SPRINGER BRIEFS IN PROBABILITY
AND MATHEMATICAL STATISTICS

J. Adolfo Minjárez-Sosa

# Zero-Sum Discrete-Time Markov Games with Unknown Disturbance Distribution

## Discounted and Average Criteria

Bernoulli Society
for Mathematical Statistics
and Probability

Springer

# SpringerBriefs in Probability and Mathematical Statistics

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic. Briefs are characterized by fast, global electronic dissemination, standard publishing contracts, standardized manuscript preparation and formatting guidelines, and expedited production schedules.

Typical topics might include:

- A timely report of state-of-the art techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- Lecture of seminar notes making a specialist topic accessible for non-specialist readers
- SpringerBriefs in Probability and Mathematical Statistics showcase topics of current relevance in the field of probability and mathematical statistics

Manuscripts presenting new results in a classical field, new field, or an emerging topic, or bridges between new results and already published works, are encouraged. This series is intended for mathematicians and other scientists with interest in probability and mathematical statistics. All volumes published in this series undergo a thorough refereeing process.

The SBPMS series is published under the auspices of the Bernoulli Society for Mathematical Statistics and Probability.

More information about this series at http://www.springer.com/series/14353

J. Adolfo Minjárez-Sosa

# Zero-Sum Discrete-Time Markov Games with Unknown Disturbance Distribution

## Discounted and Average Criteria

Springer

Bernoulli Society
for Mathematical Statistics
and Probability

J. Adolfo Minjárez-Sosa
Department of Mathematics
University of Sonora
Hermosillo, Sonora, Mexico

ISSN 2365-4333 ISSN 2365-4341 (electronic)
SpringerBriefs in Probability and Mathematical Statistics
ISBN 978-3-030-35719-1 ISBN 978-3-030-35720-7 (eBook)
https://doi.org/10.1007/978-3-030-35720-7

Mathematics Subject Classification: 91A15, 90C40, 62G05

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

*To my two women: Francisca and Camila*

# Preface

Discrete-time zero-sum Markov games constitute a class of stochastic games introduced by Shapley in [65] whose evolution over time can be described as follows. At each stage, players 1 and 2 observe the current state $x$ of the game and independently choose actions $a$ and $b$, respectively. Then, player 1 receives a payoff $r(x,a,b)$ from player 2 and the game moves to a new state $y$ in accordance with a transition probability or a transition function $F$ as in (1), below. The payoffs are accumulated throughout the evolution of the game in a finite or infinite horizon under a specific optimality criterion.

Even though there are now many studies in this field under multiple variants, it is mostly assumed that all components of the game are completely known by the players. However, the environment itself in which it evolves could make this assumption unrealistic or too strong. Hence, the availability of approximation and estimation algorithms that provide players with some insights on the evolution of the game is important, so that they can select their actions more accurately.

An important feature of this book is that it will deal with a class of Markov games with Borel state and action spaces, and possibly unbounded payoffs, under discounted and average criteria, whose state process $\{x_t\}$ evolves according to a stochastic difference equation of the form

$$x_{t+1} = F(x_t, a_t, b_t, \xi_t), \quad t = 0, 1, \ldots \tag{1}$$

Here, the pair $(a_t, b_t)$ represents the actions chosen by players 1 and 2, respectively, at time $t$, and $\{\xi_t\}$ is the disturbance process which is an observable sequence of independent and identically distributed random variables with *unknown* distribution $\theta$ for both players. In this scenario, our concern is in a game played over an infinite horizon evolving as follows. At stage $t$, once the players have observed the state $x_t$, and before choosing the actions $a_t$ and $b_t$, players 1 and 2 implement a statistical estimation process to obtain estimates $\theta_t^1$ and $\theta_t^2$ of $\theta$, respectively. Then, independently, the players adapt their decisions to such estimators to select actions $a = a_t(\theta_t^1)$ and $b = b_t(\theta_t^2)$. Next the game jumps to a new state according to the transition probability determined by Eq. (1) and the unknown distribution $\theta$, and the process is repeated over and over again.

This book is the first part of a project whose objective is to make a systematic analysis on recent developments in this kind of games. Specifically, in this first part we will provide the theoretical foundations on the procedures combining statistical estimation and control techniques for the construction of strategies of the players. We generically call this combination "estimation and control" procedures. The second part of the project will deal with another class of games models, as well as with approximation and computational aspects.

The statistical estimation process will be studied from two approaches. In the first one, we assume that the distribution $\theta$ has a density $\rho$ on $\Re^k$. In this case, there is a vast literature (see, e.g., [9–11, 27] and references therein) that provides different density estimation methods that might be easily adapted to the conditions imposed by the problem being analyzed. Among these we can mention kernel density estimation, $L_q$ estimation for $q \geq 1$, and projection estimation, through which it is possible to obtain several important properties such as the rate of convergence. The second approach is provided by the empirical distribution $\theta_t$ defined by the random disturbance process $\{\xi_t\}$. This method is very general in the sense that both the random variables $\xi_t$ and the distribution $\theta$ can be arbitrary. The price that must be paid due to this generality is that its applicability is restricted because it is necessary to impose stronger conditions than those of the previous case on the game model. Anyhow, the use of the empirical distribution has the additional advantage that it provides an approximation method of the value of the game and optimal strategies for players, in cases where the distribution $\theta$ is difficult to handle, by replacing $\theta$ with a simpler distribution given by $\theta_t$. In general terms, our approach to obtain estimation and control procedures for both discounted and average criteria consists of combining a statistical estimation method suitable for $\theta$ with game theory techniques. Our starting point is to, first, prove the existence of a value of the game as well as measurable minimizers/maximizers in the Shapley equation. To this end, some conditions are imposed on the game model which fall within the weighted-norm approach proposed by Wessels in [76] and then fully studied in [23, 24, 31] for Markov decision processes (MDPs) and recently for zero-sum stochastic games in [32, 40, 41, 44, 48]. Thereby, the estimation method is adapted to these conditions to obtain appropriate convergence properties.

Clearly, the good behavior of the strategies obtained through the estimation and control procedures depends on the accuracy of the estimation method, and even more on the optimality criterion with which their performance is measured. For instance, it is well known that the discounted criterion strongly depends on the decisions selected in the early stages of the game, just where the estimation process yields deficient information about the unknown distribution $\theta$. So, neither player

1 nor player 2 can generally ensure the existence of discounted optimal strategies. Hence the optimality under a discounted criterion is studied in an asymptotic sense. The notion of asymptotic optimality used in this book for Markov games was motivated from Schäl [67], who introduced this concept to study adaptive MDPs. In contrast, in view of the necessary asymptotic analysis in the study of the average criterion, the strategies obtained by means of estimation and control procedures turn out to be average optimal, providing suitable ergodicity conditions.

According to the historical development of the theories of stochastic control and Markov games, the problem of estimation and control for MDPs, also known as an adaptive Markov control problem, has received considerable attention in recent years (see, e.g., [2, 7, 22, 25, 26, 28, 29, 33–35, 52–55, 67] and references therein). In fact, even though approximation algorithms for stochastic games and games with partial information have been studied from several points of view (see, e.g., [8, 17, 20, 43, 46, 59, 60, 63], and references therein), in the field of statistical estimation and control procedures for Markov games the literature remains scarce; we can cite, for instance, [50, 56–58, 69, 70]. In particular, [56] deals with semi-Markov zero-sum games with unknown sojourn time distribution. The works [69, 70] study repeated games assuming that the transition law depends on an unknown parameter which is estimated by the maximum likelihood method, whereas [50, 56–58] deal with the theory developed in the context of this book.

The book is organized as follows. In Chap. 1 the class of Markov game models we deal with is introduced, together with the main elements necessary to define the game problem. Chapters 2 and 3 are devoted to analyze the discounted and the average criteria, respectively, where estimation and control procedures are presented under the assumption that the distribution $\theta$ has a density on $\Re^k$. Empirical estimation-approximation methods are given in Chap. 4. In this case, by using the empirical distribution to estimate $\theta$ both discounted and average criteria are analyzed. Finally, several examples of the class of Markov games studied throughout the book are given in Chap. 5. In this part we focus, mainly, on illustrating our assumptions on the game model, as well as on the numerical implementation of the estimation and control algorithm in specific examples.

**Acknowledgments.** The work of the author has been partially supported by Consejo Nacional de Ciencia y Tecnología (CONACYT-México) grant CB/2015-254306. Special thanks to Onésimo Hernández-Lerma who has read a draft version; his valuable comments and suggestions helped to improve the book. I also want to thank my colleagues Fernando Luque-Vásquez, Oscar Vega-Amaya, and Carmen Higuera-Chan, with whom I form the "controllers team" of the University of Sonora; undoubtedly, many of the ideas discussed in our long talks are included in this book. Finally, I deeply thank Donna Chernyk, Associate Editor at Springer, for her help.

Hermosillo, Mexico J. Adolfo Minjárez-Sosa
August 2019

# Contents

# Summary of Notation and Terminology

## Symbols and Abbreviations

| | |
|---|---|
| $\mathbb{N}$ | Set of positive integers |
| $\mathbb{N}_0$ | Set of nonnegative integers |
| $\Re$ | Set of real numbers |
| $\Re^+$ | Set of nonnegative real numbers |
| $1_D(\cdot)$ | Indicator function of the set $D$ |
| $:=$ | Equality by definition |
| a.e. | Almost everywhere |
| a.s. | Almost surely |
| i.i.d. | Independent and identically distributed |
| r.v. | Random variable |
| p.m. | Probability measure |
| l.s.c. | Lower semicontinuous |
| u.s.c. | Upper semicontinuous |

## Spaces of Functions

- The space $L_q = L_q(\Re^k)$, for $1 \leq q < \infty$, consists of all real-valued measurable functions on $\Re^k$ with finite $L_q$-norm:

$$\|\rho\|_{L_q} := \left( \int_{\Re^k} |\rho|^q \, d\mu \right)^{1/q}$$

  with respect to the Lebesgue measure $\mu$.
- A Borel space is a Borel subset of a complete separable metric space.

For a Borel space $X$, we use the following notation:

| | |
|---|---|
| $\mathscr{B}(X)$ | Borel $\sigma$-algebra in $X$, and "measurable," for either sets or functions, means "Borel measurable." |
| $\mathbb{B}(X)$ | Space of real-valued bounded measurable functions on $X$ with the supremum norm: $\|v\|_B := \sup_{x \in X} |v(x)|$ . |
| $\mathbb{C}(X) \subset \mathbb{B}(X)$ | Subspace of bounded continuous functions. |
| $\mathbb{L}(X)$ | Space of lower semicontinuous functions and bounded from below. |
| $\mathbb{B}_W(X)$ | For a function $W : X \to [1, \infty)$, space of measurable functions with finite weighted norm ($W$-norm): $\|v\|_W := \sup_{x \in X} \dfrac{|v(x)|}{W(x)}$. |
| $\mathbb{C}_W(X) \subset \mathbb{B}_W(X)$ | Subspace of $W$-bounded continuous functions. |
| $\mathbb{L}_W(X) \subset \mathbb{B}_W(X)$ | Subspace of $W$-bounded lower semicontinuous functions. |
| $\mathbb{P}(X)$ | Space of probability measures on $X$ endowed with the weak topology (see Appendix B). |
| $\mathbb{P}(X|Y)$ | Family of stochastic kernels on $X$ given $Y$, where $X$ and $Y$ are Borel spaces. |