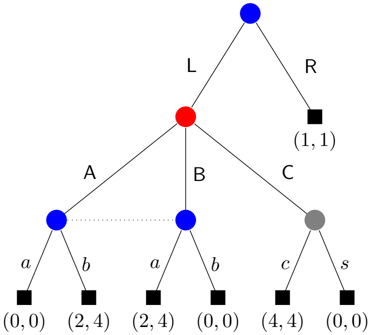


Elementos

1. Historias o nodos.
   Ej: $\emptyset$, LA, LBb, R.
2. Función que asigna a cada historia (nodo) no terminal un jugador.
   - Nodos de azar.
3. Conjuntos de información.
4. Función que asigna por cada historia (nodo) terminal y cada jugador una utilidad.
5. Distribución de probabilidad sobre el conjunto de acciones en cada nodo de azar.

Forma Extensiva: Modelo

Juegos secuenciales



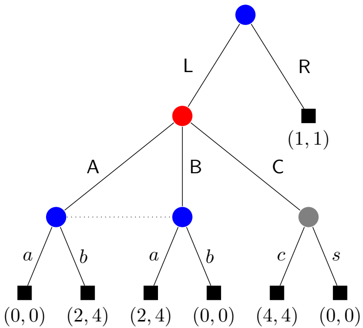
Perfect Recall

Se recuerda de forma perfecta todo lo que se ha visto.

Elementos

1. Historias o nodos.
   Ej: $\emptyset$, LA, LBb, R.
2. Función que asigna a cada historia (nodo) no terminal un jugador.
   - Nodos de azar.
3. Conjuntos de información.
4. Función que asigna por cada historia (nodo) terminal y cada jugador una utilidad.
5. Distribución de probabilidad sobre el conjunto de acciones en cada nodo de azar.

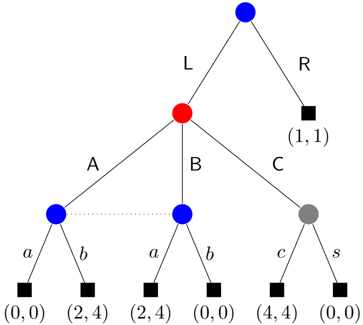
Forma Extensiva: Estrategias



Forma Normal

	A	B	C
(L, a)	$0, 0$	$2, 4$	$2, 2$
(L, b)	$2, 4$	$0, 0$	$2, 2$
(R, a)	$1, 1$	$1, 1$	$1, 1$
(R, b)	$1, 1$	$1, 1$	$1, 1$

Forma Extensiva: Estrategias



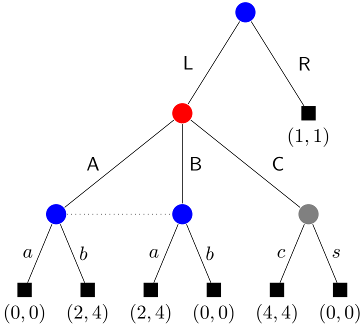
Estrategias

1. Estrategias Puras.

Forma Normal

	A	B	C
(L, a)	$0, 0$	$2, 4$	$2, 2$
(L, b)	$2, 4$	$0, 0$	$2, 2$
(R, a)	$1, 1$	$1, 1$	$1, 1$
(R, b)	$1, 1$	$1, 1$	$1, 1$

Forma Extensiva: Estrategias



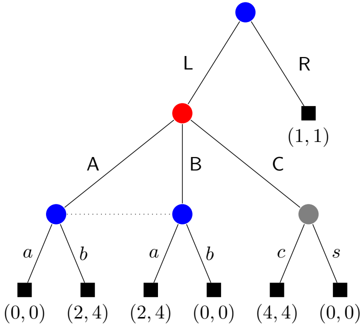
Estrategias

1. Estrategias Puras.

Forma Normal

	A	B	C
(L, a)	$0, 0$	$2, 4$	$2, 2$
(L, b)	$2, 4$	$0, 0$	$2, 2$
(R, a)	$1, 1$	$1, 1$	$1, 1$
(R, b)	$1, 1$	$1, 1$	$1, 1$

Forma Extensiva: Estrategias



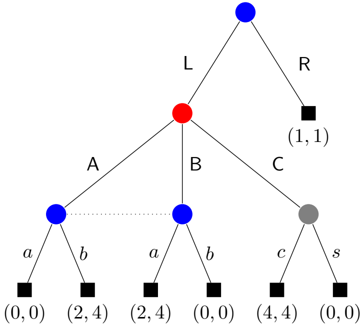
Forma Normal

	A	B	C
(L, a)	$0, 0$	$2, 4$	$2, 2$
(L, b)	$2, 4$	$0, 0$	$2, 2$
(R, a)	$1, 1$	$1, 1$	$1, 1$
(R, b)	$1, 1$	$1, 1$	$1, 1$

Estrategias

1. Estrategias Puras.

Forma Extensiva: Estrategias



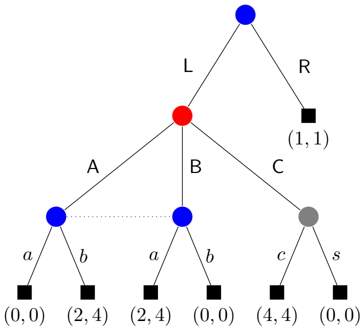
Estrategias

1. Estrategias Puras.
2. Estrategias Mixtas.

Forma Normal

	A	B	C
(L, a)	$0,0$	$2,4$	$2,2$
(L, b)	$2,4$	$0,0$	$2,2$
(R, a)	$1,1$	$1,1$	$1,1$
(R, b)	$1,1$	$1,1$	$1,1$

Forma Extensiva: Estrategias



Estrategias

1. Estrategias Puras.
2. Estrategias Mixtas.

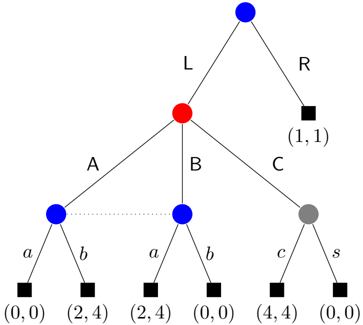
(L, a)	(L, b)	(R, a)	(R, b)
0.45	0.30	0.00	0.25

A	B	C
0.25	0.25	0.50

Forma Normal

	A	B	C
(L, a)	$0, 0$	$2, 4$	$2, 2$
(L, b)	$2, 4$	$0, 0$	$2, 2$
(R, a)	$1, 1$	$1, 1$	$1, 1$
(R, b)	$1, 1$	$1, 1$	$1, 1$

Forma Extensiva: Estrategias



Forma Normal

	A	B	C
(L, a)	$0,0$	$2,4$	$2,2$
(L, b)	$2,4$	$0,0$	$2,2$
(R, a)	$1,1$	$1,1$	$1,1$
(R, b)	$1,1$	$1,1$	$1,1$

Estrategias

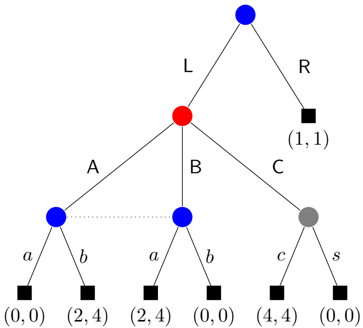
1. Estrategias Puras.
2. Estrategias Mixtas.

(L, a)	(L, b)	(R, a)	(R, b)
0.45	0.30	0.00	0.25

A	B	C
0.25	0.25	0.50

3. Estrategias de Comportamiento.

Forma Extensiva: Estrategias



Forma Normal

	A	B	C
(L, a)	$0, 0$	$2, 4$	$2, 2$
(L, b)	$2, 4$	$0, 0$	$2, 2$
(R, a)	$1, 1$	$1, 1$	$1, 1$
(R, b)	$1, 1$	$1, 1$	$1, 1$

Estrategias

1. Estrategias Puras.
2. Estrategias Mixtas.

(L, a)	(L, b)	(R, a)	(R, b)
0.45	0.30	0.00	0.25

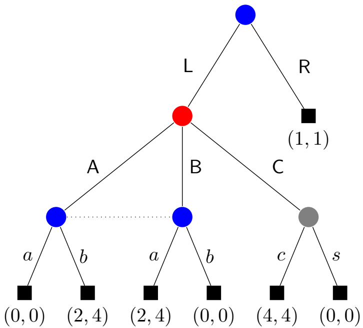
A	B	C
0.25	0.25	0.50

3. Estrategias de Comportamiento.

L	R
0.65	0.35

a	b
0.40	0.60

Forma Extensiva: Estrategias



Forma Normal

	A	B	C
(L, a)	$0, 0$	$2, 4$	$2, 2$
(L, b)	$2, 4$	$0, 0$	$2, 2$
(R, a)	$1, 1$	$1, 1$	$1, 1$
(R, b)	$1, 1$	$1, 1$	$1, 1$

Estrategias

1. Estrategias Puras.
2. Estrategias Mixtas.

(L, a)	(L, b)	(R, a)	(R, b)
0.45	0.30	0.00	0.25

A	B	C
0.25	0.25	0.50

3. Estrategias de Comportamiento.

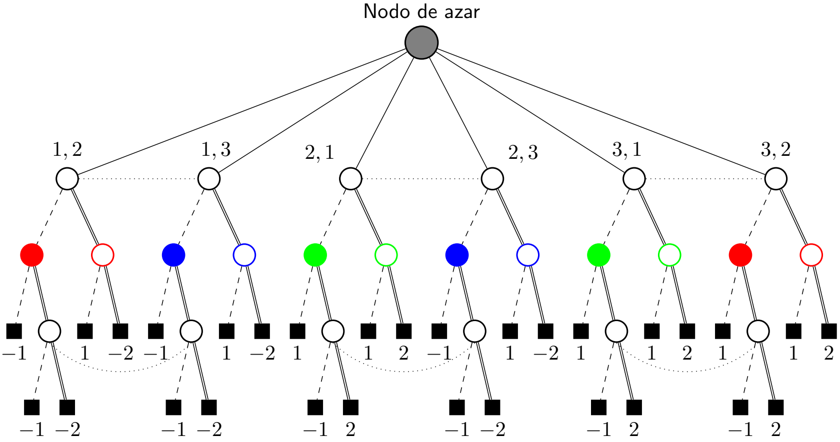
L	R
0.65	0.35

a	b
0.40	0.60

Equilibrio de Nash

Mejor respuesta frente a la estrategias de sus oponentes.

Kuhn Poker


Nodo de azar
1, 2
1, 3
2, 1
2, 3
3, 1
3, 2
−1 1 −2 −1 1 −2 1 1 2 −1 1 −2 1 1 2 1 1 2
−1 −2 −1 −2 −1 2 −1 −2 −1 2 −1 2

Counterfactual Regret Minimization (CFR)

1. El árbol del juego se recorre repetidamente.
2. Inicia con una estrategia uniforme en cada conjunto de información.
3. En cada iteración se elige una mejor estrategia revisando las decisiones pasadas, utilizando las métricas de regret.
4. La estrategia promedio converge a un equilibrio de Nash; en la práctica el algoritmo es detenido después de cierto número de iteraciones.

Counterfactual Regret Minimization (CFR)

1. El árbol del juego se recorre repetidamente.
2. Inicia con una estrategia uniforme en cada conjunto de información.
3. En cada iteración se elige una mejor estrategia revisando las decisiones pasadas, utilizando las métricas de regret.
4. La estrategia promedio converge a un equilibrio de Nash; en la práctica el algoritmo es detenido después de cierto número de iteraciones.

¿Cómo se mejora la estrategia en cada iteración?

- Sumar el regret contrafactual que se tiene en cada conjunto de información por cada acción.
- **Regret contrafactual**: cuánto mejor lo habría hecho en todos los juegos hasta ahora si siempre hubiera jugado esta acción en este conjunto de información.
- **Regret Matching**: en la nueva estrategia las acciones son elegidas con probabilidades proporcionales a los regrets positivos.

CFR: Experimentos

1. CFR con muestreo en los nodos de azar.
2. Implementación propia del algoritmo.
   - Input: definición del juego.
   - **Definición del juego**: implementación de las funciones que permiten recorrer el árbol de forma ímplicita.
3. Tres clases de juegos, cada uno con diferentes parámetros.
4. 10 horas de entrenamiento.
5. Gráfica del regret con respecto al número de iteraciones.
6. Verificación: la explotabilidad mide la distancia entre la estrategia actual y un equilibrio de Nash.
7. Un juego se considera resuelto si la explotabilidad es menor que el 1% de la mínima ganancia positiva posible.

One Card Poker (OCP)

- Generalización del Juego Kuhn Poker.
- N: número de cartas.
- OCP(N).

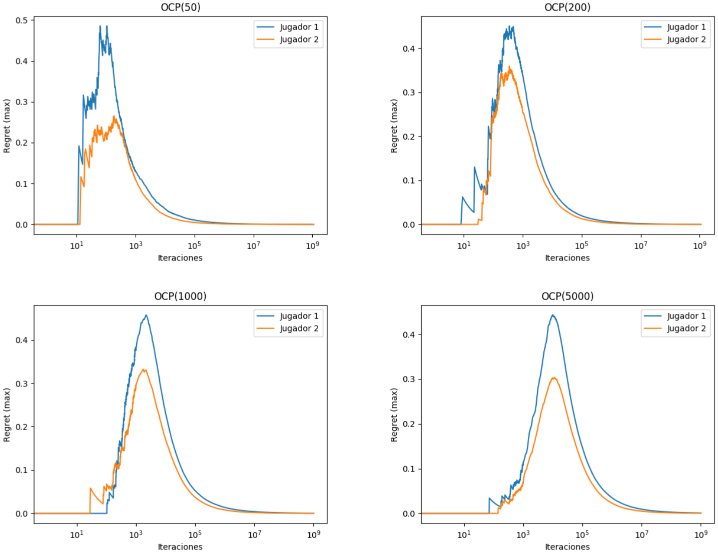
One Card Poker (OCP)

- Generalización del Juego Kuhn Poker.
- N: número de cartas.
- OCP(N).

Tabla de Resultados

Juego	N	I	Iteraciones	$u(\sigma)$	ε_σ (%)	Resuelto
OCP(3)	55	12	1.181.763.638	-0,056	0,0098	✓
OCP(12)	1.189	48	1.147.919.240	-0,062	0,0032	✓
OCP(50)	22.051	200	1.145.291.974	-0,058	0,0099	✓
OCP(200)	358.201	800	1.128.993.847	-0,056	0,0078	✓
OCP(1000)	8.991.001	4.000	1.087.573.694	-0,056	0,0098	✓
OCP(5000)	224.955.001	20.000	1.038.367.354	-0,056	0,0241	✓

One Card Poker (OCP)



Dudo

- Juego de dados y apuestas.
- K: número de caras de los dados.
- D_1, D_2: número de dados del primer y segundo jugador, respectivamente.
- Dudo(K, D_1, D_2).

Dudo

- Juego de dados y apuestas.
- K: número de caras de los dados.
- D_1, D_2: número de dados del primer y segundo jugador, respectivamente.
- Dudo(K, D_1, D_2).

Tabla de Resultados

Juego	N	I	Iteraciones	$u(\sigma)$	ε_σ (%)	Resuelto
Dudo$(4,1,1)$	8.177	512	18.697.532	-0,125	0,0259	✓
Dudo$(4,1,2)$	327.641	14.366	1.215.600	-0,508	0,0971	✓
Dudo$(4,2,1)$	327.641	14.366	1.213.799	0,552	0,3701	✓
Dudo$(4,2,2)$	13.107.101	327.680	63.109	0,0069	2,1132	✗
Dudo$(5,1,1)$	51.176	2.560	4.521.208	-0,120	0,1186	✓
Dudo$(5,1,2)$	4.915.126	163.840	151.235	-0,565	0,6197	✓
Dudo$(5,2,1)$	4.915.126	163.840	143.698	0,581	0,0122	✓
Dudo$(5,2,2)$	471.858.976	7.864.320	3.826	0,836	15,1963	✗
Dudo$(6,1,1)$	294.877	12.288	1.067.782	-0,111	0,0975	✓
Dudo$(6,1,2)$	66.060.163	1.769.472	17.702	-0,593	4,5781	✗
Dudo$(6,2,1)$	66.060.163	1.769.472	17.221	0,592	3,9594	✗

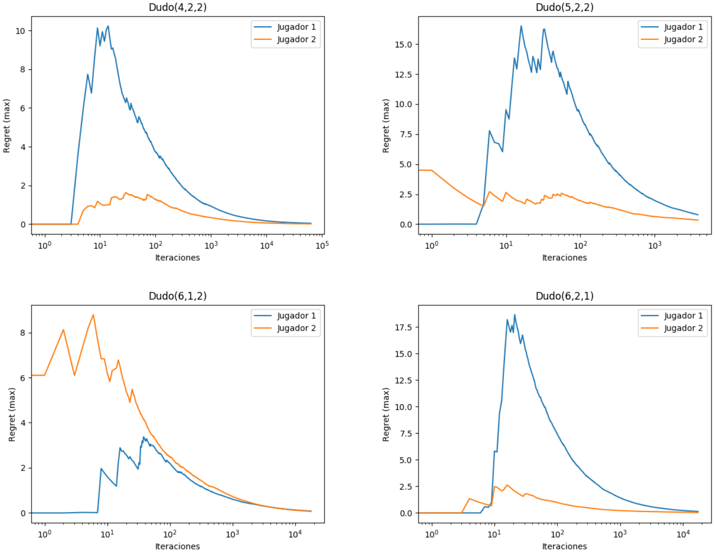
Dudo

- Juego de dados y apuestas.
- K: número de caras de los dados.
- D_1, D_2: número de dados del primer y segundo jugador, respectivamente.
- Dudo(K, D_1, D_2).

Tabla de Resultados

Juego	N	I	Iteraciones	$u(\sigma)$	ε_σ (%)	Resuelto
Dudo$(4,1,1)$	8.177	512	18.697.532	-0,125	0,0259	✓
Dudo$(4,1,2)$	327.641	14.366	1.215.600	-0,508	0,0971	✓
Dudo$(4,2,1)$	327.641	14.366	1.213.799	0,552	0,3701	✓
Dudo$(4,2,2)$	13.107.101	327.680	63.109	0,0069	2,1132	✗
Dudo$(5,1,1)$	51.176	2.560	4.521.208	-0,120	0,1186	✓
Dudo$(5,1,2)$	4.915.126	163.840	151.235	-0,565	0,6197	✓
Dudo$(5,2,1)$	4.915.126	163.840	143.698	0,581	0,0122	✓
Dudo$(5,2,2)$	471.858.976	7.864.320	3.826	0,836	15,1963	✗
Dudo$(6,1,1)$	294.877	12.288	1.067.782	-0,111	0,0975	✓
Dudo$(6,1,2)$	66.060.163	1.769.472	17.702	-0,593	4,5781	✗
Dudo$(6,2,1)$	66.060.163	1.769.472	17.221	0,592	3,9594	✗

Dudo



Dudo(4,2,2)
Jugador 1
Jugador 2
Regret (max)
Iteraciones
Dudo(5,2,2)
Jugador 1
Jugador 2
Regret (max)
Iteraciones
Dudo(6,1,2)
Jugador 1
Jugador 2
Regret (max)
Iteraciones
Dudo(6,2,1)
Jugador 1
Jugador 2
Regret (max)
Iteraciones

Dominó

- Versión para dos jugadores.
- M: máximo número de puntos en una cara de una pieza.
- N: número de piezas de la mano inicial para cada jugador.
- Domino(M, N).

Dominó

- Versión para dos jugadores.
- M: máximo número de puntos en una cara de una pieza.
- N: número de piezas de la mano inicial para cada jugador.
- Domino(M, N).

Tabla de Resultados

Juego	N	I	Iteraciones	$u(\sigma)$	ε_σ (%)	Resuelto
Domino$(2, 2)$	7.321	102	540.186.366	2,4000	0,0000	✓
Domino$(3, 2)$	46.534.657	88.947	400.047.334	2,8767	0,0315	✓
Domino$(3, 3)$	246.760.993	107.854	72.492.951	2,1539	0,3854	✓
Domino$(3, 4)$	1.547.645.185	104.050	11.213.463	3,2034	1,4871	✗

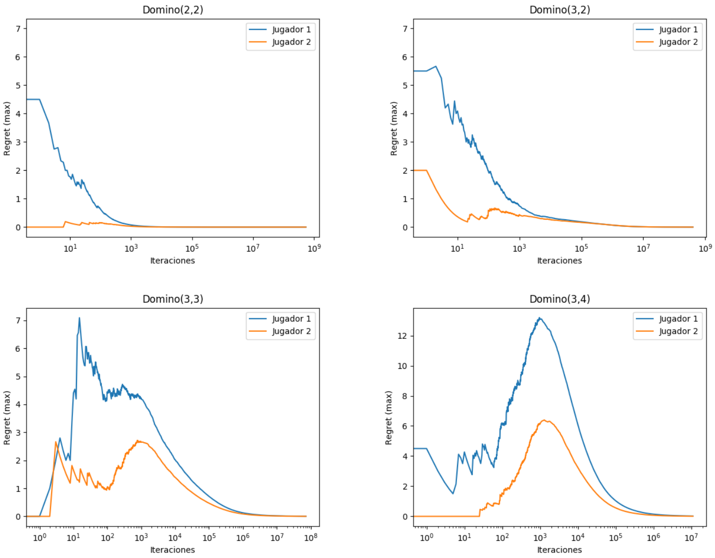
Dominó

- Versión para dos jugadores.
- M: máximo número de puntos en una cara de una pieza.
- N: número de piezas de la mano inicial para cada jugador.
- Domino(M, N).

Tabla de Resultados

Juego	N	I	Iteraciones	$u(\sigma)$	ε_σ (%)	Resuelto
Domino$(2,2)$	7.321	102	540.186.366	2,4000	0,0000	✓
Domino$(3,2)$	46.534.657	88.947	400.047.334	2,8767	0,0315	✓
Domino$(3,3)$	246.760.993	107.854	72.492.951	2,1539	0,3854	✓
Domino$(3,4)$	1.547.645.185	104.050	11.213.463	3,2034	1,4871	✗

Dominó



Domino(2,2)
Domino(3,2)
Domino(3,3)
Domino(3,4)
Jugador 1
Jugador 2
Regret (max)
Iteraciones

Experimentos Adicionales

- Juegos no resueltos con 10 horas de entrenamiento.
- 200 horas de entrenamiento.

Experimentos Adicionales

- Juegos no resueltos con 10 horas de entrenamiento.
- 200 horas de entrenamiento.

Tabla de Resultados

Juego	N	I	Iteraciones	$u(\sigma)$	ε_σ (%)	Resuelto
Dudo$(4,2,2)$	13.107.101	327.680	2.276.259	0,00875	0,2382	✓
Dudo$(5,2,2)$	471.858.976	7.864.320	133.863	-0,00004	1,7695	✗
Dudo$(6,1,2)$	66.060.163	1.769.472	543.485	-0,597	0,5102	✓
Dudo$(6,2,1)$	66.060.163	1.769.472	513.786	0,597	0,6727	✓
Domino$(3,4)$	1.547.645.185	104.050	365.484.932	3.2027	0,1812	✓

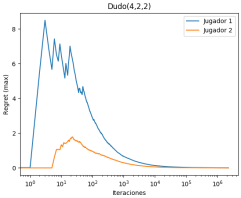
Experimentos Adicionales

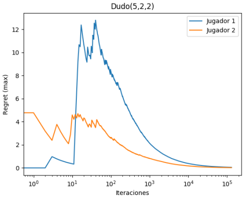
- Juegos no resueltos con 10 horas de entrenamiento.
- 200 horas de entrenamiento.

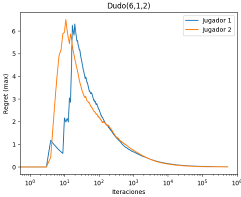
Tabla de Resultados

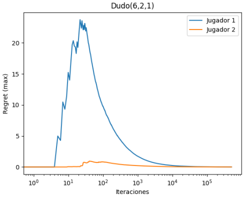
Juego	N	I	Iteraciones	$u(\sigma)$	ε_σ (%)	Resuelto
Dudo$(4,2,2)$	13.107.101	327.680	2.276.259	0,00875	0,2382	✓
Dudo$(5,2,2)$	471.858.976	7.864.320	133.863	-0,00004	1,7695	✗
Dudo$(6,1,2)$	66.060.163	1.769.472	543.485	-0,597	0,5102	✓
Dudo$(6,2,1)$	66.060.163	1.769.472	513.786	0,597	0,6727	✓
Domino$(3,4)$	1.547.645.185	104.050	365.484.932	3.2027	0,1812	✓

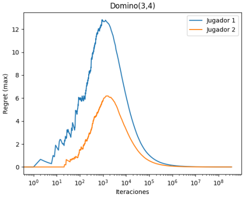
Experimentos Adicionales











CFR: Conclusiones

1. La forma extesiva es un modelo adecuado para representar juegos secuenciales con información incompleta y no determinismo.
2. El algoritmo CFR permite encontrar un equilibrio de Nash en juegos en forma extensiva de dos jugadores de suma cero.
3. Se resolvieron diversos juegos captados por el modelos:
   - OCP: el segundo tiene ventaja sobre el primer jugador: $u \in (-0.7, -0.5)$.
   - Dudo: el jugador con mayor número de dados tiene ventaja.
   - Domino: el primer jugador tiene ventaja sobre el segundo jugador: $u \in (2, 3.5)$.

Conclusiones y Recomendaciones

Conclusiones

1. Los modelos utilizados son adecuados para los juegos planteados.
2. El equilibrio de Nash es un concepto de solución satisfactorio en juegos de dos jugadores de suma cero. No lo es cuando el juego no es de suma cero o tiene más de dos jugadores.
3. Se utilizó la explotabilidad como métrica para medir la distancia entre la estrategia obtenida y un equilibrio de Nash.
4. Resolución de juegos: se encontraron aproximaciones con una explotabilidad no mayor que el 1% de la mínima ganancia positiva posible.

Conclusiones y Recomendaciones

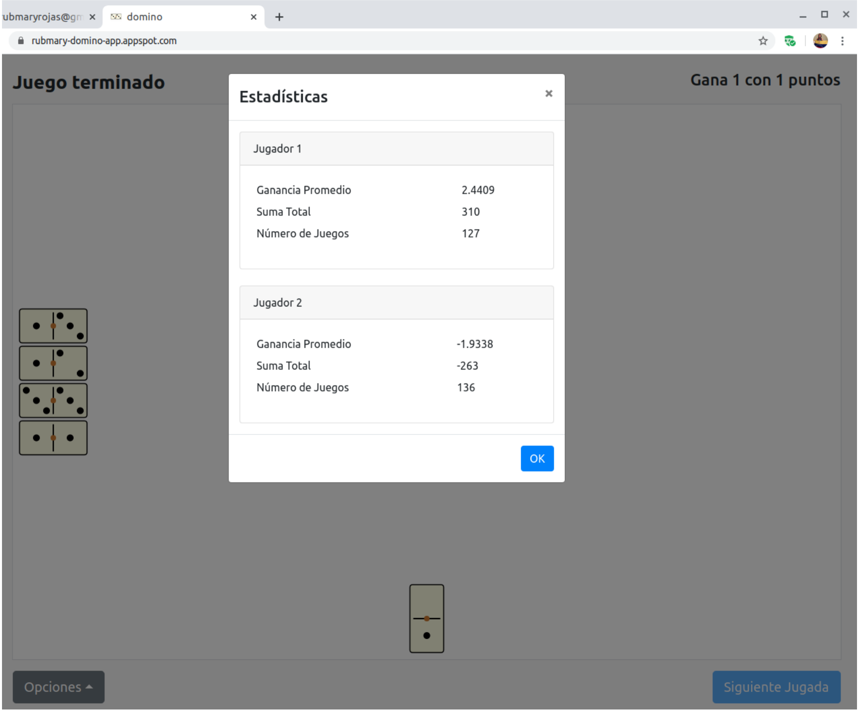
Conclusiones

1. Los modelos utilizados son adecuados para los juegos planteados.
2. El equilibrio de Nash es un concepto de solución satisfactorio en juegos de dos jugadores de suma cero. No lo es cuando el juego no es de suma cero o tiene más de dos jugadores.
3. Se utilizó la explotabilidad como métrica para medir la distancia entre la estrategia obtenida y un equilibrio de Nash.
4. Resolución de juegos: se encontraron aproximaciones con una explotabilidad no mayor que el 1% de la mínima ganancia positiva posible.

Recomendaciones

1. Resolver instancias mayores del juego de dominó para 2 personas considerando abstracciones.
2. Experimentos sobre el juego para 4 personas considerando cada pareja como un único jugador.

Demo



ubmaryrojas@gm
domino
rubmary-domino-app.appspot.com
Juego terminado
Gana 1 con 1 puntos
Estadísticas
Jugador 1
Ganancia Promedio 2.4409
Suma Total 310
Número de Juegos 127
Jugador 2
Ganancia Promedio -1.9338
Suma Total -263
Número de Juegos 136
OK
Opciones
Siguiente Jugada

Gracias por su atención

¿Preguntas?

Anexos

Dilema del prisionero.

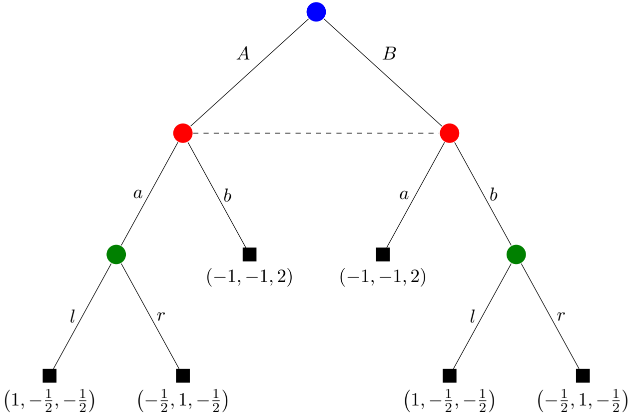
(Albert W. Tucker, 1950) Dos prisionerso son sospechosos de un crimen y se interrogan en celdas separadas.

- Si los dos sospechosos se acusan entre entre sí ambos reciben una pena de 9 años de prisión.
- Si ninguno acusa al otro, la sentencia será de 1 a no de prisión.
- Si uno de ellos acusa al otro y el otro no lo hace, quien acusó queda en libertad y el otro prisionero es condenado a 15 años de prisión.

	Acusar	No Acusar
Acusar	$-1, -1$	$-15, 0$
No Acusar	$0, -15$	$-9, -9$

Si el juego no es de suma cero, el equilibrio de Nash no es un concepto de solución satisfactorio.

Ejemplo de Juego con 3 Jugadores



- Los jugadores 1 y 2 maximizan su ganancia cuando cooperan.
- Cuando los jugadores 1 y 2 eligen la misma acción el tercer jugador siempre pierde, pero su decisión afecta las ganancias de los otros 2 jugadores.

Aplicaciones: Seguro de desempleo

(Rasmusen, 2000) Suponga un individuo. Él está desempleado y seguirá sin buscar empleo mientras el gobierno le ayude a mantener su estilo de vida. El gobierno tratará de darle una ayuda de forma que el individuo esté incentivado a buscar un empleo.

	Buscar Trabajo	Descansar
Ayudar	$5, 4$	$-3, 5$
No Ayudar	$-3, 3$	$0, 0$

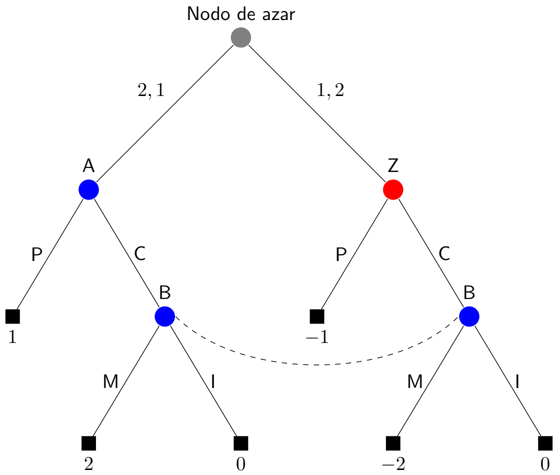
Aplicaciones: Monopolio Natural

En este juego existen dos empresas que participan en un monopolio natural. Es decir, el mercado no es lo suficientemente grande para soportar dos empresas. Cada una de las empresas tiene la opción de salir o permanecer en el mercado.

	Salir	Permanecer
Salir	$0, 0$	$0, 4$
Permanecer	$4, 0$	$-2, -2$

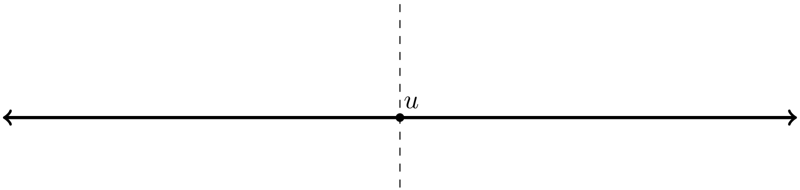
Perfect Recall

1. El jugador recuerda lo que sabía.
2. El jugador recuerda lo que eligió.



Explotabilidad

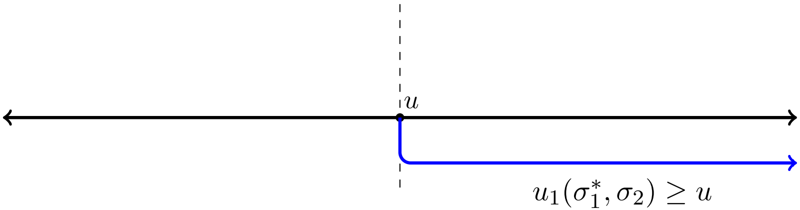
Equilibrio de Nash $\sigma^* = (\sigma_1^*, \sigma_2^*)$



Explotabilidad

Equilibrio de Nash $\sigma^* = (\sigma_1^*, \sigma_2^*)$

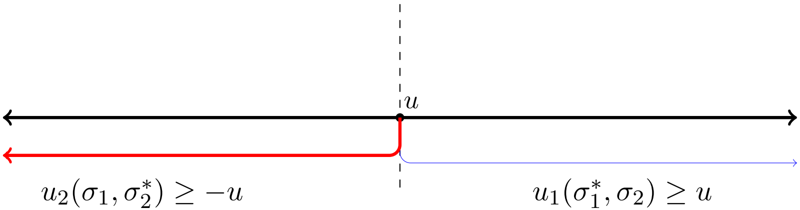
1. Primer jugador garantiza una ganancia esperada de al menos u.



Explotabilidad

Equilibrio de Nash $\sigma^* = (\sigma_1^*, \sigma_2^*)$

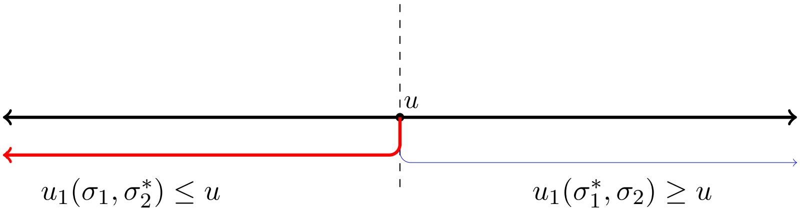
1. Primer jugador garantiza una ganancia esperada de al menos u.
2. Segundo jugador garantiza una ganancia esperada de al menos $-u$.



Explotabilidad

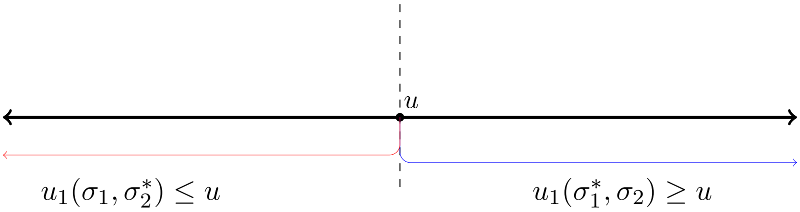
Equilibrio de Nash $\sigma^* = (\sigma_1^*, \sigma_2^*)$

1. Primer jugador garantiza una ganancia esperada de al menos u.
2. Segundo jugador garantiza una ganancia esperada de a lo sumo u para el primer jugador.



Explotabilidad

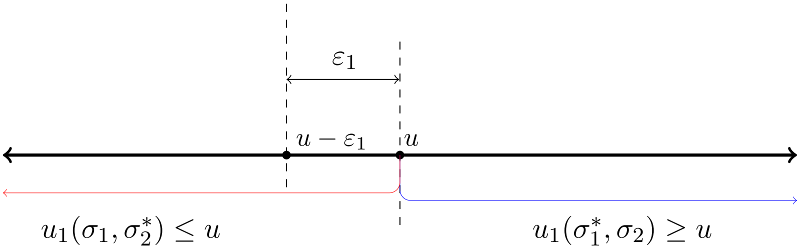
Aproximación $\sigma' = (\sigma_1', \sigma_2')$



Explotabilidad

Aproximación $\sigma' = (\sigma_1', \sigma_2')$

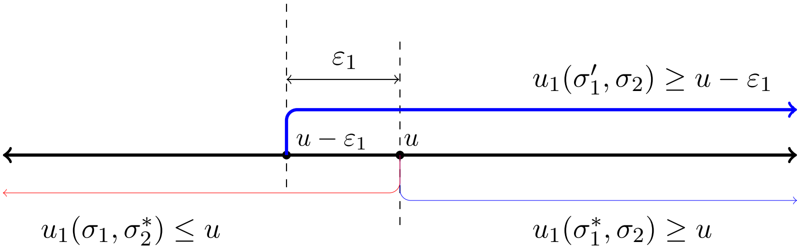
1. Primer jugador garantiza una ganancia esperada de al menos $u - \varepsilon_1$.



Explotabilidad

Aproximación $\sigma' = (\sigma'_1, \sigma'_2)$

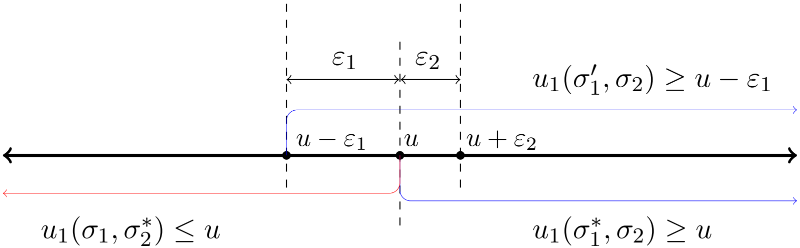
1. Primer jugador garantiza una ganancia esperada de al menos $u - \varepsilon_1$.



Explotabilidad

Aproximación $\sigma' = (\sigma'_1, \sigma'_2)$

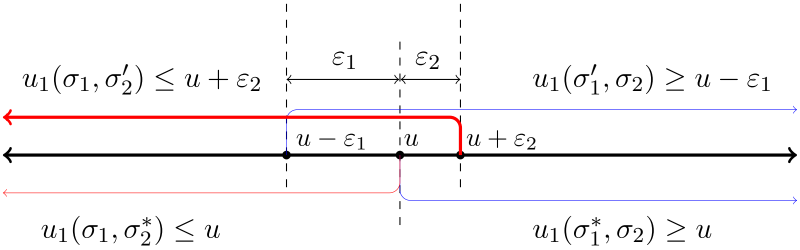
1. Primer jugador garantiza una ganancia esperada de al menos $u - \varepsilon_1$.
2. Segundo jugador garantiza una ganancia esperada de a lo sumo $u + \varepsilon_2$ para el primer jugador.



Explotabilidad

Aproximación $\sigma' = (\sigma'_1, \sigma'_2)$

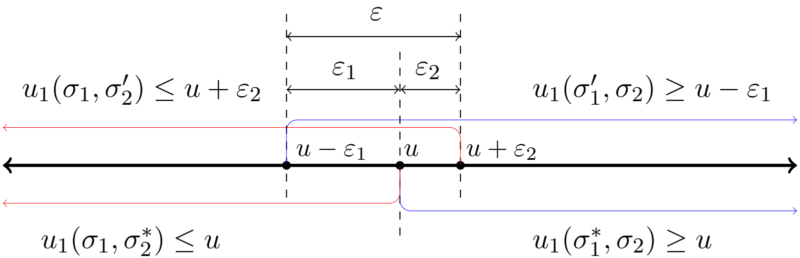
1. Primer jugador garantiza una ganancia esperada de al menos $u - \varepsilon_1$.
2. Segundo jugador garantiza una ganancia esperada de a lo sumo $u + \varepsilon_2$ para el primer jugador.



Explotabilidad

Aproximación $\sigma' = (\sigma'_1, \sigma'_2)$

1. Primer jugador garantiza una ganancia esperada de al menos $u - \varepsilon_1$.
2. Segundo jugador garantiza una ganancia esperada de a lo sumo $u + \varepsilon_2$ para el primer jugador.



$$\varepsilon = \varepsilon_1 + \varepsilon_2$$

Definiciones y Teoremas

Juegos en Forma Normal o Estratégica I

Definición

Un juego de N personas en **forma normal** *(o estratégica) es una tupla* $\Gamma = (N, (S_i)_{i \in N}, (u_i)_{i \in N})$*, donde:*

- $N = \{1, 2, \ldots, N\}$ *es el conjunto de jugadores.*
- S_i *es el conjunto de* **estrategias puras** *(o acciones) del jugador* i.
- $u_i : \Pi_{i \in N} S_i \rightarrow \mathbb{R}$ *es la* **función de pago** *del jugador* i.

Definición

Un **perfil estratégico** *(o perfil de acción) es una* N*-tupla formada por una estrategia pura para cada jugador.* $S = \Pi_{i \in N} S_i$ *es el conjunto de perfiles estratégicos y* $s = (s_i)_{i \in N}$ *representa un elemento genérico de* S.

Juegos en Forma Normal o Estratégica II

Definición

Una **estrategia mixta** *del jugador i, denotada con σ_i, es una distribución de probabilidad sobre el conjunto S_i; es decir $\sigma_i \in \Delta(S_i)$. Se denota con $\sigma_i(s_i)$ la probabilidad de que el jugador i elija la acción $s_i \in S_i$.*

Definición

El **soporte** *(support) de una estrategia mixta $\sigma_i \in \Delta(S_i)$ del jugador i es el conjunto de estrategias puras con una probabilidad positiva de ser elegidas:*

$$\textit{support}(\sigma_i) \;=\; \{s_i : \sigma_i(s_i) > 0\}\,.$$

Definición

Un **perfil estratégico mixto** *σ consiste en una estrategia mixta para cada jugador; es decir, $\sigma \in \Pi_{i\in N}\Delta(S_i)$ es una tupla de forma $\sigma = (\sigma_i)_{i\in N}$.*

Juegos en Forma Normal o Estratégica III

Definición

La **ganancia esperada** *del jugador i dado un perfil estratégico mixto σ es*

$$\begin{aligned} u_i(\sigma) \ &= \ \sum_{s\in S} u_i(s)\sigma(s) = \sum_{s\in S} u_i(s) \prod_{j\in N} \sigma_j(s_j) \\ &= \ \sum_{s\in S} u_i(s)\sigma_i(s_i)\sigma_{-i}(s_{-i})\,. \end{aligned}$$

Teorema

La ganancia esperada $u_i(\sigma)$ del jugador i dado el perfil estratégico σ satisface:

$$u_i(\sigma) \ = \ \sum_{s_i\in S_i} \sigma_i(s_i) \sum_{s_{-i}\in S_{-i}} \sigma_{-i}(s_{-i})u_i(s_i, s_{-i})\,.$$

Juegos en Forma Normal o Estratégica IV

Definición

Sea $i \in N$ *un jugador,* σ_i *una estrategia mixta para el jugador* i*, y* σ_{-i} *un perfil estratégico mixto para el resto de los jugadores. Se dice que* σ_i *es una* **mejor respuesta** *con respecto a* σ_{-i} *si y sólo si* $u_i(\sigma_i, \sigma_{-i}) \geq u_i(\sigma_i', \sigma_{-i})$ *para toda estrategia mixta* σ_i' *para el jugador* i.

Definición

Sea σ_i^* *una estrategia mixta para el jugador* i *que es mejor respuesta a* σ_{-i}*. Cualquier estrategia mixta* σ_i *para el jugador* i *cuyo soporte sea un subconjunto del soporte de* σ_i^* *es también una mejor respuesta a* σ_{-i}.

Juegos en Forma Normal o Estratégica V

Definición

Un perfil estratégico mixto σ *es un* **equilibrio de Nash** *si y sólo si para todo jugador* i*, la estrategia* σ_i *es mejor respuesta del jugador* i *para* σ_{-i}*.*

Teorema

Todo juego finito tiene al menos un equilibrio de Nash.

Definición

Una distribución $\psi \in \Delta(S)$ *es un* **equilibrio correlacionado** *si y sólo si para cualquier jugador* i*, y para cualesquiera estrategias puras* $x, y \in S_i$*,*

$$\sum_{s_{-i} \in S_{-i}} \psi(x, s_{-i})[u_i(x, s_{-i}) - u_i(y, s_{-i})] \ \geq \ 0\,.$$

Juegos en Forma Normal o Estratégica VI

Teorema

Si σ es un equilibrio de Nash, entonces σ es un equilibrio correlacionado.

Teorema

Sea $\psi \in \Delta(S)$ un equilibrio correlacionado. Si ψ se factoriza como $\psi = \prod_{i \in N} \sigma_i$ donde $\{\sigma_i\}_{i \in N}$ es un conjunto de estrategias mixtas para cada jugador (i.e., $\psi(s) = \prod_{i \in N} \sigma_i(s_i)$ para todo $s \in S$), entonces ψ es un equilibrio de Nash.

Teorema

Sean σ y σ' dos equilibrios correlacionados, y α un número real en $(0, 1)$. Entonces, la distribución $\alpha\sigma + (1 - \alpha)\sigma'$ es un equilibrio correlacionado.

Juegos en Forma Extesiva I

Definición I

Un juego finito en **forma extensiva** *con* **información incompleta** *tiene los siguientes componentes:*

- *Un conjunto finito* N *de* **jugadores**.
- *Un conjunto finito* H *de secuencias, las posible* **historias** *de acciones, tal que la secuencia vacía está en* H*, y cada prefijo de una secuencia en* H *también está en* H*.* $Z \subseteq H$ *son las historias terminales (aquellas que no son prefijo de ninguna otra secuencia).* $A(h) = \{a : (h, a) \in H\}$ *son las acciones disponibles después de una historia no terminal* $h \in H$*. Si la historia* h *es prefijo de la historia* h' *escribimos* $h \sqsubseteq h'$*, y si* h *es un prefijo propio de* h' *escribimos* $h \sqsubset h'$*.*

Juegos en Forma Extesiva II

Definición II

- *Una función P que asigna a cada historia no terminal (cada elemento de $H \setminus Z$) un elemento de $N \cup \{c\}$. P es la* **función de jugador**. *$P(h)$ es el jugador que toma una acción después de la historia h. Si $P(h) = c$ entonces la acción tomada después de la historia h es determinada por el azar. Este tipo de nodos serán denominados* **nodos de azar**.
- *Una función f_c que asocia con cada historia h, para la cual $P(h) = c$, una medida de probabilidad $f_c(\cdot|h)$ sobre $A(h)$: $f_c(a|h)$ es la probabilidad de que la acción a ocurra dado h. Cada medida de probabilidad es independiente de cualquier otra de estas medidas.*

Juegos en Forma Extesiva III

Definición III

- *Para cada jugador* $i \in N$*, una partición* $\mathcal{I}_i$ *de* $\{h \in H : P(h) = i\}$ *con la propiedad de que* $A(h) = A(h')$ *siempre que* h *y* h' *estén en el mismo bloque de la partición. Para* $I_i \in \mathcal{I}_i$ *se denota por* $A(I_i)$ *el conjunto* $A(h)$ *y por* $P(I_i)$ *el jugador* $P(h)$ *para cualquier* $h \in I_i$*.* $\mathcal{I}_i$ *es la* **partición de información** *del jugador* i*, un conjunto* $I_i \in \mathcal{I}_i$ *es un* **conjunto de información** *del jugador* i*.*
- *Para cada jugador* $i \in N$*, una función de utilidad* u_i *de los estados terminales* Z *a los reales* $\mathbb{R}$*. Si* $N = \{1, 2\}$ *y* $u_1 = -u_2$*, se dice que se tiene un* **juego de dos jugadores de suma cero en forma extensiva***. Se define* $\Delta_{u,i} = \max_z u_i(z) - \min_z u_i(z)$ *como el rango de utilidades del jugador* i*.*

Juegos en Forma Extesiva IV

Definición

Una **estrategia pura** *para el jugador i es una función $s_i : \mathcal{I}_i \to \bigcup_{I_i \in \mathcal{I}_i} A(I_i)$ tal que $s_i(I_i) \in A(I_i)$, donde $A(I_i) = A(h)$ para cualquier $h \in I_i$ es el conjunto de acciones permitidas después de la historia h.*

Definición

Una **estrategia mixta** *σ_i^m para el jugador i es una distribución de probabilidad sobre S_i. Es decir, $\sigma_i^m \in \Delta(S_i)$.*

Definición

Un **perfil estratégico mixto** *$\sigma^m \in \prod_{i \in N} \Delta(S_i)$ consiste en una estrategia mixta para cada jugador de forma $\sigma^m = (\sigma_1^m, \sigma_2^m, ..., \sigma_N^m)$.*

Juegos en Forma Extesiva V

Definición

Una **estrategia de comportamiento** *para el jugador i consiste en una distribución de probabilidad para cada conjunto de información $I_i \in \mathcal{I}_i$ sobre el conjunto de acciones $A(I_i)$ que pueden ejecutarse en I_i. Es decir, una estrategia de comportamiento es una tupla $(\sigma_i^b(I_i))_{I_i \in \mathcal{I}_i}$ donde $\sigma_i^b(I_i) \in \Delta(A(I_i))$.*

Definición

Un **perfil estratégico de comportamiento** σ^b *es una estrategia de comportamiento para cada jugador.*

Juegos en Forma Extesiva VI

Definición

Sea $\Sigma = \prod_{i \in N} \Sigma_i$ *el conjunto de perfiles mixtos o de comportamiento, según sea el caso, para los jugadores en* N *(*Σ_i *es el conjunto de estrategias del jugador* i*). Para* $\varepsilon \geq 0$*, se dice que un perfil estratégico* $\sigma \in \Sigma$ *es un* ε**-equilibrio de Nash** *si y sólo si para todo jugador* i *y perfil* $\sigma_i' \in \Sigma_i$*,*

$$u_i(\sigma) + \varepsilon \ \geq \ u_i(\sigma_i', \sigma_{-i})\,.$$

El perfil $\sigma \in \Sigma$ *es un* **equilibrio de Nash** *si y sólo si* σ *es un* 0*-equilibrio de Nash.*

Juegos en Forma Extesiva VII

Definición

Se dice que el jugador i tiene **perfect recall** *en el juego Γ (en forma extensiva) si para cualquier par de historias h_1, h_2 con $P(h_1) = P(h_2) = i$, tales que $I(h_1) = I(h_2)$ las siguientes condiciones se cumplen:*

$$h \sqsubseteq h_1 \implies \left(\exists h' \sqsubseteq h_2 : I(h) = I(h')\right),$$
$$(h_1, a) \sqsubseteq h \;\wedge\; (h_2, b) \sqsubseteq h' \;\wedge\; a \neq b \implies I(h) \neq I(h').$$

Definición

Se dice que dos estrategias σ y σ' son equivalentes si la probabilidad de alcanzar cualquier historia terminal es la misma; i.e., $\pi^{\sigma}(z) = \pi^{\sigma'}(z)$ para todo $z \in Z$.

Juegos en Forma Extesiva VIII

Definición

Sea $s_i \in S_i$ *una estrategia pura del jugador* i *e* $I_i \in \mathcal{I}$ *un conjunto de información de dicho jugador. Se dice que* I_i *es alcanzable bajo* s_i *si existe una historia* $h \in H$ *tal que* $h \in I_i$ *y para toda historia (prefijo)* $h' \sqsubset h$ *se cumple que: si* $P(h') = i$*, entonces* $(h', s_i(I(h'))) \sqsubset h$ *y si* $P(h') = c$ *entonces existe una acción* a *tal que* $(h', a) \sqsubset h$ *y la probabilidad de elegir la acción* a *en* h' *es positiva, i.e.,* $f_c(a|h') > 0$*.*

Juegos en Forma Extesiva IX

Teorema

Si para el jugador i se cumple que $I(h') \neq I(h)$ para cualquier par de historias h y h' tal que $h' \sqsubset h$ y $P(h) = P(h') = i$, entonces para cualquier estrategia de comportamiento $\sigma_i^b \in B^i$ para el jugador i, existe una estrategia mixta σ_i^m que es **equivalente** *a σ_i^b. En particular, la estrategia mixta σ_i^m viene dada por:*

$$\sigma_i^m(s_i) \; := \; \prod_{I_i \in \mathcal{I}_i} \sigma_i^b(I_i)(s_i(I_i)) \, .$$

Teorema

Considere un juego finito de N personas. Si el jugador i tiene "perfect recall", entonces para cada estrategia mixta $\sigma_i^m \in \Delta(S_i)$ del jugador i, existe una estrategia de comportamiento $\sigma_i^b \in B^i$ equivalente a σ_i^m.

Juegos en Forma Extesiva X

Teorema

Considere un juego finito de N personas. Si el jugador i tiene "perfect recall", entonces para cada estrategia mixta $\sigma_i^m \in \Delta(S_i)$ del jugador i, existe una estrategia de comportamiento $\sigma_i^b \in B^i$ equivalente a σ_i^m.

Teorema

En un juego con perfect recall, cualquier estrategia mixta de un agente dado puede ser remplazada por una estrategia de comportamiento equivalente, y cualquier estrategia de comportamiento puede ser remplazada por una estrategia mixta equivalente. Dos estrategias son equivalentes en el sentido en que inducen los mismos resultados de probabilidades, para cualquier perfil estratégico fijo (mixto o de comportamiento) del resto de los agentes.

Explotabilidad I

Teorema

Sea $\sigma^* = (\sigma_1^*, \sigma_2^*)$ *un equilibrio de Nash de un juego de dos jugadores de suma cero, tal que* $u_1(\sigma) = u$. *Entonces* $u_i(\sigma^*) \leq u_i(\sigma_i^*, \sigma_{-i})$, *para cualquier estrategia* σ_{-i}.

Teorema

Sean $\sigma = (\sigma_1, \sigma_2)$ *y* $\sigma' = (\sigma_1', \sigma_2')$ *equilibrios de Nash en un juego de dos jugadores con suma cero. Entonces* $\sigma'' = (\sigma_1, \sigma_2')$ *y* $\sigma''' = (\sigma_1', \sigma_2)$ *son también equilibrios de Nash. Además,* $u_i(\sigma) = u_i(\sigma') = u_i(\sigma'') = u_i(\sigma''')$, *para* $i \in \{1, 2\}$.

Explotabilidad II

Definición

Sea $\sigma^* = (\sigma_1^*, \sigma_2^*)$ *un equilibrio de Nash en un juego de dos jugadores de suma cero. Sea* $\sigma = (\sigma_1, \sigma_2)$ *una aproximación a* σ^**. Sean* σ_1' *y* σ_2' *las mejores respuesta a* σ_2 *y* σ_1*, respectivamente. Se definen* $\varepsilon_1 = u_2(\sigma_1, \sigma_2') - u_2(\sigma^*)$ *y* $\varepsilon_2 = u_1(\sigma_1', \sigma_2) - u_1(\sigma^*)$*. Luego, la explotabilidad de* σ *es igual a* $\varepsilon_\sigma = \varepsilon_1 + \varepsilon_2$

Teorema

Sean σ *y* σ' *las estrategias definidas en la definición anterior. Luego:*

$$\varepsilon_\sigma = u_2(\sigma_1, \sigma_2') + u_1(\sigma_1', \sigma_2)$$

Regret Matching I

Regret Condicional

$$
\begin{aligned}
D_i^t(j,k) &= \frac{1}{t} \sum_{\substack{1 \le \tau \le t \\ s_i^\tau = j}} u_i(k, s_{-i}^\tau) - u_i(s^\tau) \\
R_i^t(j,k) &= [D_i^t(j,k)]^+ = \max\{0, D_i^t(j,k)\}
\end{aligned}
$$

$$
\begin{cases}
p_{t+1}^i(k) := \frac{1}{\mu} R_i^t(j,k) & \text{si } k \neq j \\
p_{t+1}^i(j) := 1 - \sum_{k \in S_i, k \neq j} p_{t+1}^i(k) & \text{si } k = j
\end{cases}
$$

Teorema

Si cada jugador juega de acuerdo al procedimiento descrito arriba, entonces la distribución empríca del juego z_t converge (a.s.) cuando $t \to \infty$ al conjunto de equilibrios correlacionado del juego Γ.

Regret Matching II

Teorema

Sea $(s_t)_{t=1,2,\ldots}$ *una secuencia de juegos de* Γ*. Entonces,* $R_i^t(j,k)$ *converge a* 0 *para cada* i *y cada* $j,k \in S_i$*, con* $j \neq k$*, si y sólo si la secuencia de distribuciones empŕicas* z_t *converge al conjunto de equilibrio correlacionado.*

Regret Matching III

Vector Invariante de Probabilidad

$$\begin{cases} M_t^i(j,k) & := & \frac{1}{\mu} R_i^t(j,k) & \text{si } k \neq j \\ M_t^i(j,j) & := & 1 - \sum_{k \in S_i, k \neq j} R_i^t(j,k) & \text{si } k = j \end{cases}$$

Considere un vector (fila) invariante de probabilidad q_t^i (dicho vector siempre existe), donde $q_t^i \in \Delta(S_i)$, para la matriz M^t. Es decir, q_t^i satisface $q_t^i \times M_t^i = q_t^i$.

Teoream

Sea $R_t^i(j,j) = 0$. El vector q_t^i, definido previamente, cumple que:

$$q_t^i(j) \sum_{k \in S_i} R_i^t(j,k) \; = \; \sum_{k \in S_i} q_t^i(k) R_i^t(k,j) \, .$$

Regret Matching IV

Teorema

Supongamos que a cada período $t+1$*, el jugador* i *elige las estrategias acorde a un vector de distribución de probabilidad* q_t^i *que satisface la ecuación anterior. Entonces,* $R_t^i(j,k)$ *converge a cero (a.s.) para todo* $j,k \in S_i$ *con* $j \neq k$*.*

Regret Matching V

Regret Incondicional

Definition

Un procedimiento adaptativo es **universalmente consistente** *para el jugador i si:*

$$\limsup_{t\to\infty}\left[\max_{k\in S_i}\frac{1}{t}\sum_{\tau=1}^{t}u_i(k,s_{-i}^{\tau})-\frac{1}{t}\sum_{\tau=1}^{t}u_i(s_{\tau})\right] \;\leq\; 0 \quad (a.s.)\,.$$

$$D_i^t(k) \;=\; \frac{1}{t}\sum_{\tau=1}^{t}u_i(k,s_{-i}^{\tau})-u_i(s_\tau)$$

$$R_i^t(k) \;=\; [D_i^t(k)]^+ \;=\; \max\{0,D_i^t(k)\}$$

Regret Matching VI

Teorema

El procedimiento adaptativo definido previamente es universalmente consistente para el jugador i.

Teorema

En un procedimiento adaptativo de Regret Matching, si el regret condicional converge a 0, *entonces el procedimiento es universalmente consistente.*

Regret Matching VII

Teorema

Sea Γ *un juego de dos jugadores de suma cero y sea* $(s^t)_{t=1,2,\ldots,T}$ *una secuencia de juegos de* Γ*, tales que, para todo* $s_i \in S_i$*, para todo* $i \in 1,2$*:*

$$\frac{1}{T}\sum_{t=1}^{T} u_i(s_i, s_{-i}^t) - \frac{1}{T}\sum_{t=1}^{T} u_i(s^t) \;\leq\; \varepsilon$$

para algún $\varepsilon > 0$*. Sea* $\bar{\sigma}^T = (\bar{\sigma_1}^T, \bar{\sigma_2}^T)$*, donde:*

$$\bar{\sigma}_i^T(s_i) \;=\; \frac{|\{t \leq T : s_i^t = s_i\}|}{T},$$

es decir, $\bar{\sigma}^T$ *es la distribución empírica de probabilidad, note que* $|\{t \leq T : s_i^t = s_i\}|$ *es igual al número de veces que se eligió* s_i *hasta el tiempo* T*. Entonces,* $\bar{\sigma}^T$ *es un* 2ε*-equilibrio de Nash.*

Counterfactual Regret Minimization I

Definición

Considere T repeticiones de un juego en forma extensiva indexadas en tiempo por $t = 1, 2, \ldots, T$. Sea σ^t el perfil estratégico de comportamiento utilizado por los jugadores a tiempo t (i.e., σ^t consiste de una estrategia de comportamiento σ_i^t para cada jugador i) . El **regret promedio general** *del jugador i a tiempo T es:*

$$R_i^T = \max_{\sigma_i^* \in B_i} \frac{1}{T} \sum_{t=1}^{T} u_i(\sigma_i^*, \sigma_{-i}^t) - u_i(\sigma^t)$$

donde B_i es el conjunto de estrategias de comportamiento para el jugador i.

Counterfactual Regret Minimization II

Se denotará con $\bar{\sigma}_i^T$ la estrategia promedio del jugador i, i.e, para cada conjunto de información $I \in \mathcal{I}_i$ y para cada acción $a \in A(I)$ se define:

$$\bar{\sigma}_i^T(I)(a) \;=\; \frac{\sum_{t=1}^{T} \pi_i^{\sigma^t}(I)\, \sigma^t(I)(a)}{\sum_{t=1}^{T} \pi_i^{\sigma^t}(I)}$$

Teorema

En un juego de dos jugadores de suma cero si el regret promedio general a tiempo T es menor que ε entonces $\bar{\sigma}^T = (\bar{\sigma}_1^T, \bar{\sigma}_2^T)$ es un 2ε-equilibrio de Nash.

Counterfactual Regret Minimization III

Definición

Sea σ un perfil estratégico e $I \in \mathcal{I}_i$ un conjunto de información del jugador i, la **utilidad contrafactual** *del par (σ, I) es la ganancia esperada dado que el conjunto I es alcanzado y todos los jugadores juegan con la estrategia σ con excepción del jugador i que juega para alcanzar I. Formalmente:*

$$u_i(\sigma, I) \; = \; \frac{\sum_{h \in I, z \in Z} \pi^{\sigma}_{-i}(h) \, \pi^{\sigma}(h, z) \, u_i(z)}{\pi^{\sigma}_{-i}(I)}$$

donde Z denota el conjunto de historias terminales, $\pi^{\sigma}_{-i}(h)$ denota la probabilidad de alcanzar h dado que todos los jugadores utilizan σ excepto el jugador i que juega para alcanzar h, y $\pi^{\sigma}(h, h')$ denota la probabilidad de ir de la historia h a la historia h' dado el perfil estratégico σ; i.e., $\pi^{\sigma}(h, h') = \pi^{\sigma}(h')/\pi^{\sigma}(h)$ si $h \sqsubseteq h'$, y $\pi^{\sigma}(h, h') = 0$ en caso contrario.

Counterfactual Regret Minimization IV

Definición

El **regret contrafactual inmediato** *es:*

$$R^T_{i,imm}(I) \;=\; \max_{a \in A(I)} \frac{1}{T} \sum_{t=1}^{T} \pi^{\sigma^t}_{-i}(I) \Big[u_i(\sigma^t|_{I \to a}, I) - u_i(\sigma^t, I) \Big] .$$

Teorema

$R^T_i \;\leq\; \sum_{I \in \mathcal{I}_i} R^{T,+}_{i,imm}(I)$

Counterfactual Regret Minimization V

Regret Minimization

Definición

El regret promedio de no haber elegido la acción $a \in A$ *hasta tiempo* T*, y en su lugar elegir la acción* a' *con probabilidad* $p^t(a')$ *para* $t = 1, 2, \ldots, T$*, se define como:*

$$R^T(a) \ = \ \frac{1}{T} \sum_{t=1}^{T} \left[u^t(a) - \sum_{a' \in A} p^t(a') u^t(a') \right] .$$

Counterfactual Regret Minimization VI

Sea $R^{t,+}(a) = \max\{R^t(a), 0\}$, el algoritmo de *Regret Minimization* consiste en utilizar a tiempo t una distribución de probabilidad p^t, definida por:

$$p^t(a) \;=\; \begin{cases} \frac{R^{t,+}(a)}{\sum_{a' \in A} R^{t,+}(a')} & \text{si} \quad \sum_{a' \in A} R^{t,+}(a') > 0\,, \\ \frac{1}{|A|} & \text{en otro caso}\,. \end{cases}$$

Teorema

Si $|u| = \max_{t \in \{1,2,\ldots T\}} \max_{a,a' \in A} \left[u^t(a) - u^t(a')\right]$*, entonces el regret del algoritmo de Regret Minimization está acotado por:*

$$\max_{a \in A} R^t(a) \;\leq\; |u| \sqrt{\frac{|A|}{T}}\,.$$

Counterfactual Regret Minimization VII

CFR

Se mantiene para cada conjunto de información $I \in \mathcal{I}_i$ y cada acción $a \in A(I)$, la medida de regret:

$$R_i^T(I,a) \;=\; \frac{1}{T}\sum_{t=1}^{T} \pi_{-i}^{\sigma^t}(I)\Big[u_i(\sigma^t|_{I\to a}, I) - u_i(\sigma^t, I)\Big]$$

Para $R_i^{T,+}(I,a) = \max\{R_i^T(I,a), 0\}$, definimos la estrategia σ_i^{T+1} elegida por el jugador i a tiempo $T+1$:

$$\sigma_i^{T+1}(I)(a) \;=\; \begin{cases} \frac{R_i^{T,+}(I,a)}{\sum_{a' \in A(I)} R_i^{T,+}(I,a')} & \text{si } \sum_{a' \in A(I)} R_i^{T,+}(I,a') > 0\,, \\ \frac{1}{|A(I)|} & \text{en otro caso.} \end{cases}$$

Counterfactual Regret Minimization VIII

Teorema

Si el jugador i selecciona las acciones de acuerdo a la Ecuación anterior, entonces $R^T_{i,\text{imm}}(I) \leq \Delta_{u,i}\sqrt{|A_i|/T}$, donde $|A_i| = \max\{|A(h)| : P(h) = i\}$ es el máximo número de acciones que el jugador i tiene disponibles en una historia dada y $\Delta_{u,i} = \max_{z\in Z} u_i(z) - \min_{z\in Z} u_i(z)$ es el rango de las utilidades del jugador i. Luego:

$$R_i^T \;\leq\; \sum_{I\in\mathcal{I}_i} R^{T,+}_{i,imm}(I) \;\leq\; \Delta_{u,i}\,|\mathcal{I}_i|\,\sqrt{\frac{|A_i|}{T}}\,.$$

Counterfactual Regret Minimization IX

Monte Carlo CFR

$$\tilde{u}_i(\sigma, I|j) \; = \; \sum_{h \in I, z \in Q_j} \frac{\pi^{\sigma}_{-i}(h)\, \pi^{\sigma}(h, z)\, u_i(z)}{q(z)\, \pi^{\sigma}_{-i}(I)} \, .$$

Teorema

$E_{j \sim q_j}[\tilde{u}_i(\sigma, I|j)] \; = \; u_i(\sigma, I).$