Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Posgrado en Ciencias Matemáticas

Procesos de decisión de Markov difusos: Caso total y caso descontado

Tesis que se presenta como requisito final para obtener el título de

Doctorado en Ciencias (Matemáticas)

Presenta: Karla Andreina Carrero Vera

Director de tesis: Dr. Hugo Adán Cruz Suárez
Asesor de tesis: Dr. Raúl Montes de Oca Machorro

Puebla, Puebla, JULIO 2023

Dedicatoria

Con todo mi amor, a mi hijo, el dueño de corazón, quién me ha acompañado desde que entré al doctorado estando en mi pancita en el comienzo, quién me agota la energía pero que me da más fuerzas para terminar y salir adelante juntos.

Isaac

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por patrocinar este proyecto, sin su apoyo habría sido posible la realizar este trabajo de esta tesis.

A mi director de tesis Dr. Hugo Adán Cruz Suárez por su inmensa paciencia conmigo, por la confianza depositada, por el tiempo dedicado para contribuír a la elaboración de este trabajo y constantes revisiones y por todo el apoyo brindado desde un principio.

Al Dr. Raúl Montes De Oca por su grán colaboración en la elaboración de este trabajo.

A mis sinodales: Dra. Hortensia Josefina Reyes Cervantes, Dr. Fernando Velasco Luna, Dr. Bulmaro Juárez Hernández y al Dr. Victor Hugo Vázquez Guevara por contribuir en las correcciones y su deseo de ayudar siempre.

A mi abuelita Rosa por todo su amor desde siempre.

A mis compañeros Rubén, Jaicer, Alberto, Camilo, Roberto (el beto) y a Germán por su apoyo y compañia.

Índice general

Introducción

En este trabajo se trata la modelación de programas matemáticos acoplando el ambiente estocástico con el difuso. Específicamente estudiamos los procesos de decisión de Markov (PDMs) estacionarios considerando como criterios de rendimiento a la recompensa total esperada y el caso descontado total, ambos en tiempo discreto con espacio de estado finito y conjunto de acciones finito y compacto. Las funciones de recompensas se plantearon en una versión difusa [6], con una forma conveniente de tipo trapezoidal en función de una recompensa estándar nítida.

La razón por la cual recurrimos al proceso matemático de difuminar a las recompensas, los cuales son elementos de un conjunto de referencia, se debe a que resolver el problema en versión nítida implica conocer los valores de los coeficientes en la función objetivo, pero en muchas ocasiones no conocemos esta o puede que esta información sea imprecisa o incierta, lo que hace que el sistema sea mucho más complejo y por lo tanto, mas difícil de resolver, o que simplemente no se pueda resolver. Pero al considerar los valores de las recompensas como valores de un conjunto difuso trapezoidal, ya no estamos restringiendo dichos valores a ser específicos, sino que les estamos permitiendo encontrarse dentro de un rango de valores, de esta manera estaríamos modelando el desconocimiento de información precisa. Así que ahora la modelación y la metodología de solución desarrollados para PDMs en este trabajo, consideran incertidumbre en las recompensas.

Plantear a las recompensas como funciones difusas trapezoidales nos lleva a la necesidad de utilizar herramientas de la teoría de los conjuntos difusos propuesta por L. Zadeh en su artículo: [33], la cual surgió justamente de la necesidad de una nueva forma de representar la imprecisión y la incertidumbre, y así solucionar problemas complejos con información de este tipo, en la que la lógica tradicional no es suficiente. Esta teoría está bien establecida y ha sido extendida a varios campos de las ciencias matemáticas, como la teoría de control ([5] y [10]), y también ha sido de alto impacto en áreas aplicadas (ver por ejemplo ([15] y [21]). Además en el control de sistemas, principalmente de tráfico, trenes, metros, mecatrónica, lavadoras, aires a condicionado, ascenso-

res, robótica y en muchos otros sistemas ([12] y [13]). Estas son solo algunas de las tantas situaciones donde se puede aplicar la teoría de números difusos.

Ahora bien, se usarán las herramientas de la teoría de números difusos lo cuál nos permitirá representar a la función objetivo como un número trapezoidal difuso, pero el problema a resolver no deja de ser un problema de maximización: el de encontrar una política que maximice a la función objetivo ahora difusa, por lo que es necesario una relación de orden, en el sentido difuso, que nos permita decidir si un valor difuso de la función objetivo es mayor o menor que otro, cuando esta es evaluada en diferentes políticas y, de esta manera, comparar políticas y encontrar las óptimas. Dicha maximización se estableció con respecto al orden parcial en los $\alpha-$cortes de números difusos (ver [14]).

El resultado obtenido de las operaciones con números difusos trapezoidales también será un número difuso trapezoidal, por lo que el resultado no será radicalmente un valor óptimo específico, sino que se encontrará en un rango de posibilidades.

La motivación de este trabajo surgió del hecho de que nuestro lenguaje es impreciso, a diario usamos expresiones con rangos como; angosto, no tan angosto, más o menos grueso, grueso y muy grueso, o cuando decimos poco, mucho o bastante, estamos usando palabras que contienen ambiguedad e imprecisión, estos conceptos no tienen límites perfectamente definidos, de esto podemos observar que razonamos de forma difusa, esta es la razón por la cual los conjuntos con esta naturaleza se presentan con mucha frecuencia en el mundo real, así que en muchos problemas matemáticos, los datos son imprecisos y es muy complicado operarar con ellos, así que el proceso de la modelación y de resolución es más complejo. Dado que la teoría de números difusos representa la imprecisión de cada dato considerándolos como intervalos de posibles valores con cierto nivel de certeza, esta mejora en gran medida la clasificación y consigue acertar más en la resolución de problemas que presentan este tipo de información, por lo que se convierte en un método mucho más efectivo ya que se adapta mejor a las expresiones del ser humano. Además, la teoría difusa no solo permite efectuar cálculos cuando hay información con incertidumbre, sino también cuando tengamos que combinar información cuantitativa y cualitativa, trata a la vez datos numéricos e información categórica con jerarquía, mediante aproximación matemática, permitiéndonos tomar decisiones en situaciones donde se requiera razonar de forma imprecisa o aproximada, lo que nos permite caracterizar de una mejor manera las distintas aplicaciones [3]. Los trabajos de investigación relacionados con el tema aquí desarrollado son los siguientes: [18] y [29]. En ambos trabajos, versiones del problema de control difuso descontado total con espacios de estados y acciones finitos.

El contenido de este trabajo está estructurado de la forma siguiente: el

Capítulo 1 introduce los conceptos básicos de la teoría de conjuntos difusos, destacando a los números difusos trapezoidales junto con su aritmética y propiedades de interés fundamentales para desarrollar los resultados que se aplicarán en los capítulos posteriores. Se describe el orden entre números difusos y la métrica utilizada en las que se basó el procedimiento propuesto. Finalmente, dado que los estados del sistemas son aleatorios, lo que hace que las recompensas sean aleatorias difusas y la esperanza en la función objetivo se trate de la esperanza de variable aleatoria difusa, se establece la definición de los elementos aleatorios con valores de números difuso y sus correspondientes valores esperados, [9], [31] y [33]. Con esto estaríamos proporcionando las herramientas necesarias de la parte de la teoría de números difusos. En la primera parte Capítulo 2, se brindan los conceptos básicos de la teoría estándar sobre los PDMs [24] con espacio estado finitos tanto con criterio de recompensa total y recompensa descontada. Para tales tipos de PDMs, la función de recompensa se plantea difusa trapezoidal conveniente en función de una recompensa estándar nítida. El problema de control difuso consiste en determinar una política de control que maximice la recompensa total esperada difusa y una que maximice la recompensa descontada esperada difusa. La política óptima y la función de valor óptimo para el problema de control difuso se caracterizan por medio de la ecuación de programación dinámica del problema de control óptimo estándar y, se obtiene que la política de control óptimo del problema estándar y del difuso coinciden. Además la función de valor óptimo difuso tiene una forma trapezoidal afín en función de la función de valor óptimo estándar, quedando caracterizada su solución por la solución del problema estándar. Por lo tanto, problema de control difuso se reduce al problema de control óptimo estándar. Este es el principal aporte de este trabajo en el campo del control difuso. En los Capítulo 3 y 4 ilustramos la teoría desarrollada, proporcionando aplicaciones de esta a problemas en extensiones difusas para un problema de inventario [22], de paro óptima, de apuesta, selección de portafolio y un juego entre dos personas [30]. Finalmente se brindan las conclusiones.

Capítulo 1

Conceptos básicos de teoría difusa

Antes de proporcionar las definiciones y resultados básicos sobre la teoría de lógica difusa que son fundamentales en el desarrollo de esta tesis, iniciaremos dando una breve explicación sobre los conjuntos difusos y la teoría de lógica difusa con el fin de poder distinguir mejor entre estos conjuntos y los que no lo son.

Conjuntos como por ejemplo, el de las computadoras, sabemos muy bien quienes son sus elementos, este incluye a todas las computadoras, pero excluye a los celulares. Conjuntos como estos se conocen como conjuntos nítidos, certeros o clásicos, ya que cada elemento del conjunto de referencia o pertenece o no pertenece a él, y bien sabemos que dicha pertenencia está determinada por la función indicadora, la cual toma solo uno de los valores del conjunto $\{0, 1\}$ para cada elemento del conjunto de referencia, esta es la manera en que la función indica si el elemento pertenece o no al conjunto. Pero también existen conjuntos difusos, por ejemplo, el de las personas sabias, el de las personas altas o el de los vasos anchos entre otros. Si consideramos los juicios declarativos:

- Una persona de 30 años es sabia.
- Una persona de 170 cm es alta.
- Un vaso de 8 cm de diámetro es ancho.

No podríamos responder ni que son absolutamente verdaderos ni que son completamente falsos de forma objetiva, por lo cual no podemos definir claramente la pertenencia de los elementos al conjunto de las personas sabias, al de las personas altas ni al de los vasos anchos porque, ¿a partir de qué momento decidimos que la persona deja de ser sabia o alta, o que el vaso deja de ser ancho

y pasan a ser de la otra clasificación?, si una persona de 170 cm es considerada alta y le quitamos 5 mm ¿ya no es considerada como alta sino como una persona de baja estatura?. Recordemos que la altura promedio de una persona mexicana se encuentra entre 1.58 y 1.64 metros. En estos casos no solo vamos a considerar dos opciones, que es sabia o que no es sabia, que es alta o que no es alta, que es ancho o que no es ancho. El conjunto de las personas altas es un subconjunto difuso del conjunto de todas las personas, y nos permitirá considerar toda una gama de opciones, personas muy altas, altas, de estatura promedio, de estatura baja y de estatura muy baja. No hay una transición clara entre lo que es falso y lo que es verdadero, contrariamente a si los consideráramos como conjuntos clásicos.

La teoría de lógica difusa se aplica a conceptos que ni son totalmente ciertos ni completamente falsos, considerando una tercera posibilidad de pertenencia, la pertenencia parcial, que es cuando un elemento puede pertenecer parcialmente a un subconjunto dado. Esta es la diferencia fundamental entre los conjuntos difusos y los nítidos, que un elemento puede estar parcialmente ausente o presente, y esto no sucede en los conjuntos nítidos, donde la pertenencia y la ausencia de un elemento a un conjunto son mutuamente excluyentes. La teoría de conjuntos difusos considera la pertenencia de los elementos de un conjunto como una transición que es gradual al permitir que sus valores de veracidad estén dentro del intervalo $[0, 1]$, donde 0 indica la falsedad total, 1 indica la verdad absoluta, y cualquier valor de pertenencia entre cero permite medir pertenencia parcial de un elemento del conjunto de referencia al subconjunto difuso dado. Por esta razón, la teoría de conjuntos difusos es conocida como una lógica de múltiples valores, ya que permite definir a los valores intermedios entre verdadero o falso, o como en el ejemplo de la altura de las personas, define a los valores intermedios entre alto o bajo, o entre sí o no, es decir, traslada la transición entre la pertenencia y no pertenencia a un conjunto que es gradual y mientras mayor sea el grado de pertenencia (más cercano a 1), más pertenece el elemento al subconjunto difuso.

La definición formal de un subconjunto difuso se muestra a continuación.

1.1. Conjuntos difusos

Definición 1.1. *Sea Λ un conjunto no vacío. Entonces un subconjunto difuso Γ en Λ se define en términos de una función de pertenencia $\widetilde{\Gamma} : \Lambda \longrightarrow [0, 1]$.*

Esta función de pertenencia $\widetilde{\Gamma}$, no es más que una función que permite entrelazar los elementos del conjunto de referencia Λ con los elementos del intervalo $[0, 1]$, ya que asigna a cada elemento x de Λ un valor real $\widetilde{\Gamma}(x)$ dentro del intervalo $[0, 1]$, el cual mide qué tanto pertenece x del conjunto de referencia Λ

al subconjunto Γ con características de impresición.

El grado de pertenencia considera la transición gradual desde la no pertenencia hasta la pertenencia total de $x \in \Lambda$ al conjunto difuso Γ. Así, $\widetilde{\Gamma}(x) = 0$ significa que x no pertenece a Γ, $\widetilde{\Gamma}(x) = 1$ significa pertenencia total de x en Γ y $0 < \widetilde{\Gamma}(x) < 1$ significa pertenencia parcial de x en Γ. Mientras más cercano a 1 sea el grado de pertenencia de x, más pertenece x al subconjunto difuso Γ de Λ.

Así que, cuando trabajamos con un subconjunto difuso, lo primero que necesitamos hacer es representarlo de la manera más precisa posible definiendo una función de pertenencia que caracterice a dicho conjunto, esta no es única, ya que va a depender de la realidad que pretendamos describir, sin embargo, suelen usarse algunas funciones clásicas comunes como las que se muestran más adelante las cuales dan flexibilidad a la modelización que utiliza expresiones linguísticas.

1.2. $\alpha-$cortes

Uno de los conceptos más convenientes por su gran utilidad dentro de la teoría de conjuntos difusos para realizar operaciones aritméticas entre ellos, es la de sus $\alpha-$cortes, ya que permiten descomponerlos y para así determinar de manera más simple algunas propiedades de las operaciones aritméticas entre números difusos.

Definición 1.2. *El* α-corte *de un conjunto difuso* Γ*, denotado por* Γ_α*, se define como el conjunto* $\Gamma_\alpha := \{x \in \Lambda \mid \widetilde{\Gamma}(x) \geq \alpha\}$ $(0 < \alpha \leq 1)$ *y* Γ_0 *se define como la cerradura de* $\{x \in \Lambda \mid \widetilde{\Gamma}(x) > 0\}$ *denotado por* $cl\{x \in \Lambda \mid \widetilde{\Gamma}(x) > 0\}$.

Denotaremos por Γ_0 y Γ_1 al soporte y al nucleo de cualquier conjunto difuso Γ, respectivamente.

La definición de los $\alpha-$cortes indica que son las proyecciones de los cortes a través del gráfico de un conjunto difuso sobre el conjunto de referencia Λ. Estos permiten describir a todos los niveles con el que se tiene una seguridad de que los elementos pertenecen o no al conjunto difuso.

1.3. Números difusos

Los números difusos son una clase especial de conjuntos difusos.

Definición 1.3. *Un* número difuso Γ *es un subconjunto difuso definido en el conjunto de números reales* $\mathbb{R}$ *(es decir, tomando* $\Lambda = \mathbb{R}$ *en la Definición 1.1), que satisface:*

a) $\widetilde{\Gamma}$ *es normal, es decir, existe* $x_0 \in \mathbb{R}$ *con* $\widetilde{\Gamma}(x_0) = 1$*;*

b) $\widetilde{\Gamma}$ *es convexa, lo que implica que* $\widetilde{\Gamma}_\alpha$ *es convexo para todo* $\alpha \in [0, 1]$*;*

c) $\widetilde{\Gamma}$ *es semicontinua superiormente;*

c) Γ_0 *es compacto.*

La función de pertenencia debe representar a los números reales cercanos a un número real específico r, y ya que r satisface la condición, entonces se debe cumplir que $\widetilde{\Gamma}(r) = 1$, esta es la razón por la cual la función de pertenencia de un número difuso es normal.

Las propiedades citadas en la Definición 1.3 implican que la función de pertenencia corresponde a un números difuso si y solo si, es de la siguiente forma:

$$\Gamma'(x) = \begin{cases} 0 & \text{si} \quad x \leq w_1 \\ l(x) & \text{si } x \in (w_1, a) \\ 1 & \text{si } x \in [a, b] \\ r(x) & \text{si } x \in (b, w_2) \\ 0 & \text{si} \quad x \geq w_2 \end{cases} \tag{1.1}$$

con $0 \leq w_1 \leq a \leq b \leq w_2 \in \mathbb{R}$, [17], donde $l(x)$ es una función continua a la derecha y creciente en (w_1, a) y $r(x)$ una función decreciente y continua a la izquierda en (b, w_2), ambas con rango $[0, 1]$.

Al conjunto de los números difusos se denotará por $\mathfrak{F}(\mathbb{R})$, este conjunto es una extensión de los números reales.

Existe una gran diversidad de formas para las funciones de pertenencia asociadas a un número difuso. Ejemplos de las más usadas son la trapezoidal y la triangular, por tener formas gráficas más simples, lo que permite que se le asocie una interpretación más natural [8]. La ecuación de un número difuso trapezoidal se muestra a continuación.

Definición 1.4. *Un número difuso* Γ *se llama* **número difuso trapezoidal**

si su función de pertenencia tiene la siguiente forma:

$$\widetilde{\Gamma}(x) = \begin{cases} 0 & si\ x \leq l \\ \frac{x-l}{m-l} & si\ l < x \leq m \\ 1 & si\ m < x \leq n \\ \frac{p-x}{p-n} & si\ n < x \leq p \\ 0 & si\ p < x\ , \end{cases} \tag{1.2}$$

donde l, m, n y p son números reales conocidos, con $l < m \leq n < p$. Un número difuso trapezoidal simplemente se denota por (l, m, n, p).

El caso en el que $m = n$ en (1.4) se llamará número difuso triangular y se denotará simplemente por (l, m, p).

Ejemplos gráficos de cómo es un número trapezoidal y triangular difuso se muestran en las Figuras 3.2 y 1.2 respectivamente.

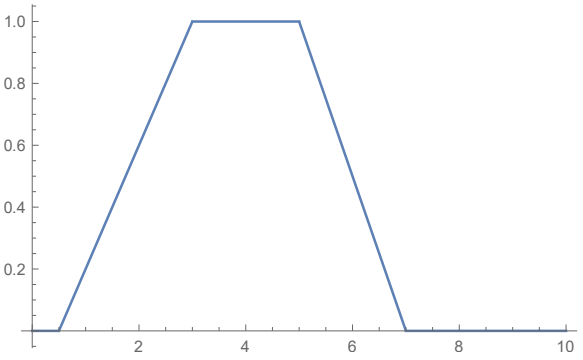

Figura 1.1: Número difuso trapezoidal $(0.5, 3, 5, 7)$.

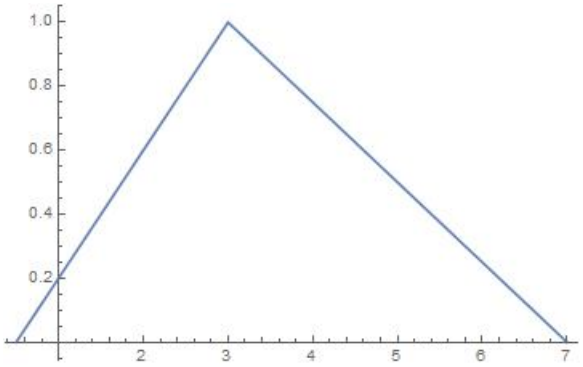

Figura 1.2: Número difuso triangular $(0.5, 3, 7)$.

Las funciones de pertenencia triangulares se usan para describir valores intermedios como el concepto de tibio sin considerar un margen de aproximación o de tolerancia alrededor del valor que se toma como el mejor representante del concepto lingüístico asociado al conjunto difuso y también para aproximar cualquier número difuso [3] y [34]. Por ejemplo, la temperatura del agua tibia se puede representar con el número difuso triangular $(18^0, 24^0, 30^0)$.

Cuando se pretende describir valores intermedios como tibio, maduro o altura promedio pero implicando un margen de aproximación o de tolerancia alrededor del valor que se toma como el mejor representante del concepto lingüístico asociado al conjunto difuso, se usa la función de pertenencia trapezoidal. Por ejemplo, una persona es considerada madura si su edad está comprendida entre 35 y 55 años. Así que el conjunto de las personas maduras se puede representar con el número trapezoidal $(0, 35, 55, 85)$.

Podemos considerar el caso degenerado en el que $l = m = p$, obteniéndose la *representación difusa* del número real m con la función de pertenencia dada por:

$$\widetilde{m}(x) = \begin{cases} 1 & \text{si } x = m \\ 0 & \text{si } x \neq m. \end{cases} \tag{1.3}$$

Estas funciones de pertenencia de números difusos que acabamos de presentar, son de las más simples de asociarles una interpretación de manera muy natural, por lo que son de las más especiales. Esto es lo que las hace ser de las más estudiadas, usadas y generalizadas en sistemas difusos [1] y [20].

Podemos observar claramente que para el caso de los números difusos, las proyecciones de sus $\alpha-$cortes son intervalos cerrados y acotados en $\mathbb{R}$, todos los $\alpha-$cortes son subconjuntos del soporte Γ_0, este es el intervalo más grande y a medida que α aumenta, los $\alpha-$cortes se van haciendo intervalos más pequeños, siendo Γ_1 el más pequeño de todos. Así, la familia de los $\alpha-$cortes forma una sucesión decreciente de conjuntos nítidos compactos.

Más específicamente, para un número difuso trapezoidal, los $\alpha-$cortes se observan en la Figura 1.3.

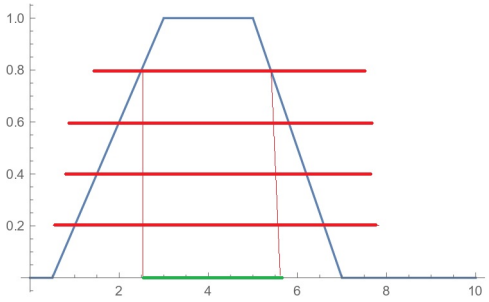

Figura 1.3: Representación gráfica para los $\alpha-$cortes de un número difuso trazoidal.

Recordemos que una de las condiciones que debe complir un conjunto difuso para ser un número difuso, es que el soporte sea compacto, y debido a que es subconjunto de $\mathbb{R}$ por tratarse de un número difuso, entonces es un intervalo cerrado y acotado, lo cual implica que todos sus $\alpha-$cortes, también serán compactos, específicamente intervalos cerrados y acotados pues forman una sucesión decreciente de conjuntos nítidos. Esto facilitará concretar las operaciones aritméticas de números difusos en términos de las operaciones aritméticas de los intervalos cerrados.

Lema 1.1. *Para un número difuso trapezoidal $\Gamma = (l, m, n, p)$, los $\alpha-$cortes correspondientes están dados por $\Gamma_\alpha = [(m-l)\alpha + l, p - (p-n)\alpha]$, $\alpha \in [0, 1]$:*

Demostración. Usando la Definición 1.2, $(l, m, n, p)(x) \geq \alpha$ si y solo si

$$\frac{x-l}{m-l} \geq \alpha \quad \text{y} \quad \frac{p-x}{p-n} \geq \alpha.$$

Esto es equivalente a

$$x \geq (m-l)\alpha + l \Leftrightarrow \qquad x \leq p - (p-n)\alpha,$$

por lo tanto

$$(l, m, n, p)_\alpha = [(m-l)\alpha + l, p - (p-n)\alpha].$$

Se puede observar fácilmente que los $\alpha-$cortes para un número difuso triangular (l, m, p), son de la forma:

$$(l, m, p)_\alpha = [(m-l)\alpha + l, p - (p-m)\alpha].$$

□

Ejemplo 1.1. *El $\alpha-$corte del número difuso trapezoidal $(2,5,7,10)$ es*

$$(2,5,7,10)_\alpha = [(5-2)\alpha+2, 10-(10-7)\alpha] = [3\alpha+2, 10-3\alpha] \quad \forall\alpha\in[0,1].$$

Para $\alpha = 0.5$,

$$(2,5,7,10)_{0.5} = [3.5, 8.5].$$

1.4. Aritmética de los números difusos

Necesitamos un teorema de representación [11] y [32] que es una herramienta básica para el análisis de números difusos, ya que nos permite descomponer a cualquier conjunto difuso en una familia de conjuntos no difusos utilizando los $\alpha-$cortes. Así también nos permite a partir de una familia de $\alpha-$cortes anidados, reconstruír a un conjunto difuso, por lo que si un problema es formulado en el marco de los conjuntos difusos, este puede ser resuelto transformando esos conjuntos difusos en su correspondiente familia de $\alpha-$cortes para determinar la solución mediante técnicas no difusas.

Teorema 1.1. *Sea $\mathfrak{C}(\mathbb{R})$ el conjunto de todos los subconjuntos convexos compactos de $\mathbb{R}$ que cumole:*

a) Para cualquier $\Gamma\in\mathfrak{F}(\mathbb{R})$, $\Gamma(x) = \sup_{\alpha\in[0,1]}\{min(\alpha, \mathbb{1}_{\Gamma_\alpha}(x))\}, x\in\mathbb{R}$.

b) Recíprocamente, para una familia de subconjuntos decreciente $\{D_\alpha \in \mathfrak{C}(\mathbb{R})|\alpha\in[0,1]\}$, el conjunto $\Gamma(x) := \sup_{\alpha\in[0,1]}\{min(\alpha, \mathbb{1}_{D_\alpha}(x)), x\in\mathbb{R}\}, x\in\mathbb{R}$ satisface que $\Gamma\in\mathfrak{F}(\mathbb{R})$.

Esto significa que todo número difuso se puede representar totalmente por los $\alpha-$cortes.

Definición 1.5. *Sean Γ y Υ conjuntos difusos. Si " $\star$ " denota la suma, resta, multiplicación o divivión entre números difusos, entonces se define un conjunto difuso en $\mathbb{R}$, $\Gamma\star\Upsilon$, mediante la función de membresía: $\widetilde{(\Gamma\star\Upsilon)}(u) = \sup_{u=x\star y}\min\{\widetilde{\Gamma}(x), \widetilde{\Upsilon}(y)\}$, para todo $u\in\mathbb{R}$.*

Proposición 1.1. *Se puede probar que si Γ y Υ son números difusos, entonces*

1) $\Gamma\star\Upsilon$ es también un número difuso.

2) $(\Gamma\star\Upsilon)_\alpha = \Gamma_\alpha\star\Upsilon_\alpha$ (para el caso del cociente, siempre que el cero no pertenezca a Υ_α para todo α).

Esto se prueba mediante la definición estandar de las operaciones entre conjuntos en $\mathbb{R}$.

De los incisos 1), 2) y el Teorema 1.1, podemos concluir que

$$\begin{aligned}(\Gamma \star \Upsilon)(x) &= \sup_{\alpha\in[0,1]} \{min(\alpha, \mathbb{1}_{(\Gamma\star\Upsilon)_\alpha}(x)), x \in \mathbb{R}\} \\ &= \sup_{\alpha\in[0,1]} \{min(\alpha, \mathbb{1}_{(\Gamma_\alpha\star\Upsilon_\alpha)}(x)), x \in \mathbb{R}\}.\end{aligned}$$

Esta forma de operar aritméticamente entre los números difusos a través de las operaciones de sus $\alpha-$cortes es muy conveniente, ya que los $\alpha-$cortes de los números difusos son intervalos cerrados y acotados, así definimos las operaciones entre los números difusos en términos de las operaciones entre intervalos, por lo cual, la raíz de los cálculos entre números difusos se encuentra en el análisis de intervalos.

Definición 1.6. *La aritmética de intervalos está definida a través de:*

a) $[a,b] + [c,d] = [a+c, b+d]$

b) $[a,b] - [c,d] = [a-d, b-c]$

c) $[a,b].[c,d] = [min(ac,ad,bc,bd), max(ac,ad,bc,bd)]$

d) $[a,b]/[c,d] = [min(\frac{a}{c}, \frac{a}{d}, \frac{b}{c}, \frac{b}{d}), max(\frac{a}{c}, \frac{a}{d}, \frac{b}{c}, \frac{b}{d})]$

Como consecuencia de esto, es posible obtener el siguiente resultado para números difusos trapezoidales [26].

En lo que sigue, usaremos las notaciones $+^\star$ y $\sum^\star$ para aclarar que estamos operando con conjuntos difusos.

Lema 1.2. *Si* $\Gamma = (l_1, m_1, n_1, p_1)$ *y* $\Upsilon = (l_2, m_2, n_2, p_2)$ *son dos números difusos trapezoidales y* λ *un número real positivo, entonces se sigue que*

a) $\Gamma +^* \Upsilon = (l_1 + l_2, m_1 + m_2, n_1 + n_2, p_1 + p_2)$.

b) Si $\{(l_k, m_k, n_k, p_k) : 1 \leq k \leq M\}$ *es un conjunto finito de* M *números difusos trapezoidales entonces*

$$\sum_{k=1}^{M}{}^{*}(l_k, m_k, n_k, p_k) = \left(\sum_{k=1}^{M} l_k, \sum_{k=1}^{M} m_k, \sum_{k=1}^{M} n_k, \sum_{k=1}^{M} p_k\right).$$

y

c) $\lambda\Gamma = (\lambda l, \lambda m, \lambda n, \lambda p)$.

Demostración. a) Usando (1.1) se tiene que

$$\Gamma_\alpha = [(m_1 - l_1)\alpha + l_1, p_1 - (p_1 - n_1)\alpha]$$

$$\Upsilon_\alpha = [(m_2 - l_2)\alpha + l_2, p_2 - (p_2 - n_2)\alpha].$$

Entonces, por la Definición 1.6 a) de suma de intervalos y por el inciso 2) de la Proposición 1.1, se tiene que

$$\begin{aligned}(\Gamma + \Upsilon)_\alpha &= \Gamma_\alpha + \Upsilon_\alpha \\ &= [((m_1 + m_2) - (l_1 + l_2))\alpha + (l_1 + l_2), (p_1 + p_2) - ((p_1 + p_2) - (n_1 + n_2))\alpha].\end{aligned}$$

Por lo tanto, la función de membresía de la suma es

$$(\widehat{r} + \widehat{s})(x) = \begin{cases} 0 & \text{si } x \leq (l_1 + l_2) \\ \frac{x-(l_1+l_2)}{(m_1+m_2)-(l_1+l_2)} & \text{si } (l_1 + l_2) \leq x \leq (m_1 + m_2) \\ 1 & \text{si } (m_1 + m_2) \leq x \leq (n_1 + n_2) \\ \frac{(p_1+p_1)-x}{(p_1+p_2)-(n_1+n_2)} & \text{si } (n_1 + n_2) \leq x \leq (p_1 + p_2) \\ 0 & \text{si } (p_1 + p_2) \leq x. \end{cases} \quad (1.4)$$

La función de pertenencia en (1.4) está asociada al número difuso

$$(l_1 + l_2, m_1 + m_2, n_1 + n_2, p_1 + p_2).$$

b) Esta prueba se realiza por inducción.

c) Usando el inciso b), se realiza la prueba.

□

1.5. Orden máximo en $\mathfrak{F}(\mathbb{R})$

En optimización difusa o en la toma de decisiones en entornos difusos, es de fundamental importancia ordenar o clasificar conjuntos difusos. En este trabajo emplearemos el orden máximo de números difuso el cual se basa en el orden de los $\alpha-$cortes, por lo que se define en términos del orden de intervalos cerrados y acotados en $\mathbb{R}$ definido de la siguiente forma:

Definición 1.7. *Sea* $\Gamma, \Upsilon \in \mathfrak{F}(\mathbb{R})$, $\Gamma_\alpha = [\Gamma_\alpha^L, \Gamma_\alpha^U]$ *y* $\Upsilon_\alpha = [\Upsilon_\alpha^L, \Upsilon_\alpha^U]$. *Entonces* $\Gamma \leqslant^* \Upsilon$ *si y solo si* $\Gamma_\alpha \leq \Upsilon_\alpha$ *para todo* $\alpha \in [0, 1]$, *es decir,* $\Gamma \leqslant^* \Upsilon$ *si y solo si* $\Gamma_\alpha^L \leq \Upsilon_\alpha^L \quad y \quad \Gamma_\alpha^U \leq \Upsilon_\alpha^U$ *para todo* $\alpha \in [0, 1]$.

Ejemplo 1.2. $(2,5,6,10)_\alpha = [3\alpha+2, 10-3\alpha]$ *y* $(7,9,13,17)_\alpha = [2\alpha+7, 17-3\alpha]$. *Además, ya que* $3\alpha+2 \leq 2\alpha+7$ *y* $10-3\alpha \leq 17-3\alpha$ *para todo* $\alpha \in [0,1]$, *entonces* $(2,5,6,10) \leq^* (7,9,13,17)$.

No es difícil verificar que el orden " $\leq^*$ " es un orden parcial en $\mathfrak{F}(\mathbb{R})$.

Observación 1.1. *Tomamos* $z_1, z_2 \in \mathbb{R}$, *y sean* Γ *y* Υ *números difusos con funciones de pertenencia dadas por* $\Gamma(x) = \Upsilon(x) = 1$, $x = z_k$ *y* $\Gamma(x) = \Upsilon(x) = 0$, $x \neq z_k$, $k = 1, 2$, *respectivamente. Entonces, es fácil ver que* $\Gamma \leqslant^* \Upsilon$ *es equivalente a* $z_1 \leq z_2$.

1.6. Métrica en el conjunto de números difusos

Definición 1.8. *Sea* $\mathcal{C}(\mathbb{R})$ *el conjunto de todos los intervalos acotados Y cerrados en* $\mathbb{R}$. *Para* $\Psi = [a_l, a_u]$, $\Phi = [b_l, b_u] \in \mathcal{C}(\mathbb{R})$ *definamos*

$$\rho_{\mathcal{C}(\mathbb{R})}(\Psi, \Phi) = \text{máx}(|a_l - b_l|, |a_u - b_u|). \tag{1.5}$$

Lema 1.3. *[2] La función d y el conjunto* $\mathcal{C}(\mathbb{R})$ *cumplen las siguientes propiedades:*

a) $\rho_{\mathcal{C}(\mathbb{R})}$ *define una métrica sobre* $\mathcal{C}(\mathbb{R})$.

b) $(\mathcal{C}(\mathbb{R}), \rho_{\mathcal{C}(\mathbb{R})})$ *es un espacio métrico completo.*

Ahora si $\Gamma, \Upsilon \in \mathfrak{F}(\mathbb{R})$, entonces Γ_α y Υ_α son conjuntos compactos porque su función de pertenencia es semicontinua superior y tiene soporte compacto. Por lo tanto, se define $\widetilde{\rho}_{\mathfrak{F}(\mathbb{R}):\mathfrak{F}(\mathbb{R})\times\mathfrak{F}(\mathbb{R})\longrightarrow\mathbb{R}}$ por

$$\widetilde{\rho}_{\mathfrak{F}(\mathbb{R})}(\Gamma, \Upsilon) = \sup_{\alpha\in[0,1]} d(\Gamma_\alpha, \Upsilon_\alpha). \tag{1.6}$$

Lema 1.4. *[25]* $\widetilde{\rho}_{\mathfrak{F}(\mathbb{R})}$ *es una métrica en* $\mathfrak{F}(\mathbb{R})$.

Definición 1.9. *Se dice que una sucesión* $\{\Gamma^n\}$ *de números difusos es convergente al número difuso* μ, *escrito como* $\lim^*_{n\longrightarrow\infty} \Gamma^n = \mu$.

Hacieindo uso del Lema 1.2 b) para números difusos trapezoidales y de la métrica de Hausdorf 1.6, se puede verificar que se cumple la siguiente afirmacion:

Lema 1.5. *Si* $\{y_k = (l_k, m_k, n_k, p_k) : k \geq 1\}$ *es una sucesión de números difusos trapezoidales tales que* $\sum_{k=1}^{\infty} l_k$,$\sum_{k=1}^{\infty} m_k$,$\sum_{k=1}^{\infty} n_k$ *y* $\sum_{k=1}^{\infty} p_k$ *convergen, entonces*

$$\sum_{k=1}^{t}{}^* y_k,$$

converge cuando $t \longrightarrow \infty$ *al número difuso trapezoidal:*

$$\left(\sum_{k=1}^{\infty} l_k, \sum_{k=1}^{\infty} m_k, \sum_{k=1}^{\infty} n_k, \sum_{k=1}^{\infty} p_k\right).$$

Lema 1.6. *[25]* $(\mathfrak{F}(\mathbb{R}), \widetilde{\rho}_{\mathcal{F}(\mathbb{R})})$ *es un espacio métrico completo.*

1.7. Variable aleatoria difusa

Siguiendo [19] y [23] se establecen las siguientes definiciones sobre variables aleatorias difusas y sus esperanzas. Para esto, $\mathfrak{C}(\mathbb{R})$ denota la clase de subconjuntos compactos no vacíos de $\mathbb{R}$, y si $(\Omega_1, \mathcal{A}_1)$ y $(\Omega_2, \mathcal{A}_2)$ son espacios medibles, entonces $\mathcal{A}_1 \otimes \mathcal{A}_2$ denota la $\sigma-$álgebra generada por el producto de las $\sigma-$álgebras $\mathcal{A}_1$ y $\mathcal{A}_2$.

En muchos de los problemas que se presentan en la realidad que involucran aleatoriedad, los datos que se requieren considerar son imprecisos. Este tipo de datos es lo que se conoce como variables aleatorias difusas, es decir, además de estar presente la aleatoriedad, también está la incertidumbre que se debe a la imprecisión en la definición de los datos. Los problemas que se consideran en los próximos capítulos son de este tipo, por tal razón, incluimos los conceptos relacionados con variables aleatorias difusa y fundamentar la formalidad de dicho concepto.

Las variables aleatorias difusas en el sentido de [23], representan elementos aleatorios cuyos valores son números y han sido un modelo útil para un gran cantidad de elementos aleatorios con valores imprecisos.

Ahora, se definirá una variable aleatoria difusa. En este caso, se adoptará la definición propuesta en [23].

Definición 1.10. *Sea* $(\Omega, \mathcal{A})$ *un espacio medible y* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *el espacio medible del conjunto de los números reales. Una función* $\tilde{Y} : \Omega \longrightarrow \mathfrak{F}(\mathbb{R})$ *se dice que es una* variable aleatoria difusa *asociada con* $(\Omega, \mathcal{A})$*, si la sección* $\tilde{Y}_\alpha : \Omega \longrightarrow \mathfrak{C}(\mathbb{R})$ *que es la función de nivel* α *definida por* $\tilde{Y}_\alpha(\omega) = (\tilde{Y}(\omega))_\alpha$ *para todo* $\omega \in \Omega$ *y* $\alpha \in [0,1]$ *satisface que* $Gr(\tilde{Y}_\alpha) = \{(\omega, x) \in \Omega \times \mathbb{R} \mid x \in (\tilde{Y}(\omega))_\alpha\} \in \mathcal{A} \otimes \mathcal{B}(\mathbb{R})$*, para todo* $\alpha \in [0,1]$*. Equivalentemente,* $\tilde{Y}$ *debe verse como un intervalo generalizado con una función de pertenencia* μ *y* $\alpha-$*corte:* $Y(\omega)_\alpha = [Y^-(\omega), Y^+(\omega)]$.

Definición 1.11. *Sea* $(\Omega, \mathcal{F}, P)$ *un espacio de probabilidad y* $\tilde{X}$ *una variable aleatoria discreta con rango* $\{\tilde{s}_1, \tilde{s}_2, ..., \tilde{s}_l\} \subseteq \mathfrak{F}(\mathbb{R})$*. La esperanza matemática*

de $\tilde{Y}$ es un número difuso, $E(\tilde{Y})$, tal que

$$E(\tilde{Y}) = \sum_{i=1}^{l} \tilde{s}_i P(\tilde{Y} = \tilde{s}_i). \tag{1.7}$$

Definición 1.12. *Dado un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ una variable aleatoria difusa $\tilde{Y}$ asociada a $(\Omega, \mathcal{A})$ se dice que es una variable aleatoria difusa integrable acotada con respecto a $(\Omega, \mathcal{A}, P)$ si existe una función $h : \Omega \longrightarrow \mathbb{R}$, $h \in L^1(\Omega, \mathcal{A}, P)$ tal que para todo $(\omega, x) \in \Omega \times \mathbb{R}$ con $x \in \tilde{Y}_0(\omega)$, se cumple que $|x| \leq h(\omega)$.*

1.8. Esperanza de una variable aleatoria difusa

Ahora presentemos la formalización del valor esperado de una variable aleatoria difusa, destacando una de sus propiedades relevantes para nuestro estudio en los próximos capítulos.

Definición 1.13. *Dada una variable aleatoria difusa acotada e integrable $\tilde{Y}$ asociada con respecto el espacio de probabilidad $(\Omega, \mathcal{A}, P)$, entonces el valor esperado difuso de $\tilde{Y}$ en el sentido de Aumann es el único conjunto difuso de $\mathbb{R}$, $E^*[\tilde{Y}]$ tal que para cada $\alpha \in [0, 1]$:*

$$\left(E^*[\tilde{Y}]\right)_\alpha = \left\{ \int_\Omega f(\omega) dP(\omega) \mid f : \Omega \longrightarrow \mathbb{R}, f \in L^1(P), f(\omega) \in (\tilde{Y}(\omega))_\alpha \quad [P] \right\}. \tag{1.8}$$

Lema 1.7. *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad. Sea Y una variable aleatoria discreta no negativa asociada a $(\Omega, \mathcal{A}, P)$ tal que $E[Y]$ existe. Entonces, $\tilde{Y} = Y(B, C, D, F)$ es una variable aleatoria difusa asociada a $(\Omega, \mathcal{A}, P)$, y*

$$E^*[\tilde{Y}] = E[Y](B, C, D, F). \tag{1.9}$$

Demostración. Sea Y una variable aleatoria discreta no negativa con rango finito o numerable denotada por $Y[\Omega] = \{y_1, y_2, ...\}$ y sea $[Y = y_j] := \{\omega \in \Omega \mid Y(\omega) = y_j\}$, $j = 1, 2, ...$ Tomemos $\Theta = (B, C, D, F)$ con α-cortes $\Theta_\alpha = [q(\alpha), s(\alpha)]$, $\alpha \in [0, 1]$. Fijemos $\alpha \in [0, 1]$. Considere la multifunción dada por

$$\tilde{Y}_\alpha(\omega) := (\tilde{Y}(\omega))_\alpha = (Y(\omega)\Theta)_\alpha = Y(\omega)[q(\alpha), s(\alpha)], \tag{1.10}$$

$\omega \in \Omega$.

Ahora, notemos que

$$\begin{aligned} Gr(\tilde{Y}_\alpha) &= \{(\omega, x) \in \Omega \times \mathbb{R} \mid x \in \tilde{Y}_\alpha(\omega)\} \\ &= \{(\omega, x) \in \Omega \times \mathbb{R} \mid x \in Y(\omega)[q(\alpha), s(\alpha)]\} \\ &= \bigcup_j ([Y = y_j] \times y_j[q(\alpha), s(\alpha)]). \end{aligned} \tag{1.11}$$

Por lo tanto, $Gr(\tilde{Y}_\alpha) \in \mathcal{A}\otimes\mathcal{B}(\mathbb{R})$. Como α es arbitraria, de la Definición 1.10 se deduce que $\tilde{Y}$ es una variable aleatoria difusa. A continuación, tenga en cuenta que, para cada $\omega \in \Omega$,

$$\tilde{Y}_0(\omega) = Y(\omega)[B, F].$$

Definamos $h : \Omega \longrightarrow \mathbb{R}$ dado por

$$h(\omega) := Y(\omega)F,$$

$\omega \in \Omega$. Entonces, trivialmente:

$$|x| \leq h(\omega),$$

$(\omega, x) \in \Omega\times\mathbb{R}$ con $x \in Y(\omega)[B, F]$. Además, claramente $E[h] = FE[Y]$ es finito. Por lo tanto, a partir de la Definición 1.12, $\tilde{Y}$ es una variable aleatoria difusa integrable acotada con respecto a $(\Omega, \mathcal{A}, P)$. Ahora, a partir de la Definición 1.13, existe un único valor esperado difuso $E^*[\tilde{Y}]$, para cada α,

$$\begin{aligned}
E[Y]\Theta_\alpha &= \int_\Omega Y(\omega)dP(\omega)[q(\alpha), s(\alpha)] \\
&= \int_\Omega Y(\omega)dP(\omega)\{x : x \in [q(\alpha), s(\alpha)]\} \\
&= \left\{\int_\Omega Y(\omega)xdP(\omega) : Y(\omega)x \in Y(\omega)[q(\alpha), s(\alpha)]\right\} \\
&= \left\{\int_\Omega f(\omega)dP(\omega) : f : \Omega \longrightarrow \mathbb{R}, f \in L^1(P), f(\omega) \in (\widetilde{Y}(\omega))_\alpha a.s[P]\right\} \\
&= \left(E^*[\widetilde{Y}]\right)_\alpha \qquad (1.12)
\end{aligned}$$

por lo cual, $(E^*[\tilde{Y}])_\alpha = E[Y][q(\alpha), s(\alpha)]$ para cada α, es el α-corte del número trapezoidal dado para

$$E[Y](B, C, D, F),$$

es decir,

$$E^*[\tilde{Y}] = E[Y](B, C, D, F).$$

□

Lema 1.8. *Sean $\tilde{X}$ y $\tilde{Y}$ variables aleatorias difusas de tipo trapezoidales. Entonces*

a) $E[\tilde{X}] \in \mathfrak{F}(\mathbb{R})$.

b) $E[\tilde{X} + \tilde{Y}] = E[\tilde{X}] + E[\tilde{Y}]$.

c) $E[\lambda\tilde{X}] = \lambda E[\tilde{X}]$, $\lambda \geq 0$.

Los conceptos y resultados sobre números difusos que hemos presentado hasta ahora, son la base para los resultados obtenidos en el Capítulo 2 sobre PDMs con criterios de rcompensa esperada total y el caso descontado.

Capítulo 2

PDMs: Versión nítida y difusa

Cuando se intenta resolver problemas relacionados con sistemas que evolucionan de forma aleatoria y que consideran las recompensas que se obtendrán de las decisiones actuales y a las posibles oportunidades de toma de decisiones en el futuro, puede suceder que la probabilidad de que ocurra un evento esté en función solamente de lo ocurrido en la etapa inmediata anterior que se ha observado (estado actual del sistema) y no de toda la historia observada en el pasado, es decir, que satisface la propiedad de Markov. Adicional a esto, los datos requeridos para la modelación podrían ser imprecisos. Por tal razón, en este capítulo se presentará primeramente la teoría de Procesos de Decisión de Markov (PDMs) en su versión nítida, los cuales son una clase muy especial de modelos de decisión secuencial que están planteados con algún componente estocástico y que modelan la evolución temporal de muchos sistemas aleatorios que satisfacen la propiedad de Markov. Luego introduciremos un MDP difuso conveniente de PDM difuso en las próximas secciones. La literatura detallada sobre la teoría de procesos de decisión de Markov se puede consultar en las referencias: [16] y [24].

2.1. Modelo de decisión de Markov: caso nítido

A continuación, explicamos el modelo de decisión de Markov, el cual es un modelo de toma de decisiones secuenciales con la propiedad de Markov, es decir, es un modelo que consiste de una serie de etapas llamadas épocas de decisión en las que en cada una de ellas, se observa el estado del sistema, se toma una decisión bajo la incertidumbre sobre el estado del sistema en la próxima época de desición, y que además la probabilidad de que ocurra un evento dependa

solamente del estado actual del sistema. El estado actual entonces proporciona toda la información útil para pronósticos, por lo que el desarrollo pasado puede ser olvidado porque sólo el presente influye.

Definición 2.1. *Un modelo de decisión de Markov es una quíntupla que consiste de los siguientes elementos:*

$$M := (X, A, \{A(x) : x \in X\}, Q, R) \tag{2.1}$$

donde

a) X es un conjunto finito, el cual es llamado el espacio de estados del sistema.

b) A es un espacio de Borel denominado el espacio de control o de acciones factibles.

c) Definimos $\{A(x) : x \in X\}$ es una familia no vacía de subconjuntos $A(x)$ de A, donde los elementos son las acciones factibles cuando el estado del sistema es x.

d) Q es una ley de transición, el cual es un kernel estocástico en X dado $\mathbb{K} := \{(x, a) : x \in X, a \in A(x)\}$. Donde $\mathbb{K}$ es denominado el conjunto de pares de estado−acciones factibles del sistema.

d) $R : \mathbb{K} \longrightarrow \mathbb{R}$ es una función de recompensa en un paso.

2.1.1. Políticas

Dado un Modelo de control de Markov, introduciremos el concepto de política.

Definición 2.2. *Una política es una sucesión $\pi = \{\pi_t : t = 0, 1, ...\}$ de kérneles estocásticos π_t en el conjunto de control A dada la historia $\mathbb{H}_t$ del proceso hasta el tiempo t, donde $\mathbb{H}_t := \mathbb{K} \times \mathbb{H}_{t-1}, t = 1, 2, ...$ y $\mathbb{H}_0 = X$.*

Las políticas o estrategias son fórmulas que eligen una acción en cualquier evento que ocurra en el futuro. El conjunto de todas las políticas es denotado por Π.

Definición 2.3. *Una política de Markov determinística es una sucesión $\pi := \{f_t\}$ tal que $f_t \in \mathbb{F}$ para $t = 0, 1, ...$, donde $\mathbb{F}$ denota el conjunto de todas las funciones $f : X \longrightarrow A$ tales que $f(x) \in A(x)$, para toda $x \in X$.*

Al conjunto de todas las políticas Markovianas las denotaremos por $\mathbb{M}$.

Definición 2.4. *Una política de Markov $\pi = \{f_t\}$ se dice que es estacionaria si f_t es independiente de t, es decir que $f_t = f$ para toda $t = 0, 1....$*

En este caso, π es identificada como f y $\mathbb{F}$ denota el conjunto de políticas estacionarias.

2.1.2. Construcción del proceso de Markov

El modelo de control de Markov y las políticas generan el espacio de probabilidad que da lugar al Proceso estocástico de interés (el Proceso de Decisión de Markov). Dicho espacio de probabilidad es $(\Omega', \mathcal{F}')$, el cual consiste del espacio muestral canónico $\Omega' = \mathbb{H}_\infty := (X \times A)^\infty$ y $\mathcal{F}'$ la correspondiente $\sigma-$álgebra producto. Los elementos de Ω' son sucesiones de las forma $\omega = (x_0, a_0, x_1, a_1 ...)$ con $x_t \in X$ y $a_t \in A$ para toda $t = 0, 1, ...$ Las proyecciones x_t y a_t son llamadas las variables de estados y acciones, respectivamente.

Sea $\pi = \{\pi_t\}$ una política arbitraria y μ una medida de probabilidad arbitraria en X llamada la distribución inicial. Entonces, por el teorema de C. Ionescu-Tulcea [24], existe una única medida de probabilidad P_μ^π en $(\Omega', \mathcal{F}')$ la cual tiene soporte en $\mathbb{H}_\infty$, es decir, $P_\mu^\pi(\mathbb{H}_\infty) = 1$ y tal que, para cada $B \in \mathcal{B}(X)$, $C \in \mathcal{B}(A)$ y $h_t \in \mathbb{H}_t$

$$\begin{aligned} P_\mu^\pi(x_0 \in B) &= \mu(B), \\ P_\mu^\pi(a_t \in C|h_t) &= \pi_t(C|h_t), \\ P_\mu^\pi(x_{t+1} \in B|h_t, a_t) &= Q(B|x_t, a_t). \end{aligned} \tag{2.2}$$

La tercera ecuación en (2.2) se llama propiedad de Markov, así que con conocimiento del presente, el pasado ejerce ninguna influencia en el futuro.

El proceso estocástico $(\Omega', \mathcal{F}', P_\mu^\pi\{x\})$ es llamado **Proceso de Decisión de Markov a tiempo discreto o Proceso de Decisión de Markov**.

Observación 2.1. *El operador esperanza con respecto a P_μ^π lo denotaremos por $E_{\mu,\pi}$. Si μ está concentrada en un estado inicial $x \in X$, entonces P_μ^π y $E_{\mu,\pi}$ son escritas como P_x^π y $E_{x,\pi}$, respectivamente.*

2.1.3. Ley de transición para un modelo de ecuaciones en diferencias

Con frecuencia, la ley de transición de un proceso de control de Markov es especificado por una ecuación en diferencias de la forma

$$x_{t+1} = F(x_t, a_t, \xi_t), \tag{2.3}$$

$t = 0, 1, 2, ...$, con $x_0 = x \in X$ conocida, donde $\{\xi_t\}$ es una sucesión de variables aleatorias independientes e identicamente distribuidas (i.i.d.) con valores en un espacio finito S y una distribución común Δ independiente del estado inicial x_0 y $F : \mathbb{K} \times S \longrightarrow X$ es una función medible conocida. En tal caso, la ley de transición Q está dada por:

$$\begin{aligned} Q(B|x,a) &= P(x_{t+1} \in B | x_t = x, a_t = a) \\ &= P(F(x_t, a_t, \xi_t) \in B | x_t = x, a_t = a) \\ &= P(F(x, a, \xi_t) \in B) \\ &= \int_X I_B(F(x,a,s)) d\mu(s) \\ &= E[I_B(F(x,a,\xi))], \end{aligned} \tag{2.4}$$

con $B \in B(X)$ y $(x, a) \in \mathbb{K}$, donde I_B es la función indicadora del conjunto $B \subseteq X, E$ es la esperanza con respecto a la distribución μ y ξ es un elemento genérico de la sucesión $\{\xi_t\}$.

2.2. PDMs con recompensa total esperada

En esta sección consideremos un Modelo de Decisión de Markov estacionario a tiempo discreto y un conjunto de políticas Π, definimos a continuación el criterio de rendimiento conocido como **recompensa total esperada**.

Definición 2.5. *Para cada $x \in X$ y $\pi \in \Pi$, la recompensa total esperada en la etapa N es la ganancia cuando se ha usado la estrategia π, dado que el estado inicial del sistema es x y se define por*

$$v(\pi, x) := E_{\pi,x}\left[\sum_{t=0}^{N-1} R(X_t, a_t) + R_N(X_n)\right] \pi \in \Pi, x \in X. \tag{2.5}$$

El criterio de recompensa total esperada cuando se ha usado la estrategia π, dado que el estado inicial del sistema es x se define por

$$v(\pi, x) := E_{\pi,x}\left[\sum_{t=0}^{\infty} R(X_t, a_t)\right]. \tag{2.6}$$

Definición 2.6. *El máximo beneficio es entonces la función de valor óptimo y se defne como:*

$$V(x) := \sup_{\pi \in \Pi} v(\pi, x), x \in X. \tag{2.7}$$

2.2.1. Problema de control óptimo para el modelo

El problema de control óptimo consiste en encontrar una política $\pi^* \in \Pi$ tal que

$$v(\pi^*, x) = V(x) \tag{2.8}$$

$x \in X$, en tal caso, π^* es llamada una **política óptima**. Esta es una sucesión especial ya que cuando operamos con ella, se obtiene el mejor beneficio.

El siguiente teorema, proporciona un algoritmo para encontrar la función de valor $V(x)$ y a una política óptima π^*. Bajo condiciones adecuadas sobre la función de recompensa en un paso y la ley de transición se caracterizan las funciones de valores óptimos V mediante una ecuación funcional. El conocimiento de V permite obtener una política óptima determinista Markoviana estacionaria.

Teorema 2.1. *Sean $V_0, V_1, ..., V_N$ funciones en X definidas hacia atrás por*

$$V_N(x) := R_N(x) \tag{2.9}$$

y para $t = N-1, N-2, ..., 1, 0$

$$V_t(x) := máx_{A(x)} \left[R(x,a) + \int V_{t+1}(y) Q(dy|x,a) \right]. \tag{2.10}$$

Supongamos que estas funciones son medibles y que para cada $t = 0, 1, 2, 3..., N-1$, existe un selector $f_t \in \mathbb{F}$ tal que $f_t(x) \in A(x)$ alcanza el máximo en la ecuación (2.10) para todo $x \in X$. Esto es que $\forall x \in X$ y $t = 0, 1, ..., N-1$,

$$V_t(x) := R(x, f_t) + \int V_{t+1}(y) Q(dy|x, f_t). \tag{2.11}$$

Entonces, la política de Markov determinística $\pi^* = \{f_0, f_1, ..., f_{N-1}\}$ es óptima y la función de valor V es igual a V_0, es decir,

$$V(x) = V_0(x) = v(\pi^*, x) \quad \forall x \in X. \tag{2.12}$$

Este teorema impone al modelo de control de Markov una importante suposición, la cual se denomina condición selección medible, el cual puede enunciarse de varias maneras. Los siguientes supuestos son una recopilación de las condiciones necesarias para resolver los PDMs con la técnica de Programación Dinámica, que se basan en garantizar que se cumpla la Condición de Selección Medible y el método de aproximaciones sucesivas si se trabaja en horizonte infinito.

Lema 2.1. *Dado un modelo de control de Markov y sea* $u : X \longrightarrow R$ *una función medible, entonces*

$$u^*(x) := máx_{A(x)} \left[R(x,a) + \int u(y)Q(dy|x,a) \right] \tag{2.13}$$

es medible y existe un selector $f \in \mathbb{F}$ *tal que la función entre corchetes alcanza su máximo en* $f(x) \in A(x)$ *para todo* x*, es decir*

$$u^*(x) := R(x,f) + \int u(y)Q(dy|x,f). \tag{2.14}$$

Enunciamos algunas condiciones generales bajo las cuales se sostiene el supuesto del Lema 2.1.

Condición 2.1. *a) El conjunto de restricciones de control* $A(x)$ *es compacto para todo* $x \in X$*;*

b) La recompensa R *es tal que* $R(x,.)$ *es l.s.c. (semicontinua inferior) en* $A(x)$ *para cada* $x \in X$*;*

c) La función $\int_X v(y)Q(dy|x,a)$ *definida en* $\mathbb{K}$ *satisface una de las dos condiciones siguientes:*

1) $\int_X v(y)Q(dy|x,.)$ *es l.s.c. en* $A(x)$ *para cada* $x \in X$ *y cada función acotada continua* v *en* X*;*

2) $\int_X v(y)Q(dy|x,.)$ *es l.s.c. en* $A(x)$ *para cada* $x \in X$ *y cada función acotada medible* v *en* X*.*

Condición 2.2. *a)* $A(x)$ *es compacto para todo* $x \in X$ *y la multifunción* $x \mapsto A(x)$ *es l.s.c.*

b) La recompensa R *es l.s.c. y acotada por debajo.*

c) La ley de transición cumple una de los siguientes supuestos:

1) Débilmente continua, es decir, $\int_X v(y)Q(dy|x,a)$ *es continua y acotada en* $\mathbb{K}$ *para cada función acotada continua* v *en* X*.*

2) Fuertemente continua, $\int_X v(y)Q(dy|x,a)$ es continua y acotada en $\mathbb{K}$ para cada función medible acotada medible v en X.

Definición 2.7. *Una función $v : \mathbb{K} \longrightarrow \mathbb{R}$ se dice que es inf−compacta en $\mathbb{K}$, si para todo $x \in X$ y $r \in \mathbb{R}$, el conjunto $\{a \in A(x)|v(x,a) \leq r\}$ es compacto.*

Condición 2.3. *a) La recompensa R es inf−compacta en $\mathbb{K}$, l.s.c. y acotada inferiormente*

b) La ley de transición cumple uno de los siguientes supuestos:

1) Débilmente continua.

2) Fuértemente continua.

Teorema 2.2. *a) Cada una de las Condiciones 2.1 y 2.2, implican el Supuesto 2.1 para cualquier función medible no−negativa.*

b) La Condition 2.3 implica el Lema 2.1 si, bajo (b1), v es non−negativa l.s.c., o, bajo (b2), si u es una función medible nonnegativa. Si, adicionalmente, la multifunctión

$$x \mapsto A^*(x) := \{a \in A(x)|u^* = c(x,a) + \int u(y)Q(dy|x,a)\}$$

es semicontinua inferiormente, entonces u^* es semicontinua inferiormente.

En las dos secciones siguientes consideremos el espacio de estados de tipo discreto, por lo que las ecuaciones (2.5-2.7) quedan de la siguiente forma:

- Recompensa total esperada:

$$V(i,\pi) = E_{i,\pi}\left[\sum_{t=0}^{\infty} R(x_t, a_t)\right]. \tag{2.15}$$

- Recompensa en T−etapas.

$$V(i,\pi) = E_{i,\pi}\left[\sum_{t=0}^{T} R(x_t, a_t)\right]. \tag{2.16}$$

Por lo tanto, una estrategia π^* es óptima si para todo $i \in X$

$$V(i) = \sup_{\pi \in \Pi} V(i,\pi^*). \tag{2.17}$$

La función V es llamada la función de valor óptimo.

Supuesto 2.1. *1. Para cada* $(i,a) \in \mathbb{K}$, $R(i,a) \geq 0$

2. $V(i) < \infty$, *para cada* $i \in X$.

Lema 2.2. *[24] y [27]*

- *La función de valor óptimo satisface la siguiente ecuación de optimalidad: para cada* $i \in X$,

$$V(i) = \sup_{a \in A(i)} \left[R(i,a) + \sum_{j \in X} p_{i.j}(a) V_0(j) \right]. \tag{2.18}$$

- *Si* $W : X \longrightarrow [0,\infty)$ *satisface que* $W(i) \geq sup_{a \in A(i)}[R(i,a) + \sum p_{i.j}(a) W(j)]$

para cada $i \in X$, *entonces* $W \geq V_0$.

Lema 2.3. *[7]. Bajo el supuesto 2.1, existe una política estacionaria óptima* f_0.

2.3. PDMs con recompensa total esperada difusa

Ahora presentamos el nuevo modelo de decisión de Markov difuso.

Definición 2.8. *Un modelo de decisión de Markov es una quíntupla que consiste de los siguientes elementos:*

$$\left(X, A, \{A(i) \mid i \in X\}, \{p_{ij}(a) \mid i,j \in X\}, a \in A(i)\}, \widetilde{R} \right), \tag{2.19}$$

El modelo de decisión de Markov difuso tiene los mismos componentes que el modelo de decisión de Markov nítido (2.1), solo que ahora se considera una función de recompensa difusa. Así, la evolución de un sistema difuso estocástico es la siguiente: si el sistema está en el estado $x_t = x \in X$ en el tiempo t y se aplica el control $a_t = a \in A(x)$, entonces pasan dos cosas:

a) se obtiene una recompensa difusa $\tilde{R}(x,a)$.

b) el sistema transita al siguiente estado x_{t+1} de acuerdo con la ley de transición Q, es decir,

$$Q(B|x,a) = Prob(x_{t+1} \in B | x_t = x, a_t = a),$$

con $B \subseteq X$.

Los conjuntos de políticas estacionarias y aleatorias coinciden para ambos modelos, además, para cada $i \in X$ y $\pi \in \Pi$ existe un espacio canónico $(\Omega, \mathcal{A}, P_{i,\pi})$ con la sucesión correspondiente $\{x_0, a_0, x_1, a_1, ...\}$ de estados y decisiones respectivamente.

Antes de definir una función objetivo para el modelo (2.19), se establecerá la función de recompensa difusa $\widetilde{R}$ a utilizar, la que será específicamente de tipo trapezoidal bajo el siguiente supuesto.

Supuesto 2.2. *Sea $R(i,a) \geq 0$ para cada $(i,a) \in \mathbb{K}$ y sean B, C, D, y F números no negativos tales que: $0 \leq B < C \leq D < F$. Se supondrá que*

$$\widetilde{R}(i,a) = R(i,a)\,(B,C,D,F)\,, \tag{2.20}$$

para todo $i \in X$ y $a \in A(i)$, donde $R : \mathbb{K} \longrightarrow \mathbb{R}$ es una función de recompensa no negativa como se consideró en la sección anterior y (B,C,D,F) es un número trapezoidal como en (1.4).

Observación 2.2. *Observemos que $\widetilde{0} \leqslant^* \widetilde{R}(i,a)$, para todo $i \in X$ y $a \in A(i)$.*

Lema 2.4. *Sea $i \in X$ y $\pi \in \Pi$, y sea $(\Omega, \mathcal{A}, P_{i,\pi})$ el espacio canónico correspondiente fijo. Sea Y una variable aleatoria discreta no negativa asociada a $(\Omega, \mathcal{A}, P_{i,\pi})$ tal que $E_{i,\pi}[Y]$ existe. Supongamos que se cumple el Supuesto 2.2. Tome $i \in X$ y $\pi \in \Pi$, y sea $(\Omega, \mathcal{A}, P_{i,\pi})$ el espacio canónico correspondiente fijo. Entonces,*

a) Para cada $T \geq 0$,

$$\sum_{t=0}^{T}{}^{*}\, \widetilde{R}\,(x_t, a_t) := \sum_{t=0}^{T} R\,(x_t, a_t)\,(B,C,D,F), \tag{2.21}$$

es una variable aleatoria difusa y

$$E^{*}_{i,\pi}\left[\sum_{t=0}^{T}{}^{*}\, \widetilde{R}\,(x_t, a_t)\right] = E_{i,\pi}\left[\sum_{t=0}^{T} R\,(x_t, a_t)\right](B,C,D,F). \tag{2.22}$$

b) Definamos:

$$H_{finite} = \left\{\omega \in \Omega \mid \sum_{t=0}^{\infty} R\,(x_t, a_t)\,(\omega) < +\infty\right\}$$

y

$$H_{\infty} = \left\{\omega \in \Omega \mid \sum_{t=0}^{\infty} R\,(x_t, a_t)\,(\omega) = +\infty\right\}.$$

Entonces,

$$\sum_{t=0}^{\infty}{}^{*}\widetilde{R}\left(x_t,a_t\right)(\omega)=\begin{cases}\sum_{t=0}^{\infty}R\left(x_t,a_t\right)(\omega)(B,C,D,F), & \omega\in H\\ \widetilde{0}, & \omega\in H_{\infty}\end{cases} \tag{2.23}$$

es una variable aleatoria difusa, y

$$E_{i,\pi}^{*}\left[\sum_{t=0}^{\infty}{}^{*}\widetilde{R}\left(x_t,a_t\right)\right]=E_{i,\pi}\left[\sum_{t=0}^{\infty}R\left(x_t,a_t\right)\right](B,C,D,F). \tag{2.24}$$

Demostración. a) Observe que para cada $T\geq 0$, $\sum_{t=0}^{T}R\left(x_t,a_t\right)$ es una variable aleatoria discreta no negativa (recuerde que X y A son conjuntos finitos). En consecuencia, la parte a) se sigue del Lema 1.7, con $Y=\sum_{t=0}^{T}R\left(x_t,a_t\right)$.

b) Consideremos para cada $T\geq 0$, $S_T:=\sum_{t=0}^{T}R\left(x_t,a_t\right)$, con rango finito dado por

$$S_T[\Omega]=\{y_1^T,y_2^T,...,y_{k_T}^T\},$$

y considere los conjuntos medibles $[S_T=y_j^T]:=\{\omega\in\Omega\mid Y(\omega)=y_j^T\}$, $j=1,2,...,k^T$. Sean $S=\sum_{t=0}^{\infty}R\left(x_t,a_t\right)$ y $\widetilde{S}=\sum_{t=0}^{*\infty}\widetilde{R}\left(x_t,a_t\right)$.

Notemos que por el Supuesto 2.2, $0\leq E_{i,\pi}[S]<\infty$ lo que implica que S es finito a.s. $[P_i^{\pi}]$ (ver Ejercicio 4Q, p. 39 en [4]), es decir, el conjunto medible H_{∞} satisface que $P_i^{\pi}(H_{\infty})=0$. Ahora, del Lema 1.2 en [4] se sigue que

$$\mathring{S}(\omega)=\begin{cases}S(\omega), & \omega\in H\\ 0, & \omega\in H_{\infty},\end{cases} \tag{2.25}$$

es medible, y con esto se tiene que $\widetilde{S}(\omega)=\mathring{S}(\omega)(B,C,D,F)$, $\omega\in\Omega$, que es

$$\widetilde{S}(\omega)=\begin{cases}S(\omega)(B,C,D,F), & \omega\in H_{finite}\\ \tilde{0}, & \omega\in H_{\infty}.\end{cases} \tag{2.26}$$

Observe que, para $\omega\in H$, es decir, si $S(\omega)<\infty$ resulta que

$$\lim_{t\longrightarrow\infty}R(x_t(\omega),a_t(\omega))=0,$$

y recordando que X y A son conjuntos finitos, se sigue que existe un entero positivo $\tau=\tau(\omega)$ tal que

$$R(x_t(\omega),a_t(\omega))=0,$$

para todo $t > \tau$, o

$$S(\omega) = S_\tau(\omega),$$

y en este caso también se cumple que:

$$\widetilde{S}(\omega) = S_\tau(\omega). \tag{2.27}$$

Ahora, tengamos en cuenta la multifunción dada por

$$\tilde{S}_\alpha(\omega) := (\tilde{S}(\omega))_\alpha = (\mathring{S}(\omega)\Theta)_\alpha = S(\omega)[q(\alpha), s(\alpha)], \tag{2.28}$$

$\omega \in \Omega$. (Recordar que $\Theta = (B, C, D, F)$ con α-cortes $\Theta_\alpha = [q(\alpha), s(\alpha)]$, $\alpha \in [0, 1]$.)

$$\begin{aligned} Gr(\tilde{S}_\alpha) &= \left\{(\omega, x) \in \Omega \times \mathbb{R} \mid x \in \mathring{S}(\omega)[q(\alpha), s(\alpha)]\right\} \\ &= \left[\bigcup_{T=0}^{+\infty} \bigcup_{j=1}^{k_T} [S_T = y_j^T] \times y_j^T[q(\alpha), s(\alpha)]\right] \cup [H_\infty \times \{0\}]. \end{aligned} \tag{2.29}$$

Por lo tanto, $Gr(\tilde{S}_\alpha) \in \mathcal{A} \otimes \mathcal{B}(\mathbb{R})$. Dado que α es arbitrario, de la Definición 1.10 resulta que $\tilde{S}$ es una variable aleatoria difusa. Y similar a la prueba de la parte a) de este lema se sigue que (2.24) se cumple.

□

Lema 2.5. *Supongamos que el Supuesto 2.2 se cumple. Tomemos $i \in X$ Y $\pi \in \Pi$ y sea $(\Omega, \mathcal{A}, P_{i.\pi})$ el espacio canónico correspondiente fijo. Entonces, para cada $T \geq 0$,*

$$\widetilde{S}_T = \sum_{t=0}^{T} {}^*\widetilde{R}(x_t, a_t) = \sum_{t=0}^{T} R(x_t, a_t)(B, C, D, F), \tag{2.30}$$

y

$$E^*_{i,\pi}\left[\sum_{t=0}^{T} {}^*\widetilde{R}(x_t, a_t)\right] = E_{i,\pi}\left[\sum_{t=0}^{T} R(x_t, a_t)\right](B, C, D, F), \tag{2.31}$$

Además,

$$\widetilde{S} = \sum_{t=0}^{\infty} {}^*\widetilde{R}(x_t, a_t) = \sum_{t=0}^{\infty} R(x_t, a_t)(B, C, D, F), \tag{2.32}$$

y

$$E^*_{i,\pi}\left[\sum_{t=0}^{\infty}{}^*\widetilde{R}(x_t,a_t)\right]=\begin{cases}E_{i,\pi}\left[\sum_{t=0}^{\infty}R(x_t,a_t)\right](B,C,D,F),\\ \widetilde{0}\end{cases}$$

Observación 2.3. *a) El caso (degenerado) en el que en el modelo de decisión (2.19) $\widetilde{R}(i,a)$ tiene una función de pertenencia dada por:*

$$(\widetilde{R}(i,a))'(x)=\begin{cases}1 & si\ x=R(i,a)\\ 0 & si\ x\neq R(i,a),\end{cases} \tag{2.33}$$

para todo $i\in X$ y $a\in A(i)$ implica que $\widetilde{R}$ es una variable aleatoria difusa y

$$E^*_{i,\pi}[\widetilde{R}]=E_{i,\pi}[R]\widetilde{1}, \tag{2.34}$$

para todo $i\in X$ y $\pi\in\Pi$.

b) Tenga en cuenta que el Lema 2.4 es válido para todas las variables aleatorias difusas y sus esperanzas.

2.3.1. Problema de control óptimo para el modelo difuso

Definición 2.9. *Para cada $i\in X$ y $\pi\in\Pi$, la esperanza difusa correspondiente viene dada por:*

$$\widetilde{V}(i,\pi):=E^*_{i,\pi}\left[\sum_{t=0}^{\infty}{}^*\,\widetilde{R}\left(x_t,a_t\right)\right]=E_{i,\pi}\left[\sum_{t=0}^{\infty}R\left(x_t,a_t\right)\right](B,C,D,F). \tag{2.35}$$

Ahora, sea $i\in X$ y $\pi\in\Pi$, y $T\geq 0$:

$$\tilde{V}_T(i,\pi):=\sum_{t=0}^{T}{}^*\,E^*_{i,\pi}\left[\widetilde{R}\left(x_t,a_t\right)\right], \tag{2.36}$$

V^*_T se conoce como recompensa total esperada difusa de T etapas.

Observación 2.4. *Nótese que la recompensa total esperada difusa de T etapas (ver (2.36)) es un número difuso trapezoidal, específicamente,*

$$\tilde{V}_T(i,\pi)=\left(BV_T(i,\pi),CV_T(i,\pi),DV_T(i,\pi),FV_T(i,\pi)\right), \tag{2.37}$$

para $\pi\in\Pi$ y $i\in X$, donde $\tilde{V}_T$ es la recompensa total nítida de la etapa T.

Lema 2.6. *Supongamos que se cumple el Supuesto 2.2. Entonces, para cada $i\in X$ y $\pi\in\Pi$, $\{\widetilde{V}_T(i,\pi)\}$ converge y*

$$\begin{aligned}\widetilde{V}(i,\pi) := lim^{*}_{T\to\infty}\widetilde{V}_T(i,\pi) &= \sum_{t=0}^{\infty}{}^{*} E^{*}_{i,\pi}\left[\widetilde{R}\left(x_t,a_t\right)\right] \\ &= (BV(i,\pi), CV(i,\pi), DV(i,\pi), FV(i,\pi)), \end{aligned} \quad (2.38)$$

donde

$$V(i,\pi) = \sum_{t=0}^{\infty} E_{i,\pi}\left[R\left(x_t,a_t\right)\right] \in \mathbb{R}.$$

Demostración. Sean $\pi \in \Pi$ y $x \in X$ fijos. Para simplificar la notación en esta prueba se denotará $V = V(\pi, x)$ and $V_T = V_T(\pi, x)$. Entonces, los $\alpha-$cortes de (2.36), están dados por

$$\begin{aligned}\Delta^T :&= (BV_T, CV_T, DV_T, FV_T)_\alpha \\ &= [B(1-\alpha)V_T + \alpha CV_T, F(1-\alpha)V_T + \alpha DV_T].\end{aligned}$$

Analogamente,

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

Por lo tanto, por (1.6), se obtiene que

$$\hat{d}(\Delta^T, \Delta) = sup_{\alpha\in[0,1]} d(\Delta^T, \Delta).$$

Ahora, debido a la identidad máx$(c, b) = (c+b+|b-c|)/2$ con $b, c \in \mathbb{R}$, resulta que

$$d(\Delta^T, \Delta) = (1-\alpha)D(v - v_T) + \alpha C(v - v_T).$$

Entonces,

$$\begin{aligned}\hat{d}(\Delta_T, \Delta) &= sup_{\alpha\in[0,1]}(v - v_T)(D - \alpha(D - C)) \\ &= (v - v_T)D.\end{aligned} \quad (2.39)$$

Por tanto, cuando T tiende a infinito en (2.39), se concluye que

$$\begin{aligned}\lim_{T\longrightarrow\infty} \rho(\tilde{v}_T, \tilde{v}) &= \lim_{T\longrightarrow\infty}(v - v_T)D \\ &= 0.\end{aligned}$$

□

Ahora, el *problema de control óptimo difuso* es el siguiente: determine $\pi_o \in \Pi$ (si existe) tal que:

$$\widetilde{V}(i,\pi) \leq^{*} \widetilde{V}(i,\pi_o), \quad (2.40)$$

para todo $i \in X$ y $\pi \in \Pi$. En este caso es posible escribir

$$\widetilde{V}(i, \pi_o) = sup^*_{\pi \in \Pi} \widetilde{V}(i, \pi), \tag{2.41}$$

$i \in X$ y se dice que π_o es *óptimo*. Además, la función $\widetilde{V_o}(i) = \widetilde{V}(i, \pi_o)$ para $i \in X$ se llamará *función de valor óptimo difusa*.

Observación 2.5. *Usando las Observaciones 1.1 y 2.3 a) es directo ver que en el caso (degenerado) cuando en el modelo de decisión (2.19) $\widetilde{R}(i, a)$ tiene una función de pertenencia dada por:*

$$(\widetilde{R}(i, a))'(x) = \begin{cases} 1 & si\ x = R(i, a) \\ 0 & si\ x \neq R(i, a), \end{cases} \tag{2.42}$$

para todo $i \in X$ y $a \in A(i)$, entonces el problema de control óptimo difuso descrito en (2.40) y (2.41) se reduce al problema de control óptimo descrito en (2.1).

Lema 2.7. *Supongamos que se cumple el Supuesto 2.2. Entonces, para cada $i \in X$, $\widetilde{V}_o(i)$ es una función acotada, es decir, existe $\widetilde{K} \in \mathcal{F}(\mathbb{R})$ tal que $\widetilde{V}_o(i) \leq^* \widetilde{K}$, $i \in X$.*

Demostración. Tomemos $\pi \in \Pi$ y $i \in X$ fijos. Entonces, como consecuencia de (2.38), el $\alpha-$corte de $\widetilde{V}(i, \pi)$ está dado por

$$\widetilde{V}(i, \pi)_\alpha = [B\widetilde{V}(i, \pi) + \alpha\widetilde{V}(i, \pi)(C - B), F\widetilde{V}(i, \pi) - \alpha\widetilde{V}(i, \pi)(F - D)].$$

Notemos que ya que X es finito, podemos encontrar un $K > 0$ tal que $V(\pi, i) \leq K$. (Observe que debido a que X es finito, es posible tomar K para obtener $V(\pi, i) \leq K$, para todo $i \in X$.) En consecuencia, observe que

$$B\widetilde{V}(i, \pi) + \alpha\widetilde{V}(i, \pi)(C - B) \leq BK + \alpha(C - B)K,$$

y

$$\begin{aligned} F\widetilde{V}(i, \pi) - \alpha\widetilde{V}(i, \pi)(F - D) &\leq FK(1 - \alpha) + \alpha DK \\ &= FK - \alpha(F - D)K. \end{aligned}$$

En consecuencia, $\widetilde{V}(i, \pi) \leqslant^* \widetilde{K} := (BK, CK, DK, FK)$. Por lo tanto, $\widetilde{V}_o(i) \leqslant^* \widetilde{K}$ (ver (2.41)). Como i y π son arbitrarios, el resultado es el siguiente. □

Teorema 2.3. *Bajo el Supuesto 2.2 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 óptima está dada por

$$\widetilde{V}(i) = (BV(i), CV(i), DV(i)), i \in X. \tag{2.43}$$

Demostración. a) Sean $\pi \in \Pi$ y $i \in X$ fijos. Primero observemos que (2.35) es equivalente a

$$\tilde{V}(i,\pi) := (BV(i,\pi), CV(i,\pi), DV(i,\pi), FV(i,\pi)),$$

como consecuencia del supuesto 2.2. Entonces, el α-corte de $\tilde{V}(\pi, x)$ viene dado por

$$\tilde{V}(\pi,x)_\alpha = [(C-B)V(i,\pi)\alpha + BV(i,\pi), FV(i,\pi) - (F-C)\alpha V(i,\pi)].$$

Ahora, por el Teorema 2.1, existe $f_o \in \mathbb{F}$ tal que

$$(C-B)V(i,\pi)\alpha + BV(i,\pi) \leq (C-B)V(i,f_o(i))\alpha + BV(i,f_o(i)).$$

y

$$FV(i,\pi)\alpha - (F-C)\alpha V(i,\pi) \leq FV(i,f_o(i))\alpha - (F-C)V(i,f_o(i)).$$

Dado que $i \in X$ y $\pi \in \Pi$ son arbitrarios, el resultado se cumple debido a (2.41).

b) Por la parte a) de este Lema, se sigue que

$$\tilde{V}(x) = (BV(i,f_o(i)), CV(i,f_o(i)), DV(i,f_o(i)), FV(i,f_o(i))),$$

para cada $i \in X$, aplicando así el Teorema 2.1, se concluye que

$$\tilde{V}(i) = (BV(i), CV(i), DV(i), FV(i)), i \in X.$$

□

2.4. PDMs con recompensa total descontada (caso nítido)

En esta sección, se consideran los Procesos de decisión de Markov con recompensa descontada total en tiempo discreto con espacios de estados finitos, conjuntos de acción compactos tanto en el caso de horizonte finito e infinito.

Definición 2.10. *$(X, A, \{A(x) : x \in X\}, Q, R)$, un modelo de Markov, entonces la recompenza descontada total esperada se define como sigue:*

$$v(\pi, x) := E_{x,\pi}\left[\sum_{t=0}^{\infty} \beta^t R(X_t, a_t)\right], \tag{2.44}$$

$\pi \in \Pi, x \in X$, donde $\beta \in (0,1)$ es un factor de descuento dado. Además, la recompensa descontada total esperada con un horizonte finito es definida de la forma siguiente:

$$v_T(\pi, x) := E_{x,\pi}\left[\sum_{t=0}^{T-1} \beta^t R(X_t, a_t)\right], \tag{2.45}$$

para cada $x \in X$ y $\pi \in \Pi$ donde T es un entero positivo.

La función de valor óptimo está definida como

$$V(x) := sup_{\pi\in\Pi} v(\pi, x), \tag{2.46}$$

$x \in X$.

2.4.1. Problema de control óptimo para el modelo

El problema de control óptimo es encontrar una política $\pi^* \in \Pi$ tal que

$$v(\pi^*, x) = V(x), \tag{2.47}$$

$x \in X$, en tal caso, π^* es llamada la política óptima. Definiciones similares pueden ser establecidas análogamente para v_T. En este caso, V_T denota la función de valor óptimo para el problema de control óptimo con un horizonte finito.

Supuesto 2.3. *a) Para cada $x \in X$, $A(x)$ es un conjunto compacto en $\mathcal{B}$, donde $\mathcal{B}$ es la $\sigma-$álgebra de Borel del espacio A.*

b) La función de Recompensa R es una función acotada y no$-$negativa.

c) Para cada $x, y \in X$. los mapeos $a \longmapsto R(x,a)$ y $a \longmapsto Q(\{y\}\}x, a)$ son continuas en $a \in A(x)$

La prueba del siguiente teorema que proporciona el teorema de Programación Dinámica puede ser consultado en [16] y [24].

Teorema 2.4. *Bajo el supuesto 2.3, las siguientes afirmaciones se cumplen:*

a) Definamos $W_T(x) = 0$ y para $n = T-1, ..., 1, 0$, consideremos

$$W_n(x) := max_{a\in A(x)}\{R(x,a) + \beta E[W_{n+1}(F(x,A,\xi))]\}. \tag{2.48}$$

$x \in X$. Entonces para cada $n = 0, 1, ..., T-1$, existe una $f_n \in \mathbb{F}$ tal que

$$W_n(x) = R(x,a) + \beta E[W_{n+1}(F(x, f_n(x), \xi))], \tag{2.49}$$

$x \in X$. En este caso, $\pi^* = \{f_0, ..., f_{T-1}\} \in \mathbb{M}$ es la política óptima y $V_T(x) = v_T(\pi^*, x) = W_0(x), x \in X$.

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

$$V(x) = max_{a\in A(x)}\{R(x,a) + \beta E[V(F(x,a,\xi))]\}, \tag{2.50}$$

$x \in X$.

c) Existe una política $f^* \in \mathbb{F}$ tal que el control $f^*(x) \in A(x)$ que alcanza el máximo en (2.54) es decir, para todo $x \in X$,

$$V(x) = R(x, f^*(x)) + \beta E[V(F(x, f^*(x), \xi))]. \tag{2.51}$$

d) Definamos la función de iteración de valor como sigue:

$$V_n(x) = min_{a\in A(x)}\{C(x,a) + \beta E[V_{n-1}(F(x, f^*(x), \xi))]\}, \tag{2.52}$$

para todo $x \in X$ y $n = 1, 2, ...$, con $V_0(.) = 0$. Entonces la secuencia de puntos $\{V_n\}$ de funciones de iteracion de valor converge puntualmente a la función de valor óptimo V, es decir,

$$\lim_{n\to\infty} V_n(x) = V(x)$$

$x \in X$.

Observación 2.6. *Como consecuencia del Teorema 2.4, los siguientes hechos se mantienen:*

a) Por la parte a) del Teorema 2.4, en el caso de recompensa esperada descontada con un horioznte finito, el óptimo es alcanzado en una política Markoviana, por lo tanto,

$$sup_{\pi\in\Pi} v_T(\pi, x) = sup_{\pi\in\Pi}\{v_T(\pi, x)\}, \tag{2.53}$$

$x \in X$.

b) Por la parte c) del Teorema 2.4, en el caso de recompensa esperada descontada con un horizonte infinito, el óptimo es alcanzado en una política óptima estacionaria. Entonces se sigue que:

$$sup_{\pi\in\Pi} v_T(\pi, x) = sup_{f\in\mathbb{F}} v_T(f, x)\}, \quad x \in X. \tag{2.54}$$

2.5. PDMs descontado con recompensa difusa

En esta sección presentamos los procesos de decisión de Markov descontados en tiempo discreto con espacios de estados finitos, conjuntos de acción compactos de horizontes finitos e infinitos y recompensa difusa de tipo trapezoidal bajo el criterio de recompenza difusa descontada total esperada. El problema de control óptimo correspondiente se establece con respecto al orden máximo difuso. La solución óptima difusa está relacionada a un PDM con descuento conveniente con una recompensa no difusa. En el Capítulo 4 se proporcionan aplicaciones de la teoría desarrollada en un modelo de horizonte finito de un sistema de inventario en el que se utiliza un algoritmo para calcular la solución óptima, y, adicionalmente para el caso de horizonte infinito, un PDM y un competitivo PDM (también conocido como juego estocástico) se suministran en un contexto económico y financiero.

Consideremos un modelo de decisión de Markov difuso como en (2.19), donde los primeros cuatro componentes son los mismos que en el modelo dado en (2.1). La componente $\tilde{R}$, corresponde a una función de recompensa difusa en $\mathbb{K}$.

2.5.1. Criterio de recompenza difusa descontada total esperada

Para cada política $\pi \in \mathbb{M}$ y estado $x \in X$, sea

$$\tilde{v}(i, \pi) = \sum_{t=0}^{T-1} \beta^t E^*_{i,\pi}\left[\tilde{R}(x_t, a_t)\right], \tag{2.55}$$

donde T es un entero positivo y $E^*_{i,\pi}$ es la esperanza con respecto a $\tilde{P}^{\pi}_{x}$ la cual está definida por la expresión (1.7). La expresión dada en (2.55) se denomina recompensa difusa descontada total esperada con un horizonte finito.

$$\tilde{V}(i,\pi) = \sum_{t=0}^{\infty} \beta^t E^*_{i,\pi}\left[\tilde{R}(x_t, a_t)\right], \tag{2.56}$$

y, la esperanza en (2.56) está definida en (1.7), cuando $\{a_t\}$ es inducida por una política estacionaria π.

De esta forma, el problema de control de interés es la maximización de la recompensa difusa total descontada esperada en un horizonte finito o infinito (ver (2.55) y (2.56), respectivamente). Se considera la siguiente suposición para la función de recompensa del modelo difuso.

Supuesto 2.4. *Sea $\gamma_1, \gamma_2, \gamma_3$ y γ_4 números reales tales que $0 < \gamma_1 < \gamma_2 \leq \gamma_3 < \gamma_4$. Supondremos que la recompensa difusa es un número difuso trapezoidal (ver la Definición 1.4), específicamente*

$$\tilde{R}(x,a) = R(x,a)(\gamma_1, \gamma_2, \gamma_3, \gamma_4) \tag{2.57}$$

para cada $(x,a) \in \mathbb{K}$, donde $R : \mathbb{K} \longrightarrow \mathbb{R}$ es la función de recompensa del modelo 2.19.

Observación 2.7. *Observemos que, bajo el Supuesto 2.4 y la parte b) del Lema 1.2, la recompensa difusa (2.55) es un número difuso trapezoidal.*

2.5.2. Problema de control óptimo para el modelo

En esta sección, se presentarán los resultados de la convergencia de la recompensa difusa (2.55) a la recompensa difusa descontada total esperada en el horizonte infinito (2.56), cuando T tiende al infinito. Posteriormente se verificará la existencia de políticas óptimas y la validez de la programación dinámica.

Lema 2.8. *Supongamos que (2.4) se cumple. Entonces, para cada $i \in X$, $\pi \in \mathbb{F}$ (ver Observación 2.6), $\{\widetilde{V}_T(\pi, i) : T = 0, 1, ...\}$ converge y*

$$\widetilde{v}(i,\pi) = lim_{T\to\infty}\widetilde{v}_T(\pi,i) = \sum_{t=0}^{\infty} E^*_{i,\pi}\left[\widetilde{R}\left(x_t, a_t\right)\right] = v(i,\pi)(\gamma_1,\gamma_2,\gamma_3,\gamma_4),$$

donde $v(i,\pi) = \sum_{t=0}^{\infty} E_{i,\pi}\left[R\left(x_t, a_t\right)\right] \in \mathbb{R}$.

Demostración. Sean $\pi \in \Pi$ y $x \in X$ fijos. Para simplificar la notación en esta demostración, denotaremos por $v = v(\pi, x)$ y $v_T = v_T(\pi, x)$ (ver (2.55) y (2.56)). Entonces, el α-corte de (2.55) está dado por

$$\begin{aligned}\Delta^T :&= (\gamma_1 v_T, \gamma_2 v_T, \gamma_3 v_T, \gamma_4 v_T)_\alpha \\ &= [\gamma_1(1-\alpha)v_T + \alpha\gamma_2 v_T, \gamma_4(1-\alpha)v_T + \alpha\gamma_3 v_T].\end{aligned}$$

Análagamente,

$$\begin{aligned}\Delta :&= (\gamma_1 v, \gamma_2 v, \gamma_3 v, \gamma_4 v)_\alpha \\ &= [\gamma_1(1-\alpha)v + \alpha\gamma_2 v, \gamma_4(1-\alpha)v + \alpha\gamma_3 v].\end{aligned}$$

Por lo tanto, por (1.6), se obtiene que

$$\hat{d}(\Delta^T, \Delta) = sup_{\alpha\in[0,1]} d(\Delta^T, \Delta).$$

Ahora, debido a la identidad $\max(c, b) = (c + b + |b - c|)/2$ con $b, c \in \mathbb{R}$, se tiene como resultado que

$$d(\Delta^T, \Delta) = (1-\alpha)\gamma_3(v - v_T) + \alpha\gamma_2(v - v_T).$$

Entonces,

$$\begin{aligned}\hat{d}(\Delta_T, \Delta) &= sup_{\alpha\in[0,1]}(v - v_T)(\gamma_3 - \alpha(\gamma_3 - \gamma_2)) \\ &= (v - v_T)D.\end{aligned} \tag{2.58}$$

Por lo tanto, donde T tiende a infinito en (2.58), y concluimos que

$$\begin{aligned}\lim_{T\longrightarrow\infty} \rho(\tilde{v}_T, \tilde{v}) &= \lim_{T\longrightarrow\infty}(v - v_T)\gamma_3 \\ &= 0.\end{aligned}$$

La segunda ecuación es una consecuencia de (2.44) y (2.45). □

Definición 2.11. *El problema de control óptimo difuso con horizonte infinito consiste en determinar una política $\pi^* \in \mathbb{F}$ tal que*

$$\tilde{v}(\pi, x) \leq^* \tilde{v}(\pi^*, x),$$

para toda $\pi \in \mathbb{F}$ y $x \in X$. En consequencia (ver Observación 2.6 (b)),

$$\tilde{v}(\pi^*, x) = sup_{\pi\in\mathbb{F}} \tilde{v}(\pi, x),$$

para todo $x \in X$ (ver Observación 1.1). En este caso, la función difusa de valor óptimo es definida de la siguiente forma:

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

$x \in X$ y π^* es llamada la política óptima para el problema de control óptimo difuso.

Observación 2.8. *Definiciones similares pueden ser establecidas para $\tilde{v}_T$, la recompensa difusa descontada total esperada con un horizonte finito T. En este caso, el valor difuso óptimo es denotado por $\tilde{V}_T$, y (Ver Observación 2.6(a)),*

$$\tilde{V}_T(x) = \tilde{v_T}(\pi^*, x) = sup_{\pi\in\mathbb{M}}\tilde{v_T}(\pi, x),$$

para toda $x \in X$, por supuesto, si tal π^* existe, entonces esta es llamada la política óptima para el problema de control óptimo difuso con un horizonte T.

Una consecuencia directa de la Definición 2.11, Observación 2.8, y el Teorema 2.8 es el próximo resultado.

Teorema 2.5. *Bajo los Supuestos 2.3 y 2.4, las siguientes afirmaciones se mantienen:*

a) La política óptima π^* del problema de control óptimo finito nítido (ver (2.55)) es la política óptima para $\tilde{v}_T$, es decir. $\tilde{v}_T(\pi^*, x) = sup_{\pi\in\mathbb{M}}\tilde{v}_T(\pi, x)$ para todo $\pi \in \Pi$ y $x \in X$.

b) la función de valor difusa óptima está dada por

$$\tilde{V}_T(x) = V_T(x)(\gamma_1, \gamma_2, \gamma_3, \gamma_4), \tag{2.59}$$

$x \in X$, donde $\tilde{V}_T(x) = sup_{\pi\in\mathbb{M}}\tilde{v}_T(\pi, x)$, $x \in X$.

Demostración. a) Sean $\pi \in \mathbb{M}$ y $x \in X$ fijos. Entonces, por (2.55), se obtiene que

$$\tilde{v}_t(\pi, x) = v_T(\pi, x)(\gamma_1, \gamma_2, \gamma_3, \gamma_4),$$

donde el Supuesto 2.4 y el Lema 1.2 fueron aplicados. Ahora, observemos que los $\alpha-$corte de $\tilde{v}_T(\pi, x)$ están dados por los siguientes intervalos cerrados:

$$\tilde{v}_t(\pi, x)_\alpha = [\gamma_1 v_T(\pi, x)+\alpha(\gamma_2-\gamma_1)v_T(\pi, x), \gamma_4 v_T(\pi, x)-\alpha v_T(\pi, x)(\gamma_4-\gamma_3)].$$

Por otro lado, por el Teorema 2.4, existe una política óptima $\pi^* \in \mathbb{M}$ tal que, $v_T(\pi, x) \leq v_T(\pi^*, x)$. Entonces, notemos que los extremos de $\tilde{v}_t(\pi, x)_\alpha$ satisfacen las siguientes inecuaciones:

$$\gamma_1 v_T(\pi, x) + \alpha(\gamma_2 - \gamma_1)v_T(\pi, x) \leq \gamma_1 v_T(\pi^*, x) + \alpha(\gamma_2 - \gamma_1)v_T(\pi^*, x)$$
$$\gamma_4 v_T(\pi, x) - \alpha v_T(\pi^*, x)(\gamma_4 - \gamma_3) \leq \gamma_4 v_T(\pi^*, x) - \alpha v_T(\pi^*, x)(\gamma_4 - \gamma_3).$$

En consecuencia, $\tilde{v}_T(\pi, x) \leq^* \tilde{v}_T(\pi^*, x)$. Ya que $x \in X$ y $\pi \in \Pi$ son arbitrarios, el resultado sigue, debido a la Definición 2.11.

b) Por el Teorema 2.6 a), se sigue que

$$\tilde{V}_T(x) = v(\pi^*, x)(\gamma_1, \gamma_2, \gamma_3, \gamma_4),$$

para cada $x \in X$, de esta manera, aplicando el Teorema 2.4, se concluye que

$$\tilde{V}_T(x) = V_T(x)(\gamma_1, \gamma_2, \gamma_3, \gamma_4),$$

$x \in X$.

□

La prueba del Teorema 2.6 es similar a la prueba del Teorema 2.5, por eso se omite.

Teorema 2.6. *Bajo los Supuesto 2.3 y 2.4, la siguiente afirmación se cumple:*

a) La política óptima del problema de control difuso es la misma que la política óptima del problema de control óptimo nítido.

b) La función de valor difuso óptimo está dada por

$$\tilde{V}(x) = V(x)(\gamma_1, \gamma_2, \gamma_3, \gamma_4), x \in X. \tag{2.60}$$

En el Capítulo 4, los Teoremas 2.5 y 2.6 se ilustrarán en varios ejemplos.

Capítulo 3

Aplicaciones de PDMs con recompensa total difusa

Este capítulo se refiere a los procesos de decisión de Markov (PDM) en los que tanto el estado como los espacios de decisión son finitos y la función objetivo es la recompensa total esperada. Para este tipo de PDM, asumimos que la función de recompensa es de tipo difuso. Específicamente, esta función de recompensa difusa tiene una forma trapezoidal adecuada que es una función de una recompensa estándar no difusa. Además, esta recompensa difusa se aproxima, en un sentido difuso, a la recompensa no difusa correspondiente. El problema de control difuso consiste en determinar una política de control que maximice la recompensa total esperada difusa, donde la maximización se realiza con respecto al orden parcial en los α-cortes de números difusos. La política óptima y la función de valor óptimo para el problema de control óptimo difuso se caracterizan mediante una versión de la ecuación de programación dinámica y, como principales conclusiones, se obtiene que la política óptima del problema estándar y el difuso coinciden y la función de valor óptimo difuso tiene una forma trapezoidal conveniente. Como ilustraciones, se presentan extensiones difusas de un problema de parada óptima y de un modelo de juego red-black.

3.1. Un problema de paro óptima

Aquí proporcionamos un ejemplo de un problema de paro óptimo visto como un PDM de recompensa total, el cual es una versión similar del Ejemplo 7.2.6 en [24] y su extensión al entorno difuso.

Consideremos el problema de determinar una política de paro óptimo para la cadena de Markov finita, en el que el sistema se mueve entre los estados $X = \{i_1, i_2, i_3, i_4\}$, y matriz de transición:

$$P = \begin{bmatrix} 0 & 1/3 & 2/3 & 0 \\ 4/5 & 1/5 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{3.1}$$

Cada entrada de P describe la probabilidad de transición p_{ij}, para $i, j \in X^{'}$.

Supongamos que en cada época de decisión, el controlador tiene dos acciones admisibles: parar (Q) o continuar (C).

Si además, para cada $i \in X^{'}$, $A(i) = \{C, Q\}$. Si en el estado i elegimos el control C, el sistema se mueve al estado $j \in X^{'}$ con probabilidad p_{ij}, y si elegimos Q, el sistema se mueve al estado δ, en el que no recibimos recompensa. Observemos que $X = X' \cup \delta$ y $A(\delta) = \{C\}$, por lo que δ es un estado recurrente.

Observe que $X = \{i_1, i_2, i_3, i_4, \delta\}$ y $A(\delta) = \{C\}$. En particular, supongamos que $R(i_1, Q) = g(i_1) = 8$, $R(i_2, Q) = g(i_2) = 5$, $R(i_3, Q) = g(i_3) = 3$, $R(i_4, Q) = g(i_4) = 0$, y $R(\delta, C) = 0$.

El objetivo consiste en determinar una política que maximice la recompensa total esperada, bajo el supuesto de que las recompensas se reciben solo al final. Entonces, debido al Teorema 7.2.3 (a) en [24], la función de valor óptimo V_0 es la solución mínima en la clase funciones $V^{+} : \{V : X \longrightarrow \mathbb{R} : V \geq 0 \quad y \quad V(x) < \infty \quad para \quad cada \quad s \in S\}$, la función de valor óptimo V_o es la solución mínima $w : X^{'} \longrightarrow \mathbb{R}$ con $w \geq 0$ de la siguiente ecuación de programación dinámica:

$$w(i) = max\{g(i), \sum_{j \in X'} w(j) p_{ij}\}. \tag{3.2}$$

$i \in X^{'}$ con $V_o(\delta) = 0$. Entonces, aplicando el enfoque de programación lineal, (3.2) es equivalente al siguiente programa lineal:

$$MINIMIZAR : w(i_1) + w(i_2) + w(i_3) + w(i_4) \tag{3.3}$$

sujeto a

$$w(i) \geq g(i), \tag{3.4}$$

$$w(i) \geq \sum_{j \in X'} w(j) p_{ij}, \tag{3.5}$$

$i \in X^{'}$. Entonces, las desigualdades (3.4) y (3.5) son equivalentes a

$$\begin{aligned} 3w(i_1) - w(i_2) - 2w(i_3) &\geq 0, \\ w(i_2) - V(i_1) &\geq 0, \\ 2w(i_3) - w(i_1) - w(i_4) &\geq 0, \\ w(i_1) \geq 8, \quad w(i_2) \geq 5, \quad w(i_3) \geq 3, \quad w(i_4) &\geq 0. \end{aligned}$$

Aplicando el algoritmo símplex se obtiene que el valor óptimo es $V_o(i_1) = 8$, $V_o(i_2) = 8$, $V_o(i_3) = 4$ y $V_o(i_4) = 0$.

Teorema 3.1. *Si el espacio de estados es finito,* $g(s) < \infty \quad \forall s \in S'$, *y* $g(s) \geq 0$ *si* s *es un estado, entonces el valor del problema de parada óptima* v^* *es la solución mínima no negativa* $v \geq g$ *que satisface (3.5). Además, la política estacionaria* $(d^*)^\infty$ *definida mediante*

$$d^*(i) = \begin{cases} Q & si \quad i \in \{i' \in X' : v'(i) = g(i)\} \\ \\ C & \qquad\qquad en\ otro\ caso. \end{cases} \tag{3.6}$$

es óptima.

En consecuencia, la política estacionaria óptima f_o viene dada por:

$$f_o(i) = \begin{cases} Q & si \quad i \in \{i_1, i_4\} \\ \\ C & si \quad i \in \{i_2, i_3\} \end{cases}. \tag{3.7}$$

Ahora, considere la siguiente función de recompensa difusa trapezoidal:

$$\widetilde{R}(i, Q) = R(i, Q)(0, \frac{9}{10}, \frac{11}{10}, 2),$$

$i \in X^{'}$ con la interpretación de que el número trapezoidal $(0, 0, 0, 0)$ es igual a $\widetilde{0}$, y $\widetilde{R}(\delta, C) = \widetilde{0}$.

Concretamente, para la decisión Q las recompensas difusas vienen dadas por:

- $\widetilde{R}(i_1, Q) = (0, 7{,}2, 8{,}8, 16)$,
- $\widetilde{R}(i_2, Q) = (0, 4{,}5, 5{,}5, 10)$,
- $\widetilde{R}(i_3, Q) = (0, 2{,}7, 3{,}3, 6)$,
- $\widetilde{R}(i_4, Q) = \widetilde{0}$.

Observación 3.1. *Tenga en cuenta que, por ejemplo,* $\widetilde{R}(i_1, Q) = (0, 7{,}2, 8{,}8, 16)$ *modela el hecho de que en el estado* i_1, *la recompensa recibida solo al finalizar está aproximadamente en el intervalo [7.2,8.8] en lugar de recibir la cantidad exacta de* $g(i_1) = 8$ *en el PDM estándar; el resto de las recompensas difusas tienen una interpretación similar.*

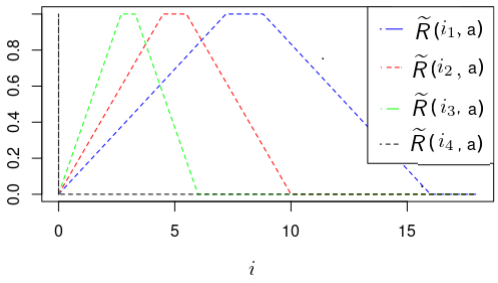

Figura 3.1: Recompensas difusas trapezoidales.

Ahora, la política óptima del problema de control difuso es la misma que la política óptima f_o del problema de control óptimo dada en (3.7), y la función de valor difuso óptimo $\widetilde{V_o}(i)$ es:

$$\widetilde{V_o}(i) = (0, \frac{9}{10}V_o(i), \frac{11}{10}V_o(i), 2V_o(i)),$$

$i \in X^{'}$. Y,

$$\widetilde{V_o}(\delta) = 0.$$

En consecuencia,

- $\widetilde{V_o}(i_1) = (0, 7{,}2, 8{,}8, 16)$,
- $\widetilde{V_o}(i_2) = (0, 7{,}2, 8{,}8, 16)$,
- $\widetilde{V_o}(i_3) = (0, 4{,}5, 4{,}4, 8)$,
- $\widetilde{V_o}(i_4) = \widetilde{0}$.

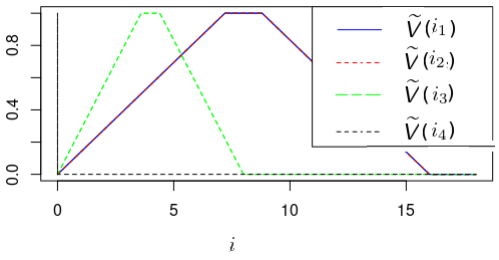

Figura 3.2: Función de valor óptimo difuso trapezoidal.

3.2. Modelo de apuesta

La primera parte de esta sección se basa en [28], pp. 73-83, y luego se proporciona la extensión aproximada.

Una persona que posee i dólares ingresa a un casino de juego que permite cualquier apuesta de la siguiente manera: si posee i dólares, entonces puede apostar cualquier número entero positivo menor o igual a i. Además, si apuesta j entonces

(a) gana j con probabilidad p o

(b) pierde j con probabilidad $1-p$.

La pregunta establecida en [28] es: ¿Qué estrategia de juego maximiza la probabilidad de que el individuo alcance una fortuna de N antes de quebrar? La respuesta a esta pregunta encaja en el marco de los PDMs con la recompensa total dada en la subsección anterior, donde el estado es la fortuna de los jugadores, ya que si se supone que se gana una recompensa terminal de 1 si alguna vez alcanzamos el estado N y todas las demás recompensas son cero, entonces la recompensa total esperada es igual a la probabilidad de alcanzar el estado N. En concreto, este modelo de juego se formula de la siguiente manera:

Descripción del modelo

(a) $X = \{0, 1, \ldots, N\}$, donde decimos que el estado es i cuando la fortuna actual es i.

(b) Sea $[k]$ la parte entera de k. Si la fortuna presente es i, entonces nunca valdría la pena apostar más de $N-i$, es decir,

$$A = \{0, 1, \ldots, [N/2]\},\ A(0) = \{0\},\ A(i) = \{1, 2, \ldots, \text{mín}\{i, N-i\}\},\ i \neq 0.$$

(c) $p_{ii+a}(a) = p,\ p_{ii-a}(a) = q = 1-p,\ p_{N0}(a) = 1,\ p_{00}(0) = 1.$

(d) $R(i,a) = 0,\ i \neq N,\ a \in A(i)$, y $R(N,0) = 1$.

Observación 3.2. *Sea G el conjunto de llegar alguna vez al estado N. Note que, para cada estrategia $\pi \in \Pi$ y $i \in X$, $V(i,\pi) = P_{i,\pi}[G]$ [28].*

Se define la *estrategia tímida* τ como aquella estrategia que siempre apuesta a 1, y se define la *estrategia audaz* β como la estrategia que, si la fortuna presente es i,

(a) apuestas i si $i \leq \frac{N}{2}$,

(b) apuestas $N - i$ si $i \geq \frac{N}{2}$.

De las Proposiciones 2.1 y el Corolario 2.6 en [28] se obtiene el siguiente lema.

Lema 3.1. *(a) Si $p \geq \frac{1}{2}$, entonces τ maximiza la probabilidad de alcanzar alguna vez una fortuna N, es decir, en este caso, $V_o(i) = V(i, \tau)$, para todo $i \in X$.*

(b) Si $p \leq \frac{1}{2}$, entonces β maximiza la probabilidad de alcanzar alguna vez una fortuna N, es decir, en este caso, $V_o(i) = V(i, \beta)$, para todo $i \in X$.

Ahora, se presentará el resultado sobre el modelo difuso red-black.

Teorema 3.2. *Supongamos que se cumple el Supuesto 2.2.*

(a) Si $p \geq \frac{1}{2}$, entonces $\widetilde{V}(i, \pi) \leq^* \widetilde{V}(i, \tau)$, para todo $\pi \in \Pi$ y $i \in X$. Por lo tanto τ es óptima y

$$\widetilde{V}(i, \tau) = (BV(i, \tau), CV(i, \tau), DV(i, \tau), FV(i, \tau)), \tag{3.8}$$

$i \in X$.

(b) Si $p \leq \frac{1}{2}$, entonces $\widetilde{V}(i, \pi) \leq^* \widetilde{V}(i, \beta)$, para todo $\pi \in \Pi$ y $i \in X$. Para todo, β es óptima y

$$\widetilde{V}(i, \beta) = (BV(i, \beta), CV(i, \beta), DV(i, \beta), FV(i, \beta)), \tag{3.9}$$

$i \in X$.

Observación 3.3. *Observe que en el modelo red-black no difuso, el objetivo del jugador es alcanzar al final del juego una cierta fortuna N. Ahora, siguiendo la descripción del modelo no borroso rojo-negro y el Supuesto 2.2 se obtiene que para el modelo borroso: $\widetilde{R}(i, a) = \widetilde{0}$, $i \neq N$, $a \in A(i)$, y $\widetilde{R}(N, 0) = (B, C, D, F)$; por lo tanto, tomando $C \leq N \leq D$, podría interpretarse que el jugador recibe al final del juego una cantidad entre los límites C y D en lugar de que el jugador obtenga la cantidad exacta N como en el modelo no difuso.*

Capítulo 4

Aplicaciones de PDMs descontados difuso

En este capítulo se proporcionan aplicaciones de la teoría desarrollada en la Sección 2.5 en la que se trató a los DMPs con un espacio de estado finito, conjuntos de acción compactos con recompensa descontada de tipo trapezoidal difusa, tanto con horizonte finito e infinito.

Ya que la principal motivación para analizar este tipo de PDMs fue predominantemente económico, se tratará un modelo de horizonte finito de un sistema de inventario en el que se utiliza un algoritmo para calcular la solución óptima, y, adicionalmente para el caso de horizonte infinito, un MDP en un contexto económico y financiero es presentado.

4.1. Un sistema de control de inventario difuso

En esta sección, primero se presentará un ejemplo clásico de sistema de control de inventario [24], luego se introducirá un sistema de control de inventario difuso trapezoidal. La solución óptima del inventario difuso se obtiene mediante una aplicación del Teorema 2.5 y la solución del sistema de inventario nítido.

El siguiente ejemplo se aborda en [24], a continuación se presenta un resumen de los puntos de interés para presentar su versión difusa. Considere la siguiente situación: un almacén donde cada cierto período de tiempo el gerente realiza un inventario para determinar la cantidad de producto almacenado. Con base en dicha información, se toma la decisión de pedir o no una cierta cantidad de producto adicional a un proveedor. El objetivo del gestor es maximizar el beneficio obtenido. Se supone que la demanda del producto es una distribución de probabilidad conocida y aleatoria. Se tratarán los siguientes supuestos para

proponer el modelo matemático.

Supuestos en el inventario

a) La decisión de una orden adicional se toma al principio del periodo y se entrega de inmediato.

b) Las demandas de productos se reciben a lo largo del período de tiempo, pero se cumplida en el último instante del tiempo del plazo.

c) No hay pedidos pendientes.

c) Los ingresos y la distribución de la demanda no varían con el período.

d) El producto solo se vende en unidades enteras.

e) El almacén tiene una capacidad para M unidades, donde M es un número entero positivo.

Entonces, bajo la suposición anterior, el espacio de estado está dado por $X := \{0, 1, 2, ..., M\}$, el espacio de acción y el conjunto de acción admisible están dados por $A := \{0, 1, 2, ...\}$ y $A(x) := \{0, 1, 2, ..., M - x\}, x \in X$, respectivamente.

Ahora, considere las siguientes variables: sea x_t el inventario en el tiempo $t = 0, 1, ...$, la evolución del sistema está modelada por una dinámica que sigue un proceso de Lindley

$$x_{t+1} = (x_t + a_t - D_{t+1})^+, \tag{4.1}$$

con $x_0 = x \in X$ conocido, donde $(z)^+ = max\{0, z\}$, $z \in \mathbb{R}$, y

a) a_t denota el control o decisión aplicada en el instante t y representa la cantidad ordenada por el gerente de inventario (o tomador de decisiones).

b) La secuencia $\{D_t\}$ está conformada por variables aleatorias no negativas independientes e idénticamente distribuidas con distribución común $p_j := \mathbb{P}(D = j), j = 0, 1, ...$, donde D_t denota la demanda en el periodo de tiempo t.

Observe que la ecuación en diferencias dada en (4.1) induce un kernel estocástico definido en X dado $\mathbb{K} := \{(x, a) : x \in X, a \in A(x)\}$, como sigue

$$Q(x_{t+1} \in (-\infty, y])|x_t = x, a_t = a) = 1 - \Delta(x + a - y),$$

donde Δ es la distribución de D con $x \in X$, $y, a \in \{0, 1, ...\}$ y $Q(x_{t+1} \in (-\infty, y])|x_t = x, a_t = a) = 0$, si $x \in X, a \in \{0, 1, ...\}$ y $y < 0$. Entonces se sigue que

$$Q(\{x_{t+1}=y\}|x,a)=\begin{cases}0 & if \quad M\geq y>x+a\\ p_{x+a-y} & if \quad M\geq x+a\geq y>0\\ q_{x+a} & if \quad M\geq x+a, y=0.\end{cases}$$

La función de recompensa escalonada viene dada por $R(x,a)=E[H(x+a-(x+a-D)^{+})]$, $(x,a)\in\mathbb{K}$, dónde $H:\{0,1,...\}\to\{0,1,...\}$ es la función de ingresos, que es una función conocida y D es un elemento genérico de la secuencia $\{D_t\}$. De manera equivalente, $R(x,a)=F(x+a)$, $(x,a)\in\mathbb{K}$, donde

$$F(u):=\sum_{k=0}^{u-1}H(k)p_k+H(u)q_u, \tag{4.2}$$

con $q_u:=\sum_{k=u}^{\infty}p_k$. El objetivo de esta sección es maximizar la recompensa total descontada con un horizonte finito, ver (2.55).
En particular, suponga que el horizonte es $T=156$, el espacio de estado $X=\{0,1,...,9\}$, la función de ingreso $H(u)=5u$ y la ley de transición se da en la Tabla 4.1.

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	1.0000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
[2,]	1.0000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
[3,]	0.9777778	0.02222222	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
[4,]	0.9333333	0.04444444	0.02222222	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
[5,]	0.8666667	0.06666667	0.04444444	0.02222222	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
[6,]	0.7777778	0.08888889	0.06666667	0.04444444	0.02222222	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
[7,]	0.6666667	0.11111111	0.08888889	0.06666667	0.04444444	0.02222222	0.00000000	0.00000000	0.00000000	0.00000000
[8,]	0.5333333	0.13333333	0.11111111	0.08888889	0.06666667	0.04444444	0.02222222	0.00000000	0.00000000	0.00000000
[9,]	0.3777778	0.15555556	0.13333333	0.11111111	0.08888889	0.06666667	0.04444444	0.02222222	0.00000000	0.00000000
[10,]	0.2000000	0.17777778	0.15555556	0.13333333	0.11111111	0.08888889	0.06666667	0.04444444	0.02222222	0.00000000

Tabla 4.1: Ley de transición.

Algoritmo Para calcular el valor óptimo y la política óptima.

Input: MDP

Output: El vector de valor óptimo.

Una política óptima

Inicializar $W_T(x,A)=0$, $W_T^*(x)=0$,

$K_T(x)=W_T^*(x)$.

$t=T-1$

repeat

for $x \in S$ **do**
  $f_x = 0$
  $a(x) = f_x$
  $W(x, a(x)) = R(x, a(x)) +$
    $\beta \sum_{i=0}^{Z} Q(y|x + a(x)) W_{t+1}(y, 0)$

  $A(x) = 1, ..., M - x$
  for $a \in A(x)$ **do**
    $W_t(x, a) = R(x, a) +$
      $\beta \sum_{y=0}^{Z} Q(y|x + a) W_{t+1}(y, 0)$
    if $W_t(x, a) \geq W(x, a(x))$ **do**
      $W(x, a(x)) = W_t(x, a)$
      $f_x = a$
  end for

  $W_t(x) = W_t(x, f_x)$

  $W_t(x, 0) = W_t(x)$

  if $W_t(x) \geq K_{t+1}(x)$ **do**
    $K_t(x) = W_t(x)$

  $W^*(x) = K_t(x)$
end for

$t = n - 1$

until $t = 0$

En consecuencia, la salida del programa se obtiene como se ilustra en la Tabla 4.2. En esta matriz, la última columna representa la política óptima y la penúltima columna, la función de valor, para cada estado $x \in \{0, 1, ..., 9\}$. La otra entrada de la matriz representa lo siguiente:

$$G(x, a) := R(x, a) + \alpha E[W_1(F(x, a, D))],$$

$(x, a) \in \mathbb{K}$.

										V_t	π^*
[1,]	285.0000	290.0000	294.8889	299.5555	303.8889	307.7778	311.1111	313.7778	315.6666	316.6666	9
[2,]	290.0000	294.8889	299.5555	303.8889	307.7778	311.1111	313.7778	315.6666	316.6666	0.0000	8
[3,]	294.8889	299.5555	303.8889	307.7778	311.1111	313.7778	315.6666	316.6666	0.0000	0.0000	7
[4,]	299.5555	303.8889	307.7778	311.1111	313.7778	315.6666	316.6666	0.0000	0.0000	0.0000	6
[5,]	303.8889	307.7778	311.1111	313.7778	315.6666	316.6666	0.0000	0.0000	0.0000	0.0000	5
[6,]	307.7778	311.1111	313.7778	315.6666	316.6666	0.0000	0.0000	0.0000	0.0000	0.0000	4
[7,]	311.1111	313.7778	315.6666	316.6666	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	3
[8,]	313.7778	315.6666	316.6666	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	2
[9,]	315.6666	316.6666	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1
[10,]	316.6666	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0

Tabla 4.2: Solución óptima.

En conclusión, la función de valor óptimo es $V_T(x) = 316.6$ para cada $x \in X$ y la política óptima está dada por $f_t(x) = M - x$, $t = 0, 1, ..., T-1$, $x \in X$ con $M = 9$.

Ahora bien, considerando que en la investigación de operaciones a menudo es difícil para un gerente controlar los sistemas de inventario, debido a que los datos en cada etapa de observación no siempre son ciertos, entonces se debe aplicar un enfoque difuso. De esta forma, se tendrá en cuenta el sistema de inventario anterior en un entorno difuso, es decir, se considerará la función de recompensa dada en el Supuesto 2.4:

$$\tilde{R}(x, a) = (BR(x, a), CR(x, a), DR(x, a), FR(x, a)),$$

con $0 < B < C \leq D < F$. Entonces, por el Teorema 2.5, se deduce que la política óptima del problema de control óptimo difuso está dada por $\tilde{\pi}^* = \{f_0, ..., f_{T-1}\}$, donde $f_t(x) = M - x$, $t = 0, 1, ..., T-1$, $x \in X$ y la función de valor óptimo está dada por

$$\tilde{V}_T(x) = V_T(x)(B, C, D, F),$$

$x \in X$.

4.2. Un problema de selección de portafolio

Sean $X = \{\chi_0, \chi_1\}$, $0 < \chi_0 < \chi_1$, $A(\chi) = [0, 1]$, $\chi \in X$. La ley de transición está dada por

$$Q(\{\chi_0\}|\chi_0, a) = p, \tag{4.3}$$
$$Q(\{\chi_1\}|\chi_0, a) = 1 - p, \tag{4.4}$$
$$Q(\{\chi_1\}|\chi_1, a) = q, \tag{4.5}$$
$$Q(\{\chi_0\}|\chi_1, a) = 1 - q, \tag{4.6}$$

para toda $a \in [0,1]$, donde $0 \leq p \leq 1$ y $0 \leq q \leq 1$. La recompensa está dada por $R(\chi, a)$, $(\chi, a) \in \mathbb{K}$ que satisface:

Supuesto 4.1. **(a)** *R depende solo de a, que es $R(\chi, a) = U(a)$, para todo $(\chi, a) \in \mathbb{K}$, donde U es no−negativa y continua.*

(b) *Existe un $a^* \in [0,1]$ tal que*

$$max_{a\in[0,1]} U(a) = U(a^*),$$

para todo $\chi \in X$.

Una interpretación de este ejemplo se da en la siguiente observación.

Observación 4.1. *Los estados χ_0 y χ_1 representan el comportamiento de ciertos mercados bursátiles, lo cual es malo ($\equiv \chi_0$) y bueno ($\equiv \chi_1$). Si asumimos que, para cada a y $t = 0, 1, ...$, la probabilidad de ir de χ_0 a χ_0 es p (resp. la probabilidad de χ_0 to χ_1 es $1 - p$); además, para cada a y $t = 0, 1, \cdots$, la probabilidad de ir de χ_1 a χ_1 es q (resp. la probabilidad de ir desde χ_1 hasta χ_0 es $1-q$). Ahora, especificamente, supongamos que en un problema dinámico de elección de cartera, hay dos activos disponibles para un inversionista: uno está libre de riesgo y se supone que la tasa de riesgo $r > 0$ es conocida y constante a lo largo del tiempo y una varianza σ^2. Siguiendo el Ejemplo 1.24 en [?], la utilidad esperada del inversionista podría ser dada por la expresión:*

$$U(a) = a\mu + (1-a)r - \frac{k}{2}a^2\sigma^2, \tag{4.7}$$

donde $a \in [0,1]$ es la fracción de su dinero que el inversionista invierte en el activo riesgoso y el resto $1-a$, lo invierte en el activo sin riesgo. En (4.7) , k representa el valor que el inversor le da a la varianza relativa a la experanza. Observe que si $\mu > \frac{k\sigma^2}{2}$, entonces U definido en (4.7) es positivo en $[0,1]$ (de hecho, en este caso $U(0) = r > 0$ y $U(1) = \mu - \frac{k\sigma^2}{2} > 0$); además, es posible probar (ver [?]) que si $0 < \mu - r < k\sigma^2$, entonces $max_{a\ en[0,1]} U(a)$ se obtiene para $a^* \in (0,1)$ dado por

$$a^* = \frac{\mu - r}{k\sigma^2}.$$

Por lo tanto, tomando $R(\chi, a) = U(a)$, $\chi \in X$ y $a \in [0,1]$, donde U viene dado por ([33]), y considerando las dos últimas desigualdades dadas en el párrafo anterior, se cumple el Supuesto 4.1.

Lema 4.1. *Supongamos que el Supuesto 4.1 se mantiene. Entonces, por el Teorema 2.4 ,*

$$V(\chi) = \frac{U(a^*)}{1-\alpha}$$

y $f^*(\chi) = a^*$, para toda $\chi \in X$.

Demostración. En primer lugar, se encontrarán las funciones de iteración de valor: V_n, para $n = 1, 2, \ldots$.
Por Teorema 2.4,

$$V_1(\chi_0) = \max_{a\in[0,1]} U(a),$$

esto implica que $V_1(\chi_0) = U(a^*)$. En un camino similar, es posible obtener que $V_1(\chi_1) = U(a^*)$.

Ahora, para $n = 2$,

$$\begin{aligned} V_2(\chi_0) &= \max_{a\in[0,1]} \{U(a) + \alpha[V_1(\chi_1)(1-p) + V_1(\chi_0)p]\} \\ &= U(a^*) + \alpha[V_1(\chi_1)(1-p) + V_1(\chi_0)p] \\ &= U(a^*) + \alpha[U(a^*)(1-p) + U(a^*)p] \\ &= U(a^*) + \alpha U(a^*). \end{aligned}$$

Análogamente, $V_2(\chi_1) = U(a^*) + \alpha U(a^*)$. Continuando en este sentido, se obtiene que

$$V_n(\chi_0) = V_n(\chi_0) = U(a^*) + \alpha U(a^*) + \ldots + \alpha^{n-1}U(a^*),$$

para toda $n = 1, 2, \ldots$.

Por Teorema 2.4, $V_n(\chi) \to V(\chi)$, $n \to \infty$, $\chi \in X$, el cual implica que $V(\chi) = \dfrac{U(a^*)}{1-\alpha}$, $\chi \in X$. Y, de la Ecuación de Programación Dinámica (ver (2.54)), se sigue que $f^*(\chi) = a^*$, para todo $\chi \in X$.

Ahora, supongamos que la función de recompensa difusa está dada por

$$\tilde{R}(x, a) = (B, C, D, F)R(x, a),$$

with $(x, a) \in \mathbb{K}$. Entonces, como consecuencia del Teorema 2.6 se obtienen los resultados. □

Lema 4.2. *Para la versión difusa del problema de elección de cartera, resulta que* $\widetilde{V}(\chi) = V(\chi)(B, C, D, F)$ *and* $f^*(\chi) = a^*$, *para todo* $\chi \in X$.

4.3. Un juego de dos personas

Ahora, se presentan un modelo de un juego estocástico entre dos jugadores que buscan maximizar sus recompensas totales descontadas. Denotemos

por J_1 y J_2 los jugadores/inversores. Cada uno de ellos sigue un modelo de decisión similar al propuesto en la Sección (4.2). Esto es, J_1 tiene un modelo de decisión del tipo: (X, A, Q, R_1), donde $X = \{X_0, X_1\}, 0 < X_0 < X_1$, $B = B(\mathcal{X}) = [0, 1], \mathcal{X} \in X$. La ley de transición Q se da como en la Sección (4.2) es independiente de la decisión a), y la recompensa viene dada por la función $R_1 = U_2$ con

$$U_1(a) = a\mu_1 + (1 - a)r_1 - \frac{k_1}{2}a^2\sigma_1^2,$$

con $a \in [0, 1]$. Además, se asume que $0 < \mu_1 - r_1 < k_1\sigma_1^2$, entonces $max_{a\in[0,1]}U_1(a)$ es alcanzado en $a^* \in (0, 1)$ dado por

$$a^* = \frac{\mu_1 - r_1}{k\sigma_1^2}.$$

Sea $\mathbb{F}$ el correspondiente conjunto de estrategias estacionarias para J_1. Observemos que

$$\mathbb{F} = \left\{ \sum_{i=1}^{n} \lambda_i f_i : \sum_{i=1}^{n} \lambda_i = 1, \lambda_i \geq 0,\ f_i \in \mathbb{F},\ n \geq 1 \right\}.$$

Entonces, $\mathbb{F}$ también puede ser visto como el conjunto de estrategias mixtas para J_1.

Ahora, para J_2, el modelo de decisión es de la forma (X, B, Q, R_2) ,donde $X = \{\chi_0, \chi_1\}$, $0 < \chi_0 < \chi_1$, $B = B(\chi) = [0, 1]$, $\chi \in X$. La ley de transición Q es dada como en (4.3)-(4.6), y la recompensa está dada por $R_2 = U_2$ con

$$U_2(b) = a\mu_2 + (1 - b)r_2 - \frac{k_2}{2}b^2\sigma_2^2,$$

donde $b \in [0, 1]$, y también se supone que $0 < \mu_2 - r_2 < k_2\sigma_2^2$. Por lo tanto, $max_{b\in[0,1]}U_2(b)$ es alcanzado en $b^* \in (0, 1)$ dado por

$$b^* = \frac{\mu_2 - r_2}{k_2\sigma_2^2}.$$

Sea $\mathbb{G}$ el conjunto correspondiente de estrategias estacionarias (o mixtas) para J_2.

El juego se desarrolla de la siguiente manera. Dado un estado inicial $x_0 \in X$, ambos jugadores toman una decisión $a_0 \in A(x_0)$ y $b_0 \in B(x_0)$ de acuerdo a las estrategias mixtas f y g. Entonces cada jugador recibe una recompensa esperada $E_{x_0}^{f,g}[U_1(x_0, a_0, b_0)]$ and $E_{x_0}^{f,g}[U_2(x_0, a_0, b_0)]$, respectivamente. El juego entonces cambia a un nuevo estado $x_1 \in X$ de acuerdo con la transición $Q(\cdot|x_0)$ y luego el proceso se repite. Con el tiempo, ambos jugadores recibirán el total

de sus recompensas esperadas por cada decisión tomada durante el juego, es decir, recibirán

$$V_{J_1}(\chi, f, g) = \sum_{t=0}^{\infty} \alpha^t E_{\chi,f,g}[U_1(f(x_t)] \text{ and } V_{J_2}(\chi, f, g) = \sum_{t=0}^{\infty} \alpha^t E_{\chi,f,g}[U_2(g(x_t)],$$

respectivamente, donde $x_0 = \chi$. Tenga en cuenta que el juego descrito constituye un juego estocástico descontado entre dos jugadores en el que toman decisiones de forma independiente y simultánea.

A continuación, un par de estrategias (f^*, g^*) es llamadoa *equilibrio de Nash* si

$$V_{J_1}(\chi, f^*, g^*) = \sup_{f' \in \mathbb{F}} V_{J_1}(\chi, f', g^*)$$

y

$$V_{J_2}(\chi, f^*, g^*) = \sup_{g' \in \mathbb{G}} V_{J_2}(\chi, f^*, g'),$$

para cada $\chi \in X$.

Lema 4.3. *Para el juego de dos personas, el par (f^*, g^*) con $f^*(\chi) = a^*$ y $g^*(\chi) = b^*$, para toda $\chi \in X$ es un equilibrio de Nash, y*

$$V_{J_1}(\chi, f^*, g^*) = \frac{U_1(a^*)}{1-\alpha}$$

y

$$V_{J_2}(\chi, f^*, g^*) = \frac{U_2(b^*)}{1-\alpha},$$

para todo $\chi \in X$.

Demostración. Observemos que, bajo las condiciones del problema de selección de portafolio, para $f \in \mathbb{F}$, $g \in \mathbb{G}$, y $x_0 = \chi$,

$$V_{J_1}(\chi, f, g) = \sum_{t=0}^{\infty} \alpha^t E_{\chi,f}[U_1(f(x_t)]$$

y

$$V_{J_2}(\chi, f, g) = \sum_{t=0}^{\infty} \alpha^t E_{\chi,g}[U_2(g(x_t)].$$

Por lo tanto, una aplicación directa del Lema 4.2 y el Teorema 2.6 permiten obtener la demostración de los siguientes resultados.

□

Lema 4.4. *Supongamos que la función de recompensa difusa está dada por*

$$\tilde{R}(x,a) = (B,C,D,F)R(x,a),$$

con $(x,a) \in \mathbb{K}$. *Entonces, la versión del juego de dos personas, el equilibrio de Nash viene dado por* (a^*, b^*) *y* $\tilde{V}_{J_1}(\chi) = V_{J_1}(\chi, f, g)(B,C,D,F)$ *and* $\tilde{V}_{J_2}(\chi) = V_{J_2}(\chi, f, g)(B,C,D,F)$, *para todo* $\chi \in X$.

Como observación final, siguiendo ideas similares dadas en la Sección 2.4, es posible obtener la solución óptima del siguiente problema de control óptimo.

Considere los modelos decisión en versión estándar dados por

$$M_1 \quad = \quad (X, A, \{A(x) : x \in X\}, Q, R_1), \tag{4.8}$$

y

$$M_2 \quad = \quad (X, A, \{A(x) : x \in X\}, Q, R_2), \tag{4.9}$$

donde ambos modelos satisfacen los supuestos dados en la Sección 2.3.1, y

$$0 < R_1(x,a) \leq R_2(x,a) < \gamma, \tag{4.10}$$

para todo $x \leq X, a \in A(x), \gamma$ es una constante positiva y $R_2 = zR_1, z > 1$.

Ahora, tenga en cuenta el problema de control óptimo difuso de horizonte infinito con modelo decisión:

$$\widetilde{M} \quad = \quad (X, A, \{A(x) : x \in X\}, Q, \widetilde{R}) \tag{4.11}$$

con

$$\widetilde{R}(x,a) := (0, R_1(x,a), R_2(x,a), \gamma) \tag{4.12}$$

$x \in X, a \in A(x)$. Nótese que R dada en (4.12) modela el hecho de que, en sentido difuso, "la recompensa está aproximadamente en el intervalo

$$[R_1(x,a), R_2(x,a)], x \in X, a \in A(x)".$$

Sean v_i, V_i y f_i^* ser la función objetivo, la función de valor óptimo y la política estacionaria óptima, respectivamente para el modelo $M_i, i = 1, 2$ y sea V por el función de valor óptimo para M. Como en la demostración del Teorema 2.5, usando eso para cada $\pi \in \mathbb{F}$ y $x \in X$,

$$v_1(\pi, x) = v_2(\pi, x) \tag{4.13}$$

y es directo obtener que, para cada $\pi \in \mathbb{F}, x \in X$ y α :

$$\alpha v_1(\pi, x) \leq \alpha v_2(f^*, x) = \alpha v_2(x)) \tag{4.14}$$

y

$$\begin{aligned} \alpha v_1(\pi, x) + (1-\alpha)(\frac{\gamma}{1-\beta}) &\leq \alpha v_2(f^*, x) + (1-\alpha)(\frac{\gamma}{1-\beta}) && (4.15) \\ &= \alpha v_2(x). && (4.16) \end{aligned}$$

Por tanto, de (4.14) y (4.15) resulta que

$$\widetilde{V}(x) = \left(0, V_2(x), V_2(x), \frac{\gamma}{1-\beta}\right) \tag{4.17}$$

$x \in X$, y $f_1^* = f_2^*$ es óptimo para M. Observe que V puede verse como el tipo triangular:

$$\widetilde{V}(x) = \left(0, V_2(x), \frac{\gamma}{1-\beta}\right). \tag{4.18}$$

Resumen y conclusiones

En resumen, la teoría presentada en este trabajo tiene en cuenta la imprecisión o ambigüedad en la función de recompensa, lo cual nos permitió ampliar la teoría estándar de PDMs dando solución a dos problemas en tiempo discreto con espacios de estados finitos:

- El primero de ellos con conjunto de acciones finito y criterio de recompensa total esparada difusa.
- El segundo considera un conjunto de acciones compacto para el caso de recompensa descontado total esperada difusa.

Ambos criterios tanto en horizontes finito e infinito. En ambos, las funciones de recompensas fueron planteadas en forma difusa para modelar la incertidumbre, específicamente de tipo trapezoidal con una forma conveniente en función de una recompensa estándar no difusa como está dada en el Supuesto 2.4.

Para la realización de este trabajo fue necesario estudiar la teoría de Programación Dinámica y los conceptos elementales de la Teoría de lógica difusa. Dentro de la parte de lógica difusa, se expusieron los conceptos principales que se usaron durante el desarrollo de la tesis, como el de números difusos, $\alpha-$cortes, operaciones entre números difusos trapezoidales, orden máximo difuso, métrica en el conjunto de los números difusos, variable aleatoria difusa y esperanza de una variable aleatoria difusa trapezoidal dada por una variable aleatoria multiplicada por un número trapezoidal. Dichos conceptos fueron necesarios ya que al trabajar con recompensas difusas trapezoidales, los criterios de rendimientos se convierten en una variable aleatoria difusa de tipo trapezoidal mediante la operación de sumas tanto finitas como infinitas y por el hecho de que el estado del sistema es una variable aleatoria.

Del área de PDMs se exhibió el Modelo de Control de Markov y los tipos de políticas, los cuales generan el espacio de probabilidad del Proceso estocástico de interés, que es el Proceso de Decisión de Markov. Adicionalmente se recopiló una lista de las condiciones necesarias para resolver los PDMs con la técnica de Programación Dinámica, las cuales garantizan que se cumpla la Condición

de Selección Medible y el método de aproximaciones sucesivas para cuando se trabaja en horizonte infinito.

Cada uno de los problemas se presentaron en una versión nítida y fueron transformados en una versión difusa trapezoidal, y con los supuestos provistos en la Sección 2.2 para los problema óptimos no difusos, las principales consecuencias que se obtienen son que la política óptima del problema difuso coincide con el problema estándar nítido y la función de valor óptimo difusa tiene una forma trapezoidal conveniente.

Con la intención de ilustrar la teoría desarrollada en este trabajo, se adicionaron cinco problemas de aplicaciones relacionados con PDMs en este contexto difuso, de los cuales dos se abordaron a través de un PDM bajo recompensa total esperada, siendo uno de parada óptima y el otro de apuesta. Los tres problemas restantes estuvieron relacionados con recompensa total descontada, uno de un sistema de control de inventario difuso, debido a que en la investigación de operaciones a menudo es difícil para un gerente controlar los sistemas de inventario, debido a que los datos en cada etapa de observación no siempre son ciertos, otro problema se trató con la selección de portafolio y el último problema de aplicación se refirió a un juego de dos personas. Es relevante señalar que, en la versión difusa del modelo de juego dado, las estrategias audaces y tímidas, que son bien conocidas en el contexto del juego, aparecen como las estrategias óptimas para el jugador, y que la fortuna N que al final del juego recibirá el jugador puede ser sustituida por el hecho de que N pertenece a un cierto intervalo.

Bibliografía

[1] Abbasbandy, S., Hajjari, T. (2009). A new approach for ranking of trapezoidal fuzzy numbers. Computers and mathematics with applications **57**(3), 413-419.

[2] Aliprantis, C.D., Border, K. (2006). Infinite Dimensional Analysis. Springer, Heidelberg . https://doi.org/10.1007/3-540-29587-9.pdf

[3] Ban A. I. (2009). Triangular and parametric approximations of fuzzy numbers inadvertences and corrections. Fuzzy Sets and Systems, 160(21), 3048–3058.

[4] Bartle, R.G. (1995). The Elements of Integration. Wiley.

[5] Bellman R. E. and Zadeh L. A. (1970) Decision-making in a fuzzy enviroment. Management Sciences, Vol. 17, No. 4, 141-164.

[6] Carrero-Vera, K., Cruz-Suárez, H., Montes-de-Oca, R. (2021). Discounted Markov decision processes with fuzzy rewards induced by non-fuzzy systems. In: Parlier G.H., Liberatore F., Demange, M. (eds.) ICORES 2021, Proceedings of the 10th International Conference on Operations Research and Enterprise Systems, pp. 49–59. SCITEPRESS

[7] Cavazos-Cadena R. and Montes-de-Oca R. (2001) Existence of optimal stationary policies in finite dynamic programs with nonnegative rewards. Probabiliy in the Engineering and Informational Sciences, Vol. 15, 557-564.

[8] Chen S.H. (1998). Operations of fuzzy numbers with step form membership function using function principle. Journal of Information Sciences 108, 149-155.

[9] Diamond, P. and Kloeden, P. (1994). Metric Spaces of Fuzzy Sets. World Scientific.

[10] Driankov, D., Hellendoorn, H., and Reinfrank, M. (2013). An Introduction to Fuzzy Control. Springer Science and Business Media.

[11] D. Dubois and H. Prade (1980). Fuzzy Sets and Systems: Theory and Applications. AcademicPress.

[12] Efendi, N. Arbaiy, and M. M. Deris. (2018). A new procedure in stock market forecasting based on fuzzy random auto-regression time series model. Information Sci. 441, 113-132. DOI:10.1016/j.ins.2018.02.016

[13] M. Fakoor, A. Kosari, and M. (2016). Jafarzadeh: Humanoid robot path planning with fuzzy Markov decision processes. J. Appl. Res. Tech. 14, 300-310. DOI:10.1016/j.jart.2016.06.006.

[14] Furukawa, N. (1997). Parametric orders on fuzzy numbers and their roles in fuzzy optimization problems. Optimization **40**, 171-192.

[15] Guo, Y., Jiao, L., Wang, S., Wang, S., Liu, F., Hua, W. (2018). Fuzzy superpixels for polarimetric SAR images classification. IEEE Trans. Fuzzy Syst. 26(5), 2846–2860.

[16] Hernández-Lerma, O.: Adaptive Markov Control Processes. Springer Science Business Media, Heidelberg (1989). https://doi.org/10.1007/978-1-4419-8714-3

[17] Klir, G.J., Yuan, B. (1995). Fuzzy Sets and Fuzzy Logic: Theory and Applications. Prentice Hall, Upper Saddle River.

[18] Kurano M., Yasuda M., Nakagami J. and Yoshida Y. (2002). Markov decision processes with fuzzy rewards Proc. Int. Conf. on Nonlinear Analysis (Hirosaki, Japan) , 221-232.

[19] López-Díaz, M and Ralescu, D. A. (2006). Tools for fuzzy random variables: embeddings and measurabilities, Comput. Statist. Data Anal., 109-114.

[20] Pedrycz,W.(1994). Why triangular membership functions? Fuzzy Sets and Systems, 64(1), 21–30.

[21] Phuong, N.H., Kreinovich, V. (2001). Fuzzy logic and its applications in medicine. Int. J. Med. Inf. 62(2–3), 165–173.

[22] Porteus, E.L. (2002). Foundations of Stochastic Inventory Theory. Stanford Business Books, Stanford.

[23] Puri, M. L. and Ralescu, D. A.: Fuzzy random variable, J. Math. Anal. Appl. Vol. 114 , 402-422, 1986.

[24] Puterman M. L. 2005. Markov Decision Processes: Discrete Stochastic Dynamic. First Edition, Wiley-Interscience, California.

[25] J.Ramík and J.Rimánek. (1985). Inequality relation between fuzzy numbers and its use in fuzzy optimization, Fuzzy Sets and Systems, 16, 123-138.

[26] Rezvani S. and Molani M. (2014). Representation of trapezoidal fuzzy numbers with shape function, Annals of Fuzzy Mathematics and Informatics, Vol. 8, No. 1, 89-112.

[27] Ross S. (1974). Dynamic programming and gambling models. Advances in Applied Probability, Vol. 6,No. 3, 593-606.

[28] Ross S. (1983). Introduction to Stochastic Dynamic Programming. Academic Press.

[29] Semmouri, A., Jourhmane, M., and Belhallaj, Z. (2020). Why triangular membership functions? Annals of Operations Research, pages 1–18.

[30] Shapley, L.S. (1953). Stochastic games. Proc. Natl. Acad. Sci. 39(10), 1095–1100.

[31] Syropoulos, A. and Grammenos, T. (2020). A Modern Introduction to Fuzzy Mathematics. Wiley.

[32] V. Noviik. (1989). Fuzzy Sets and their Applications (Adam Hilder, Bristol-Boston).

[33] Zadeh, L. A.: Fuzzy sets. Information and Control, Vol. 8, 338-353.

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