STOCHASTIC GAMES

Ulrich Rieder
University of Ulm, Germany

Keywords: minimax theorem, optimal strategies, Nash-equilibria, zero-sum games, general-sum games, discounted and average reward games, mathematical programming.

Contents

1. Introduction
2. Basic Definitions and Notations
3. Zero-Sum Stochastic Games
3.1 Discounted Stochastic Games
3.2 Average Reward Stochastic Games
4. General-Sum Stochastic Games
4.1 Discounted Stochastic Games
4.2 Average Reward Stochastic Games
5. Further Topics
Glossary
Bibliography
Biographical Sketch

Summary

Stochastic games are a generalization of Markov decision processes to the case of two or more controllers. In this paper we discuss the main existence results on optimality and equilibria in two-person stochastic games with finite state and action spaces. Moreover, we present some algorithms for computing optimal strategies. Stochastic games are classified in discounted or limiting average reward problems and zero-sum or general-sum games. In each section, we provide several examples to illustrate the most important phenomena.

1. Introduction

Stochastic games are a generalization of *Markov Decision Processes* to the case of two or more controllers. Sometimes these games are therefore also called Markov games.

Surprisingly, stochastic games preceded Markov decision processes and in the early years, the two theories evolved independently and yet in a somewhat parallel fashion.

Here we restrict to two-person stochastic games. As for Markov decision processes the players are allowed to choose actions according to rules and depending on their information. The objective is to select the action in such a way as to maximize a performance criterion. Stochastic games can be classified in

- Continuous-time or discrete-time games
- Discounted or average reward problems

- Zero-sum or general-sum games

In this exposition, we will restrict to discrete-time games with finite state and action spaces. In Section 3, we discuss the main existence results and algorithms for zero-sum stochastic games. In Section 4, we focus on general-sum stochastic games. In each section we discuss several examples to illustrate the most important phenomena. (For a detailed discussion of general game theoretic aspects, see *Game Theory*).

2. Basic Definitions and Notations

In 1928, von Neumann proved the so-called *minimax theorem* which says that for each finite (m, n)-matrix $A = (a_{ij})$ there exist probability vectors $x_0 \in \mathbb{R}^m$ and $y_0 \in \mathbb{R}^n$ such that for all x and y

$$x^T Ay_0 \leq x_0^T Ay_0 \leq x_0^T Ay. \tag{1}$$

In other words:

$$\max_x \min_y x^T Ay = \min_y \max_x x^T Ay. \tag{2}$$

This statement can be interpreted in a way that each *matrix game* has a value. A matrix game A is played as follows. Simultaneously, and independent from each other, player 1 chooses a row i and player 2 chooses a column j of A. Then player 2 has to pay the amount a_{ij} to player 1. Each player is allowed to randomize over his available actions and we assume that player 1 wants to maximize his expected payoff, while player 2 wants to minimize the expected payoff to player 1. The minimax theorem tells us that, for each matrix A there is a unique amount val(A), which player 1 can guarantee as his minimal expected payoff, while at the same time player 2 can guarantee that the expected payoff to player 1 will be at most this amount.

Later Nash (1951) considered the n-person extension of matrix games, in the sense that all n players simultaneously and independently choose actions that determine a payoff for each and every one of them. Nash showed that in such games there always exists at least one *Nash-equilibrium*: a set of strategies such that each player is playing a best reply against the joint strategy of his opponents. For the two-player case this boils down to a *bimatrix game* where player 1 and 2 receive a_{ij} and b_{ij} respectively in case their choices determine entry (i, j). The result of Nash says that there exist probability vectors $x_0 \in \mathbb{R}^m$ and $y_0 \in \mathbb{R}^n$ such that for all x and y

$$x_0^T Ay_0 \geq x^T Ay_0 \quad \text{and} \quad x_0^T By_0 \geq x_0^T By \tag{3}$$

where $A = (a_{ij})$ and $B = (b_{ij})$ are finite (m, n)-matrices. Then (x_0, y_0) is called a Nash-equilibrium.

In 1953, Shapley introduced dynamics into game theory by considering the situation that

at discrete stages the players play one of finitely many matrix games, where the choices of the players determine a payoff to player 1 (by player 2) as well as a stochastic transition to go to a next matrix game. He called these games *stochastic games*, which brings us to the topic of this paper.
Formally, a two-person stochastic game with finite state and action spaces can be represented by the data

$$(S, A, B, p, r_1, r_2, \beta) \tag{4}$$

with the following meaning:

1. S is the finite *state space*.
2. $A(s)$ and $B(s)$ are the finite *sets of admissible actions* in state s for player 1 and player 2, respectively.
3. p *is* a *transition probability*, i.e. $p(s, a, b, z)$ is the (conditional) probability that the next state will be $z \in S$ given state $s \in S$ and actions $a \in A(s)$ and $b \in B(s)$.
4. r_k is the *reward function* for each player k, $k \in \{1, 2\}$, i.e. $r_k(s, a, b)$ is the payoff for player k in state $s \in S$ whenever actions $a \in A(s)$ and $b \in B(s)$ are selected.
5. $ß \in (0, 1)$ is the *discount factor*.

The stochastic game can start in an arbitrary state of S. By choosing actions $a_n \in A(s_n)$ and $b_n \in B(s_n)$ independently (where s_n denotes the state visited at stage n), the players receive the rewards $r_1(s_n, a_n, b_n)$ and $r_2(s_n, a_n, b_n)$ respectively and determine the transition probability $p(s_n, a_n, b_n, z)$ of visiting the next state z. In case

$$r_1(s,a,b) + r_2(s,a,b) = 0 \quad \text{for all} \quad s \in S,\, a \in A(s),\, b \in B(s) \tag{5}$$

the game is called *zero-sum*, otherwise it is called *general-sum*. In zero-sum games players have strictly opposite interests, since they are paying each other.

A *strategy* for a player is a rule that tells him for any history $h_n = (s_0, a_0, b_0, s_1, ..., s_{n-1}, a_{n-1}, b_{n-1}, s_n)$ up to stage n, what randomized action to use in state s_n at stage n. Such strategies will be denoted by π for player 1 and σ for player 2. For initial state s and any pair of strategies (π, σ) the *limiting average reward* and the *ß-discounted reward* to player $k \in \{1, 2\}$ are respectively given by

$$G_{\pi\sigma}^k(s) := \liminf_{T\to\infty} \frac{1}{T} \mathbb{E}_{s\pi\sigma}\left[\sum_{n=0}^{T-1} r_k(X_n, A_n, B_n)\right] \tag{6}$$

$$V_{\pi\sigma}^k(s) := \mathbb{E}_{s\pi\sigma}\left[\sum_{n=0}^{\infty} \beta^n r_k(X_n, A_n, B_n)\right] \tag{7}$$

where X_n, A_n, B_n are random variables for the state and actions at stage n.

A *stationary strategy* for a player consists of a randomized action for each state, to be used whenever that state is being visited, regardless of the history. Stationary strategies for player 1 are denoted by $f: S \to \mathbb{P}(A)$ where $f(s)$ is the randomized action to be used in state $s \in S$. More formally, $f(s) \in \mathbb{P}(A(s))$ and $\mathbb{P}(A(s))$ is the set of all probability vectors on $A(s)$. For player 2's stationary strategies we write g. A pair (f, g) of stationary strategies determines a Markov chain with transition matrix $P(f, g)$ on S, where entry (s, z) of $P(f, g)$ is given by

$$p_{fg}(s,z) = \sum_{a\in A(s)} \sum_{b\in B(s)} f(s,a)g(s,b)p(s,a,b,z). \tag{8}$$

Moreover, we use the notation

$$r_{fg}^k(s) = \sum_{a\in A(s)} \sum_{b\in B(s)} f(s,a)g(s,b)r_k(s,a,b) \tag{9}$$

for the sth component of the vector $r_k(f, g)$, $k \in \{1, 2\}$.

Then we have

$$V_{fg}^k = \left(I - \beta P(f,g)\right)^{-1} r_k(f,g) \tag{10}$$

where I is the identity matrix and

$$G_{fg}^k = P^*(f,g) r_k(f,g) \tag{11}$$

With

$$P^*(f,g) = \lim_{T\to\infty} \frac{1}{T} \sum_{n=0}^{T-1} P^n(f,g) \tag{12}$$

$P^*(f, g)$ is the Césaro-limit of the transition matrices (see *Markov Models*), and it is well-known that

$$P^*(f,g) = P^*(f,g)P(f,g) \tag{13}$$

$$P^*(f,g) = \lim_{\beta\uparrow 1}(1-\beta)\left(1-\beta P(f,g)\right)^{-1}. \tag{14}$$

Hence, we obtain from (10) and (11)

$$G_{fg}^k = \lim_{\beta\uparrow 1}(1-\beta)V_{fg}^k(\beta). \tag{15}$$

Note that $V_{fg}^k(\beta) = V_{fg}^k$ depends on the discount factor *ß*. Finally, we mention one more type of strategies, namely *Markov strategies*. These are strategies which, at any stage of play, prescribe actions that only depend on the current state and stage. Thus, the past actions of the opponent are not taken into account.

3. Zero-Sum Stochastic Games

In zero-sum stochastic games, it is customary to consider only the payoffs to player 1. Then we write

$$r(s,a,b) := r_1(s,a,b) = -r_2(s,a,b). \tag{16}$$

In the first part of this section, we present the main results and algorithms for discounted games, while in the second part we consider average reward stochastic games.

-
-
-

Bibliography

Filar J. and Vrieze K. (1997). *Competitive Markov Decision Processes*, 393 pp. New York: Springer-Verlag. [This textbook presents the main topics for stochastic games.]

Heyman D.P. and Sobel M. (1984). *Stochastic Models in Operations Research, II.* New York: McGraw-Hill. [Chapter 9 provides early results and references.]

Mertens J.F. (1992). Stochastic Games. In R.J. Aumann and S. Hart (eds.): *Handbook of Game Theory with Economic Applications*. Amsterdam: North-Holland. [Recent survey on stochastic games.]

Thuijsman F. (1992). *Optimality and Equilibria in Stochastic Games. CWI-Tract* **82.** Amsterdam: Center of Mathematics and Computer Science. [Recent thesis on the main existence results for stochastic games.]

Biographical Sketch

Ulrich Rieder, born in 1945, received the Ph.D. degree in mathematics in 1972 (University of Hamburg) and the Habilitation in 1979 (University of Karlsruhe). Since 1980, he has been Full Professor of Mathematics and head of the Department of Optimization and Operations Research at the University of Ulm. His research interests include the analysis and control of stochastic processes with applications in telecommunication, logistics, applied probability, finance and insurance. Dr. Rieder is Editor-in-Chief of the journal *Mathematical Methods of Operations Research.*