CAMBRIDGE

Cambridge University Press & Assessment
978-1-316-51633-1 — A Course in Stochastic Game Theory
Eilon Solan
Excerpt
More Information

Introduction

Stochastic games are a mathematical model that is used to study dynamic interactions among agents who influence the evolution of the environment. These games were first presented and studied by Lloyd Shapley (1953).[1,2] Since Shapley's seminal work, the literature on stochastic games expanded considerably, and the model was applied to numerous areas, such as arms race, fishery wars, and taxation.

A stochastic game is played in discrete time by a finite set I of players, and it consists of a finite number of states. In each state s, each player $i \in I$ has a given set of actions, denoted $A^i(s)$. In every stage $t \in \mathbb{N}$, the play is in one of the states, denoted s_t. Each player $i \in I$ chooses an action $a_t^i \in A^i(s_t)$ that is available to her at the current stage, receives a stage payoff, which depends on the current state s_t as well as on the actions $(a_t^j)_{j \in I}$ chosen by the players, and a new state s_{t+1} is chosen, according to a probability distribution that depends on the current state and on the actions of the players $(a_t^j)_{j \in I}$.

In a stochastic game, the players have two, seemingly contradicting, goals. First, they need to ensure that their future opportunities remain high. At the same time, they should make sure that their stage payoff is also high. This dichotomy makes the analysis of stochastic games intriguing and not trivial.

The study of stochastic games uses tools from many mathematical branches, such as probability, analysis, algebra, differential equations, and combinatorics. The goal of this book is to present the theory through the mathematical techniques that it employs. Thus, each chapter presents mathematical results

[1] Lloyd Stowell Shapley (Cambridge, Massachusetts, June 2, 1923 – Tucson, Arizona, March 12, 2016) was an American mathematician who made many influential contributions to Game Theory, like the Shapley value, stochastic games, and the defer-acceptance algorithm for stable marriages. Shapley shared the 2012 Nobel Prize in Economics together with game theorist Alvin Roth.

[2] All commentary is taken from Wikipedia.

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-1-316-51633-1 — A Course in Stochastic Game Theory
Eilon Solan
Excerpt
More Information

from some branch of mathematics, and uses them to prove results on stochastic games. The goal is not to prove the most general theorems in stochastic games, but rather to present the beauty of the theory. Accordingly, we sometimes restrict the scope of the results that is proven, to allow for simpler proofs that bypass technical difficulties.

The material in this book is summarized by the following table:

Chapter	Tool + Result
1	Contracting mappings Stationary optimal strategies in Markov decision problems
2	Tauberian Theorem Uniform ϵ-optimality in hidden Markov decision problems
5	Contracting mappings Stationary discounted optimal strategies in zero-sum stochastic games
6	Semi-algebraic mappings Existence of the limit of the discounted value
7	B-graphs Continuity of the limit of the discounted value
8	Kakutani's fixed point theorem Stationary discounted equilibria in multiplayer stochastic games
9	Existence of the uniform value in zero-sum stochastic games
10	The vanishing discount factor approach Existence of uniform equilibrium in absorbing games
11	Ramsey's Theorem Existence of undiscounted equilibrium in two-player deterministic stopping games
12	Approximating infinite orbits Existence of undiscounted equilibrium in multiplayer quitting games
13	Linear complementarity problems Existence of undiscounted equilibrium in multiplayer quitting games

Each chapter contains exercises. Solutions are available as supplementary material on the book's page on the publisher's website. The book is based on a graduate level course that I taught at Tel Aviv University for more than

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-316-51633-1 — A Course in Stochastic Game Theory
Eilon Solan
Excerpt
More Information

a decade. I hope that the readers, as my students, will like the diversity of the topics and the elegance of the proofs. For the benefit of readers who would like to expand their knowledge in stochastic games, I added references to related results at the end of each chapter. Books and surveys that include material on different aspects of stochastic games include Raghavan et al. (1991), Raghavan and Filar (1991), Filar and Vrieze (1997), Başar and Olsder (1998), Mertens (2002), Vieille (2002), Neyman and Sorin (2003), Solan (2008), Chatterjee et al. (2009, 2013), Chatterjee and Henzinger (2012), Laraki and Sorin (2015), Mertens et al. (2015), Solan and Vieille (2015), Solan and Ziliotto (2016), Başar and Zaccour (2017), Jaśkiewicz and Nowak (2018a,b), and Renault (2019).

I end the introduction by thanking Ayala Mashiah-Yaakovi, who read the manuscript and the solution manual and made many comments that improved the text; Andrei Iacob, who copyedited the text; and John Yehuda Levy, Andrzej Nowak, Robert Simon, Bernhard von Stengel, Uri Zwick, and my students throughout the years for providing comments and spotting typos.

Notation

The set of positive integers is

$$\mathbb{N} := \{1, 2, 3, \ldots\}.$$

The number of elements in a finite set K is denoted by $|K|$. For every finite set K, the set of probability distributions over K is denoted by $\Delta(K)$. We identify each element $k \in K$ with the probability distribution in $\Delta(K)$ that assigns probability 1 to k. For a probability distribution $\mu \in \Delta(K)$, the *support* of μ, denoted $\mathrm{supp}(\mu)$, is the set of all elements $k \in K$ that have positive probability under μ:

$$\mathrm{supp}(\mu) := \{k \in K : \mu[k] > 0\}.$$

A probability distribution is *pure* if $\mathrm{supp}(\mu)$ contains only one element: $|\mathrm{supp}(\mu)| = 1$.

Let I be a finite set, and, for each $i \in I$, let A^i be a set. We denote by $A^I := \prod_{i \in I} A^i$ the cartesian product, and denote $A^{-i} := \prod_{j \in I \setminus \{i\}} A^j$. Similarly, if $a = (a^i)_{i \in I} \in A^I$, we denote by $a^{-i} := (a^j)_{j \in I \setminus \{i\}} \in A^{-i}$ the vector a with its i'th coordinate removed.

We will use two norms, the L_1-*norm* and the L_∞-*norm* (or the maximum norm). For a vector $x \in \mathbb{R}^n$, we define

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-1-316-51633-1 — A Course in Stochastic Game Theory
Eilon Solan
Excerpt
More Information

$$\|x\|_1 := \sum_{i=1}^{n} |x_i|,$$

and

$$\|x\|_\infty := \max_{i=1,\dots,n} |x_i|.$$

For a function $f : X \to \mathbb{R}$, $\text{argmax}_{x \in X} f(x)$ is the set of all points in X that maximize f:

$$\text{argmax}_{x \in X} f(x) := \left\{ y \in X : f(y) = \max_{x \in X} f(x) \right\}.$$

When the set X is compact and the function f is continuous, the set $\text{argmax}_{x \in X} f(x)$ is non-empty.

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-1-316-51633-1 — A Course in Stochastic Game Theory
Eilon Solan
Excerpt
More Information

1

Markov Decision Problems

In this chapter, we introduce Markov decision problems, which are stochastic games with a single player. They serve as an appetizer. On the one hand, the basic concepts and basic proofs for zero-sum stochastic games are better understood in this simple model. On the other hand, some of the conclusions that we draw for Markov decision problems are different from those drawn for zero-sum stochastic games. This illustrates the inherent difference between single-player decision problems and multiplayer decision problems (=games). The interested reader is referred to, for example, Ross (1982) or Puterman (1994) for an exposition of Markov decision problems.

We will study both the T-stage evaluation and the discounted evaluation. We will introduce and study contracting mappings,[1] and will use such mappings to show that the decision maker has a stationary discounted optimal strategy. We will also define the concept of uniform optimality, and show that the decision maker has a stationary uniformly optimal strategy.

Definition 1.1 A *Markov decision problem*[2] is a vector $\Gamma = \langle S, (A(s))_{s \in S}, q, r \rangle$ where

- S is a finite set of states.
- For each $s \in S$, $A(s)$ is a finite set of actions available at state s. The set of pairs (state, action) is denoted by

$$SA := \{(s,a) \colon s \in S, a \in A(s)\}.$$

- $q \colon SA \to \Delta(S)$ is a transition rule.
- $r \colon SA \to \mathbf{R}$ is a payoff function.

[1] We adhere to the convention that a mapping is a function whose range is a general space or $\mathbb{R}^n$, while a function is always real-valued.

[2] Andrey Andreyevich Markov (Ryazan, Russia, June 14, 1856 – St. Petersburg, Russia, July 20, 1922) was a Russian mathematician. He is best known for his work on the theory of stochastic processes that now bear his name: Markov chains and Markov processes.

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-1-316-51633-1 — A Course in Stochastic Game Theory
Eilon Solan
Excerpt
More Information

A Markov decision problem involves a decision maker, and it evolves as follows. The problem lasts for infinitely many stages. The initial state $s_1 \in S$ is given. At each stage $t \geq 1$, the following happens:

- The current state s_t is announced to the decision maker.
- The decision maker chooses an action $a_t \in A(s_t)$ and receives the stage payoff $r(s_t, a_t)$.
- A new state s_{t+1} is drawn according to $q(\cdot \mid s_t, a_t)$, and the game proceeds to stage $t+1$.

Example 1.2 Consider the following situation. The technological level of a country can be High (H), Medium (M), or Low (L). The annual investment of the country in technological advances can also be high (2 billion dollars), medium (1 billion dollars), or low (0.5 billion dollars). The annual gain from technological level is increasing: the high, medium, and low technological level yield 10, 6, and 2 billion dollars, respectively. The technological level changes stochastically as a function of the investment in technological advancement, according to the following table:[3]

Technology level	High investment	Medium investment	Low investment
H	H	$\left[\frac{1}{2}(H), \frac{1}{2}(M)\right]$	$\left[\frac{1}{4}(H), \frac{3}{4}(M)\right]$
M	$\left[\frac{3}{5}(H), \frac{2}{5}(M)\right]$	M	$\left[\frac{2}{5}(M), \frac{3}{5}(L)\right]$
L	$\left[\frac{3}{5}(M), \frac{2}{5}(L)\right]$	$\left[\frac{2}{5}(M), \frac{3}{5}(L)\right]$	L

The situation can be presented as a Markov decision problem as follows:

- There are three states, which represent the three technological levels: $S = \{H, M, L\}$.
- There are three actions in each state, which represent the three investment levels: $A(s) = \{h, m, l\}$ for each $s \in S$.
- The transition rule is given by

[3] Here and in the sequel, a probability distribution is denoted by a list of probabilities and outcomes in square brackets, where the outcomes are written within round brackets. Thus, $\left[\frac{2}{3}(H), \frac{1}{3}(M)\right]$ means a probability distribution that assigns probability $\frac{2}{3}$ to H and probability $\frac{1}{3}$ to M.

© in this web service Cambridge University Press & Assessment www.cambridge.org

$$q(H \mid H,h) = 1, \quad q(M \mid H,h) = 0, \quad q(L \mid H,h) = 0,$$
$$q(H \mid H,m) = \tfrac{1}{2}, \quad q(M \mid H,m) = \tfrac{1}{2}, \quad q(L \mid H,m) = 0,$$
$$q(H \mid H,l) = \tfrac{1}{4}, \quad q(M \mid H,l) = \tfrac{3}{4}, \quad q(L \mid H,l) = 0,$$
$$q(H \mid M,h) = \tfrac{3}{5}, \quad q(M \mid M,h) = \tfrac{2}{5}, \quad q(L \mid M,h) = 0,$$
$$q(H \mid M,m) = 0, \quad q(M \mid M,m) = 1, \quad q(L \mid M,m) = 0,$$
$$q(H \mid M,l) = 0, \quad q(M \mid M,l) = \tfrac{2}{5}, \quad q(L \mid M,l) = \tfrac{3}{5},$$
$$q(H \mid L,h) = 0, \quad q(M \mid L,h) = \tfrac{3}{5}, \quad q(L \mid L,h) = \tfrac{2}{5},$$
$$q(H \mid L,m) = 0, \quad q(M \mid L,m) = \tfrac{2}{5}, \quad q(L \mid L,m) = \tfrac{3}{5},$$
$$q(H \mid L,l) = 0, \quad q(M \mid L,l) = 0, \quad q(L \mid L,l) = 1.$$

- The payoff function (in billions of dollars) is given by

$$r(H,h) = 8, \quad r(H,m) = 9, \quad r(H,l) = 9\tfrac{1}{2},$$
$$r(M,h) = 4, \quad r(M,m) = 5, \quad r(M,l) = 5\tfrac{1}{2},$$
$$r(L,h) = 0, \quad r(L,m) = 1, \quad r(L,l) = 1\tfrac{1}{2}. \qquad \blacklozenge$$

Example 1.3 The Markov decision problem that is illustrated in Figure 1.1 is formally defined as follows:

- There are three states: $S = \{s(1), s(2), s(3)\}$.
- In state $s(1)$, there are two actions: $A(s(1)) = \{U, D\}$; in states $s(2)$ and $s(3)$, there is one action: $A(s(2)) = A(s(3)) = \{D\}$.
- Payoffs appear at the center of each entry and are given by:

$$r(s(1), U) = 10; \quad r(s(1), D) = 5; \quad r(s(2), D) = 10; \quad r(s(3), D) = -100.$$

- Transitions appear in parentheses next to the payoff and are given by:
  - If in state $s(1)$ the decision maker chooses U, the process moves to state $s(2)$, that is, $q(s(2) \mid s(1), U) = 1$.
  - If in state $s(1)$ the decision maker chooses D, the process remains in state $s(1)$, that is, $q(s(1) \mid s(1), D) = 1$.

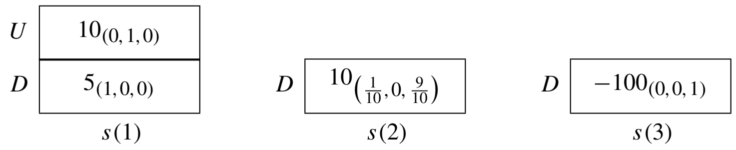

Figure 1.1 The Markov decision problem in Example 1.3.

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-1-316-51633-1 — A Course in Stochastic Game Theory
Eilon Solan
Excerpt
More Information

- From state $s(2)$, the process moves to state $s(1)$ with probability $\frac{1}{10}$ and to state $s(3)$ with probability $\frac{9}{10}$, that is, $q(s(1) \mid s(2), D) = \frac{1}{10}$ and $q(s(3) \mid s(2), D) = \frac{9}{10}$.
- Once the process reaches state $s(3)$, it stays there, that is, $q(s(3) \mid s(3), D) = 1$. ♦

1.1 On Histories

For $t \in \mathbb{N}$, the set of *histories of length t* is defined by

$$H_t := (SA)^{t-1} \times S,$$

where by convention $(SA)^0 = \emptyset$. This is the set of all histories that may occur until stage t. A typical element in H_t is denoted by h_t. The last state of history h_t is denoted by s_t. The set H_1 is identified with the state space S, and the history (s_1) is simply denoted by s_1.

We denote the set of all *histories* by

$$H := \bigcup_{t \in \mathbb{N}} H_t,$$

and the set of all *infinite histories* or *plays* by

$$H_\infty := (SA)^{\mathbb{N}}.$$

The set of plays H_∞ is a measurable space, with the sigma-algebra generated by the cylinder sets, which are defined as follows. For a history $\widetilde{h}_t = (\widetilde{s}_1, \widetilde{a}_1, \dots, \widetilde{s}_t) \in H_t$, the *cylinder set* $C(\widetilde{h}_t) \subset H_\infty$ is the collection of all plays that start with $\widetilde{h}_t$, that is,

$$C(\widetilde{h}_t) := \{h = (s_1, a_1, s_2, a_2, \dots) \in H_\infty \colon s_1 = \widetilde{s}_1, a_1 = \widetilde{a}_1, \dots, s_t = \widetilde{s}_t\}.$$

For every $t \in \mathbb{N}$, the collection of all cylinder sets $(C(\widetilde{h}_t))_{\widetilde{h}_t \in H_t}$ defines a finite partition, or an algebra, on H_∞. We denote by $\mathcal{H}_t$ this algebra and by $\mathcal{H}$ the sigma-algebra on H_∞ generated by the algebras $(\mathcal{H}_t)_{t \in \mathbb{N}}$.

1.2 On Strategies

A *mixed action* at state s is a probability distribution over the set of actions $A(s)$ available at state s. The set of mixed actions at state s is therefore $\Delta(A(s))$. A strategy of the decision maker specifies how the decision maker should play after each possible history.

Definition 1.4 A *strategy* is a mapping σ that assigns to each history $h = (s_1, a_1, \dots, a_{t-1}, s_t)$ a mixed action in $\Delta(A(s_t))$.

The set of all strategies is denoted by Σ.

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-316-51633-1 — A Course in Stochastic Game Theory
Eilon Solan
Excerpt
More Information

A decision maker who follows a strategy σ behaves as follows: at each stage t, given the past history $(s_1, a_1, \dots, s_t)$, the decision maker chooses an action a_t according to the mixed action $\sigma(\cdot \mid s_1, a_1, \dots, s_t)$.

Comment 1.5 A strategy as defined in Definition 1.4 is termed in the literature *behavior strategy*.

Comment 1.6 The fact that the choice of the decision maker depends on past play implicitly assumes that the decision maker knows the past play; that is, the decision maker observes (and remembers) all past states that the process visited, and she remembers all her past choices. In Chapter 2, we will study the model of Markov decision problems when the decision maker does not observe the state.

Comment 1.7 A strategy contains a lot of irrelevant information. Indeed, when the initial state is $s_1 = s$, it is not important what the decision maker would play if the initial state were $s' \neq s$. Similarly, if in the first stage the decision maker played the action $a_1 = a$, it is irrelevant what she would play in the second stage if she played the action $a' \neq a$ in the first stage. We nevertheless regard a strategy as a mapping defined on the set of *all* histories, because of the simplicity of the definition; otherwise we would have to define for every strategy σ and every positive integer t the set of all histories of length t that can occur with positive probability when the decision maker follows strategy σ (which depend on the definition of σ up to stage $t-1$), and define σ at stage t only for those histories.

Every strategy σ, together with the initial state s_1, defines a probability distribution $\mathbf{P}_{s_1,\sigma}$ on the space of measurable space $(H_\infty, \mathcal{H})$. To define this probability distribution formally, we define it on the collection of cylinder sets that generate $(H_\infty, \mathcal{H})$ by the rule

$$\mathbf{P}_{s_1,\sigma}(C(\widetilde{s}_1, \widetilde{a}_1, \dots, \widetilde{s}_{t-1}, \widetilde{a}_{t-1}, \widetilde{s}_t)) \tag{1.1}$$

$$:= \mathbf{1}_{\{s_1 = \widetilde{s}_1\}} \cdot \prod_{k=1}^{t-1} \sigma(\widetilde{a}_k \mid \widetilde{s}_1, \widetilde{a}_1, \dots, \widetilde{s}_1) \cdot \prod_{k=1}^{t-1} q(\widetilde{s}_{k+1} \mid \widetilde{s}_k, \widetilde{a}_k).$$

Let $\mathbf{P}_{s_1,\sigma}$ be the unique probability distribution on H_∞ that agrees with this definition on cylinder sets. The fact that, in this way, we indeed obtain a unique probability distribution is guaranteed by the Carathéodory[4] Extension Theorem (see, e.g., theorem 3.1 in Billingsley (1995)).

[4] Constantin Carathéodory (Berlin, Germany, September 13, 1873 – Munich, Germany, February 2, 1950) was a Greek mathematician who spent most of his career in Germany. He made significant contributions to the theory of functions of a real variable, the calculus of variations, and measure theory. His work also includes important results in conformal representations and in the theory of boundary correspondence.

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-316-51633-1 — A Course in Stochastic Game Theory
Eilon Solan
Excerpt
More Information

Two simple classes of strategies are pure strategies that involve no randomization, and stationary strategies that depend only on the current state and not on the whole past history.

Definition 1.8 A strategy σ is *pure* if $|\text{supp}(\sigma(h_t))| = 1$ for every history $h_t \in H$.

The set of pure strategies is denoted by Σ_{P}.

Definition 1.9 A strategy σ is *stationary* if, for every two histories $h_t = (s_1, a_1, s_2, \ldots, a_{t-1}, s_t)$ and $\widehat{h}_k = (\widehat{s}_1, \widehat{a}_1, \widehat{s}_2, \ldots, \widehat{a}_{k-1}, \widehat{s}_k)$ that satisfy $s_t = \widehat{s}_k$, we have $\sigma(h_t) = \sigma(\widehat{h}_k)$.

The set of stationary strategies is denoted by Σ_{S}.

A pure stationary strategy assigns to each state $s \in S$ an action in $A(s)$. Since the number of actions in $A(s)$ is $|A(s)|$, we can express the number of pure stationary strategies in terms of the data of the Markov decision problem.

Theorem 1.10 *The number of pure stationary strategies is* $\prod_{s \in S} |A(s)|$.

One can identify a stationary strategy σ with a vector $x \in \prod_{s \in S} \Delta(A(s))$. With this identification, $x(s)$ is the mixed action chosen when the current state is s. Thus, the set of stationary strategies Σ_{S} can be identified with the space $X := \prod_{s \in S} \Delta(A(s))$, which is convex and compact. For every element $x \in X$, the stationary strategy that corresponds to x is still denoted by x.

In Definition 1.4 we defined a strategy to be a mapping from histories to mixed actions. We now present another concept of a strategy that involves randomization – a mixed strategy.

Definition 1.11 A *mixed strategy* is a probability distribution over the set Σ_{P} of pure strategies.

Every strategy is equivalent to a mixed strategy. Indeed, a strategy σ is defined by $\aleph_0$ lotteries: to each history $h_t \in H$, it assigns a lottery $\sigma(h_t) \in \Delta(A(s_t))$. If the decision maker performs all the $\aleph_0$ lotteries before the play starts, then the realizations of the lotteries define a pure strategy. In particular, the strategy defines a probability distribution over the set of pure strategies.

Conversely, every mixed strategy is equivalent to a strategy. Indeed, given a mixed strategy τ, one can calculate for each history h_t the conditional probability $\sigma(a_t \mid h_t)$ that the action chosen after h_t is $a_t \in A(s_t)$. If the history h_t occurs with probability 0 under $\mathbf{P}_{s_1,\sigma}$, we set $\sigma(a_t \mid h_t)$ arbitrarily. One can show that the strategy σ is equivalent to the mixed strategy τ.

© in this web service Cambridge University Press & Assessment www.cambridge.org