

¿Cooperación o Competencia? ¡Juegos diferenciales de todos modos!

José Daniel López Barrientos
Facultad de Ciencias Actuariales
Universidad Anáhuac México
daniel.lopez@anahuac.mx

Resumen

En este trabajo presentamos una perspectiva de algunos juegos dinámicos competitivos y colaborativos. Nuestro recorrido empieza en los juegos matriciales, y visita los diferenciales, los diferenciales estocásticos, y llega a los diferenciales estocásticos con modos múltiples en horizonte infinito con el criterio del pago/costo descontado. Terminamos nuestro viaje estudiando el célebre procedimiento de distribución de recompensas para juegos dinámicos colaborativos de Leon Petrosyan y lo aplicamos al caso de los juegos diferenciales estocásticos con modos múltiples. Deseamos que los resultados presentados sirvan para ilustrar aplicaciones en los contextos natural y social, y que el lector los aproveche como guía en el apasionante mundo de la teoría de los juegos dinámicos.

Je vous écris une longue lettre parce que je n'ai pas le temps d'en écrire une courte.
Blaise Pascal

1. Introducción

Desde hace algunos años, la ciencia y la tecnología han prestado una atención especial al comportamiento estratégico en las sociedades, tanto humanas, como animales. Ejemplo de esto es el impulso que se ha dado al desarrollo de programas computacionales para optimizar rutas de tránsito a través de la coordinación de una red social virtual. Desde el punto de vista de quien toma decisiones, resulta muy importante considerar las interacciones y las interdependencias entre las decisiones

de los agentes que afectan un sistema. Esto ha derivado en el nacimiento y desarrollo de la teoría de juegos como un instrumento clave en las investigaciones pura y aplicada. Sin embargo, es frecuente que se enfoque esta disciplina desde el punto de vista de la competencia entre los agentes, dejando de lado los modelos de colaboración entre ellos.



Una razón por la que este fenómeno está tan difundido en nuestra cultura es que los tiempos que vivimos nos han habituado a aceptar con normalidad (y casi indiferencia o resignación) los sistemas en los que existen conflictos entre los intereses de los participantes involucrados. Y es que, desde el final de la Segunda Guerra Mundial el mundo ha conocido un sinnúmero de aplicaciones que sirven para *evitar* la colaboración entre los tomadores de decisiones, y que más aún, fomentan la competencia entre ellos. Por ejemplo, en el año 1948 nació en los Estados Unidos de América la Corporación RAND como un laboratorio de ideas para formar a las Fuerzas Armadas de ese país (visite [32]). Algunos consideran que el nombre de la compañía es un acrónimo de la frase «Research And Development», y ciertamente, la investigación que se llevaba a cabo ahí nutrió significativamente a la teoría de juegos desde el período de la Guerra Fría, y ha seguido haciéndolo hasta los tiempos del neoliberalismo y la globalización.

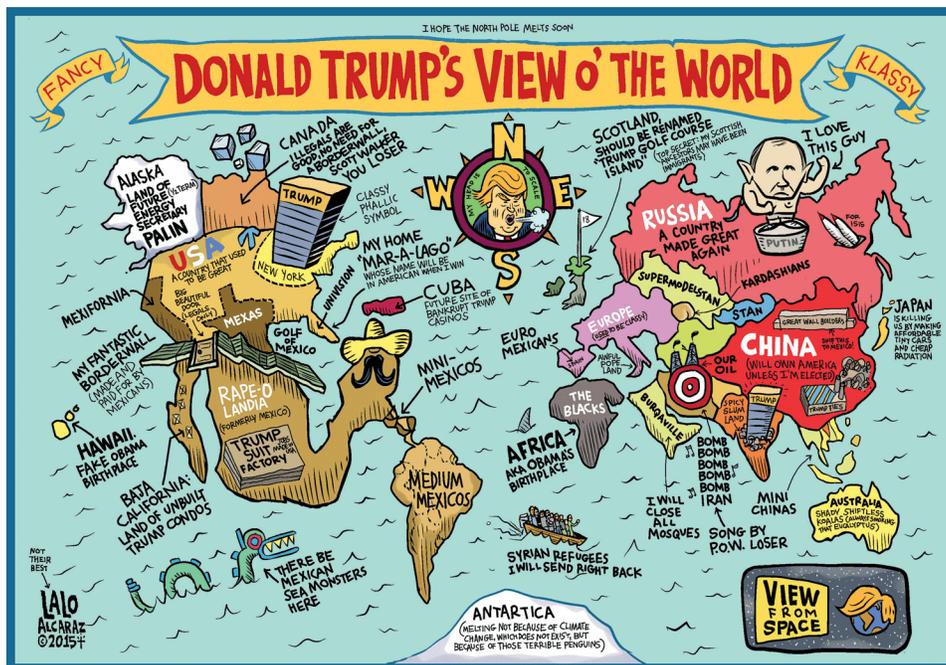
Por otro lado, medios como el cine e industrias como las de los juegos sociales han contribuido a que nuestra visión de la teoría de juegos esté enfocada —casi exclusivamente— en la competencia. La imagen de la derecha muestra un ejemplo típico de juego competitivo: la persecución de cazas imperiales por parte de las naves rebeldes (¡incluso puede conseguir un cuadro con esta decoración en [23]!). La figura 1 ilustra dos juegos sociales en los que derrotar al adversario es la meta. El de la izquierda es un juego *de autor*. Lo debemos a Ananda Gupta y Jason Matthews (puede ver más sobre él en [14]). La de la derecha, es una captura de pantalla de un videojuego que vio la luz en la década de los 90s, y que sigue dando de qué hablar aún en nuestros días. Lo debemos a Takashi Nishiyama y Hiroshi Matsumoto (vea [28] para saber más de este juego). Desde hace varias décadas, los expertos en teoría de juegos saben que las conductas no-colaborativas pueden llevar a resultados indeseables (como en el dilema del prisionero), o simplemente devastadores (como en la tragedia de los comunes —vea el ejemplo 4





Figura 1. Twilight Struggle y Street Fighter II.

en [31, cap. 3]—). Más aún, en un mundo en constante evolución, las conductas no-colaborativas, guiadas por la racionalidad individual sin visión de futuro, pueden ser una fuente de consecuencias desastrosas. Para muestra, ¿qué tal la siguiente obra de Lalo Alcaráz (vea [1])?



La colaboración, en cambio, da la posibilidad de alcanzar soluciones eficientes con resultados óptimos para la sociedad que las promueve. Sin embargo, la búsqueda de tales soluciones exige una planeación de la repartición de las ganancias que la coalición obtendrá. Pues, si no se toma en cuenta esto, las consecuencias pueden ser igualmente severas que en el caso competitivo. Vea la figura 2.

Una alternativa ante la problemática que planteamos fue propuesta por Leon Petrosyan en 1977 [30]. Se trata de un procedimiento de asignación de recompensas instantáneas que permite considerar el tamaño



Figura 2. En esta pintura, obra del artista gráfico Yapim Aşamasi (vea [3]), podemos ver a Luke Skywalker tratando de salvar de la destrucción a su acérrimo rival en la Estrella de la Muerte. Ciertamente, los agentes colaboraron para derrocar al Emperador Palpatine, pero no planearon qué harían una vez que su alianza prevaleciera sobre el mal. Esto llevó a la aniquilación de uno de los agentes.

y la participación de los agentes que conforman la coalición. Este método ha sido utilizado por muchos expertos a partir de entonces, e incluso le valió a su descubridor el reconocimiento de la Sociedad Internacional de Juegos Dinámicos con el Premio Isaacs en 2014.

Nuestro trabajo está dividido de la manera siguiente. En la sección que sucede a ésta presentamos una revisión de algunos juegos competitivos. Introducimos primeramente un juego matricial que sirve para fijar ideas en un contexto de competencia, y expone progresivamente ideas que complementan ese modelo hasta arribar a la noción de juegos diferenciales estocásticos con modos múltiples. La sección termina con un resultado *de verificación* que relaciona el principio de programación dinámica con la solución del problema de competencia.

La sección 3 utiliza algunos fenómenos presentes en la naturaleza y en la vida social (humana y animal) para exhibir los supuestos necesarios para definir adecuadamente la colaboración entre los individuos en el contexto diferencial estocástico con modos múltiples. Tales hipótesis son: racionalidades individual y de grupo, y la consistencia de los subjuegos. La sección finaliza con un segundo resultado de verificación, análogo al de la sección 2.

La sección 4 expone el procedimiento de asignación de recompensas de Petrosyan para el caso de un juego diferencial estocástico de modos múltiples. Nuestras conclusiones están en la sección 5.

Los resultados que presentamos están destinados a ilustrar sus aplicaciones en los contextos natural y social, y esperamos sirvan al lector como una guía en el apasionante mundo de los juegos dinámicos.

2. Juegos competitivos

Dentro de la literatura convencional sobre teoría de juegos dinámicos, es común encontrar herramientas que se usan para modelar situaciones de competencia entre dos o más agentes. Por ejemplo, la clase de los juegos llamados *diferenciales* —introducida por Rufus Isaacs mientras trabajaba para RAND Corporation, de 1948 a 1955 (vea [16, §7.13], [18] y [40])— surgió a partir de la situación en la que un jugador trata de atrapar a otro, mientras que éste último procura escaparse. Como es de esperarse, estos juegos llevan el nombre de *juegos de persecución-evasión*.



La pintura de la derecha es un retrato de Rufus Isaacs y es obra de la pintora Esther Freeman (está tomada de [40]). El trabajo de Isaacs es tan importante, que el premio que otorga la Sociedad Internacional de Juegos Dinámicos a los investigadores más destacados de cada año lleva su nombre.

A continuación tomamos de [31, ej. 1.1.9] una versión matricial de un juego de persecución-evasión.

Ejemplo 2.1. *Suponga que el jugador P busca al jugador E , que está en movimiento, escondiéndose del jugador P . El jugador P puede moverse a las velocidades β_1 (caminar), β_2 (trotar) y β_3 (correr), mientras que el jugador E puede hacerlo a las velocidades α_1 (caminar), α_2 (trotar) y α_3 (correr). La habilidad de detección del jugador P está dada en la tabla que mostramos a continuación.*

Velocidades	β_1	β_2	β_3
α_1	4	5	6
α_2	3	4	5
α_3	1	2	3

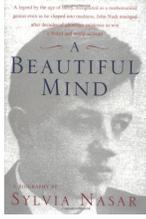
Las estrategias de los jugadores son sus respectivas velocidades, mientras que el pago del jugador P es la eficiencia de su búsqueda $a_{ij} = \omega_i d_{ij}$; $i, j = 1, 2, 3$, donde d_{ij} es la (i, j) -ésima entrada de la tabla anterior; y $\omega_1 = 1$, $\omega_2 = 2$ y $\omega_3 = 3$. Entonces el problema de elegir velocidades puede representarse mediante un juego de suma cero con matriz de

pagos (para el jugador P) dada por

$$A = \begin{pmatrix} 4 & 5 & 6 \\ 6 & 8 & 10 \\ 3 & 6 & 9 \end{pmatrix}.$$

Es fácil ver que el par (α_2, β_1) es la estrategia que garantiza la *ganancia mínima más grande* para el jugador P en el juego del ejemplo 2.1. En efecto, si el jugador P usa α_2 , y el jugador E usa β_1 , el jugador P obtendrá una recompensa de 6. Pero si el jugador P cambia su estrategia por α_1 o por α_3 , mientras que el jugador E mantiene su estrategia en β_1 , su recompensa será de 4 o de 3, respectivamente. ¡Ambas recompensas son menores que la que habría obtenido usando α_2 !

La pareja (α_2, β_1) también garantiza el *castigo máximo más pequeño* para el jugador E. Efectivamente, si el jugador E usa las estrategias β_2 o β_3 mientras que su oponente usa α_2 , su castigo será de 8 o 10, respectivamente. ¡Ambos castigos son mayores que el que le habría correspondido si hubiera usado la velocidad β_1 !



Esta noción subyace a la definición siguiente, y es el fundamento que le valió a su descubridor, John Forbes Nash Jr., obtener el premio Nobel de Economía en 1994. A la izquierda puede verse la portada del libro de Sylvia Nasar sobre la vida y obras de John Nash (vea [26]). Es una ficción que abarca el período de la vida de Nash que comprende desde su infancia, y hasta su premiación, incluyendo el período en que trabajó para

RAND Corporation de 1950 a 1954 con Rufus Isaacs. John Nash falleció el 23 de mayo de 2015, en New Jersey, a causa de un accidente automovilístico.

Definición 2.1. Sea $V^\ell(\pi^1, \pi^2)$ el pago que recibe el ℓ -ésimo jugador ($\ell = 1, 2$) en un juego de dos jugadores que usan las estrategias π^1 y π^2 , respectivamente. Decimos que el par (π^{1*}, π^{2*}) es un *equilibrio de Nash* para el juego en cuestión si

$$V^1(\pi^1, \pi^{2*}) \leq V^1(\pi^{1*}, \pi^{2*}) \text{ y } V^2(\pi^{1*}, \pi^2) \leq V^2(\pi^{1*}, \pi^{2*}).$$

Asimismo, llamamos a $V^1(\pi^{1*}, \pi^{2*})$ y $V^2(\pi^{1*}, \pi^{2*})$ los valores del juego para el jugador 1 y el jugador 2, respectivamente. En el caso de un juego se suma-cero, tenemos que $V^1(\pi^{1*}, \pi^{2*}) = -V^2(\pi^{1*}, \pi^{2*}) := V(\pi^{1*}, \pi^{2*})$ y decimos que esta cantidad es el *valor del juego*.

La figura 3 muestra un juego de persecución-evasión. Se trata de una versión modificada de la que se encuentra en [13].

El ejemplo 2.1 representa un caso en que las trayectorias del perseguidor y del evasor son irrelevantes, pero si deseamos captar la idea que los agentes se mueven y asumen posiciones distintas a medida que el

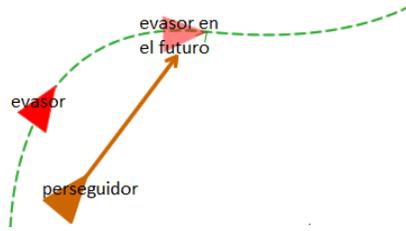


Figura 3. Juego de persecución-evasión.

tiempo pasa, será necesario echar mano de alguna herramienta capaz de modelar sistemas dinámicos.

Ejemplo 2.2. [18, ejs. 1.9.1 y 2.1.1] *Suponga que los jugadores P y E viajan en el plano con rapidez constantes¹ W_1 y w_1 , con $W_1 > w_1$. El pago para el perseguidor será el tiempo de captura. ¿Cuáles son las estrategias óptimas de cada jugador?*

En este sencillo ejemplo, la solución es evidente. Cada jugador debe moverse (el jugador E huyendo y el jugador P persiguiéndolo) a lo largo de la línea que une sus posiciones iniciales. Es fácil ver que se moverán hasta el momento en que el perseguidor dé alcance al evasor.

La figura 4 nos muestra que el *valor del juego* es 15.5 unidades de tiempo (representamos la magnitud de las rapidez en la parte derecha de la imagen). Esto quiere decir que, sin importar qué haga el evasor, no puede escapar por más de 15.5 unidades de tiempo cuando es perseguido por el otro jugador. En efecto, si hace cualquier cosa que no sea tratar de escapar de P en dirección opuesta a él, el jugador P lo atrapará en menos de 15.5 unidades de tiempo. Recíprocamente, el perseguidor no puede lograr la captura en menos de 15.5 unidades, y si no persigue inmediatamente, el evasor quedará en libertad por más de 15.5 unidades. Desde este punto de vista, las estrategias que constituyen el equilibrio de Nash son óptimas.

Observación 2.1. *A partir de ahora, cuando nos refiramos a una estrategia, estaremos hablando de una función de las variables de estado (a menos de que indiquemos lo contrario).*

En este caso, resulta natural considerar que las variables de estado son las coordenadas en el plano de cada jugador. Sean $\vec{x}_E := (x_E, y_E)$ y $\vec{x}_P := (x_P, y_P)$ las posiciones del evasor y el perseguidor, respectivamente. El movimiento de cada agente puede caracterizarse mediante el

¹Recuerde que la rapidez es una magnitud escalar que relaciona la distancia recorrida con el tiempo empleado, además, no tiene en cuenta la dirección. La velocidad, por el contrario puede representar un cambio en la dirección, y es una magnitud vectorial que compara el desplazamiento con el tiempo empleado.

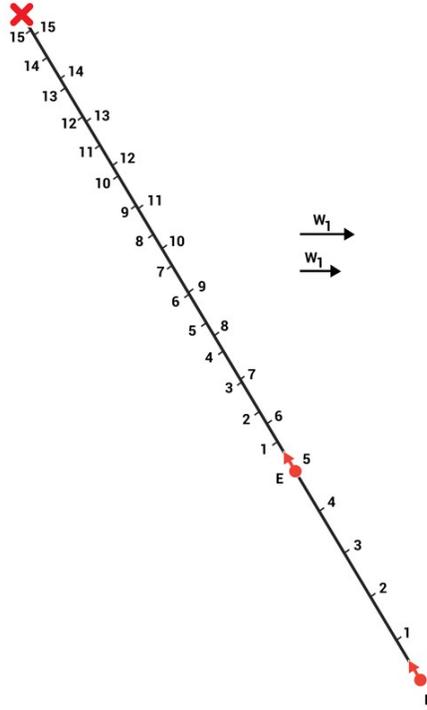


Figura 4. La captura ocurre en el punto \times y el pago es de 15.5 unidades de tiempo. (Fuente: [18, fig. 1.9.1(a)].)

sistema

$$\dot{x}_E = w_1 \cos \phi, \quad (1)$$

$$\dot{y}_E = w_1 \sen \phi, \quad (2)$$

$$\dot{x}_P = W_1 \cos \psi, \quad (3)$$

$$\dot{y}_P = W_1 \sen \psi, \quad (4)$$

donde ϕ y ψ son los ángulos que forman los vectores \vec{x}_E y \vec{x}_P , respectivamente, con el eje de las abscisas.

En el juego del ejemplo 2.2, la estrategia óptima del perseguidor será viajar siempre hacia el evasor. Poco importa en dónde se encuentre este último. En la figura 5 podemos ver al evasor comportándose de un modo no-óptimo: viaja a lo largo de la línea L , mientras que su perseguidor mantiene una conducta óptima, capturándole en el punto \times , y obteniendo así un pago de 6 unidades. Esto constituye una ilustración más de la definición 2.1, pues, al dejar de usar la estrategia óptima, el evasor es capturado en un tiempo menor que 15.5 unidades.

Observación 2.2. Las ecuaciones (1)-(4) son la razón por la que la clase de los juegos que admiten que las variables de estado sean descritas

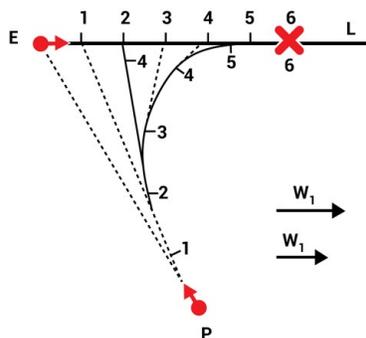


Figura 5. El evasor **no** usa la estrategia de Nash y es capturado en un tiempo de 6. (Fuente: [18, fig. 1.9.1(c)].)

por un sistema de ecuaciones diferenciales se llamen juegos diferenciales. En ese juego, las estrategias para los jugadores son ϕ y ψ . Note que, en efecto, pueden escribirse en términos de las variables de estado \vec{x}_E y \vec{x}_P .

Las observaciones 2.1 y 2.2 hacen referencia al hecho de que las estrategias de los jugadores deben escribirse en términos del estado en que cada cual se encuentra. Pero en el ejemplo 2.2, el estado en tiempo t está determinado por las ecuaciones diferenciales (1)-(4). Esto quiere decir que las estrategias deben depender del tiempo también. En general, los juegos diferenciales para n personas se juegan de la siguiente manera. En cada tiempo $t > 0$, cada jugador observa el estado del sistema $\vec{x}(t)$, e independientemente de los demás, elige acciones de control $u_\ell(t)$ (para $\ell = 1, \dots, n$) en un conjunto adecuado. Para cada estado inicial \vec{x} , la meta de cada jugador es maximizar cierto índice de rendimiento a través del tiempo. La definición siguiente establece el tipo de estrategias en el que estamos interesados (vea la sección 3.3 en [2] para conocer más detalles).

Definición 2.2. Sea Ω el conjunto de estados asociados con un juego diferencial (estocástico), y U_1, \dots, U_n los conjuntos de acciones puras disponibles para cada jugador. Una estrategia para el ℓ -ésimo jugador es una familia de funciones $\pi^\ell := \{\pi_t^\ell : \mathcal{B}(U_\ell) \times \Omega \rightarrow [0, 1], t > 0\}$, donde $\mathcal{B}(U_\ell)$ es la σ -álgebra de Borel de Ω . Cada miembro de esta familia de funciones satisface lo siguiente:

- (a) Para cada $t \geq 0$ y $x \in \Omega$, $\pi_t^\ell(\cdot|x)$ es una medida de probabilidad en U_ℓ tal que $\pi_t^\ell(U_\ell|x) = 1$, y para cada $D \in \mathcal{B}(U_\ell)$, $\pi_t^\ell(D|\cdot)$ es una función de Borel sobre Ω ;
- (b) Para cada $D \in \mathcal{B}(U_\ell)$ y $x \in \Omega$, la función $\pi_t^\ell(D|x)$ es medible en el sentido de Borel para $t \geq 0$.

En este trabajo supondremos la existencia de una medida de probabilidad $\pi^\ell(\cdot|x)$ soportada en U_ℓ , tal que $\pi^\ell(\cdot|x) = \pi_t^\ell(\cdot|x)$ para todo $x \in \Omega$ y $t \geq 0$.

Denotamos el conjunto de estrategias para el ℓ -ésimo jugador a través del símbolo Π^ℓ , $\ell = 1, \dots, n$.

No debe sorprender al lector que, en realidad la estrategia π^ℓ no esté dada simplemente en términos de las acciones disponibles para el ℓ -ésimo jugador. Considere, por ejemplo el sencillo juego *piedra-papel-tijeras*: en ese juego, la estrategia óptima de cada jugador es la distribución de probabilidad uniforme sobre las acciones posibles (vea, por ejemplo, [16, §2.4.3] y [31, §1.4]). O sea que la decisión óptima de cada jugador es asignar a cada una de las tres acciones admisibles una probabilidad de $1/3$. De hecho, al definir una estrategia como una función de las variables de estado (vea la observación 2.1), en el caso de los juegos diferenciales estocásticos, las estrategias están escritas, además, en términos *del tiempo* en que se da tal estado. Por esta razón, cuando decimos que «los jugadores emplean el par de estrategias (π^1, π^2) a lo largo del juego», escribimos (π_t^1, π_t^2) para especificar la distribución de probabilidad que cada uno debe usar en el tiempo t .

2.1 Ecuaciones diferenciales estocásticas

Suponga que, en el juego de persecución-evasión que presentamos en el ejemplo 2.2, cada jugador desconoce la posición exacta de su contraparte en cada momento. Una manera de modelar esta situación en las variables de estado es la utilización de ecuaciones diferenciales *estocásticas*.² Para ilustrar esta idea, considere el siguiente método para describir los precios de los activos.

Ejemplo 2.3. [36, cap. 2] *Seguramente usted ha oído decir que los precios de los activos financieros evolucionan aleatoriamente a causa de las llamadas hipótesis de mercado eficiente. ¿Qué dicen estas condiciones? En esencia, dicen dos cosas:*

- *La historia pasada está reflejada en el precio actual, pero esto no nos da información sobre el futuro;*
- *Los mercados responden inmediatamente a informaciones nuevas sobre el activo en cuestión.*

Así, la correcta modelación de los precios de los activos depende absolutamente de nuestra aptitud para abstraer las informaciones nuevas que afecten tales precios. Las dos hipótesis recién mencionadas implican

²Las obras de Fima Klebaner [20] y Bernt Øksendal [29] constituyen excelentes introducciones al tema.

que los cambios imprevistos en el precio de un activo son un proceso de Markov.

¿A qué nos referimos cuando hablamos de los cambios en el precio de un activo? Primeramente, note que no nos es útil trabajar con los cambios absolutos. En efecto, un movimiento de un peso es mucho más significativo cuando el precio es de 20 pesos, que cuando es de 200.

Para reflejar esta idea, al cambio de precio en un activo le asociaremos un rendimiento, definido como el cambio en el precio dividido por el precio original.

Ahora suponga que en el tiempo $t \geq 0$, el precio del activo es $S(t)$. Consideremos un pequeño intervalo temporal subsecuente de longitud dt , en el que el precio cambia de $S(t)$ a $S(t) + dS(t)$ como se ve en la figura 6.

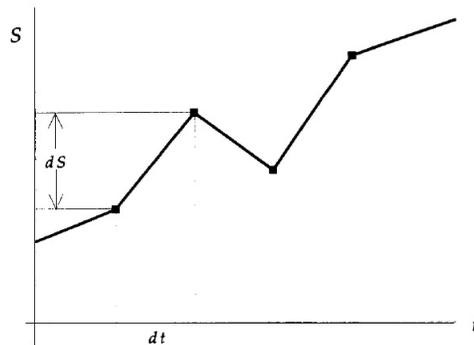


Figura 6. Como el diferencial en el precio del activo es incierto, dS podría ser de tamaños variables, y estar orientado en una dirección distinta a la que mostramos aquí. Por esta razón es que llamamos *caminatas aleatorias* a las trayectorias de los precios de un activo a través del tiempo. (Fuente: [36, cap. 2].)

El modelo más común para describir el rendimiento de un activo $\frac{dS(t)}{S(t)}$ considera dos partes. La primera es predecible y determinista; por esto se llama «libre de riesgo». Su contribución es de tamaño

$$\mu dt,$$

donde μ representa una media del crecimiento promedio del precio del activo (también se le conoce con el nombre de «deriva» o coeficiente de tendencia).

La segunda parte de la contribución al cambio en el precio es de naturaleza aleatoria, y representa la respuesta del precio de un activo a efectos exógenos, como noticias inesperadas. Su contribución es $\sigma dW(t)$, donde σ se llama típicamente «volatilidad» (o coeficiente de difusión) y mide la desviación estándar de los rendimientos. La cantidad $dW(t)$ es una variable aleatoria con distribución Normal con media

0 y varianza dt . Ahora sumamos estas contribuciones y obtenemos la **ecuación diferencial estocástica**:

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dW(t). \quad (5)$$

Esta es una representación matemática de una sencilla receta para generar precios aleatorios. Esta receta recibe el nombre de movimiento browniano geométrico.

El modelo de cambio de precios (5) y el modelo de movimiento visto en (1)-(4) están descritos por ecuaciones diferenciales. De hecho, si $\sigma \equiv 0$ y μ es una constante, la ecuación diferencial (5) puede resolverse exactamente para dar paso al llamado *modelo de interés compuesto convertible continuamente*

$$S(t) = S_0 e^{\mu(t-t_0)},$$

donde S_0 es el precio del activo al tiempo t_0 (vea [19, cap. 1.9]). Así, si la volatilidad es nula, es posible predecir el precio futuro del activo con certeza.

La diferencia entre (1)-(4) y (5) es el símbolo $dW(t)$. ¿Qué representa esta cantidad? La respuesta es: la incertidumbre inherente a los precios de los activos. Se trata de un movimiento browniano, y tiene las características mencionadas anteriormente.

Para ver cómo aplicar la fórmula (5) en teoría de juegos, considere el problema en el que dos inversionistas compiten eligiendo estrategias de inversión, uno para maximizar cierto índice de rendimiento, y otro para minimizarlo. Si suponemos que los agentes tienen a su disposición dos activos, uno riesgoso (cuyo precio evoluciona de acuerdo con (5)), y uno sin riesgo (cuyo precio se actualiza según la ecuación diferencial ordinaria $\frac{dB(t)}{B(t)} = rdt$, donde $r > 0$ es predecible y determinista), entonces podemos definir las riquezas de los jugadores en términos de estas cantidades. En efecto, sea $\pi^\ell(t)$ la proporción de la fortuna del ℓ -ésimo inversionista que se destina al activo riesgoso en el tiempo $t \geq 0$ (naturalmente, $\ell = 1, 2$). Defina $x_\ell(t)$ como la riqueza del ℓ -ésimo jugador en el tiempo $t \geq 0$, entonces la riqueza del ℓ -ésimo jugador obedece a la dinámica

$$\begin{aligned} dx_\ell(t) &= \pi^\ell(t)x_\ell(t)\frac{dS(t)}{S(t)} + (1 - \pi^\ell(t))x_\ell(t)\frac{dW(t)}{W(t)} \\ &= x_\ell(t) [r + \pi^\ell(t)(\mu - r)dt + \pi^\ell(t)\sigma dW(t)], \end{aligned}$$

con $x(0) = x_0$. Esto es exactamente lo que se hace en, por ejemplo [6] y [29, cap. 12].

2.2 Juegos diferenciales estocásticos con modos múltiples

¿Qué pasaría si las posiciones de los agentes en los juegos de los ejemplos 2.1 y 2.2 fueran inciertas para el jugador rival, y además hubiera cambios repentinos en las características de sus trayectorias? Estas características están presentes en el juego *Helicopters vs. Submarines* (la imagen de la derecha es una captura de pantalla —vea [24]—), donde un submarino y un aeroplano utilizan sendos radares para localizarse y dispararse mutuamente, mientras que ambos cambian de ruta, velocidad o altura/profundidad para evitar ser hallados y, por ende, atacados.



Utilizaremos ecuaciones diferenciales estocásticas *con modos múltiples* para modelar esta situación. Para fijar ideas, pensemos nuevamente en el juego de persecución-evasión en el plano del ejemplo 2.2. Usaremos el lado derecho de las ecuaciones (1)-(2) para caracterizar el coeficiente de deriva del evasor *en modo 1*, y el lado derecho de (3)-(4) para el coeficiente correspondiente del perseguidor. Esto es

$$b_E(\vec{x}_P, 1, \phi) = w_1 \begin{pmatrix} \cos[\phi(\vec{x}_P)] \\ \text{sen}[\phi(\vec{x}_P)] \end{pmatrix}, \quad b_P(\vec{x}_E, 1, \psi) = W_1 \begin{pmatrix} \cos[\psi(\vec{x}_E)] \\ \text{sen}[\psi(\vec{x}_E)] \end{pmatrix},$$

donde $\vec{x}_P = (x_P, y_P)$ y $\vec{x}_E = (x_E, y_E)$ y $\phi, \psi : \mathbb{R}^2 \rightarrow \mathbb{R}$; como en la observación 2.2.

Para reflejar la multiplicidad de modos, pensemos que puede ocurrir que, *por alguna razón*, el perseguidor disminuye su velocidad, mientras que el evasor aumenta la suya, o sea $w_2 > W_2$. De esta manera, los coeficientes de deriva de nuestros agentes en el modo 2 lucen así:

$$b_E(\vec{x}_P, 2, \phi) = w_2 \begin{pmatrix} \cos[\phi(\vec{x}_P)] \\ \text{sen}[\phi(\vec{x}_P)] \end{pmatrix}, \quad b_P(\vec{x}_E, 1, \psi) = W_2 \begin{pmatrix} \cos[\psi(\vec{x}_E)] \\ \text{sen}[\psi(\vec{x}_E)] \end{pmatrix}.$$

Note que hemos incorporado la observación 2.2 al escribir la estrategia de cada jugador en función de la posición del otro, i.e. $\phi(\vec{x}_P)$ y $\psi(\vec{x}_E)$. Caracterizamos el cambio entre un modo y otro con una cadena de Markov a tiempo continuo³ $\Theta(\bullet) := \{\theta(t) : t \geq 0\}$. Para efectos de ilustración, supongamos que la matriz de transición de este proceso es

$$P(t) = \frac{1}{2} \begin{pmatrix} 1 + e^{-2\lambda t} & 1 - e^{-2\lambda t} \\ 1 - e^{-2\lambda t} & 1 + e^{-2\lambda t} \end{pmatrix}.$$

³Los capítulos 1, 4 y 5 de [33]; y 4, 5 y 6 de [35] constituyen una excelente introducción al tema.

El llamado *generador infinitesimal*⁴ del proceso $\Theta(\bullet)$ está dado por $\Lambda = \begin{pmatrix} -\lambda & \lambda \\ \lambda & -\lambda \end{pmatrix}$.

Haremos patente la incertidumbre de un jugador sobre la posición del otro (de ahí la necesidad de usar radares) sumando un término no-nulo de difusión a los coeficientes de deriva. Este número dependerá del modo en que se encuentre nuestro sistema. Así, el cambio en la posición de los jugadores quedará determinado por el sistema de ecuaciones diferenciales estocásticas siguiente:

$$d\vec{x}_E(t) = w_\theta \begin{pmatrix} \cos[\phi(\vec{x}_P)] \\ \text{sen}[\phi(\vec{x}_P)] \end{pmatrix} dt + \sigma_\theta dW(t), \quad (6)$$

$$d\vec{x}_P(t) = W_\theta \begin{pmatrix} \cos[\psi(\vec{x}_E)] \\ \text{sen}[\psi(\vec{x}_E)] \end{pmatrix} dt + \sigma_\theta dW(t). \quad (7)$$

Aquí, $\theta = 1, 2$ representa el modo en el que se encuentra el sistema, σ_θ es una constante positiva que simboliza la dificultad que enfrenta cada jugador para rastrear a su oponente, y $dW(t)$ es un movimiento browniano de dimensión 2. De esto se sigue que el cambio en las posiciones de los jugadores queda descrito en realidad, ¡por ocho ecuaciones diferenciales estocásticas!

2.3 Dos resultados clásicos

En general, el sistema de estados de un juego diferencial estocástico con modos múltiples puede describirse a través de las relaciones siguientes.

$$d\vec{x}(t) = b(\vec{x}(t), \theta(t), \pi_t^1, \pi_t^2) dt + \sigma(\vec{x}(t), \theta(t)) dW(t), \quad \vec{x}(0) = \vec{x}_0, \quad (8)$$

$$\begin{aligned} & \mathbb{P}[\theta(t + \Delta t) = j \mid \theta(t) = i; \vec{x}(s), \theta(s) \text{ para } s \leq t] \\ &= \lambda_{ij}^{\pi_t^1, \pi_t^2}(\vec{x}(t)) \Delta t + o(\Delta t), \end{aligned} \quad (9)$$

$$\sum_{j \in \mathcal{S}} \lambda_{ij}^{\pi_t^1, \pi_t^2}(\vec{x}(t)) = 0. \quad (10)$$

⁴El generador infinitesimal, o matriz de intensidades de transición sirve para determinar, bajo ciertas condiciones, la existencia de medidas estacionarias para una cadena de Markov a tiempo continuo. Dada una cadena de Markov a tiempo continuo con matriz de transición $P(t)$, definimos la intensidad de transición entre dos estados (digamos i y j) como la derivada de $P_{ij}(t)$ en $t = 0$ por la derecha. O sea:

$$\lambda_{ij}(t) := \begin{cases} \lim_{t \downarrow 0} \frac{P_{ij}(t) - 1}{t} & \text{si } i = j; \\ \lim_{t \downarrow 0} \frac{P_{ij}(t)}{t}, & \text{si } i \neq j. \end{cases}$$

Si este límite existe, entonces escribiremos $P_{ij}(t) = \lambda_{ij}t + o(t)$ para $i \neq j$, y $P_{ii}(t) = 1 - \lambda_i t + o(t)$; con $\lambda_i := -\lambda_{ii}$; en donde la función $o(t)$ es tal que $o(t)/t \rightarrow 0$ cuando $t \rightarrow 0$ y depende, en general, de los estados i y j . En otras palabras, cuando $t \rightarrow 0$, las probabilidades $P_{ij}(t)$ y $1 - P_{ii}(t)$ son proporcionales a t , excepto por los términos $o(t)$, con constantes de proporcionalidad λ_{ij} y λ_i , respectivamente. Vea [33, cap. 5] y [35, cap. 6.4].

En (8)-(10), supondremos que $\mathcal{S} \subseteq \mathbb{N}$ es el conjunto de modos del sistema, y que Π^1 y Π^2 son las colecciones de estrategias de los jugadores. Asimismo, la función $\theta(t) \in \mathcal{S}$ designará el *modo* en que se encuentran los coeficientes de deriva y difusión del sistema dinámico $\vec{x}(t)$ para $t \geq 0$. Tales coeficientes están dados mediante las funciones $b : \mathbb{R}^m \times \mathcal{S} \times \Pi^1 \times \Pi^2 \rightarrow \mathbb{R}^m$ y $\sigma : \mathbb{R}^m \times \mathcal{S} \rightarrow \mathbb{R}^{m \times d}$, respectivamente. Denotamos con el símbolo $\lambda_{ij}^{\pi_t^1, \pi_t^2}(x(t))$ en (9) a la intensidad de la transición del modo i al modo j cuando los jugadores emplean el par de estrategias $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$.

Así como fue posible definir un valor para el juego del ejemplo 2.1, podemos hacerlo para los juegos que estamos estudiando ahora. Suponga que el ℓ -ésimo jugador cobra $r_\ell(\vec{x}(t), \theta(t), \pi_t^1, \pi_t^2)$ por jugar cuando el sistema se encuentra en $x(t)$. De esta manera podemos definir el índice de rendimiento del ℓ -ésimo jugador en el juego que inicia en el estado (\vec{x}_s, θ_s) como:

$$V^\ell(\vec{x}_s, \theta_s, \pi^1, \pi^2) = \mathbb{E}_{\vec{x}_s, \theta_s}^{\pi^1, \pi^2} \left[\int_s^\infty e^{-\delta(t-s)} r_\ell(\vec{x}(t), \theta(t), \pi_t^1, \pi_t^2) dt \right], \quad (11)$$

donde $\mathbb{E}_{\vec{x}_s, \theta_s}^{\pi^1, \pi^2}[\dots]$ representa el valor esperado (promedio) de la integral de las recompensas descontadas para el ℓ -ésimo jugador a través del tiempo cuando el sistema $(\vec{x}(\bullet), \Theta(\bullet))$ comienza en el estado (\vec{x}_s, θ_s) y los jugadores utilizan las estrategias (π^1, π^2) .

El índice de rendimiento (11) es muy común en la literatura de juegos diferenciales (deterministas, estocásticos, y estocásticos con modos múltiples). La razón es que describe *la suma* de los pagos en valor presente que el ℓ -ésimo jugador recibe a través del tiempo (vea la figura 7).

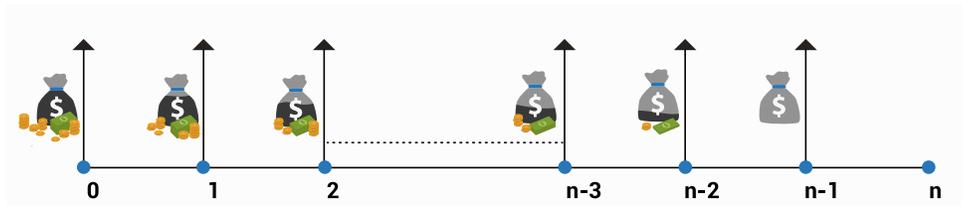


Figura 7. Este dibujo representa los valores presentes de un peso pagadero en $t = 0, 1, \dots, n-1$ (cf. [19, cap. 1]). Note que, a medida que avanza el tiempo, el efecto del *factor de descuento* $e^{-\delta(t-s)}$ se manifiesta en el hecho de que las bolsas de dinero lucen cada vez menos llenas. La expresión (11) representa *la suma* de las bolsas aquí mostradas en *tiempo continuo*.

La maximización de (11) por parte del ℓ -ésimo jugador se conoce como **criterio del pago descontado esperado en horizonte infinito**.

Esto es, el objetivo del primer jugador es hallar

$$V^1(\vec{x}_s, \theta_s, \pi^{1*}, \pi^2) := \sup_{\pi^1 \in \Pi^1} V^1(\vec{x}_s, \theta_s, \pi^1, \pi^2),$$

mientras que el segundo jugador desea encontrar

$$V^2(\vec{x}_s, \theta_s, \pi^1, \pi^{2*}) := \sup_{\pi^2 \in \Pi^2} V^2(\vec{x}_s, \theta_s, \pi^1, \pi^2).$$

Si existe una pareja de estrategias $(\pi^{1*}, \pi^{2*}) \in \Pi^1 \times \Pi^2$ tal que

$$V^1(\vec{x}_s, \theta_s, \pi^1, \pi^{2*}) \leq V^1(\vec{x}_s, \theta_s, \pi^{1*}, \pi^{2*}) \quad \text{y} \quad (12)$$

$$V^2(\vec{x}_s, \theta_s, \pi^{1*}, \pi^2) \leq V^2(\vec{x}_s, \theta_s, \pi^{1*}, \pi^{2*}), \quad (13)$$

decimos que (π^{1*}, π^{2*}) es un equilibrio de Nash para el juego competitivo

$$\Gamma(\vec{x}_s, \theta_s) := (\vec{x}(\bullet), \Theta(\bullet), \vec{x}_s, \theta_s, r_1, r_2, \Pi^1, \Pi^2, \mathcal{S}), \quad (14)$$

y definimos también las *funciones de valor de los jugadores* como

$$V^1(\vec{x}_s, \theta_s) := V^1(\vec{x}_s, \theta_s, \pi^{1*}, \pi^{2*}),$$

$$V^2(\vec{x}_s, \theta_s) := V^2(\vec{x}_s, \theta_s, \pi^{1*}, \pi^{2*}).$$

Observación 2.3. (a): *Note que el juego (14) inicia en el tiempo $t = s$.*

(b): *Compare las desigualdades (12)-(13) con la definición 2.1 y vea que difieren solo en que tales desigualdades incluyen el estado inicial del sistema (\vec{x}_s, θ_s) . Ello nos habla del carácter dinámico del juego.*

Ahora mostramos el resultado principal de esta sección. Se trata de uno de los llamados *Teoremas de verificación* comunes en Teoría de Juegos. Tales teoremas sirven para establecer una conexión entre un sistema de ecuaciones en derivadas parciales *deterministas* (vea (15)-(16) más abajo) con el problema original (maximizar (11) sujeto a (8)-(10)), que es estocástico. Aquí, los símbolos $V_{\vec{x}}$ y $V_{\vec{x}\vec{x}}$ denotan, respectivamente, el vector gradiente y la matriz Hessiana de la función V . Asimismo, $\langle b, c \rangle$ es el producto punto entre los vectores b y c ; σ' representa la transpuesta de la matriz σ y $\text{Tr}(A)$ es la suma de los elementos de la diagonal principal de la matriz A .

Teorema 2.1. (cf. [11, teo. 3.2], [37, teo. 5.8.1].) *Bajo ciertas condiciones⁵, un par de estrategias (π^{1*}, π^{2*}) es un equilibrio de Nash para*

⁵Estas condiciones son:

- que los conjuntos de estrategias Π^1 y Π^2 sean compactos (para garantizar la convergencia de —al menos— una subsucesión de estrategias),
- que las ecuaciones (8)-(10) admitan una solución en el sentido de Markov-Feller (vea [17, cap. 9], [22, cap. 3] y [29, §5.2-3]),
- que el sistema (8)-(10) sea estable en el *sentido de Lyapunov* (vea [15], [22, cap. 5] y [25]),
- que el término de difusión sea positivo-definido (note que esto quiere decir que, en particular, $\sigma \neq 0$),

el juego competitivo $\Gamma(\vec{x}_0, \theta_0)$ si existen funciones $V^1 : \mathbb{R}^m \times \mathcal{S} \rightarrow \mathbb{R}$ y $V^2 : \mathbb{R}^m \times \mathcal{S} \rightarrow \mathbb{R}$ tales que

$$\begin{aligned} \delta V^1(\vec{x}, \theta) &= \frac{1}{2} \text{Tr}[V_{\vec{x}\vec{x}}^1(\vec{x}, \theta) \cdot \sigma(\vec{x}, \theta) \sigma'(\vec{x}, \theta)] \\ &= \sup_{\pi^1} \left\{ r_1(\vec{x}, \theta, \pi^1, \pi^{2*}) + \langle b(\vec{x}, \theta, \pi^1, \pi^{2*}), V_{\vec{x}}^1(\vec{x}, \theta) \rangle \right. \\ &\quad \left. + \sum_{i \neq \theta} \lambda_{i, \theta}^{\pi^1, \pi^{2*}}(\vec{x}) V^1(\vec{x}, \theta) \right\}, \quad (15) \end{aligned}$$

$$\begin{aligned} \delta V^2(\vec{x}, \theta) &= \frac{1}{2} \text{Tr}[V_{\vec{x}\vec{x}}^2(\vec{x}, \theta) \cdot \sigma(\vec{x}, \theta) \sigma'(\vec{x}, \theta)] \\ &= \sup_{\pi^2} \left\{ r_2(\vec{x}, \theta, \pi^{1*}, \pi^2) + \langle b(\vec{x}, \theta, \pi^{1*}, \pi^2), V_{\vec{x}}^2(\vec{x}, \theta) \rangle \right. \\ &\quad \left. + \sum_{i \neq \theta} \lambda_{i, \theta}^{\pi^{1*}, \pi^2}(\vec{x}) V^2(\vec{x}, \theta) \right\}, \quad (16) \end{aligned}$$

para toda pareja $(\vec{x}, \theta) \in \mathbb{R}^m \times \mathcal{S}$. En particular

$$V^\ell(\vec{x}_s, \theta_s) = \mathbb{E}_{\vec{x}_s, \theta_s}^{\pi^{1*}, \pi^{2*}} \left[\int_s^\infty e^{-\delta(t-s)} r_\ell(\vec{x}(t), \theta(t), \pi^{1*}, \pi^{2*}) dt \right]$$

para $\ell = 1, 2$.



En el teorema 2.1, las relaciones (15)-(16) son las llamadas *ecuaciones de Bellman*. Estas ecuaciones fueron bautizadas con este nombre precisamente en honor a su descubridor, Richard Bellman (vea la fotografía a la izquierda —tomada de [4]—). Estas ecuaciones resumen el *principio de Programación Dinámica (PPD)*, que es fruto del trabajo de Bellman en RAND Corporation, de 1949 a 1965. Este principio sirve para un sinnúmero de aplicaciones, que abarca desde el hallazgo de la ruta más corta/larga en un mapa dado, y hasta, por ejemplo, ¡la resolución de juegos diferenciales estocásticos con modos múltiples! El PPD reza así:

- que las funciones $x \mapsto r_1$ y $x \mapsto r_2$ cumplan una condición de Lipschitz uniformemente con respecto a θ , π^1 y π^2 (vea [29, cap. 5]),
- que las funciones $x \mapsto r_1$ y $x \mapsto r_2$ estén dominadas por una función integrable conocida (esto asegura, por el teorema de Fubini, que la esperanza y la integral en (11) son intercambiables).

Sea $\pi^* = \{\pi_t : t \geq 0\}$ una estrategia óptima para un problema de Programación Dinámica y suponga que cuando se usa π^* , el estado $x(r)$ ocurre en el tiempo $r > 0$ con probabilidad positiva. Considere pues el subproblema en el que nos encontramos en el estado $x(r)$ si deseamos maximizar la recompensa (11) a partir del tiempo r . Entonces la estrategia *trunca*

$$\varphi^* := \{\pi_t : t \geq r\}$$

es óptima para este subproblema.

Para ver cómo las ecuaciones (15)-(16) equivalen al PPD (al menos en el caso diferencial estocástico), vea [9, §I.4-5, III.7-9] y [39, §3.2 y 3.3]. La clave está en la expresión

$$\begin{aligned} \langle b(\vec{x}, \theta, \pi_t^1, \pi_t^2), V_{\vec{x}}^\ell(\vec{x}, \theta) \rangle + \frac{1}{2} \text{Tr}[V_{\vec{x}\vec{x}}^\ell(\vec{x}, \theta) \cdot \sigma(\vec{x}, \theta) \sigma'(\vec{x}, \theta)] \\ + \sum_{i \neq \theta} \lambda_{i, \theta}^{\pi_t^1, \pi_t^2}(\vec{x}) V^\ell(\vec{x}, \theta) \end{aligned} \quad (17)$$

para $\ell = 1, 2$ en las ecuaciones de Bellman (15)-(16), pues (17) es, precisamente el generador infinitesimal del proceso $(x(\bullet), \Theta(\bullet))$, aunque también puede interpretarse como la esperanza de la *derivada* —en el sentido de Itô— de la función $V^\ell(x(t), \theta(t))$. (Vea [22, cap. 3] y [29, cap. 5] para conocer los detalles.)

Cerramos la sección con un resultado clásico más. Se trata de una consecuencia del teorema 2.1 y nos dice que, en un juego de *suma-cero* las funciones de valor son soluciones *clásicas* de las ecuaciones (15)-(16); mientras que en el caso de los juegos de suma *no-cero*, tales funciones simplemente son soluciones débiles de las mismas ecuaciones.

Corolario 2.1. *Si $r_1 = -r_2$, entonces las funciones V^1, V^2 son dos veces continuamente diferenciables en \vec{x} (cf. [11, teo. 3.1]). Si $r_1 \neq -r_2$ entonces V^1, V^2 son miembros del espacio de Sobolev⁶ $\mathcal{W}^{2,p}(\mathbb{R}^m \times \mathcal{S})$ para $2 \leq p < \infty$ (vea [11, lema 4.1]).*

3. ¡Júntense!

En la sección anterior presentamos una panorámica del uso de ciertos juegos como auxiliar para analizar situaciones en las que la competencia es natural. Sin embargo, éste no es el único ejemplo en que las herramientas de la teoría de juegos son útiles. Ciertamente, puede ser complicado imaginar una situación no-competitiva en la que se use la Teoría de Juegos: ¡hasta en los juegos de mesa es difícil! Pero eche un vistazo a la figura 8 ...

⁶El capítulo 5 de [8] es una excelente introducción a este tema.

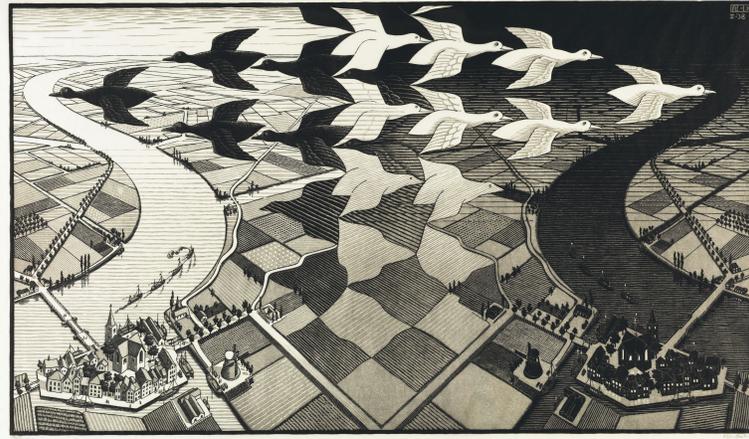


Figura 8. Aquí mostramos dos juegos de mesa que tienen la característica de ser colaborativos. Ambos son juegos de autor. El que mostramos en la parte izquierda de la gráfica se llama *Pandemic* y es obra de M. Leacock; el de la derecha es *Mysterium*, y se lo debemos a O. Nevskiy y O. Sidorenko. (Puede ver más sobre estos juegos en [21], y [27], respectivamente.)

También es posible usar los resultados para juegos diferenciales estocásticos de modos múltiples de los que hemos hablado en un contexto de colaboración. Ahora mostramos un par de ejemplos.

3.1 Parvadas viajeras

La coordinación entre los animales sociales requiere una transferencia rápida y eficiente de información entre los individuos, que puede depender crucialmente de la estructura subyacente de la red de comunicación. Establecer los circuitos de decisión y las redes que dan lugar al comportamiento individual ha sido un objetivo central de la neurociencia. En [34], por ejemplo, se estudia el problema análogo de determinar la estructura de la red de comunicación entre los organismos que da lugar a un comportamiento colectivo coordinado, como el que exhiben las aves migratorias representadas aquí (vea [7]). Trabajos como el de Rosenthal et al. han permitido llevar un poco más allá las redes visuales, y pueden ser clave para saber qué pájaro inicia un cambio de dirección de la parvada.

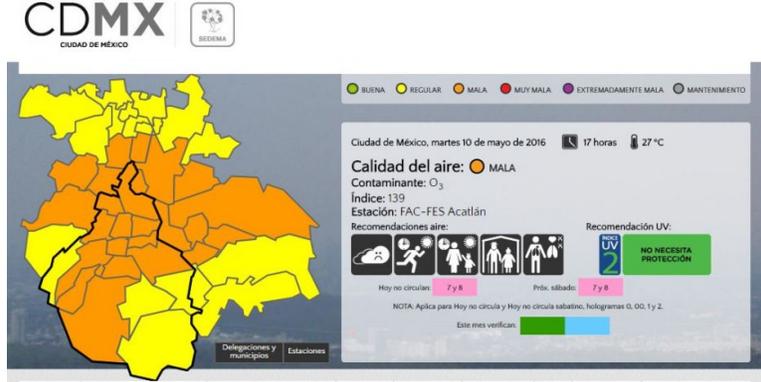


Note que la situación que estamos describiendo corresponde, precisamente, a un juego diferencial estocástico con modos múltiples. Ciertamente, la posición de cada paloma queda determinada por un sistema como (8), donde la estrategia π^ℓ que determina el movimiento del ℓ -ésimo pájaro está afectada por las posiciones de los pájaros de la izquierda, la derecha, arriba, abajo y adelante (uno de los resultados que se presentan en [34] es que las aves que vuelan en parvada se ubican siguiendo simplemente a sus vecinos más cercanos). La aleatoriedad modelada a través del movimiento browniano en (8) puede ser, en este caso, el efecto «impredecible» del viento o de los contaminantes en el aire. El modo θ en que se encuentra el sistema (8)-(10) es, en términos del trabajo [34], si la parvada se encuentra evadiendo un peligro, o no.

3.2 El principio de racionalidad grupal

Enfrentémoslo. La CDMX y el área metropolitana conforman una ciudad muy contaminada. Pero, ¿y si las demarcaciones que la componen fijaran como función de «recompensa» el costo que les impone tener un alto⁷ Índice Metropolitano de la Calidad del Aire (IMECA), y trataran de minimizar los de todas ellas de manera conjunta? (Entre a [12] para ver información en tiempo real.)

⁷<http://www.aire.df.gob.mx>



Esto es:

$$\mathbb{E}_{\vec{x}_0, \theta_0}^{\phi^1, \dots, \phi^m} \left[\int_0^\infty e^{-\delta t} (r_1(\vec{x}(t), \theta(t), \phi_t^1, \dots, \phi_t^m) + \dots + r_m(\vec{x}(t), \theta(t), \phi_t^1, \dots, \phi_t^m)) dt \right]. \quad (18)$$

Aquí, la ℓ -ésima entrada del vector $\vec{x} \in \mathbb{R}^m$ representa el IMECA de la ℓ -ésima demarcación que compone la CDMX, mientras que el movimiento browniano puede representar los cambios «impredecibles» del viento. Asimismo, el modo del sistema θ es la calidad del aire en que se encuentra el sistema (buena, regular, mala, muy mala, extremadamente mala y mantenimiento).

La clave de la expresión (18) está en la suma de las «recompensas» (costos) de las m demarcaciones: los agentes deciden colaborar para minimizar de manera conjunta un costo que les afecta a todos. En Teoría de Juegos esto se llama **racionalidad de grupo**.

A continuación planteamos el criterio de optimización en que estamos interesados a partir del concepto de racionalidad de grupo. La definición siguiente está pensada para el caso en que cada agente está interesado en maximizar su propia recompensa, y especifica lo que ocurre para solo dos jugadores, pero sus extensiones para el caso en que lo que desean los agentes es minimizar un costo, y para más «demarcaciones» es evidente.

Definición 3.1. Sea

$$W(\vec{x}_0, \theta_0, \phi^1, \phi^2) := \mathbb{E}_{\vec{x}_0, \theta_0}^{\phi^1, \phi^2} \left[\int_0^\infty e^{-\delta t} (r_1(\vec{x}(t), \theta(t), \phi_t^1, \phi_t^2) + r_2(\vec{x}(t), \theta(t), \phi_t^1, \phi_t^2)) dt \right]$$

el pago que percibe una coalición en un juego colaborativo de dos jugadores que usan las estrategias $\phi^1 \in \Pi^1$ y $\phi^2 \in \Pi^2$. Decimos que el par

$(\phi^{1*}, \phi^{2*}) \in \Pi^1 \times \Pi^2$ es *óptimo* para el juego colaborativo

$$\Gamma_c(x_0, \theta_0) := (\vec{x}(\bullet), \Theta(\bullet), \vec{x}_s, \theta_s, r_1 + r_2, \Pi^1, \Pi^2, \mathcal{S})$$

si

$$W(\vec{x}_0, \theta_0, \phi^{1*}, \phi^{2*}) := \sup_{(\phi^1, \phi^2)} W(\vec{x}_0, \theta_0, \phi^1, \phi^2).$$

Cerramos esta sección con un teorema de verificación análogo al teorema 2.1.

Teorema 3.1. (cf. [37, teo 5.8.2].) *Bajo ciertas condiciones⁸, un par de estrategias (ϕ^{1*}, ϕ^{2*}) es óptimo para el juego colaborativo $\Gamma_c(x_0, \theta_0)$ si existe una función $W : \mathbb{R}^m \times \mathcal{S} \rightarrow \mathbb{R}$ tal que*

$$\begin{aligned} & \delta W(\vec{x}, \theta) - \frac{1}{2} \text{Tr}[(W_{\vec{x}\vec{x}}(\vec{x}, \theta) \cdot \sigma(\vec{x}, \theta) \sigma'(\vec{x}, \theta))] \\ &= \sup_{(\phi_t^1, \phi_t^2)} \left\{ [r_1(\vec{x}, \theta, \phi_t^1, \phi_t^2) + r_2(\vec{x}, \theta, \phi_t^1, \phi_t^2)] \right. \\ & \quad \left. + \langle b(\vec{x}, \theta, \phi_t^1, \phi_t^2), W_{\vec{x}}(\vec{x}, \theta) \rangle + \sum_{i \neq \theta} \lambda_{i, \theta}^{\phi_t^1, \phi_t^2}(\vec{x}) W(\vec{x}, \theta) \right\} \quad (19) \end{aligned}$$

para todo $(\vec{x}, \theta) \in \mathbb{R}^m \times \mathcal{S}$. En particular

$$W(\vec{x}_s, \theta_s) = W(\vec{x}_s, \theta_s, \phi^{1*}, \phi^{2*}).$$

Observación 3.1. *Por el corolario 6.2 en [10], sabemos que la función $W(\vec{x}, \theta)$ a la que se refiere el teorema 3.1 es dos veces continuamente diferenciable.*

3.3 Racionalidad individual y consistencia de los subjuegos

Una de las condiciones suficientes para que el teorema 3.1 rija es la racionalidad individual, otra es la consistencia de los subjuegos. Presentamos ambas hipótesis en esta sección. El principio de racionalidad individual se requiere para asegurar que la ganancia de cada participante de la coalición sea al menos la que obtendría en el caso competitivo (vea [37, §4.2 y 5.2] y [38, cap. 1]). La figura 9 ilustra esta idea.

Definimos implícitamente las *recompensas instantáneas* $(B_1(t), B_2(t))$ para el juego $\Gamma_c(x_0^*, \theta_0^*)$ mediante las relaciones

$$\xi^\ell(\vec{x}^*(\tau), \theta^*(\tau)) = \mathbb{E}_{\vec{x}^*(\tau), \theta^*(\tau)} \left[\int_{\tau}^{\infty} e^{-\delta(t-\tau)} B_\ell(t) dt \right], \quad (20)$$

para $\ell = 1, 2$, donde $(\vec{x}^*(\cdot), \theta^*(\cdot))$ es la trayectoria que obtenemos al insertar el par óptimo de estrategias (ϕ^{1*}, ϕ^{2*}) en (8)-(10).

⁸En realidad son las mismas hipótesis del teorema 2.1, la racionalidad de grupo y las condiciones que dan nombre a la sección 3.3.

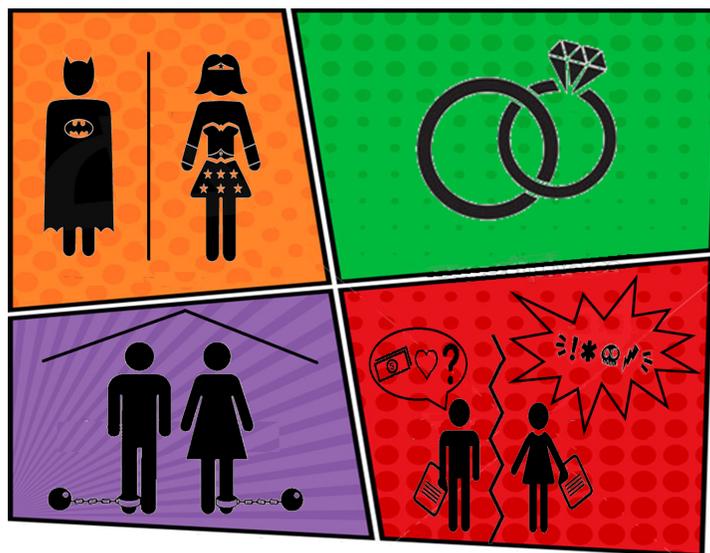


Figura 9. Suponga que usted es un súper-héroe (o una súper-heroína). ¿Por qué participar en una coalición si ésta no le reporta —al menos— las mismas ventajas que tendría luchando solo(a) contra la injusticia? Más aún, si la coalición terminara, usted esperaría recuperar esas ventajas. Si no, incurriría usted en una acción irracional como individuo. ¿No es así? (Fuente: Elaboración propia.)

Note que las funciones de recompensa instantánea B_1 y B_2 **no** dependen (explícitamente) de las estrategias de los jugadores, sino, simplemente, del instante t . La razón es que los jugadores trabajan para que el sistema (8)-(10) siga la trayectoria $(x(t)^*, \theta(t)^*)$ en cada instante.

El **principio de racionalidad individual** puede expresarse matemáticamente de la manera siguiente:

$$\xi^\ell(\vec{x}^*(t), \theta^*(t)) \geq V^\ell(\vec{x}^*(t), \theta^*(t)) \quad (21)$$

para $\ell = 1, 2$ y $t \geq 0$.

La función ξ^ℓ se llama *reemplazo de la ganancia descontada* (11) para $\ell = 1, 2$ en el juego $\Gamma_c(\vec{x}_0, \theta_0)$ (vea [37, §3.2] y [38, §3.2]).

Para introducir la consistencia de los subjuegos, tomamos $\tau \in [0, \infty[$ y $t \geq \tau$. Ahora definimos $\gamma(\tau; t, \vec{x}^*(t), \theta^*(t))$ como la esperanza del *valor presente en tiempo τ de una sucesión de pagos que comienzan en el tiempo $t \geq \tau$ para el ℓ -ésimo jugador* (vea la figura 10). Esto es

$$\begin{aligned} & \gamma(\tau; t, \vec{x}^*(t), \theta^*(t)) \\ & := \mathbb{E}_{\vec{x}^*(\tau), \theta^*(\tau)} \left[\int_t^\infty e^{-\delta(s-\tau)} B_\ell(s) ds \mid \vec{x}^*(t) = \vec{x}_t^*, \theta^*(t) = \theta_t^* \right]. \quad (22) \end{aligned}$$

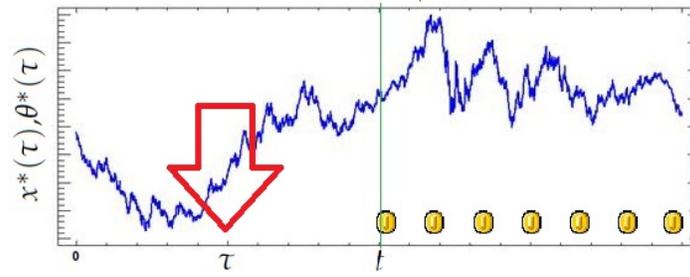


Figura 10. Aquí mostramos una abstracción del valor presente de una anualidad contingente diferida por $t-\tau$ períodos pagadera continuamente a partir de t , suponiendo que el sistema (8)-(10) sigue la trayectoria $(\vec{x}^*(t), \theta^*(t))$. Vea [5, cap. 5] y [19, cap. 3]. (Fuente: Elaboración propia.)

Note que, si las recompensas instantáneas (B_1, B_2) están bien definidas, el PPD (sección 2.3) nos da lo siguiente.

$$\begin{aligned}
 \gamma^\ell(\tau; t, \vec{x}^*(t), \theta^*(t)) &= e^{-\delta(t-\tau)} \mathbb{E}_{\vec{x}^*(t), \theta^*(t)} \left[\int_t^\infty e^{-\delta(s-t)} B_\ell(s) ds \right] \\
 &= e^{-\delta(t-\tau)} \xi^\ell(\vec{x}^*(t), \theta^*(t)) \\
 &= e^{-\delta(t-\tau)} \gamma^\ell(t; t, \vec{x}^*(t), \theta^*(t)).
 \end{aligned} \tag{23}$$

Cuando esta igualdad rige para todo $t \geq 0$ decimos que los **subjuegos** de $\Gamma_c(\vec{x}_0, \theta_0)$ son **consistentes en el tiempo**.

4. El procedimiento de asignación de recompensas instantáneas

En la sección 3.3 establecimos que, por el principio de racionalidad individual, las recompensas de cada uno de los miembros de una coalición deben ser al menos tan grandes como lo serían si estuvieran compitiendo entre ellos (cf. (21)). Más aún, al introducir (20) utilizamos la expresión «definir implícitamente» las recompensas instantáneas de los agentes. Pero hasta ahora no sabemos quiénes son esas funciones, ni cómo se calculan.

En esta sección vamos a despejar esa cuestión mediante el procedimiento de asignación de recompensas instantáneas introducido por Leon Petrosyan en 1977 (vea [30] y la fotografía a la derecha). El método consiste, en primer lugar, en una invocación a nuestra hipótesis de consistencia de los subjuegos, i.e. a la ecuación (23); y después, en



la correcta utilización de la expresión (17) interpretada como la derivada de una función valuada en el proceso $(\vec{x}(\bullet), \Theta(\bullet))$. Esto nos dejará con un sistema de ecuaciones diferenciales en términos de las funciones de reemplazo de la ganancia descontada $\xi^\ell(\vec{x}, \theta)$ ($\ell = 1, 2$), de manera que terminaremos por simplemente elegir una forma para estas funciones.

Note que, gracias a la consistencia de los subjuegos en $\Gamma_c(\vec{x}_\tau^*, \theta_\tau^*)$, podemos escribir

$$\begin{aligned} & \gamma^\ell(\tau; \tau, \vec{x}^*(\tau), \theta^*(\tau)) \\ = & \mathbb{E}_{\vec{x}^*(\tau), \theta^*(\tau)} \left[\int_{\tau}^{\tau+\Delta t} e^{-\delta(s-\tau)} B_\ell(s) ds \right. \\ & \left. + e^{-\delta\Delta t} \gamma^\ell(\tau + \Delta t; \tau + \Delta t, \vec{x}_\tau^* + \Delta \vec{x}_\tau^*, \theta^*(\tau + \Delta t)) \mid \vec{x}^*(\tau) = \vec{x}_\tau^*, \theta^*(\tau) = \theta_\tau^* \right] \\ = & \mathbb{E}_{\vec{x}^*(\tau), \theta^*(\tau)} \left[\int_{\tau}^{\tau+\Delta t} e^{-\delta(s-\tau)} B_\ell(s) ds \right. \\ & \left. + \gamma^\ell(\tau; \tau + \Delta t, \vec{x}_\tau^* + \Delta \vec{x}_\tau^*, \theta^*(\tau + \Delta t)) \mid \vec{x}^*(\tau) = \vec{x}_\tau^*, \theta^*(\tau) = \theta_\tau^* \right]. \end{aligned}$$

La última igualdad se sigue de (23). Un despeje nos da

$$\begin{aligned} & \mathbb{E}_{\vec{x}^*(\tau), \theta^*(\tau)} \left[\int_{\tau}^{\tau+\Delta t} e^{-\delta(s-\tau)} B_\ell(s) ds \mid \vec{x}^*(\tau) = \vec{x}_\tau^*, \theta^*(\tau) = \theta_\tau^* \right] \\ = & \mathbb{E}_{\vec{x}^*(\tau), \theta^*(\tau)} \left[\gamma^\ell(\tau; \tau + \Delta t, \vec{x}_\tau^*, \theta_\tau^*) - \gamma^\ell(\tau; \tau + \Delta t, \vec{x}_\tau^* + \Delta \vec{x}_\tau^*, \theta^*(\tau + \Delta t)) \right]. \end{aligned}$$

Divida por Δt , tome el límite cuando $\Delta t \downarrow 0$, y use (17) para ver que

$$\begin{aligned} B_\ell(\tau) &= \delta \xi^\ell(\vec{x}_\tau^*, \theta_\tau^*) - \frac{1}{2} \text{Tr}[(\xi_{\vec{x}\vec{x}}^\ell(\vec{x}_\tau^*, \theta_\tau^*) \cdot \sigma(\vec{x}, \theta) \sigma'(\vec{x}_\tau^*, \theta_\tau^*)) \\ &\quad - \langle b(\vec{x}_\tau^*, \theta_\tau^*, \phi_\tau^{1*}, \phi_\tau^{2*}), \xi_{\vec{x}}^\ell(\vec{x}_\tau^*, \theta_\tau^*) \rangle \\ &\quad - \sum_{i \neq \theta} \lambda_{i, \theta}^{\phi_\tau^{1*}, \phi_\tau^{2*}}(\vec{x}_\tau^*) \xi^\ell(\vec{x}_\tau^*, \theta_\tau^*)] \end{aligned} \quad (24)$$

es una expresión funcional de las recompensas instantáneas que depende de la forma de la función de reemplazo de la ganancia descontada $\xi^\ell(\vec{x}_\tau^*, \theta_\tau^*)$ (cf. [37, teo. 5.8.3]).

Finalmente, elija alguna forma para las funciones de reemplazo de la ganancia descontada $\xi^\ell(\vec{x}_\tau^*, \theta_\tau^*)$. Por ejemplo, si la coalición está compuesta de partes de tamaños iguales, una opción interesante puede ser la llamada función de reemplazo de ganancia descontada *de negociación de Nash*. O sea:

$$\begin{aligned} \xi_{NB}^\ell(\vec{x}_\tau^*, \theta_\tau^*) &= V^\ell(\vec{x}_\tau^*, \theta_\tau^*) \\ &\quad + \frac{1}{2} [W(\vec{x}_\tau^*, \theta_\tau^*, \tau) - V^1(\vec{x}_\tau^*, \theta_\tau^*) - V^2(\vec{x}_\tau^*, \theta_\tau^*)]. \end{aligned} \quad (25)$$

Si no es el caso, y los agentes deciden dividir las ganancias de acuerdo a su nivel de participación en la coalición, es posible usar la función de reemplazo de ganancia descontada *de Pareto*:

$$\begin{aligned} \xi_P^\ell(\vec{x}_\tau^*, \theta_\tau^*) &= V^\ell(\vec{x}_\tau^*, \theta_\tau^*) \\ &+ \frac{V^\ell(\vec{x}_\tau^*, \theta_\tau^*)}{V^1(\vec{x}_\tau^*, \theta_\tau^*) + V^2(\vec{x}_\tau^*, \theta_\tau^*)} [W(\vec{x}_\tau^*, \theta_\tau^*, \tau) - V^1(\vec{x}_\tau^*, \theta_\tau^*) - V^2(\vec{x}_\tau^*, \theta_\tau^*)]. \end{aligned} \quad (26)$$

Las relaciones (24) y (25); y (24) y (26) representan soluciones completas al problema de distribuir las ganancias a través del tiempo entre los miembros de una coalición. La diferencia entre cada una de estas soluciones radica en la forma en la que los agentes contribuyen a la alianza.

Vale la pena destacar las hipótesis que nos condujeron a estas soluciones.

- Como tanto (25), como (26) están escritas en términos de las soluciones a las ecuaciones de Bellman (15)-(16) y (19), es necesario que los teoremas 2.1 y 3.1 rijan. De tal manera que sus condiciones son fundamentales.
- La racionalidad de grupo (18) es necesaria para definir el juego colaborativo $\Gamma_c(\vec{x}_0, \theta_0)$.
- La racionalidad individual (21) debe cumplirse para que los agentes que conforman la coalición no tengan incentivos para dejarla en ningún momento $t \geq 0$.
- La condición de consistencia de los subjugos a través del tiempo (23) es crucial para asegurar la posibilidad de escribir (24).

Finalizamos esta sección con un comentario sobre la diferenciabilidad de las funciones de valor de los juegos competitivo y colaborativo.

Observación 4.1. *Suponga que estamos interesados en obtener una ganancia monetaria para la coalición. En virtud del corolario 2.1, sabemos que cuando tenemos un juego de **suma cero**, las funciones de valor del juego son dos veces continuamente diferenciables. Cosa que simplificaría el cálculo de las recompensas instantáneas a través de las relaciones (24)-(25) y (24)-(26). Sin embargo, esto no tiene ningún sentido en términos de la racionalidad grupal. La razón es que, por el principio mencionado, lo primero que debemos hacer es sumar las recompensas originales! Esto nos dejará con un pago nulo para la coalición, y eliminaría, por ende, la posibilidad de la cooperación.*

El caso verdaderamente no-trivial es el de la suma no-cero. El problema que este caso entraña es que, desafortunadamente (por el mismo

corolario 2.1), no es posible encontrar soluciones clásicas⁹ a las ecuaciones (15)-(16). Esto nos llevará inexorablemente a trabajar en espacios de soluciones más débiles. Una propuesta es la de hacerlo con espacios de Sobolev.

Por otro lado, si lo que nos interesa es simplemente evitar la tragedia de los comunes (o sea que no nos interese obtener una ganancia monetaria para la coalición), el contexto competitivo bastará, siempre y cuando todos los competidores tengan en cuenta las restricciones existentes sobre el sistema.

5. Conclusiones

En este trabajo hemos presentado juegos competitivos de varios grados de complejidad desde la perspectiva del problema de persecución-evasión. Sin embargo, hemos argumentado la conveniencia de la consideración de juegos en los que los agentes colaboren entre ellos para lograr la optimización de un índice de rendimiento común. Más aún, hemos discutido la necesidad de asignar recompensas instantáneas a cada miembro de la coalición y hemos presentado el método propuesto por Petrosyan para asignar las recompensas instantáneas. Hemos decidido hacer la presentación de los juegos colaborativos y del algoritmo de asignación de recompensas instantáneas en términos de los juegos diferenciales estocásticos de modos múltiples. Debe tomarse en cuenta que, sin abundar en los detalles técnicos, hemos exhibido las principales hipótesis que deben usarse para garantizar la existencia de funciones de valor, equilibrios de Nash y equilibrios colaborativos.

Nuestra presentación ha privilegiado los juegos diferenciales estocásticos de modos múltiples en horizonte infinito porque se trata de un tema de actualidad en Teoría de Juegos y por ello se alinea con la intención del autor de hacer un breve recorrido por algunos de los resultados y las técnicas que utilizan los expertos de hoy en esta apasionante disciplina.

Agradecimiento especial

El autor agradece los valiosos comentarios de la editora en jefe de Miscelánea Matemática, la Dra. Ana Meda Guardiola, así como de los dos árbitros anónimos que revisaron este texto, y, por supuesto, las observaciones de las Mtras. Araceli Sánchez Venegas y Sylvanna Barona; y del Dr. Enrique Lemus Rockdríguez. Todos ellos contribuyeron a mejorar la presentación final de esta obra. Asimismo, el autor agradece la

⁹O sea, «dos veces continuamente diferenciables».

dedicación y el esfuerzo de la Lic. Priscilla Carmago Bacha, diseñadora de la Facultad de Ciencias Actuariales de la Universidad Anáhuac México, por su invaluable apoyo con la elaboración de las imágenes de las figuras 4, 5 y 7.

Por último, este autor desea reconocer la trayectoria del Dr. Juan González Hernández, pues fue uno de sus inspiradores para seguir el apasionante camino de la investigación en Matemáticas, y en particular, en teoría de juegos.

Bibliografía

- [1] E. Alcaráz, «Donald Trump's view o' the world», 2015, Dibujo. Recuperado de <https://lalo-alcaraz-art-shop.myshopify.com/products/donald-trumps-view-o-the-world-print>.
- [2] A. Arapostathis y V. Borkar, «Uniform recurrence properties of controlled diffusions and applications to optimal control», *SIAM J. Control Optim.*, vol. 48, 2010, 4181–4223.
- [3] Y. Aşamasi, «Star Wars Artwork No. 2», 2015, Pintura al óleo. Recuperado de: <http://www.animasyongastesi.com/star-wars-artwork-yapim-asamasi/>.
- [4] R. Bellman, *Eye of the Hurricane: An Autobiography*, World Scientific Pub. Co. Inc., 1984.
- [5] N. Bowers, H. Gerber, J. Hickman, D. Jones y C. Nesbitt, *Actuarial Mathematics*, The Society of Actuaries, Schaumburg, Illinois, 1997.
- [6] S. Browne, «Stochastic Differential Portfolio Games», *Journal of Applied Probability*, vol. 37(1), 2000, 126–147.
- [7] M. Escher, «Day and Night», 1938, Fotografía. Recuperado de: <https://www.escherinhetpaleis.nl/showpiece/day-and-night/?lang=en>.
- [8] L. Evans, *Partial Differential Equations*, Graduate Studies in Mathematics, vol. 19, American Mathematical Society, 2010.
- [9] W. Fleming y M. Soner, *Controlled Markov process and viscosity solutions*, Springer, New York, 2006.
- [10] M. Ghosh, A. Arapostathis y S. Marcus, «Optimal control of switching diffusions with application to flexible manufacturing systems», *SIAM J. Control Optim.*, vol. 30, 1992, 1–23.
- [11] M. Ghosh y S. Marcus, «Stochastic differential games with multiple modes», *Stochastic Analysis and Applications*, vol. 16, 1998, 91–105.
- [12] Gobierno de la Ciudad de México, «Sistema de Monitoreo Atmosférico de la Ciudad de México», 2017, Captura de pantalla. Recuperado de: <http://www.aire.cdmx.gob.mx/default.php>.
- [13] Y. Gujjar, «Steering Pursue and Evade», 2018, Captura de pantalla. Recuperado de: <https://yashasgujjar.com/tech-demos/>.
- [14] A. Gupta y J. Matthews, «Twilight Struggle», 2009, Fotografía. Recuperado de <https://boardgamegeek.com/boardgame/12333/twilight-struggle>.
- [15] R. Has'minskii, *Stochastic Stability of Differential Equations*, Sijthoff and Noordhoff, Germantown, Md., 1980.
- [16] A. Haurie, J. Krawczyk y G. Zaccour, *Games and Dynamic Games*, World Scientific Publishing, Singapore, 2012.
- [17] O. Hernández-Lerma, *Lecture Notes on Continuous-time Markov Control Processes*, Aportaciones Matemáticas, Sociedad Matemática Mexicana. México, 1994.
- [18] R. Isaacs, *Differential Games: A Mathematical Theory with Applications to Welfare and Pursuit, Control and Optimization*, John Wiley and Sons, Inc., New York, 1965.

- [19] S. Kellison, *The Theory of Interest*, McGraw Hill/Irwin Professional Pub, 2008.
- [20] F. Klebaner, *Introduction to Stochastic Calculus with Applications*, Imperial College Press, UK, 2012.
- [21] M. Leacock, «Pandemic», 2008, Fotografía. Recuperado de: <https://boardgamegeek.com/boardgame/30549/pandemic>.
- [22] X. Mao y C. Yuan, *Stochastic Differential Equations with Markovian Switching*, World Scientific Publishing, Singapore, 2006.
- [23] Media Graffiti Studio, «X-wing and Tie fighter», 2010, Cartel. Recuperado de: <https://www.etsy.com/ie/listing/206903818/star-wars-art-x-wing-and-tie-fighter>.
- [24] Melocoton Wars, «Helicopters vs Submarines–War Machines Battle», 2016, Captura de pantalla. Recuperado de <https://appgrooves.com/android/com.melocotonwars.helicoptersvssubmarines/helicopters-vs-submarines-war-machines-battle/melocoton-wars/>.
- [25] S. Meyn y R. Tweedie, «Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes», *Adv. Appl. Prob.*, vol. 25, 1993, 518–548.
- [26] S. Nasar, *A beautiful mind: a biography of John Forbes Nash, Jr., winner of the Nobel Prize in economics 1994*, Simon & Schuster, USA, 1998.
- [27] O. Nevskiy y O. Sidorenko, «Mysterium», 2015, Fotografía. Recuperado de: <https://boardgamegeek.com/boardgame/181304/mysterium>.
- [28] T. Nishiyama y H. Matsumoto, «Street Fighter II», 1991, Captura de pantalla. Recuperado de <http://www.capcom.com>.
- [29] B. Øksendal, *Stochastic Differential Equations. An Introduction with Applications*, Springer-Verlag, Berlin, 2003.
- [30] L. Petrosyan, «Stability of the solutions in differential games with several players (en ruso)», *Vestnik Leningrad. Univ.*, vol. 19, 1977, 46–52.
- [31] L. Petrosyan y N. Zenkevich, *Game Theory*, World Scientific, Singapore, 1996.
- [32] Rand Corporation, «Emblema», 1949, Captura de Pantalla. Recuperado de: <https://www.rand.org/about.html>.
- [33] S. Resnick, *Adventures in Stochastic Processes*, Birkhäuser, Basel, 2002.
- [34] S. Rosenthal, C. Twomey, A. Hartnett, H. Wu y I. Couzinb, «Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion», *Proc. Natl. Acad. Sci.*, vol. 112(15), 2015, 4690–4695.
- [35] S. Ross, *Introduction to Probability models*, Academic Press, San Diego, CA, 2003.
- [36] P. Wilmott, P. Dewynne y S. Howison, *The Mathematics of Financial Derivatives: A Student Introduction*, Cambridge University Press, 1995.
- [37] D. Yeung y L. Petrosyan, *Cooperative Stochastic Differential Games*, Springer, New York, 2006.
- [38] ———, *Subgame Consistent Cooperation. A Comprehensive Treatise*, Springer, Singapore, 2016.
- [39] J. Yong y X. Zhou, *Stochastic controls*, Springer-Verlag, New York, 1999.
- [40] P. Yu, «An appreciation of professor Rufus Isaacs», *Journal of Optimization Theory and Applications*, vol. 27(1), 1979, 1–6.