Decimocuarta Semana Internacional de la Estadística y la Probabilidad
14-18 de junio de 2021

Procesos de decisión de Markov con recompensa trapezoidal difusa.

Karla Carrero-Vera.[a], Hugo Cruz-Suárez.[b], Raúl Montes-de-Oca.[c]
[a,b] *Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias Físico Matemáticas, Puebla, Puebla, México.*
[c] *Departamento de Matemáticas, Universidad Autónoma Metropolitana-Iztapalapa, Av. San Rafael Atlixco 186, Col. Vicentina, México D.F. 09340, México*
[a] karla.carrero@alumno.buap.mx, [b]hcs@fcfm.buap.mx, [c] momr@xanum.uam.mx.

Resumen

Este artículo se refiere a los procesos de decisión de Markov descontados con una función de recompensa difusa de forma trapezoidal. Partimos de un modelo de control de Markov habitual y no difuso (ver Hernandez-Lerma, 1989) con conjuntos de acciones compactos y recompensa $\boldsymbol{R}$, se induce un modelo de control solo sustituyendo $\boldsymbol{R}$ en el modelo habitual por una función difusa trapezoidal adecuada que involucra a $\boldsymbol{R}$. De esta manera, para este modelo inducido se considera un problema de control óptimo descontado, teniendo en cuenta tanto un horizonte finito como infinito, y funciones objetivo difusas. Para obtener la solución óptima, se utiliza el orden parcial en los cortes de números difusos, y la función de valor óptimo y la política óptima para los procesos de decisión de Markov difusos inducidos se relacionan con la función de valor óptimo y la política óptima correspondiente a los procesos de decisión de Markov habituales.

Palabra claves: Procesos de decisión de Markov, política óptima, números difusos trapezoidales.

1 Introducción

En diversas áreas aplicadas, como ingeniería, investigación de operaciones, economía, finanzas e inteligencia artificial, entre otras, los datos requeridos para proponer un modelo matemático presentan ambigüedad, vaguedad o características aproximadas del problema de interés (ver, por ejemplo, (Fakoor et al., 2016), (Efendi et al., 2018)). En este contexto, es posible encontrar en la literatura el enfoque de los números difusos para incorporar este tipo de características o afirmaciones a modelos matemáticos. La teoría básica del tema de los números difusos fue propuesta por L. Zadeh en su artículo fundamental escrito en 1965, que se titula Conjuntos difusos (Zadeh,1965). Posteriormente, en la literatura sobre el tema se pueden encontrar diversos artículos de investigación y textos referentes a la teoría difusa, además, es posible ubicar extensiones de la teoría en otros campos de las ciencias matemáticas, como la teoría de control, ver (Zadeh, 1965).

En este manuscrito, se proporciona un proceso de decisión de Markov (MDP) con un espacio de estado finito, conjuntos de acciones compactos y características difusas en su función de pago o recompensa. La idea es la siguiente: se considera un modelo de control de Markov nítido (MCM), es decir, un MCM del tipo que se ha analizado en (Hernandez-Lerma, 1989), con recompensa R como base y se induce

un nuevo MCM cambiando solo R por una función de recompensa con valores difusos $\hat{R}$. Específicamente, se asume que la función de recompensa difusa es trapezoidal. De esta forma, el problema de control difuso consiste en determinar una política de control que maximice la recompensa difusa descontada total esperada, donde la maximización se realiza con respecto al orden parcial en los cortes a de números difusos.

La metodología que se sigue en este artículo para garantizar la existencia de políticas óptimas en el problema difuso consiste en aplicar la existencia de políticas óptimas y la validez de la programación dinámica para el problema de control nítido, así como ciertas propiedades de los números trapezoidales difusos.

2 Teoría básica de números difusos

Definición 2.1 Sea $\boldsymbol{\Theta}$ un conjunto no vacío, entonces un conjunto difuso A en $\boldsymbol{\Theta}$ esta definido en términos de la función de membresía $\hat{\mu}$: $\boldsymbol{\Theta} \rightarrow$ [0,1]. En consecuencia, un conjunto difuso A puede ser expresado como un conjunto de pares ordenados: $\{(\mathrm{x}, \hat{\mu}(\mathrm{x})) : \mathrm{x} \in \boldsymbol{\Theta}\}$. Diremos que un conjunto difuso A en $\boldsymbol{\Theta}$ es normal si existe un $x \in \boldsymbol{\Theta}$ tal que $\hat{\mu}(x) = 1$.

Definición 2.2 Un número difuso en el conjunto $\mathbb{R}$, se define en términos de la función de pertenencia $\hat{\mu}$, que asigna a cada elemento de $\mathbb{R}$, un valor real del intervalo [0,1] y está dada por la siguiente forma

$$\hat{\mu}(x) = \begin{cases} 0 & si \quad x \leq a \\ l(x) & si \ a \leq x \leq b \\ 1 & si \ b \leq x \leq c \\ \mathrm{r(x)} & si \ c \leq x \leq d \\ 0 & si \ d \leq x \end{cases} \qquad (1)$$

donde a, b, c, d son números reales, $l(x)$ es una función no-decreciente y $r(x)$ es una función no-creciente.

Definición 2.3 Un número difuso trapezoidal, es un conjunto difuso definido en los números reales caracterizado por la función de pertenencia

$$\hat{\mu}(x) = \begin{cases} 0 & si \quad x \leq a \\ \dfrac{\mathrm{x} - a}{\mathrm{b} - a} & si \ a \leq x \leq b \\ 1 & si \ b \leq x \leq c \\ \dfrac{\mathrm{d} - \mathrm{x}}{\mathrm{d} - \mathrm{c}} & si \ c \leq x \leq d \\ 0 & si \ d \leq x \end{cases} \qquad (2)$$

donde a, b, c, d son números reales que $0 \leq a \leq b \leq c \leq d$.

Definición 2.4 Un número difuso trapezoidal se denota por $\hat{\mu} = (a, b, c, d)$ y su α-corte denotado por $\hat{\mu}_{\alpha}$, se define como el conjunto $\{x \in \Theta: \hat{\mu}(x) \geq \alpha\}$.

Observación 2.5 Para un número difuso trapezoidal (a, b, c, d), su α-corte es el intervalo cerrado

$$(a, b, c, d)_{\alpha} = [(b - a)\alpha + a, \ d - (d - c)\alpha]$$

Definición 2.6 Denotemos por • a cualquiera de las cuatro operaciones aritméticas básicas y sean A y B dos números difusos. Entonces es definido el conjunto difuso en $\mathbb{R}$, $A \bullet B$, por la expresión

$$\hat{\mu}_{A \bullet B}(x) \sup_{x = y \bullet z} \min\{\hat{\mu}_A(y), \hat{\mu}_B(z)\}$$

Lema 2.7 Si $A = (a_1, a_2, a_3, a_4)$ y $B = (b_1, b_2, b_3, b_4)$ dos números difusos trapezoidales, entonces, de la definición 2.6 se tiene que:

$$A + B = (a_1 + b_1, a_2 + b_2, a_3 + b_3, a_4 + b_4)$$

Sea I el conjunto de todos los intervalos acotados cerrados $A = [a_l, a_{\mathrm{u}}]$ en la línea real $\mathbb{R}$. Para $A, B \in I$ se define

$$d(A, B) = \text{máx}[a_l - b_l, a_u - b_u] \qquad (3)$$

Es posible comprobar que (I, d) es un espacio métrico completo.

Además, para A, B en I definamos: $A \leq B$ si y solo si $a_l \leq b_l$ y $a_u \leq b_u$ donde $A = [a_l, a_u]$ y $B = [a_l, a_u]$, se tiene que $\leq$ es un orden parcial en I.

Sea $\mathbf{F}(\mathbb{R})$ el conjunto de todos los números difusos con función de membresía semi-continuas superiormente, convexas, normales y tienen soporte compacto. Definamos la función real-valuada $\rho: \mathbf{F}(\mathbb{R}) \times \mathbf{F}(\mathbb{R}) \to \mathbb{R}$ por

$$\rho(\hat{\mu}, \hat{v}) = \sup_{\alpha \in [0,1]} d(\mu_\alpha, v_\alpha) \qquad (4)$$

con $\hat{\mu}, \hat{v} \in \mathbf{F}(\mathbb{R})$

Es sencillo ver que ρ es una métrica en $\mathbf{F}(\mathbb{R})$. Además, para $\hat{\mu}, \hat{v} \in \mathbf{F}(\mathbb{R})$ definimos

$$\hat{\mu} \leq \hat{v} \text{ si y solo si } \hat{\mu}_\alpha \leq \hat{v}_\alpha \qquad (5)$$

con $\alpha \in [0,1]$.

Lema 2.8 El espacio métrico $(\mathbf{F}(\mathbb{R}), \rho)$ es completo.

Definición 2.9 Se dice que una secuencia $\{\hat{X}_n\}$ de números difusos es convergente al número difuso $\hat{X}$, si para cada $\epsilon > 0$ existe un entero positivo N tal que $\rho(\hat{X}_n, \hat{X}) < \epsilon$ para n > N.

Lema 2.10 Para un número difuso trapezoidal la siguiente declaración se mantiene:

a) Si $\{(a_n, b_n, c_n, d_n): 1 \leq n \leq N\}$ donde N es un entero positivo

$$\sum_{n=1}^{N} (a_n, b_n, c_n, d_n) = (\sum_{n=1}^{N} a_n, \sum_{n=1}^{N} b_n, \sum_{n=1}^{N} c_n, \sum_{n=1}^{N} d_n)$$

b) Si $\hat{\mu}_n = \{(a_n, b_n, c_n, d_n): 1 \leq n\}$ y $\sum_{n=1}^{\infty} a_n$, $\sum_{n=1}^{\infty} b_n$, $\sum_{n=1}^{\infty} c_n$, $\sum_{n=1}^{\infty} d_n < \infty$, entonces $S_n := \sum_{m=1}^{n} \hat{\mu}_m$, $n \leq 1$ converge al número difuso trapezoidal: $\left(\sum_{n=1}^{\infty} a_n, \sum_{n=1}^{\infty} b_n, \sum_{n=1}^{\infty} c_n, \sum_{n=1}^{\infty} d_n\right)$.

Definición 2.11 Sea $(\Omega, \mathcal{F})$ un espacio medible y $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ el espacio medible de los números reales. Una variable aleatoria difusa es una función $\hat{X}: \Omega \to \mathbf{F}(\mathbb{R})$ tal que para todo $(\alpha, B) \in [0,1] \times \mathcal{B}(\mathbb{R})$, $\{\omega \in \Omega: \hat{X}(\omega)_\alpha \cap B \neq \emptyset\} \in \mathcal{F}$. De manera equivalente, $\hat{X}$ debe verse como un intervalo generalizado con función de pertenencia μ y α-corte:

$$\hat{X}: (\omega)_\alpha = [\hat{X}^-(\omega), \hat{X}^+(\omega)].$$

Definición 2.12 Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y sea $\hat{X}$ una variable aleatoria difusa discreta con rango $\{s_1, s_2, \ldots, s_n\} \subseteq \mathbf{F}(\mathbb{R})$. La esperanza matemática de es un número difuso, $E(\hat{X})$, tal que

$$E(\hat{X}) = \sum_{k=0}^{\infty} s_i P(\hat{X} = s_i) \qquad (6)$$

Sea $\hat{X}$ y $\hat{Y}$ variables aleatorias difusas discretas con rango finito. Entonces

a) $E(\hat{X}) \in \mathbf{F}(\mathbb{R})$
b) $E[\hat{X} + \hat{Y}] = E[\hat{X}] + E[\hat{Y}]$
c) $E[\lambda \hat{X}] = \lambda E[\hat{X}]$

3 Problema de control de Markov óptimo con descuento con funciones de recompensa difusas.

Considere un modelo de decisión de Markov

$$(X, A, \{A(x): x \in X\}, Q, \hat{R}) \qquad (7)$$

donde

a) X es un conjunto finito, llamado espacio de estado.

b) A es un espacio de Borel, denominado control o espacio de acción.

c) $\{A(x): x \in X\}$ es una familia de subconjuntos no vacíos $A(x)$ de A, cuyos elementos son las acciones factibles.

d) Q es la ley de transición, que es un núcleo estocástico en X dado $K := \{(x, a): x \in$

$X, a \in A(x)\}$, este conjunto se denomina el conjunto de pares de estados-acciones factibles.

e) $\hat{R}$ es una función de recompensa difusa en K en un solo paso.

La evolución del sistema estocástico difuso es la siguiente: si el sistema está en el estado $x_t = x$ en el momento t y se aplica el control $a_t = a \in A(x)$, entonces dos cosas suceden:

a) Una recompensa difusa $\hat{R}(x, a)$ es obtenida.

b) El sistema salta al próximo estado x_{t+1} de acuerdo con la ley de transición Q, i.e.

$Q(x \in B|x, a)$=Prob($|x \in B|x_t = x, a_t = a$)

con $B \subseteq X$.

Para un modelo no difuso $(X, A, \{A(x): x \in X\}, Q, R)$, la recompensa en un solo paso es una función $R: K \to \mathbb{R}$. Una política es una secuencia $\pi = \{\pi_t: t = 0,1, \dots\}$ de kérneles estocásticos π_t en el conjunto de control A dado el historial $\mathbb{H}_t$ del proceso hasta el momento t, donde $\mathbb{H}_t \coloneqq K \times X$ y $\mathbb{H}_0 \coloneqq X$. El conjunto de todas las políticas se indicará con Π. $\mathbb{F}$ denota el conjunto de funciones $f: X \to A$ tales que $f(x) \in A(x)$, para todo $x \in X$. Una política determinista de Markov es una secuencia $\pi = f_t$ tal que $f_t \in \mathbb{F}$, $t = 0,1, \dots$. Se dice que una política de Markov $\pi = f_t$ es estacionaria si f_t es independiente de t, es decir, $f_t = f$, para todo $t = 0,1, \dots$. En este caso, π se denota por f y $\mathbb{F}$ se denomina el conjunto de políticas estacionarias.

Sea $(\Omega, \mathcal{F})$ el espacio medible que consta del espacio canónico $\Omega = \mathbb{H}_\infty \coloneqq (X \times A)^\infty$ y $\mathcal{F}$ la correspondiente σ –algebra producto. Los elementos de Ω son secuencias de la forma $\omega = (x_0, a_0, x_1, a_1, \dots)$ con $x_t \in X$ y $a_t \in A$ para todo $t = 0,1, \dots$. Las proyecciones x_t y a_t desde Ω a los conjuntos X y A se denominan variables de estado y acción, respectivamente.

Sea $\pi = \pi_t$ una política arbitraria y μ una medida de probabilidad arbitraria en X llamada distribución inicial. Entonces, según el teorema de C. Ionescu-Tulcea, hay una medida de probabilidad única P_μ^π en (Ω, F) que es compatible con $\mathbb{H}_\infty$, es decir, $P_\mu^\pi(\mathbb{H}_\infty) = 1$.

El proceso estocástico $(\Omega, \mathcal{F}, P_\mu^\pi, x_t)$ se denomina Proceso de control de Markov o proceso de decisión de Markov. El operador de esperanza con respecto a P_μ^π se denota por E_μ^π. Si μ se concentra en el estado inicial $x \in X$, entonces P_μ^π y E_μ^π se escriben como P_x^π y E_x^π respectivamente.

La ley de transición de un proceso de control de Markov a menudo se especifica por una ecuación en diferencias de la forma $x_{t+1} = F(x_t, a_t, \xi_t)$, $t = 0,1, \dots$, con $x_0 = x \in X$. conocido, donde ξ_t es una secuencia de variables aleatorias independientes e idénticamente distribuidas (iid) con valores en un espacio de Borel S y una distribución común Δ, independiente del estado inicial. En este caso, la ley de transición Q viene dada por $Q(x, a) = EI_B[F(x, a, \xi)], B \subset X$, $(x, a) \in K$. E es la esperanza con respecto a la distribución Δ, I_B denota la función indicadora del conjunto B.

Definición 3.1 Sea $\left(X, A, \{A(x): x \in X\}, Q, \hat{R}\right)$ un modelo de Markov con recompensa difusa, para una política π y cada estado $x \in X$, se define el costo total descontado esperado con recompensa difusa de la siguiente manera

$$\hat{v}(\pi, x) \coloneqq \sum_{t=0}^{\infty} \alpha^t \hat{E}_x^\pi[\hat{R}(x, a)] \qquad (8)$$

Además, se define la recompensa difusa en la etapa N de la siguiente manera

$$\hat{v}_N(\pi, x) \coloneqq \sum_{t=0}^{N-1} \alpha^t \hat{E}_x^\pi[\hat{R}(x, a)] \qquad (9)$$

Definición 3.2 La función de valor óptimo se define como

$$\hat{v}(x) \coloneqq \sup_{\pi \in \Pi} \hat{v}(\pi, x), \qquad (10)$$

$x \in X$. Entonces el problema de control óptimo es encontrar una política π^* tal que

$$\hat{v}(\pi^*, x) = \hat{v}(x). \quad (11)$$

Para el modelo no difuso $(X, A, \{A(x): x \in X\}, Q, R)$, el costo total descontado esperado con recompensa y la recompensa en la etapa N (con recompensa no-difusa) están definidos de la siguiente manera

$$V(\pi, x) := E_x^{\pi}[\sum_{t=0}^{\infty} \alpha^t R(x, a)] \quad (13)$$

$$V_N(\pi, x) := E_x^{\pi}[\sum_{t=0}^{N-1} \alpha^t R(x, a)] \quad (14)$$

El problema de control de interés es la maximización de la recompensa difusa descontada total esperada del horizonte finito/infinito (ver (8) y (9)). En la siguiente sección demostrará que (9) converge a la función objetivo (8) con respecto a la métrica ρ (ver (4)). Se considera el siguiente supuesto para la función de recompensa del modelo difuso.

Supuesto 3.3 Sea B, C, D y E números reales, tales que $B < C < D < E$. Asumimos que la recompensa difusa es un número trapezoidal (ver definición 2.3), específicamente

$$\hat{R}(x, a) = (BR(x, a), CR(x, a), DR(x, a), ER(x, a)) \quad (12)$$

para cada $(x, a) \in K$ donde $R: K \to \mathbb{R}$ es la función recompensa del modelo no difuso. observemos que, bajo el Supuesto 3.3 y el Lema 2.10, la recompensa difusa en N-pasos es un número difuso trapezoidal

El siguiente supuesto garantiza la existencia de una política optima y proporciona la función de valor para el caso no difuso.

Supuesto 3.4 a) Para todo $x \in X$, $A(x)$ es un conjunto compacto en $\mathscr{B}(A)$, donde $\mathscr{B}(A)$ es la σ-algebra de Borel del espacio A.

b) La función recompensa no difusa R es una función no-negativa y acotada.

c) Para cada $x, y \in X$, los mapeos $a \to \mathrm{R}(\mathrm{x}, \mathrm{a})$ y $a \to \mathrm{Q}(\mathrm{y}|\mathrm{x}, \mathrm{a})$ son continuos en $A(x)$.

Teorema 3.5 [Programación dinámica]: Bajo el supuesto 3.2, la siguiente afirmación se mantiene:

a) Definamos $W_N(x) = 0$ y para cada $t = N-1, N-2, \ldots, 1, 0$, consideremos

$$W_N(x) = \max_{a \in A(x)} \{R(x, a)\, \alpha E[W_{n+1}(F(x, a, \xi))]\} \quad (15)$$

$x \in X$. Entonces para cada $t = 0,1,2 \ldots$ existe $f_t \in \mathrm{F}$ tal que

$$W_N(x) = \{R(x, f_t(x)) + \alpha E[W_{n+1}(F(x, f_t(x), \xi))]\} \quad (16)$$

y $\pi^* = \{f_0, f_1, \ldots, f_{N-1}\}$ es una política óptima markoviana y $v_N(\pi^*, x) = W_0, x \in X$.

b) La función de valor óptimo V, satisface la siguiente ecuación de programación dinámica:

$$\mathrm{V}(x) = \max_{a \in A(x)} \{R(x, a) + \alpha E[\mathrm{V}(F(x, a, \xi))]\} \quad (17)$$

c) Existe una política $f^* \in \mathrm{F}$ tal que el control $\mathrm{f}^*(x) \in A(x)$ y

$$\mathrm{V}(x) = \{R(x, f^*(x)) + \alpha E[\mathrm{V}(F(x, f^*(x), \xi))]\} \quad (18)$$

d) Definamos las funciones de iteración de valor como sigue:

$$V_t(x) = \max_{a \in A(x)} \{R(x, f_t^*(x)) + \alpha E[\mathrm{V}_{t-1}(F(x, f^*(x), \xi)])\} \quad (19)$$

para toda $x \in X$ y $n = 1,2, \ldots$, con $V_0(.) = 0$. Entonces las funciones de iteración de valor convergen puntualmente a la función de valor óptimo V,i.e.

$$\lim_{\mathrm{x} \to \infty} V_n(x) = V(x), \quad x \in X.$$

4 Resultados para recompensa difusa

Los siguientes resultados están referidos a la convergencia de la recompensa difusa de la etapa N a la recompensa difusa descontada total esperada en el horizonte infinito. Se verificará la existencia de políticas óptimas y vigencia de la programación dinámica.

Lema 4.1 Para cada $\pi \in \Pi$ y $x \in X$ fijos,

$$\lim_{\mathrm{N}\to\infty} \rho(\hat{v}_N, \hat{v}) = 0$$

donde ρ es la métrica de Hausdorff (ver (4)).

Demostración: Sean $\pi \in \Pi$ y $x \in X$ fijos. Para simplificar la notación denotaremos por $\hat{v}$ y $\hat{v}_N$ a $\hat{v}(\pi, \mathrm{x})$ y a $\hat{v}_N\ (\pi, \mathrm{x})$ respectivamente. Entonces, de acuerdo a (17) el α-corte de la función de recompensa difusa, está dado por

$$\begin{aligned}\Delta_{\mathrm{N}} &:= \ (B\mathrm{V}_{\mathrm{N}}, C\mathrm{V}_{\mathrm{N}}, D\mathrm{V}_{\mathrm{N}}, E\mathrm{V}_{\mathrm{N}})_{\alpha} \\ &=[\, B(1-\alpha)\mathrm{V}_{\mathrm{N}} + \alpha C\mathrm{V}_{\mathrm{N}}, E(1-\alpha)\mathrm{V}_{\mathrm{N}} + \alpha D\mathrm{V}_{\mathrm{N}}]\end{aligned}$$

Análogamente, el α-corte de (8) está dado por

$$\begin{aligned}\Delta &:= \ (BV, CV, DV, EV)_{\alpha} \\ &=[\, B(1-\alpha)V + \alpha CV,\ E(1-\alpha)V + \alpha DV]\end{aligned}$$

Por lo tanto, por (4), se obtiene que

$$\rho(\Delta_{\mathrm{N}}, \Delta) = \sup_{\alpha\in[0,1]} d(\Delta_{\mathrm{N}}^{\alpha}, \Delta^{\alpha})$$

Ahora, debido a la identidad:

$$\max(c, b) = (c + b + |b - c|)/2$$

con $b, c \in \mathbb{R}$, se produce que

$$d(\Delta_{\mathrm{N}}, \Delta) = (1-\alpha)E(V - V_N) + \alpha D(V - V_N)$$

Entonces,

$$\rho(\Delta_{\mathrm{N}}, \Delta) = \sup_{\alpha\in[0,1]} (V - V_N)(E - \alpha(E - D))$$

$$= (V - V_N)E \qquad (18)$$

Por lo tanto, cuando N tiende a infinito en (9), se concluye que

$$\lim_{\mathrm{N}\to\infty} \rho(\hat{v}_N, \hat{v}) = \lim_{\mathrm{N}\to\infty} (V - V_N)E = 0.$$

La segunda igualdad es una consecuencia de (13) y (14).

El problema de control óptimo difuso consiste en determinar una política π^* tal que;

$$\hat{v}(\pi, x) \le \hat{v}(\pi^*, x),$$

para todo $\pi \in \Pi$, y $x \in X$. En consecuencia,

$$\hat{v}(\pi^*, x) = \sup_{\pi\in\Pi} \hat{v}(\pi, x)$$

para todo $x \in X$. En este caso, se define la función de valor óptimo difusa como

$$\hat{v}(x) = \hat{v}(\pi^*, x),$$

$x \in X$ y π^* es llamada la política óptima del problema de control óptimo difuso. Se pueden establecer definiciones similares de manera análoga para V_N, intercambiando V por V_N.

Observación 4.2: Una consecuencia directa de la definición anterior y la aplicación del Teorema 3.3 y los Supuestos 3.4 y 3.5 es el siguiente resultado.

Teorema 4.3: Bajo los Supuestos 3.4 y 3.5 se cumplen las siguientes afirmaciones:

a) La política óptima π^* del problema de control óptimo finito no-difuso (ver (14)) es la política óptima para $\hat{v}_N$,, es decir, $\hat{v}_N(\pi, x) \le \hat{v}_N(\pi^*, x)$ para toda $\pi \in \Pi$ y $x \in X$.

b) La función de valor difuso óptimo finito está dada por

$$\hat{v}_N(x) = \big(BV_N(x), CV_N(x), DV_N(x), EV_N(x)\big), \qquad (19)$$

donde $V_N(x) = \sup_{\pi \in \Pi} V_N(\pi, x)$, $x \in X$.

Teorema 4.4: Bajo los Supuestos 3.4 y 3.5 se cumplen las siguientes afirmaciones:

a) La política óptima del problema de control difuso es la misma que la política óptima del problema de control óptimo.
b) La función de valor difuso óptimo está dada por

$$\hat{v}(\mathrm{x}) = (\mathrm{BV}(\mathrm{x}), \mathrm{CV}(\mathrm{x}), \mathrm{DV}(\mathrm{x}), \mathrm{DV}(\mathrm{x})), \quad x \in X. \tag{20}$$

Demostración:

a) Sean $\pi \in \Pi$, y $x \in X$, fijos, Primero, observemos que (8) es equivalente a

$$\hat{v}(\pi, x) = (\mathrm{B}V(\pi, \mathrm{x}), \mathrm{C}V(\pi, \mathrm{x}), \mathrm{D}V(\pi, \mathrm{x}), \mathrm{E}V(\pi, \mathrm{x})),$$

Como una consecuencia de supuesto (3.4). Entonces el α-corte de $\hat{v}(\pi, x)$ está dado por

$$\hat{v}(\pi, x)_\alpha = [BV(\pi, \mathrm{x}) + \alpha(C - B)V(\pi, \mathrm{x}), EV(\pi, \mathrm{x}) + \alpha(D - E)V(\pi, \mathrm{x})]$$

Ahora, por teorema 3.5, existe $f^* \in \mathbb{F}$ tal que

$$BV(\pi, \mathrm{x}) + \alpha(C - B)V(\pi, \mathrm{x}) \leq BV(f^*(x), \mathrm{x}) + \alpha(C - B)V(f^*(x), \mathrm{x})$$

y

$$EV(\pi, \mathrm{x}) + \alpha(D - E)V(\pi, \mathrm{x}) \leq EV(f^*(x), \mathrm{x}) + \alpha(D - E)V(f^*(x), \mathrm{x}).$$

y como $x \in X$ y $\pi \in \Pi$ son arbitrarios, resulta lo siguiente.

b) Por teorema 4.4 a)

$$\hat{v}(x) = \left(BV(f^*(x), \mathrm{x}), CV(f^*(x), \mathrm{x}), DV(f^*(x), \mathrm{x}), EV(f^*(x), \mathrm{x})\right)$$

para cada $x \in X$, de esta manera, aplicando (3.5) se concluye que

$$\hat{v}(x) = (BV(x), CV(x), DV(x), EV(x)), \quad x \in X.$$

5 Conclusiones

En este artículo, los procesos de decisión de Markov fueron estudiado bajo el criterio de recompensa con descuento total esperado y considerando una función de recompensa difusa, específicamente del tipo trapezoidal. Este proceso fue inducido a partir de un proceso nítido, teniendo en cuenta algunas de sus propiedades para inducir ciertas propiedades en el caso difuso. Trabajo futuro en la dirección de este trabajo consiste en aplicar la metodología a otros criterios de optimalidad como el caso promedio o los criterios sensibles al riesgo.

Referencias

Aliprantis, C. D. and Border, K. (2006). Infinite dimensional analysis. Springer.

Ban, A. I. (2009). Trapezoidal and parametric approximations of fuzzy numbers—inadvertences and corrections. Fuzzy Sets and Systems, 160(21):3048–3058.

Driankov, D., Hellendoorn, H., and Reinfrank, M. (2013). An introduction to fuzzy control. Springer Science & Business Media.

Efendi, R., Arbaiy, N., and Deris, M. M. (2018). A new procedure in stock market forecasting based on fuzzy random auto-regression time series model. Information Sciences, 441:113–132.

Fakoor, M., Kosari, A., and Jafarzadeh, M. (2016). Humanoid robot path planning with fuzzy Markov decision processes. Journal of applied research and technology, 14(5):300–310.

Hernandez-Lerma, O. (1989). ´ Adaptive Markov control processes, volume 79. Springer Science & Business Media.

Kurano, M., Yasuda, M., Nakagami, J.-i., and Yoshida, Y. (2003). Markov decision processes with fuzzy rewards. Journal of Nonlinear and Convex Analysis, 4(1):105–116.

Pedrycz, W. (1994). Why triangular membership functions? Fuzzy sets and Systems, 64(1):21–30.

Puri, M. L., Ralescu, D. A., and Zadeh, L. (1993). Fuzzy random variables. In Readings in fuzzy sets for intelligent systems, pages 265–271. Elsevier.

Puterman, M. L. (1994). Markov decision processes: discrete stochastic dynamic programming. John Wiley & Sons.

Semmouri, A., Jourhmane, M., and Belhallaj, Z. (2020). Discounted Markov decision processes with fuzzy costs. Annals of Operations Research, pages 1–18.

Topkis, D. M. (1998). Supermodularity and complementarity. Princeton university press.

Webb, J. N. (2007). Game theory: decisions, interaction and Evolution. Springer Science & Business Media.

Zadeh, L. A. (1965). Fuzzy sets. Information and control, 8(3):338–353.

Zeng, W. and Li, H. (2007). Weighted triangular approximation of fuzzy numbers. International Journal of Approximate Reasoning, 46(1):137–150