TRABAJOS DE INVESTIGACION OPERATIVA
Vol. 7, núm. 1, 1990, pp. 63 a 76

JUEGOS ESTOCASTICOS CONTINUOS: VALOR Y ESTRATEGIAS OPTIMAS

M. ANGELES MURUAGA Y R. VÉLEZ
UNED

RESUMEN

El objeto de este trabajo es analizar los juegos estocásticos cuyo espacio de estados y de acciones son métricos compactos, con adecuadas condiciones de continuidad acerca de las funciones de pago y de transición. Tras describir el modelo e introducir las hipótesis de continuidad, se trata el problema con horizonte finito, a fin de probar que existe valor y estrategias óptimas para ambos jugadores, que pueden ser determinados recurrentemente. También se considera el caso de horizonte infinito en presencia de un factor de descuento. El resultado final, en este caso, fue obtenido en Maitra-Parthasarathy (1970) mediante una demostración considerablemente más complicada. La simplificación se basa en la introducción de un pago terminal, que puede ser adecuadamente elegido a fin de obtener las conclusiones deseadas.

ABSTRACT

The aim of this paper is to analyse stochastic games with state and action spaces wich are metric and compact and with suitable continuity conditions about the payoff and transition functions. After the description of the model and the introduction of the continuity assumptions, the finite horizon problem is analyzed, in order to prove that a value exists and both players have optimal strategies, that can be recursively determined. The discounted infinite horizon case is also considered. The final result in this case was obtained in Maitra-Parthasarathy (1970) with a considerably more cumbersome proof. The simpli-

Recibido: Marzo 1991.
Revisado: Septiembre 1991.

fication is based in the introduction of a terminal payoff, that can be suitably choosed in order to get the desired conclusions.
Key words: Continuos stochastic games. Discount factor. Value functions. Optimal strategies.
AMS 1980-Subject Classification: 90D15-93E05.

1. INTRODUCCION

Los juegos estocásticos discretos fueron introducidos por Shapley (1953). Consisten básicamente en una sucesión de juegos finitos a los que se juega en sucesivas etapas. En cada una de ellas se elige al azar el juego siguiente con probabilidades que dependen de las estrategias escogidas por los jugadores en el juego anterior.

Shapley consideró tanto el caso de número definido de etapas (horizonte finito), como el caso de un número ilimitado de ellas (horizonte infinito), suponiendo entonces una probabilidad positiva de finalizar en cada una, lo cual da lugar al criterio del pago descontado. Demostró que en ambos casos dichos juegos poseen valor y que ambos jugadores tienen estrategias óptimas, que en el segundo caso son estacionarias, es decir que no dependen de la etapa en que se realice cada juego, sino sólo del juego que se lleve a cabo.

Posteriormente, siempre con el criterio del pago descontado, se han realizado extensiones del modelo de Shapley para incluir tanto el caso en que el número A de juegos no sea finito, como el caso en que cada juego tenga un número infinito de estrategias para cada jugador. Kushner y Chamberlain (1969) estudiaron el caso A finito, con conjuntos de estrategias compactos: Wessels (1977) el caso en que A es numerable pero los conjuntos de estrategias finitos; Groenewegen y Wessels (1976) el caso en que tanto A como los conjuntos de estrategias son numerables... El caso en que tanto el conjunto de juegos posibles como sus espacios de estrategias constituyan espacios métricos compactos, que nos proponemos analizar en este trabajo, ha sido ya considerado por Maitra y Parthasarathy (1970, 1971).

Rogers (1969) extiende el modelo de Shapley a juegos bipersonales de suma no nula y demuestra la existencia de un punto de equilibrio de estrategias estacionarias. Parthasarathy (1971) considera el modelo de Rogers con número de juegos posibles numerable.

El concepto de función de pago no descontada o criterio de pago promedio fue introducido por Gillette (1957). La mayoría de los trabajos que estudian esta clase de juegos estocásticos no descontados tratan de encontrar condiciones suficientes para la existencia de un valor y estrategias óptimas; ver, por ejemplo, Hoffman y Karp (1966), Kohlberg (1974), Stern (1975), Parthasarathy y Raghavan (1981), Bewley y Kohlberg (1978), Filar (1981) entre otros otros.

Los resultados más precisos y más generales fueron los de Bewley y Kohlberg (1976a, 1976b, 1978) que exponen una manera de relacionar los juegos estocásticos descontados con los no descontados. Ello permitió a Monash (1979) y más tarde a Mertens y Neyman (1981) demostrar que todos los juegos estocásticos finitos, no descontados, tienen un valor para cualquier criterio razonable de pago promedio con el que se valoren las estrategias de los jugadores en el juego con horizonte infinito.

2. DESCRIPCION DE LOS JUEGOS ESTOCASTICOS

Un juego estocástico queda caracterizado por:

1. Una familia $\Gamma(a)$, $a \in A$, de juegos bipersonales de suma nula, cuyos conjuntos de estrategias puras para cada uno de los jugadores, X_a e Y_a, pueden depender de a y con funciones de pago $M(a, \cdot, \cdot)$.

   Tanto el espacio de parámetros (o de estados) A, como los distintos conjuntos de estrategias, X_a e Y_a, se supone que son espacios medibles.

   Para cada $a \in A$ se supone que $M(a, \cdot, \cdot)$ es una función medible acotada por una constante M.
2. Para cada $a \in A$ y para cada $x \in X_a$, $y \in Y_a$, una probabilidad $P(x, y, a, \cdot)$ sobre A tal que $P(\cdot, \cdot, a, B)$ sea medible.
3. Un número N de etapas que se jugarán, y si $N < \infty$, un pago adicional $w(a)$ que recibirá el primer jugador tras la última partida.

   Impondremos en principio que w sea una función medible en A, acotada por una constante que, sin pérdida de generalidad, puede suponerse que coincide con M.

El juego estocástico Γ_N, determinado por los datos anteriores, consiste en que, para cada etapa $n = 1, 2, ..., N - 1$, los jugadores eligen estrategias $x \in X_a$ e $y \in Y_a$ para el juego $\Gamma(a)$ que se desarrolle en dicha etapa; el jugador J_1 recibe el pago $M(a, x, y)$ y a continuación se sortea, con probabilidad $P(x, y, a, \cdot)$ el juego $\Gamma(b)$ al que jugará en la etapa $n + 1$.

En la etapa N la situación es la misma salvo que se sortea con probabilidad $P(x, y, a, \cdot)$ el pago final $w(b)$ que recibirá el primer jugador.

2.1. Estrategias para un juego estocástico

Sean $\bar{X}_a$ e $\bar{Y}_a$ los conjuntos de estrategias mixtas para el juego $\Gamma(a)$, cuyos elementos serán representados típicamente por ξ y η respectivamente.

En cada etapa, las estrategias de los jugadores deben precisar las estrategias mixtas $\xi(a)$ y $\eta(a)$ que emplearían en el caso de que el juego a llevar a cabo fuese $\Gamma(a)$. En principio, dichas estrategias podrían depender del desarrollo del juego estocástico en todas las etapas anteriores, sin embargo, en general, bastará considerar estrategias markovianas en las que la estrategias mixta empleada por cada jugador sea sólo función del estado $a \in A$ que se presente.

Así, las estrategias de cada jugador para una etapa determinada serán elementos $\xi(\cdot) \in \Pi_{a \in A} \bar{X}_a$ y $\eta(\cdot) \in \Pi_{a \in A} \bar{Y}_a$ respectivamente. No obstante, para que la evolución y la función de pago del juego multietápico estén definidas, los jugadores habrán de utilizar en cada etapa estrategias dentro de los subconjuntos $U \subset \Pi_{a \in A} \bar{X}_a$ y $V \subset \Pi_{a \in A} \bar{Y}_a$, tales que

a) $P(\xi(a), \eta(a), a, B) = \int_{X_a} \int_{Y_a} P(x, y, a, B) \xi(a)(dx) \eta(a)(dy)$ sea una función de transición.

b) $M(a, \xi(a), \eta(a)) = \int_{X_a} \int_{Y_a} M(a, x, y) \xi(a)(dx) \eta(a)(dy)$ sea una función medible en A.

Para el juego estocástico Γ_N, que se desarrollará durante N etapas, los jugadores habrán de elegir estrategias $\{\xi_n(\cdot)\}_{n=1}^N$ y $\{\eta_n(\cdot)\}_{n=1}^N$, siendo pra cada $n = 1...N$, $\xi_n \in U$ y $\eta_n \in V$.

Aquellas estrategias del primer jugador tales que $\xi_n(\cdot) \equiv \xi(\cdot)$ para cada $n = 1...N$ se denominan estrategias estacionarias. Análoga definición sirve para el segundo jugador.

2.2. Función de pago de un juego estocástico

Consideremos la sucesión aleatoria $\{\alpha_n\}_{n=1}^{N}$ que indica los subíndices de los subjuegos que habrán de desarrollarse en las etapas $n = 1, 2, 3, ..., N$ en un juego estocástico Γ_N. La distribución de $\{\alpha_n\}_{n=1}^{N}$ queda determinada cuando se conoce el subjuego $\Gamma(a)$ que se jugará inicialmente y ambos jugadores han fijado las estrategias que emplearán: $\{\xi_n(\cdot)\}_{n=1}^{N} \in \mathrm{U}^N$ y $\{\eta_n(\cdot)\}_{n=1}^{N} \in \mathrm{V}^N$.

El pago asociado a las estrategias $\{\xi_n(\cdot)\}_{n=1}^{N}$ y $\{\eta_n(\cdot)\}_{n=1}^{N}$, en un juego multietápico Γ_N, iniciado por el subjuego $\Gamma(a)$, será

$$m_{\beta,N}(a, \{\xi_n(\cdot)\}_{n=1}^{N}, \{\eta_n(\cdot)\}_{n=1}^{N}) =$$

$$= E\left[\sum_{n=1}^{N} \beta^{n-1} M(\alpha_n, \xi_n(\alpha_n), \eta_n(\alpha_n)) + \beta^N w(\alpha_{N+1}) \middle| \alpha_1 = a\right]$$

En la expresión anterior, $\beta \in (0, 1]$ representa un factor de descuento que actúa entre los pagos recibidos en una etapa y los pagos de la etpa siguiente.

En el caso $N < \infty$, la presencia de un factor de descuento, que puede valer 1, supone una generalidad extra. Pero, la razón para su introducción es que, en el caso $N = \infty$ si $\beta = 1$, la serie

$$\sum_{n=1}^{\infty} E[M(\alpha_n, \xi_n(\alpha_n), \eta_n(\alpha_n))]$$

será en general divergente y no puede utilizarse como criterio de valoración de las estrategias.

En este caso habrá que adoptar un criterio más elaborado, que puede basarse en el estudio de las propiedades límites del modelo con $N = \infty$ y $\beta \uparrow 1$; o en las propiedades asintóticas del modelo con $\beta = 1$ y $N \uparrow \infty$. El análisis del caso $N = \infty$ y $\beta = 1$ habrá de posponerse, por tanto, al estudio de los modelos con $N < \infty$ *o* $\beta < 1$.

En el caso $N = \infty$, $\beta < 1$, simplificaremos la notación representando la función de pago por $m_\beta(a, \{\xi_n(\cdot)\}_{n=1}^{\infty}, \{\eta_n(\cdot)\}_{n=1}^{\infty})$.

La función de pago $m_{\beta,N}$ admite una expresión recurrente, más útil tanto para su obtención como para las consideraciones teóricas basadas en ella. De hecho es fácil establecer que

$$m_{\beta,N}(a, \{\xi_n(\cdot)\}_{n=1}^{N}, \{\eta_n(\cdot)\}_{n=1}^{N}) = M(a, \xi_1(a), \eta_1(a)) +$$

$$+ \beta \int_A m_{\beta,N-1}(b, \{\xi_{n+1}(\cdot)\}_{n=1}^{N-1}, \{\eta_{n+1}(\cdot)\}_{n=1}^{N-1}) P(\xi_1(a), \eta_1(a), a, db)$$

3. JUEGOS ESTOCASTICOS CONTINUOS

El tipo de juegos estocásticos que se pretende analizar son los juegos estocásticos continuos, caracterizados por las siguientes condiciones:

i) Para cada $a \in A$ es $X_a = X$ e $Y_a = Y$ y tanto A como X e Y son espacios métricos compactos.

ii) La función de pago $M(a, x, y)$ es continua en $A \times X \times Y$.

iii) Si $x_k \to x$, $y_k \to y$ y $a_k \to a$, entonces $P(x_k, y_k, a_k, \cdot) \xrightarrow{w} P(x, y, a, \cdot)$.

iv) El pago terminal $w(a)$ es una función continua en A.

Nótese que si X_a e Y_a son, por ejemplo, intervalos de la recta real, pueden transformarse con continuidad en un intervalo fijo; de manera que la restricción de que sean independientes de a no es sustancial. Por otra parte, si X e Y son finitos, puede considerarse en ellos la métrica discreta, cumpliéndose automáticamente las hipótesis anteriores; sin embargo, la restricción de que X_a e Y_a no dependan de a, sí es ahora relevante.

En el contexto de los juegos estocásticos continuos, se puede precisar los conjuntos de estrategias, U y V, permitidas para los jugadores en cada etapa. Se puede probar:

Teorema 1: *Para un juego estocástico continuo, los conjuntos U y V de estrategias permitidas para cada jugador en cada etapa contienen todas las funciones medibles $\xi(\cdot)$ y $\eta(\cdot)$ de A en $\bar{X}$ e $\bar{Y}$ respectivamente (considerando en $\bar{X}$ e $\bar{Y}$ la σ-álgebra de Borel correspondiente a la topología de la convergencia débil).*

Puede establecerse también otra propiedad de interés de los juegos estocásticos: *Para un juego estocástico continuo, si f es una función continua de A en $\mathbb{R}$*

$$\int_A f(b)P(\xi, \eta, a, db)$$

es una función continua en $A \times \bar{X} \times \bar{Y}$.

4. VALOR Y ESTRATEGIAS OPTIMAS DE JUEGOS ESTOCASTICOS CONTINUOS CON HORIZONTE FINITO

El objetivo que se persigue en primer lugar es el análisis de los juegos estocásticos continuos, en el caso en que el número N de etapas a jugar sea un número finito dado. En tal caso, el factor de descuento β no juega un papel primordial, de forma que podemos admitir la posibilidad de que sea $\beta = 1$.

Puesto que, en cada etapa, para fijar sus estrategias, los jugadores han de resolver una familia de juegos dependiente de un parámetro, debemos estudiar inicialmente tales familias de juegos. El siguiente resultado, que simplifica y completa los razonamientos de Maitra y Parthasarathy (1970 y 1971) cubre este objetivo.

Teorema 2: *Sea $\Gamma(a)$ una familia de juegos tales que sus conjuntos de estrategias puras, X e Y, sean métricos compactos y con función de pago $M(a, x, y)$ continua en $A \times X \times Y$. Entonces*

$$v(a) = \text{Val } [\Gamma(a)]$$

existe y es una función continua de A.

Además los jugadores disponen de estrategias óptimas $\xi^*(a)$ y $\eta^*(a)$, que constituyen funciones medibles de A en los espacios de estrategias mixtas $\bar{X}$ e $\bar{Y}$ respectivamente (dotados de la σ-álgebra de Borel asociada a la topología de la convergencia débil).

La demostración utiliza el teorema fundamental de los juegos continuos para garantizar la existencia de valor; su continuidad puede probarse usando la compacidad débil de los espacios de estrategias mixtas $\bar{X}$ e $\bar{Y}$. Además el teorema de selección de Dubins y Savage, garantiza la existencia de estrategias óptimas medibles para ambos jugadores.

Mediante el teorema anterior, podemos abordar la solución recurrente de los juegos estocásticos continuos. La idea, típica en programación dinámica, es resolver primero el juego estocástico Γ_1, correspondiente a una única etapa, aprovechar la solución resolver el juego Γ_2 y, en general, utilizar la solución del juego Γ_N para resolver el juego Γ_{N+1}. La introducción del pago terminal w, que omiten Maitra y Parthasarathy, permite realmente reducir cada uno de los pasos al primero de ellos.

Así para el caso de una única etapa, la solución viene dada en el teorema siguiente:

Teorema 3: *Para cualquier $\beta \leqslant 1$, el juego estocástico continuo Γ_1, de función de pago*

$$m_{\beta,1}(a, x, y) = M(a, x, y) + \beta \int_A w(b)P(x, y, a, db)$$

tiene valor, que será representado por $v_{\beta,1}(a)$ y que define una función continua de A en $\mathbb{R}$.

Además, los jugadores disponen de estrategias óptimas, $\xi^*_{\beta,1}(a)$ y $\eta^*_{\beta,1}(a)$, medibles como aplicaciones de A en $\bar{X}$ e $\bar{Y}$ respectivamente (en el sentido del teorema anterior).

La demostración se reduce a comprobar que las hipótesis ii), iii) y iv) aseguran la continuidad de la función de pago $m_{\beta,1}(a, x, y)$, de manera que se puede utilizar el teorema 2.

Aplicando recurrentemente el teorema anterior (con la sustitución de w por $v_{\beta,N-1}$) se obtiene:

Teorema 4: *Bajo las hipótesis i)-iv) y para cualquier $\beta \leqslant 1$, el juego de función de pago*

$$m_{\beta,N}(a, x, y) = M(a, x, y) + \beta \int_A v_{\beta,N-1}(b)P(x, y, a, db)$$

tiene valor, representado por $v_{\beta,N}(a)$, que constituye una función continua de a.

Además, ambos jugadores disponen de estrategias óptimas: $\xi^*_{\beta,N}(a)$ y $\eta^*_{\beta,N}(a)$, medibles en el sentido del teorema 1.

En el caso $\beta = 1$, simplificaremos la notación, omitiendo la variable β y representando el valor del juego definido en el teorema anterior por $v_N(a)$ y las estrategias óptimas de cada jugador por $\xi^*_N(a)$ y $\eta^*_N(a)$ respectivamente.

Naturalmente, una vez resuelto el juego correspondiente a cada etapa, en función del resultado del juego de la etapa anterior, hay que comprobar que el procedimiento proporciona, para cada N, la solución del juego estocástico Γ_N. En ese sentido se verifica:

Teorema 5: *Para cualquier $\beta \leqslant 1$, el juego estocástico continuo Γ_N, con factor de descuento β, iniciado a través del subjuego $\Gamma(a)$, tiene valor $v_{\beta,N}(a)$. Es decir*

$$\inf_{\{\eta_n(\cdot)\}_{n=1}^N \in V^N} \sup_{\{\xi_n(\cdot)\}_{n=1}^N \in U^N} m_{\beta,N}(a, \{\xi_n(\cdot)\}_{n=1}^N, \{\eta_n(\cdot)\}_{n=1}^N) = v_{\beta,N}(a)$$

y

$$\sup_{\{\xi_n(\cdot)\}_{n=1}^N \in U^N} \inf_{\{\eta_n(\cdot)\}_{n=1}^N \in V^N} m_{\beta,N}(a, \{\xi_n(\cdot)\}_{n=1}^N, \{\eta_n(\cdot)\}_{n=1}^N) = v_{\beta,N}(a)$$

Además, las estrategias $\{\xi^*_{\beta,N}(\cdot)\}_{n=N}^1$ y $\{\eta^*_{\beta,N}(\cdot)\}_{n=N}^1$ son óptimas para cada uno de los jugadores (donde $\{\xi^*_{\beta,N}(\cdot)\} \in U$ y $\{\eta^*_{\beta,N}(\cdot)\} \in V$ son las estrategias definidas en el teorema anterior).

La comprobación puede hacerse por inducción, gracias a la relación recurrente establecida en la sección 2.2, para la función de pago del juego Γ_N.

5. VALOR Y ESTRATEGIAS OPTIMAS DE JUEGOS ESTOCASTICOS CONTINUOS CON HORIZONTE INFINITO Y FACTOR DE DESCUENTO

La presencia de un factor de descuento, β, estrictamente menor que 1, que asegure la convergencia de la serie de los pagos, hace que los juegos estocásticos continuos con horizonte infinito tengan una solución simple y general.

Al igual que en el caso discreto, la observación básica es que en cada etapa el valor del juego es una contracción del valor en la etapa anterior. En el caso continuo, el hecho aparece explícitamente en el trabajo de (Maitra y Parthasarathy; 1970, Lema 2.5):

Lema: *Si w y w' son dos funciones continuas, y $v_{\beta,1}$ y $v'_{\beta,1}$ son los valores de los juegos estocásticos continuos Γ_1 y Γ'_1, con factor de descuento β y pago terminal w y w' respectivamente, se verifica*

$$\|v_{\beta,1} - v'_{\beta,1}\| \leqslant \beta \|w - w'\|$$

siendo $\|\cdot\|$ la norma del supremo en el espacio de las funciones continuas de A en $\mathbb{R}$.

El lema anterior permite analizar el comportamiento límite de la sucesión de valores $v_{\beta,N}(a)$ de los juegos estocásticos Γ_N, con factor de descenso $\beta < 1$, al crecer N hacia infinito.

Teorema 6: *Para un juego estocástico continuo, con factor de descuento $\beta < 1$, cuando N tiende a infinito, $v_{\beta,N}$ converge uniformemente hacia una función continua v_β, que es la única solución de la ecuación*

$$v_\beta(a) = \text{Val}\ \{M(a, x, y) + \beta \int_A v_\beta(b)P(x, y, a, db)\}$$

Demostración. Según el lema anterior

$$\|v_{\beta,N+1} - v_{\beta,N}\| \leqslant \beta\|v_{\beta,N} - v_{\beta,N-1}\|$$

y por consiguiente

$$\|v_{\beta,N+1} - v_{\beta,N}\| \leqslant \beta^N\|v_{\beta,1} - w\|.$$

Entonces, para cualquier $r \in N$

$$\|v_{\beta,N+r} - v_{\beta,N}\| \leqslant (\beta^{N+r-1} + \beta^{N+r-2} + \cdots + \beta^N)\|v_{\beta,1} - w\| \leqslant$$
$$\leqslant \frac{\beta^N}{1-\beta}\|v_{\beta,1} - w\|$$

y, como el último miembro converge a cero cuando N tiende a infinito, resulta que $v_{\beta,N}$ converge uniformemente a una función, que denominaremos v_β, y que será función continua en A.

Por otra parte, de acuerdo con el teorema 3, el juego de función de pago

$$M(a, x, y) + \beta \int_A v_\beta(b)P(x, y, a, db)$$

tiene valor, función continua en A. Si su valor fuese $\bar{v}_\beta(a)$, según el lema anterior se tendría $\|\bar{v}_\beta - v_{\beta,N+1}\| \leqslant \beta\|v_\beta - v_{\beta,N}\|$ cualquiera que fuese N. Luego, como $\|v_\beta - v_{\beta,N}\|$ se hace arbitrariamente pequeño al crecer N, $v_{\beta,N+1}$ converge a $\bar{v}_\beta$ y ha de ser, por consiguiente, $\bar{v}_\beta \equiv v_\beta$. Es decir, que se verifica

$$v_\beta(a) = \text{Val}\ \{M(a, x, y) + \beta \int_A v_\beta(b)P(x, y, a, db)\}$$

Por último, si la ecuación anterior tuviese otra solución, v'_β, en virtud nuevamente del lema anterior, sería $\|v'_\beta - v_\beta\| \leqslant \beta \|v'_\beta - v_\beta\|$; con lo cual $\|v'_\beta - v_\beta\| = 0$ y $v'_\beta \equiv v_\beta$.

Podemos reformular el resultado anterior, aprovechando las conclusiones del teorema 3, en la forma:

Corolario: *Dado un juego estocástico continuo, para cada $\beta < 1$, existe una única función, v_β, de A en $\mathbb{R}$, tal que el juego de función de pago*

$$M(a, x, y) + \beta \int_A v_\beta(b) P(x, y, a, db)$$

tiene valor $v_\beta(a)$. Además, ambos jugadores disponen en dicho juego de estrategias óptimas, que representaremos por $\xi^*_\beta(a)$ y $\eta^*_\beta(a)$, que constituyen aplicaciones medibles de A en $\bar{X}$ e $\bar{Y}$ respectivamente.

Se puede establecer entonces la solución general de los juegos continuos con factor de descuento, $\beta < 1$, y horizonte infinito; en el sentido de que basta resolver el juego formulado en el corolario anterior y los jugadores deben mantener fijas las estrategias óptimas de dicho juego:

Teorema 7: *Para cada $\beta < 1$, el juego estocástico continuo Γ_∞ con factor de descuento β, iniciado en el subjuego $\Gamma(a)$, tiene valor $v_\beta(a)$ (definido en el teorema 6). Además las estrategias estacionarias $\{\xi^*_\beta(\cdot)\}_{n=1}^\infty$ y $\{n^*_\beta(\cdot)\}_{n=1}^\infty$ son óptimas para cada uno de los jugadores. (Siendo $\xi^*_\beta(a)$ y $\eta^*_\beta(a)$ las estrategias determinadas en el corolario anterior).*

Demostración. La función de pago del juego estocástico Γ_∞ con factor de descuento β, iniciado a través del subjuego $\Gamma(a)$, es

$$m_\beta(a, \{\xi_n(\cdot)\}_{n=1}^\infty, \{\eta_n(\cdot)\}_{n=1}^\infty) = \sum_{n=1}^{\infty} \beta^{n-1} E[M(\alpha_n, \xi_n(\alpha_n), \eta_n(\alpha_n)]$$

Como la función de pago se supone acotada por una constante M, será

$$m_\beta(a, \{\xi_n(\cdot)\}_{n=1}^\infty, \{\eta_n(\cdot)\}_{n=1}^\infty) = \sum_{n=1}^{N} \beta^{n-1} E[M(\alpha_n, \xi_n(\alpha_n), \eta_n(\alpha_n))] - \frac{\beta^N}{1-\beta} M$$

El primer sumando es la función de pago $m_{\beta,N}(a, \{\xi_n(\cdot)\}_{n=1}^N, \{\eta_n(\cdot)\}_{n=1}^N)$ del juego estocástico Γ_N, salvo el término del pago terminal:

$\beta^N E[w(\alpha_{N+1})]$. Puesto que el pago terminal se supone también acotado por M, se obtiene

$$m_\beta(a, \{\xi_n(\cdot)\}_{n=1}^{\infty}, \{\eta_n(\cdot)\}_{n=1}^{\infty}) \geqslant m_{\beta,N}(a, \{\xi_n(\cdot)\}_{n=1}^{N}), \{\eta_n(\cdot)\}_{n=1}^{N}) - \beta^N \frac{2-\beta}{1-\beta} M$$

Entonces

$$\sup_{\{\xi_n(\cdot)\}_{n=1}^N \in U^N} \inf_{\{\eta_n(\cdot)\}_{n=1}^N \in V^N} m_{\beta,N}(a, \{\xi_n(\cdot)\}_{n=1}^{N}, \{\eta_n(\cdot)\}_{n=1}^{N}) \geqslant v_{\beta,N}(a) - \beta^N \frac{2-\beta}{1-\beta} M$$

De manera similar

$$\inf_{\{\eta_n(\cdot)\}_{n=1}^N \in V^N} \sup_{\{\xi_n(\cdot)\}_{n=1}^N \in U^N} m_{\beta,N}(a, \{\xi_n(\cdot)\}_{n=1}^{N}, \{\eta_n(\cdot)\}_{n=1}^{N}) \leqslant v_{\beta,N}(a) + \beta^N \frac{2-\beta}{1-\beta} M$$

Como ello es cierto para cualquier $N \in N$ y $v_{\beta,N}(a)$ converge a $v_\beta(a)$ cuando N tiende a infinito, resulta que $v_\beta(a)$ es el valor de Γ_∞ iniciado a través del subjuego $\Gamma(a)$.

Nótese que la conclusión es cierta para cualquier pago terminal, acotado. De hecho, la función de pago $m_\beta(a, \{\xi_n(\cdot)\}_{n=1}^{\infty}, \{\eta_n(\cdot)\}_{n=1}^{\infty})$ del juego Γ_∞ no depende del pago terminal; mientras que la determinación recursiva de $v_{\beta,N}$ y de las correspondientes estrategias óptimas, depende de la elección de w.

Supongamos entonces que $w \equiv v_\beta$, con lo cual, según el teorema 6, se tiene $v_{\beta,N} \equiv v_\beta$ para cada N y además $(\xi_\beta^*(\cdot)\}_{n=1}^{N}$es estrategia óptima para el primer jugador en el juego Γ_N (de acuerdo con el teorema 5). Entonces, para cualquier estrategia $\{\eta_n(\cdot)\}_{n=1}^{\infty} \in V^{\infty}$ del segundo jugador

$$\begin{aligned} m_\beta(a, \{\xi_\beta^*(\cdot)\}_{n=1}^{\infty}, \{\eta_n(\cdot)\}_{n=1}^{\infty}) &\geqslant \sum_{n=1}^{N} \beta^{n-1} E[M(\alpha_n, \xi_\beta^*(\alpha_n), \eta_n(\alpha_n))] - \frac{\beta^N}{1-\beta} M \\ &\geqslant m_{\beta,N}(a, \{\xi_\beta^*(\cdot)\}_{n=1}^{N}, \{\eta_n(\cdot)\}_{n=1}^{N} - \beta^N \frac{2-\beta}{1-\beta} M \end{aligned}$$

de forma que

$$\inf_{\{\eta_n(\cdot)\}_{n=1}^N \in V^N} m_\beta(a, \{\xi_\beta^*(\cdot)\}_{n=1}^{\infty}) \geqslant v_\beta(a) - \beta^N \frac{2-\beta}{1-\beta} M$$

y, como ello es cierto para cualquier N, haciendo N tender a infinito, se obtiene que $\{\xi_\beta^*(\cdot)\}_{n=1}^{\infty}$ es estrategia óptima para el primer jugador en el juego Γ_∞.

De manera análoga puede probarse que $\{\eta_\beta^*(\cdot)\}_{n=1}^{\infty}$ es la estrategia óptima para el segundo jugador.

El resultado anterior, y más en concreto, la existencia de estrategias óptimas estacionarias para el juego Γ_∞, constituye el teorema final de Maitra y Parthasarathy, que se obtiene en el estudio de los procesos de decisión markovianos que se generan cuando se fuerza a uno de los jugadores a emplear una estrategia estacionaria determinada. Es sorprendente que, contando con el lema anterior, no se haya empleado en la referencia citada, el recurso a hacer una elección conveniente del pago terminal de los juegos Γ_N, de horizonte finito, cuyo valor aproxima el valor de Γ_∞.

REFERENCIAS

BEWLEY, T., y KOHLBERG, E. (1976a): «The asymptotic theory of stochastic games», *Math. of O. R.*, vol. 1, 197-208.

BEWLEY, T., y KOHLBERG, E. (1976b): «The asymptotic solution of a recursive equation arising in stochastic games», *Math. of O. R.*, vol. 1, 321-336.

BEWLEY, T., y KOHLBERG, E. (1978): «On stochastic games with stationary optimal strategies», *Math. of* 0. *R.*, vol. 3, 104-125.

DUBINS, L. E., y SAVAGE, L. J. (1965): *How to gamble in you must*, McGraw-Hill Book Company, New York.

GILLETTE, D. (1957): «Stochastic games with zero stop probabilities», en: Dresher, M., A. W. Tucker y P. Wolfe (eds.), *Contributions to the theory of games*, vol. III, *Ann. of Math. Stud.*, 39, Princeton Univ. Press, Princeton.

GROENEWEGEN, L., y WESSELS, J. (1976): *On the relation between optimality and saddle-conservation in Markov games*, Eindhoven Univ. of Technology, Dep. of Math., Memorandum COSOR, 76-14.

HOFFMAN, A., y KARP, R. (1966): «On nonterminating stochastic games», *Man. Science*, vol. 12, 359-370.

KUSHNER, H. J., y CHAMBERLAIN, S. G. (1969): *Finite state stochastic games: existence theorems and computational procedures*, IEEE, Trans Automatic Control, AC-14, 248-255.

MAITRA, A., y PARTHASARATHY, T. (1970): «On stochastic games», *J.O.T.A.*, vol. 5, 289-300.

MAITRA, A., y PARTHASARATHY, T. (1970): «On stochastic games II», *J.O.T.A.*, vol. 8, 154-160.

MERTENS, J. F., NEYMAN, A. (1981): «Stochastic games», *Int. J. of Game Theory*, vol. 10, 53-56.

MONASH, C. A. (1979): *Stochastic games: The minmax theorem*, Thesis submitted to the dep. of Math. of the Harvard Univ., Cambridge, Massachusetts.

PARTHASARATHY, T. (1971): «Discounted and positive stochastic games», *Bull. Amer. Math. Soc.*, vol. 77, 134-136.

PARTHASARATHY, T., y RAGHAVAN, T. E. S. (1981): «An orderfield property for stochastic games when one player controls transition probabilities», *J.O.T.A.*, vol. 33, 375-392.

ROGERS, P. D. (1969): *Non-zero-sum stochastic games*, Ph. D. Dissertation, Report ORC 69-8, Operations Research Centre, Univ. of California, Berkeley.

SHAPLEY, L. S. (1953): «Stochastic games», *Proc. Nat. Acad. Sci. U.S.A.*, vol. 39, 1095-1100.

STERN, M. A. (1975): *On stochastic games with limiting average payoff*, Ph. D. Dissertation, Univ. of Illinois, Chicago.

WESSELS, J. (1977): «Markov games with unbounded rewards», en: M. Schäl, *Dynamische optimierung*, *Bonner Mathematische Schriften*, n.° 98.