# UNIVERSIDAD DE SONORA

División de Ciencias Exactas y Naturales

**Programa de Posgrado en Matemáticas**


# Modelos de control de campo medio para Sistemas Estocásticos de Interacción de Objetos con Distribución Desconocida

# T E S I S

que presenta:


M.C. Carmen Geraldi Higuera Chan


para obtener el título de:

Doctor en Ciencias **Matemáticas**

Director de tesis: Dr. Jesús Adolfo Minjárez Sosa
Dr. Héctor Jasso Fuentes


Hermosillo, Sonora, México, 18 de marzo de 2016

# SINODALES

Dr. Evgueni I. Gordienko
Universidad Autónoma Metropolitana-Iztapalapa, Ciudad de México.

Dr. Héctor Jasso Fuentes
Centro de Investigación y Estudios Avanzados del I.P.N., Ciudad de México.

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

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

Dr. Óscar Vega Amaya
Universidad de Sonora, Hermosillo, México.

# Índice general

# Agradecimientos

Sin duda alguna tengo mucho a quien dar las gracias.

Agradezco profundamente a los sinodales por la revisión de este trabajo de tesis y por las sugerencias realizadas al mismo.

De una manera muy especial agradezco también a mis directores de tesis Jesús Adolfo Minjárez Sosa y Héctor Jasso Fuentes por su apoyo, guía, paciencia y por aceptar ser parte de este proyecto.

Reconozco el apoyo incondicional de mi familia, sin ellos no valdría la pena todo el esfuerzo. A mis amigos, por siempre estar conmigo. A mis maestros, por siempre saber motivarme y guiarme por este camino. Al Departamento de Matemáticas de la Universidad de Sonora que siempre me ha abierto sus puertas y a CONACYT por su apoyo.

Gracias Dios por estar en mi...

# Abstract

The Mean field theory has its origins on particle physics with Pierre Weiss's work in the 1900's. The goal of this theory is to simplify systems analysis, which are composed of a large number on interacting particles, proposing a new system, calling mean field, which it is easier to analyze. This system is formed by the effect of a group of particles that share certain property. In general terms, mean field theory has been adopted for describing microscopic phenomena from a macroscopic point of view, that is, instead of analyzing each of particles, the average behavior of a group of them is studied.

Given this idea, applications of mean field theory have spread to other areas as economics, finance, ecology, biology and operations research (see [3, 6, 8, 9, 11, 12, 13, 14, 24, 25, 28, 32, 37, 39]) where each particle is identified in a general form as an object or as an agent. For example:

- models in economics and finance, where objects represent each small investor;
- The study of the behavior of large populations in Biology;
- reforestation policy design, maintenance and exploitation of forests, in which case each tree is represented by an object.

Recently, mean field theory has been applied in game theory to study the in-

teractions between a large number of objects, which represent each player. These games are known as mean field games and were introduced in 2006, independently, by Minyi Huang, Rowland Malhamé and Peter Caines in Montreal, and by Jean-Michel Lasry and Pierre-Louis Lions in París (véase [7, 29, 30, 31]). Nowadays, mean field game theory is a branch of game theory (see [1, 4, 15, 18, 26, 28, 31]). It is therefore a set of concepts, mathematical tools, theorems, simulation methods and algorithms.

Another class of applications of the mean field theory is in optimization problems of dynamic systems which consist of a large number of objects interacting with each other. That is, systems such as those mentioned above where it is possible to formulate a problem of sequential decisions under uncertainty conditions. In this case, a stochastic control problem is posed, where there is a central controller with specific interests, and in order to determine optimal control policies under certain optimality criteria.

This is the kind of problems we deal with in this work. That is, control problems of systems composed by a large number of objects interacting with each other, under a mean field approach.

We consider a controlled stochastic system composed of a large number of $N$ interacting objects which share a common environment. At each step, the objects are randomly distributed in a finite number of classes, and there is a central controller whose decisions affect the behavior of objects. Once the controller takes a decision, objects change to another class randomly. In particular we will assume that evolution of the object $n$ at time $t$ is determined by a difference equation, homogeneous in $N$, of the form

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

where $F$ is a known function, $X_n^N(t)$ is the class of the object $n$ at time $t$, $C^N(t)$ is the context of the environment, $a_t$ is the control or action selected by a central controller, and $\xi_t^n$ is the random disturbance. In addition, at each stage, a cost resulting from the movement of the objects and the selected control is generated. In this sense, we propose a suitable Markov control model to study this class of systems, in which the controller aims to select actions to minimize a given discounted cost criterion.

The facts that $N$ is too large ($N \sim \infty$), lead to formulate an alternative scheme to analyze the corresponding optimal control problem. Indeed, our approach to follow will be framed in the context of the *mean field theory*, under which, instead of analyzing a single object, we focus on the number or proportion of objects occupying certain class at each stage. This defines a control model $\mathcal{M}_N$ whose states are precisely the proportions of objects evolving according to a suitable stochastic difference equation, depending on $N$. Then, by taking limit as $N$ goes to infinity, we obtain the so-called mean field control model $\mathcal{M}$, whose states are probability measures, resulting of the limit of the aforementioned proportions, which in turn satisfy a *deterministic* difference equation. In this way $\mathcal{M}$ can be considered as an approximating model for $\mathcal{M}_N$, in the sense that any optimal control policy $\pi^*$ associated to $\mathcal{M}$ can be used to control the original process (the $N-$system) on $\mathcal{M}_N$, and therefore the objective is to measure its optimality deviation. Clearly, the good performance of $\pi^*$ on $\mathcal{M}_N$ depends of the accuracy of the mean field limit of $\mathcal{M}_N$ to $\mathcal{M}$ as $N \to \infty$.

The control problem in mean field model is studied depending on the conditions in the system. For example, the main feature of our work is to assume that the distribution of random variables $\{\xi_t^n\}$ is unknown. This implies that the corresponding dynamic on models $\mathcal{M}_N$ and $\mathcal{M}$ is unknown, and therefore the problem is analyzed under the following two scenarios:

**(i)** The process $\{\xi_t^n\}$ is formed by observable, independent and identically distributed random variables for each object.

**(ii)** The process $\{\xi_t^n\}$ is unobservable, and random variables are independent but not necessarily identically distributed.

In case (i), the dynamic of the objects as well as the mean field process depend on the unknown density $\rho$. Thus, besides the analysis of the limit behavior as $N \to \infty$, the controller must implement a statistical estimation procedure for $\rho$ in order to get some information about the dynamic of the objects. To this end, at each stage $t$, $\rho$ is estimated from historical observations $\xi_0, \xi_1, ...\xi_t$, collected during the evolution of the system. Such estimation procedure is then combined with the minimization task for obtaining control policies. However, as is well-known, the discounted criterion depends strongly on the decisions selected in the first stages, precisely where the information on $\rho$ is rather poor or deficient. This fact implies that, in general, under discounted criteria, procedures of estimation and control do not provide optimal policies (see, e.g., [17, 19, 23, 36]). Thus, in this paper we seek optimality results in a weaker sense, the so-called eventually asymptotically optimality.

In case (ii), we will appeal to the minimax theory. This allow us to define as a first step a game against nature model $\mathcal{M}^N$ whose states are now the proportions of objects, as in case (i). Then, by letting $N \to \infty$ and applying an appropriate type of laws of large numbers, we obtain a game against nature $\mathcal{M}$, independent of $N$, whose states are precisely the probability measures obtained as limit of the proportions. The model $\mathcal{M}$ is referred to as the mean field game against nature. Under this scheme, we can establish solutions for the game model $\mathcal{M}$, and obtain a "worst case" optimal control policy $\pi^*$. Then, our approach consists in analyzing the minimax deviation of $\pi^*$ when it is used to control the original process defined on $\mathcal{M}^N$. Specifically we aim to show that $\pi^*$ is a (worst case) optimal control policy in an asymptotic sense as $N \to \infty$.

According to cases (i) y (ii), our research consists in the following problems:

**1.-** Estimation and control problem for system of interacting objects under mean field approach.

**2.-** Minimax control problem for system of interacting objects under mean field approach.

Therefore, this thesis is a combination of mean field theory with techniques of control processes and adaptive Markov control for problem 1 [5, 17, 19, 20, 23, 36, 38] and minimax control techniques for problem 2 [34, 16, 33, 35]. With the results of our research, papers [21, 22] were developed.

# Capítulo 1

# Introducción

La teoría de campo medio tiene sus orígenes en el área de la física con los trabajos de Pierre Weiss en la década de 1900. Esta teoría tiene como objetivo simplificar el análisis de sistemas (o campos) que se componen de un número grande de partículas que interactúan entre sí, proponiendo un sistema, llamado campo medio, más fácil de analizar, el cual se forma por el efecto promedio de un grupo de partículas que comparten cierta propiedad. En términos generales, la teoría de campo medio se ha adoptado para describir fenómenos microscópicos desde un punto de vista macroscópico, en el siguiente sentido: en lugar de analizar cada partícula de forma individual, lo cual sería prácticamente imposible, se estudia el comportamiento promedio de un grupo de ellas, dando lugar al campo medio.

Tomando en cuenta esta idea, las aplicaciones de la teoría de campo medio se ha extendido a otras áreas como economía, finanzas, ecología, biología e investigación de operaciones (ver [3, 6, 8, 9, 11, 12, 13, 14, 24, 25, 28, 32, 37, 39]) donde cada partícula es identificada genéricamente con un objeto o agente. Por ejemplo:

- modelos en economía y finanzas, donde los objetos representan a cada pequeño inversionista;

- estudio del comportamiento de grandes poblaciones de seres vivos en Biología;
- diseño de políticas de reforestación, mantenimiento y explotación de bosques, en cuyo caso cada árbol es representado por un objeto.

Recientemente, la teoría de campo medio se ha aplicado en el área de teoría de juegos para estudiar las interacciones entre un número muy grande de agentes económicos, los cuales representan a cada jugador. A estos juegos se les conoce como juegos de campo medio y fueron introducidos en el año 2006, de forma independiente, por Minyi Huang, Rowland Malhamé y Peter Caines en Montreal, y por Jean-Michel Lasry y Pierre-Louis Lions en París (véase [7, 29, 30, 31]). Actualmente, los juegos de campo medio es un área bien establecida y estudiada dentro de la teoría de juegos (véase [1, 4, 15, 18, 26, 28, 31]), la cual incluye sus propios teoremas, conceptos y técnicas de simulación.

Otra clase de aplicaciones de la teoría de campo medio es en problemas de optimización de sistemas dinámicos que se componen de un número grande de objetos que interactúan entre sí. Es decir, sistemas como los mencionados anteriormente donde es posible formular un problema de decisiones secuenciales, y en condiciones de incertidumbre. En este caso, es posible plantear un problema de control estocástico donde existe un controlador central con intereses bien específicos, de tal manera que el objetivo es determinar políticas de control óptimas con respecto ciertos criterios de optimalidad.

Esta es la clase de problemas que estudiaremos en el presente trabajo. Es decir, problemas de control de sistemas de interacción de objetos bajo un esquema de campo medio, los cuales describimos a continuación.

Consideremos un sistema compuesto por una gran cantidad $N$ de objetos que interactuan entre sí, los cuales se pueden clasificar y comparten un ambiente en común. En cada etapa los objetos se encuentran distribuidos aleatoriamente en un número finito de clases, y además existe un controlador central cuyas decisiones afectan el comportamiento de los objetos. Esto es, una vez que el controlador toma una decisión, los objetos cambian de clase de manera aleatoria. En particular supondremos que la evolución del $n$-ésimo objeto está dada por una ecuación en diferencias de la forma

$$X_n^N(t+1) = F(X_n^N(t), C^N(t), a_t, \xi_t^n),\ t = 0, 1, \ldots,$$

donde $F$ es una función conocida, $X_n^N(t)$ representa la clase del $n$-ésimo objeto, $C^N(t)$ es el contexto del ambiente que comparten los objetos, $a_t$ la acción (o control) seleccionada por el controlador, y $\{\xi_t^n\}$ es una sucesión de variables aleatorias. Además, en cada etapa, se genera un costo como resultado del movimiento de los objetos y de la acción elegida. En este sentido, nos enfrentamos a un problema de control estocástico, en el cual el objetivo del controlador es tomar acciones que minimicen el criterio de costo descontado.

El hecho de que el número de objetos $N$ sea muy grande ($N \sim \infty$) nos lleva seguir el enfoque de la teoría de campo medio, la cual, en el contexto del problema de control estocástico descrito previamente, consiste en lo siguiente. En lugar de analizar cada objeto, nos centramos en el análisis de la proporción de objetos que ocupan cada clase. Tomando en cuenta este hecho, es posible definir un modelo de control $\mathcal{M}_N$ cuyos estados son precisamente las proporciones de objetos. Entonces, tomando límite cuando $N \to \infty$, y aplicando un tipo de ley de los grandes números, obtenemos un nuevo modelo, el llamado modelo de control de campo medio, $\mathcal{M}$ (independiente de $N$), cuyos estados son las medidas de probabilidad resultantes del límite de las proporciones. Por lo tanto, $\mathcal{M}$ puede ser considerado como un modelo aproximante para el modelo original $\mathcal{M}_N$ de tal forma que una política óptima $\pi^*$ asociada a $\mathcal{M}$ puede ser usada para controlar el proceso original correspondiente al modelo $\mathcal{M}_N$. Por lo tanto, el objetivo es medir la desviación de optimalidad de $\pi^*$ en un sentido asintótico cuando $N \to \infty$. Es claro que el buen desempeño de $\pi^*$ en $\mathcal{M}_N$ dependerá de la precisión del proceso límite de $\mathcal{M}_N$ a $\mathcal{M}$ cuando $N \to \infty$ (el límite de campo medio).

El problema de control en el modelo de campo medio se estudia dependiendo de las condiciones en el sistema de interacción de objetos. Por ejemplo, la característica principal de nuestro trabajo es el suponer que la distribución de las variables aleatorias $\{\xi_t^n\}$ es desconocida. Esto implica que la dinámica de las proporciones o de los estados en el modelo $\mathcal{M}_N$, así como la dinámica de las probabilidades en $\mathcal{M}$ también sea desconocida, y por lo tanto el problema se analizará bajo los siguientes dos escenarios:

**(i)** El proceso $\{\xi_t^n\}$ esta formado por variables aleatorias observables, independientes e idénticamente distribuidas para cada objeto.

**(ii)** El proceso $\{\xi_t^n\}$ es no observable y las variables aleatorias son independientes pero no necesariamente idénticamente distribuidas.

El caso (i) se estudiará mediante métodos de estimación y control en el modelo $\mathcal{M}$. Esto es, sea $\rho$ la densidad común y desconocida de las variables aleatorias $\{\xi_t^n\}$ para cada $n$. En cada etapa, tomando en cuenta el historial de observaciones $\xi_0^n, \xi_1^n, \ldots, \xi_t^n$ recopiladas durante la evolución del sistema, el controlador obtiene un estimador $\rho_t$ de $\rho$, y por lo tanto se obtiene una estimación de la dinámica. Tal proceso de estimación se combina con el proceso de minimización para obtener políticas de control en el campo medio. Sin embargo, como se conoce, el criterio descontado depende fuertemente de las decisiones tomadas en las primeras etapas, precisamente donde la información acerca de $\rho$ es pobre o deficiente. Este hecho implica que, en general, bajo el criterio descontado, los métodos de estimación y control no proveen políticas óptimas (vea por ejemplo, [17, 19, 23, 36]). Por lo tanto se recurre a resultados de optimalidad en un sentido más débil, los cuales combinan el proceso de estimación ($t \to \infty$) con el límite de campo medio ($N \to \infty$), obteniendo las llamadas *políticas eventualmente-asintóticamente óptimas*.

La hipótesis de observabilidad del proceso $\{\xi_t^n\}$ en el caso (i) es determinante para la implementación del método de estimación y control, y por lo tanto no es aplicable al caso (ii). Bajo este último escenario, el problema de control del sistema de interacción de objetos lo estudiaremos como un problema de control minimax, también conocido como *juego contra la naturaleza*. En este caso supondremos que la única información que posee el controlador es que las distribuciones de las variables aleatorias $\xi_t^n$'s pertenecen a un conjunto apropiado de medidas de probabilidad $\Gamma$. Además, el controlador tiene un oponente, la “naturaleza ”, quien, en cada etapa selecciona una distribución del conjunto $\Gamma$ para la correspondiente variable aleatoria $\xi_t^n$. Por lo tanto el objetivo del controlador será minimizar el máximo costo incurrido por la naturaleza.

Al igual que en el caso (i), para el modelo de juegos contra la naturaleza, aplicamos la teoría de campo medio, y de esta manera se define el modelo de control minimax $\mathcal{M}_{mxm}^N$ cuyos estados son las proporciones de objetos en cada clase. Entonces, haciendo $N \to \infty$ obtenemos un nuevo modelo de control minimax $\mathcal{M}_{mxm}$, que es independiente de $N$, y cuyos estados son medidas de probabilidad obtenidas como límite de las proporciones. Al modelo $\mathcal{M}_{mxm}$ lo llamamos modelo de control minimax de campo medio. Bajo este esquema, podemos resolver el problema de control minimax asociado al modelo $\mathcal{M}_{mxm}$ y obtener una política de control minimax $\pi^*$. Así, nuestra aproximación consiste en analizar la desviación minimax de $\pi^*$ cuando ésta se aplica en el proceso de control original $\mathcal{M}^N_{mxm}$. Específicamente mostramos que $\pi^*$ es una política de control minimax en un sentido asintótico cuando $N \to \infty$.

Considerando los casos (i) y (ii), nuestra investigación se centra en el análisis de los siguientes problemas:

**1.-** Problema de estimación y control en sistemas de interacción de objetos bajo en esquema de campo medio.

**2.-** Problema de control minimax en sistemas de interacción de objetos bajo un esquema de campo medio.

Estos problemas tienen como antecedente el trabajo en [12] en el cual se aborda un problema similar considerando completamente conocidas las componentes que definen la dinámica de los objetos y bajo criterios con horizonte finito. Dicha investigación se centra principalmente en las tasas de convergencia al de campo medio, más que en los resultados de optimalidad, como es nuestro caso.

Por lo tanto, el presente trabajo es una combinación de la teoría de campo medio con las técnicas de los procesos de control de Markov y control adaptado para el Problema 1 [5, 17, 19, 20, 23, 36, 38] y con las técnicas de control minimax para el Problema 2 [34, 16, 33, 35]. Con los resultados obtenidos en nuestra investigación se desarrollaron los artículos [21, 22] los cuales corresponden a los problemas 1 y 2 respectivamente.

El trabajo está estructurado de la siguiente manera. Los Capítulos 2 y 3 se desarrollan considerando el escenario (i). En particular, en el Capítulo 2 se introducen los elementos para la descripción del sistema de interacción de objetos, y en el Capítulo

3 se introducen las hipótesis que garantizan la existencia de estrategias óptimas bajo el criterio descontado. Además se presenta la construcción de las políticas adaptadas eventualmente-asintóticamente óptimas para el controlador. Después, bajo el escenario del caso (ii), en el Capítulo 4 estudiamos el problema de control minimax bajo el criterio de costo descontado. De igual forma, se construyen estrategias, las cuales resultan ser asintóticamente minimax para el modelo original. Por último en el Capítulo 5, se presenta un ejemplo para ilustrar nuestras hipótesis.

## Notación y terminología

Denotaremos por $\mathbb{N}$ (respectivamente $\mathbb{N}_0$) al conjunto de los enteros positivos (resp. enteros no negativos); $\mathbb{R}$ (resp. $\mathbb{R}_+$) representa el conjunto de los números reales (resp. reales no negativos).

Por otro lado, dado un espacio de Borel $Z$ (esto es, un subconjuto de un espacio métrico completo y separable), denotamos a su $\sigma-$álgebra de Borel por $\mathcal{B}(Z)$ y el atributo de "medible"será aplicado para referirse a conjuntos Borel medibles o funciones Borel medibles.

Sea $\mathbb{M}(Z)$ el conjunto de las medidas con signo finitas en $Z$. Si $Z \subset \mathbb{R}$ es finito, es decir $Z = \{1, 2, ..., z\}$, identificaremos a cualquier $p \in \mathbb{M}(Z)$ como un vector $p := (p(1), p(2), ..., p(z))$. En particular, denotaremos por $\mathbb{P}(Z) \subset \mathbb{M}(Z)$ al conjunto de medidas de probabilidad en $Z$. En este caso, cualquier medida $p \in \mathbb{P}(Z)$ puede ser expresada en términos de su distribución de probabilidad $\{p(i) : i \in Z\}$ donde $p(i) \geq 0$, $i \in Z$, y $\sum_{i=1}^{z} p(i) = 1$. Observemos que con la topología heredada de $\mathbb{R}$, $Z = \{1, 2, \cdots, z\}$ es un conjunto de Borel, por lo que el conjunto $\mathbb{P}(Z)$ también lo es. Como es usual, $|\cdot|$ será la notación de la norma en $\mathbb{R}$.

Definiremos la norma $L_\infty$ en $\mathbb{M}(Z) \times \mathbb{R}^d$, para un conjunto finito $Z$ de la siguiente manera. Para cada vector $(p, c) \in \mathbb{M}(Z) \times \mathbb{R}^d$:

$$\|(p, c)\|_\infty := \text{máx}\left\{\|p\|_\infty^1, \|c\|_\infty^2\right\},$$

donde $\|p\|_\infty^1 := \text{máx}\{|p(1)|, ..., |p(z)|\}$ y $\|c\|_\infty^2 := \text{máx}\{|c_1|, ..., |c_d|\}$, donde $c = (c_1, \cdots, c_d)$. Además, para un espacio de Borel $A$, denotaremos por $d_A$ a la métrica heredada. Para cada $(p, c, a), (p', c', a') \in \mathbb{P}(Z) \times \mathbb{R}^d \times A$ la correspondiente $L_\infty-$distancia se define

$$\|(p, c, a) - (p', c', a')\|_\infty^3 := \text{máx}\left\{\|p - p'\|_\infty^1, \|c - c'\|_\infty^2, d_A(a, a')\right\}.$$

Sea $W : \mathbb{P}(Z) \times \mathbb{R}^d \times A \to \mathbb{R}$. Diremos que $W$ satisface una condición de tipo Lipschitz en $\mathbb{P}(Z) \times \mathbb{R}^d$ si existe una constante $L_W$ tal que para cada $(p,c),(p',c') \in \mathbb{P}(Z) \times \mathbb{R}^d$ y $a \in A$ se satisface

$$|W(p,c,a) - W(p',c',a)| \leq L_W \max\left\{ \|p-p'\|_\infty^1, \|c-c'\|_\infty^2 \right\}.$$

Observe que esta propiedad no es exactamente la condición usual de Lipschitz. Sin embargo, a lo largo del trabajo emplearemos el término función de Lipschitz para esta situación, especificando en cada caso la propiedad a la que nos referimos.

Para una matriz $A_{n\times n}$ definimos a su correspondiente norma como

$$\|A\|_\infty^0 := \max_{i,j} |A_{ij}|.$$

Sean $Z$ y $A$ espacios de Borel. Un kernel estocástico $Q(\cdot|\cdot)$ es una función $Q : \mathcal{B}(Z) \times A \to [0,1]$, tal que $Q(\cdot|a)$ es una medida de probabilidad en $\mathcal{B}(Z)$ para cada $a \in A$ fija, y $Q(B|\cdot)$ es una función medible en $A$ para cada $B \in \mathcal{B}(Z)$ fijo. Finalmente, $\mathbb{B}(Z)$ denota la clase de las funciones de $Z$ con valores en los reales que son acotadas y lo dotamos con la norma del supremo $\|v\| := \sup_{z\in Z} |v(z)|$, mientras que $\mathbb{C}_b(Z)$ es el subespacio de $\mathbb{B}(Z)$ que consiste en lan funciones continuas y acotadas.

Para una fácil lectura, a continuación resumiremos la notación empleada a lo largo de este trabajo.

## Lista de variables principales

| | |
|---|---|
| $N \in \mathbb{N}$ | Número de objetos |
| $S = \{1, 2, \ldots, s\}$ | Espacio de las clases |
| $X_n^N(t)$ | Clase del $n$-ésimo objeto al tiempo $t$ |
| $F : S \times \mathbb{R}^d \times A \times \mathbb{R} \to \mathbb{R}$ | Dinámica del proceso $\{X_n^N(t)\}$ |
| $\xi_t^n$ | Ruido aleatorio del $n$-ésimo objeto al tiempo $t$ |
| $C^N(t), c(t) \in \mathbb{R}^d$ | Contexto del ambiente |
| $M_i^N(t)$ | Proporción de objetos en el estado $i$ al tiempo $t$ |
| $\vec{M}^N(t) \in \mathbb{P}_N$ | Vector de proporciones |
| $a_t \in A$ | Acción seleccionada por el controlador al tiempo $t$ |
| $g$ | Dinámica del contexto del ambiente |
| $y^N(t) := (M^N(t), C^N(t))$ | Estado del sistema con $N$ objetos al tiempo $t$ |
| $\mathbb{Y}_N = \mathbb{P}_N(S) \times \mathbb{R}^d$ | Espacio de estados del sistema con $N$ objetos |
| $r$ | Función de costo |
| $\alpha \in (0, 1)$ | Factor de descuento |
| $\pi^N = \{\pi_t^N\} \in \Pi^N$ | Política para el controlador |
| $\widetilde{\pi} = \{\mu_t\} \in \Pi_\Gamma$ | Política para el oponente |
| $V^N$ | Costo total $\alpha$-descontado para el sistema de $N$ objetos |
| $V_*^N$ | $N$-función de valor (minimax) |
| $\pi^N = \{\pi_t^N\} \in \Pi^N$ | Política para el controlador |
| $\vec{m}$ | Medida de ocupación en el conjunto $\mathbb{P}(S)$ |
| $y(t) := (\vec{m}(t), c(t))$ | Estado del sistema determinista |

| | |
|---|---|
| $\Phi^N$ | Función de discrepancia en el modelo de $N$ objetos |
| $\Phi$ | Función de discrepancia en el modelo determinista |
| $\mathbb{Y} = \mathbb{P}(S) \times \mathbb{R}^d$ | Espacio de estados del modelo determinista |
| $H$ | Dinámica del proceso $\{y(t)\}$ |
| $v$ | Costo total $\alpha$-descontado para el sistema determinista |
| $v_*$ | Función de valor (minimax) para el sistema determinista |
| $\mathcal{M}^N$ | Modelo de control asociado al sistema de $N$ objetos |
| $\mathcal{M}$ | Modelo de control límite |
| $G_\rho^N$ | Dinámica del vector de proporciones $\vec{M}^N$ |
| $H_\rho^N$ | Dinámica del proceso $\{y^N(t)\}$ en $\mathcal{M}^N$ |
| $G_\rho$ | Dinámica de $\vec{m}$ |
| $H_\rho$ | Dinámica del proceso $\{y(t)\}$ en $\mathcal{M}$ |
| $K_{ij}^\rho(a,c)$ | Probabilidad de transición de la clase $i$ a la clase $j$ |
| $\mathcal{M}_{mxm}^N$ | Modelo de control minimax asociado al sistema de $N$ objetos |
| $\mathcal{M}_{mxm}$ | Modelo límite del modelo de control minimax |
| $G^N$ | Dinámica del vector de proporciones $\vec{M}^N$ |
| $H^N$ | Dinámica del proceso $\{y^N(t)\}$ en $\mathcal{M}_{mxm}^N$ |
| $G$ | Dinámica de $\vec{m}$ |
| $H$ | Dinámica del proceso $\{y(t)\}$ en $\mathcal{M}_{mxm}$ |
| $\mu_t$ | Distribución de $\xi_t$ que pertenecen al conjunto $\Gamma$ |
| $K_{ij}(a,\mu,c)$ | Probabilidad de transición de la clase $i$ a la clase $j$ para el caso minimax |
| $\widetilde{\pi} = \{\mu_t\} \in \Pi_\Gamma$ | Política para el oponente |

# Capítulo 2

# Modelo de control de Markov para Sistemas de interacción de objetos

## 2.1. Introducción

En este capítulo se describe el sistema de interacción de objetos considerando que las v.a. que definen la dinámica de los objetos son observables y que son i.i.d.. En particular, se introducen las condiciones necesarias para aplicar la teoría de campo medio. Con estos elementos se introduce el modelo de control correspondiente y se plantea el problema de control óptimo bajo el criterio de costo descontado. Por último se establece un conjunto de hipótesis que garantizan la existencia de la solución a dicho problema.

## 2.2. Descripción del Sistema de objetos

Consideremos un sistema de control estocástico a tiempo discreto compuesto por un número grande $N$ de objetos que interactúan entre sí como se describe a continuación. Sea $X_n^N(t)$, $n = 1, 2, \ldots, N$, $t \in \mathbb{N}_0$, la clase o estado del $n$-ésimo objeto en el tiempo $t$ y que toma valores en el conjunto $S = \{1, 2, \ldots, s\} \subseteq \mathbb{N}$. Existe un controlador central quien, en cada etapa, influye en el comportamiento de los objetos mediante la elección de una acción o control $a_t$ elegidas dentro de un conjunto de Borel $A$. Suponemos también que los objetos comparten un ambiente en común que influye en el comportamiento del sistema. Llamemos $C^N(t) \in \mathbb{R}^d$ al contexto del ambiente al tiempo $t \in \mathbb{N}$. Una vez que el ambiente es especificado, el comportamiento y la evolución de los objetos se consideran independientes. Concretamente, consideremos que la evolución del proceso $\left\{X_n^N(t)\right\}_{t\in\mathbb{N}_0}$ está dado de acuerdo a una ecuación en diferencias estocástica, homogénea en $N$, de la forma

$$X_n^N(t+1) = F(X_n^N(t), C^N(t), a_t, \xi_t^n), \quad t = 0, 1, ..., \tag{2.1}$$

donde $F : S \times \mathbb{R}^d \times A \times \mathbb{R} \to S$ es una función medible (conocida para el controlador) y $\{\xi_t^n : t \geq 0, n = 1, 2, \ldots N\}$ es una familia de variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.) con densidad común $\rho$ *desconocida* para el controlador, y definida en un espacio de probabilidad $(\Omega, \mathcal{F}, P)$. Como consecuencia de las definiciones anteriores, es posible definir la ley de transición $K_\rho$ de cada objeto en términos de la función $F$, de la siguiente manera: para todo $n = 1, 2, \ldots, N$,

$$\begin{aligned} K_{ij}^\rho(a,c) &:= P\left[X_n^N(t+1) = j \,\middle|\, X_n^N(t) = i, a_t = a, C^N(t) = c\right] \\ &= \int_{\mathbb{R}} I_j[F(i,c,a,z)]\rho(z)dz, \quad i,j \in S,\ (a,c) \in A \times \mathbb{R}^d, \end{aligned} \tag{2.2}$$

donde $I_B$ representa la función indicadora para el conjunto $B$. Esta relación define la ley de transición mediante el kernel estocástico $K_\rho = K_\rho(a,c) = \left[K_{ij}^\rho(a,c)\right]$. Notemos que $K_\rho$ representa la distribución condicional común de los estados.

A lo largo de este trabajo supondremos que los objetos son observables por el controlador de acuerdo a la clase en que se encuentran, de tal manera que el controlador puede determinar el número de objetos en cada una de las clases $i \in S$ y calcular su proporción. En este sentido, el comportamiento del sistema puede ser reformulado analizando la proporción de objetos en cada clase. Sea $M_i^N(t)$ la proporción de objetos en la clase $i \in S$ al tiempo $t$, esto es,

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

Denotamos por $\vec{M}^N(t)$ al vector cuyas componentes son las proporciones $M_i^N(t)$'s:

$$\vec{M}^N(t) = (M_1^N(t), M_2^N(t), \ldots, M_s^N(t)). \tag{2.3}$$

Observemos que $\vec{M}^N(t) \in \mathbb{P}_N(S) := \{p \in \mathbb{P}(S) : Np(i) \in \mathbb{N}, \forall i \in S\} \subset \mathbb{P}(S)$ y que $\mathbb{P}_N(S)$ es un conjunto finito.

Supondremos que el contexto del ambiente es un sistema dinámico cuya evolución está determinada por la ecuación en diferencias:

$$C^N(t+1) = g(C^N(t), \vec{M}(t+1), a_t),\ t \in \mathbb{N}_0, \tag{2.4}$$

donde $g : \mathbb{R}^d \times \mathbb{P}(S) \times A \to \mathbb{R}^d$ es una función conocida. También supondremos que la evolución de $\vec{M}^N(\cdot)$ depende de la ley de transición $K_\rho$ de los objetos. Concretamente, supondremos que existe una función medible $G_\rho^N : \mathbb{P}_N(S) \times \mathbb{R}^d \times A \times \mathbb{R}^N \to \mathbb{P}_N(S)$ tal que

$$\vec{M}^N(t+1) = G_\rho^N(\vec{M}^N(t), C^N(t), a_t, \vec{w}_t), \tag{2.5}$$

donde $\{\vec{w}_t\}$ es una sucesión de vectores aleatorios independientes e idénticamente distribuidos (i.i.d.) en $\mathbb{R}^N$ con distribución común $\theta$.

Sea $\mathbb{Y}_N := \mathbb{P}_N(S) \times \mathbb{R}^d$ y $H_\rho^N : \mathbb{Y}_N \times A \times \mathbb{R}^N \to \mathbb{Y}_N$ la función definida como

$$H_\rho^N(y, a, \vec{w}) := (G_\rho^N(y, a, \vec{w}), g(c, G_\rho^N(y, a, \vec{w}), a)). \tag{2.6}$$

Entonces, tomando $y^N(t) := (M^N(t), C^N(t))$, de acuerdo a (2.4) y (2.5), $H_\rho^N$ define la dinámica del proceso $\{y^N(t)\}$; esto es,

$$\begin{aligned} y^N(t+1) &= (G_\rho^N(y^N(t), a_t, \vec{w}_t), g(C^N(t), \vec{M}^N(t+1), a_t)) \\ &= (G_\rho^N(y^N(t), a_t, \vec{w}_t), g(C^N(t), G_\rho^N(y^N(t), a_t, \vec{w}_t), a_t)) \\ &= H_\rho^N\left(y^N(t), a_t, \vec{w}_t\right). \end{aligned} \tag{2.7}$$

Finalmente, asumiremos que el costo por etapa depende de la configuración del sistema dada por $y^N(t)$ y de las acciones seleccionadas por el controlador central. Definiendo el espacio $\mathbb{Y} := \mathbb{P}(S) \times \mathbb{R}^d$ y observando que $\mathbb{Y}_N := \mathbb{P}_N(S) \times \mathbb{R}^d \subset \mathbb{Y}$, el costo por etapa puede ser representado por una función medible

$$r : \mathbb{Y} \times A \to \mathbb{R}. \tag{2.8}$$

## 2.3. $N$-Modelo de control de Markov

De acuerdo a la descripción del sistema de interacción de objetos, es posible aplicar la teoría de los procesos de control de Markov para formularlo como un problema de control óptimo de Markov.

Definamos el modelo de control de Markov asociado al sistema de $N$ objetos ($N$-MCM) como:

$$\mathcal{M}_N := \left(\mathbb{Y}_N, A, H_\rho^N, \theta, r\right) \tag{2.9}$$

El modelo de control $\mathcal{M}_N$ representa un sistema estocástico controlado que se observa en los tiempos $t = 0, 1, \ldots$, el cual evoluciona de la siguiente manera: sea $y^N(t) = (\vec{M}^N(t), C^N(t)) \in \mathbb{Y}_\mathbb{N}$ el estado del sistema al tiempo $t$ y $a_t$ la acción aplicada por el controlador al tiempo $t$. Si $y^N(t) = y \in \mathbb{Y}_N$ y $a_t = a \in A$ entonces: 1) Se incurre un costo $r(y, a)$; 2) el sistema se mueve a un nuevo estado

$y' = y^N(t+1) = (\vec{M}^N(t+1), C^N(t+1))$ de acuerdo a la ley de transición

$$\begin{aligned} Q_\rho(B|y,a) &= Prob\left[y^N(t+1) \in B | y^N(t) = y, a_t = a\right], \quad B \in \mathcal{B}(\mathbb{Y}_N) \\ &= \int_{\mathbb{R}^N} I_B\left[H_\rho^N(y,a,\vec{w})\right] \theta(d\vec{w}), \end{aligned}$$

donde $H_\rho^N$ es la función definida en (2.6). Una vez que el sistema se encuentra en el estado $y'$, se aplica un nuevo control $a' \in A(y')$ y se repite el proceso. Además, vamos a suponer que los costos por etapa son acumulados durante la evolución del sistema con un horizonte infinito usando un criterio de costo descontado, por lo que las acciones tomadas por el controlador central están dirigidas a minimizar el costo esperado total descontado que introduciremos mas adelante.

## 2.4. Optimalidad en el $N$-MCM

En esta parte presentaremos las componentes que definen el problema de control óptimo, así como las hipótesis que garantizan la existencia de políticas óptimas bajo el criterio de costo descontado asociado al $N$-MCM (2.9).

**Políticas de control:** Como es usual, las acciones aplicadas por el controlador son seleccionadas de acuerdo a reglas conocidas como políticas de control, las cuales defnimos a continuación.

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

$$h_t^N = (y^N(0), a_0, \vec{w}_0, \ldots, y^N(t-1), a_{t-1}, \vec{w}_{t-1}, y^N(t)),$$

donde $y^N(t) = \left(\vec{M}^N(t), C^N(t)\right)$. Una política de control es una sucesión $\pi^N = \left\{\pi_t^N\right\}$ de kérneles estocásticos $\pi_t^N$ en $A$ dado $\mathbb{H}_t^N$ tales que $\pi_t^N\left(A|h_t^N\right) = 1$ para cada $h_t^N \in \mathbb{H}_t^N$, $t \in \mathbb{N}_0$. Denotaremos por $\Pi^N$ al conjunto de todas las políticas de control.

Ahora, sea $\mathbb{F}$ el conjunto de las funciones medibles $f : \mathbb{Y} \to A$ y $\mathbb{F}^N := \mathbb{F}|_{\mathbb{Y}_N}$ la restricción de $\mathbb{F}$ en $\mathbb{Y}_N$. Diremos que una política $\pi^N \in \Pi^N$ es una política de Markov (determinista) si existe una sucesión $\{f_t^N\} \subseteq \mathbb{F}^N$ tal que para cada $t \in \mathbb{N}_0$ y $h_t^N \in \mathbb{H}_t^N$, $\pi_t^N\left(\cdot|h_t^N\right) = \delta_{f_t^N(y^N(t))}(\cdot)$. En este caso $\pi^N$ toma la forma $\pi^N = \{f_t^N\}$. En particular, si $f_t^N \equiv f^N$ para alguna $f^N \in \mathbb{F}$ y para todo $t \in \mathbb{N}_0$, decimos que $\pi^N$ es una política *estacionaria*. Denotamos por $\Pi_M^N$ al conjunto de las políticas de Markov, y siguiendo una convención estándar, usaremos la misma notación de $\mathbb{F}^N$ para denotar al conjunto de las políticas estacionarias.

*Observación* 2.4.1. **(a)** *Denotamos por $\Pi_M$ al conjunto de las políticas de Markov deterministas cuando usamos $\mathbb{F}$ en lugar de $\mathbb{F}^N$; esto es, $\Pi_M$ es la familia de sucesiones de funciones $\{f_t\} \subset \mathbb{F}$. Observemos que una política $\pi = \{f_t\} \in \Pi_M$ cuyos elementos $f_t$ están restringidos a $\mathbb{Y}_N$ se convierte en un elemento de $\Pi^N$.*

**(b)** *Por argumentos estándares (vea [20]), para cada $\pi^N \in \Pi^N$ y estado inicial $y^N(0) = y \in \mathbb{Y}_N$, existe un espacio de probabilidad $\left(\Omega', \mathcal{F}', P_y^{\pi^N}\right)$ donde $\Omega' := \left(\mathbb{Y}_N \times A \times \mathbb{R} \times \mathbb{R}^N\right)^\infty$, $\mathcal{F}'$ su respectiva $\sigma-$álgebra producto y $P_y^{\pi^N}$ es una medida de probabilidad que satisface las siguientes propiedades. Para cada $t \in \mathbb{N}_0$:*

**(i)** $P_y^{\pi^N}(y^N(0) \in B) = \delta_y(B), \quad B \in \mathcal{B}(\mathbb{Y}_N)$,

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

**(iii)** *Propiedad de Markov:*

$$\begin{aligned} P_y^{\pi^N}\left[y^N(t+1) \in B|h_t^N, a_t\right] &= Q_\rho(B|y^N(t), a_t) \\ &= \int_{\mathbb{R}^N} I_B\left[H_\rho^N\left(y^N(t), a_t, \vec{w}\right)\right]\theta(d\vec{w}), \quad (2.10) \\ &\quad B \in \mathcal{B}(\mathbb{Y}_N). \end{aligned}$$

**Criterio de optimalidad descontado.** Para cada política de control $\pi^N \in \Pi^N$ y estado inicia $y^N(0) = y \in \mathbb{Y}_N$, definimos el costo total esperado descontado como

$$V^N(\pi^N, y) := E_y^{\pi^N} \sum_{t=0}^{\infty} \alpha^t r(y^N(t), a_t), \tag{2.11}$$

donde $\alpha \in (0,1)$ es el factor de descuento y $E_y^{\pi^N}$ representa el operador esperanza respecto a la medida de probabilidad $P_y^{\pi^N}$ inducida por la política $\pi^N$ y el estado inicial $y^N(0) = y$. Diremos que $\pi_*^N$ es óptima para el $N$-MCM si y solo si

$$V_*^N(y) := \inf_{\pi^N \in \Pi^N} V^N(\pi^N, y) = V^N(\pi_*^N, y) \quad \forall y \in \mathbb{Y}_N. \tag{2.12}$$

En este caso, $V_*^N$ le llamaremos la $N-$*función de valor*.

Por lo tanto, el problema de optimización asociado al sistema de interacción de $N$ objetos en el que estamos interesados en analizar es precisamente encontrar una política $\pi_*^N$ que satisfaga (2.12), el cual, en la terminología de la teoría de los procesos de control de Markov, constituye el problema de control óptimo en el modelo $\mathcal{M}_N$.

Ahora impondremos a nuestro modelo $\mathcal{M}_N$ algunas condiciones de compacidad y continuidad para establecer posteriormente el resultado que proporciona una caracterización de las políticas óptimas y de la $N$-función de valor en términos de la solución de cierta ecuación funcional llamada la $N$-ecuación de optimalidad (vea, por ejemplo, [19, 38]):

*Hipótesis* 1. **(a)** El espacio de control $A$ es un espacio compacto de Borel, cuya métrica será denotada por $d_A$.

**(b)** La función $g$ en (2.4) es una función Lipschitz con constante $L_g$; esto es, para $c, c' \in \mathbb{R}^d$, $\vec{m}, \vec{m}' \in \mathbb{P}(S)$, $a, a' \in A$,

$$\|g(c, \vec{m}, a) - g(c', \vec{m}', a')\|_\infty^2 \leq L_g \max\left\{\|c - c'\|_\infty^2, \|\vec{m} - \vec{m}'\|_\infty^1, d_A(a, a')\right\}. \tag{2.13}$$

Sin pérdida de generalidad, podemos suponer que $L_g \geq 1$.

**(c)** La función $a \longmapsto H_\rho^N(y,a,w)$ definida en (2.6) es continua para toda $y \in \mathbb{Y}_N$ y $w \in \mathbb{R}^N$.

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

$$|r(y,a)| \leq R \ \ \forall (y,a) \in \mathbb{Y} \times A,$$

y para cada $y, y' \in \mathbb{Y}$,

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

Como consecuencia de esta hipótesis tenemos el siguiente resultado, el cual es bien conocido en la literatura de los procesos de control de Markov (ver [5, 19, 20, 38]).

**Proposición 2.4.1.** Suponiendo que la Hipótesis 1 se cumple. Entonces

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

$$V_*^N(y) = \min_{a \in A} \left\{ r(y,a) + \alpha \int_{\mathbb{R}^N} V_*^N \left[ H_\rho^N\left(y,a,\vec{w}\right) \right] \theta(d\vec{w}) \right\}, \quad y \in \mathbb{Y}_N. \tag{2.14}$$

Además,

$$\left| V_*^N(y) \right| \leq \frac{R}{1-\alpha}, \quad y \in \mathbb{Y}_N.$$

**(b)** Existe $f_*^N \in \mathbb{F}^N$ tal que $f_*^N(y) \in A$ alcanza el mínimo en (2.14), i.e.,

$$V_*^N(y) = r(y, f_*^N) + \alpha \int_{\mathbb{R}^N} V_*^N \left[ H_\rho^N\left(y, f_*^N, \vec{w}\right) \right] \theta(d\vec{w}), \quad y \in \mathbb{Y}_N, \tag{2.15}$$

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

La Proposición 2.4.1 representa la ventaja principal de formular el problema de control de sistemas de iteracción de objetos como un problema de control de Markov, ya que ésta proporciona una manera sencilla de modelar el problema, al menos formalmente, por medio de la programación dinámica. Sin embargo, en el contexto de nuestro trabajo, es decir, cuando $N$ es un número muy grande y la densidad $\rho$ es desconocida, la aplicación de la Proposición 2.4.1 es muy limitada. Esto obliga a recurrir a métodos alternativos para resolver el problema o proporcionar métodos de aproximación razonables a su solución. Estos temas serán tratados en el próximo capítulo.

# Capítulo 3

# Estimación y Control en Sistemas de Interacción de objetos bajo un esquema de campo medio

## 3.1. Introducción

El problema que abordamos en este capítulo es de optimalidad bajo el supuesto de que el proceso de perturbaciones aleatorias $\{\xi_t\}$ que definen la dinámica de los objetos (2.2) son v.a.i.i.d. y observables pero con densidad desconocida $\rho$.

Como se comentó en el capítulo anterior, la Proposición 2.4.1 proporciona un marco de trabajo flexible para el análisis de optimalidad en el sistema de interacción de objetos. Sin embargo, desde el punto de vista práctico, su utilidad es muy limitada cuando $N$ es muy grande ($N \sim \infty$) o por desconocimiento parcial o total de la densidad $\rho$. Por ejemplo, para analizar las ecuaciones (2.14) y (2.15), nos enfrentamos con una integral múltiple de dimensión $N$ la cual podría resultar muy difícil de calcular, si no es que imposible; esto sumado al hecho de que la dinámica del sistema depende de la densidad $\rho$ la cual suponemos desconocida. Para superar ambos obstáculos recurrimos a la teoría de campo medio y a técnicas de estimación estadística de $\rho$. Especificamente, como un primer paso, introduciremos un modelo de control adecuado $\mathcal{M}$ que representa el "modelo límite"de $\mathcal{M}_N$ cuando $N \to \infty$; este nuevo modelo es conocido como *modelo de control de campo medio*, el cual también depende de la densidad desconocida $\rho$. Una vez que se tiene bien definido el modelo $\mathcal{M}$, se plantea el problema de control correspondiente, es decir el problema de control de campo medio. Sin embargo, como $\mathcal{M}$ depende de $\rho$, proponemos métodos de estimación y control para la construcción de las políticas, y el objetivo es medir su desviación de optimalidad cuando estas son usadas para controlar el sistema original correspondiente al modelo $\mathcal{M}_N$. En otras palabras, en el modelo de campo medio nos enfrentamos a un problema de control adaptado, cuyas políticas resultan ser "cercanamente"óptimas en $\mathcal{M}_N$ en un sentido asintótico cuando $N \to \infty$.

En términos generales, nuestro método consiste en considerar a $\mathcal{M}$ como un modelo de aproximación para $\mathcal{M}_N$.

## 3.2. El modelo de control de campo medio

Recordemos que $\mathbb{Y} := \mathbb{P}(S) \times \mathbb{R}^d$. Consideremos un sistema de control determinista $\{(\vec{m}(t), c(t))\}$ con valores en $\mathbb{Y}$ que depende implícitamente de la densidad $\rho$ en (2.2), y cuya dinámica se rige por medio de las ecuaciones en diferencias

$$\vec{m}(t+1) = G_\rho(\vec{m}(t), c(t), a_t), \tag{3.1}$$

$$c(t+1) = g(c(t), \vec{m}(t+1), a_t), \tag{3.2}$$

donde $(\vec{m}(0), c(0)) = (\vec{m}, c) \in \mathbb{Y}$ representa la condición inicial, $a_t \in A$ es el control (o acción) tomada al tiempo $t$, $g : \mathbb{R}^d \times \mathbb{P}(S) \times A \to \mathbb{R}^d$ es la función definida en (2.4) y $G_\rho : \mathbb{P}(S) \times \mathbb{R}^d \times A \to \mathbb{P}(S)$ es una función de Lipschitz conocida (que depende de $\rho$) con constante $L_G$; esto es, para $\vec{m}, \vec{m}' \in \mathbb{P}(S)$, $c, c' \in \mathbb{R}^d$, y $a, a' \in A$,

$$\|G_\rho(\vec{m}, c, a) - G_\rho(\vec{m}', c', a')\|_\infty^1 \leq L_G \operatorname{máx}\left\{ \|\vec{m} - \vec{m}'\|_\infty^1, \|c - c'\|_\infty^2, d_A(a, a') \right\}. \tag{3.3}$$

Debido a la naturaleza determinista del proceso (3.1)-(3.2), es claro que el proceso $\{\vec{m}(t), c(t)\}$ está completamente determinado por la sucesión de acciones $\{a_t\} \subset A$ y por la condición inicial $(\vec{m}, c) \in \mathbb{Y}$. Además, supondremos (ver Hipótesis 3 mas adelante) que el proceso $y(t) := (\vec{m}(t), c(t))$ representa el límite de campo medio; esto es, $y(t)$ será el proceso límite de $y^N(t) := (\vec{M}^N(t), C^N(t))$ en (2.7) cuando $N$ tiende a infinito.

Sea $H_\rho : \mathbb{Y} \times A \to \mathbb{Y}$ la función que define la dinámica del proceso $\{(\vec{m}(t), c(t))\}$; esto es,

$$H_\rho(y, a) := (G_\rho(\vec{m}, c, a), g(c, G_\rho(\vec{m}, c, a), a)), \quad y = (\vec{m}, c) \in \mathbb{Y}, \; a \in A. \tag{3.4}$$

De (3.1) and (3.2), podemos escribir

$$\begin{aligned} y(t+1) &= (G_\rho(\vec{m}(t), c(t), a_t), g(c(t), \vec{m}(t+1), a_t)) \\ &=: H_\rho(y(t), a_t), \quad t \geq 0, \end{aligned} \tag{3.5}$$

con $y(0) = (\vec{m}, c) \in \mathbb{P}(S) \times \mathbb{R}^d$. Un cálculo directo nos lleva a que $H_\rho$ es una función de Lipschitz como lo establece el siguiente resultado.

**Lema 3.2.1.** Bajo la Hipótesis 1, la función $H_\rho : \mathbb{Y} \times A \to \mathbb{Y}$ es Lipschitz con constante $L_{H_\rho} = \operatorname{máx}\{L_g, L_g L_G\}$.

*Demostración.* Sea $(y,a),(y',a') \in \mathbb{Y}\times A$ con $y=(\vec{m},c)$ y $y=(\vec{m}',c')$, y recordemos que $G_\rho$ y $g$ son funciones de Lipschitz. Entonces, por (3.4),

$$\|H_\rho(y,a)-H_\rho(y',a')\|_\infty =$$

$$\begin{aligned}
&\|(G_\rho(\vec{m},c,a),g(c,G_\rho(\vec{m},c,a),a))-(G_\rho(\vec{m}',c',a'),g(c',G_\rho(\vec{m}',c',a'),a'))\|_\infty\\
&=\text{máx}\{\|G_\rho(\vec{m},c,a)-G_\rho(\vec{m}',c',a')\|_\infty^1, \|g(c,G_\rho(\vec{m},c,a),a)-g(c',G_\rho(\vec{m}',c',a')\|_\infty^2\}\\
&\leq \text{máx}\{L_G,L_gL_G\}\,\text{máx}\{\|\vec{m}-\vec{m}'\|_\infty^1, \|c-c'\|_\infty^2, d_A(a,a')\}\\
&=\text{máx}\{L_G,L_gL_G\}\,\text{máx}\{\|y-y'\|_\infty, d_A(a,a')\}
\end{aligned}$$

Por lo que

$$\|H_\rho(y,a)-H_\rho(y',a')\|_\infty \leq L_{H_\rho}\,\text{máx}\left\{\|y-y'\|_\infty, d_A(a,a')\right\}, \tag{3.6}$$

con $L_{H_\rho}=\text{máx}\left\{L_g, L_gL_G\right\}$. □

Usando la función de costo por etapa $r$ definida para el $N$-MCM (2.8), podemos entonces definir el modelo de control de campo medio como

$$\mathcal{M}=(\mathbb{Y},A,H_\rho,r),$$

el cual tiene una interpretación similar al $N-$MCM $\mathcal{M}_N$, que describimos en (2.9).

## 3.3. Optimalidad en el campo medio

En esta sección presentamos resultados ya conocidos relativos a la optimalidad del sistema controlado (3.1)-(3.2) cuando se usa el criterio descontado determinista. Básicamente estos resultados muestran caracterizaciones de las políticas óptimas y de la función de valor correspondiente, como soluciones de una ecuación funcional asociada al modelo de control de campo medio.

Es bien conocido (consulte, por ejemplo, [5]) que para sistemas de control deterministas, una política de control $\pi$ es una sucesión de reglas de decisión (o selectores) $\pi = \{f_t\} \subset \mathbb{F}$. Por lo tanto, de acuerdo a la Observación 2.4.1(a), podemos naturalmente considerar el conjunto $\Pi_M$ como el conjunto de las políticas de control para el modelo $\mathcal{M}$. Así, dada una política de control $\pi \in \Pi_M$ junto con la condición inicial $y(0) = y \in \mathbb{Y}$, definimos el costo total descontado para el modelo de campo medio como

$$v(\pi, y) = \sum_{t=0}^{\infty} \alpha^t r\left(y(t), a_t\right). \tag{3.7}$$

Entonces, el problema de control de campo medio es encontrar una política $\pi_* \in \Pi_M$ tal que

$$v_*(y) := \inf_{\pi \in \Pi_M} v(\pi, y) = v(\pi_*, y), \quad y \in \mathbb{Y}, \tag{3.8}$$

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

Observemos que de la continuidad de la función $H_\rho$ (vea Lema 3.2.1), la compacidad del espacio de control $A$ y la continuidad de la función de costo por etapa $r$, podemos establecer el siguiente resultado relacionado con la función de valor (ver, por ejemplo, [5, 19, 38]).

**Proposición 3.3.1.** a) La función de valor $v_*$ satisface la ecuación de optimalidad de campo medio

$$v_*(y) = \min_{a \in A} \left\{ r(y, a) + \alpha v_* \left[ H_\rho\left(y, a\right) \right] \right\}, \quad y \in \mathbb{Y}. \tag{3.9}$$

Equivalentemente,

$$\min_{a \in A} \Phi(y, a) = 0, \quad y \in \mathbb{Y},$$

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

$$\Phi(y,a) := r(y,a) + \alpha v_* \left[ H_\rho\left(y,a\right)\right] - v_*(y). \tag{3.10}$$

Además,

$$|v_*(y)| \leq \frac{R}{1-\alpha}, \quad y \in \mathbb{Y},$$

donde $R$ es la cota uniforme de la función de costo $r$. b) Existe $f^* \in \mathbb{F}$ que alcanza el mínimo en (3.9), es decir,

$$v_*(y) = r(y,f^*) + \alpha v_* \left[ H_\rho\left(y,f^*\right)\right], \quad y \in \mathbb{Y}, \tag{3.11}$$

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

*Observación* 3.3.1. *Sea* $\{(y_t,a_t)\}$ *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,* $\pi^*$ *es una política óptima si y solo si* $\Phi(y_t, f^*(y_t)) = 0$, *para toda* $t \in \mathbb{N}_0$.

La función de valor y la política óptima están bien caracterizadas por la Proposición 3.3.1 y la Observación 3.3.1, pero no son directamente aprovechables pues dependen de la densidad $\rho$, la cual en nuestro caso es desconocida. Sin embargo, mostraremos que bajo ciertas condiciones, y usando métodos adecuados de estimación y control se pueden encontrar políticas con buenas propiedades de optimalidad. Este punto lo estudiamos en la siguiente sección.

## 3.4. Estimación y control en el campo medio

Si la densidad $\rho$ es desconocida, entonces la función $H_\rho$ que define la dinámica del proceso de control de campo medio (3.5), también es desconocida. Luego, bajo la hipótesis de observabilidad del proceso de perturbaciones aleatorias $\{\xi_t\}$, es posible plantear un problema de control adaptado en el modelo de campo medio $\mathcal{M}$. Es decir, el controlador puede implementar métodos estadśiticos de estimación de $\rho$ y por lo tanto de $H_\rho$, junto con técnicas de control, con el fin de obtener información sobre la evolución del sistema para obtener políticas óptimas. Específicamente, antes de elegir la acción $a_t$ al tiempo $t$, el controlador obtiene un estimador $\rho_t$ de $\rho$, y por lo tanto un estimador $H_t = H_{\rho_t}$ de la función $H_\rho$, y entonces sus decisiones son adaptadas a este estimador para obtener un control $a_t = a_t(\rho_t)$.

Para fijar ideas, supongamos que $\xi_0, \xi_1, ..., \xi_{k-1}$ son realizaciones independientes de una variable aleatoria con densidad desconocida $\rho$ y $\rho_k(\cdot) := \rho_k(\cdot; \xi_0, \xi_1, ..., \xi_{k-1})$ un estimador de esta densidad tal que,

$$\int_{\mathbb{R}} |\rho_k(z) - \rho(z)|\, dz \to 0 \ \text{ c.s.} \tag{3.12}$$

y

$$\sup_{(y,a)\in\mathbb{Y}\times A} \|G_{\rho_k}(y,a) - G_\rho(y,a)\|_\infty^1 \to 0 \ \text{ c.s.}, \tag{3.13}$$

donde $y = (\vec{m}, c)$, y $G_{\rho_k}$, $k \in \mathbb{N}$ es la función que define la dinámica del proceso $\{\vec{m}(t)\}$ (vea (3.1)) cuando se usa la densidad $\rho_k$ en vez de $\rho$. Así, $G_{\rho_k}$ define un nuevo proceso estimado el cual es generado por la función (vea (3.4) y (3.5))

$$H_k(y,a) := \left(G_{\rho_k}(y,a), g(c, G_{\rho_k}(y,a), a)\right), \quad y = (\vec{m}, c) \in \mathbb{Y},\ a \in A.$$

**Lema 3.4.1.** Bajo la Hipótesis 1, se cumplen las siguientes afirmaciones:

$$\sup_{(y,a)\in\mathbb{Y}\times A} \|H_k(y,a) - H_\rho(y,a)\|_\infty \to 0 \ \text{ c.s., cuando } \ k \to \infty, \tag{3.14}$$

$$E_y^\pi\left[\sup_{(x,a)\in\mathbb{Y}\times A}\left\|H_k(x,a)-H_\rho(x,a)\right\|_\infty\right]\to 0,\ \text{cuando } k\to\infty. \tag{3.15}$$

*Demostración.* Dado que $g$ es una función de Lipschitz, tenemos que para cada $y=(\vec{m},c)\in\mathbb{Y},\ a\in A$,

$$\left\|g(c,G_{\rho_k}(y,a),a)-g(c,G_\rho(y,a),a)\right\|_\infty^2\le L_g\left\|G_{\rho_k}(y,a)-G_\rho(y,a)\right\|_\infty^1. \tag{3.16}$$

Entonces, combinando (3.13) y (3.16), obtenemos

$$\sup_{(y,a)\in\mathbb{Y}\times A}\left\|g(c,G_{\rho_k}(y,a),a)-g(c,G_\rho(y,a),a)\right\|_\infty^2\to 0 \text{ c.s.},\ \text{ cuando } k\to\infty.$$

De esta manera podemos ver fácilmente que (3.14) se cumple. Además, para cada $\pi\in\Pi_M$ y $y\in\mathbb{Y}$, de (3.14) y el teorema de la convergencia dominada, podemos concluir (3.15) debido a que $\rho_k$ no depende de $\pi$ ni de $y$. $\square$

Sea $\{v_k\}$ una sucesión de funciones $v_k:\mathbb{Y}\to\mathbb{R}$ en el espacio $\mathbb{C}_b(\mathbb{Y})$ definidas como:

$$\begin{aligned} v_0&\equiv 0;\\ v_k(y)&=\min_{a\in A}\left\{r(y,a)+\alpha v_{k-1}\left[H_k\left(y,a\right)\right]\right\},\quad k\in\mathbb{N},\ y\in\mathbb{Y}. \end{aligned} \tag{3.17}$$

Entonces, tomando en cuenta que la función $(y,a)\to H_k\left(y,a\right),\ k\in\mathbb{N}$, es continua y que $A$ es compacto, por teoremas de selección medible de uso estándar (vea, por ejemplo, Proposition D5(a) en [20]), para cada $k\in\mathbb{N}$, existe $\hat{f}_k\in\mathbb{F}$ (que depende de $\rho_k$), tal que

$$v_k(y)=r(y,\hat{f}_k)+\alpha v_{k-1}\left[H_k\left(y,\hat{f}_k\right)\right],\quad y\in\mathbb{Y}. \tag{3.18}$$

**Definición 3.4.1.** Sean $\hat{f}_k,\ k\in\mathbb{N}$, los selectores que satisfacen la relación (3.18). Definimos la política de control $\hat{\pi}=\{\hat{f}_k\}\in\Pi_M, k\in\mathbb{N}_0$, donde $\hat{f}_0(y)$ es cualquier acción fija en $A$.

Observemos que esta política es completetamente calculable para el controlador, y por tanto, de acuerdo a nuestro objetivo, estamos interesados en el estudio de su optimalidad. A las políticas que se definen por medio de procesos de estimación y control se les conoce como *políticas adaptadas* en la literatura de los procesos de control de Markov. Generalmente, este tipo de políticas no resultan ser óptimas bajo el criterio de costo descontado ya que depende fuertemente de las decisiones tomadas en las primeras etapas, que en nuestro caso, es justo donde el proceso de estimación estadística tiene poca información de la dinámica desconocida. Esto implica que, en general, no es posible asegurar que $\hat{\pi}$ es una política óptima en el sentido usual para el modelo de campo medio. Por lo tanto, necesitamos hacer uso de un criterio de optimalidad más débil para analizar su optimalidad, el cual es motivado por el comentario en la Observación 3.3.1 (vea, por ejemplo, [17, 19, 23, 36] para mas información acerca de este criterio de optimalidad).

**Definición 3.4.2.** Decimos que una política $\pi \in \Pi_M$ es *eventualmente óptima* para el modelo de control de campo medio (o simplemente eventualmente óptima) si y sólo si,

$$\lim_{t\to\infty} E_y^{\pi}\Phi(y(t), a_t) = 0, \quad y \in \mathbb{Y},$$

para cualquier condición inicial $y(0) = y \in \mathbb{Y}$. Donde $\Phi$ es la función de discrepancia definida en (3.10).

Antes de establecer un resultado importante en este capítulo, necesitamos imponer el siguiente requerimiento técnico.

*Hipótesis* 2. La constante $L_{H_\rho}$ definida en (3.6) satisface que $\alpha L_{H_\rho} < 1$.

**Teorema 3.4.2.** Bajo las Hipótesis 1 y 2, la política $\hat{\pi}$ obtenida mediante el método iterativo descrito en (3.18), es eventualmente óptima.

El resto de esta sección está enfocada a la demostración del Teorema 3.4.2, la cual esta basada en los siguientes lemas 3.4.3 y 3.4.4.

Sean $\{u_t\} \subset \mathbb{C}_b(\mathbb{Y})$ las funciones de iteración de valores en el campo medio definidas como:

$$u_0 \equiv 0; \tag{3.19}$$

$$u_t(y) = \min_{a \in A} \{r(y,a) + \alpha u_{t-1}[H_\rho(y,a)]\}, \quad t \in \mathbb{N},\ y \in \mathbb{Y}. \tag{3.20}$$

Como se muestra en [5, 19, 38], nuestras hipótesis nos llevan a

$$v_*(y) = \lim_{t \to \infty} u_t(y), \quad y \in \mathbb{Y}, \tag{3.21}$$

donde $v_*$ es la función de valor de campo medio, la cual satisface (3.9).

**Lema 3.4.3.** Suponga que la Hipótesis 1 se cumple. Entonces:

**(a)** Para cada $t \in \mathbb{N}_0$, las funciones $u_t$ generadas por medio de las iteraciones (3.19)-(3.20) son Lipschits continuas con constante

$$L_{u_t} := L_r \sum_{l=0}^{t-1} \left(\alpha L_{H_\rho}\right)^l. \tag{3.22}$$

**(b)** Además, si se satisface la Hipótesis 2, entonces la función de valor de campo medio $v_*$ es Lipschitz continua con constante

$$L_{v_*} = \frac{L_r}{1 - \alpha L_{H_\rho}}, \tag{3.23}$$

donde $L_r$ y $L_{H_\rho}$ son las constantes de Lipschitz de las Hipótesis 1(d) y (3.6), respectivamente.

*Demostración.* (a) Procederemos por inducción. Primero, de (3.19), claramente la parte (a) se cumple para $t=0$. Ahora supongamos que $u_t$ es una función de Lipschitz cuya constante es como en (3.22). Entonces, para $y, y' \in \mathbb{Y}$, de (3.20) tenemos

$$|u_{t+1}(y) - u_{t+1}(y')| \leq \sup_{a \in A} \{|r(y,a) - r(y',a)| + \alpha\, |u_t\,[H_\rho(y,a)] - u_t\,[H_\rho(y',a)]|\}\,.$$

Así, dado que $r$ y $H_\rho$ son funciones de Lipschitz (por las Hipótesis 1(d) y (3.6)), usando la relación (3.22), obtenemos

$$\begin{aligned}
&|u_{t+1}(y) - u_{t+1}(y')| \leq L_r \|y - y'\|_\infty + \alpha L_{u_t} L_{H_\rho} \|y - y'\|_\infty \\
&\leq \left( L_r + \alpha L_{H_\rho} L_r \sum_{l=0}^{t-1} \left(\alpha L_{H_\rho}\right)^l \right) \|y - y'\|_\infty \leq L_r \left( 1 + \sum_{l=0}^{t-1} \left(\alpha L_{H_\rho}\right)^{l+1} \right) \|y - y'\|_\infty \\
&= L_r \sum_{l=0}^{t} \left(\alpha L_{H_\rho}\right)^l \|y - y'\|_\infty\,.
\end{aligned}$$

Por lo tanto, $u_{t+1}$ es una función de Lipschitz con constante

$$L_{u_{t+1}} := L_r \sum_{l=0}^{t} \left(\alpha L_{H_\rho}\right)^l.$$

Este hecho prueba la parte (a).

(b) Para $y, y' \in \mathbb{Y}$, sumando y restando los términos $u_t(y)$ y $u_t(y')$ obtenemos

$$\begin{aligned}
|v_*(y) - v_*(y')| &\leq |v_*(y) - u_t(y)| + |u_t(y) - u_t(y')| + |u_t(y') - v_*(y')| \\
&\leq |v_*(y) - u_t(y)| + L_{u_t} \|y - y'\|_\infty + |u_t(y') - v_*(y')|\,, \quad \forall t \in \mathbb{N}_0,
\end{aligned} \tag{3.24}$$

donde la última desigualdad se debe a la parte (a). Ahora observemos que bajo la Hipótesis 2

$$\lim_{t\to\infty} L_{u_t} = L_r \sum_{l=0}^{\infty} \left(\alpha L_{H_\rho}\right)^l = \frac{L_r}{1 - \alpha L_{H_\rho}}. \tag{3.25}$$

Por lo tanto, haciendo $t \to \infty$ en (3.24), tenemos que de (3.21) y (3.25) se obtiene

$$|v_*(y) - v_*(y')| \leq \frac{L_r}{1 - \alpha L_{H_\rho}} \|y - y'\|_\infty, \qquad y, y' \in \mathbb{Y},$$

esto es, $v_*$ es una función continua de Lipschitz. $\square$

**Lema 3.4.4.** Sea $v_k$ la familia de funciones generada por las iteraciones (3.17), y $v_*$ la función de valor en (3.8) (ver (3.9)). Entonces, bajo las Hipótesis 1 y 2, para cada $\pi \in \Pi_M$ y $y \in \mathbb{Y}$, $E_y^\pi \|v_* - v_k\| \to 0$, cuando $k \to \infty$.

*Demostración.* De (3.9) y (3.17), tenemos, para cada $k \in \mathbb{N}$ y $y \in \mathbb{Y}$,

$$\begin{aligned}
&|v_*(y) - v_k(y)| \leq \alpha \sup_{a \in A} |v_*[H_\rho(y,a)] - v_{k-1}[H_k(y,a)]| \\
&\leq \alpha \sup_{a \in A} |v_*[H_\rho(y,a)] - v_*[H_k(y,a)]| + \alpha \sup_{a \in A} |v_*[H_k(y,a)] - v_{k-1}[H_k(y,a)]|,
\end{aligned}$$

donde en la última desigualdad sumamos y restamos el término $v_*[H_k(y,a)]$. Así, del Lema 3.4.3 y el hecho que $v_*, v_k \in \mathbb{B}(\mathbb{Y})\ \forall k \in \mathbb{N}$,

$$0 \leq \|v_* - v_k\| \leq L_{v_*} \sup_{(y,a) \in \mathbb{Y} \times A} \|H_\rho(y,a) - H_k(y,a)\|_\infty + \alpha \|v_* - v_{k-1}\|, \tag{3.26}$$

lo cual implica

$$E_y^\pi \|v_* - v_k\| \leq L_{v_*} E_y^\pi \left[ \sup_{(y,a) \in \mathbb{Y} \times A} \|H_\rho(y,a) - H_k(y,a)\|_\infty \right] + \alpha E_y^\pi \|v_* - v_{k-1}\|, \tag{3.27}$$

para cada $\pi \in \Pi_M$ y $y \in \mathbb{Y}$. Sea $l := \limsup_{k \to \infty} E_y^\pi \|v_* - v_k\| < \infty$. Así, haciendo $k \to \infty$ in (3.27) y usando la convergencia en (3.15), obtenemos $l \leq \alpha l$. Finalmente, dado que $\alpha < 1$, podemos deducir que $\lim_{k \to \infty} E_y^\pi \|v_* - v_k\| = 0$, lo cual demuestra el resultado. $\square$

### 3.4.1. Demostración del Teorema 3.4.2.

Definamos para cada $k \in \mathbb{N}$, la función de discrepancia aproximada $\Phi_k : \mathbb{Y} \times A \to \mathbb{R}$ como

$$\Phi_k(y,a) := r(y,a) + \alpha v_{k-1}\left[H_k\left(y,a\right)\right] - v_k(y), \quad (y,a) \in \mathbb{Y} \times A.$$

Ahora observemos que para cada $k \in \mathbb{N}$ y $(y,a) \in \mathbb{Y} \times A$,

$$\left|\Phi(y,a) - \Phi_k(y,a)\right| \leq \left|v_*\left[H_\rho(y,a)\right] - v_{k-1}\left[H_k(y,a)\right]\right| + \left|v_*(y) - v_k(y)\right|.$$

Entonces, del Lema 3.4.4, haciendo $k \to \infty$ obtenemos

$$E_y^{\hat{\pi}}\left[\sup_{(y,a)\in\mathbb{Y}\times A}\left|\Phi(y,a) - \Phi_k(y,a)\right|\right] \to 0. \tag{3.28}$$

Por otro lado, observando que $\Phi_k(y, \hat{f}_k(y)) = 0$, $y \in \mathbb{Y}$, tenemos

$$\begin{aligned} 0 \leq \Phi(y(k), \hat{f}_k(y(k))) &= \left|\Phi(y(k), \hat{f}_k(y(k))) - \Phi_k(y(k), \hat{f}_k(y(k)))\right| \\ &\leq \sup_{(y,a)\in\mathbb{Y}\times A}\left|\Phi(y,a) - \Phi_k(y,a)\right|, \end{aligned}$$

Así, de (3.28), obtenemos

$$\lim_{k\to\infty} E_y^{\hat{\pi}}\Phi(y(k), a_k) = 0. \quad \blacksquare$$

## 3.5. Convergencia en el campo medio

En esta sección estudiamos la eficiencia de la política eventualmente óptima $\hat{\pi}$ obtenida en la Sección 3.4; esto es, estamos interesados en analizar la desviación de optimalidad de $\hat{\pi}$ cuando se usa para controlar el proceso original $\left\{y^N(t)\right\}$. Cláramente, tal desviación de optimalidad debe ser medida en términos de la diferencia entre las correspondientes funciones de valor óptimo $V_*^N$ y $v_*$ de los modelos $\mathcal{M}_N$

y $\mathcal{M}$ respectivamente y como fue señalado en la Sección 2, tiene que ser analizada en un sentido asintótico. Este análisis constituye el resultado principal de este trabajo. Para probar tal resultado imponemos las siguientes hipótesis relacionadas con la convergencia de las trayectorias $y^N(\cdot)$ a las trayectorias $y(\cdot)$, definidas en (2.7) y (3.5), respectivamente.

Observemos que de acuerdo a las Proposiciones 2.4.1 y 3.3.1, y de la definición de la política $\hat{\pi}$, podemos restringir nuestro análisis a la clase todas las políticas de Markov $\Pi_M$. Además recordemos que para cualquier $t \in \mathbb{N}_0$ dado,

$$\|y^N(t) - y(t)\|_\infty = \text{máx}\left\{\|\vec{M}^N(t) - \vec{m}(t)\|_\infty^1 \ , \ \|C^N(t) - c(t)\|_\infty^2\right\}. \tag{3.29}$$

*Hipótesis* 3. **(a)** $(\vec{M}^N(0), C^N(0)) = (\vec{m}(0), c(0)) = (\vec{m}_0, c_0) = y \in \mathbb{Y}_N$, para todo $N \in \mathbb{N}$.

**(b)** For cada $y \in \mathbb{Y}_N$, $T \in \mathbb{N}$ y $\varepsilon > 0$, existen constantes positivas $K$ y $\lambda$ tales que

$$\sup_{\pi\in\Pi_M} P_y^\pi\left\{\sup_{0\le t\le T}\left\|y^N(t) - y(t)\right\|_\infty \ge \gamma_T(\varepsilon)\right\} \le KTe^{-\lambda N\varepsilon^2}, \tag{3.30}$$

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

Usaremos la siguiente notación:

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

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

las acciones al tiempo $t$ correspondientes a la aplicación de la política $\pi$ bajo el proceso $\left\{y^N(t)\right\}$ y $\{y(t)\}$, respectivamente.

- Para cada $T \in \mathbb{N}$,

$$Y_T := \sup_{0 \leq t \leq T} \|y^N(t) - y(t)\|_\infty \tag{3.31}$$

y

$$\mathcal{K}(T) := (L_g)^T \text{máx}\{L_g, diam(A)\}, \tag{3.32}$$

donde $L_g \geq 1$ es la constante de Lipschitz en Hipótesis 1(b) y $diam(A) := \sup_{(a,a') \in A \times A} d(a, a')$.

Ahora estamos en condiciones de establecer uno de nuestros resultados principales. Primeramente, proporcionamos una cota para la diferencia entre las funciones de valor $V_*^N$ y $v_*$, que a su vez definen un esquema de aproximación cuando $N \to \infty$. Luego mostramos que la política control $\hat{\pi}$ es eventualmente óptima en el modelo de control $\mathcal{M}_N$ en un sentido asintótico.

**Teorema 3.5.1.** Supongamos que se cumplen las Hipótesis 1, 2 y 3. Entonces:

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

$$\begin{aligned} \sup_{\varphi \in \Pi_M} E_y^{\varphi} \left| V_*^N(y^N(t)) - v_*(y(t)) \right| &\leq \frac{2R\alpha^T}{1-\alpha} \\ &\quad + L_r \frac{1-\alpha^T}{1-\alpha} \left[ KTe^{-\lambda N \varepsilon^2}(1 + \mathcal{K}(T)) + \gamma_T(\varepsilon) \right]. \end{aligned} \tag{3.33}$$

**(b)** La política de control $\hat{\pi} \in \Pi_M$ definida en Definición 3.4.1 es eventualmente asintóticamente óptima para el $N$ modelo de control de Markov $\mathcal{M}_N$ cuando $N \to \infty$; esto es

$$\lim_{t \to \infty} \lim_{N \to \infty} E_y^{\hat{\pi}} \Phi^N(y^N(t), \hat{f}_t) = 0, \tag{3.34}$$

donde

$$\Phi^N(y^N,a) := r(y^N,a) + \alpha \int_{\mathbb{R}^N} V_*^N \left[ H_\rho^N \left( y^N, a, w \right) \right] \theta(dw) - V_*^N(y^N), \quad y^N \in \mathbb{Y}_N \tag{3.35}$$

es la función de discrepancia en el $N-$MCM $\mathcal{M}_N$ (ver (3.10)).

En el resto de esta sección supondremos que las Hipótesis 1, 2 y 3 se cumplen. Basados en este hecho, la prueba del Teorema 3.5.1 será consecuencia de las siguientes proposiciones.

**Proposición 3.5.2.** **(a)** Para cada $\pi \in \Pi_M$ y $T \in \mathbb{N}$,

$$Y_T := \sup_{0 \le t \le T} \|y^N(t) - y(t)\|_\infty \le \mathcal{K}(T). \tag{3.36}$$

**(b)** Para cada $y \in \mathbb{Y}_N$ y $T \in \mathbb{N}$,

$$\sup_{\pi \in \Pi_M} E_y^\pi \left[ \sup_{0 \le t \le T} \|y^N(t) - y(t)\|_\infty \right] \le K T e^{-\lambda N \varepsilon^2}(1 + \mathcal{K}(T)) + \gamma_T(\varepsilon). \tag{3.37}$$

*Demostración.* **(a)** Para obtener (3.36) es suficiente mostrar que para cada $t \in \mathbb{N}_0$ y $\pi \in \Pi_M$

$$\|y^N(t) - y(t)\|_\infty \le (L_g)^{t-1} \max\{L_g, diam(A)\}. \tag{3.38}$$

Entonces nos enfocaremos en obtener (3.38). Notemos que bajo la Hipótesis 3(a) tenemos $a_0^{\pi,N} = a_0^\pi =: a_0 \in A$ y $\|y^N(0) - y(0)\|_\infty = 0$. Por otro lado, dado que $\vec{M}^N(t)$ y $\vec{m}(t)$ son medidas de probabilidad, se sigue que $\|\vec{M}^N(t) - \vec{m}(t)\|_\infty^1 \le 1$, pata todo $t \in \mathbb{N}_0$. Por lo tanto, dado que $L_g \ge 1$, la demostración se reduce a analizar la norma $\|\cdot\|_\infty^2$ en (3.29). En particular, obtendremos (3.38) si mostramos que

$$\|C^N(t) - c(t)\|_\infty^2 \le (L_g)^{t-1} \max\{L_g, diam(A)\}, \quad \forall t \in \mathbb{N}_0. \tag{3.39}$$

Para esto procederemos por inducción. Primero, observemos que de (2.4) y (3.2) obtenemos

$$\begin{aligned}\|C^N(1)-c(1)\|_\infty^2 &= \|g(c_0,\vec{M}^N(1),a_0)-g(c_0,\vec{m}(1),a_0)\|_\infty^2\\ &\le L_g\|\vec{M}^N(1)-\vec{m}(1)\|_\infty^1 \le L_g \quad \text{(por (2.13))}.\end{aligned}$$

También,

$$\begin{aligned}\|C^N(2)-c(2)\|_\infty^2 &= \|g(C^N(1),\vec{M}^N(2),a_1^{\pi,N})-g(c(1),\vec{m}(1),a_1^{\pi})\|_\infty^2\\ &\le L_g\,\text{máx}\left\{\|C^N(1)-c(1)\|_\infty^2,\ \|\vec{M}^N(2)-\vec{m}(2)\|_\infty^1,\ d_A(a_1^{\pi,N},a_1^{\pi})\right\}\\ &\le L_g\,\text{máx}\left\{L_g,\ 1,\ diam(A)\right\} = L_g\,\text{máx}\left\{L_g, diam(A)\right\}.\end{aligned}$$

Ahora, supongamos que (3.39) se cumple para algún $t\in\mathbb{N}$. Entonces

$$\begin{aligned}\|C^N(t+1)-c(t+1)\|_\infty^2 &= \|g(C^N(t),\vec{M}^N(t+1),a_t^{\pi,N})-g(c(t),\vec{m}(t+1),a_t^{\pi})\|_\infty^2\\ \le L_g\,\text{máx}&\left\{\|C^N(t)-c(t)\|_\infty^2,\ \|\vec{M}^N(t+1)-m(t+1)\|_\infty^1,\ d_A(a_t^{\pi,N},a_t^{\pi})\right\} \quad \text{(by (2.13))}\\ &\le L_g\,\text{máx}\left\{(L_g)^{t-1}\,\text{máx}\{L_g,diam(A)\},1,diam(A)\right\} \quad \text{(by (3.39))}\\ &\le (L_g)^t\,\text{máx}\{L_g,diam(A)\}.\end{aligned}$$

Esto demuestra (3.39) lo que a su vez implica (3.38) y (3.36).

**(b)** Observemos que para cada $y\in\mathbb{Y}_N$, $\pi\in\Pi_M$, $T\in\mathbb{N}$ y $\varepsilon>0$, la esperanza en (3.37) satisface (ver (3.31))

$$\begin{aligned}&E_y^\pi[Y_T] = E_y^\pi\left[Y_T\mathbb{I}_{\{Y_T\ge\gamma_T(\varepsilon)\}}+Y_T\mathbb{I}_{\{Y_T<\gamma_T(\varepsilon)\}}\right]\\ &\le E_y^\pi\left[Y_T\mathbb{I}_{\{Y_T\ge\gamma_T(\varepsilon)\}}\right]+\gamma_T(\varepsilon)P_y^\pi(Y_T<\gamma_T(\varepsilon)) \le E_y^\pi\left[Y_T\mathbb{I}_{\{T_T\ge\gamma_T(\varepsilon)\}}\right]+\gamma_T(\varepsilon).\end{aligned} \tag{3.40}$$

Por otro lado, tanto por (3.36) y la no negatividad de $Y_T$, tenemos

$$\frac{Y_T}{1+\mathcal{K}(T)} \le \frac{Y_T}{1+Y_T} \le 1,$$

lo cual implica

$$\frac{Y_T}{1+\mathcal{K}(T)}\mathbb{I}_{\{Y_T\geq\gamma_T(\varepsilon)\}}\leq\mathbb{I}_{\{Y_T\geq\gamma_T(\varepsilon)\}}.$$

Aplicando este hecho junto con la definición de $Y_T$ y la Hipótesis 3(b) tiene como consecuencia

$$\frac{1}{1+\mathcal{K}(T)}E_y^\pi[Y_T\mathbb{I}_{\{Y_T\geq\gamma_T(\varepsilon)\}}]\leq P_y^\pi(Y_T\geq\gamma_T(\varepsilon))\leq KTe^{-\lambda N\varepsilon^2},\quad \pi\in\Pi_M.$$

Finalmente, de (3.40) obtenemos

$$E_y^\pi\left[Y_T\right]\leq KTe^{-\lambda N\varepsilon^2}(1+\mathcal{K}(T))+\gamma_T(\varepsilon),\quad \pi\in\Pi_M, \tag{3.41}$$

y tomando supremo sobre $\pi\in\Pi_M$ en (3.41) demostramos la parte (b). □

A continuación presentaremos algunos resultados relacionados con el criterio de costo descontado con horizonte finito para los modelos $\mathcal{M}_N$ y $\mathcal{M}$ definidos de la siguiente manera: Para cualquier $\pi\in\Pi_M$, $y\in\mathbb{Y}_N\subset\mathbb{Y}$ y $T\in\mathbb{N}$,

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

**Proposición 3.5.3.** Sean $L_r$ y $R$ las constantes en Hipótesis 1(d). Entonces, para cada $y\in\mathbb{Y}_N$, $\varepsilon>0$, $T\in\mathbb{N}$ y $0\leq t\leq T$, los siguientes enunciados se cumplen:

**(a)**

$$\sup_{\pi\in\Pi}E_y^\pi\left|r(y^N(t),a_t^{\pi,N})-r(y(t),a_t^\pi)\right|\leq L_r\left(KTe^{-\lambda N\varepsilon^2}(1+\mathcal{K}(T))+\gamma_T(\varepsilon)\right); \tag{3.42}$$

**(b)**

$$\sup_{\varphi\in\Pi}E_y^\varphi\left[\sup_{\pi\in\Pi}\left|V_T^N(\pi,y^N(t))-v_T(\pi,y(t))\right|\right]$$
$$\leq L_r\frac{1-\alpha^T}{1-\alpha}\left[KTe^{-\lambda N\varepsilon^2}(1+\mathcal{K}(T))+\gamma_T(\varepsilon)\right]; \tag{3.43}$$

**(c)**

$$\sup_{\varphi\in\Pi} E_y^{\varphi}\left[\sup_{\pi\in\Pi}\left|V^N(\pi,y^N(t))-V_T^N(\pi,y^N(t))\right|\right]\leq\frac{R\alpha^T}{1-\alpha}; \tag{3.44}$$

**(d)**

$$\sup_{\varphi\in\Pi} E_y^{\varphi}\left[\sup_{\pi\in\Pi}\left|v(\pi,y(t))-v_T(\pi,y(t))\right|\right]\leq\frac{R\alpha^T}{1-\alpha}. \tag{3.45}$$

*Demostración.* **(a)** Fijemos una política $\pi\in\Pi_M$ y $T\in\mathbb{N}$. Entonces, la Hipótesis 1(d) junto con la Proposición 3.5.2, nos llevan a las siguientes relaciones

$$E_y^{\pi}\left|r(y^N(t),a_t^{\pi,N})-r(y(t),a_t^{\pi})\right|\leq L_rE_y^{\pi}\left[\|y^N(t)-y(t)\|_\infty\right] \tag{3.46}$$
$$\leq L_rE_y^{\pi}\left[\sup_{0\leq t\leq T}\|y^N(t)-y(t)\|_\infty\right]\leq L_r\left(KTe^{-\lambda N\varepsilon^2}(1+\mathcal{K}(T))+\gamma_T(\varepsilon)\right).$$

Esto implica la parte (a).

**(b)** Para cada $\pi\in\Pi_M$,

$$|V_T(\pi,y^N(t))-v_T(\pi,y(t))|=\left|E_{y^N(t)}^{\pi}\left\{\sum_{k=0}^{T-1}\alpha^kr(y^N(k),a_k^{\pi,N})-\sum_{k=0}^{T-1}\alpha^kr(y(k),a_k^{\pi})\right\}\right|$$
$$\leq\sum_{k=0}^{T-1}\alpha^kE_{y^N(t)}^{\pi}\left|r(y^N(k),a_k^{\pi,N})-r(y(t),a_k^{\pi})\right|$$
$$\leq L_r\frac{1-\alpha^T}{1-\alpha}\left[KTe^{-\lambda N\varepsilon^2}(1+\mathcal{K}(T))+\gamma_T(\varepsilon)\right],$$

donde la última desigualdad se sigue de (3.46). Esto implica

$$\sup_{\pi\in\Pi}\left|V_T^N(\pi,y^N(t))-v_T(\pi,y(t))\right|\leq L_r\frac{1-\alpha^T}{1-\alpha}\left[KTe^{-\lambda N\varepsilon^2}(1+\mathcal{K}(T))+\gamma_T(\varepsilon)\right],\ \ \forall t\in\mathbb{N}_0.$$

Tomando esperanza $E_y^{\varphi}$ en ambos lados de esta expresión, y tomando supremo sobre $\varphi\in\Pi_M$, obtenemos (3.43).

(c) Para cada $\pi\in\Pi_M$, tenemos

$$\left|V^N(\pi,y^N(t))-V_T^N(\pi,y^N(t))\right|\leq\left|E_{y^N(t)}^{\pi}\left\{\sum_{k=0}^{\infty}\alpha^kr(y^N(k),a_k^{\pi,N})-\sum_{k=0}^{T-1}\alpha^kr(y^N(k),a_k^{\pi,N})\right\}\right|$$
$$\leq\sum_{k=T}^{\infty}\alpha^kE_{y^N(t)}^{\pi}|r(y^N(k),a_k^{\pi,N})|\leq R\sum_{k=T}^{\infty}\alpha^k\leq\frac{R\alpha^T}{1-\alpha}.$$

Por lo tanto, podemos observar que fácilmente se cumple (3.44).

(d) Se sigue usando los mismos agumentos que en (c). □

### 3.5.1. Demostración del Teorema 3.5.1(a)

Sea $\pi_*^N = \left\{ f_*^N \right\} \in \Pi_M^N$ una política óptima estacionaria para el $N-$MCM $\mathcal{M}_N$ (ver Proposicion 2.4.1(b)), y para un selector arbitrario $\tilde{f} \in \mathbb{F}$, definimos la política estacionaria $\bar{\pi} = \left\{ \bar{f} \right\} \in \Pi_M$, donde $\bar{f} : \mathbb{Y} \to A$ esta dada por

$$\bar{f}(y) = f_*^N(y) I_{\mathbb{Y}_N}(y) + \tilde{f}(y) I_{[\mathbb{Y}_N]^c}(y).$$

Además, sea $\varphi \in \Pi_M$ una política arbitraria y denotemos $y_\varphi^N(t) = y^N(t) \in \mathbb{Y}_N$ y $y_\varphi(t) := y(t) \in \mathbb{Y}$. Observemos que para cada $t \in \mathbb{N}_0$,

$$V_*^N(y^N(t)) = V^N(\pi_*^N, y^N(t)) = V^N(\bar{\pi}, y^N(t)) \leq \sup_{\pi \in \Pi_M} V^N(\pi, y^N(t)).$$

Por lo tanto,

$$V_*^N(y^N(t)) - v_*(y(t)) \leq \sup_{\pi \in \Pi_M} V^N(\pi, y^N(t)) - \inf_{\pi \in \Pi_M} v(\pi, y(t))$$

lo cual a su vez implica

$$\left| V_*^N(y^N(t)) - v_*(y(t)) \right| \leq \sup_{\pi \in \Pi_M} \left| V^N(\pi, y^N(t)) - v(\pi, y(t)) \right|, \quad t \in \mathbb{N}_0.$$

Por lo tanto, para cada $y \in \mathbb{Y}_N$ y $0 \leq t \leq T$,

$$
\begin{aligned}
&E_y^{\varphi} \left| V_*^N(y^N(t)) - v_*(y(t)) \right| \leq E_y^{\varphi} \left[ \sup_{\pi \in \Pi_M} \left| V^N(\pi, y^N(t)) - v(\pi, y(t)) \right| \right] \\
&\leq E_y^{\varphi} \left[ \sup_{\pi \in \Pi_M} \left\{ \left| V^N(\pi, y^N(t)) - V_T^N(\pi, y^N(t)) \right| + \left| V_T^N(\pi, y^N(t)) - v_T(\pi, y(t)) \right| \right.\right. \\
&\left.\left. + \left| v_T(\pi, y(t)) - v(\pi, y(t)) \right| \right\} \right] \\
&\leq E_y^{\varphi} \left[ \sup_{\pi \in \Pi_M} \left| V^N(\pi, y^N(t)) - V_T^N(\pi, y^N(t)) \right| \right] + E_y^{\varphi} \left[ \sup_{\pi \in \Pi_M} \left| V_T^N(\pi, y^N(t)) - v_T(\pi, y(t)) \right| \right] \\
&+ E_y^{\varphi} \left[ \sup_{\pi \in \Pi_M} \left| v_T(\pi, y(t)) - v(\pi, y(t)) \right| \right] \\
&\leq \frac{2R\alpha^T}{1-\alpha} + L_r \frac{1-\alpha^T}{1-\alpha} \left[ KTe^{-\lambda N \varepsilon^2}(1 + \mathcal{K}(T)) + \gamma_T(\varepsilon) \right],
\end{aligned}
$$

donde la última desigualdad se debe a la Proposición 3.5.3. Finalmente, tomando supremo sobre $\varphi \in \Pi_M$ obtenemos (3.33).■

### 3.5.2. Demostración del Teorema 3.5.1(b)

Para facilitar la notación, sea $\hat{a}_t^N := a_t^{\hat{\pi}, N}$ y $\hat{a}_t := a_t^{\hat{\pi}}$. Entonces, consideremos $\left\{ (y^N(t), \hat{a}_t^N) \right\} \in \mathbb{Y}_N \times A$ y $\{ (y(t), \hat{a}_t) \} \in \mathbb{Y} \times A$ las sucesiones de pares estado-acción correspondiente a la aplicación de la política $\hat{\pi}$ (ver Definición 3.4.1). Para cada $t \in \mathbb{N}_0$, definamos la variable aleatoria

$$
\Delta_t^N := \left| \Phi^N(y^N(t), \hat{a}_t^N) - \Phi(y(t), \hat{a}_t) \right|.
$$

Entonces, de la definición de las funciones de discrepancia $\Phi^N$ y $\Phi$ dadas en (3.35) y (3.10), respectivamente, para cada $t \in \mathbb{N}_0$ tenemos

$$\begin{aligned}\Delta_t^N &\le \left|r(y^N(t),\hat{a}_t^N)-r(y(t),\hat{a}_t)\right|+\left|V_*^N(y^N(t))-v_*(y(t))\right|\\ &+\alpha\left|\int_{\mathbb{R}^N}V_*^N\left[H_\rho^N\left(y^N(t),\hat{a}_t^N,w\right)\right]\theta(dw)-v_*(H_\rho\left(y(t),\hat{a}_t\right))\right|\\ &\le \left|r(y^N(t),\hat{a}_t^N)-r(y(t),\hat{a}_t)\right|+\left|V_*^N(y^N(t))-v_*(y(t))\right|\\ &+\left|\int_{\mathbb{R}^N}\left\{V_*^N\left[H_\rho^N\left(y^N(t),\hat{a}_t^N,w\right)\right]-v_*(y(t+1))\right\}\theta(dw)\right| \quad \text{(by (3.5))}\\ &= \left|r(y^N(t),\hat{a}_t^N)-r(y(t),\hat{a}_t)\right|+\left|V_*^N(y^N(t))-v_*(y(t))\right|\\ &+\left|E_y^{\hat{\pi}}\left[V_*^N(y^N(t+1))-v_*(y(t+1))\mid h_t^N,\hat{a}_t^N\right]\right| \quad \text{(by (2.10))}\\ &\le \left|r(y^N(t),\hat{a}_t^N)-r(y(t),\hat{a}_t)\right|+\left|V_*^N(y^N(t))-v_*(y(t))\right|\\ &+E_y^{\hat{\pi}}\left[\left|V_*^N(y^N(t+1))-v_*(y(t+1))\right|\mid h_t^N,\hat{a}_t^N\right]. \end{aligned} \tag{3.47}$$

Tomando esperanza $E_y^{\hat{\pi}}$ en (3.47) y usando propiedades de la esperanza condicional tenemos

$$\begin{aligned}E_y^{\hat{\pi}}\left[\Delta_t^N\right] &\le E_y^{\hat{\pi}}\left|r(y^N(t),\hat{a}_t^N)-r(y(t),\hat{a}_t)\right|+E_y^{\hat{\pi}}\left|V_*^N(y^N(t))-v_*(y(t))\right|\\ &+E_y^{\hat{\pi}}\left|V_*^N(y^N(t+1))-v_*(y(t+1))\right|.\end{aligned}$$

Además, por la Proposición 3.5.3 y el Teorema 3.5.1 obtenemos

$$\begin{aligned}E_y^{\hat{\pi}}\left[\Delta_t^N\right] &\le L_r\left[KTe^{-\lambda N\varepsilon^2}(1+\mathcal{K}(T)+\gamma_T(\varepsilon))\right]+\frac{4R\alpha^T}{1-\alpha}\\ &+2L_r\frac{1-\alpha^T}{1-\alpha}\left[KTe^{-\lambda N\varepsilon^2}(1+\mathcal{K}(T))+\gamma_T(\varepsilon)\right],\end{aligned}$$

para cualquier $\varepsilon>0$ arbitrario y $T>t$.

También, observemos que

$$\begin{aligned}&E_y^{\hat{\pi}}\left[\Phi^N(y^N(t),\hat{a}_t^N)\right]\le E_y^{\hat{\pi}}\left[|\Phi^N(y^N(t),\hat{a}_t^N)-\Phi(y(t),\hat{a}_t)|\right]+E_y^{\hat{\pi}}\left[\Phi(y(t),\hat{a}_t)\right]\\ &=E_y^{\hat{\pi}}[\Delta_t^N]+E_y^{\hat{\pi}}\left[\Phi(y(t),\hat{a}_t)\right]\le L_r\left[KTe^{-\lambda N\varepsilon^2}(1+\mathcal{K}(T))+\gamma_T(\varepsilon)\right]+\frac{4R\alpha^T}{1-\alpha}\\ &+2L_r\frac{1-\alpha^T}{1-\alpha}\left[KTe^{-\lambda N\varepsilon^2}(1+\mathcal{K}(T))+\gamma_T(\varepsilon)\right]+E_y^{\hat{\pi}}\left[\Phi(y(t),\hat{a}_t)\right].\end{aligned}$$

Así, tomando límite cuando $N \to \infty$ obtenemos

$$\begin{aligned} 0 \leq \lim_{N\to\infty} E_y^{\hat{\pi}} \left[\Phi^N(y^N(t), \hat{a}_t^N)\right] &\leq L_r \gamma_T(\varepsilon) + \frac{4R\alpha^T}{1-\alpha} + 2L_r \frac{1-\alpha^T}{1-\alpha}\gamma_T(\varepsilon) \\ &+ E_y^{\hat{\pi}} \left[\Phi(y(t), \hat{f}_t(y(t)))\right]. \end{aligned} \tag{3.48}$$

Finalmente, como $\varepsilon$ y $T$ son arbitrarios, haciendo $t \to \infty$ en (3.48), un simple uso del Teorema 3.4.2 muestra que

$$\lim_{t\to\infty} \lim_{N\to\infty} E_y^{\hat{\pi}} \left[\Phi^N(y^N(t), \hat{a}_t^N)\right] = 0$$

lo cual demuestra el resultado deseado.■

# Capítulo 4

# Sistemas de interacción de objetos modelados como juegos contra la naturaleza bajo un esquema de campo medio

## 4.1. Introducción

En este capítulo estamos interesados en estudiar el problema de control de sistemas de interacción de objetos, descrito en el Capítulo anterior, considerando que el proceso de perturbaciones aleatorias $\{\xi_t^n\}$, que interviene en la dinámica del sistema (2.1) está formado por v.a. independientes pero no necesariamente idénticamente distribuidas ni observables, y con distribución desconocida.

Como podemos observar, la hipótesis de observabilidad así como el hecho de asumir que $\{\xi_t^n\}$ es una familia de v.a.i.i.d., fue determinante en el capítulo anterior para poder implementar métodos de estimación y control con el fin de construir políticas, lo cual claramente no es posible bajo el contexto actual.

Ahora, la hipótesis principal será asumir que la única información que posee el controlador en cada etapa es que la distribución desconocida $\mu_t$, correspondiente a la familia de v.a. $\{\xi_t^n\}_{n=1}^N$, pertenece a un conjunto apropiado $\Gamma$ de medidas de probabilidad. Este hecho nos permite modelar el problema de control asociado al sistema de interacción de objetos como un juego contra la naturaleza. Estos juegos son una clase de sistemas de control minimax los cuales consisten en asumir que el controlador tiene un oponente, la "naturaleza", quien en cada etapa selecciona la distribución $\mu_t$ (desconocida por el controlador), de las v.a. $\{\xi_t^n\}_{n=1}^N$, sobre el conjunto $\Gamma$. De esta manera, el objetivo del controlador es minimizar el máximo costo que genera la naturaleza.

Por el hecho de tratar con un número grande de objetos ($N \sim \infty$) no es posible calcular políticas minimax. Entonces recurrimos a la teoría de campo medio y seguiremos un procedimiento similar al del capítulo anterior. Es decir, primero definimos el modelo minimax $\mathcal{M}_{mxm}^N$ correspondiente al juego contra la naturaleza cuyos estados son las proporciones de objetos en cada clase. Entonces, haciendo $N \to \infty$ obtenemos un nuevo modelo $\mathcal{M}_{mxm}$, que corresponde al juego contra la naturaleza de campo medio. Sobre el modelo $\mathcal{M}_{mxm}$ es posible garantizar la existencia de una política minimax $\pi^*$. Entonces analizamos la desviación minimax de $\pi^*$ cuando ésta se usa para controlar el proceso original definido en $\mathcal{M}_{mxm}$. Por lo tanto, nuestro objetivo es demostrar que $\pi^*$ es una política minimax en un sentido asintótico cuando $N \to \infty$.

## 4.2. Descripción del sistema

En esta ocasión describiremos el sistema de interacción de objetos bajo este nuevo contexto. Para una fácil referencia y claridad en la exposición, parte de la notación y descripción introducidos en los capítulos anteriores los reescribiremos.

Consideremos un sistema de control estocástico a tiempo discreto compuesto por un número grande de objetos que interactúan entre sí, donde $X_n^N(t)$, $n = 1, 2, \ldots, N$, $t \in \mathbb{N}_0$ representa la clase del objeto $n$ al tiempo $t$, tomando valores en un conjunto dado $S = \{1, 2, \ldots, s\} \subseteq \mathbb{N}$. En cada etapa, un controlador central selecciona una acción o control $a_t$ de un conjunto de Borel $A$, la cual influye en el comportamiento de los objetos. Supongamos que los objetos comparten un ambiente en común, y sea $C^N(t) \in \mathbb{R}^d$ el contexto del ambiente al tiempo $t \in \mathbb{N}_0$. El comportamiento así como la evolución de los objetos son consideradas independientes. Esto es, el proceso $\left\{X_n^N(t)\right\}_{t\in\mathbb{N}_0}$ evoluciona de acuerdo a la ecuación en diferencias:

$$X_n^N(t+1) = F(X_n^N(t), C^N(t), a_t, \xi_t^n), \quad t = 0, 1, ..., \tag{4.1}$$

donde $F : S \times \mathbb{R}^d \times A \times \mathbb{R} \to S$ es una función dada (conocida por el controlador) y $\{\xi_t^n : t \geq 1, n = 1, 2, \ldots, N\}$ es una familia de variables aleatorias independientes, posiblemente no observables definidas en un espacio de probabilidad $(\Omega, \mathcal{F}, P')$ tomando valores en $D \subset \mathbb{R}$, con correspondiente distribución $\mu_t^n \in \Gamma \subset \mathbb{P}(D)$. Esto es,

$$\mu_t^n(B) := P'\left[\xi_t^n \in B\right], \quad t \in \mathbb{N}_0,\ n = 1, 2, \ldots, N, \quad B \in \mathcal{B}(D). \tag{4.2}$$

Supondremos que para cada $t \geq 0$ fija,

$$\mu_t^1 = \mu_t^2 = \ldots = \mu_t^N =: \mu_t$$

es decir, $\mu_t$ es la distribución común al tiempo $t$ del ruido que afecta a cada objeto. Además, supongamos que $\mu_t$ es desconocida para el controlador, la cual puede cambiar de etapa a etapa. Esto es, para cada $t \geq 0$ tendrá una distribución $\mu_t$ la cual influirá en la configuración del sistema al tiempo $t$.

Podemos escribir la ley de transición $K$ de cada objeto en términos de la función $F$ como:

$$\begin{aligned} K_{ij}(a,\mu,c) &:= P\left[X_n^N(t+1)=j \,\middle|\, X_n^N(t)=i, a_t=a, \mu_t^n=\mu, C^N(t)=c\right] \\ &= \int_{\mathbb{R}} I_j[F(i,c,a,z)]\mu(dz), \quad i,j\in S,\ (a,c)\in A\times\mathbb{R}^d,\ \mu\in\Gamma. \end{aligned} \tag{4.3}$$

Esta relación define una ley de transición mediante el kernel estocástico $K = K(a,\mu,c) = [K_{ij}(a,\mu,c)]$.

Al igual que en los capítulos anteriores supondremos que los objetos son observables de acuerdo a las clases donde se encuentran localizadas, así que el controlador solo puede determinar el *número* de objetos en cada una de las clases $i \in S$. Así, el comportamiento del sistema puede ser reformulado mediante la *proporción* de objetos en cada clase. Ciertamente, sea $M_i^N(t)$ la proporción de objetos en la clase $i \in S$ al tiempo $t$:

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

y denotemos por $\vec{M}^N(t)$ al vector cuyas componentes son los $M_i^N(t)$'s:

$$\vec{M}^N(t) = \left(M_1^N(t), M_2^N(t), \ldots, M_s^N(t)\right).$$

Notemos que $\vec{M}^N(t)$ pertenece al conjunto $\mathbb{P}_N(S) := \{p\in\mathbb{P}(S) : Np(i)\in\mathbb{N},\ \forall i\in S\} \subset \mathbb{P}(S)$.

Por otro lado, supongamos que el contexto del ambiente $C^N(\cdot)$ y las proporciones $\vec{M}^N(\cdot)$ son sistemas dinámicos definidos por las ecuaciones en diferencias:

$$C^N(t+1) = g(C^N(t), \vec{M}^N(t+1), a_t),\ \ t\in\mathbb{N}_0, \tag{4.4}$$

$$\vec{M}^N(t+1) = G^N(\vec{M}^N(t), C^N(t), a_t, \mu_t, \vec{w}_t), \tag{4.5}$$

donde $g : \mathbb{R}^d \times \mathbb{P}(S) \times A \to \mathbb{R}^d$ y $G^N : \mathbb{P}_N(S) \times \mathbb{R}^d \times A \times \Gamma \times \mathbb{R}^N \to \mathbb{P}_N(S)$ son funciones conocidas y $\{\vec{w}_t\}$ es una sucesión de v.a.i.i.d. en $\mathbb{R}^N$ con distribución común $\theta$.

Claramente, el comportamiento del sistema de interacción de partículas está determinado por la evolución del proceso $\left\{(\vec{M}^N(t), C^N(t))\right\}$, cuya dinámica puede ser expresada como sigue. Recordando que $\mathbb{Y}_N := \mathbb{P}_N(S) \times \mathbb{R}^d$, sea $H^N : \mathbb{Y}_N \times A \times \Gamma \times \mathbb{R}^N \to \mathbb{Y}_N$ la función definida como

$$H^N(y, a, \mu, w) := (G^N(y, a, \mu, w), g(c, G^N(y, a, \mu, w), a)). \tag{4.6}$$

Denotando a $y^N(t) := (\vec{M}^N(t), C^N(t))$, de acuerdo a (4.4) y (4.5), $H^N$ define la dinámica del proceso $\left\{y^N(t)\right\}$ :

$$\begin{aligned} y^N(t+1) &= (G^N(y^N(t), a_t, \mu_t, \vec{w}_t), g(C^N(t), \vec{M}^N(t+1), a_t)) \\ &= (G^N(y^N(t), a_t, \mu_t, \vec{w}_t), g(C^N(t), G^N(y^N(t), a_t, \mu_t, \vec{w}_t), a_t)) \\ &= H^N\left(y^N(t), a_t, \mu_t, \vec{w}_t\right). \end{aligned} \tag{4.7}$$

Asumiremos que la función de costo por etapa depende de la configuración del sistema dado por $y^N(t)$ y de la acción seleccionada por el controlador. Entonces, el costo será representado por una función medible $r : \mathbb{Y} \times A \to \mathbb{R}$, donde $\mathbb{Y} := \mathbb{P}(S) \times \mathbb{R}^d$. Recuerde que $\mathbb{Y}_N := \mathbb{P}_N(S) \times \mathbb{R}^d \subseteq \mathbb{Y}$.

Por lo tanto el problema de control óptimo para el controlador es encontrar una política dirigida a minimizar el criterio de costo descontado. Sin embargo, es importante enfatizar que en cada etapa la única información que se tiene acerca de la distribución $\mu_t$ es que pertenece al conjunto $\Gamma$, por lo cual las técnicas tradicionales no son aplicables. De esta manera el problema se modelará como un juego contra la naturaleza, mismo que describiremos en la siguiente sección.

## 4.3. Formulación del problema como un juego contra la naturaleza

En términos generales, un juego contra la naturaleza consiste en suponer que el controlador tiene un oponente quien selecciona la distribución $\mu_t$ en cada etapa, de tal manera que el objetivo es minimizar el máximo costo incurrido. Los elementos que intervienen para formular el sistema de interacción de objetos como un juego contra la naturaleza lo podemos agrupar en el siguiente modelo de control minimax (abreviado $N$-Minimax):

$$\mathcal{M}_{mxm}^{N} := \left(\mathbb{Y}_N, A, \Gamma, H^N, \theta, r\right). \tag{4.8}$$

El modelo $\mathcal{M}_{mxm}^{N}$ tiene la siguiente interpretación. Al tiempo $t$, el controlador observa la configuración del sistema por medio del estado $y = y^N(t) = (\vec{M}^N(t), C^N(t)) \in \mathbb{Y}_N$, el cual está comupesto por las proporciones de los objetos y por el contexto del ambiente. Entonces el controlador selecciona una acción $a = a_t \in A$ mientras que el oponente (la naturaleza) elige una distribución $\mu = \mu_t \in \Gamma$ de la variable aleatoria $\xi_t$. Como consecuencia ocurre lo siguiente: (1) un costo $r(y, a)$ es incurrido, y (2) el sistema se mueve a un nuevo estado $y' = y^N(t+1) = (\vec{M}^N(t+1), C^N(t+1))$ de acuerdo a la ley de transición

$$\begin{aligned} Q(B|y, a, \mu) &:= P\left[y^N(t+1) \in B | y^N(t) = y, a_t = a, \mu_t = \mu\right] \\ &= \int_{\mathbb{R}^N} I_B\left[H^N(y, a, \mu, w)\right] \theta(dw), \end{aligned}$$

con $H^N$ como en (4.6). Una vez que la transición al estado $y'$ ocurre, el proceso se repite.

Observemos que los cambios de la confguración del sistema, gobernado por la ley de transición $Q$, depende fuertemente de las distribuciones $\{\mu_t\}$ seleccionadas por el oponente. Por lo tanto las acciones seleccionadas por el controlador estan dirigidas a minimizar el máximo *costo total esperado descontado* introducido mas adelante.

A fin de asegurar la existencia de minimizadores, estableceremos las siguientes condiciones de continuidad y compacidad.

*Hipótesis* 4. **(a)** El espacio de control $A$ es un espacio de Borel compacto, cuya métrica es denotada por $d_A$.

**(b)** La función $g$ en (4.4) es una función de Lipschitz con constante $L_g$; esto es, para $c, c' \in \mathbb{R}^d$, $\vec{m}, \vec{m}' \in \mathbb{P}(S)$, $a, a' \in A$,

$$\|g(c, \vec{m}, a) - g(c', \vec{m}', a')\|_\infty^2 \leq L_g \operatorname{máx} \left\{ \|c - c'\|_\infty^2, \|\vec{m} - \vec{m}'\|_\infty^1, d_A(a, a') \right\}. \tag{4.9}$$

Sin pérdida de generalidad, podemos suponer que $L_g \geq 1$.

**(c)** El conjunto $\Gamma \subset \mathbb{P}(D)$ (ver (4.2)) es compacto.

**(d)** La función $a \longmapsto H^N(y, a, \mu, w)$ definida en (4.6), es continua, para toda $y \in \mathbb{Y}_N$, $\mu \in \Gamma$ y $w \in \mathbb{R}^N$.

**(e)** La función de costo por etapa $r$ es una función acotada y uniformemente de Lipschitz con constante $L_r$; esto es, existe una constante $R > 0$ tal que

$$|r(y, a)| \leq R \quad \forall (y, a) \in \mathbb{Y} \times A,$$

y para cada $a, a' \in A$ y $y, y' \in \mathbb{Y}$,

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

### 4.3.1. Criterio de optimalidad descontado

Sea $\mathbb{H}_0^N := \mathbb{Y}_N$ y $\mathbb{H}_t^N := \left(\mathbb{Y}_N \times A \times \Gamma \times \mathbb{R}^N\right)^t \times \mathbb{Y}_N$, $t \geq 1$, el espacio de las historias hasta el tiempo $t$. Un elemento $h_t^N$ de $\mathbb{H}_t^N$ es escrito como

$$h_t^N = (y^N(0), a_0, \mu_0, \vec{w}_0, \ldots, y^N(t-1), a_{t-1}, \mu_{t-1}, \vec{w}_{t-1}, y^N(t)),$$

donde $y^N(t) = \left(\vec{M}^N(t), C^N(t)\right)$. Una política de control para el controlador es una sucesión $\pi^N = \left\{\pi_t^N\right\}$ de kerneles estocásticos $\pi_t^N$ en $A$ dado $\mathbb{H}_t^N$ tales que $\pi_t^N\left(A|h_t^N\right) = 1$ para toda $h_t^N \in \mathbb{H}_t^N$, $t \in \mathbb{N}_0$. Denotaremos por $\Pi^N$ al conjunto de todas las políticas para el controlador.

De igual forma como se hizo en el Capítulo 2, denotaremos por $\mathbb{F}$ al conjunto de todas las funciones medibles $f : \mathbb{Y} \to A$ y $\mathbb{F}^N := \mathbb{F}|_{\mathbb{Y_N}}$ la restricción de $\mathbb{F}$ en el conjunto $\mathbb{Y}_N$. Entonces, las políticas markovianas y estacionarias se definen de manera similar. Denotemos por $\Pi_M^N$ y $\mathbb{F}^N$ al conjunto de las políticas de Markov y al conjunto de las políticas estacionarias para el controlador, respectivamente.

Además denotemos por $\Pi_M$ al conjunto de las políticas de Markov deterministas para el controlador cuando se usa $\mathbb{F}$ en lugar de $\mathbb{F}^N$; esto es, $\Pi_M$ es la familia de sucesiones de funciones $\{f_t\} \subset \mathbb{F}$. Por lo tanto cualquier política $\pi = \{f_t\} \in \Pi_M$ cuyos elementos $f_t$ son restringidos a $\mathbb{Y}_N$ se convierte en un elemento de $\Pi^N$.

Por otro lado, supondremos que las acciones del oponente dependen solo de las decisiones tomadas a través del tiempo. Esto es, una política de control para el oponente es una sucesión $\tilde{\pi} = \{\mu_t\}$, con $\mu_t \in \Gamma$ y $t \in \mathbb{N}_0$. Denotaremos por $\Pi_\Gamma$ al conjunto de las políticas para el oponente.

Para cada par de políticas $\left(\pi^N, \widetilde{\pi}\right) \in \Pi^N \times \Pi_\Gamma$ y estado inicial $y^N(0) = y \in \mathbb{Y}_N$, definimos el costo esperado total descontado como

$$V^N(\pi^N,\widetilde{\pi},y^N) := E_y^{\pi^N,\widetilde{\pi}} \sum_{t=0}^{\infty} \alpha^t r(y^N(t),a_t), \tag{4.10}$$

donde $E_y^{\pi^N,\widetilde{\pi}}$ denota el operador esperanza con respecto a la medida de probabilidad $P_y^{\pi^N,\widetilde{\pi}}$ inducida por $(\pi^N,\widetilde{\pi}) \in \Pi^N \times \Pi_\Gamma$ dado $y^N(0)=y$. Por lo tanto, el problema de control minimax para el controlador es encontrar una política $\pi_*^N \in \Pi^N$ tal que

$$V_*^N(y^N) := \inf_{\pi^N\in\Pi^N} \sup_{\widetilde{\pi}\in\Pi_\Gamma} V^N(\pi^N,\widetilde{\pi},y^N) = \sup_{\widetilde{\pi}\in\Pi_\Gamma} V^N(\pi_*^N,\widetilde{\pi},y^N), \quad y^N \in \mathbb{Y}_N. \tag{4.11}$$

En este caso, $V_*^N$ se dice ser la $N-$*función de valor minimax*.

Bajo las condiciones impuestas en el modelo $N$-Minimax $\mathcal{M}_{mxm}^N$, podemos establecer los siguientes resultados, los cuales nos dan una caracterización de la política minimax y de la $N$-función de valor minimax en términos de la solución de una ecuación funcional llamada la $N$-ecuación minimax (ver por ejemplo, [16, 35]):

Para $u \in \mathbb{B}(\mathbb{Y})$ y $(y,a,\mu) \in \mathbb{Y}\times A\times\Gamma$, definimos

$$T_{(a,\mu)}u(y) := r(y,a) + \alpha \int_{\mathbb{R}^N} u\left[H^N(y,a,\mu,w)\right]\theta(dw), \tag{4.12}$$

$$\widehat{T}_a u(y) := \sup_{\mu\in\Gamma} T_{(a,\mu)}u(y), \tag{4.13}$$

y

$$Tu(y) := \min_{a\in A}\sup_{\mu\in\Gamma} T_{(a,\mu)}u(y) = \min_{a\in A} \widehat{T}_a u(y), \tag{4.14}$$

Ahora establecemos el siguiente resultado

**Proposición 4.3.1.** Si la Hipótesis 4 se cumple, entonces:

**(a)** El operador $T$ es de contracción con módulo $\alpha$.

**(b)** $T$ es un operador de $\mathbb{C}_b(\mathbb{Y}_N)$ en sí mismo.

**(c)** Para cada $u \in \mathbb{C}_b(\mathbb{Y}_N)$ existe $f_*^N \in \mathbb{F}$ tal que

$$Tu\left(y^N\right) = \sup_{\mu \in \Gamma} T_{(f_*^N,\mu)} u(y^N),\ \ y^N \in \mathbb{Y}_N. \tag{4.15}$$

*Demostración.* (a) Sean $u, u' \in \mathbb{B}(\mathbb{Y})$ y $(y, a, \mu) \in \mathbb{Y} \times A \times \Gamma$. Entonces

$$\begin{aligned}
|Tu(y) - Tu'(y)| &= \left| \min_{a \in A} \sup_{\mu \in \Gamma} T_{(a,\mu)} u(y) - \min_{a \in A} \sup_{\mu \in \Gamma} T_{(a,\mu)} u'(y) \right| \\
&\leq \sup_{a \in A} \left| \sup_{\mu \in \Gamma} T_{(a,\mu)} u(y) - \sup_{\mu \in \Gamma} T_{(a,\mu)} u'(y) \right| \\
&\leq \sup_{a \in A} \ \sup_{\mu \in \Gamma} \left| T_{(a,\mu)} u(y) - T_{(a,\mu)} u'(y) \right| \\
&= |r(y,a) + \alpha \int_{\mathbb{R}^N} u\left[H^N(y,a,\mu,w)\right] \theta(dw) \\
&\quad - r(y,a) - \alpha \int_{\mathbb{R}^N} u'\left[H^N(y,a,\mu,w)\right] \theta(dw)| \\
&\leq \alpha \int_{\mathbb{R}^N} |u\left[H^N(y,a,\mu,w)\right] - u'\left[H^N(y,a,\mu,w)\right]| \theta(dw) \\
&\leq \alpha \int_{\mathbb{R}^N} \sup_{y \in \mathbb{Y}_N} |u(y) - u'(y)| \, \theta(dw) \\
&\leq \alpha \left\| u - u' \right\|.
\end{aligned} \tag{4.16}$$

Luego, tomando supremo sobre $y \in \mathbb{Y}_N$ obtenemos

$$\left\| Tu - Tu' \right\| \leq \alpha \left\| u - u' \right\|.$$

(b) Para probar esta parte, primero mostraremos que el operador $T$ es acotado.

Para $u \in \mathbb{C}_b(\mathbb{Y}_N)$, $y \in \mathbb{Y}_N$ y $a \in A$ tenemos

$$\begin{aligned} |\widehat{T}_a u\,(y)\,| &= |\sup_{\mu\in\Gamma} T_{(a,\mu)} u\,(y)\,| \le \sup_{\mu\in\Gamma} |T_{(a,\mu)} u\,(y)\,| \\ &\le \sup_{\mu\in\Gamma} |r\,(y,a) + \alpha \int_{\mathbb{R}^N} u\left[H^N\,(y,a,\mu,w)\right]\theta\,(dw)\,| \\ &\le \sup_{\mu\in\Gamma} |r\,(y,a)\,| + \alpha \sup_{\mu\in\Gamma} \int_{\mathbb{R}^N} |u\left[H^N\,(y,a,\mu,w)\right]|\theta\,(dw) \\ &\le M + \alpha\,\|u\|\,. \end{aligned} \tag{4.17}$$

Ahora tomando supremo sobre todo $\mu \in \Gamma$ llegamos a que

$$|Tu\,(y)\,| \le \widehat{M} \quad \forall y \in \mathbb{Y}_N,$$

donde $\hat{M} := M + \alpha\,\|u\|$ lo cual implica $\|Tu\| < \infty$.

Ahora mostraremos la continuidad de $\widehat{T}_a$ en $\mathbb{Y}_N$. Observemos primero que, para $u \in \mathbb{C}_b(\mathbb{Y}_N)$, el operador $T_{(a,\mu)} u\left(y^N\right)$ es continuo en $\mathbb{Y}_N \times A$. Entonces, usando el hecho de que $\Gamma$ es un conjunto compacto, por el Teorema de Berge (ver [16], [35]) obtenemos que $\sup_{\mu\in\Gamma} T_{(a,\mu)} u\left(y^N\right) =: \widehat{T}_a u\left(y^N\right)$ es continuo en $a \in A$. Esto implica, por la Hipótesis 4(a), que el operador $T$ es continuo.

(c) De la continuidad del operador $T$, la Hipótesis 4, y aplicando un teorema de selección medible (ver [20]) podemos garantizar la existencia de minimizadores $f_*^N \in \mathbb{F}$ tales que

$$Tu\left(y^N\right) = \min_{a\in A}\left\{\widehat{T}_a u\left(y^N\right)\right\} = \widehat{T}_{f_*^N} u\left(y^N\right), \tag{4.18}$$

obteniendo así la parte (c).

□

*Observación* 4.3.1. *Se puede demostrar que $\mathbb{C}_b(\mathbb{Y}_N)$ es un subconjunto cerrado del espacio de Banach $\mathbb{B}(\mathbb{Y}_N)$, por lo que $\mathbb{C}_b(\mathbb{Y}_N)$ subespacio completo de $\mathbb{B}(\mathbb{Y}_N)$. Ahora, dado que $T$ es un operador de contracción de $\mathbb{C}_b(\mathbb{Y}_N)$ en sí mismo y $\mathbb{C}_b(\mathbb{Y}_N) \subset \mathbb{B}(\mathbb{Y}_N)$ es completo, por el Teorema de Punto fijo de Banach, existe una única función $\widetilde{u} \in \mathbb{C}_b(\mathbb{Y}_N)$ tal que para todo $y^N \in \mathbb{Y}_N$,*

$$\widetilde{u}\left(y^N\right) = T\widetilde{u}\left(y^N\right),$$

*y*

$$\|T^n u - \widetilde{u}\| \leq \alpha^n \|u - \widetilde{u}\| \;\; \forall u \in \mathbb{C}_b(\mathbb{Y}_N), \;\; n \in \mathbb{N}_0. \tag{4.19}$$

Definimos la sucesión $\{v_n\} \subset \mathbb{C}_b(\mathbb{Y}_N)$ como $v_0 = 0$, y para $n \geq 1$,

$$v_n\left(y^N\right) = Tv_{n-1}\left(y^N\right), \;\; y^N \in \mathbb{Y}_N.$$

Ahora establecemos el siguiente resultado que nos asegura la existencia de una política minimax para el modelo $\mathcal{M}^N_{mxm}$.

**Teorema 4.3.2.** Supongamos que la Hipótesis 4 se satisface. Entonces:

**(a)** La función de valor minimax (4.11) es la única solución en $\mathbb{C}_b(\mathbb{Y}_N)$ que satisface

$$V^N_*\left(y^N\right) = TV^N_*\left(y^N\right), \;\; y^N \in \mathbb{Y}_N.$$

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

$$\left\|v_n - V^N_*\right\| \leq M\frac{\alpha^n}{1-\alpha}. \tag{4.20}$$

**(c)** Existe $f^N_* \in \mathbb{F}$ tal que para toda $y^N \in \mathbb{Y}_N$,

$$V^N_*\left(y^N\right) = \sup_{\mu \in \Gamma}\left\{r\left(y^N, f^N_*\right) + \alpha\int_{\mathbb{R}^{\mathbb{N}}} V^N_*\left[H^N\left(y^N, f^N_*, \mu, w\right)\right]\theta(dw)\right\}. \tag{4.21}$$

Equivalentemente

$$\Phi^N\left(y^N, f_*^N\right) = 0, \tag{4.22}$$

donde

$$\Phi^N\left(y^N, a\right) := \sup_{\mu\in\Gamma}\left\{r(y^N, a) + \alpha\int_{\mathbb{R}^N} V_*^N\left[H^N(y^N, a, \mu, w)\right]\theta\left(dw\right)\right\} - V_*^N\left(y^N\right), \tag{4.23}$$

es la función de discrepancia en el $N-$MCMinimax $\mathcal{M}_{mxm}^N$. Más aún $\left\{f_*^N\right\}$ es una política minimax para el controlador, esto es,

$$V_*^N\left(y^N\right) = \sup_{\widetilde{\pi}\in\Pi_\Gamma} V^N\left(y^N, f_*^N, \widetilde{\pi}\right).$$

*Demostración.* (a)-(b) Primero observemos que $v_n = T^n v_0$ para todo $n \in \mathbb{N}_0$. Así, tomando $u = v_0$ en la desigualdad (4.19) tenemos

$$\left\|v_n - \widetilde{u}\right\| \le \alpha^n \left\|\widetilde{u}\right\| \;\; \forall n \in \mathbb{N}_0. \tag{4.24}$$

Por lo tanto, si mostramos que $\widetilde{u} = V_*^N$ las partes (a) y (b) serán demostradas. Sea $f^N \in \mathbb{F}$ un selector tal que

$$\widetilde{u}\left(y^N\right) = \sup_{\mu\in\Gamma} T_{(f^N,\mu)} u\left(y^N\right) \quad y^N \in \mathbb{Y}_N. \tag{4.25}$$

Entonces,

$$\widetilde{u}\left(y^N\right) \ge r\left(y^N, f^N\right) + \alpha\int_{\mathbb{R}^N} \widetilde{u}\left[H^N\left(y^N, f^N, \mu, w\right)\right]\theta(dw) \; \forall y^N \in \mathbb{Y}_N,\; \mu\in\Gamma. \tag{4.26}$$

Ahora, sea $\widetilde{\pi} = \{\mu_t\} \in \Pi_\Gamma$ una política arbitraria para el oponente y $\left\{\left(y_n^N, f^N\left(y_n^N\right), \mu_n\right)\right\}$ una sucesión de ternas estado-acciones correspondiente a la aplicación de las estrategias $f^N$ y $\widetilde{\pi}$. Entonces, iterando la desigualdad (4.26) obtenemos

$$\widetilde{u}\left(y^N\right) \ge E_{y^N}^{f^N,\widetilde{\pi}}\left[\sum_{t=0}^{n-1}\alpha^t r(y^N(t), f^N(y^N t(t)))\right] + \alpha^n E_{y^N}^{f^N,\widetilde{\pi}}\widetilde{u}\left(y^N\left(n\right)\right), \tag{4.27}$$

lo cual, haciendo $n \to \infty$, implica

$$\widetilde{u}\left(y^N\right) \geq V^N\left(y^N, f^N, \widetilde{\pi}\right). \tag{4.28}$$

Como $\tilde{\pi} \in \Pi_\Gamma$ es arbitraria, de (4.11) y (4.27) tenemos

$$\tilde{u}(y^N) \geq V_*^N(y^N),\ \ y^N \in \mathbb{Y}_N \tag{4.29}$$

Por otro lado, para una política arbitraria $\pi^N \in \Pi^N$, $\widetilde{\pi} \in \Pi_\Gamma$ y estado $y^N \in \mathbb{Y}_N$ tenemos

$$\begin{aligned}
E_{y^N}^{\pi,\widetilde{\pi}}\left[\alpha^{t+1}\widetilde{u}\left(y^N(t+1)\right)|h_t,a_t\right] &= \alpha^{t+1}\int_{\mathbb{R}^N}\widetilde{u}\left[H^N\left(y^N(t),a_t,\mu_t,w\right)\right]\theta(dw)\\
&= \alpha^t[r\left(y^N(t),a_t\right)+\alpha\int_{\mathbb{R}^N}\widetilde{u}\left[H^N\left(y^N(t),a_t,\mu_t,w\right)\right]\theta(dw)\\
&-r\left(y^N(t),a_t\right)]\\
&\geq \alpha^t\left[\widetilde{u}\left(y^N(t)\right)-r\left(y^N(t),a_t\right)\right]
\end{aligned} \tag{4.30}$$

por lo tanto

$$\alpha^t r\left(y^N(t),a_t\right) \geq -E_{y^N}^{\pi,\widetilde{\pi}}\left[\alpha^{t+1}\widetilde{u}\left(y^N(t+1)\right)-\alpha^t\widetilde{u}\left(y^N(t)\right)|h_t,a_t\right].$$

Entonces, tomando esperanza $E_{y^N}^{\pi,\widetilde{\pi}}$ y sumando sobre $t = 0, \ldots, n-1$, tenemos

$$E_{y^N}^{\pi,\widetilde{\pi}}\sum_{t=0}^{n-1}\alpha^t r\left(y^N(t),a_t\right) \geq \widetilde{u}\left(y^N\right)-\alpha^n E_{y^N}^{\pi,\widetilde{\pi}}\widetilde{u}\left(y^N(t)\right)\ \ \forall n \in \mathbb{N}.$$

Finalmente, haciendo $n \to \infty$, se sigue que

$$V^N\left(\pi^N,\widetilde{\pi},y^N\right) \geq \widetilde{u}\left(y^N\right)\ \ y^N \in \mathbb{Y}_N, \tag{4.31}$$

lo cual implica que

$$V_*^N\left(y^N\right) \geq \widetilde{u}\left(y^N\right)\ \ y^N \in \mathbb{Y}_N. \tag{4.32}$$

Combinando (4.29) y (4.32) demostramos que $\tilde{u} = V_*^N$.

(c) La existencia de $f^N \in \mathbb{F}$ se sigue de la parte (a) y de la Proposición 4.3.1(c). Además, aplicando argumentos similares a la demostración de (4.28), tenemos que para cualquier política arbitraria $\tilde{\pi} \in \Pi_\Gamma$

$$V_*^N(y^N) \geq V^N\left(y^N, , f^N, \tilde{\pi}\right) \;\; \forall y^N \in \mathbb{Y}_N.$$

lo cual implica

$$V_*^N\left(y^N\right) = \sup_{\widehat{\pi} \in \Pi_\Gamma} V^N\left(f^N, \widehat{\pi}, y^N\right) \;\; \forall y^N \in \mathbb{Y}_N,$$

esto es, $f^N$ es una estrategia minimax. □

Siguiendo un procedimiento similar al del capítulo anterior, en la siguiente sección introduciremos el modelo correspondiente al juego contra la naturaleza en campo medio con el fin de evitar la dependencia en $N$. Este modelo se obtendrá como límite del modelo $\mathcal{M}_{mxm}^N$ cuando $N \to \infty$.

## 4.4. Juegos contra la naturaleza de campo medio

Consideremos el proceso de control determinista $\{(\vec{m}(t), c(t))\} \in \mathbb{Y} = \mathbb{P}(S) \times \mathbb{R}^d$ definido por el sistema de ecuaciones

$$\vec{m}(t+1) = G(\vec{m}(t), c(t), a_t, \mu_t); \tag{4.33}$$

$$c(t+1) = g(c(t), \vec{m}(t+1), a_t), \tag{4.34}$$

donde $(\vec{m}(0), c(0)) = (\vec{m}, c) \in \mathbb{Y}$ representa la condición inicial, $a_t \in A$ es el control seleccionado al tiempo $t$, $\mu_t$ es la distribución del ruido aleatorio $\xi_t$, al tiempo $t$, $g : \mathbb{R}^d \times \mathbb{P}(S) \times A \to \mathbb{R}^d$ es la función definida en (4.4) y $G : \mathbb{P}(S) \times \mathbb{R}^d \times A \times \Gamma \to \mathbb{P}(S)$

es una función de Lipschitz con constante $L_G$; esto es, para $\vec{m}, \vec{m}' \in \mathbb{P}(S)$, $c, c' \in \mathbb{R}^d$, y $a, a' \in A$,

$$\|G(\vec{m}, c, a, \mu) - G(\vec{m}', c', a', \mu)\|_\infty^1 \leq L_G \max \left\{ \|\vec{m} - \vec{m}'\|_\infty^1, \|c - c'\|_\infty^2, d_A(a, a') \right\}. \tag{4.35}$$

Con el fin de definir el juego contra la naturaleza de campo medio, consideremos la función $H : \mathbb{Y} \times A \times \Gamma \to \mathbb{Y}$ que determina la dinámica del proceso $\{(\vec{m}(t), c(t))\}$; esto es,

$$H(y, a, \mu) := (G(\vec{m}, c, a, \mu), g(c, G(\vec{m}, c, a, \mu), a)), \quad y = (\vec{m}, c) \in \mathbb{Y},\ a \in A,\ \mu \in \Gamma. \tag{4.36}$$

Entonces, de (4.33) y (4.34),

$$\begin{aligned} y(t+1) &= (G(\vec{m}(t), c(t), a_t, \mu_t), g(c(t), \vec{m}(t+1), a_t)) \\ &=: H(y(t), a_t, \mu_t), \quad t \geq 0, \end{aligned} \tag{4.37}$$

con $y(0) = (\vec{m}, c) \in \mathbb{P}(S) \times \mathbb{R}^d = \mathbb{Y}$.

Las siguientes hipótesis se refieren a la convergencia de las trayectorias $y^N(\cdot)$ a las trayectorias $y(\cdot)$ definidas en (4.7) y (4.37), respectivamente.

*Hipótesis* 5. **(a)** $(\vec{M}^N(0), C^N(0)) = (\vec{m}(0), c(0)) = (\vec{m}_0, c_0) = y \in \mathbb{Y}_N$, para todo $N \in \mathbb{N}$.

**(b)** Para cualquier $y \in \mathbb{Y}_N$, $T \in \mathbb{N}$ y $\varepsilon > 0$,

$$\sup_{(\pi,\widetilde{\pi}) \in \Pi_M \times \Pi_\Gamma} P_y^{\pi,\widetilde{\pi}} \left\{ \sup_{0 \leq t \leq T} \left\| y^N(t) - y(t) \right\|_\infty > \gamma_T(\varepsilon) \right\} \leq KTe^{-\lambda N \varepsilon^2}, \tag{4.38}$$

donde $\gamma_T(\varepsilon) \to 0$ cuando $\varepsilon \to 0$ y $T \to \infty$.

Aplicando argumentos similares a la demostración de la Proposición 2.4.2 (b), es fácil mostrar que bajo la Hipótesis 5, para cada $y \in \mathbb{Y}_N$, $(\pi, \widetilde{\pi}) \in \Pi_M \times \Pi_\Gamma$, $T \in \mathbb{N}$ y $\varepsilon > 0$,

$$E_y^{\pi,\widetilde{\pi}} \left[ \sup_{0 \leq t \leq T} \| y^N(t) - y(t) \|_\infty \right] \leq KTe^{-\lambda N \varepsilon^2}(1 + \mathcal{K}(T)) + \gamma_T(\varepsilon), \tag{4.39}$$

donde

$$\mathcal{K}(T) := (L_g)^T \max\{L_g, diam(A)\},$$

$L_g \geq 1$ es la constante de Lipschitz en Hipótesis 4(b) y $diam(A) := \sup_{(a,a') \in A \times A} d(a, a')$.

Usando la misma función de costo por etapa $r$ definida en (4.8) y tomando en cuenta la Hipótesis 5, podemos definir el modelo de control minimax asociado al juego contra la naturaleza de campo medio como

$$\mathcal{M}_{mxm} = (\mathbb{Y}, A, H, r), \tag{4.40}$$

el cual es interpretado como el límite, cuando $N \to \infty$, del modelo $N-$MCMinimax $\mathcal{M}_{mxm}^N$.

Considerando la naturaleza determinista del proceso (4.33)-(4.34), podemos considerar al conjunto $\Pi_M$ como el conjunto de las políticas para el controlador en el modelo $\mathcal{M}_{mxm}$. Entonces, para cada $(\pi, \widetilde{\pi}) \in \Pi_M \times \Pi_\Gamma$ y condición inicial $y(0) = y \in \mathbb{Y}$, definimos el costo total descontado para el modelo de campo medio como

$$v(\pi, \widetilde{\pi}, y) = \sum_{t=0}^{\infty} \alpha^t r\left(y(t), a_t\right). \tag{4.41}$$

Así, el problema de control minimax es encontrar una política $\pi_* \in \Pi_M$ tal que

$$v_*(y) := \inf_{\pi \in \Pi_M} \sup_{\widetilde{\pi} \in \Pi_\Gamma} v(\pi, \widetilde{\pi}, y) = \sup_{\widetilde{\pi} \in \Pi_\Gamma} v(\pi_*, \widetilde{\pi}, y), \quad y \in \mathbb{Y}, \tag{4.42}$$

donde $v_*$ es la *función de valor minimax de campo medio* y $\pi_*$ se dice ser una política minimax para el modelo de campo medio $\mathcal{M}_{mxm}$.

De acuerdo a la descripción del problema, podemos establecer el siguiente teorema minimax de campo medio cuya demostración es similar a la de la Proposición 3.3.1.

**Teorema 4.4.1.** Si la Hipótesis 4 se cumple, entonces:

**(a)** La función de valor minimax (4.42) es la única solución en $\mathbb{C}_b(\mathbb{Y})$ que satisface

$$v_*(y) = Tv_*(y), \ y \in \mathbb{Y},$$

donde, para $u \in \mathbb{C}_b(\mathbb{Y})$ y $y \in \mathbb{Y}$,

$$Tu(y) := \min_{a \in A} \sup_{\mu \in \Gamma} \{r(y,a) + \alpha u [H(y,a,\mu)]\}.$$

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

$$\|v_n - v_*\| \leq M \frac{\alpha^n}{1-\alpha}.$$

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

$$v_*(y) = \sup_{\mu \in \Gamma} \{r(y, f^*) + \alpha v_* [H(y, f^*, \mu)]\},$$

y más aún $\pi^* = \{f^*\} \in \Pi_M$ es una política minimax para el controlador, esto es,

$$v_*(y) = \sup_{\widetilde{\pi} \in \Pi_\Gamma} v(y, f^*, \widetilde{\pi}).$$

Observe que la existencia de la política minimax $\pi^*$ no depende de $N$, por lo que es calculable. Así que a continuación vamos establecer el resultado principal de este capítulo.

**Teorema 4.4.2.** Bajo las Hipótesis 4 y 5 la política de control minimax $\pi^* = \{f^*\}$ para el modelo de control de campo medio $\mathcal{M}$ es asintóticamente minimax para el modelo minimax $\mathcal{M}^N$, cuando $N \to \infty$. Esto es

$$\lim_{N\to\infty} \Phi^N(y, f^*) = 0, \quad y \in \mathbb{Y}_N, \tag{4.43}$$

donde $\Phi^N$ es la función de discrepancia definida en (4.23).

El Teorema 4.4.2 establece que si usamos la política $\pi^*$ para controlar el proceso original (4.7), correspondiente al modelo $\mathcal{M}^N$, la desviación minimax se desvanece a cero, cuando $N \to \infty$. Esto es, para $N$ suficientemente grande, $\pi^*$ es una política minimax para el modelo $N-$MCMinimax $\mathcal{M}^N_{mxm}$.

### 4.4.1. Demostración del Teorema 4.4.2

La demostración del Teorema 4.4.2 está basada en los siguientes resultados que son una adaptación al caso minimax de los resultados usados en en Capítulo 2. Por lo tanto, solo incluiremos la parte esencial de sus demostraciones.

Para cada $(\pi, \widetilde{\pi}) \in \Pi_M \times \Pi_\Gamma$, $y \in \mathbb{Y}_N \subset \mathbb{Y}$ y $T \in \mathbb{N}$, definimos el costo descontado con horizonte finito $T$ como:

$$V_T^N(\pi, \widetilde{\pi}, y) := E_y^{\pi,\widetilde{\pi}}\left[\sum_{k=0}^{T-1} \alpha^k r(y^N(k), a_k)\right], \qquad v_T(\pi, \widetilde{\pi}, y) := \sum_{k=0}^{T-1} \alpha^k r(y(k), a_k).$$

Usaremos la siguiente notación. Para una política fija $\pi = \{f_t\} \in \Pi_M$, denotamos

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

a las acciones al tiempo $t$ correspondientes a la aplicación de la política $\pi$ bajo el proceso $\left\{y^N(t)\right\}$ y $\{y(t)\}$, respectivamente.

**Lema 4.4.3.** Sean $L_r$ y $R$ las constantes en la Hipótesis 4(e). Entonces, bajo las Hipótesis 4 y 5, para cada $y \in \mathbb{Y}_N$, $\varepsilon > 0$, $T \in \mathbb{N}$, $0 \leq t \leq T$, and $(\varphi, \tilde{\varphi}) \in \Pi_M \times \Pi_\Gamma$, los siguientes enunciados se cumplen:

**(a)**

$$\begin{aligned} E_y^{\varphi,\widetilde{\varphi}} & \left[ \sup_{(\pi,\widetilde{\pi})\in\Pi_M\times\Pi_\Gamma} \left| V_T^N(\pi, \widetilde{\pi}, y^N(t)) - v_T(\pi, \widetilde{\pi}, y(t)) \right| \right] \\ & \leq L_r \frac{1-\alpha^T}{1-\alpha} \left[ KTe^{-\lambda N \varepsilon^2}(1 + \mathcal{K}(T)) + \gamma_T(\varepsilon) \right]; \end{aligned} \tag{4.44}$$

**(b)**

$$E_y^{\varphi,\widetilde{\varphi}} \left[ \sup_{(\pi,\widetilde{\pi})\in\Pi_M\times\Pi_\Gamma} \left| V^N(\pi, \widetilde{\pi}, y^N(t)) - V_T^N(\pi, \widetilde{\pi}, y^N(t)) \right| \right] \leq \frac{R\alpha^T}{1-\alpha}; \tag{4.45}$$

**(c)**

$$E_y^{\varphi,\widetilde{\varphi}} \left[ \sup_{(\pi,\widetilde{\pi})\in\Pi_M\times\Pi_\Gamma} |v(\pi, \widetilde{\pi}, y(t)) - v_T(\pi, \widetilde{\pi}, y(t))| \right] \leq \frac{R\alpha^T}{1-\alpha}. \tag{4.46}$$

**(d)**

$$E_y^{\varphi,\widetilde{\varphi}} |V_*^N \left( y^N(t) \right) - v_* \left( y(t) \right)| \leq \frac{2R\alpha^T}{1-\alpha} + L_r \frac{1-\alpha^T}{1-\alpha} \left[ KTe^{-\lambda N \varepsilon^2} \left( 1 + \mathcal{K}(T) \right) + \gamma_T(\varepsilon) \right]. \tag{4.47}$$

*Demostración.* (a) Notemos que bajo la Hipótesis 4(e) y (4.39) para cada $(\pi, \widetilde{\pi}) \in \Pi_M \times \Pi_\Gamma$

$$\begin{aligned} & E_y^{\pi,\widetilde{\pi}} \left| r(y^N(t), a_t^{\pi,N}) - r(y(t), a_t^\pi) \right| \leq L_r E_y^{\pi,\widetilde{\pi}} \left[ \|y^N(t) - y(t)\|_\infty \right] \\ & \leq L_r E_y^{\pi,\widetilde{\pi}} \left[ \sup_{0\leq t\leq T} \|y^N(t) - y(t)\|_\infty \right] \leq L_r \left( KTe^{-\lambda N \varepsilon^2}(1 + \mathcal{K}(T)) + \gamma_T(\varepsilon) \right). \end{aligned} \tag{4.48}$$

Ahora, para cada par de políticas $(\varphi, \widetilde{\varphi}) \in \Pi_M \times \Pi_\Gamma$, denotemos las correspondientes trayectorias como $y^N(t) := y_{\varphi,\widetilde{\varphi}}^N(t)$ y $y(t) := y_{\varphi,\widetilde{\varphi}}(t)$. Entonces tenemos

$$
\begin{aligned}
&\sup_{(\pi,\widetilde{\pi})\in\Pi_M\times\Pi_\Gamma} |V_T(\pi,\widetilde{\pi},y^N(t)) - v_T(\pi,\widetilde{\pi},y(t))| \\
&= \sup_{(\pi,\widetilde{\pi})\in\Pi_M\times\Pi_\Gamma} \left| E^{\pi,\widetilde{\pi}}_{y^N(t)} \left[ \sum_{k=0}^{T-1} \alpha^k r(y^N(k), a_k^{\pi,N}) - \sum_{k=0}^{T-1} \alpha^k r(y(k), a_k^{\pi}) \right] \right| \\
&\leq \sum_{k=0}^{T-1} \alpha^k \sup_{(\pi,\widetilde{\pi})\in\Pi_M\times\Pi_\Gamma} E^{\pi,\widetilde{\pi}}_{y^N(t)} \left| r(y^N(k), a_k^{\pi,N}) - r(y(t), a_k^{\pi}) \right| \\
&\leq L_r \frac{1-\alpha^T}{1-\alpha} \left[ KTe^{-\lambda N\varepsilon^2}(1+\mathcal{K}(T)) + \gamma_T(\varepsilon) \right] \quad P_y^{\varphi,\widetilde{\varphi}}\text{-c.s.},
\end{aligned} \tag{4.49}
$$

donde la última desigualdad se sigue de (4.48). Así (4.44) se tiene tomando esperanza $E_y^{\varphi,\widetilde{\varphi}}(\cdot)$ en ambos lados de (4.49).

(b) De nuevo, para cada par de políticas $(\varphi,\widetilde{\varphi}) \in \Pi_M \times \Pi_\Gamma$, denotemos las correspondientes trayectorias como $y^N(t) := y^N_{\varphi,\widetilde{\varphi}}(t)$ y $y(t) := y_{\varphi,\widetilde{\varphi}}(t)$. Así, obtenemos

$$
\begin{aligned}
&\sup_{(\pi,\widetilde{\pi})\in\Pi_M\times\Pi_\Gamma} \left| V^N(\pi,\widetilde{\pi},y^N(t)) - V_T^N(\pi,\widetilde{\pi},y^N(t)) \right| \\
&\leq \sup_{(\pi,\widetilde{\pi})\in\Pi_M\times\Pi_\Gamma} \left| E^{\pi,\widetilde{\pi}}_{y^N(t)} \left[ \sum_{k=0}^{\infty} \alpha^k r(y^N(k), a_k^{\pi,N}) - \sum_{k=0}^{T-1} \alpha^k r(y^N(k), a_k^{\pi,N}) \right] \right| \\
&\leq \sum_{k=T}^{\infty} \alpha^k \sup_{(\pi,\widetilde{\pi})\in\Pi_M\times\Pi_\Gamma} E^{\pi,\widetilde{\pi}}_{y^N(t)} |r(y^N(k), a_k^{\pi,N})| \leq R \sum_{k=T}^{\infty} \alpha^k \leq \frac{R\alpha^T}{1-\alpha} \quad P_y^{\varphi,\widetilde{\varphi}}\text{-c.s.},.
\end{aligned} \tag{4.50}
$$

Por lo tanto, (3.44) se obtiene tomando esperanza $E_y^{\varphi,\widetilde{\varphi}}(\cdot)$ en ambos lados de (4.50).

(c) Por argumentos similares usados en la demostración de la parte (b) se prueba la parte (c).

(d) Definamos la política estacionaria $\bar{\pi} = \{\bar{f}\} \in \Pi_M$, donde $\bar{f} : \mathbb{Y} \to A$ está dada por

$$
\bar{f}(y) = f_*^N(y) I_{\mathbb{Y}_N}(y) + \tilde{f}(y) I_{[\mathbb{Y}_N]^c}(y),
$$

siendo $\pi_*^N = \left\{f_*^N\right\} \in \Pi_M^N$ una política minimax estacionaria para el $N-$Minimax $\mathcal{M}_{mxm}^N$ (ver Teorema 4.3.2), y $\tilde{f} \in \mathbb{F}$, un selector arbitrario. Además, para $\varphi \in \Pi_M$ una política arbitraria fija, denotemos $y_\varphi^N(t) = y^N(t) \in \mathbb{Y}_N$ y $y_\varphi(t) := y(t) \in \mathbb{Y}$. Observemos que para cada $t \in \mathbb{N}_0$ y $\widetilde{\pi} \in \Pi_\Gamma$,

$$V_*^N(y^N(t)) = \sup_{\widetilde{\pi} \in \Pi_\Gamma} V^N(\pi_*^N, \widetilde{\pi}, y^N(t)) = \sup_{\widetilde{\pi} \in \Pi_\Gamma} V^N(\bar{\pi}, \widetilde{\pi}, y^N(t)) \leq \sup_{(\pi, \widetilde{\pi}) \in \Pi_M \times \Pi_\Gamma} V^N(\pi, \widetilde{\pi}, y^N(t)).$$

Por lo que,

$$V_*^N(y^N(t)) - v_*(y(t)) \leq \sup_{(\pi, \widetilde{\pi}) \in \Pi_M \times \Pi_\Gamma} V^N(\pi, \widetilde{\pi}, y^N(t)) - \inf_{\pi \in \Pi_M} \sup_{\widetilde{\pi} \in \Pi_\Gamma} v(\pi, \widetilde{\pi}, y(t)),$$

que a su vez implica

$$\left|V_*^N(y^N(t)) - v_*(y(t))\right| \leq \sup_{(\pi, \widetilde{\pi}) \in \Pi_M \times \Pi_\Gamma} \left|V^N(\pi, \widetilde{\pi}, y^N(t)) - v(\pi, \widetilde{\pi}, y(t))\right|, \quad t \in \mathbb{N}_0.$$

Por lo tanto, para cada $y \in \mathbb{Y}_N, 0 \leq t < T$,

$$\begin{aligned}
\left|V_*^N(y^N(t)) - v_*(y(t))\right| &\leq \sup_{(\pi, \widetilde{\pi}) \in \Pi_M \times \Pi_\Gamma} \left\{|V^N(\pi, \widetilde{\pi}, y^N(t)) - V_T^N(\pi, \widetilde{\pi}, y^N(t))|\right. \\
&\quad + |V_T^N(\pi, \widetilde{\pi}, y^N(t)) - v_T(\pi, \widetilde{\pi}, y(t))| \\
&\quad \left. + |v_T(\pi, \widetilde{\pi}, y(t)) - v(\pi, \widetilde{\pi}, y(t))|\right\} \\
&\leq \sup_{(\pi, \widetilde{\pi}) \in \Pi_M \times \Pi_\Gamma} |V^N(\pi, \widetilde{\pi}, y^N(t)) - V_T^N(\pi, \widetilde{\pi}, y^N(t))| \\
&\quad + \sup_{(\pi, \widetilde{\pi}) \in \Pi_M \times \Pi_\Gamma} |V_T^N(\pi, \widetilde{\pi}, y^N(t)) - v_T(\pi, \widetilde{\pi}, y(t))| \\
&\quad + \sup_{(\pi, \widetilde{\pi}) \in \Pi_M \times \Pi_\Gamma} |v_T(\pi, \widetilde{\pi}, y(t)) - v(\pi, \widetilde{\pi}, y(t))| \\
&\leq \frac{2R\alpha^T}{1-\alpha} + L_r \frac{1-\alpha^T}{1-\alpha} \left[KTe^{-\lambda N \varepsilon^2}(1 + \mathcal{K}(T)) + \gamma_T(\varepsilon)\right] P_y^{\varphi, \widetilde{\varphi}}\text{-c.s.},
\end{aligned} \tag{4.51}$$

donde la última desigualdad se debe a las partes (a)-(c). Finalmente, tomando esperanza $E_y^{\varphi, \widetilde{\varphi}}(\cdot)$ en ambos lados de (4.51), obtenemos (4.47). □

*Observación* 4.4.1. *Observe que de (2.7) y (3.5) tenemos*

$$y^N(1) = H^N(y^N(0), a_0, \mu_0, \vec{w}_0) := y^N_{\mu_0}(1)$$

*y*

$$y(1) = H(y(0), a_0, \mu_0) := y_{\mu_0}(1).$$

*Entonces, (4.47) implica que para cada* $y \in \mathbb{Y}_N \subset \mathbb{Y}$, $\varepsilon > 0$, $T \in \mathbb{N}$, *y para todo* $(\varphi, \tilde{\varphi}) \in \Pi_M \times \Pi_\Gamma$,

$$\sup_{\mu \in \Gamma} E_y^{\varphi,\tilde{\varphi}} |V_*^N \left(y_\mu^N (1)\right) - v_* \left(y_\mu (1)\right)| \leq \frac{2R\alpha^T}{1-\alpha} + L_r \frac{1-\alpha^T}{1-\alpha} \left[ KTe^{-\lambda N \varepsilon^2} \left(1 + \mathcal{K}(T)\right) + \gamma_T (\varepsilon) \right]. \tag{4.52}$$

**Prueba del Teorema 4.4.2**.

Sea $\Phi(\cdot,\cdot) \geq 0$ la función de discrepancia para el modelo de juego contra la naturaleza de campo medio definida como

$$\Phi(y,a) := \max_{\mu \in \Gamma} \{ r(y,a) + \alpha v_* \left[ H(y,a,\mu) \right] \} - v_*(y), \;\; y \in \mathbb{Y}. \tag{4.53}$$

(compare (4.23)) Notemos que de la definición de política minimax de campo medio $\pi^* = \{f^*\}$ (ver Teorema 4.4.1), tenemos $\Phi(y, f^*) = 0$.

Sea $\tilde{\pi} \in \Pi_\Gamma$ una política arbitraria para el oponente, con $\mu_0 = \mu \in \Gamma$. Entonces, para cada $y^N(0) = y(0) = y \in \mathbb{Y}_N \subset \mathbb{Y}$, de la propiedad de Markov tenemos

$$\int_{\mathbb{R}^N} V_*^N \left[ H^N(y, f^*, \mu, w) \right] \theta(dw) - v_* \left[ H(y, f^*, \mu) \right] = E_y^{\pi^*, \tilde{\pi}} \left[ V_*^N (y^N(1)) - v_*(y(1)) | h_0^N, \mu \right] \tag{4.54}$$

lo cual a su vez implica

$$\left| \int_{\mathbb{R}^N} V_*^N \left[ H^N(y, f^*, \mu, w) \right] \theta(dw) - v_* \left[ H(y, f^*, \mu) \right] \right| = E_y^{\pi^*, \tilde{\pi}} \left| V_*^N (y^N(1)) - v_*(y(1)) \right|. \tag{4.55}$$

Esto nos lleva a que

$$
\begin{aligned}
\Phi^{N}(y,f^{*}) &= |\Phi^{N}(y,f^{*})-\Phi(y,f^{*})| \\
&= \left|\sup_{\mu\in\Gamma}\left\{r(y,f^{*})+\alpha\int_{\mathbb{R}^N}V_{*}^{N}\left[H^{N}(y,f^{*}\,\mu,w)\right]\theta(dw)\right\}-V_{*}^{N}(y)\right. \\
&\left.-\sup_{\mu\in\Gamma}\left\{r(y,f^{*})+\alpha v_{*}\left[H(y,f^{*},\mu)\right]\right\}+v_{*}(y)\right| \\
&\leq |V_{*}^{N}(y)-v_{*}(y)| \\
&+\alpha\sup_{\mu\in\Gamma}\left|\int_{\mathbb{R}^N}V_{*}^{N}\left[H^{N}(y,f^{*},\mu,w)\right]\theta(dw)-v_{*}\left[H(y,f^{*},\mu)\right]\right| \\
&\leq |V_{*}^{N}(y)-v_{*}(y)| \\
&+\alpha\sup_{\mu\in\Gamma}E_{y}^{\pi^{*},\widetilde{\pi}}|V_{*}^{N}\left(y^{N}(1)\right)-v_{*}(y(1))|,
\end{aligned}
\tag{4.56}
$$

donde la última desigualdad se debe a (4.55). Usando este hecho junto con la relación (4.52), el Lema 4.4.3 implica

$$
0\leq\Phi^{N}(y,f^{*})\leq\frac{4R\alpha^{T}}{1-\alpha}+2L_{r}\frac{1-\alpha^{T}}{1-\alpha}\left[KTe^{-\lambda N\varepsilon^{2}}(1+\mathcal{K}(T))+\gamma_{T}(\varepsilon)\right],
$$

para $\varepsilon>0$ y $T\in\mathbb{N}$ arbitrarios. Por lo tanto, haciendo $N\to\infty$ demostramos el Teorema 4.4.2, esto es

$$
\lim_{N\to\infty}\Phi^{N}(y,f^{*})=0.\blacksquare
$$

# Capítulo 5

# Ejemplo: Modelo de consumo-inversión

## 5.1. Introducción

Con el fin de ilustrar nuestros resultados, presentaremos el siguiente sistema de consumo-inversión que se compone de un número grande de pequeños inversionistas. Mostraremos que las hipótesis asumidas en este trabajo se satisfacen. Esencialmente nos enfocaremos a definir las dinámicas de las proporciones y del contexto del ambiente, en verificar las Hipótesis 1-3, y mostrar que la función $G_\rho$ es Lipschitz (ver 3.3), todo esto para el problema de estimación y control. La verificación de las Hipótesis 4 y 5 establecidas para el modelo del juego contra la naturaleza se hace mediante procedimientos completamente similares, y por lo tanto las omitiremos.

## 5.2. Un modelo de consumo-inversión con subsidios/impuestos controlado

Consideremos un sistema de consumo-inversión compuesto por $N$ "pequeños "inversionistas, es decir, agentes económicos cuyas decisiones individuales no tienen influencia en el mercado de precios. Estos agentes invierten en varios activos con diferentes tasas de retorno y consumen algún producto específico. Además, supondremos que existe un controlador central, por ejemplo, el gobierno o algúna organización pública que apoya con un subsidio para incentivar a los inversionistas o impone una impuesto que los inversionistas deben pagar. Por simplicidad, consideraremos solo dos activos para los inversionistas: uno de ellos es libre de riesgo con tasa de retorno fijo $\tau$, y el otro tiene tasa de retorno aleatoria $\xi_t$ que toma valores en un conjunto acotado $Z \subseteq \mathbb{R}$. La fracción de capital asociada que se invierte en la opción con riesgo es una función $\varphi_1 : \mathbb{N} \to [0, 1]$ que depende del contexto del ambiente, el cual toma valores en el conjunto $\mathbb{N}$. Naturalmente, $(1-\varphi_1)$ representará la fracción de capital invertida en la opción libre de riesgo. Por otro lado, supondremos que cada inversionista consume una cantidad $\varphi_2 : \mathbb{N} \to \mathbb{R}_+$ que también es una función acotada que depende del contexto del ambiente.

En el escenario de nuestro trabajo, dado que el espacio de estados $S$ es 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 $a_t \in A := \{0, \pm 1, ... \pm a^*\}$ para algún $a^* \geq 0$. Esto es, al tiempo $t$

$$a_t = \text{ impuesto de tamaño } -a_t \text{ (si } a_t < 0) \text{ o subsidio de tamaño } a_t \text{ (si } a_t > 0).$$

Denotando por $X_n^N(t) \in S := \{0, 1, \cdots, s\}$ el capital del inversionista $n$ al tiempo

$t$. Entonces

$$X_n^N(t{+}1) = \operatorname{int}\left\{\left[(1-\varphi_1(C^N(t)))(1+\tau)+\varphi_1(C^N(t))\xi_t\right]\left[X_n^N(t)-\varphi_2(C^N(t))+a_t\right]\right\}, \tag{5.1}$$

donde $\operatorname{int}\{x\}$ representa la parte entera de $x$. Suponemos que $s \in \mathbb{N}_0$ es lo suficientemente grande y las funciones $\varphi_m$, $m = 1, 2$, satisfacen una condición de Lipschitz con constantes $L_{\varphi_m}$, $m = 1, 2$, respectivamente, tomando valores en conjuntos apropiados tales que se satisface lo siguiente

$$F(i,c,a,z) := \operatorname{int}\left\{\left[(1-\varphi_1(c))(1+\tau)+\varphi_1(c)z\right]\left[i-\varphi_2(c)-a\right]\right\} \in S. \tag{5.2}$$

**Lema 5.2.1.** La función $F$ definida en (5.2) es Lipschitz con constante

$$L_F := 2 + L_{\varphi_2} + (1+\tau+z)\left(iL_{\varphi_1} + a^* L_{\varphi_1} + \bar{L}_{\varphi_2} + \bar{L}_{\varphi_1} L_{\varphi_2} + L_{\varphi_1}\bar{L}_{\varphi_2}\right),$$

donde $\bar{L}_{\varphi_m}$ es una cota de $\varphi_m$, $m = 1, 2$ y $a^* = \max_{a \in A} |a|$.

*Demostración.* Observemos primero que para cualquier par de números reales $a$ y $b$ se satisface

$$\begin{aligned} |\operatorname{int}\{a\} - \operatorname{int}\{b\}| &= |a+\alpha-b-\beta| = |a-b+\alpha-\beta|. \\ &\leq |a-b| + 1 \end{aligned} \tag{5.3}$$

De esta manera, para $(a,c),(a',c') \in A\times\mathbb{N}$, un estado $i\in S$ y $z\in\mathbb{R}$ fijos por (5.3)

$$\begin{aligned}|F(i,c,a,z)-F(i,c',a',z)| &\leq |[(1-\varphi_1(c))(\tau+1)][i-\varphi_2(c)-a]\\ &-[(1-\varphi_1(c'))(\tau+1)][i-\varphi_2(c')-a']|+1\\ &\leq 1+|\varphi_2(c)-\varphi_2(c')|+|a-a'|+i(\tau+1)|\varphi_1(c)-\varphi_1(c')|\\ &+(\tau+1)|\varphi_1(c)\varphi_2(c)-\varphi_1(c')\varphi_2(c')|\\ &+(1+\tau)|a\varphi_1(c)-a'\varphi_1(c')|+iz|\varphi_1(c)-\varphi_1(c')|\\ &+z|\varphi_1(c)\varphi_2(c)-\varphi_1(c')\varphi_2(c')|+z|a\varphi_1(c)-a'\varphi_1(c')|\\ &\leq 1+L_{\varphi_2}\left\|c-c'\right\|_\infty^2+|a-a'|+i(\tau+1)L_{\varphi_1}|c-c'|\\ &+(\tau+1)(\bar{L}_{\varphi_1}L_{\varphi_2}+L_{\varphi_1}\bar{L}_{\varphi_2})\left\|c-c'\right\|_\infty^2\\ &+(\tau+1)(a^*L_{\varphi_1}\left\|c-c'\right\|_\infty^2+\bar{L}_{\varphi_1}|a-a'|)+izL_{\varphi_1}\left\|c-c'\right\|_\infty^2\\ &+z\left\|c-c'\right\|_\infty^2(\bar{L}_{\varphi_1}L_{\varphi_2}+\bar{L}_{\varphi_2}L_{\varphi_1})\\ &\leq \max\{\left\|c-c'\right\|_\infty^2,|a-a'|\}\,[2+L_{\varphi_2}+(z+\tau+1)(iL_{\varphi_1}\\ &+a^*L_{\varphi_1}+\bar{L}_{\varphi_2}+\bar{L}_{\varphi_1}L_{\varphi_2}+\bar{L}_{\varphi_2}L_{\varphi_1})]\end{aligned}$$

De aquí podemos deducir que $F$ satisface la siguiente condición de Lipschitz:

$$|F(i,c,a,z)-F(i,c',a',z)| \leq L_F \max\left\{\left\|c-c'\right\|_\infty^2\ ,\ |a-a'|\right\}, \tag{5.4}$$

donde

$$L_F = 1+(1+\tau+\max\nolimits_{z\in Z}|z|)\left(L_{\varphi_1}s+\bar{L}_{\varphi_1}L_{\varphi_2}+\bar{L}_{\varphi_2}L_{\varphi_1}+a^*L_{\varphi_1}+L_{\varphi_2}\right)+(1+\tau)(1+L_{\varphi_2}).$$

□

Ahora, si $\rho$ es la densidad de la tasa de retorno $\xi_t$, la ley de transición toma la forma

$$K_{ij}^{\rho}(a,c)=\int_{\mathbb{R}} I_j[F(i,c,a,z)]\rho(z)dz, \tag{5.5}$$

para cada $i, j \in S$ y $(a, c) \in A \times \mathbb{N}$. Además, dado que $F$ es una función con valores en $S := \{0, 1, \cdots, s\}$, es fácil ver que para todo $i, j \in S$, $a, a' \in A$, $c, c' \in \mathbb{N}$, la función indicadora satisface la desigualdad

$$\begin{aligned}\left|1_j[F(i,c,a,z)] - 1_j[F(i,c',a',z)]\right| &\leq \left|F(i,c,a,z) - F(i,c',a',z)\right| \\ &\leq L_F \,\text{máx}\left\{\|c-c'\|_\infty^2, |a-a'|\right\},\end{aligned}$$

debido a la propiedad de Lipschitz de $F$ dada en (5.2.1). Así, de (2.2),

$$\begin{aligned}\left|K_{ij}^\rho(a,c) - K_{ij}^\rho(a',c')\right| &\leq \int \left|I_j[F(i,c,a,z)] - I_j[F(i,c',a',z)]\right| \rho(z)dz \\ &\leq L_F \,\text{máx}\left\{\|c-c'\|_\infty^2, |a-a'|\right\}, \qquad (5.6)\end{aligned}$$

lo cual implica que $K_\rho$ es de Lipschitz.

A continuación procederemos a mostrar la ecuación en diferencias que determina la dinámica de las proporciones $\{\vec{M}^N(t)\}$ (vea (2.5)). La base para obtener la función $G_\rho^N$ es el siguiente resultado, el cual se demuestra en [13], Teorema 1.

**Teorema 5.2.2.** Sea $\{a_t\}$ cualquier sucesión de acciones elegidas del conjunto $A$, entonces el proceso $\vec{M}^N(t)$ es de Markov.

Puesto que $\vec{M}^N(t)$, $t \in \mathbb{N}$, pertenece al conjunto $\mathbb{P}_N(S)$ (ver (2.3)), es una cadena de Markov con espacio de estados finito. Por otra parte, note que (ver [12])

$$M_i^N(t+1) = \frac{1}{N}\sum_{k=0}^{s}\sum_{n=1}^{NM_k^N(t)} I_{\{A_{ki}^\rho(a_t, C^N(t))\}}(w_n^k(t)), \qquad (5.7)$$

donde

$$A_{ki}^\rho(a,c) := [\Gamma_{ki}^\rho(a,c), \Gamma_{ki+1}^\rho(a,c)] \subseteq [0,1], \qquad (5.8)$$

$$\Gamma_{ki}^\rho(a,c) := \sum_{l=0}^{i-1} K_{kl}^\rho(a,c), k, i \in S. \qquad (5.9)$$

y $w_k^n(t)$ son v.a. uniformes en $[0,1]$ independientes.

Para cada $i \in S$ y $t \in \mathbb{N}_0$, denotemos

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

y

$$\vec{w}_t := (\vec{w}^0(t), \cdots, \vec{w}^s(t)).$$

Cabe notar que $\sum_{i=0}^s NM_i^N(t) = N$, así $\vec{w}_t \in [0,1]^N$. Esta afirmación implica que el número de v.a. (uniformes) involucradas en la dinámica (5.7) coincide con el número $N$ de pequeños inversionistas; un hecho que es presentado de manera general mediante la dinámica (2.5).

Reescribiendo las expresiones anteriores como en (2.7), y considerando que $\mathbb{Y}_N := \mathbb{P}_N(S) \times \mathbb{N}$ y $\mathbb{Y} := \mathbb{P}(S) \times \mathbb{N}$, definamos, para $y^N(t) = (M^N(t), C^N(t)) \in \mathbb{Y}_N$,

$$G_{\rho,i}^N(y^N(t), a_t, \vec{w}_t) := \frac{1}{N} \sum_{k=0}^{s} \sum_{n=1}^{NM_k^N(t)} I_{\{A_{ki}^\rho(C^N(t), a_t)\}}(w_n^k), \quad i \in S.$$

La función $G_\rho^N$ toma la siguiente forma vectorial

$$G_\rho^N(y,a,w) = \left(G_{\rho,0}^N(y,a,w), ..., G_{\rho,s}^N(y,a,w)\right), \quad (y,a,w) \in \mathbb{Y}_N \times [-a^*, a^*] \times [0,1]^N, \tag{5.10}$$

dando lugar a la siguiente expresión

$$\vec{M}^N(t+1) = G_\rho^N(\vec{M}^N(t), C^N(t), a_t, \vec{w}_t). \tag{5.11}$$

Finalmente, para definir la dinámica del proceso $\{y^N(t)\}$, supondremos que $g : \mathbb{N} \times \mathbb{P}(S) \times [-a^*, a^*] \to \mathbb{R}$ es una función arbitraria que satisface la Hipótesis 1(b), de manera que el contexto del ambiente evoluciona de acuerdo a la siguiente ecuación:

$$C^N(t+1) = g(C^N(t), \vec{M}^N(t+1), a_t), \;\; t \in \mathbb{N}_0. \tag{5.12}$$

Entonces, (5.11) y (5.12) definen la función

$$H_\rho^N(y,a,w) := (G_\rho^N(y,a,w), g(c, G_\rho^N(y,a,w), a)), \tag{5.13}$$

la cual determina la dinámica del proceso $\left\{y^N(t)\right\}$ similar a (ver (2.7)).

Ahora, supongamos que la función de costo por etapa $r$ es acotada y uniformemente de Lipschitz con constante $L_r$; esto es, para alguna constante $R > 0$

$$|r(y,a)| \leq R \quad \forall (y,a) \in \mathbb{Y} \times A,$$

y para cada $a, a' \in A$ y $y, y' \in \mathbb{Y}$,

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

Finalmente, dado que el espacio de acciones $A$ es finito, por tanto compacto, la continuidad de $a \longmapsto H_\rho^N(\cdot, a, \cdot)$, requerida en Hipótesis 1(c), se cumple trivialmente. Este hecho, junto con la consideración de que la función en (5.12) es de Lipschitz, implica la Hipótesis 1.

### 5.2.1. Proceso de Campo Medio

Continuando con nuestro ejemplo, en esta parte definimos el sistema de control determinista $\{(\vec{m}(t), c(t))\} \in \mathbb{Y}$ como

$$\vec{m}(t+1) = \vec{m}(t) K_\rho(a_t, c(t)) \tag{5.14}$$

$$c(t+1) = g(c(t), \vec{m}(t+1), a_t), \tag{5.15}$$

donde $K_\rho$ es la matriz $[K_{ij}^\rho]$, cuyos elementos son kérneles estocásticos definidos en (5.5) y $g : \mathbb{N} \times \mathbb{P}(S) \times A \to \mathbb{N}$ es la función definida en (5.12). Observemos que $\vec{m}(t+1)$ es el vector con componentes

$$m_j(t+1) = \sum_{i=1}^{s} m_i(t) K_{ij}^\rho(a_t, c(t)),$$

donde $\vec{m}(0) = m \in \mathbb{P}(S)$. En este caso la función $G_\rho$ en (3.1) toma la forma

$$G_\rho(\vec{m}, c, a) = \vec{m} K_\rho(a, c), \quad (\vec{m}, c) \in \mathbb{Y}, a \in A, \tag{5.16}$$

y como el kernel $K_\rho$ es de Lipschitz (ver (5.6)), también lo es $G_\rho$, como fue establecido en (3.3), con constante de Lipschitz $L_G$.

Con el fin de definir el proceso de control determinista, consideremos la función $H_\rho : \mathbb{Y} \times A \to \mathbb{Y}$ que define la dinamica del proceso $\{(\vec{m}(t), c(t))\}$; esto es,

$$H_\rho(y, a) := (G_\rho(\vec{m}, c, a), g(c, G(\vec{m}, c, a), a)), \quad y = (\vec{m}, c) \in \mathbb{Y}, \ a \in A, \ \mu \in \Gamma. \tag{5.17}$$

De (3.1) y (3.2), podemos escribir

$$\begin{aligned} y(t+1) &= (G_\rho(\vec{m}(t), c(t), a_t), g(c(t), \vec{m}(t+1), a_t)) \\ &=: H_\rho(y(t), a_t), \quad t \geq 0, \end{aligned} \tag{5.18}$$

con $y(0) = (\vec{m}, c) \in \mathbb{P}(S) \times \mathbb{N} = \mathbb{Y}$.

Por el Lema 3.2.1, tenemos que la función definida en (5.18) es de Lipschitz con constante $L_{H_\rho}$, y para garantizar la Hipótesis 2 supondremos que se satisface $\alpha L_{H_\rho} < 1$ donde $\alpha$ es el factor de descuento.

Más adelante mostraremos que (5.14)-(5.15) es, en efecto, el proceso límite de (5.11)-(5.12).

### 5.2.2. Proceso de estimación

Sea $\rho_k$ un estimador de la densidad $\rho$ que satisface (3.12). El objetivo es mostrar la convergencia establecida en (3.13)

Para el estimador $\rho_k$, definimos, de manera similar a (5.5), el kernel de transición estimado $K_k(a,c) = \left[K_{ij}^k(a,c)\right]$ con componentes (ver (5.2))

$$K_{ij}^k(a,c) =: \int_{\mathbb{R}} I_j[F(i,c,a,z)]\rho_k(z)dz, \quad i,j \in S, \ (a,c) \in A \times \mathbb{R}.$$

También definimos

$$G_{\rho_k}(\vec{m},c,a) := \vec{m}K_k(a,c), \quad (\vec{m},c) \in \mathbb{Y}, a \in A,$$

y

$$H_k(y,a) := (\vec{m}K_k(a,c), g(c,\vec{m}K_k(a,c),a)), \quad y = (\vec{m},c) \in \mathbb{Y}, a \in A.$$

Observemos que para cada $i,j \in S, \ (a,c) \in A \times \mathbb{N}$,

$$\left|K_{ij}^k(a,c) - K_{ij}^\rho(a,c)\right| \leq \int_{\mathbb{R}} |\rho_k(z) - \rho(z)|\, dz.$$

Por lo tanto, de acuerdo a (3.12)

$$\sup_{(a,c)\in A\times\mathbb{N}} \|K_k(a,c) - K_\rho(a,c)\|_\infty^0 \to 0 \ \text{ c.s., cuando } \ t \to \infty, \tag{5.19}$$

lo cual implica (ver (5.16))

$$\sup_{(y,a)\in\mathbb{Y}\times A} \|G_{\rho_k}(y,a) - G_\rho(y,a)\|_\infty^1 =$$

$$\sup_{(y,a)\in\mathbb{Y}\times A} \|\vec{m}K_k(a,c) - \vec{m}K_\rho(a,c)\|_\infty^1 \to 0 \ \text{ c.s., cuando } \ k \to \infty. \tag{5.20}$$

### 5.2.3. Convergencia en Campo Medio

Para concluir, demostraremos que la Hipótesis 3(b) se cumple, es decir, (5.14)-(5.15) es el proceso límite de (5.11)-(5.12). Sea $\pi = \{f_t\} \in \Pi_M$ una política arbitraria y $y \in \mathbb{Y}_N \subset \mathbb{Y}$ el estado inicial. Definimos

$$B_{inj}^{N\rho}(t) := I_{\left\{A_{ij}^\rho\left(a_t^{\pi,N}, C^N(t)\right)\right\}}(w_n^i(t)), \quad i,j \in S, n \in \mathbb{N},$$

donde $C^N(t)$ es como en (5.12) y $w_n^i(t)$ son v.a.i.i.d. con distribución uniforme en $[0,1]$. Observemos que para cada $t \in \mathbb{N}_0$, $\left\{B_{inj}^{N\rho}(t)\right\}_{inj}$ son v.a.i.i.d. con distribución Bernoulli cuya media es

$$E_y^\pi\left[B_{inj}^{N\rho}(t)|a_t^{\pi,N}=a, C^N(t)=c\right] = K_{ij}^\rho(a,c)$$
$$= \int I_j[F(i,c,a,z)]\rho(z)dz, \quad i,j \in S, \ (a,c) \in A \times \mathbb{N}.$$

Entonces, para un $\varepsilon > 0$ fijo, por la desigualdad de Hoeffding, tenemos

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

Consideremos el conjunto

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

(ver Observación 2.4.1(b)) y sea $\varepsilon_t$ un número positivo tal que $\left\|y^N(t) - y(t)\right\|_\infty \leq \varepsilon_t$; esto es,

$$\left\|\vec{M}^N(t) - \vec{m}(t)\right\|_\infty^1 \leq \varepsilon_t \ \text{ y } \ \left\|C^N(t) - c(t)\right\|_\infty^2 \leq \varepsilon_t. \tag{5.21}$$

Así, de (5.7), (5.14) y (5.21), tenemos que las siguientes relaciones son ciertas en $\bar{\Omega}$:

$$\begin{aligned}
\left|M_j^N(t+1) - m_j(t+1)\right| &= \left|\sum_{i=0}^{s} \frac{1}{N}\left[\sum_{n=1}^{NM_i^N(t)} B_{inj}^{N\rho}(t) - Nm_i(t)K_{ij}^\rho\left(a_t^{\pi,N}, c(t)\right)\right]\right| \\
&\leq \sum_{i=0}^{s} \frac{1}{N}\left|\sum_{n=1}^{NM_i^N(t)} B_{inj}^{N\rho}(t) - Nm_i(t)K_{ij}^\rho\left(a_t^{\pi,N}, c(t)\right)\right| \\
\leq \sum_{i=0}^{s} \frac{1}{N}\left|\sum_{n=1}^{NM_i^N(t)} B_{inj}^{N\rho}(t) - NM_i^N(t)K_{ij}^\rho\left(a_t^{\pi,N}, C^N(t)\right)\right| &+ \sum_{i=0}^{s}\left|M_i^N(t) - m_i(t)\right| K_{ij}^\rho\left(a_t^{\pi,N}, C^N(t)\right) \\
+\sum_{i=0}^{s} m_i(t)\left|K_{ij}^\rho\left(a_t^{\pi,N}, C^N(t)\right) - K_{ij}^\rho\left(a_t^{\pi,N}, c(t)\right)\right| &< (s+1)\varepsilon + (s+1)\varepsilon_t + L_K\varepsilon_t.
\end{aligned} \tag{5.22}$$

Por lo tanto, dado que el lado derecho de la última desigualdad no depende de $j$, tenemos

$$\left\|\vec{M}^N(t+1)-\vec{m}(t+1)\right\|_\infty^1 \leq (s+1)\varepsilon+(s+1)\varepsilon_t+L_K\varepsilon_t.$$

Por otro lado, dado que $g$ es una función de Lipschitz (ver Hipótesis 1), las expresiones (5.12) y (5.15)) junto con (5.21) y (5.22) nos conducen a

$$\left\|C^N(t+1)-c(t+1)\right\|_\infty^2 \leq \left\|g(C^N(t),\vec{M}^N(t+1),a_t^{\pi,N})-g(c(t),\vec{m}(t+1),a_t^{\pi,N})\right\|_\infty^2$$
$$< L_g \max\{\varepsilon_t,(s+1)\varepsilon+(s+1)\varepsilon_t+L_K\varepsilon_t\} = L_g\left((s+1)\varepsilon+(s+1)\varepsilon_t+L_K\varepsilon_t\right),$$

lo cual implica que en el conjunto $\bar{\Omega}$ (recordemos que $L_g \geq 1$)

$$\left\|y^N(t+1)-y(t+1)\right\|_\infty < L_g\left((s+1)\varepsilon+(s+1)\varepsilon_t+L_K\varepsilon_t\right).$$

Considerando ahora $\left\|y^N(0)-y(0)\right\|_\infty=\varepsilon_0=0$ (ver Hipótesis 3(a)) y aplicando un procedimiento inductivo, un cálculo directo nos lleva a que en el conjunto $\bar{\Omega}$,

$$\left\|y^N(t+1)-y(t+1)\right\|_\infty < L_g(s+1)\varepsilon\beta_t, \quad t\in\mathbb{N}_0,$$

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

$$\left\|y^N(t+1)-y(t+1)\right\|_\infty < L_g(s+1)\varepsilon\beta_T, \quad \forall 0\leq t\leq T,$$

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

$$P_y^\pi\left[\sup_{0\leq t\leq T}\left\|y^N(t+1)-y(t+1)\right\|_\infty < L_g(s+1)\varepsilon\beta_T\right] \geq 1-2e^{-2N\varepsilon^2}.$$

De aquí, si hacemos $\gamma_T(\varepsilon):=L_g(s+1)\varepsilon\beta_T$, $K=\lambda=2$, implica que

$$\sup_{\pi\in\Pi_M} P_y^\pi\left\{\sup_{0\leq t\leq T}\left\|y^N(t)-y(t)\right\|_\infty \geq \gamma_T(\varepsilon)\right\} \leq KTe^{-\lambda N\varepsilon^2}.$$

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

Por lo que la Hipótesis 3 se cumple.

# Capítulo 6

# Conclusiones y problemas abiertos

En los últimos años la teoría de campo medio ha tenido un gran desarrollo por la extensa variedad de sus aplicaciones en distintas áreas como economía, finanzas, ecología, biología e investigación de operaciones (ver [3, 6, 8, 9, 11, 12, 13, 14, 24, 25, 28, 32, 37, 39]). Dentro de estas aplicaciones se encuentra el análisis de sistemas dinámicos que se componen de un número grande de objetos (partículas o agentes) que interactúan entre sí. Mas aún, suponiendo que existe un controlador que puede influir en el comportamiento de los objetos, es posible plantear un problema de control óptimo. Este tipo de problemas fue la motivación principal para el desarrollo del trabajo.

En particular se estudió un sistema compuesto por una gran cantidad $N$ de objetos que interactúan entre sí, los cuales se pueden clasificar y además comparten un ambiente en común. Se asumió que la evolución de cada objeto está dada por una ecuación en diferencias que depende, entre otras variables, de un ruido aleatorio $\xi_t$ con distribución desconocida. A partir de este hecho, y aplicando las teorías de procesos de control de Markov, control adaptado y control minimax, se propuso un modelo de control apropiado, y considerando que el número de objetos $N$ es muy grande ($N \sim \infty$) nos llevó a pensar en un esquema de la teoría de campo medio para proponer una solución al problema de control óptimo correspondiente.

Después de realizar un análisis del planteamiento del problema en general, nuestra investigación se centró en el estudio de los siguientes problemas:

**1.-** Problema de estimación y control en sistemas de interacción de objetos bajo en esquema de campo medio.

**2.-** Problema de control minimax en sistemas de interacción de objetos bajo un esquema de campo medio.

Como resultado de nuestra investigación se logró demostrar la existencia de políticas eventualmente-asintóticamente óptimas para el Problema 1 y políticas minimax para el Problema 2, dando lugar al desarrollo de los artículos [21, 22]. El artículo [21] se encuentra aceptado para su publicación, mientras que [22] esta sometido.

La diferencia escencial entre estos dos problemas está en las hipótesis respecto al proceso de perturbaciones aleatorias $\{\xi_t\}$. En el primer caso $\{\xi_t\}$ es una sucesión de variables aleatorias observables independientes e idénticamente distribuidas, mientras que el segundo solo asume independencia. Dependiendo de la situación a modelar uno puede determinar la teoría mas adecuada por aplicar: estimación y control o minimax. Por ejemplo, en problemas de medio ambiente, finanzas y economía el asumir que el ruido aleatorio es observable podría ser una hipótesis un poco realista, situación que no se presenta en muchos problemas de investigación de operaciones donde la variable aleatoria es observable.

Por otro lado, el hecho de haber supuesto que los objetos evolucionan de acuerdo a una ecuación en diferencias implica que los estados deben ser cantidades númericas que satisfacen la relación funcional correspondiente, lo cual restringe el potencial de aplicaciones. El caso mas general sería asumir que la dinámica es determinada por un kérnel de transición. Sin embargo, es importante observar que nuestros resultados son aplicables a este caso general siempre y cuando el kérnel sea completamente conocido o si depende de un parárametro desconocido, ya sea observable o no observable.

## 6.1. Problemas abiertos

- El criterio de optimalidad estudiado fue el de costo descontado. Usualmente, este criterio es aplicado en economía y finanzas. Sin embargo, en muchos problemas se necesita analizar el comportamiento asintótico del sistema. En estos casos, el criterio mas adecuado es el de costo promedio, cuyo estudio constituye un problema abierto. El reto matemático que se presenta en este caso es el de combinar la convergencia de los procesos del $N$-sistema al de campo medio, con el límite de los promedios y el de estimación.

- Una posible extensión de nuestro trabajo podría considerar el problema con dos controladores que toman decisiones para influir en el comportamiento de los objetos, y que tienen intereses opuestos. En este caso, el problema podría analizarce como un juego estocástico de suma cero, y en el contexto de ecuaciones en diferencia, considerar el caso en el que la distribución de las variables aleatorias es desconocida por uno o ambos controladores.

- En una dirección distinta a la anterior, pero dentro de la teor¿$\frac{1}{2}$ia de juegos, podemos plantear el problema donde cada objeto representa un jugador. Es decir, tendriamos un juego con una cantidad muy grande de jugadores, con un objetivo común u opuesto. Esto da lugar a los juegos de campo medio, área en la que aún existen muchos problemas abiertos.

- Como se puede observar, nuestro trabajo se desarrollo bajo las siguientes hipótesis: costos acotados, espacio de estados numerable o finito, condiciones de compacidad y de acotamiento (Lipschitz) en las funciones, entre otros. Por lo tanto, un problema abierto, de forma natural, es debilitar dichas hipótesis.

- Como podemos observar, una de las características principales de nuestro trabajo fue la de modelar el sistema de interacción de objetos como un problema de control óptimo markoviano. Por lo tanto, tomando en cuenta el amplio desarrollo de la teoría de procesos de control de Markov y lo joven de las aplicaciones de la teoría de campo medio a problemas de optimizacï¿½n, es posible proponer nuevos problemas. Por ejemplo, sistemas parcialmente observables, procesos a tiempo continuo, sistemas no estacionarios, desarrollo de nuevas teorías de control adaptado, etc., así como la combinación de estos.

# Bibliografía


[1] Achdou I, Capuzzo-Dolcetta I (2010) *Mean field games: numerical methods*, SIAM J. Numer. Anal., 48:1136-1162.

[2] Altman A, Hordijk A (1995) *Zero-sum Markov games and worst-case optimal control queueing systems, Queueing Systems*, Theory and appl 21:415-447.

[3] Aoki M (1994) *New macroeconomic modeling approaches*, Hierarchical dynamics and mean field approximation, J. Econ. Dyn. Control 18:865-877.

[4] Bensoussan A, Frehse J, Yam P (2010) *Mean Field Games and Mean Field Control Theory*, Springer, New York.

[5] Bertsekas DP (1987) *Dynamic Programming: Deterministic and Stochastic Models*, Prentice-Hall, Englewood Cliffs, N.J.

[6] Budhiraja A, Dupuis P, Fischer M (2012) *Large deviation properties of weakly interacting processes via weak convergence methods*, Annals Probab. 40:74-102.

[7] Caines PE, Malhamé RP, Nourian M (2006) *Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle*, Comunications in Informations and Systems 6:221-252.

[8] Caines PE, Malhamé RP, Nourian M (2013) *Nash, Social and Centralized to Consensus Problems via Mean Field Control Theory*, IEEE Transactions on Automatic Control 58:3.

[9] Carmona R, Fouque JP, Vestal D (2009) *Interacting particle systems for the computation of rare credit portfolio losses*, Finance Stoch 13:613-633.

[10] Coraluppi SP, Marcus SI (2000) *Mixed risk-neutral/minimax control of discrete-time finite state Markov decision processes*, IEEE Transactions on Automatic Control 45:528-532.

[11] Darabi-Sahnneh F, Scoglio C, Van-Mieghem P (2013) *Generalized Epidemic Mean-Field Model for Spreading Processes Over Multilayer Complex Networks*, IEEE Transactions on Networking 21:5.

[12] Gast N, Gaujal B (2011) *A mean field approach for optimization in discrete time*, Discrete Event Dyn. Syst. 21:63-101.

[13] Gast N, Gaujal B, Le Boudec JY (2012) *Mean field for Markov decision processes: from discrete to continuous optimization*, IEEE Trans. Autom. Control 57:2266-2280.

[14] Cross R, Grinfeld M, Lamba H (2006) *A mean field model of investor behaivor*, Journal of Physics: Conference Series 55.

[15] Gomes DA, Mohr J, Souza RR (2010) *Discrete time, finite state space mean field games*, J. Math. Pures Appl. 93:308-328.

[16] González-Trejo TJ, Hernández-Lerma O, Hoyos-Reyes LF(2003) *Minimax control of discrete-time stochastic systems* SIAM Jounal on Control and Optim. 41:1626-1659.

[17] Gordienko EI, Minjárez-Sosa JA (1998) *Adaptive control for discrete-time Markov processes with unbounded costs: discounted criterion*, Kybernetika 34:217-234.

[18] Guéant O, Lasry JM, Lions PL(2010)*Mean Field Games and Applications*,Paris-Princeton Lectures on Mathematical Finance 2010.

[19] Hernández-Lerma O (1989) *Adaptive Markov Control Processes*, Springer, New York.

[20] Hernández-Lerma O, Lasserre JB (1996) *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer, New York.

[21] Higuera-Chan CG, Jasso-Fuentes H, Minjárez-Sosa JA, *Discrete-Time Control for Systems of Interacting Objects with Unknown Random Disturbance Distributions: A Mean Field Approach*, Applied Math. Optim. DOI 10.1007/s00245-015-9312-6..

[22] Higuera-Chan CG, Jasso-Fuentes H, Minjárez-Sosa JA, *Control systems of interacting objects modeled as a game against nature under a mean field approach*,(sometido para su publicación).

[23] Hilgert N, Minjárez-Sosa JA (2006) *Adaptive control of stochastic systems with unknown disturbance distribution: discounted criterion*, Math. Meth. Oper. Res. 63:443-460.

[24] Horst U (1999) *Ergodic fluctuations in a stock market model with interacting agents: the mean field case.* Discussion paper No. 106, Sonderforschungbereich 373, Humboldt Universitat, Berlin.

[25] Huang M (2010) *Large-population LQG games involving a major player: the Nash certainty equivalence principle*, SIAM J. Control Optim. 48:3318-3353.

[26] Kolokoltsov VN, Troeva M, Yang W (2014) *On the rate of convergence for the mean-field approximation of controlled diffusions with large number of players*, Dyn. Games Appl. 4:208-230.

[27] M. Kurano (1987) *Minimax strategies for averange cost stochastic games with an application to inventory models.* J. Oper. Res. Soc. Jpn. 30:232-247.

[28] Lachapelle A, Salomon J, Turinici G (2010) *Computation of mean field equilibria in economics*, Math. Models Meth. Appl. Sci. 20:567-588.

[29] Lasry JM, Lions PL (2006) *Jeux a champ moyen i. le cas stationnaire*, C.R: Acad. Sci. Paris 343(9):619-625.

[30] Lasry JM, Lions PL (2006) *Jeux a champ moyen ii. horizon fini et controle optimal*, C.R: Acad. Sci. Paris 343(10):679-684.

[31] Lasry JM, Lions PL (2007) *Mean field games*, Jap. J. Math. 2:229-260.

[32] Le Boudec JY, McDonald D, Mundinger J (2007) *A generic mean field convergence result for systems of interacting objects* 4th Int. Conf. Quantitative Evaluation of Systems DOI 0-7695-2883-X/07

[33] López-Barrientos JD, Jasso-Fuentes H, (2015)*On the use of stochastic differential games against nature to ergodic control problems with unknown parameters*, Internat. J. Control 88:897-909.

[34] López-Barrientos JD, Jasso-Fuentes H, Escobedo Trujillo BA, (2015) *Discounted robust control for Markov diffusion processes*, TOP 23:53-76.

[35] Luque-Vásquez F, Minjárez-Sosa JA, Rosas-Rosas LC (2011) *Semi-Markov control models with partially known holding times distribution: Discounted and Average criteria*, Appl. Math. 114:135-156.

[36] Minjárez-Sosa JA, Vega-Amaya O (2009) *Asymptotically optimal strategies for adaptive zero-sum discounted Markov games*, SIAM J. Control Optim. 48:1405-1421.

[37] Peyrard N, Sabbadin R (2006) *Mean field approximation of the policy iteration algorithm for graph-based Markov decision processes, in: G. Biewka, S. Coradeschi, A. Perini, P. Traverso (Eds), Frontiers in Artificial Intelligence and Applications*, IOS Press, 595-599.

[38] Puterman ML (1994) *Markov Decision Processes*, Discrete Stochastic Dynamic Programming, John Wiley & Sons.

[39] Schield A (2008) *Robust optimal control for a consumption-investment problem*, Mathemathical Methods of Oper. Research 76:1-20.