"El saber de mis hijos hará mi grandeza"

UNIVERSIDAD DE SONORA

División de Ciencias Exactas y Naturales

Programa de Posgrado en Matemáticas

Modelos de control en sistemas estocásticos de interacción de objetos bajo un enfoque de campo medio

T E S I S

Que para obtener el título de:

Maestra en Ciencias (Matemáticas)

Presenta:

María Elena Martínez Manzanares

Directores de tesis: Dr. Jesús Adolfo Minjárez Sosa

Dra. Carmen Geraldi Higuera Chan

Hermosillo, Sonora, México, 21 de julio del 2020

SINODALES

Dra. Carmen Geraldi Higuera Chan
Universidad de Sonora, Hermosillo, México

Dr. Fernando Luque Vásquez
Universidad de Sonora, Hermosillo, México

Dr. Jesús Adolfo Minjárez Sosa
Universidad de Sonora, Hermosillo, México

Dr. Saúl Mendoza Palacios
El Colegio de México, Ciudad de México, México

Agradecimientos

Quiero agradecer a las siguientes personas.

A mi mamá, Betsabé, y a mi hermano, Isaac.

A mi abuela, María Luisa.

A mi abuelo, José Manzanares.
Bueno, mejor dicho, a la memoria de mi abuelo José.

A toda mi familia y amigos.

A mis directores por su paciencia.

Índice general

Introducción

La teoría de campo medio tiene como objetivo estudiar el comportamiento de sistemas que se componen de un número grande de objetos que interactúan entre sí a través de una clase apropiada de promedios. Por lo tanto, esta teoría es útil para analizar sistemas de dimensiones muy grandes por medio de un modelo simplificado, llamado modelo de campo medio. Esta teoría tiene sus inicios en el área de física a inicios del siglo XX con el estudio de sistemas de interacción de partículas.

Las aplicaciones de la teoría de campo medio se han extendido a otras áreas como por ejemplo, economía, investigación de operaciones y biología (ver, *e.g.* [1, 15, 23]). Particularmente en matemáticas se ha utilizado ampliamente en el área de optimización, principalmente en control estocástico y teoría de juegos (ver [5, 6, 12–14, 24]). Es precisamente en esta área donde se centra nuestro trabajo.

Específicamente, la tesis aborda el estudio de sistemas que se componen de N objetos que interactúan entre sí a tiempo discreto. Existe un controlador con intereses bien definidos y en cada etapa toma una acción que afecta el comportamiento de los objetos. En este contexto, supondremos lo siguiente:

- El número de objetos N es del orden del infinito, lo cual denotaremos como $N \sim \infty$.
- Los objetos se clasifican de acuerdo a características determinadas por el controlador dando lugar a un conjunto de clases o categorías de objetos.
- El número de clases, que denotaremos por s, es finito.

El sistema evoluciona de la siguiente manera. Los objetos estarán distribuidos en las s clases; la evolución de cada objeto se produce de acuerdo a una probabilidad de transición K_{ij} de la forma

$$K_{ij}(a) = P[X_n^N(t+1) = j | X_n^N(t) = i, a_t = a], \quad i, j \in S.$$

donde $X_n^N(t)$ es la clase del objeto n al tiempo t, a_t la acción del controlador y S es el conjunto de clases. Además se genera un costo y una vez que se tiene la nueva configuración de los objetos en las clases, el proceso se repite.

La configuración del sistema estará dada por un vector $M^N(t)$ cuyas componentes son la proporción de objetos en cada clase.

Lo anterior define un modelo de control $\mathcal{M}_N$ cuyos estados son los vectores $M^N(t)$. Entonces, el objetivo es estudiar el problema de control óptimo asociado a $\mathcal{M}_N$. Sin embargo, el hecho de que $N \sim \infty$ nos lleva a enfrentar la llamada "maldición de la dimensión", lo cual implica que en términos prácticos el problema sea imposible de resolver. Es a partir de esta situación donde se aplica la teoría de campo medio para presentar una solución computacionalmente factible. Concretamente calculamos el límite cuando $N \to \infty$ del modelo $\mathcal{M}_N$, al cual se le conoce como "límite de campo medio", y obtenemos un nuevo modelo $\mathcal{M}$ cuyos estados son las medidas de probabilidad correspondientes al límite del vector de proporciones $M^N(t)$. A $\mathcal{M}$ se le conoce como el modelo de campo medio y es independiente de N. Más aún, el problema de control asociado es determinista, y por lo tanto, más fácil de resolver. En este sentido, podemos calcular una política óptima π_* en $\mathcal{M}$. Usando a π_* para controlar el proceso original en $\mathcal{M}_N$, el objetivo es demostrar que para N suficientemente grande, como es nuestro caso, π_* es óptima. En otras palabras, mostraremos que la desviación de optimalidad de π_* en $\mathcal{M}_N$ converge a cero cuando $N \to \infty$.

Este tipo de problemas han sido estudiados en diferentes contextos. Por ejemplo, en [5, 6] se analiza, principalmente, la tasa de convergencia en el límite de campo medio. En [13, 14] se estudia el sistema de objetos cuando la variable aleatoria que define la dinámica tiene distribución desconocida. Además, en [12] se analiza una clase de juegos de suma cero asociado al sistema de objetos.

Otra clase de problemas que estudiaremos en la tesis son los problemas de control óptimo donde existe la posibilidad de que el sistema de objetos puede ser interrumpido, ya sea por acciones seleccionadas por el controlador o por agentes externos. En este sentido, en cada etapa, supondremos que existe una probabilidad positiva, que dependerá de la configuración del sistema, de que el sistema se detenga o muera. Esto nos conduce a analizar un problema de control con horizonte aleatorio, el cual será una variable aleatoria τ que representa el tiempo cuando el sistema muere o cuando el sistema es absorbido a una configuración especial considerada como el "cementerio".

Además de enfrentarnos con la "maldición de la dimensión", el estudio de este problema presenta el obstáculo del horizonte aleatorio. Entonces, para su estudio, primero transformaremos este problema en un problema con horizonte infinito, para luego aplicar la teoría de campo medio. El estudio de este problema dio lugar al artículo [18] el cual se encuentra sometido.

El presente trabajo está estructurado en cuatro capítulos. En el Capítulo 1 establecemos los elementos del modelo del control de Markov asociado al sistema de N objetos $\mathcal{M}_N$, su respectivo problema de control óptimo y condiciones bajo las cuales se puede garantizar la existencia de soluciones.

En el Capítulo 2 presentaremos los elementos del modelo de campo medio $\mathcal{M}$ y el problema de control óptimo correspondiente. Posteriormente estudiamos la desviación de óptimalidad de la solución obtenida en $\mathcal{M}$ en el modelo de control original $\mathcal{M}_N$.

En el Capítulo 3 estudiamos el modelo de control de Markov asociado al sistema de N objetos con probabilidad positiva de existinguirse. Para esto, introducimos el modelo de control de campo medio apropiado.

Finalmente, en el Capítulo 4 presentaremos una serie de ejemplos en donde se describirán sistemas de N objetos en diferentes contextos, con el fin de ilustrar nuestra teoría.

Notación

Símbolos

$\mathbb{N}$	Conjunto de los enteros positivos
$\mathbb{N}_0$	Conjunto de los enteros no negativos
$\mathbb{R}$	Conjunto de los números reales
$\mathbb{R}_+$	Conjunto de los números reales positivos
$\mathbb{1}_B(\cdot)$	Función indicadora del conjunto B
$:=$	Iguales por definición
$X_n^N(t)$	Clase del objeto n al tiempo t
S	Conjunto de s clases
$\mathbb{P}(S)$	Conjunto de medidas de probabilidad en S
$\mathbb{P}_N(S)$	Subconjunto de $\mathbb{P}(S)$ que satisface $(N)(p(i)) \in \mathbb{N}_0$
m	Un vector en $\mathbb{P}(S)$
$m(t)$	La configuración del sistema de objetos al tiempo t en el campo medio
$M^N(t)$	La configuración del sistema de N objetos al tiempo t
$\mathcal{M}_N$	Modelo de control de Markov asociado al sistema de N objetos
$\mathcal{B}(Z)$	σ-álgebra de Borel asociada al espacio de Borel Z (ver sección *Definiciones* de este apartado)

Abreviaciones

v.a.	variable aleatoria
i.i.d.	independientes idénticamente distribuidas
PCO	problema de control óptimo
N-PCO	N-problema de control óptimo (asociado al sistema con N objetos)
MCM	Modelo de control de Markov

Definiciones

- Sea $m \in \mathbb{P}(S)$, definimos la norma

$$||m||_\infty := \max\{|m_1|, |m_2|, \ldots, |m_s|\}.$$

- Sea $u : \mathbb{P}(S) \to \mathbb{R}$, definimos la norma

$$||u|| := \sup_{m \in \mathbb{P}(S)} |u(m)|.$$

- Decimos que un espacio Z es de Borel si y solo si es subconjunto de un espacio métrico completo y separable.

Espacios de funciones

$\mathcal{C}(\mathbb{P}(S))$	Espacio de funciones continuas y acotadas en $\mathbb{P}(S)$
$\mathcal{C}_+(\mathbb{P}(S))$	Subconjunto de $\mathcal{C}(\mathbb{P}(S))$ de funciones no negativas
$\mathcal{C}(\mathbb{P}_N(S))$	Espacio de funciones continuas y acotadas en $\mathbb{P}_N(S)$
$\mathcal{C}_+(\mathbb{P}_N(S))$	Subconjunto de $\mathcal{C}(\mathbb{P}_N(S))$ de funciones no negativas

Los siguientes símbolos estan definidos en la Sección 1.4:

Π^N, $\mathbb{F}$	Definición 1.4.1
Π^N_M, $\mathbb{F}^N$	Definición 1.4.2
Π_M	Observación 1.4.3 (a)
Φ^N	Teorema 1.4.8

Los siguientes símbolos estan definidos en la Sección 3.3:

Γ_n	Definida en (3.9)
$\widetilde{\mathcal{M}_N}$	Definida en (3.14)

Capítulo 1

Modelo de control de Markov para sistemas de interacción de objetos

1.1. Introducción

En este capítulo introduciremos los elementos necesarios que definen un Modelo de Control de Markov para un sistema de interacción de objetos. Este incluye los espacios de estados y control, la dinámica de los objetos, así como los conjuntos de políticas. Con estos elementos definiremos el problema de control óptimo correspondiente e impondremos las condiciones necesarias que garantizan la existencia de su solución por medio de una ecuación funcional y argumentos de programación dinámica.

1.2. Descripción del sistema de objetos

Supongamos que tenemos un sistema estocástico que evoluciona a tiempo discreto el cual se compone de un número grande de N objetos que interactúan entre sí. Denotamos por $X_n^N(t)$, $n = 1, 2, \ldots, N$, $t \in \mathbb{N}_0$ la clase del objeto n al tiempo t. Esta variable toma valores en el conjunto de clases $S = \{c_1, c_2, \ldots, c_s\} := \{1, 2, \ldots, s\}$, donde c_i representa la clase $i = 1, 2, \ldots, s$. El sistema cuenta con un controlador central el cual en cada etapa toma una decisión a_t contenida en un conjunto de Borel A. Estas decisiones afectarán a los objetos en cada etapa. De aquí, $X_n^N(t) = k$ significa que el objeto n se encuentra en la clase k al tiempo t.

La evolución del proceso estará determinada por una probabilidad de transición K_{ij} homogénea para n en $\{1, 2, \ldots, N\}$ y en N de la forma

$$K_{ij}(a) = P[X_n^N(t+1) = j | X_n^N(t) = i, a_t = a], \quad i, j \in S. \tag{1.1}$$

Es decir, $K_{ij}(a)$ representa la probabilidad de que un objeto se mueva de la clase i a la clase j cuando el controlador selecciona la acción $a \in A$. Al ser S un conjunto finito, es posible definir la ley de transición mediante el kernel estocástico $K = K(a) = [K_{ij}(a)]$, cuya forma matricial es:

$$K = K(a) = [K_{ij}(a)] = \begin{bmatrix} K_{11}(a) & K_{12}(a) & \cdots & K_{1s}(a) \\ K_{21}(a) & K_{22}(a) & \cdots & K_{2s}(a) \\ \vdots & & & \\ K_{s1}(a) & K_{s2}(a) & \cdots & K_{ss}(a) \end{bmatrix}.$$

Existen casos particulares donde la dinámica de los objetos se define de acuerdo a una ecuación en diferencias estocásticas, homogénea en N, de la forma

$$X_n^N(t+1) = F(X_n^N(t), a_t, \xi_t^n), \quad t = 0, 1, \ldots \tag{1.2}$$

donde $F : S \times A \times \mathbb{R} \to S$ es una función medible y $\{\xi_t^n : t \geq 0, n = 1, 2, \ldots, N\}$ es una familia de v.a.i.i.d. con densidad común ρ y definidas en un espacio de probabilidad $(\Omega, \mathcal{F}, P)$. Ambas funciones supondremos son concidas por el controlador.

Haciendo uso de (1.2) es posible escribir la ley de transición K de cada objeto en términos de la función F de la siguiente manera. Para cada $n = 1, 2, \ldots, N$ tenemos

$$\begin{aligned} K_{i,j}(a) &= P[X_n^N(t+1) = j | X_n^N(t) = i, a_t = a], \\ &= \int_{\mathbb{R}} \mathbb{1}_{\{j\}}[F(i, a, z)]\rho(z), \quad i, j \in S, a \in A \end{aligned} \tag{1.3}$$

donde $\mathbb{1}_B$ representa la función indicadora del conjunto B.

Claramente, el hecho de que N sea demasiado grande es un obstáculo para el controlador ya que sería imposible analizar cada objeto. Por lo tanto, asumiendo que los objetos son distinguibles, solo a través de la clase en los que se encuentran, y considerando que K no depende de N, el comportamiento del sistema puede analizarse por medio de las proporciones de los objetos en cada clase.

Sea $M_i^N(t)$ la proporción de objetos en la clase $i \in S$ al tiempo t, esto es,

$$M_i^N(t) := \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{\{X_n^N(t)=i\}}, \quad i \in S. \tag{1.4}$$

Denotamos por $M^N(t)$ al vector cuyas componentes son las proporciones $M_i^N(t)$:

$$M^N(t) = (M_1^N(t), M_2^N(t), \ldots, M_s^N(t)). \tag{1.5}$$

Podemos observar que la suma de las entradas del vector $M^N(t)$ es 1. En efecto,

$$\begin{aligned}
&M_1^N(t) + M_2^N(t) + \ldots + M_s^N(t)\\
&\qquad = \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{\{X_n^N(t)=1\}} + \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{\{X_n^N(t)=2\}} + \ldots + \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{\{X_n^N(t)=s\}}\\
&\qquad = \frac{1}{N}\left(\sum_{n=1}^{N} \mathbb{1}_{\{X_n^N(t)=1\}} + \sum_{n=1}^{N} \mathbb{1}_{\{X_n^N(t)=2\}} + \ldots + \sum_{n=1}^{N} \mathbb{1}_{\{X_n^N(t)=s\}}\right)\\
&\qquad = \frac{1}{N}(N) = 1.
\end{aligned}$$

Por lo tanto, como $M_i^N(t) \geq 0$ cada $M^N(t)$ define una medida de probabilidad en S. Por otra parte, de la definición de $M_i^N(t)$, podemos notar que al multiplicarlo por N obtenemos un número natural o cero. Es decir $(N)(M_i^N(t)) \in \mathbb{N}_0$ nos dice el número de objetos en cada clase i. Esto motiva que definamos el siguiente conjunto de medidas de probabilidad

$$\mathbb{P}_N(S) := \{p \in \mathbb{P}(S) : Np(i) \in \mathbb{N}_0, \forall i \in S\} \subset \mathbb{P}(S).$$

Observemos que $M^N(t) \in \mathbb{P}_N(S)$ y que $\mathbb{P}_N(S)$ es un conjunto finito.

Lema 1.2.1. *El proceso $\{M^N(t)\}_{t\in\mathbb{N}_0}$ es una cadena de Markov con espacio de estado $\mathbb{P}_N(S)$.*

La demostración del Lema 1.2.1 puede ser consultada en la Sección 1.5.

El Lema 1.2.1 nos permite demostrar la existencia de una función medible $H^N : \mathbb{P}_N(S) \times A \times \mathbb{R}^N \to \mathbb{P}_N(S)$ tal que

$$M^N(t+1) = H^N(M^N(t), a_t, w_t), \tag{1.6}$$

donde $\{w_t\}$ es una sucesión de vectores aleatorios i.i.d en $\mathbb{R}^N$ con distribución común θ. La demostración que presentamos a continuación esta basada en un proceso de simulación, adaptado a este contexto, de una cadena de Markov (ver [6]). El proceso $\{M^N(t)\}$ puede ser expresado como

$$M_i^N(t+1) = \frac{1}{N} \sum_{k=1}^{s} \sum_{n=1}^{NM_k^N(t)} \mathbb{1}_{A_{ki}(a_t)}(w_n^k(t)), \tag{1.7}$$

donde

$$A_{ki}(a) := [\Delta_{k(i-1)}(a), \Delta_{ki}(a)] \subseteq [0,1],$$

con

$$\begin{aligned} \Delta_{k0}(a) &\equiv 0, \\ \Delta_{ki}(a) &:= \sum_{l=1}^{i} K_{kl}(a), k, i \in S. \end{aligned}$$

y $w_n^k(t)$ son v.a.i.i.d. uniformes en $[0,1]$. Para cada $i \in S$ y $t \in \mathbb{N}_0$, denotemos

$$w^i(t) := (w_1^i(t), \ldots, w_{NM_i^N}^i(t)).$$

Es decir, $w^i(t)$, $i = 1, \ldots, s$ es un vector con los ruidos que hay en cada clase $i \in S$ en cada tiempo t. Similarmente, definimos

$$w_t := (w^1(t), \ldots, w^s(t)),$$

el cual es un vector que en cada entrada tiene al vector de ruidos por clase. Observemos que $\sum_{i=1}^{s} NM_i^N(t) = N$, por lo tanto $w_t \in [0,1]^N$. Esto implica que el número de variables aleatorias uniformes en (1.7) coincide con el número N. Definiendo (1.7) como

$$H_i^N(M_i^N(t), a_t, w_t) := \frac{1}{N} \sum_{k=1}^{s} \sum_{n=1}^{NM_k^N(t)} \mathbb{1}_{A_{ki}(a_t)}(w_n^k(t)), i \in S. \tag{1.8}$$

Definiendo

$$H^N(m,a,w) := (H_1^N(m,a,w), \ldots, H_s^N(m,a,w)), \quad (m,a,w) \in \mathbb{P}_N(S) \times A \times [0,1]^N.$$

Obtenemos la expresión

$$M^N(t+1) = H^N(M^N(t), a_t, w_t). \tag{1.9}$$

La demostración de la existencia de H^N fue realizada a través de un procedimiento constructivo, sin embargo, existen otras maneras de obtener la función H^N las cuales dependen de la situación particular a modelar (ver ejemplo del modelo de reforestación del Capítulo 4). No obstante, en el resto de este trabajo supondremos que la dinámica del sistema viene dada por la función H^N que determinamos en (1.8) y (1.9).

Observemos que $M^N(t)$ representa la configuración del sistema al tiempo t. Es decir, $M^N(t)$ proporciona la distribución de los objetos entre las clases. Por lo tanto, el costo por etapa dependerá de $M^N(t)$ y a_t. Definimos la función costo por etapa como una función

$$r : \mathbb{P}(S) \times A \to \mathbb{R}. \tag{1.10}$$

1.3. N-Modelo de control de Markov

Definimos el modelo de control de Markov asociado al sistema de N objetos (N-MCM) como

$$\mathcal{M}_N := (\mathbb{P}_N(S), A, H^N, \theta, r). \tag{1.11}$$

El modelo de control $\mathcal{M}_N$ representa un sistema estocástico controlado que evoluciona de la siguiente manera. Para cada tiempo $t \in \mathbb{N}_0$, cuando el sistema se encuentra en el estado $M^N(t) = m \in \mathbb{P}_N(S)$, el controlador elige la acción $a_t = a \in A$. Esto tiene como consecuencia lo siguiente:

1) Se incurre en un costo $r(m, a)$;

2) El sistema se mueve a un nuevo estado $M^N(t+1) = m'$ de acuerdo a la ley de transición

$$\begin{aligned} Q(B|m,a) &= \text{Prob}[M^N(t+1) \in B | M^N(t) = m, a_t = a], \\ &= \int_{\mathbb{R}^N} \mathbb{1}_B[H^N(m,a,w)]\theta(dw), \quad B \in \mathcal{B}(\mathbb{P}_N(S)), \end{aligned} \tag{1.12}$$

donde H^N es la función definida en (1.9).

3) Una vez que el sistema se encuentra en el estado m' se aplica un nuevo control $a' \in A$ y se repite el proceso. Los costos $\{r(M^N(t), a_t)\}$ se acumulan a lo largo del proceso de acuerdo a un criterio de optimalidad definido en la siguiente sección.

1.4. Optimalidad en el N-MCM

Recordemos que en cada etapa t del sistema, después de observar el estado actual $M^N(t)$, el controlador debe de tomar una acción a_t del conjunto de acciones A. En esta sección describimos cómo se determina el control en cada tiempo.

Sean $\mathbb{H}_0^N := \mathbb{P}_N(S)$ y $\mathbb{H}_t^N := (\mathbb{P}_N(S) \times A \times \mathbb{R}^N)^t \times \mathbb{P}_N(S)$, $t \geq 1$, el espacio de historias hasta el tiempo t. Un elemento en $\mathbb{H}_t^N$ es de la forma

$$h_t^N = (M^N(0), a_0, w_0, \ldots, M^N(t-1), a_{t-1}, w_{t-1}, M^N(t)),$$

donde $M^N(k) \in \mathbb{P}_N(S)$, $k = 0, 1, \ldots, t$.

Definición 1.4.1. (Política de control) *Una política de control es una sucesión* $\pi^N = \{\pi_t^N\}$ *de kérneles estocásticos* π_t^N *en* A *dado* $\mathbb{H}_t^N$ *tales que* $\pi_t^N(A|h_t^N) = 1$ *para cada* $h_t^N \in \mathbb{H}_t^N$, $t \in \mathbb{N}_0$.

Denotaremos por Π^N al conjunto de todas las políticas de control.

Sea $\mathbb{F}$ el conjunto de las funciones medibles $f : \mathbb{P}(S) \to A$ y $\mathbb{F}^N := \mathbb{F}\Big|_{\mathbb{P}_N(S)}$ la restricción de $\mathbb{F}$ en $\mathbb{P}_N(S)$. Entonces, $f \in \mathbb{F}^N$ es una función $f : \mathbb{P}_N(S) \to A$.

Definición 1.4.2. (Políticas de control de Markov)

(a) *Diremos que una política* $\pi^N \in \Pi^N$ *es una política determinista de Markov si existe una sucesión* $\{f_t^N\} \subset \mathbb{F}^N$ *tal que para cada* $t \in \mathbb{N}_0$ *y* $h_t^N \in \mathbb{H}_t^N$, $\pi_t^N(\cdot|h_t^N) = \delta_{f_t^N(M^N(t))}(\cdot)$. *En este caso,* π^N *toma la forma* $\pi^N = \{f_t^N\}$.

(b) *En particular, si* $f_t^N \equiv f^N$ *para alguna* $f^N \in \mathbb{F}^N$ *y para todo* $t \in \mathbb{N}_0$, *decimos que* π^N *es una política estacionaria.*

Denotaremos por Π_M^N al conjunto de políticas de Markov y abusando de la notación, usaremos $\mathbb{F}^N$ para denotar al conjunto de políticas estacionarias.

Observación 1.4.3. (a) *Cuando usamos* $\mathbb{F}$ *en lugar de* $\mathbb{F}^N$ *denotaremos por* Π_M *al conjunto de políticas de Markov deterministas, es decir,* Π_M *es la familia de sucesiones de funciones* $\{f_t\} \subset \mathbb{F}$. *Observemos que una política* $\pi = \{f_t\} \in \Pi_M$ *cuyos elementos están restringidos a* $\mathbb{P}_N(S)$ *se convierte en un elemento de* Π_M^N.

(b) *Siguiendo argumentos estándares de la teoría de control estocástico (e.g., ver [2, 3, 9, 21]) para cada* $\pi^N \in \Pi^N$ *y estado inicial* $M^N(0) = m \in \mathbb{P}_N(S)$, *existe un espacio* *de probabilidad* $(\Omega', \mathcal{F}', P_m^{\pi^N})$ *donde* $\Omega' := (\mathbb{P}_N(S) \times A \times \mathbb{R}^N)^\infty$, $\mathcal{F}'$ *su respectiva* σ*-álgebra producto y* $P_m^{\pi^N}$ *es una medida de probabilidad que satisface las siguientes propiedades. Para cada* $t \in \mathbb{N}_0$*:*

(i) $P_m^{\pi^N}(M^N(0) \in B) = \delta_m(B)$, $B \in \mathcal{B}(\mathbb{P}_N(S))$,

(ii) $P_m^{\pi^N}(a_t \in C|h_t^N) = \pi_t^N(C|h_t^N)$, $C \in \mathcal{B}(A)$,

(iii) *Propiedad de Markov:*

$$\begin{aligned} P_m^{\pi^N}[M^N(t+1) \in B|h_t^N, a_t] &= Q(B|M^N(t), a_t), \\ &= \int_{\mathbb{R}^N} \mathbb{1}_B[H^N(M^N(t), a_t, w)]\theta(dw), B \in \mathcal{B}(\mathbb{P}_N(S)). \end{aligned} \tag{1.13}$$

1.4.1. Criterio de optimalidad descontado y PCO

Consideremos el N Modelo de Control Markoviano $\mathcal{M}_N$. Para cada política de control $\pi^N \in \Pi^N$ y estado inicial $M^N(0) = m \in \mathbb{P}_N(S)$, definimos el costo total esperado descontado como

$$V^N(\pi^N, m) := E_m^{\pi^N} \sum_{t=0}^{\infty} \alpha^t r(M^N(t), a_t), \tag{1.14}$$

donde $\alpha \in (0,1)$ es el factor de descuento y $E_m^{\pi^N}$ representa el operador esperanza respecto a la medida de probabilidad $P_m^{\pi^N}$ inducida por la política π^N y el estado inicial $M^N(0) = m$ (ver Observación 1.4.3 (b)).

Por lo tanto, el problema de control óptimo para el sistema de N objetos (N-PCO) consiste en encontrar una política $\pi_*^N \in \Pi^N$ tal que

$$V_*^N(m) := \inf_{\pi^N \in \Pi^N} V^N(\pi^N, m) = V^N(\pi_*^N, m) \quad \forall m \in \mathbb{P}_N(S). \tag{1.15}$$

A la función V_*^N le llamaremos la N-función de valor. Este es precisamente el problema que estamos interesados en estudiar.

Observación 1.4.4. (Suficiencia de las políticas de Markov)
Es conocido (ver [9], p. 47, Teorema 4.2.3) que para cualquier $\pi^N \in \Pi^N$ *existe una política de Markov* $\psi^N \in \Pi_M^N$ *tal que*

$$V^N(\psi^N, m) \leq V^N(\pi^N, m), \quad m \in \mathbb{P}_N(S).$$

Por lo tanto, para resolver el N-PCO es suficiente considerar solo la clase de políticas de Markov Π_M^N.

1.4.2. Existencia de políticas óptimas

Con el fin de garantizar una solución al problema de control óptimo para el modelo $\mathcal{M}_N$, supondremos algunas condiciones de compacidad y continuidad las cuales proporcionan una caracterización de las políticas óptimas y de la N-función de valor en términos de la N-ecuación de optimalidad la cual introducimos a continuación (ver [10]).

Definición 1.4.5. *Una función medible* $u : \mathbb{P}_N(S) \to \mathbb{R}$ *se dice ser una solución a la ecuación de optimalidad con costo* α*-descontado (*N*-ecuación de optimalidad) para el* N *modelo si satisface*

$$u(m) = \min_{a\in A}\{r(m,a) + \alpha \int_{\mathbb{R}^N} u[H^N(m,a,w)]\theta(dw)\} \quad \forall m \in \mathbb{P}_N(S).$$

En esta sección, demostraremos que la N-función de valor V_*^N es una solución de la N-ecuación de optimalidad, i.e.,

$$V_*^N(m) = \min_{a\in A}\{r(m,a) + \alpha \int_{\mathbb{R}^N} V_*^N[H^N(m,a,w)]\theta(dw)\} \quad \forall m \in \mathbb{P}_N(S).$$

Hipótesis 1.4.6. (a) *El espacio de control* A *es un espacio compacto de Borel, cuya métrica será denotada por* d_A.

(b) *El mapeo* $a \mapsto K_{ij}(a)$ *definido en (1.1) es continuo para todo* $i, j \in S$.

(c) *La función de costo por etapa* r *es continua, acotada por alguna constante* $R > 0$ *y uniformemente Lipschitz con constante* L_r*; esto es,*

$$|r(m,a)| \leq R \quad \forall (m,a) \in \mathbb{P}(S) \times A, \tag{1.16}$$

y para cada $m, m' \in \mathbb{P}(S)$,

$$\sup_{a\in A} |r(m,a) - r(m',a)| \leq L_r ||m - m'||_\infty.$$

Una primer consecuencia de la Hipótesis 1.4.6 es la siguiente.

Proposición 1.4.7. *Bajo la Hipótesis 1.4.6, la función* $a \mapsto H^N(m,a,w)$ *definida en (1.6) es continua para toda* $m \in \mathbb{P}_N(S)$ *y* $w \in \mathbb{R}^N$.

Demostración. Sea $\{a_n\} \in A$ una suceción convergente a a^* cuando n tiende a infinito. Como A es compacto, esto implica que a^* pertenece a A. Como el mapeo $a \mapsto K_{ij}(a)$ es continuo (ver Hipótesis 1.4.6 (b)), tenemos que

$$\Delta_{ki}(a) = \sum_{l=1}^{i} K_{kl}(a), \;\; \forall k, i \in S,$$

es una suma de funciones continuas. Por lo tanto,

$$\Delta_{ki}(a_n) \to \Delta_{ki}(a^*) \;\forall k, i \in S,$$

cuando n tiende a infinito. De lo anterior tenemos que

$$A_{ki}(a_n) = [\Delta_{ki}(a_n), \Delta_{k(i+1)}(a_n)] \to A_{ki}(a^*) = [\Delta_{ki}(a^*), \Delta_{k(i+1)}(a^*)] \;\forall k, i \in S.$$

Sea $\delta_w(A_{ki}(a))$ la medida de Dirac correspondiente a la función $\mathbb{1}_{\{A_{ki}(a)\}}(w)$, para k,$i \in S$, $w \in [0, 1]$. Como $\delta_w(\cdot)$ es continua ya que es una medida de probabilidad, lo cual implica que

$$\delta_w(A_{ki}(a_n)) \to \delta_w(A_{ki}(a^*))$$

cuando n tiende a infinito. Esto implica para cada $i \in S$ y $t \in \mathbb{N}_0$, la función

$$\frac{1}{N} \sum_{k=1}^{s} \sum_{n=1}^{NM_k^N(t)} \mathbb{1}_{A_{ki}(a_t)}(w_n^k(t)),$$

es continua y por lo tanto H^N es continua. ∎

Adicionalmente, bajo la Hipótesis 1.4.6 y aplicando procedimientos estándares de la teoría de procesos de control de Markov (ver [10]) obtenemos el siguiente resultado.

Proposición 1.4.8. *Supongamos que la Hipótesis 1.4.6 se cumple. Entonces*

(a) *La N-función de valor V_*^N satisface la N-ecuación de optimalidad, esto es*

$$V_*^N(m) = \min_{a \in A}\{r(m, a) + \alpha \int_{\mathbb{R}^N} V_*^N[H^N(m, a, w)]\theta(dw)\}, \quad m \in \mathbb{P}_N(S). \tag{1.17}$$

Equivalentemente,

$$\min_{a \in A} \Phi^N(m, a) = 0, \quad m \in \mathbb{P}_N(S)$$

donde $\Phi^N(\cdot,\cdot)$ es la función de discrepancia definida como

$$\Phi^N(m,a) := r(m,a) + \alpha \int_{\mathbb{R}^N} V_*^N[H^N(m,a,w)]\theta(dw) - V_*^N(m). \tag{1.18}$$

Además,

$$V_*^N(m) \leq \frac{R}{1-\alpha}, \quad m \in \mathbb{P}_N(S).$$

(b) *Existe $f_*^N \in \mathbb{F}^N$ tal que $f_*^N(m) \in A$ alcanza el mínimo en (1.17), es decir,*

$$V_*^N(m) = r(m, f_*^N) + \alpha \int_{\mathbb{R}^N} V_*^N[H^N(m, f_*^N, w)]\theta(dw), \quad m \in \mathbb{P}_N(S), \tag{1.19}$$

o equivalentemente

$$\Phi^N(m, f_*^N) = 0,$$

y además, la política estacionaria $\pi_*^N = \{f_*^N\} \in \Pi_M^N$ es óptima para el modelo de control $\mathcal{M}_N$.

La Proposición 1.4.8 establece un procedimiento para resolver el N-PCO mediante la ecuación (1.17). No obstante, esta ecuación involucra una integral de un orden muy alto ($N \sim \infty$), la cual podría resultar difícil o incluso imposible de calcular. Debido a esto, es necesaria la búsqueda de métodos alternativos para resolver el problema de control óptimo de N objetos, lo cual será tratado en el próximo capítulo.

1.5. Demostración del Lema 1.2.1

Para una prueba más general de este resultado ver [6].

Sea $X^N := (X_1^N, \ldots, X_N^N) \in S^N$ el vector que indica la configuración del sistema objeto por objeto.

Observación 1.5.1.

(a) El proceso $\{X_n^N(t)\}_{t\in\mathbb{N}_0}$ es una cadena de Markov. Esto es debido a que $\{X_n^N(t)\}_{t\in\mathbb{N}_0}$ es una cadena de Markov para cada $n = 1, \ldots, N$. Más aún, por la independecia de las X_n^N se tiene que para

$$x = (x_1, \ldots, x_N), \; y = (y_1, \ldots, y_N) \in S^N,$$

$$\Lambda_{xy}^N(a) := P[X^N(t) = y | X^N(t-1) = x, a] = \prod_{n=1}^{N} K_{x_n y_n}(a). \tag{1.20}$$

(b) El vector $M^N := (M_1^N, \ldots, M_s^N)$ *es función de* X^N*. Sea* $\mu^N : S^N \to \mathbb{P}_N(S)$ *tal que* $\mu^N(X^N) = M^N$ *mediante la relación*

$$\mu^N(X^N)(i) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{\{X_n^N = i\}}, \quad i \in S. \tag{1.21}$$

(c) Si X^N *está determinada, entonces existe un único* M^N *tal que que* $M^N = \mu^N(X^N)$*. Esto no ocurre en el sentido inverso, esto es, si* M^N *esta fijo, el vector* X^N *no está determinado de manera única pero los* X^N *tales que* $M^N = \mu^N(X^N)$ *son iguales bajo permutaciones. Esto se demostrará en el Lema 1.5.2*

Sea G^N el conjunto de permutaciones de los N objetos.

Lema 1.5.2. *Sea* x*,* $y \in S^N$ *dos configuraciones tales que* $\mu^N(x) = \mu^N(y)$*, entonces existe* $\sigma \in G^N$ *tal que* $x = \sigma(y)$*.*

Demostración. La demostración es por inducción sobre N. Para $N = 1$ el resultado es trivial. Supongamos que el lema se satisface para $N - 1$. Sean $x, y \in S^N$ tales que $\mu^N(x) = \mu^N(y)$, entonces existe al menos una coordenada, digamos n, tal que $x_1 = y_n$ ya que deben tener el mismo número de objetos en cada clase. Sea $x' = (x_2, \ldots, x_N)$ y $y' = (y_1, \ldots, y_{n-1}, y_{n+1}, \ldots, y_N)$ entonces $\mu^{N-1}(x') = \mu^{N-1}(y')$. Por hipótesis de inducción, existe $\bar{\sigma} \in G^{N-1}$ tal que $x' = \bar{\sigma}(y')$. Definiendo la permutación $\sigma \in G^N$ como

$$\sigma(1) = n, \quad \sigma(j) = \bar{\sigma}(j) + \mathbb{1}_{\bar{\sigma}(j) > n} \text{ para } j \geq 2,$$

es claro que $x = \sigma(y)$. ■

Sea $f : S^N \to E$ (E arbitrario). Decimos que f es invariante bajo permutaciones si $f \circ \sigma = f$ para toda $\sigma \in G^N$.

El siguiente resultado es clave para la demostración de que $\{M^N(t)\}_{t \in \mathbb{N}_0}$ es cadena de Markov, pues establece que si una función definida en S^N es invariante bajo permutaciones, entonces está en función de μ^N.

Lema 1.5.3. *Si* $f : S^N \to E$ *es invariante bajo* G^N *entonces existe* $\bar{f} : \mathbb{P}_N(S) \to E$ *tal que* $f = \bar{f} \circ \mu^N$*.*

Demostración. Para cada $m \in \mathbb{P}_N(S)$, sea x_0 alguna configuración tal que $m = \mu^N(x_0)$ y definamos $\bar{f}(m) := f(x_0)$. Observe que está bien definida ya que si x es otra configuración tal que $m = \mu^N(x)$, por el Lema 1.5.2 existe $\sigma \in G^N$ tal que $x = \sigma(x_0)$ y tendríamos

$$\bar{f}(m) = f(x_0) = f(x) = \bar{f}(\mu^N(x)) = \bar{f}(m).$$

Como m es arbitrario se tiene que

$$f(x) = \bar{f}(\mu^N(x)).$$

Por lo tanto, $f = \bar{f} \circ \mu^N$. ∎

Demostración del Lema 1.2.1.

Sea $\{a_t\}_{t\in\mathbb{N}_0}$ una sucesión de acciones fija. Para probar que $\{M^N(t)\}$ es una cadena de Markov, demostraremos que para cualquier $\eta : S^N \to \mathbb{R}$ acotada se satisface

$$E\left(\eta(M^N(t))|M^N(0),\ldots,M^N(t-1)\right) = E(\eta(M^N(t))|M^N(t-1)). \tag{1.22}$$

La equivalencia de la condición (1.22) con la definición usual de cadena de Makov puede ser consultada en ([4], p.3).

Observemos que $E\left(\eta(M^N(t))|X^N(0),\ldots,X^N(t-1)\right)$ es función de $X^N(t-1)$ ya que $\{X^N(t)\}_{t\in\mathbb{N}_0}$ es cadena de Markov y M^N está en función solamente de X^N. Sea Ψ esta función, es decir,

$$\Psi(X^N(t-1)) := E\left(\eta(M^N(t))|X^N(0),\ldots,X^N(t-1)\right).$$

Notemos que Ψ es invariante bajo permutaciones. En efecto, sea $x \in S^N$

$$\begin{aligned}\Psi(x) &= E\left(\eta(M^N(t))|X^N(0),\ldots,X^N(t-1)=x\right), \\ &= \sum_{y\in S^N} \eta(\mu^N(y))\Lambda^N_{xy}(a),\end{aligned}$$

y

$$\begin{aligned}\Psi(\sigma(x)) &= \sum_{y\in S^N} \eta(\mu^N(y))\Lambda^N_{\sigma(x)y}(a),\\ &= \sum_{y\in S^N} \eta(\mu^N(\sigma(y)))\Lambda^N_{\sigma(x)\sigma(y)}(a),\\ &= \sum_{y\in S^N} \eta(\mu^N(y))\Lambda^N_{xy}(a),\end{aligned}$$

donde la última igualdad es debido a que μ^N es invariante bajo permutaciones. Por lo tanto, $\Psi(x) = \Psi(\sigma(x))$ y por el Lema 1.5.3, existe $\overline{\Psi}$ tal que $\Psi(x) = \overline{\Psi}(\mu^N(x))$ con lo cual tendríamos que

$$E\left(\eta(M^N(t))|X^N(0),\ldots,X^N(t-1)=x\right) = \overline{\Psi}(\mu^N(x)) = \overline{\Psi}(M^N(t-1)).$$

Tenemos que

$$\begin{aligned}&E\left(\eta(M^N(t))|M^N(0),\ldots,M^N(t-1)\right)\\ &\qquad = E\left(E\left(\eta(M^N(t))|X^N(0),\ldots,X^N(t-1)=x\right)|M^N(0),\ldots,M^N(t-1)\right),\\ &\qquad = E\left(\overline{\Psi}(M^N(t-1))|M^N(0),\ldots,M^N(t-1)\right),\\ &\qquad = \overline{\Psi}(M^N(t-1)). \end{aligned} \tag{1.23}$$

Similarmente

$$\begin{aligned}&E\left(\eta(M^N(t))|M^N(t-1)\right)\\ &\qquad = E\left(E\left(\eta(M^N(t))|X^N(0),\ldots,X^N(t-1)=x\right)|M^N(t-1)\right),\\ &\qquad = E\left(\overline{\Psi}(M^N(t-1))|M^N(t-1)\right),\\ &\qquad = \overline{\Psi}(M^N(t-1)). \end{aligned} \tag{1.24}$$

De (1.23) y (1.24) el resultado se sigue. ■

Capítulo 2

Control en sistemas de interacción de objetos bajo un esquema de campo medio

2.1. Introducción

Como fue mencionado anteriormente, a lo largo del capítulo trabajaremos con herramientas nuevas que nos permitirán resolver en cierto sentido el problema de control óptimo del modelo con N objetos. Esto se realizará a través de un nuevo modelo $\mathcal{M}$ que se obtiene como límite del modelo $\mathcal{M}_N$ cuando $N \to \infty$.

La idea para definir este nuevo modelo $\mathcal{M}$ surge de analizar el límite de los estados $M^N(t)$ cuando $N \to \infty$. Como supusimos anteriormente, N es un número del orden del infinito, por lo que esto podría sugerir que hacer tender N a infinito podría resultar en un modelo similar al original pero más sencillo en el sentido de que nos estaría exentando de estudiar elementos donde este involucrado el número N.

Por lo definido en el capítulo anterior, un elemento arbitrario en $\mathbb{P}_N(S)$ es un vector $M^N(t)$ cuyas entradas son

$$M_i^N(t) := \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{\{X_n^N(t)=i\}}, \quad i \in S.$$

Teniendo un tiempo t fijo, al hacer N tender a infinito, por un tipo de Ley de los Grandes Números tendremos que $M_i^N(t)$ converge a su media correspondiente, digamos $m_i(t)$. En otras palabras, para cada $t \in \mathbb{N}_0$ y para cada $i \in S$ existe m_i tal que

$$M_i^N(t) \to m_i(t).$$

Definimos $m := (m_1, \ldots, m_s)$. Notemos que este nuevo vector es una medida de probabilidad en $\mathbb{P}(S)$. En este sentido, podemos decir que $\mathbb{P}_N(S)$ converge de "cierta manera" a $\mathbb{P}(S)$.

Este procedimiento define el modelo límite $\mathcal{M}$ cuyos estados son precisamente las medidas $m \in \mathbb{P}(S)$.

Motivados por lo anterior, en este capítulo definimos formalmente los elementos del modelo límite $\mathcal{M}$ el cual es llamado modelo de campo medio. En principio, el modelo $\mathcal{M}$ tiene la ventaja que es independiente de N y además define un proceso de control determinista. Por lo tanto, el PCO en $\mathcal{M}$ será mucho más fácil de resolver que el original. Una vez definido el Modelo de Control de Markov $\mathcal{M}$ y planteado su respectivo problema de control óptimo, le impondremos las hipótesis necesarias que nos garanticen la existencia de una política óptima π_*. A partir de este hecho, nuestra teoría consiste en usar π_* para controlar el proceso original $\{M^N(t)\} \subset \mathbb{P}_N(S)$ y medir su desviación de optimalidad.

2.2. El modelo de control de campo medio

Sea $H : \mathbb{P}(S) \times A \to \mathbb{P}(S)$ definida como

$$H(m, a) := mK(a), \tag{2.1}$$

donde K es el kernel de transición de los objetos definido en (1.1). Supongamos que tenemos un sistema de control determinista $\{m(t)\}$ que toma valores en $\mathbb{P}(S)$, cuya evolución esta dada por la ecuación en diferencias

$$m(t+1) = H(m(t), a_t) \tag{2.2}$$

donde $m(0) = m \in \mathbb{P}(S)$ representa el estado inicial y $a_t \in A$ es el control al tiempo t. Observemos que $m(t+1)$ es un vector con componentes

$$m_j(t+1) = \sum_{i=1}^{s} m_i(t) K_{ij}(a_t) \tag{2.3}$$

Dado que el proceso $\{m(t)\}$ es determinista, este queda completamente determinado por medio de la sucesión de acciones $\{a_t\} \subset A$ y la condición inicial $m(0) = m \in \mathbb{P}(S)$. Debido a esto, podemos definir el modelo de control

$$\mathcal{M} = (\mathbb{P}(S), A, H, r),$$

donde r es la función de costo por etapa dada en (1.10).

Más adelante mostraremos que el proceso $\{M^N(t)\}$ definido en (1.9) converge en probabilidad a $\{m(t)\}$ donde $m(t)$ satisface (2.2). En este sentido al modelo $\mathcal{M}$ le llamaremos modelo de control de campo medio.

2.3. Optimalidad en el campo medio

Similar al trabajo realizado en el capítulo pasado, en esta sección daremos los elementos necesarios para plantear el problema de control óptimo del modelo de campo medio así como también para garantizar la existencia de soluciones al PCO correspondiente por medio de una ecuación funcional.

Recordando que $\mathcal{M}$ es un modelo determinista, la definición de política de control en esta clase de contextos se reduce al estudio de políticas en $\mathbb{F}$. Debido a esto, consideraremos al conjunto Π_M definido en la Observación 1.4.3 como el conjunto de las políticas de control para el modelo $\mathcal{M}$.

Dada una política de control $\pi \in \Pi_M$ y un estado inicial $m(0) = m \in \mathbb{P}(S)$, definimos el costo total descontado para el modelo de campo medio como

$$v(\pi, m) = \sum_{t=0}^{\infty} \alpha^t r(m(t), a_t). \tag{2.4}$$

Notemos que la definición del costo total descontado para campo medio no lleva el operador esperanza debido a que estamos trabajando con un sistema determinista. En este contexto, el problema de control de campo medio consiste en encontrar una política $\pi_* \in \Pi_M$ tal que

$$v_*(m) := \inf_{\pi \in \Pi_M} v(\pi, m) = v(\pi_*, m), \quad m \in \mathbb{P}(S), \tag{2.5}$$

donde v_* es la función de valor de campo medio. Decimos que π_* es la política óptima para el modelo de control de campo medio $\mathcal{M}$.

Por la compacidad del espacio A, la continuidad de la función de costo r y el hecho de que la función H es continua (ver Proposición 1.4.7) tenemos como consecuencia el siguiente resultado (ver [10])

Proposición 2.3.1. *Supongamos que se cumple la Hipótesis 1.4.6. Entonces:*

a) *La función de valor v_* satisface la ecuación de optimalidad de campo medio*

$$v_*(m) = \min_{a \in A}\{r(m,a) + \alpha v_*[H(m,a)]\}, \quad m \in \mathbb{P}(S). \tag{2.6}$$

Equivalentemente,

$$\min_{a \in A} \Phi(m,a) = 0, \quad m \in \mathbb{P}(S)$$

donde $\Phi(\cdot,\cdot)$ es la función de discrepancia definida como

$$\Phi(m,a) := r(m,a) + \alpha v_*[H(m,a)] - v_*(m). \tag{2.7}$$

Además,

$$|v_*(m)| \leq \frac{R}{1-\alpha}, \quad m \in \mathbb{P}(S)$$

donde R es la cota uniforme de la función de costo r.

b) *Existe $f^* \in \mathbb{F}$ que alcanza el mínimo en (2.6), es decir,*

$$v_*(m) = r(m,f^*) + \alpha v_*[H(m,f^*)], \quad m \in \mathbb{P}(S), \tag{2.8}$$

y la política estacionaria $\pi_* = \{f^*\} \in \Pi_M$ es óptima para el modelo de control $\mathcal{M}$.

Observación 2.3.2. *Sea $\{(m(t), a_t)\}$ una sucesión de pares estado-acción correspondiente a la aplicación de una política estacionaria $\pi_* = \{f^*\} \in \Pi_M$. Observemos que por el principio de optimalidad y argumentos de programación dinámica, π_* es una política óptima si y solo si $\Phi(m(t), f^*(m(t))) = 0$ para toda $t \in \mathbb{N}_0$.*

La función de valor y la política óptima están bien caracterizadas por la Proposición 2.3.1 y la Observación 2.3.2.

2.4. Convergencia en el campo medio

En la sección anterior logramos dar las condiciones necesarias para garantizar la existencia de soluciones para el problema de control óptimo del modelo de campo medio. En esta sección analizaremos la desviación de óptimalidad de la política óptima π_* en el modelo del campo medio $\mathcal{M}$ cuando esta es utilizada para controlar el proceso original $\{M^N(t)\}$. Utilizaremos la siguiente norma para calcular la distancia entre elementos de $\mathbb{P}(S)$. Para cada $t \in \mathbb{N}_0$

$$||M^N(t)-m(t)||_\infty = \text{máx}\{|M_1^N(t)-m_1(t)|, |M_2^N(t)-m_2(t)|, \ldots, |M_s^N(t)-m_s(t)|\}. \quad (2.9)$$

Por otro lado, de acuerdo a los resultados en las Proposiciones 1.4.8 y 2.3.1, podemos restringir nuestro estudio a la clase de todas las políticas de Markov Π_M (ver Observación 1.4.4).

A lo largo del capítulo supondremos que $M^N(0) = m(0) = m \in \mathbb{P}_N(S)$.

Notación. Para cada política fija $\pi = \{f_t\} \in \Pi_M$, denotamos

$$a_t^{\pi,N} := f_t(M^N(t)) \quad \text{y} \quad a_t^\pi := f_t(m(t))$$

las acciones al tiempo t correspondientes a la aplicación de la política π bajo los procesos $\{M^N(t)\}$ y $\{m(t)\}$ respectivamente.

A continuación mostraremos que el proceso $\{M^N(t)\}$ definido en (1.9) converge en probabilidad a $\{m(t)\}$ definido en (2.2).

Proposición 2.4.1. *Para cada $m \in \mathbb{P}_N(S)$, $T \in \mathbb{N}$ y $\epsilon > 0$, existen constantes positivas C, λ y $\gamma_T : \mathbb{R}_+ \to \mathbb{R}$ tales que*

$$\sup_{\pi\in\Pi_M} P_m^\pi \left\{ \sup_{0\le t\le T} ||M^N(t) - m(t)||_\infty \ge \gamma_T(\epsilon) \right\} \le CTe^{-\lambda N\epsilon^2}, \quad (2.10)$$

donde $\gamma_T(\epsilon) \to 0$ cuando $\epsilon \to 0$.

Demostración. Sea $\pi = \{f_t\} \in \Pi_M$ una política arbitraria y $M^N(0) = m \in \mathbb{P}_N(S) \subset \mathbb{P}(S)$ el estado inicial. Definimos

$$B_{ij,n}^N(t) := \mathbb{1}_{\{(A_{ij}(a_t^{\pi,N})\}}(w_n^i(t)), \quad i,j \in S, n \in \{1,\ldots,N\},$$

donde $w_n^i(t)$ son v.a.i.i.d. con distribución uniforme en $[0,1]$ (ver (1.7)). Observemos que para cada $t \in \mathbb{N}_0$, $\{B_{ij,n}^N(t)\}_{ij,n}$ son v.a.i.i.d. con distribución Bernoulli cuya media es

$$E_m^\pi[B_{ij,n}^N(t)|a_t^{\pi,N} = a] = K_{ij}(a) \quad i,j \in S, a \in A.$$

Entonces, para un $\epsilon > 0$ fijo, por la desigualdad de Hoeffding, tenemos para cada $i,j \in S$

$$P_m^\pi\left[\left|\sum_{n=1}^{NM_i^N(t)} B_{ij,n}^N(t) - NM_i^N(t)K_{ij}(a_t^{\pi,N})\right| < N\epsilon\right] > 1 - 2e^{-2N\epsilon^2}.$$

Consideremos el conjunto

$$\bar{\Omega} = \left\{\omega \in \Omega' \middle| \left|\sum_{n=1}^{NM_i^N(t)} B_{ij,n}^N(t) - NM_i^N(t)K_{ij}(a_t^{\pi,N})\right| < N\epsilon\right\} \subset \Omega'$$

(ver Observación 1.4.3(b)) y sea ϵ_t definido como

$$\epsilon_t := ||M^N(t) - m(t)||_\infty. \tag{2.11}$$

Así, de (1.7), (2.3) y (2.11), para $j \in S$, tenemos que las siguientes relaciones se cumplen en $\bar{\Omega}$:

$$\begin{aligned}
|M_j^N(t+1) - m_j(t+1)| &= \left|\sum_{i=1}^{s} \frac{1}{N}\left[\sum_{n=1}^{NM_i^N(t)} B_{ij,n}^N(t) - Nm_i(t)K_{ij}(a_t^{\pi,N})\right]\right|, \\
&\leq \sum_{i=1}^{s} \frac{1}{N}\left|\sum_{n=1}^{NM_i^N(t)} B_{ij,n}^N(t) - Nm_i(t)K_{ij}(a_t^{\pi,N})\right|, \\
&\leq \sum_{i=1}^{s} \frac{1}{N}\left|\sum_{n=1}^{NM_i^N(t)} B_{ij,n}^N(t) - NM_i^N(t)K_{ij}(a_t^{\pi,N})\right| \\
&+ \sum_{i=1}^{s} |M_i^N(t) - m_i(t)|K_{ij}(a_t^{\pi,N}), \\
&< s\epsilon + s\epsilon_t.
\end{aligned}$$

Dado que el lado derecho de la última desigualdad no depende de la clase j, tenemos que

$$||M^N(t+1) - m(t+1)||_\infty \leq s\epsilon + s\epsilon_t.$$

que es igual a

$$||M^N(t+1) - m(t+1)||_\infty \leq s\epsilon + s||M^N(t) - m(t)||_\infty.$$

Como $||M^N(0) - m(0)||_\infty = \epsilon_0 = 0$ y aplicando un procedimiento inductivo, un cálculo directo nos lleva a que en el conjunto $\bar{\Omega}$,

$$||M^N(t+1) - m(t+1)||_\infty < s\epsilon\beta_t, \quad t \in \mathbb{N}_0$$

donde $\{\beta_t\}$ es una sucesión creciente. Entonces, para un $T \in \mathbb{N}$ fijo,

$$||M^N(t+1) - m(t+1)||_\infty < s\epsilon\beta_T, \quad \forall 0 \leq t \leq T,$$

en el conjunto $\bar{\Omega}$. Por lo tanto, bajo la política $\pi \in \Pi_M$,

$$P_m^\pi\Big[\sup_{0\leq t\leq T} ||M^N(t+1) - m(t+1)||_\infty < s\epsilon\beta_T\Big] > 1 - 2e^{-2N\epsilon^2}.$$

De aquí, si hacemos $\gamma_T(\epsilon) := s\epsilon\beta_T$, $C = \lambda = 2$, tenemos que

$$\sup_{\pi\in\Pi_M} P_m^\pi\Big\{\sup_{0\leq t\leq T} ||M^N(t) - m(t)||_\infty \geq \gamma_T(\epsilon)\Big\} \leq CTe^{-\lambda N\epsilon^2}.$$

Finalmente, observemos que $\gamma_T(\epsilon) \to 0$, cuando $\epsilon \to 0$. ∎

Impondremos la siguiente hipótesis sobre la convergencia de campo medio.

Hipótesis 2.4.2. *Para N suficientemente grande se cumple que*

$$\begin{aligned}\beta_N &:= |r(m, a_t^{\pi,N}) - r(m, a_t^\pi)|, \\ &= |r(m, f_t(M^N(t))) - r(m, f_t(m(t)))| = O(\eta^{-N})\end{aligned}$$

para algún $\eta > 0$ y toda $\pi = \{f_t\} \in \Pi_M$.

La Hipótesis 2.4.2 dice que, considerando que $N \sim \infty$, el controlador elige la misma acción mediante una política de Markov $\pi = \{f_t\} \in \Pi_M$ en el campo medio, cuando el sistema se encuentra en estados muy cercanos.

Observación 2.4.3. *La Hipótesis 2.4.2 implica*

$$\bar{\beta}_N := \sup_{\pi\in\Pi_M} E_m^\pi \beta_N \to 0.$$

En su defecto, si la Hipótesis 2.4.2 no se cumple, impondremos la siguiente hipótesis.

Hipótesis 2.4.4. *La función de costo $r : \mathbb{P}(S) \times A \to \mathbb{R}$ satisface*

$$r(m,a) = r(m) \quad \forall a \in A.$$

Es decir, bajo la Hipótesis 2.4.4 estamos suponiendo que el costo por estapa depende de las acciones solo a través de la dinámica de los procesos $\{M^N(t)\}$ y $\{m(t)\}$. A continuación establecemos el resultado principal de este capítulo.

Teorema 2.4.5. *Supongamos que se cumplen las Hipótesis 1.4.6 y 2.4.2 o 2.4.4. Entonces:*

(a) *Para cada $T \in \mathbb{N}$, $0 \leq t \leq T$ y $m \in \mathbb{P}_N(S)$,*

$$\begin{aligned}\sup_{\varphi \in \Pi_M} & E_m^{\varphi} |V_*^N(M^N(t)) - v_*(m(t))| \\ & \leq \frac{2R\alpha^T}{1-\alpha} + \left(L_r\left(CTe^{-\lambda N \epsilon^2} + \gamma_T(\epsilon)\right) + \bar{\beta}_N\right)\left(\frac{1-\alpha^T}{1-\alpha}\right). \end{aligned} \tag{2.12}$$

(b) *La política de control $\pi_* \in \Pi_M$ que satisface (2.8) es asintóticamente óptima para el N-Modelo de control de Markov $\mathcal{M}_N$ cuando $N \to \infty$; esto es*

$$\lim_{N\to\infty} \Phi^N(m, f^*(m)) = 0, \tag{2.13}$$

donde

$$\Phi^N(m,a) := r(m,a) + \alpha \int_{\mathbb{R}^N} V_*^N[H^N(m,a,w)]\theta(dw) - V_*^N(m), \quad m \in \mathbb{P}_N(S) \tag{2.14}$$

es la función de discrepancia en el N-MCM $\mathcal{M}_N$, definida en (1.18)

En el resto de esta sección supondremos que se cumplen las Hipótesis 1.4.6 y 2.4.2 o 2.4.4. La demostración del Teorema 2.4.5 será consecuencia de los siguientes resultados.

Proposición 2.4.6. *Para cada $M^N(0) = m(0) = m \in \mathbb{P}_N(S)$ y $T \in \mathbb{N}$,*

$$\sup_{\pi \in \Pi_M} E_m^{\pi}\left[\sup_{0\leq t\leq T} ||M^N(t) - m(t)||_\infty\right] \leq CTe^{-\lambda N\epsilon^2} + \gamma_T(\epsilon). \tag{2.15}$$

Demostración. Dado que $M^N(t)$ y $m(t)$ son medidas de probabilidad, se sigue que para cada $\pi \in \Pi_M$ y $T \in \mathbb{N}$

$$M_T := \sup_{0\leq t\leq T} ||M^N(t) - m(t)||_\infty \leq 1. \tag{2.16}$$

Entonces, para cada $m \in \mathbb{P}_N(S)$, $\pi \in \Pi_M$ y $\epsilon > 0$ tenemos que

$$\begin{aligned} E_m^\pi[M_T] &= E_m^\pi[M_T \mathbb{1}_{\{M_T \geq \gamma_T(\epsilon)\}} + M_T \mathbb{1}_{\{M_T < \gamma_T(\epsilon)\}}], \\ &= E_m^\pi[M_T \mathbb{1}_{\{M_T \geq \gamma_T(\epsilon)\}}] + \gamma_T(\epsilon) P_m^\pi(M_T < \gamma_T(\epsilon)), \\ &= E_m^\pi[M_T \mathbb{1}_{\{M_T \geq \gamma_T(\epsilon)\}}] + \gamma_T(\epsilon). \end{aligned} \tag{2.17}$$

Por otra parte, tanto por (2.16) y por la no negatividad de M_T tenemos que

$$M_T \mathbb{1}_{\{M_T \geq \gamma_T(\epsilon)\}} \leq \mathbb{1}_{\{M_T \geq \gamma_T(\epsilon)\}}.$$

Usando esto último, la definición de M_T y la Proposición 2.4.1 tenemos que

$$E_m^\pi[M_T \mathbb{1}_{\{M_T \geq \gamma_T(\epsilon)\}}] \leq P_m^\pi(M_T \geq \gamma_T(\epsilon)) \leq CTe^{-\lambda N \epsilon^2}, \quad \pi \in \Pi_M.$$

Finalmente de (2.17) obtenemos

$$E_m^\pi[M_T] \leq CTe^{-\lambda N \epsilon^2} + \gamma_T(\epsilon), \quad \pi \in \Pi_M, m \in \mathbb{P}_N(S). \tag{2.18}$$

Tomando supremo sobre $\pi \in \Pi_M$ en (2.18) obtenemos el resultado. ∎

Para cualquier $\pi \in \Pi_M$, $m \in \mathbb{P}_N(S) \subset \mathbb{P}(S)$ y $T \in \mathbb{N}$, definimos el costo total α-descontado en T etapas para el modelo $\mathcal{M}_N$ y $\mathcal{M}$ como

$$V_T^N(\pi, m) := E_m^\pi\left[\sum_{k=0}^{T-1} \alpha^k r(M^N(k), a_k)\right] \quad \text{y} \quad v_T(\pi, m) := \sum_{k=0}^{T-1} \alpha^k r(m(t), a_k).$$

Proposición 2.4.7. *Sean L_r y R las constantes en Hipótesis 1.4.6(c). Entonces, para cada $m \in \mathbb{P}_N(S)$, $\epsilon > 0$, $T \in \mathbb{N}$ y $0 \leq t \leq T$, los siguientes enunciados se cumplen:*

(a)

$$\sup_{\pi\in\Pi_M} E_m^\pi |r(M^N(t), a_t^{\pi,N}) - r(m(t), a_t^\pi)| \leq L_r \left(CTe^{-\lambda N\epsilon^2} + \gamma_T(\epsilon)\right) + \bar{\beta}_N; \tag{2.19}$$

donde $M^N(t)$ y $m(t)$ son generado por una política arbitraria $\varphi \in \Pi_M$.

(b)

$$\sup_{\varphi\in\Pi_M} E_m^{\varphi}\left[\sup_{\pi\in\Pi_M}|V_T^N(\pi,M^N(t))-v_T(\pi,m(t))|\right]\leq$$
$$\left(L_r\left(CTe^{-\lambda N\epsilon^2}+\gamma_T(\epsilon)\right)+\bar{\beta}_N\right)\left(\frac{1-\alpha^T}{1-\alpha}\right); \quad (2.20)$$

(c)

$$\sup_{\varphi\in\Pi_M} E_m^{\varphi}\left[\sup_{\pi\in\Pi_M}|V^N(\pi,M^N(t))-V_T^N(\pi,M^N(t))|\right]\leq\frac{R\alpha^T}{1-\alpha}; \quad (2.21)$$

(d)

$$\sup_{\varphi\in\Pi_M} E_m^{\varphi}\left[\sup_{\pi\in\Pi_M}|v(\pi,m(t))-v_T(\pi,m(t))|\right]\leq\frac{R\alpha^T}{1-\alpha}. \quad (2.22)$$

Demostración. (a) Sea $\pi\in\Pi_M$ y $T\in\mathbb{N}$ fijas. De la Hipótesis 2.4.2 tenemos que

$$\begin{aligned}|r(M^N(t),a_t^{\pi,N})-r(m(t),a_t^{\pi})| &\leq |r(M^N(t),a_t^{\pi,N})-r(m(t),a_t^{\pi,N})|\\ &+|r(m(t),a_t^{\pi,N})-r(m(t),a_t^{\pi})|,\\ &\leq \sup_{a\in A}|r(M^N(t),a)-r(m(t),a)|\\ &+|r(m(t),a_t^{\pi,N})-r(m(t),a_t^{\pi})|,\\ &\leq L_r||M^N(t)-m(t)||_\infty+\beta_N.\end{aligned}$$

De la Hipótesis 1.4.6(c) y la Proposición 2.4.6 tenemos que

$$\begin{aligned}E_m^{\pi}|r(M^N(t),a_t^{\pi,N})-r(m(t),a_t^{\pi})| &\leq E_m^{\pi}\left(L_r||M^N(t)-m(t)||_\infty+\beta_N\right),\\ &\leq L_rE_m^{\pi}\left(\sup_{0\leq t\leq T}||M^N(t)-m(t)||_\infty\right)+E_m^{\pi}\beta_N,\\ &\leq L_r\left(CTe^{-\lambda N\epsilon^2}+\gamma_T(\epsilon)\right)+E_m^{\pi}\beta_N,\\ &\leq L_r\left(CTe^{-\lambda N\epsilon^2}+\gamma_T(\epsilon)\right)+\bar{\beta}_N \qquad (2.23)\end{aligned}$$

Por otra parte, si suponemos la Hipótesis 2.4.4,

$$\begin{aligned}E_m^{\pi}|r(M^N(t),a_t^{\pi,N})-r(m(t),a_t^{\pi})| &= E_m^{\pi}|r(M^N(t))-r(m(t))|,\\ &\leq L_rE_m^{\pi}(||M^N(t)-m(t)||_\infty),\\ &\leq L_r\left(CTe^{-\lambda N\epsilon^2}+\gamma_T(\epsilon)\right). \qquad (2.24)\end{aligned}$$

Por lo tanto, ya sea que estemos en el caso (2.23) o (2.24), concluimos

$$E_m^{\pi}|r(M^N(t), a_t^{\pi,N}) - r(m(t), a_t^{\pi})| \leq L_r \left(CTe^{-\lambda N\epsilon^2} + \gamma_T(\epsilon) \right) + \bar{\beta}_N. \tag{2.25}$$

Tomando supremo sobre $\pi \in \Pi_M$ llegamos al resultado deseado.

(b) Sea $\varphi \in \Pi_M$ una política arbitraria, definimos $M_{\varphi}^N(t) := M^N(t)$ y $m_{\varphi}(t) := m(t)$, las trayectorias generadas por φ. Para cada $\pi \in \Pi_M$,

$$\begin{aligned}
|V_T(\pi, M^N(t)) - v_T(\pi, m(t))| &= \left| E_{M^N(t)}^{\pi} \left\{ \sum_{k=0}^{T-1} \alpha^k r(M^N(k), a_k^{\pi,N}) - \sum_{k=0}^{T-1} \alpha^k r(m(k), a_k^{\pi}) \right\} \right|, \\
&\leq \sum_{k=0}^{T-1} \alpha^k E_{M^N(t)}^{\pi} |r(M^N(k), a_k^{\pi,N}) - r(m(t), a_k^{\pi})|, \\
&\leq \left(L_r \left(CTe^{-\lambda N\epsilon^2} + \gamma_T(\epsilon) \right) + \bar{\beta}_N \right) \left(\frac{1-\alpha^T}{1-\alpha} \right) \ P_m^{\varphi} - c.s.
\end{aligned}$$

donde la última igualdad es debido a (2.25). Por lo tanto

$$\begin{aligned}
&\sup_{\pi \in \Pi_M} |V_T(\pi, M^N(t)) - v_T(\pi, m(t))| \leq \\
&\qquad \left(L_r \left(CTe^{-\lambda N\epsilon^2} + \gamma_T(\epsilon) \right) + \bar{\beta}_N \right) \left(\frac{1-\alpha^T}{1-\alpha} \right) \ P_m^{\varphi} - c.s., \ \forall t \in \mathbb{N}_0.
\end{aligned} \tag{2.26}$$

Tomando esperanza E_m^{φ} en ambos lados y tomando supremo sobre $\varphi \in \Pi_M$, obtenemos (b).

(c) Similar al inciso anterior, para $\varphi \in \Pi_M$ una política arbitraria, definimos $M_{\varphi}^N(t) := M^N(t)$ y $m_{\varphi}(t) := m(t)$, las trayectorias generadas por φ. Para cada $\pi \in \Pi_M$, tenemos que

$$\begin{aligned}
&|V^N(\pi, M^N(t)) - V_T^N(\pi, M^N(t))| \\
&\qquad \leq \left| E_{M^N(t)}^{\pi} \left\{ \sum_{k=0}^{\infty} \alpha^k r(M^N(k), a_k^{\pi,N}) - \sum_{k=0}^{T-1} \alpha^k r(M^N(k), a_k^{\pi,N}) \right\} \right|, \\
&\qquad \leq \sum_{k=T}^{\infty} \alpha^k E_{M^N(t)}^{\pi} |r(M^N(k), a_k^{\pi,N})|, \\
&\qquad \leq R \sum_{k=T}^{\infty} \alpha^k, \\
&\qquad \leq \frac{R\alpha^T}{1-\alpha} \ P_m^{\varphi} - c.s..
\end{aligned} \tag{2.27}$$

Por argumentos similares a los realizados en (b), se sigue (c).

(d) Es análogo al procedimiento en (c). ■

2.4.1. Demostración del Teorema 2.4.5

(a) Sea $\pi_*^N = \{f_*^N\} \in \Pi_M^N$ una política óptima estacionaria para el N-MCM $\mathcal{M}$ y sea $\tilde{f} \in \mathbb{F}$ un selector arbitrario. Definimos la política estacionaria $\bar{\pi} = \{\bar{f}\} \in \Pi_M$, donde cada $\bar{f} : \mathbb{P}(S) \to A$ la definimos como

$$\bar{f}(m) = f_*^N(m)\mathbb{1}_{\mathbb{P}_N}(m) + \tilde{f}(m)\mathbb{1}_{[\mathbb{P}_N]^c}(m).$$

Sea $\varphi \in \Pi_M$ una política arbitraria y denotemos $M_\varphi^N(t) := M^N(t) \in \mathbb{P}_N(S)$ y $m_\varphi(t) := m(t) \in \mathbb{P}(S)$. Dado que la dinámica de las proporciones en el N-modelo vienen dadas por la función H^N introducida en (1.6), esto nos permite garantizar que el proceso $\{M^N(t)\}_{t\in\mathbb{N}}$ permanece en $\mathbb{P}_N(S)$ una vez dado el estado inicial $m \in \mathbb{P}_N(S)$.

Esto implica por la Observación 1.4.4 que dado un estado inicial $M^N(0) = m \in \mathbb{P}_N(S)$,

$$\inf_{\pi\in\Pi^N} V^N(\pi, m) = V_*^N(m) = V^N(\pi_*^N, m) = V^N(\bar{\pi}, m) = \inf_{\pi\in\Pi_M} V^N(\pi, m).$$

Es decir, la búsqueda del costo óptimo V_*^N no cambia al sustituir Π_M^N por Π_M. Debido a esto

$$\begin{aligned} |V_*^N(M^N(t)) - v_*(m(t))| &= |\inf_{\pi\in\Pi_M^N} V^N(\pi, M^N(t)) - \inf_{\pi\in\Pi_M} v(\pi, m(t))|, \\ &= |\inf_{\pi\in\Pi_M} V^N(\pi, M^N(t)) - \inf_{\pi\in\Pi_M} v(\pi, m(t))|, \\ &\leq \sup_{\pi\in\Pi_M} |V^N(\pi, M^N(t)) - v(\pi, m(t))|, \quad t \in \mathbb{N}_0. \end{aligned} \tag{2.28}$$

Por lo tanto, para cada $m \in \mathbb{P}_N(S)$ y $0 \leq t \leq T$,

$$
\begin{aligned}
&E_m^{\varphi}|V_*^N(M^N(t)) - v_*(m(t))| \\
&\leq E_m^{\varphi}\left[\sup_{\pi\in\Pi_M} |V^N(\pi, M^N(t)) - v(\pi, m(t))|\right], \\
&\leq E_m^{\varphi}\Bigg[\sup_{\pi\in\Pi_M} \Big\{|V^N(\pi, M^N(t)) - V_T^N(\pi, M^N(t))| \\
&+|V_T^N(\pi, M^N(t)) - v_T(\pi, m(t))| + |v_T(\pi, m(t)) - v(\pi, m(t))|\Big\}\Bigg], \\
&\leq E_m^{\varphi}\left[\sup_{\pi\in\Pi_M} |V^N(\pi, M^N(T)) - V_T^N(\pi, M^N(t))|\right] \\
&+E_m^{\varphi}\left[\sup_{\pi\in\Pi_M} |V_T^N(\pi, M^N(t)) - v_T(\pi, m(t))|\right] \\
&+E_m^{\varphi}\left[\sup_{\pi\in\Pi_M} |v_T(\pi, m(t)) - v(\pi, m(t))|\right] \\
&\leq \frac{2R\alpha^T}{1-\alpha} + \left(L_r\left(CTe^{-\lambda N\epsilon^2} + \gamma_T(\epsilon)\right) + E_m^{\varphi}\beta_N\right)\left(\frac{1-\alpha^T}{1-\alpha}\right), \\
&\leq \frac{2R\alpha^T}{1-\alpha} + \left(L_r\left(CTe^{-\lambda N\epsilon^2} + \gamma_T(\epsilon)\right) + \bar{\beta}_N\right)\left(\frac{1-\alpha^T}{1-\alpha}\right),
\end{aligned}
$$

donde la última desigualdad se debe a la Proposición 2.4.7. Tomando supremo sobre $\varphi \in \Pi_M$ obtenemos (2.16).

(b) Notemos primero que para cada $M^N(0) = m(0) = m \in \mathbb{P}_N(S) \subset \mathbb{P}(S)$, por la propiedad de Markov (ver Observación 1.4.3, (1.13)) tenemos

$$
\begin{aligned}
&\left|\int_{\mathbb{R}^N} (V_*^N[H^N(m, f^*, w)] - v_*[H(m, f^*)])\theta(dw)\right| \\
&\leq \int_{\mathbb{R}^N} \Big|V_*^N[H^N(m, f^*, w)] - v_*[H(m, f^*)]\Big|\theta(dw), \\
&= E_m^{\pi_*}\Big|V_*^N(M^N(1)) - v_*(m(1))\Big|, \\
&\leq \left(L_r\left(CTe^{-\lambda N\epsilon^2} + \gamma_T(\epsilon)\right) + E_m^{\pi_*}\beta_N\right)\left(\frac{1-\alpha^T}{1-\alpha}\right), \\
&\leq \left(L_r\left(CTe^{-\lambda N\epsilon^2} + \gamma_T(\epsilon)\right) + \bar{\beta}_N\right)\left(\frac{1-\alpha^T}{1-\alpha}\right). \qquad (2.29)
\end{aligned}
$$

donde la última desigualdad es por (2.26).

Por otra parte, de las definición de las funciones de discrepancia Φ^N y Φ dadas en (1.18) y (2.7) respectivamente, tenemos

$$\begin{aligned}
\Phi^N(m,f^*) &= \left|\Phi^N(m,f^*)-\Phi(m,f^*)\right|, \\
&= \Big|r(m,f^*)+\alpha\int_{\mathbb{R}^N} V_*^N[H^N(m,f^*,w)]\theta(dw)-V_*^N(m) \\
&\quad -(r(m,f^*)-\alpha v_*[H(m,f^*)]-v_*(m))\Big| \\
&\leq |V_*^N(m)-v_*(m)|+\alpha\Big|\int_{\mathbb{R}^N}(V_*^N[H^N(m,f^*,w)]-v_*[H(m,f^*)])\theta(dw)\Big|, \\
&\leq \sup_{\pi\in\Pi_M}|V^N(\pi,M^N(t))-v(\pi,m(t))| \\
&\quad +\alpha\Big|\int_{\mathbb{R}^N}(V_*^N[H^N(m,f^*,w)]-v_*[H(m,f^*)])\theta(dw)\Big| \quad \text{por (2.28)} \\
&\leq \sup_{\pi\in\Pi_M}\Big\{|V^N(\pi,M^N(t))-V_T^N(\pi,M^N(t))| \\
&\quad +|V_T^N(\pi,M^N(t))-v_T(\pi,m(t))|+|v_T(\pi,m(t))-v(\pi,m(t))|\Big\} \\
&\quad +\alpha\Big|\int_{\mathbb{R}^N}(V_*^N[H^N(m,f^*,w)]-v_*[H(m,f^*)])\theta(dw)\Big|, \\
&\leq \frac{2R\alpha^T}{1-\alpha}+L_r\frac{1-\alpha^T}{1-\alpha}[CTe^{-\lambda N\epsilon^2}+\gamma_T(\epsilon)] \\
&\quad +\alpha\Big|\int_{\mathbb{R}^N}(V_*^N[H^N(m,f^*,w)]-v_*[H(m,f^*)])\theta(dw)\Big|, \\
&\leq \frac{2R\alpha^T}{1-\alpha}+\left(L_r\left(CTe^{-\lambda N\epsilon^2}+\gamma_T(\epsilon)\right)+\bar\beta_N\right)\left(\frac{1-\alpha^T}{1-\alpha}\right) \\
&\quad +\left(L_r\left(CTe^{-\lambda N\epsilon^2}+\gamma_T(\epsilon)\right)+\bar\beta_N\right)\left(\frac{1-\alpha^T}{1-\alpha}\right).
\end{aligned}$$

donde la última desigualdad es por (2.26), (2.27) y (2.29). Entonces obtenemos

$$\begin{aligned}
\Phi^N(m,f^*) &\leq \frac{2R\alpha^T}{1-\alpha}+\left(L_r\left(CTe^{-\lambda N\epsilon^2}+\gamma_T(\epsilon)\right)+\bar\beta_N\right)\left(\frac{1-\alpha^T}{1-\alpha}\right) \\
&\quad +\left(L_r\left(CTe^{-\lambda N\epsilon^2}+\gamma_T(\epsilon)\right)+\bar\beta_N\right)\left(\frac{1-\alpha^T}{1-\alpha}\right).
\end{aligned}$$

Como ϵ y T son arbitrarios, tomando límite a ambos lados cuando $N\to\infty$ y considerando la Observación 2.4.3 obtenemos (2.13).

2.5. Algoritmo de iteración de valores para el modelo de campo medio

Hasta este punto, hemos podido simplificar el análisis del N Problema de Control Óptimo a través de la solución del Problema de Control Óptimo del modelo de campo medio. El objetivo de esta sección es presentar el Algoritmo de Iteración de Valores el cual es método iterativo para aproximar la función de valor en el campo medio v_* a través de la ecuación de optimalidad.

Como ha sido mencionado anteriormente en la Proposición 2.3.1, sabemos que la función de valor v_* es una solución de la ecuación de optimalidad determinista, es decir,

$$v_*(m) = \min_{a \in A}\{r(m,a) + \alpha v_*[H(m,a)]\} \quad \forall m \in \mathbb{P}(S).$$

A partir de aquí, para cada función u en $\mathcal{C}_+(\mathbb{P}(S))$ y $f \in \mathbb{F}$, definimos los operadores

$$\overline{T}u(m) := \inf_{a \in A}\{r(m,a) + \alpha u[H(m,a)]\}$$

y

$$\overline{T}_f u(m) := r(m,f) + \alpha u[H(m,f)].$$

Como consecuencia de la Hipótesis 1.4.6 y por medio de procedimientos estándares de la teoría de procesos de control de Markov (ver [9], p. 56) sabemos que $\overline{T}$ mapea $\mathcal{C}_+(\mathbb{P}(S))$ en sí mismo y es monótono. Aun más, existe existe $\tilde{f} \in \mathbb{F}$ tal que

$$\overline{T}u(m) = \overline{T}_{\tilde{f}}u(m), \;\; m \in \mathbb{P}(S). \tag{2.30}$$

Para establecer el teorema principal de la sección, definimos la sucesión de iteración de valores $\{v_n\} \in \mathcal{C}_+(\mathbb{P}(S))$ como:

$$\begin{aligned} v_0 &\equiv 0; \\ v_n(m) &= \overline{T}v_{n-1}(m), \;\; m \in \mathbb{P}(S). \end{aligned}$$

Como $\overline{T}$ es monótona, $\{v_n\}$ es una sucesión no decreciente. La continuidad de las funciones $\{v_n\}$ la obtenemos a partir de la continuidad de H (ver (2.1)) y la continuidad de la función de costos r.

Sea $f_n \in \mathbb{F}$ tal que

$$\begin{aligned} v_n(m) &= \overline{T} v_{n-1}(m), \\ &= \inf_{a\in A}\{r(m,a) + \alpha v_{n-1}[H(m,a)]\}, \\ &= r(m,f_n) + \alpha v_{n-1}[H(m,f_n)], \\ &= \overline{T}_{f_n} v_{n-1}(m),\ m \in \mathbb{P}(S). \end{aligned} \tag{2.31}$$

Definimos la política no estacionaria $\tilde{\pi} = \{f_n\}$. El objetivo es demostrar que $v_n \nearrow v_*$ cuando $n \to \infty$. A partir de este hecho, mostraremos que $\tilde{\pi}$ es cercanamente óptima, para n suficientemente grande. Esto lo establece el siguiente resultado.

Teorema 2.5.1. *Suponga que la Hipótesis 1.4.6 se cumple. Entonces:*

(a) $v_n \nearrow v_*$ *cuando* $n \to \infty$*. De hecho tenemos que*

$$\sup_{m\in\mathbb{P}(S)} |v_n(m) - v_*(m)| \to 0. \tag{2.32}$$

(b) *Para cada* $m \in \mathbb{P}(S)$*,* $\Phi(m, f_n(m)) \to 0$ *cuando* $n \to \infty$*, donde* Φ *es la función de discrepancia en el campo medio definida en (2.7).*

Demostración.
(a) Se sigue inmediatamente de ([9], p.58, Teorema 4.4.1).

(b) Para cada $n \in \mathbb{N}$, definimos la función

$$\Phi_n(m,a) := r(m,a) + \alpha v_{n-1}[H(m,a)] - v_n(m),\ (m,a) \in \mathbb{P}(S) \times A.$$

Observese que por (2.31),

$$\Phi_n(m, f_n) = 0\ \forall n \in \mathbb{N}, m \in \mathbb{P}(S).$$

Entonces

$$\begin{aligned} \Phi(m,f_n) &= \Phi(m,f_n) - \Phi_n(m,f_n), \\ &\leq \Big|\Phi(m,f_n) - \Phi_n(m,f_n)\Big|, \\ &\leq \Big|v_*[H(m,f_n)] - v_{n-1}[H(m,f_n)]\Big| + \Big|v_n(m) - v_*(m)\Big|, \\ &\leq \sup_{m\in\mathbb{P}(S)} \Big|v_*(m) - v_{n-1}(m)\Big| + \sup_{m\in\mathbb{P}(S)} \Big|v_*(m) - v_n(m)\Big|. \end{aligned}$$

Tomando límite cuando $n \to \infty$, por (2.32) obtenemos que $\Phi(m, f_n) \to 0$. ■

Capítulo 3

Sistema de N objetos con horizonte aleatorio

3.1. Introducción

En este capítulo supondremos que el sistema de N objetos introducido en el Capítulo 1 presenta una probabilidad positiva de existinguirse $1 - \alpha(M^N(t))$ en cada etapa t, la cual depende del estado. La evolución del sistema es la siguiente: al tiempo t, cuando el sistema está en el estado $M^N(t)$ y se elige el control $a_t \in A$, se genera un costo $r(M^N(t), a_t)$; entonces, hay una probabilidad positiva $1 - \alpha(M^N(t))$ de que el sistema se detenga (o “muera”), en otro caso, el sistema continuará a un nuevo estado $M^N(t+1)$ y el proceso se repetirá. En estas circunstancias, el sistema se analiza mediante un criterio de costo total con horizonte aleatorio $\tau = \tau(M^N(0), a_0, w_0, \ldots)$ de la forma

$$E \sum_{n=0}^{\tau} r(M^N(t), a_t) \tag{3.1}$$

donde r es la función costo por etapa. Dicho sistema estará definido en un nuevo modelo $\mathcal{M}^*$. Bajo este escenario, nuestro objetivo es demostrar la existencia de políticas óptimas.

Apoyándonos de [7, 17, 19] como antecedente, presentaremos una solución al PCO asociado a (3.1) demostrando que es equivalente a un índice de funcionamiento con horizonte infinito de la forma

$$E\left[\sum_{t=0}^{\infty} \prod_{k=0}^{t-1} \alpha(M^N(k)) r(M^N(t), a_t)\right]. \tag{3.2}$$

El índice (3.2) estará definido en un MCM que denotaremos por $\widetilde{\mathcal{M}}_N$, el cual compartirá muchos de los elementos del modelo $\mathcal{M}_N$. Esto nos permitirá utilizar las herramientas definidas en el Capítulo 1 y Capítulo 2.

Posteriormente, impondremos las condiciones necesarias para garantizar la solución al PCO asociado a (3.2) y recurriremos a la Teoría de Campo Medio para obtener una solución, siguiendo las ideas de los Capítulos 1 y 2.

3.2. Modelo de Control Markoviano para el sistema con N objetos y horizonte aleatorio

Consideremos el N Modelo de Control Markoviano (1.11), definido en el Capítulo 1,

$$\mathcal{M}_N := (\mathbb{P}_N(S), A, H^N, \theta, r). \tag{3.3}$$

Nuestro objetivo es demostrar la existencia de políticas óptimas para un PCO asociado a un índice de funcionamiento de la forma (3.1). Para ello, plantearemos un nuevo modelo $\mathcal{M}^*$, a partir de los elementos de $\mathcal{M}_N$, en el cual definiremos el problema de control óptimo que estamos interesados en resolver.

Definimos el modelo de control

$$\mathcal{M}^* := (\mathbb{P}_N(S)^*, A^*, Q^*, r^*, \alpha) \tag{3.4}$$

por los siguientes elementos. Sea $\mathbb{P}_N(S)^* = \mathbb{P}_N(S) \cup \{m^*\}$ y $A^* = A \cup \{a^*\}$ donde m^* es un estado ficticio que representa al estado donde el sistema se extingue y a^* una acción arbitraria correspondiente a m^*. Es decir, el conjunto de acciones admisibles para el estado $m \in \mathbb{P}_N(S)^*$ es

$$A^*(m) := \begin{cases} \{a^*\} & \text{si } m = m^*, \\ A & \text{si } m \in \mathbb{P}_N(S). \end{cases} \tag{3.5}$$

Además, $\alpha : \mathbb{P}(S) \to (0,1)$ es una función medible y la ley de transición Q^* es un kernel estocástico en $\mathbb{P}_N(S)^*$ dado el conjunto

$$\mathbb{K}^* := \{(m, a) : m \in \mathbb{P}_N(S)^*, a \in A^*(m)\}$$

definido a continuación: para $m \in \mathbb{P}_N(S)$ y $a \in A$,

$$\begin{aligned} Q^*(D|m,a) &:= \alpha(m)Q(D|m,a),\ D \in \mathcal{B}(\mathbb{P}_N(S)), \\ Q^*(\{m^*\}|m,a) &:= 1-\alpha(m),\ m \in \mathbb{P}_N(S),\ a \in A, \\ Q^*(\{m^*\}|m^*,a^*) &:= 1. \end{aligned}$$

Finalmente, la función de costo $r^* : \mathbb{K} \to \mathbb{R}$ estará dada por

$$r^*(m,a) := \begin{cases} r(m,a) & \text{si } (m,a) \in \mathbb{P}_N(S) \times A, \\ 0 & \text{si } (m,a) = (m^*,a^*). \end{cases} \tag{3.6}$$

El modelo $\mathcal{M}^*$ representa un sistema estocástico controlado que tiene la siguiente interpretación. Supongamos que al tiempo $t \in \mathbb{N}_0$ el estado del sistema es $M^N(t) = m \in \mathbb{P}_N(S)^*$. Posiblemente tomando en cuenta la historia del sistema, el controlador selecciona una acción $a_t = a \in A^*(m)$ y se genera un costo $r^*(m,a)$. Si $m = m^*$ tenemos que la única acción adimisible es $a = a^*$, y en este caso $r(m^*,a^*) = 0$. Si $m \in \mathbb{P}_N(S)$, entonces $a \in A$, y con probabilidad $1-\alpha(m)$ el sistema se detiene, es decir, el sistema se mueve al estado m^* y permanecerá allí con un costo cero. En otro caso, con probabilidad $\alpha(m)$ el sistema pasa a un nuevo estado $M^N(t+1) = m' \in \mathbb{P}_N(S)^*$ de acuerdo a la ley de transición Q^*. Una vez en este nuevo estado el proceso se repite.

Suponemos que existe $\gamma \in (0,1)$ tal que $1-\alpha(m) \geq \gamma$, $m \in \mathbb{P}(S)$. Entonces

$$\sup_{m \in \mathbb{P}(S)} \alpha(m) \leq 1-\gamma < 1. \tag{3.7}$$

La construcción del espacio de medida $(\Omega^*, \mathcal{F}^*)$ y los operadores P_m^{*,π^N} y E_m^{*,π^N} asociados al modelo $\mathcal{M}^*$ se sigue de manera usual (ver Observación 1.4.3).

Criterio de optimalidad. Definamos el tiempo de paro $\tau : \Omega^* \to \mathbb{N}_0 \cup \{+\infty\}$ como

$$\tau(M^N(0), a_0, w_0 \ldots) := \inf\{n \in \mathbb{N}_0 : M^N(n) = m^*\},$$

donde, como es usual, $\inf \emptyset = +\infty$. Para cada política $\pi^N \in \Pi^N$ y estado inicial $M^N(0) = m \in \mathbb{P}_N(S)^*$, el costo total esperado con horizonte aleatorio τ toma la forma

$$V_\tau(\pi^N, m) := E_m^{*,\pi^N} \sum_{t=0}^{\tau} r^*(M^N(t), a_t). \tag{3.8}$$

El problema de control óptimo asociado al modelo de control $\mathcal{M}^*$ consiste en encontrar una política óptima $\pi_*^N \in \Pi^N$ tal que

$$V_\tau(m) := \inf_{\pi^N \in \Pi^N} V_\tau(\pi^N, m) = V_\tau(\pi_*^N, m) \quad \forall m \in \mathbb{P}_N(S)^*.$$

A $V_\tau(\cdot)$ se le conoce como función de valor óptimo.

3.3. Planteamiento del problema en el modelo $\mathcal{M}_N$

En esta sección, introducimos un nuevo índice de funcionamiento $\widetilde{V}$, el cual estará definido por las componentes del modelo original $\mathcal{M}_N$, y demostraremos su equivalencia con V_τ. El hecho de lograr plantear el problema que estamos interesados en resolver en términos de los elementos del modelo $\mathcal{M}_N$ nos permite realizar el análisis de una manera más sencilla, ya que $\widetilde{V}$ es un índice con horizonte infinito, lo cual evita tratar con el tiempo de paro τ.

Para este fin, definamos

$$\Gamma_t^N := \prod_{k=0}^{t-1} \alpha(M^N(k)) \text{ si } n \geq 1, \text{ y } \Gamma_0 = 1. \tag{3.9}$$

Para una política $\pi^N \in \Pi_M^N$, dado un estado inicial $M^N(0) = m \in \mathbb{P}_N(S)$, definimos el índice de funcionamiento

$$\widetilde{V}(\pi^N, m) := E_m^{\pi^N} \left[\sum_{k=0}^{\infty} \Gamma_k^N r(M^N(k), a_k) \right] \tag{3.10}$$

donde $E_m^{\pi^N}$ denota el operador esperanza con respecto a la medida de probabilidad $P_m^{\pi^N}$ inducida por la política π^N, dado $M^N(0) = m \in \mathbb{P}_N(S)$, en el modelo original.

El problema de control óptimo asociado a este índice de funcionamiento es encontrar una política óptima $\pi_*^N \in \Pi_M^N$ tal que $\widetilde{V}(\pi_*^N, m) = \widetilde{V}(m)$ para toda $M^N(0) = m \in \mathbb{P}_N(S)$, donde

$$\widetilde{V}(m) := \inf_{\pi^N \in \Pi_M^N} \widetilde{V}(\pi^N, m) \tag{3.11}$$

es la función de valor óptimo.

En el siguiente resultado demostraremos que los índices (3.8) y (3.10) son equivalentes para parejas de estados y acción donde el sistema no se a extinguido.

Teorema 3.3.1. *Los índices V_τ y $\widetilde{V}$ son equivalentes, esto es,*

$$V_\tau(\pi^N, m) = \widetilde{V}(\pi^N, m) \ \forall m \in \mathbb{P}_N(S) \ y \ \pi^N \in \Pi_M^N,$$

es decir,

$$E_m^{*,\pi^N} \sum_{n=0}^{\tau} r^*(M^N(n), a_n) = E_m^{\pi^N} \sum_{n=0}^{\infty} \prod_{k=0}^{n-1} \alpha(M^N(k)) r(M^N(n), a_n) \tag{3.12}$$

donde $M^N(0) = m \in \mathbb{P}_N(S)$ y $\pi^N \in \Pi_M^N$.

Demostración. Demostraremos la equivalencia entre los índices de funcionamiento comprobando igualdad término a término de (3.12). Para $n = 0$ tenemos que

$$\begin{aligned}
E_m^{*,\pi^N} r^*(m, a_0) &= \int_{A^*} r^*(m, a_0)\pi_0^N(da_0|m), \\
&= \int_{A} r^*(m, a_0)\pi_0^N(da_0|m) \\
&= \int_{A} r(m, a_0)\pi_0^N(da_0|m), \\
&= E_m^{\pi^N} r(m, a_0),
\end{aligned}$$

Ahora, para $n = 1$ tenemos

$$\begin{aligned}
&E_m^{*,\pi^N} r^*(M^N(1), a_1) \\
&= \int_{A^*} \int_{\mathbb{P}_N(S)^*} \int_{A^*} r^*(M^N(1), a_1)\pi_1^N(da_1|M^N(1))Q^*(dM^N(1)|m, a_0)\pi_0^N(da_0|m) \\
&= \int_{A} \int_{\mathbb{P}_N(S)} \int_{A} r(M^N(1), a_1)\pi_1^N(da_1|M^N(1))\alpha(m)Q(dM^N(1)|m, a_0)\pi_0^N(da_0|m) \qquad (3.13) \\
&= \int_{A} \int_{\mathbb{P}_N(S)} \int_{A} \alpha(m)r(M^N(1), a_1)\pi_1^N(da_1|M^N(1))Q(dM^N(1)|m, a_0)\pi_0^N(da_0|m) \\
&= E_m^{\pi^N} \alpha(m)r(M^N(1), a_1),
\end{aligned}$$

donde (3.13) se sigue de

$$\begin{aligned}
&\int_{A^*}\int_{\mathbb{P}_N(S)^*}\int_{A} r^*(M^N(1),a_1)\pi_1^N(da_1|M^N(1))Q^*(dM^N(1)|m,a_0)\pi_0^N(da_0|m)\\
&\quad=\int_{A^*}\int_{\mathbb{P}_N(S)}\int_{A^*} r^*(M^N(1),a_1)\pi_1^N(da_1|M^N(1))Q^*(dM^N(1)|m,a_0)\pi_0^N(da_0|m)\\
&\quad+\int_{A^*}\int_{\{m^*\}}\int_{A^*} r^*(M^N(1),a_1)\pi_1^N(da_1|M^N(1))Q^*(dM^N(1)|m,a_0)\pi_0^N(da_0|m)\\
&\quad=\int_{A^*}\int_{\mathbb{P}_N(S)}\int_{A^*} r^*(M^N(1),a_1)\pi_1^N(da_1|M^N(1))Q^*(dM^N(1)|m,a_0)\pi_0^N(da_0|m)\\
&\quad+0 \quad \text{por (3.6)}.
\end{aligned}$$

Similarmente, podemos probar para $n = k$ que

$$E_m^{*,\pi^N} r^*(M^N(k),a_k) = E_m^{\pi^N}\prod_{t=0}^{k-1}\alpha(M^N(t))r(M^N(k),a_k).$$

Utilizando este procedimiento inductivo sobre las n, cada uno de los sumandos de V_τ pueden ser escritos como uno de $\widetilde{V}$, lo cual muestra (3.12). ∎

De lo anterior podemos concluir que el problema que estamos interesados en estudiar lo hemos transformado al análisis del índice de funcionamiento (3.10). Esto nos lleva a considerar el siguiente modelo de control

$$\widetilde{\mathcal{M}}_N := (\mathbb{P}_N(S), A, H^N, \theta, r, \alpha), \tag{3.14}$$

el cual esta asociado al problema de control óptimo planteado en (3.11).

En la siguiente sección daremos las condiciones bajo las cuales se garantiza una solución al Problema de Control Óptimo del modelo $\widetilde{\mathcal{M}}$.

3.4. El algoritmo de iteración de valores para el índice con descuento variable

En esta sección introduciremos condiciones que garantizan la existencia de políticas óptimas para el Problema de Control Óptimo del modelo $\widetilde{\mathcal{M}}_N$, las cuales se obtendran por medio del algoritmo de iteración de valores, probaremos que $\widetilde{V}$ es la solución de la ecuación de optimalidad y demostraremos la existencia de una política estacionaria óptima.

Definición 3.4.1. (Ecuación de optimalidad para $\widetilde{\mathcal{M}}_N$) *Una función medible $v : \mathbb{P}_N(S) \to \mathbb{R}$ se dice ser una solución a la ecuación de optimalidad para el modelo $\widetilde{\mathcal{M}}_N$ si satisface*

$$v(m) = \min_{a\in A}\{r(m,a) + \alpha(m)\int_{\mathbb{R}^N} v[H^N(m,a,w)]\theta(dw)\}, \quad \forall m \in \mathbb{P}_N(S).$$

Probaremos que la función de valor $\widetilde{V}$ es una solución de la ecuación de optimalidad, es decir,

$$\widetilde{V}(m) = \min_{a\in A}\{r(m,a) + \alpha(m)\int_{\mathbb{R}^N} \widetilde{V}[H^N(m,a,w)]\theta(dw)\}, \quad \forall m \in \mathbb{P}_N(S).$$

Hipótesis 3.4.2. *La función de descuento $\alpha : \mathbb{P}(S) \to (0,1)$ es Lipschitz con constante L_α; esto es, para cada m, $m' \in \mathbb{P}(S)$*

$$|\alpha(m) - \alpha(m')| \leq L_\alpha ||m - m'||_\infty. \tag{3.15}$$

En lo que resta del capítulo, supondremos adicionalmente que la Hipótesis 1.4.6 se cumple.

Observación 3.4.3.

(a) *Dado que $\mathbb{P}_N(S)$ es finito, entonces es un conjunto compacto. Esto implica que el espacio $\mathbb{P}_N(S) \times A$ es compacto.*

(b) *De la Hipótesis 1.4.6 y la condición (3.7) se sigue que existe una política π tal que $\widetilde{V}(\pi, m) < \infty$ para cada $m \in \mathbb{P}_N(S)$. De hecho tenemos $\widetilde{V}(\pi, m) \leq R/\gamma$ para todo $m \in \mathbb{P}_N(S)$ y $\pi \in \Pi_M^N$, donde R es la cota para r. Por lo tanto,*

$$\left|\widetilde{V}(m)\right| \leq \frac{R}{\gamma}, \;\; m \in \mathbb{P}_N(S). \tag{3.16}$$

(c) *Como la función H^N es continua (ver Proposición 1.4.7) tenemos que Q es débilmente continua, es decir, para una función u acotada y continua en $\mathbb{P}_N(S)$ (ver [9], p. 177, C.7) el mapeo*

$$(m,a) \to \int_{\mathbb{R}^N} u[H^N(m,a,w)]\theta(dw)$$

es continuo.

Para cada función medible u en $\mathbb{P}_N(S)$ y $f \in \mathbb{F}^N$, definimos los operadores

$$Tu(m) := \inf_{a\in A}\{r(m,a) + \alpha(m)\int_{\mathbb{R}^N} u[H^N(m,a,w)]\theta(dw)\},\ m \in \mathbb{P}_N(S)$$

y

$$T_f u(m) := r(m,f) + \alpha(m)\int_{\mathbb{R}^N} u[H^N(m,f,w)]\theta(dw),\ m \in \mathbb{P}_N(S).$$

Observemos que el operador T es monótono en el sentido de que si $v \geq u$, entonces $Tv \geq Tu$. Adicionalmente, si $u \in \mathcal{C}_+(\mathbb{P}_N(S))$, entonces de la Observación 3.4.3 (a), (c) y la continuidad de las funciones r y α, el operador T es continuo y por lo tanto tenemos que T mapea $\mathcal{C}_+(\mathbb{P}_N(S))$ en si mismo. Aún más, existe $\tilde{f} \in \mathbb{F}^N$ tal que

$$Tu(m) = T_{\tilde{f}}u(m),\ m \in \mathbb{P}_N(S). \tag{3.17}$$

Por otro lado, de (3.7), tenemos que $T : \mathcal{C}_+(\mathbb{P}_N(S)) \to \mathcal{C}_+(\mathbb{P}_N(S))$ es un operador de contracción con módulo $1-\gamma$. En efecto, para u, $u' \in \mathcal{C}_+(\mathbb{P}_N(S))$, y $m \in \mathbb{P}_N(S)$,

$$\begin{aligned} |Tu(m) - Tu'(m)| &\leq \sup_{a\in A}\left[\alpha(m)\int_{\mathbb{R}^N} |u(H^N(m,a,w)) - u'(H^N(m,a,w))|\theta(dw)\right], \\ &\leq (1-\gamma)||u-u'||. \end{aligned}$$

Esto implica que

$$||Tu - Tu'|| \leq (1-\gamma)||u-u'||.$$

Por lo tanto, por el Teorema de Punto Fijo de Banach, existe una única función $\tilde{u} \in \mathcal{C}_+(\mathbb{P}_N(S))$ tal que para todo $m \in \mathbb{P}_N(S)$,

$$\tilde{u}(m) = T\tilde{u}(m) \tag{3.18}$$

y

$$||T^n u - \tilde{u}|| \leq (1-\gamma)^n||u-\tilde{u}||\ \forall u \in \mathcal{C}_+(\mathbb{P}_N(S)),\ n \in \mathbb{N}_0. \tag{3.19}$$

Definimos la sucesión de funciones de iteración de valores $\{\tilde{v}_n\} \in \mathcal{C}_+(\mathbb{P}_N(S))$ como

$$\begin{aligned} \tilde{v}_0 &= 0, \\ \tilde{v}_n(m) &= T\tilde{v}_{n-1}(m),\ m \in \mathbb{P}_N(S),\ n \in \mathbb{N}_0. \end{aligned} \tag{3.20}$$

De (3.19) y (3.20), tomando $u = \tilde{v}_0 = 0$ obtenemos

$$||\tilde{v}_n - \tilde{u}|| \leq (1-\gamma)^n ||\tilde{u}|| \; \forall n \in \mathbb{N}_0. \tag{3.21}$$

El teorema principal de esta sección lo establecemos a continuación.

Teorema 3.4.4. *Supongamos que se cumple la Hipótesis 3.4.2. Entonces*

(a) *La función de valor $\widetilde{V}$ es la única solución en $\mathcal{C}_+(\mathbb{P}_N(S))$ que satisface la ecuación de optimalidad*

$$\widetilde{V}(m) = T\widetilde{V}(m), \;\; m \in \mathbb{P}_N(S).$$

(b) *Para cada $n \in \mathbb{N}$,*

$$||v_n - \widetilde{V}|| \leq \frac{R(1-\gamma)^n}{\gamma}.$$

Por lo tanto, $v_n \to \widetilde{V}$ cuando $n \to \infty$.

(c) *Existe $f^* \in \mathbb{F}^N$ tal que,*

$$\widetilde{V}(m) = r(m, f^*) + \alpha(m) \int_{\mathbb{R}^N} \widetilde{V}[H^N(m, f^*, w)]\theta(dw), \;\; m \in \mathbb{P}_N(S),$$

y f^* es una política óptima.

Demostración. De (3.18), (3.21) y (3.16), las partes (a) y (b) se obtienen si demostramos que $\tilde{u} = \widetilde{V}$.

Sea $f \in \mathbb{F}^N$ tal que, para $m \in \mathbb{P}_N(S)$

$$\tilde{u}(m) = r(m, f) + \alpha(m) \int_{\mathbb{R}^N} \tilde{u}[H^N(m, f, w)]\theta(dw). \tag{3.22}$$

Iterando (3.22) obtenemos

$$\begin{aligned}
\tilde{u}(m) &= E_m^f \left[r(M^N(0), f) + \sum_{t=1}^{n-1} \prod_{k=0}^{t-1} \alpha(M^N(k)) r(M^N(t), f) \right] \\
&\quad + E_m^f \left[\prod_{k=0}^{n-1} \alpha(M^N(k)) \tilde{u}(M^N(n)) \right], \\
&= E_m^f \left[\sum_{t=0}^{n-1} \Gamma_t^N r(M^N(t), f) \right] + E_m^f \left[\Gamma_n^N \tilde{u}(M^N(n)) \right]. \qquad (3.23)
\end{aligned}$$

Observemos que

$$E_m^f \left[\Gamma_n^N \tilde{u}(M^N(n)) \right] \leq (1-\gamma)^n ||\tilde{u}|| \to 0$$

cuando $n \to \infty$. Por lo tanto, tomando límite cuando $n \to 0$ en (3.23) obtenemos

$$\tilde{u}(m) = \widetilde{V}(f, m) \geq \widetilde{V}(m), \;\; m \in \mathbb{P}_N(S). \tag{3.24}$$

Por otro lado, observemos que se cumple, para todo $m \in \mathbb{P}_N(S)$ y $a \in A$,

$$\tilde{u}(m) \leq r(m, a) + \alpha(m) \int_{\mathbb{R}^N} \tilde{u}[H^N(m, a, w)]\theta(dw). \tag{3.25}$$

Similarmente como (3.23), para una política arbitaria $\pi^N \in \Pi_M^N$, iterando (3.25) obtenemos

$$\tilde{u}(m) \leq \widetilde{V}(\pi^N, m) \; \forall m \in \mathbb{P}_N(S). \tag{3.26}$$

Como π^N es arbitraria, concluimos que

$$\tilde{u} \leq \widetilde{V}(m) \; \forall m \in \mathbb{P}_N(S). \tag{3.27}$$

Por lo tanto, combinando (3.24) y (3.27) demostramos que $\tilde{u} = \widetilde{V}$.

La parte (c) se sigue de (a) y siguiendo un proceso similar de iteración como los anteriores. En efecto, existe $f^* \in \mathbb{F}^N$ tal que $\widetilde{V}(m) = T_{f^*}\widetilde{V}(m)$, $m \in \mathbb{P}_N(S)$. Iterando esta ecuación obtenemos $\widetilde{V}(m) = \widetilde{V}(f^*, m)$, es decir, f^* es óptima. ∎

3.5. Modelo de control de campo medio

Por los resultados obtenidos en el Teorema 3.4.4, podemos notar que, similarmente a lo hecho en el Capítulo 1, la solución del Problema de Control Óptimo del modelo $\widetilde{\mathcal{M}}_N$ nos quedó en términos de una integral de orden $N \sim \infty$. Esto nos lleva a plantearnos la búsqueda de una simplificación para brindar una solución computacionalmente razonable al PCO del modelo $\widetilde{\mathcal{M}}_N$.

Con este fin, realizaremos un procedimiento análogo al presentado en el Capítulo 2.

3.5.1. El modelo de control de campo medio con α dependiente del estado

Análogo al Capítulo 2, supongamos que $\{m(t)\}$ es un sistema de control determinista que toma valores en $\mathbb{P}(S)$. La evolución está dada por la ecuación en diferencias

$$m(t+1) = H(m(t), a_t)$$

donde $m(0) = m \in \mathbb{P}(S)$ representa el estado inicial, $a_t \in A$ es el control al tiempo t y $H : \mathbb{P}(S) \times A \to \mathbb{P}(S)$ es la función definida en (2.1).

Dado que el proceso $\{m(t)\}$ es determinista, este queda completamente determinado por la sucesión de acciones $\{a_t\} \in A$ y la condición inicial $m(0) = m \in \mathbb{P}(S)$.

Definimos el modelo de control

$$\mathcal{M}_\alpha := (\mathbb{P}(S), A, H, r, \alpha). \tag{3.28}$$

Observación 3.5.1. *En lo que resta de este trabajo, abreviaremos la frase "campo medio con α dependiente del estado" por "α-campo medio".*

3.5.2. Optimalidad en el α-campo medio

Para una política de control y un estado inicial, definimos el costo total descontado para el modelo α-campo medio como

$$\widetilde{v}(\pi, m) = \sum_{t=0}^{\infty} \Gamma_t r(m(t), a_t)$$

donde

$$\Gamma_t := \prod_{k=0}^{t-1} \alpha(m(k)) \text{ si } t \geq 1, \text{ y } \Gamma_0 = 1. \tag{3.29}$$

En este contexto, el problema de control α-campo medio consiste en encontrar una política $\varphi_* \in \Pi_M$ tal que

$$\widetilde{v}_*(m) := \inf_{\pi \in \Pi_M} \widetilde{v}(\pi, m) = \widetilde{v}(\varphi_*, m) \quad \forall m \in \mathbb{P}(S),$$

donde $\widetilde{v}_*$ es la función de valor del α-campo medio. Decimos que φ_* es la política óptima para el modelo de control del α-campo medio $\mathcal{M}_\alpha$.

Dado que $\mathcal{M}_\alpha$ es un modelo (determinista) más sencillo que el descrito en $\widetilde{\mathcal{M}}_N$, el siguiente resultado se sigue de los mismos argumentos presentados en la Sección 3.4 para el espacio $\mathcal{C}_+(\mathbb{P}(S))$ y considerando que $\mathbb{P}(S)$ es compacto al ser S finito (ver [20], p. 101).

Proposición 3.5.2. *Suponga que la Hipótesis 3.4.2 se cumple. Entonces*

(a) $\widetilde{v}_*$ *es la única solución de la ecuación de optimalidad del modelo* $\mathcal{M}_\alpha$*, es decir,*

$$\widetilde{v}_*(m) = T\widetilde{v}_*(m) = \inf_{a\in A}\{r(m,a) + \alpha(m)\widetilde{v}_*[H(m,a)]\} \quad \forall m \in \mathbb{P}(S).$$

(b) *Existe una política estacionaria* $f^* \in \mathbb{F}$ *tal que, para cada* $m \in \mathbb{P}(S)$, $\widetilde{v}_*(m) = T_{f^*}\widetilde{v}_*(m)$*, esto es,*

$$\widetilde{v}_*(m) = r(m,f^*) + \alpha(m)\widetilde{v}_*[H(m,f^*)]\}, \tag{3.30}$$

y f^* *es una política óptima.*

De la Proposición 3.5.2 y (3.7) tenemos un resultado similar a la Proposición 2.3.1.

- La función de valor $\widetilde{v}_*$ satisface que

$$\min_{a\in A} \Phi_\alpha(m) = 0, \quad m \in \mathbb{P}(S)$$

  donde Φ_α es la función de discrepancia definida como

$$\Phi_\alpha(m,a) := r(m,a) + \alpha(m)\widetilde{v}_*[H(m,a)] - \widetilde{v}_*(m). \tag{3.31}$$

- La función $\widetilde{v}_*$ satisface

$$|\widetilde{v}_*(m)| \leq \frac{R}{\gamma}$$

  donde R es la cota de la función de costo r.

3.5.3. Convergencia en el α-campo medio

En esta sección analizaremos la desviación de optimalidad de la política óptima φ_* del modelo $\mathcal{M}_\alpha$ cuando es utilizada en el proceso original.

Supondremos que las Hipótesis 1.4.6, 2.4.2 o 2.4.4 son válidas. El objetivo es demostrar un resultado análogo al Teorema 2.4.5. A pesar de que tenemos como ciertas las hipótesis necesarias para el Teorema 2.4.5, los resultados no son exactamente los mismos, ya que a diferencia del modelo $\mathcal{M}$, en el modelo $\mathcal{M}_\alpha$ se tiene la función α, lo que no ocurría en el modelo $\mathcal{M}$ del Capítulo 2.

Teorema 3.5.3. *Supongamos que se cumplen las Hipótesis 1.4.6 y 2.4.2 o 2.4.4. Entonces*

(a) *Para cada $T \in \mathbb{N}$, $0 \leq t \leq T$ y $m \in \mathbb{P}_N(S)$,*

$$\sup_{\pi\in\Pi_M} E_m^\pi|\widetilde{V}(M^N(t)) - \widetilde{v}_*(m(t))| \leq \frac{2R(1-\gamma)^T}{\gamma} + \left(\frac{1-(1-\gamma)^T}{\gamma}\right)\left[\left(CTe^{-\lambda N\varepsilon^2} + \gamma_T(\varepsilon)\right)(RL_\alpha T + L_r) + \bar{\beta}_N\right]. \quad (3.32)$$

(b) *La política de control $\pi_* \in \Pi_M$ que satisface (3.30) es asintóticamente óptima para de control de Markov $\widetilde{\mathcal{M}}_N$ cuando $N \to \infty$; esto es*

$$\lim_{N\to\infty} \Phi_\alpha^N(m, f^*(m)) = 0, \quad (3.33)$$

donde

$$\Phi_\alpha^N(m,a) := r(m,a) + \alpha(m)\int_{\mathbb{R}^N} \widetilde{V}[H^N(m,a,w)]\theta(dw) - \widetilde{V}(m), \quad m \in \mathbb{P}_N(S) \quad (3.34)$$

es la función de discrepancia en el $\widetilde{\mathcal{M}}_N$.

Para la demostración del Teorema 3.5.3 se aplicará aplicando un procedimiento análogo al presentado en el Teorema 2.4.5. Sin embargo, es necesario adaptar las desigualdades utilizadas en su demostración a este contexto. Para ello, lo que resta de la sección, nuestro objetivo sera mostrar los resultados necesarios para demostrar el Teorema 3.5.3.

Un resultado que es válido en este contexto y se sigue sin mayores cambios a los del Capítulo 2 es el siguiente.

Proposición 3.5.4. *Para cada $M^N \in \mathbb{P}_N(S)$ y $T \in \mathbb{N}$,*

$$\sup_{\pi\in\Pi_M} E_m^\pi\left[\sup_{0\leq t\leq T} ||M^N(t) - m(t)||_\infty\right] \leq CTe^{-\lambda N\epsilon^2} + \gamma_T(\epsilon). \quad (3.35)$$

Para cualquier $\pi \in \Pi_M$, $m \in \mathbb{P}_N(S) \subset \mathbb{P}(S)$ y $T \in \mathbb{N}$, definimos el costo total α-descontado en T etapas para el modelo $\widetilde{\mathcal{M}}_N$ y $\mathcal{M}_\alpha$ como

$$\widetilde{V}_T(\pi,m) := E_m^\pi\left[\sum_{k=0}^{T-1}\Gamma_k^N r(M^N(k),a_k)\right] \quad \text{y} \quad \widetilde{v}_T(\pi,m) := \sum_{k=0}^{T-1}\Gamma_k r(m(t),a_k).$$

Proposición 3.5.5. *Para cualquier $T \in \mathbb{N}$, $0 \leq k \leq T$, y $\pi \in \Pi_M$,*

$$E^{\pi}_{M^N(t)}|\Gamma^N_k - \Gamma_k| \leq (1-\gamma)^{k-1} L_\alpha k \left(CTe^{-\lambda N \varepsilon^2} + \gamma_T(\varepsilon) \right), \tag{3.36}$$

donde $M^N(t)$ es el estado al tiempo t correspondiente al uso de una política arbitraria $\pi' \in \Pi_M$.

Demostración. De (3.9) y (3.29), la desigualdad (3.36) es cierta para $k = 0$. Supongamos que (3.36) se cumple para $k = j$, con $0 \leq j \leq T-1$, es decir,

$$E^{\pi}_{M^N(t)} \left|\Gamma^N_j - \Gamma_j\right| \leq (1-\gamma)^{j-1} L_\alpha j \left(CTe^{-\lambda N \varepsilon^2} + \gamma_T(\varepsilon) \right). \tag{3.37}$$

Entonces

$$\begin{aligned}
E^{\pi}_{M^N(t)}|\Gamma^N_{j+1} - \Gamma_{j+1}| &= E^{\pi}_{M^N(t)} \left|\Gamma^N_j \alpha(M^N(j)) - \Gamma_j \alpha(m(j))\right| \\
&\leq E^{\pi}_{M^N(t)} \left|\Gamma^N_j \alpha(M^N(j)) - \Gamma^N_j \alpha(m(j))\right| \\
&\quad + E^{\pi}_{M^N(t)} \left|\Gamma^N_j \alpha(m(j)) - \Gamma_j \alpha(m(j))\right| \\
&\leq E^{\pi}_{M^N(t)} \Gamma^N_j \left|\alpha(M^N(j)) - \alpha(m(j))\right| \\
&\quad + E^{\pi}_{M^N(t)} \alpha(m(j)) \left|\Gamma^N_j - \Gamma_j\right|.
\end{aligned}$$

Notemos que de (3.7), tenemos que $\alpha(m(j)) \leq (1-\gamma)$ y $\Gamma^N_j \leq (1-\gamma)^j$. De (3.15) y (2.15) tenemos

$$\begin{aligned}
&E^{\pi}_{M^N(t)}|\Gamma^N_{j+1} - \Gamma_{j+1}| \\
&\qquad \leq (1-\gamma)^j L_\alpha E^{\pi}_{M^N(t)} ||M^N(j) - m(j)||_\infty + (1-\gamma) E^{\pi}_{M^N(t)} \left|\Gamma^N_j - \Gamma_j\right| \\
&\qquad \leq (1-\gamma)^j L_\alpha \left(CTe^{-\lambda N \varepsilon^2} + \gamma_T(\varepsilon) \right) + (1-\gamma)^j L_\alpha j \left(CTe^{-\lambda N \varepsilon^2} + \gamma_T(\varepsilon) \right) \\
&\qquad = (1-\gamma)^j L_\alpha (j+1) \left(CTe^{-\lambda N \varepsilon^2} + \gamma_T(\varepsilon) \right).
\end{aligned}$$

Por lo tanto, (3.36) es válido para k tal que $0 \leq k \leq T$. ∎

Observación 3.5.6. *La Proposición 3.5.5 implica que $E^{\pi}_{M^N(t)}|\Gamma^N_j - \Gamma_j| \to 0$ cuando $N \to \infty$.*

Proposición 3.5.7. *Sean L_r y R las constantes en Hipótesis 1.4.6(c). Entonces, para cada $m \in \mathbb{P}_N(S)$, $\epsilon > 0$, $T \in \mathbb{N}$ y $0 \leq t \leq T$, los siguientes enunciados se cumplen:*

(a)

$$\sup_{\pi\in\Pi_M} E_m^\pi |r(M^N(t), a_t^{\pi,N}) - r(m(t), a_t^\pi)| \le L_r \left(CTe^{-\lambda N\epsilon^2} + \gamma_T(\epsilon) \right) + \bar{\beta}_N \quad (3.38)$$

donde $M^N(t)$ *y* $m(t)$ *son generados por una política arbitraria* $\varphi \in \Pi_M$.

(b)

$$\sup_{\varphi\in\Pi_M} E_m^\varphi \left[\sup_{\pi\in\Pi_M} |\widetilde{V}_T(\pi, M^N(t)) - \widetilde{v}_T(\pi, m(t))| \right] \le \left(\frac{1-(1-\gamma)^T}{\gamma} \right) \left[\left(CTe^{-\lambda N\varepsilon^2} + \gamma_T(\varepsilon) \right) (RL_\alpha T + L_r) + \bar{\beta}_N \right]; \quad (3.39)$$

(c)

$$\sup_{\varphi\in\Pi_M} E_m^\varphi \left[\sup_{\pi\in\Pi_M} |\widetilde{V}(\pi, M^N(t)) - \widetilde{V}_T(\pi, M^N(t))| \right] \le \frac{R(1-\gamma)^T}{\gamma}; \quad (3.40)$$

(d)

$$\sup_{\varphi\in\Pi_M} E_m^\varphi \left[\sup_{\pi\in\Pi_M} |\widetilde{v}(\pi, m(t)) - \widetilde{v}_T(\pi, m(t))| \right] \le \frac{R(1-\gamma)^T}{\gamma}. \quad (3.41)$$

Demostración.
(a) Se sigue de los mismos argumentos que los presentados en la Proposición 2.4.7.

(b) Sea $\varphi \in \Pi_M$ una política arbitraria, definimos $M_\varphi^N(t) := M^N(t)$ y $m_\varphi(t) := m(t)$, las trayectorias generadas por φ. Para cada $\pi \in \Pi_M$,

$$
\begin{aligned}
|\widetilde{V}_T(\pi, M^N(t)) - \widetilde{v}_T(\pi, m(t))| &= \left| E^{\pi}_{M^N(t)} \left\{ \sum_{k=0}^{T-1} \Gamma^N_k r(M^N(k), a^{\pi,N}_k) - \sum_{k=0}^{T-1} \Gamma_k r(m(k), a^{\pi}_k) \right\} \right|, \\
&\leq \left| E^{\pi}_{M^N(t)} \left\{ \sum_{k=0}^{T-1} \Gamma^N_k r(M^N(k), a^{\pi,N}_k) - \sum_{k=0}^{T-1} \Gamma_k r(M^N(k), a^{\pi,N}_k) \right\} \right| \\
&+ \left| E^{\pi}_{M^N(t)} \left\{ \sum_{k=0}^{T-1} \Gamma_k r(M^N(k), a^{\pi,N}_k) - \sum_{k=0}^{T-1} \Gamma_k r(m(k), a^{\pi}_k) \right\} \right|, \\
&\leq \left| E^{\pi}_{M^N(t)} \left\{ \sum_{k=0}^{T-1} (\Gamma^N_k - \Gamma_k) r(M^N(k), a^{\pi,N}_k) \right\} \right| \\
&+ \left| E^{\pi}_{M^N(t)} \left\{ \sum_{k=0}^{T-1} \Gamma_k (r(M^N(k), a^{\pi,N}_k) - r(m(k), a^{\pi}_k)) \right\} \right|, \\
&\leq \left| E^{\pi}_{M^N(t)} \left\{ \sum_{k=0}^{T-1} (\Gamma^N_k - \Gamma_k) R \right\} \right| \\
&+ \left| E^{\pi}_{M^N(t)} \left\{ \sum_{k=0}^{T-1} (1-\gamma)^k (r(M^N(k), a^{\pi,N}_k) - r(m(k), a^{\pi}_k)) \right\} \right|, \\
&\leq R E^{\pi}_{M^N(t)} \left\{ \sum_{k=0}^{T-1} \left| \Gamma^N_k - \Gamma_k \right| \right\} \\
&+ E^{\pi}_{M^N(t)} \left\{ \sum_{k=0}^{T-1} (1-\gamma)^k \left| (r(M^N(k), a^{\pi,N}_k) - r(m(k), a^{\pi}_k)) \right| \right\}, \\
&\leq R E^{\pi}_{M^N(t)} \left\{ \sum_{k=1}^{T-1} \left| \Gamma^N_k - \Gamma_k \right| \right\} \\
&+ E^{\pi}_{M^N(t)} \left\{ \sum_{k=0}^{T-1} (1-\gamma)^k \left| (r(M^N(k), a^{\pi,N}_k) - r(m(k), a^{\pi}_k)) \right| \right\}, \\
&\leq R \sum_{k=1}^{T-1} (1-\gamma)^{k-1} L_\alpha k \left(C T e^{-\lambda N \epsilon^2} + \gamma_T(\epsilon) \right) \\
&+ \left(L_r \left(C T e^{-\lambda N \epsilon^2} + \gamma_T(\epsilon) \right) + E^{\pi}_m \beta_N \right) \left(\frac{1-(1-\gamma)^T}{\gamma} \right), \quad (3.42) \\
&\leq R \left(\frac{1-(1-\gamma)^{T-1}}{\gamma} \right) L_\alpha T \left(C T e^{-\lambda N \epsilon^2} + \gamma_T(\epsilon) \right) \\
&+ \left(L_r \left(C T e^{-\lambda N \epsilon^2} + \gamma_T(\epsilon) \right) + E^{\pi}_m \beta_N \right) \left(\frac{1-(1-\gamma)^T}{\gamma} \right) \\
&\leq R \left(\frac{1-(1-\gamma)^{T}}{\gamma} \right) L_\alpha T \left(C T e^{-\lambda N \epsilon^2} + \gamma_T(\epsilon) \right) \\
&+ \left(L_r \left(C T e^{-\lambda N \epsilon^2} + \gamma_T(\epsilon) \right) + E^{\pi}_m \beta_N \right) \left(\frac{1-(1-\gamma)^T}{\gamma} \right)
\end{aligned}
$$

donde la desigualdad (3.42) es debido a (3.38) y la Proposición 3.5.5. Por lo tanto

$$
\begin{aligned}
\sup_{\pi\in\Pi_M} |\widetilde{V}_T(\pi, M^N(t)) - \widetilde{v}_T(\pi, m(t))| \leq \\
& R\left(\frac{1-(1-\gamma)^T}{\gamma}\right) L_\alpha T \left(CTe^{-\lambda N\epsilon^2} + \gamma_T(\epsilon)\right) \\
& + \left(L_r\left(CTe^{-\lambda N\epsilon^2} + \gamma_T(\epsilon)\right) + E_m^\pi \beta_N\right)\left(\frac{1-(1-\gamma)^T}{\gamma}\right), \\
& = \left(\frac{1-(1-\gamma)^T}{\gamma}\right)\left[\left(CTe^{-\lambda N\varepsilon^2} + \gamma_T(\varepsilon)\right)(RL_\alpha T + L_r) + \bar{\beta}_N\right], \ P_m^\varphi - c.s. \quad (3.43)
\end{aligned}
$$

Tomando esperanza E_m^φ en ambos lados y tomando supremo sobre $\varphi \in \Pi_M$, obtenemos (b).

(c) Similar al inciso anterior, para $\varphi \in \Pi_M$ una política arbitraria, definimos $M_\varphi^N(t) := M^N(t)$ y $m_\varphi(t) := m(t)$, las trayectorias generadas por φ. Para cada $\pi \in \Pi_M$, tenemos que

$$
\begin{aligned}
|\widetilde{V}(\pi, M^N(t)) - \widetilde{V}_T^N(\pi, M^N(t))| \\
& \leq \left| E_{M^N(t)}^\pi \left\{ \sum_{k=0}^{\infty} \Gamma_k^N r(M^N(k), a_k^{\pi,N}) - \sum_{k=0}^{T-1} \Gamma_k^N r(M^N(k), a_k^{\pi,N}) \right\} \right|, \\
& \leq \sum_{k=T}^{\infty} E_{M^N(t)}^\pi \Gamma_k^N |r(M^N(k), a_k^{\pi,N})|, \\
& \leq \sum_{k=T}^{\infty} E_{M^N(t)}^\pi (1-\gamma)^k |r(M^N(k), a_k^{\pi,N})|, \\
& \leq R\sum_{k=T}^{\infty} (1-\gamma)^k, \\
& \leq \frac{R(1-\gamma)^T}{\gamma}. \ P_m^\varphi - c.s. \quad (3.44)
\end{aligned}
$$

Por un procedimiento similar al realizado en (b), se sigue (c).

(d) Es análogo al procedimiento en (c). ∎

La demostración del Teorema 3.5.3 se sigue de los mismos razonamientos presentados en el Teorema 2.4.5, haciendo uso de las cotas presentadas en la Proposición 3.5.7.

Capítulo 4

Ejemplos

4.1. Introducción

En este capítulo presentaremos una serie de ejemplos en los cuales se ilustran sistemas de interacción de objetos en el contexto de la teoría presentada en los Capítulos 1 y 2.

Dichos ejemplos abarcan diferentes áreas de aplicación, como finanzas, epidemiología, reforestación y estrategias de mercado. En todos ellos mostraremos que las hipótesis usadas a lo largo del trabajo se satisfacen. Así mismo, se pondrá mayor énfasis en el análisis de la dinámica de las proporciones $M^N(t)$ y la del campo medio dada por las relaciones

$$\begin{aligned} H(m,a) &:= mK(a), && (4.1)\\ m(t+1) &= H(m(t),a_t), && (4.2)\\ m_j(t+1) &= \sum_{i=1}^{s} m_i(t)K_{ij}(a_t), && (4.3) \end{aligned}$$

donde $m(0) = m \in \mathbb{P}(S)$ representa el estado inicial y $a_t \in A$ es el control al tiempo t.

Finalmente, con el fin de ilustrar el método de solución, resolvemos el problema de control en el ejemplo de estrategias de mercado aplicando el algoritmo de iteración de valores.

4.2. Modelo de consumo-inversión

En esta sección presentaremos un ejemplo de un sistema de N objetos los cuales tienen una dinámica en ecuaciones en diferencias estocásticas. Este tipo de dinámicas fue mencionado al inicio del Capítulo 1.

Para ver una formulación más general de este ejemplo, ver [11] o [13].

Consideremos un sistema de consumo-inversión donde los N objetos representan a pequeños inversionistas. Supondremos que las decisiones de este tipo de agentes económicos presentan un impacto imperceptible al mercado de precios. Estos agentes invierten en diferentes activos con diferentes tasas de retorno y consumen algún producto específico.

En este contexto, consideramos al gobierno como el controlador central, el cual incentiva a los pequeños inversionistas o impone impuestos que estos deben pagar. Supondremos que los inversionistas cuentan solamente con dos activos. El primero con una tasa libre de riesgo, es decir, con tasa de retorno fijo λ y el segundo tiene una tasa de retorno aleatoria ξ_t que toma valores en un conjunto acotado $Z \subset \mathbb{R}$.

Por simplicidad y dado que es deseado que el espacio de estados S sea finito, supondremos que el uso de los centavos es despreciable. Sea a_t la decisión del controlador central en el tiempo t, y supondremos que $a_t \in A := \{0, \pm 1, \ldots, \pm a^*\}$ para algún $a^* \geq 0$. Para cada t, tenemos

$$a_t = \begin{cases} \text{impuesto de tamaño } -a_t & \text{si } a_t < 0 \\ \text{subsidio de tamaño } a_t & \text{si } a_t \geq 0 \end{cases}$$

Sea $X_n^N(t) \in S := \{0, 1, \ldots, s\}$ el capital del inversionista n al tiempo $t \in \mathbb{N}$. Además, la fracción del capital que se invierte en el bien con riesgo es una función $\varphi_1 : S \to [0, 1]$. Por lo tanto, $1 - \varphi_1$ representa la fracción del capital que se invierte libre de riesgo.

Supondremos que la cantidad que consume cada inversionista es una función acotada $\varphi_2 : S \to \mathbb{R}_+$. Vamos a suponer que, en cada tiempo, el gobierno no impone impuestos mayores al capital de los inversionistas y que éstos no pueden consumir un monto mayor que su capital adicionado al subsidio/impuesto que les da el gobierno.

Tomando en cuenta lo anterior, el proceso $\{X_n^N\}$ evoluciona de acuerdo a la siguiente ecuación en diferencia

$$X_n^N(t+1) = \text{int}\Big\{[(1 - \varphi_1(X_n^N(t)))(1 + \lambda) + \varphi_1(X_n^N(t))\xi_t][X_n^N(t) - \varphi_2(X_n^N(t)) + a_t]\Big\}$$

donde $\text{int}\{x\}$ representa la parte entera de x. Definimos

$$F(i,a,z) := \text{int}\Big\{[(1-\varphi_1(i))(1+\lambda)+\varphi_1(i)z][i-\varphi_2(i)+a]\Big\}.$$

Si suponemos que ρ es la densidad de la tasa de retorno ξ_t, para toda $t \in \mathbb{N}$ la ley de transición K toma la forma

$$K_{ij}(a) = \int_{\mathbb{R}} \mathbb{1}_{\{j\}}[F(i,a,z)]\rho(z)dz,$$

para cada $i, j \in S$ y $a \in A$.

La dinámica de las proporciones H^N puede determinarse por la construcción presentada en el Capítulo 1, específicamente en la ecuación (1.9). Por otra parte, supondremos que la función de costo por etapa $r : \mathbb{P}(S) \times A \to \mathbb{R}$ es de la forma $r(m,a) = r(m)$ para toda $a \in A$, y supondremos que es acotada y uniformemente Lipschitz con constante L_r; esto es, para alguna constante $R > 0$

$$|r(m)| \leq R \quad \forall m \in \mathbb{P}(S),$$

y para cada m, $m' \in \mathbb{P}(S)$,

$$|r(m) - r(m')| \leq L_r ||m - m'||_\infty.$$

Por lo tanto, las Hipótesis 1.4.6 y 2.4.4 se satisfacen. Las dinámicas H^N y H quedan completamente determinados por las ecuaciones (1.7)-(1.9) y (4.1)-(4.3) respectivamente.

4.3. Modelo de reforestación

Como comentamos a lo largo del trabajo, hay casos donde H^N se puede obtener según el contexto del problema. En el siguiente ejemplo, veremos una situación en la que esto ocurre.

Este ejemplo es una adaptación del presentado en [16].

Supongamos que una unidad de tierra esta ocupada por una gran cantidad de árboles (N árboles) con edades que van de 1 hasta 4. Los árboles plantados pierden su valor pasados los 4 años. Sea $S = \{1,2,3,4\}$. La edad de cada uno de los árboles aumenta de i a $i+1$ o muere debido a una plaga o un desastre natural en cada período de tiempo

$t \in \mathbb{N}$. Para cada $t \in \mathbb{N}_0$, sea $\{\xi_t^i\}_{i=1}^N$ una familia de variables aleatorias Bernoulli i.i.d. con parámetro θ que toman valores 0 y 1. Estas variables aleatorias son un indicador para el n-ésimo árbol al tiempo t,

$$1 := \text{el árbol murió}, \ 0 := \text{el árbol continúa vivo}.$$

Sea $X_n^N(t)$ la edad del n-ésimo árbol al tiempo t y denotemos por $M_i^N(t)$ a la proporción de árboles de edades $i = 1, 2, 3, 4$ ocupando terreno en ese tiempo. Claramente $M^N(t) = (M_1^N(t), M_2^N(t), M_3^N(t), M_4^N(t)) \in \mathbb{P}_N(S)$.

El dueño de la tierra quien se asume es el controlador central, al final de cada tiempo t, corta una proporción $a_t^{(i)}$, $i = 1, 2, 3, 4$ de los árboles de cada grupo de edad. Sea

$$a_t := \left(a_t^{(1)}, a_t^{(2)}, a_t^{(3)}, a_t^{(4)}\right), \ 0 \le a_t^{(i)} \le M_i^N(t), \ i = 1, 2, 3, 4.$$

Como los árboles pasados los 4 años pierden su valor al morir, tenemos que $M_4(t) - \frac{1}{N}\sum_{i=1}^N \xi_t^i \mathbb{1}_{\{X_i^N(t)=4\}} = a_t^{(4)}$ para toda $t \in \mathbb{N}$. Supondremos que cuando un árbol es talado o muere, inmediatamente es colocado en su lugar un árbol nuevo. Al inicio de cada período $t+1$, el estado de las tierras esta dado por

$$\begin{aligned} M^N(t+1) = \Bigg(& \sum_{i=1}^4 a_t^{(i)} + \frac{1}{N}\sum_{i=1}^N \xi_t^i, M_1^N(t) - a_t^{(1)} - \frac{1}{N}\sum_{i=1}^N \xi_t^i \mathbb{1}_{\{X_i^N(t)=1\}}, \\ & M_2^N(t) - a_t^{(2)} - \frac{1}{N}\sum_{i=1}^N \xi_t^i \mathbb{1}_{\{X_i^N(t)=2\}}, M_3^N(t) - a_t^{(3)} - \frac{1}{N}\sum_{i=1}^N \xi_t^i \mathbb{1}_{\{X_i^N(t)=3\}}\Bigg), \end{aligned} \tag{4.4}$$

lo cual es equivalente a la forma matricial

$$M^N(t+1) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} M^N(t) + \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{pmatrix} a_t - \frac{1}{N}\begin{pmatrix} -\sum_{i=1}^N \xi_t^i \\ \sum_{i=1}^N \xi_t^i \mathbb{1}_{\{X_i^N(t)=1\}} \\ \sum_{i=1}^N \xi_t^i \mathbb{1}_{\{X_i^N(t)=2\}} \\ \sum_{i=1}^N \xi_t^i \mathbb{1}_{\{X_i^N(t)=3\}} \end{pmatrix}. \tag{4.5}$$

En este contexto, el espacio de estados es $\mathbb{P}_N(S)$ y el espacio de acciones es $A := [0, 1]^4$. Notemos que el mapeo $a \mapsto H^N(m, a, \xi)$, donde H^N representa la dinámica (4.5) y $\xi = (\xi^1, \ldots, \xi^N)$, es continua. En efecto, sea $\epsilon > 0$ y a, $a' \in A$ tales que

$$d(a, a') := \max_{i=1,2,3,4} |a^{(i)} - a'^{(i)}| < \epsilon/4.$$

De (4.4), obtenemos que

$$
\begin{aligned}
&||H^N(m,a,\xi) - H^N(m,a',\xi)||_\infty \\
&\quad = \text{máx}\left\{ \left| \sum_{i=1}^{4} a_t^{(i)} - \sum_{i=1}^{4} a_t'^{(i)} \right|, |-a^{(1)} + a'^{(1)}|, |-a^{(2)} + a'^{(2)}|, |-a^{(3)} + a'^{(3)}| \right\}, \\
&\quad \leq \text{máx}\left\{ \sum_{i=1}^{4} |a_t^{(i)} - a_t'^{(i)}|, |-a^{(1)} + a'^{(1)}|, |-a^{(2)} + a'^{(2)}|, |-a^{(3)} + a'^{(3)}| \right\}, \\
&\quad \leq \epsilon.
\end{aligned}
$$

Para definir la dinámica H en el modelo de campo medio, consideremos que la proporción de árboles que mueren en cada etapa esta dado por el parámetro $\theta \in [0,1)$. Entonces, para cada tiempo $t \in \mathbb{N}$

$$
m(t+1) = \left(\sum_{i=1}^{4} \left[\theta m_i(t) + a_t^{(i)} \right], (1-\theta)m_1(t) - a_t^{(1)}, (1-\theta)m_2(t) - a_t^{(2)}, (1-\theta)m_3(t) - a_t^{(3)} \right),
$$

la cual tiene como forma matricial

$$
m(t+1) = \begin{pmatrix} \theta & \theta & \theta & \theta \\ 1-\theta & 0 & 0 & 0 \\ 0 & 1-\theta & 0 & 0 \\ 0 & 0 & 1-\theta & 0 \end{pmatrix} m(t) + \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{pmatrix} a_t,
$$

donde $m(t) \in \mathbb{P}(S)$.

Dado que la dinámica H^N y H fueron obtenidas por medio de procedimientos distintos a los presentados a lo largo del trabajo, la convergencia probada en el Teorema 2.4.1 no se sigue inmediatamente en este caso. Por lo tanto, es necesario realizar una adaptación de los argumentos de la demostración del Teorema 2.4.1 a este ejemplo, lo cual presentamos a continuación.

Lema 4.3.1. *Para cada $m \in \mathbb{P}_N(S)$, $T \in \mathbb{N}$ y $\epsilon > 0$, existen constantes positivas C y λ tales que*

$$
\sup_{\pi \in \Pi_M} P_m^\pi \left\{ \sup_{0 \leq t \leq T} ||M^N(t) - m(t)||_\infty \geq \gamma_T(\epsilon) \right\} \leq CTe^{-\lambda N \epsilon^2}
$$

donde $\gamma_T(\epsilon) \to 0$ cuando $\epsilon \to 0$.

Demostración. Sea $\pi = \{f_t\} \in \Pi_M$ una política arbitraria y $M^N(0) = m(0) = m \in \mathbb{P}_N(S) \subset \mathbb{P}(S)$ el estado inicial. Para simplificar la notación a lo largo de la prueba, considere que

$$a_t = a_t^{\pi,N} := f_t(M^N(t)),$$

y

$$a_t = \left(a_t^{(1)}, a_t^{(2)}, a_t^{(3)}, a_t^{(4)}\right) = \left(a_t^{(1),\pi,N}, a_t^{(2),\pi,N}, a_t^{(3),\pi,N}, a_t^{(4),\pi,N}\right)$$

De la desigualdad de Hoeffding, tenemos que para $\epsilon \geq 0$ y $t \in \mathbb{N}_0$,

$$P_m^\pi\left[\left|\frac{1}{N}\sum_{i=1}^N \xi_t^i - E_m^\pi\left(\frac{1}{N}\sum_{i=1}^N \xi_t^i\right)\right| < \epsilon\right] = P_m^\pi\left[\left|\frac{1}{N}\sum_{i=1}^N \xi_t^i - \theta\right| < \epsilon\right] > 1 - 2e^{-2N\epsilon^2}.$$

Dado que la v.a. que modela en qué clase está cada árbol y la v.a. que indica si un árbol muere o no por causas naturales son independientes, tenemos que para $t \in \mathbb{N}_0$ y $j = 1, 2, 3, 4$

$$\begin{aligned} E_m^\pi\left[\frac{1}{N}\sum_{i=1}^N \xi_t^i \mathbb{1}_{\{X_i^N(t)=j\}}\right] &= \frac{1}{N}\sum_{i=1}^N E_m^\pi\left[\xi_t^i \mathbb{1}_{\{X_i^N(t)=j\}}\right], \\ &= \frac{1}{N}\sum_{i=1}^N E_m^\pi\left[\xi_t^i\right] E_m^\pi\left[\mathbb{1}_{\{X_i^N(t)=j\}}\right], \\ &= \theta M_j^N(t). \end{aligned}$$

Consideremos los conjuntos

$$\Omega_1 := \left\{\omega \in \Omega' \middle| \left|\frac{1}{N}\sum_{i=1}^N \xi_t^i - \theta\right| < \epsilon\right\} \subset \Omega'$$

$$\Omega_j := \left\{\omega \in \Omega' \middle| \left|\frac{1}{N}\sum_{i=1}^N \xi_t^i \mathbb{1}_{\{X_i^N(t)=j-1\}} - \theta M_{j-1}^N(t)\right| < \epsilon\right\} \subset \Omega', \; j = 2, 3, 4.$$

Definimos

$$\bar{\Omega} = \bigcap_{j=1}^4 \Omega_j.$$

Sea ϵ_t un número positivo tal que

$$||M^N(t) - m(t)||_\infty \leq \epsilon_t.$$

Para $j = 1$, las siguientes relaciones son ciertas en $\bar{\Omega}$,

$$\begin{aligned}
|M_1^N(t+1) - m_1(t+1)| &= \left| \sum_{i=1}^{4} a_t^{(i)} + \frac{1}{N}\sum_{i=1}^{N} \xi_t^i - \sum_{i=1}^{4} \left[\theta m_i(t) + a_t^{(i)}\right] \right|, \\
&\leq \left| \frac{1}{N}\sum_{i=1}^{N} \xi_t^i - \sum_{i=1}^{4} \theta m_i(t) \right|, \\
&= \left| \frac{1}{N}\sum_{i=1}^{N} \xi_t^i - \theta \sum_{i=1}^{4} m_i(t) \right|, \\
&= \left| \frac{1}{N}\sum_{i=1}^{N} \xi_t^i - \theta \right| < \epsilon. \qquad (4.6)
\end{aligned}$$

Para $j = 2, 3, 4$, en $\bar{\Omega}$ tenemos que

$$\begin{aligned}
&|M_j^N(t+1) - m_j(t+1)| \\
&\quad = \left| M_{j-1}^N(t) - a_t^{(j-1)} - \frac{1}{N}\sum_{i=1}^{N} \xi_t^i \mathbb{1}_{\{X_i^N(t)=j-1\}} - (1-\theta)m_{j-1}(t) + a_t^{(j-1)} \right|, \\
&\quad = \left| M_{j-1}^N(t) - \frac{1}{N}\sum_{i=1}^{N} \xi_t^i \mathbb{1}_{\{X_i^N(t)=j-1\}} - (1-\theta)m_{j-1}(t) \right|, \\
&\quad \leq |M_{j-1}^N(t) - m_{j-1}(t)| + \left| \theta m_{j-1}(t) - \frac{1}{N}\sum_{i=1}^{N} \xi_t^i \mathbb{1}_{\{X_i^N(t)=j-1\}} \right|, \\
&\quad \leq \epsilon_t + \left| \theta m_{j-1}(t) - \frac{1}{N}\sum_{i=1}^{N} \xi_t^i \mathbb{1}_{\{X_i^N(t)=j-1\}} \right|, \\
&\quad \leq \epsilon_t + \left| \theta m_{j-1}(t) - \theta M_{j-1}^N(t) \right| + \left| \theta M_{j-1}^N(t) - \frac{1}{N}\sum_{i=1}^{N} \xi_t^i \mathbb{1}_{\{X_i^N(t)=j-1\}} \right|, \\
&\quad \leq 2\epsilon_t + \epsilon. \qquad (4.7)
\end{aligned}$$

De (4.6) y (4.7), tenemos que

$$||M^N(t+1) - m(t+1)||_\infty \leq 2\epsilon_t + \epsilon.$$

Considerando que $||M^N(0) - m(0)||_\infty = \epsilon_0 = 0$ y aplicando un procedimiento inductivo, obtenemos que en $\bar{\Omega}$ se satisface

$$||M^N(t+1) - m(t+1)||_\infty < \epsilon\beta_t, \ t \in \mathbb{N}_0$$

donde $\{\beta_t\}$ es una sucesión creciente. Entonces, para un $T \in \mathbb{N}$ fijo,

$$||M^N(t+1) - m(t+1)||_\infty < \epsilon\beta_T, \ \forall\ 0 \leq t \leq T,$$

dentro del conjunto $\bar{\Omega}$. Por último, para la política $\pi \in \Pi_M$,

$$P_m^\pi \left\{ \sup_{0\leq t\leq T} ||M^N(t) - m(t)||_\infty < \epsilon\beta_T \right\} > 1 - 2e^{-2N\epsilon^2}.$$

Definiendo $\gamma_T(\epsilon) := \epsilon\beta_T$ y $C = \lambda = 2$, obtenemos

$$\sup_{\pi\in\Pi_M} P_m^\pi \left\{ \sup_{0\leq t\leq T} ||M^N(t) - m(t)||_\infty \geq \gamma_T(\epsilon) \right\} \leq CTe^{-\lambda N\epsilon^2}.$$

Notemos que si $\epsilon \to 0$ entonces $\gamma_T(\epsilon) \to 0$. ■

4.4. Modelo SIM

El gobierno de un país se encuentra tratando de solucionar cierta endemia que surgió en algunas regiones de su territorio. Para tratar de controlar esta situación, la población se clasifica en tres clases, susceptible (S), infectado (I) y muerto (M), y se calcula la proporción de ciudadanos que se encuentran en cada clasificación.

Con la intención de disminuir la incidencia de enfermos, el país puede tomar cualquiera de las siguientes acciones en $A := \{a^{(1)}, a^{(2)}, a^{(3)}, a^{(4)}\}$, donde

$a^{(1)}$ Cuarentena de individuos infectados.
$a^{(2)}$ Vacunación de individuos susceptibles.
$a^{(3)}$ Adquisición y almacenamiento preventivo de vacunas.
$a^{(4)}$ Investigación de nuevos tratamientos.

Dependiendo de las acciones que tome el gobierno, los individuos de la población cambian de clase de acuerdo a las siguientes matrices de transición:

$$K(a^{(1)}) = \begin{pmatrix} 7/8 & 1/8 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \end{pmatrix}, \quad K(a^{(2)}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3/4 & 1/4 \\ 0 & 0 & 1 \end{pmatrix},$$

$$K(a^{(3)}) = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \end{pmatrix}, \quad K(a^{(4)}) = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \end{pmatrix}.$$

De un estudio previo realizado en base a un brote de la misma enfermedad ocurrida en años anteriores, se sabe que existe una proporción-umbral m^* que nos estima si la enfermedad pasó de endemia a epidemia. Además, el país cuenta con una función de recompensa (o bienestar social), $r : \mathbb{P}(S) \times A \to \mathbb{R}$, que tiene la forma $r(m, a) = r(m)$ para toda $a \in A$, definida por

$$r(m) = ||m - m^*||_\infty.$$

Claramente

$$|r(m)| \leq 1 \ \forall m \in \mathbb{P}(S).$$

Para cada m, $m' \in \mathbb{P}(S)$

$$\begin{aligned} |r(m) - r(m')| &= \big|||m - m*||_\infty - ||m' - m^*||_\infty\big|, \\ &\leq ||m - m'||_\infty. \end{aligned}$$

Es decir,

$$|r(m) - r(m')| \leq ||m - m'||_\infty.$$

Por lo tanto, la Hipótesis 1.4.6 se satisface. Con lo anteriormente expuesto, H^N y H quedan completamente determinados por las ecuaciones (1.7)-(1.9) y (4.1)-(4.3) respectivamente. Notemos que la Hipótesis 2.4.4 se cumple trivialmente.

4.5. Estrategias de Marketing

Este ejemplo es una adaptación aún problema de control de un juego suma cero presentado en el artículo [12].

Una empresa está interesada en aumentar las ventas de un artículo que produce, para lo cual realizó un estudio de marketing para determinar qué estrategia publicitaria es más conveniente efectuar. Dicha empresa clasifica a sus posibles consumidores de la siguiente manera.

c_1 Consume el producto.

c_2 Ocasionalmente consume el producto.

c_3 Nunca consume el producto.

Denotemos $S = \{c_1, c_2, c_3\} := \{1, 2, 3\}$ el conjunto de clases. Vamos a suponer que las acciones de la empresa son solamente dos, $A = \left\{a^{(1)}, a^{(2)}\right\}$, donde

$a^{(1)}$ Aumentar la publicidad en redes sociales.

$a^{(2)}$ Realizar una intensa campaña por televisión.

Del estudio de Marketing se estimó las proporciones de respuesta de los clientes según la estrategia tomada por la empresa.

$$K\left(a^{(1)}\right) = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}, \quad K\left(a^{(2)}\right) = \begin{pmatrix} 2/3 & 1/6 & 1/6 \\ 2/3 & 1/6 & 1/6 \\ 2/3 & 1/6 & 1/6 \end{pmatrix}.$$

Recordemos que $\mathbb{P}(S)$ es el conjunto de medidas de probabilidad definidas en S. Por lo tanto, un elemento $m \in \mathbb{P}(S)$ tiene la forma $m = (m_1, m_2, m_3)$ tal que $m_i \geq 0$, $i = 1, 2, 3$ y $m_1 + m_2 + m_3 = 1$. Para $m = (m_1, m_2, m_3) \in \mathbb{P}(S)$, la función de costos de la empresa viene dado por $r : \mathbb{P}(S) \times A \to \mathbb{R}$, que es de la forma $r(m, a) = r(m)$ para toda $a \in A$, definida por

$$r(m) = m_1 + m_3.$$

Claramente

$$|r(m)| \leq 2 \ \forall m \in \mathbb{P}(S).$$

Por otro lado, para cada m, $m' \in \mathbb{P}(S)$

$$\begin{aligned} |r(m) - r(m')| &\leq |m_1 + m_3 - m_1' - m_3'|, \\ &\leq |m_1 - m_1'| + |m_3 - m_3'|, \\ &\leq 2 \max\{|m_1 - m_1'|, |m_2 - m_2'|, |m_3 - m_3'|\}, \\ &\leq 2||m - m'||_\infty = L_r||m - m'||_\infty \end{aligned}$$

con $L_r := 2$. Por lo tanto, la Hipótesis 1.4.6 se satisface. Con lo anteriormente expuesto, H^N y H quedan completamente determinados por las ecuaciones (1.7)-(1.9) y (4.1)-(4.3) respectivamente. La Hipótesis 2.4.4 se cumple trivialmente.

A continuación, utilizando el algoritmo de iteración de valores presentado en el Capítulo 2, calcularemos la función de valor en el campo medio.

Sea $m(0) = m = (m_1, m_2, m_3) \in \mathbb{P}(S)$ un estado inicial arbitrario. Las funciones de iteración de valores toman la siguiente forma

$$\begin{aligned} v_0 &= 0, \\ v_n(m) &= \inf_{a\in A} \{r(m) + \alpha v_{n-1}[mK(a)]\}. \end{aligned} \tag{4.8}$$

Como podemos observar en (4.8), el producto $mK(a)$ es constantemente utilizado en las ecuaciones del algoritmo de iteración de valores, por lo que, antes de continuar, procedemos a determinar su valor.

$$\begin{aligned} mK\left(a^{(1)}\right) &= m\begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}, \\ &= \begin{pmatrix} m_1 & m_2 & m_3 \end{pmatrix}\begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}, \\ &= \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right), \end{aligned}$$

y

$$\begin{aligned} mK\left(a^{(2)}\right) &= m\begin{pmatrix} 2/3 & 1/6 & 1/6 \\ 2/3 & 1/6 & 1/6 \\ 2/3 & 1/6 & 1/6 \end{pmatrix}, \\ &= \begin{pmatrix} m_1 & m_2 & m_3 \end{pmatrix}\begin{pmatrix} 2/3 & 1/6 & 1/6 \\ 2/3 & 1/6 & 1/6 \\ 2/3 & 1/6 & 1/6 \end{pmatrix}, \\ &= \left(\frac{2}{3}, \frac{1}{6}, \frac{1}{6}\right). \end{aligned}$$

Esto implica que

$$v_1(m) = \min_{a\in A}\{r(m)\} = m_1 + m_3;$$

$$\begin{aligned}
v_2(m) &= \min_{a\in A}\{r(m)+\alpha v_1(mK(a))\},\\
&= \min_{a\in A}\{m_1+m_3+\alpha v_1(mK(a))\},\\
&= m_1+m_3+\alpha v_1\left(mK\left(a^{(1)}\right)\right),\\
&= m_1+m_3+\alpha\left[\frac{1}{2}+\frac{1}{4}\right],\\
&= m_1+m_3+\alpha\left(\frac{3}{4}\right),\\
&= m_1+m_3+\frac{3}{4}\alpha;
\end{aligned}$$

$$\begin{aligned}
v_3(m) &= \min_{a\in A}\{r(m)+\alpha v_2(mK(a))\},\\
&= \min_{a\in A}\{m_1+m_3+\alpha v_2(mK(a))\},\\
&= m_1+m_3+\alpha v_2\left(mK\left(a^{(1)}\right)\right),\\
&= m_1+m_3+\alpha\left(\frac{1}{2}+\frac{1}{4}+\frac{3}{4}\alpha\right),\\
&= m_1+m_3+\alpha\left(\frac{3}{4}+\frac{3}{4}\alpha\right),\\
&= m_1+m_3+\frac{3}{4}\alpha+\frac{3}{4}\alpha^2.
\end{aligned}$$

En general, para $n \geq 1$ tenemos que

$$v_n(m) = m_1 + m_3 + \frac{3}{4}\sum_{k=1}^{n-1}\alpha^k.$$

Haciendo n tender a infinito obtenemos

$$v_*(m) = m_1 + m_3 + \frac{3}{4}\left(\frac{\alpha}{1-\alpha}\right).$$

Bibliografía

[1] Masanao Aoki. New macroeconomic modeling approaches: Hierarchical dynamics and mean-field approximation. *Journal of economic dynamics and control*, 18(3-4): 865–877, 1994. 1

[2] Robert B Ash and Catherine A Doleans-Dade. *Probability and measure theory*. Academic Press, 2000. 11

[3] Dimitir P Bertsekas and Steven Shreve. *Stochastic optimal control: the discrete-time case*. 2004. 11

[4] Kai Lai Chung and John B Walsh. *Markov processes, Brownian motion, and time symmetry*, volume 249. Springer Science & Business Media, 2006. 17

[5] Nicolas Gast and Bruno Gaujal. A mean field approach for optimization in discrete time. *Discrete Event Dynamic Systems*, 21(1):63–101, 2011. 1, 2

[6] Nicolas Gast, Bruno Gaujal, and Jean-Yves Le Boudec. Mean field for markov decision processes: from discrete to continuous optimization. *IEEE Transactions on Automatic Control*, 57(9):2266–2280, 2012. 1, 2, 9, 15

[7] David González-Sánchez, Fernando Luque-Vásquez, and J Adolfo Minjárez-Sosa. Zero-sum markov games with random state-actions-dependent discount factors: Existence of optimal strategies. *Dynamic Games and Applications*, 9(1):103–121, 2019. 36

[8] Evgueni I Gordienko and J Adolfo Minjárez-Sosa. Adaptive control for discrete-time markov processes with unbounded costs: discounted criterion. *Kybernetika*, 34(2): 217–234, 1998.

[9] Onésimo Hernández-Lerma and Jean B Lasserre. *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Springer, New York, 2012. 11, 12, 33, 34, 42

[10] Onésimo Hernández-Lerma and Jean B Lasserre. *Further Topics on Discrete-Time Markov Control Processes*. Springer, New York, 2012. 13, 14, 22

[11] Carmen G Higuera-Chan. *Modelos de control de campo medio para sistemas estocásticos de interacción de objetos con distribución desconocida*. Tesis de Doctorado en Ciencias (Matemáticas). Universidad de Sonora, 2016. 54

[12] Carmen G Higuera-Chan and J Adolfo Minjárez-Sosa. A mean field approach for discounted zero-sum games in a class of systems of interacting objects [artículo sometido]. 2020. 1, 2, 61

[13] Carmen G Higuera-Chan, Héctor Jasso-Fuentes, and J Adolfo Minjárez-Sosa. Discrete-time control for systems of interacting objects with unknown random disturbance distributions: a mean field approach. *Applied Mathematics & Optimization*, 74(1):197–227, 2016. 2, 54

[14] Carmen G Higuera-Chan, Héctor Jasso-Fuentes, and J Adolfo Minjárez-Sosa. Control systems of interacting objects modeled as a game against nature under a mean field approach. *Journal of Dynamics & Games*, 4(1):59, 2017. 1, 2

[15] Aime Lachapelle, Julien Salomon, and Gabriel Turinici. Computation of mean field equilibria in economics. *Mathematical Models and Methods in Applied Sciences*, 20 (04):567–588, 2010. 1

[16] Leonardo R Laura-Guarachi and Onésimo Hernández-Lerma. The mitra-wan forestry model: a discrete-time optimal control problem. *Natural Resource Modeling*, 28(2): 152–168, 2015. 55

[17] Fernando Luque-Vásquez and J Adolfo Minjárez-Sosa. Iteration algorithms in markov decision processes with state-action-dependent discount factors and unbounded costs. *Operations Research: the Art of Making Good Decisions*, page 55, 2016. 36

[18] M Elena Martínez-Manzanares and J Adolfo Minjárez-Sosa. A mean field absorbing control model for interacting objects systems [artículo sometido]. 2020. 3

[19] J Adolfo Minjárez-Sosa. Markov control models with unknown random state–action-dependent discount factors. *Top*, 23(3):743–772, 2015. 36

[20] J Adolfo Minjárez-Sosa. *Zero-Sum Discrete-Time Markov Games with Unknown Disturbance Distribution*. Springer Nature, 2020. 47

[21] Jacques Neveu. *Mathematical foundations of the calculus of probability*. Holden-day, 1965. 11

[22] Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill Publishing Co., 1976.

[23] Faryad Darabi Sahneh, Caterina Scoglio, and Piet Van Mieghem. Generalized epidemic mean-field model for spreading processes over multilayer complex networks. *IEEE/ACM Transactions on Networking*, 21(5):1609–1620, 2013. 1

[24] Jiongmin Yong. Linear-quadratic optimal control problems for mean-field stochastic differential equations. *SIAM journal on Control and Optimization*, 51(4):2809–2838, 2013. 1