STOCHASTIC GAMES WITH BOREL PAYOFFS

A. MAITRA AND W. SUDDERTH
University of Minnesota
Minneapolis, USA

1. Introduction

Consider a two-person zero-sum stochastic game with countable state space S, finite action sets A and B for players 1 and 2, respectively, and law of motion p. The history space H_∞ consists of all plays

$$h = (x_1, x_2, \ldots) = ((a_1, b_1, z_1), (a_2, b_2, z_2), \ldots).$$

Thus

$$H_\infty = (A \times B \times S) \times (A \times B \times S) \times \ldots.$$

Let f be a bounded Borel-measurable function defined on H_∞, where the sets A, B, and S are given the discrete topology and H_∞ is given the product topology. The stochastic game $\mathcal{S}(f)(z)$ begins at state $z \in S$, player 1 chooses a strategy σ, player 2 chooses a strategy τ, and then 2 pays 1 the expected value of f, namely $E_{z,\sigma,\tau} f$.

Theorem 1 (Martin [5]) *The game $\mathcal{S}(f)(z)$ has a value.*

Our object in this chapter is to sketch Martin's proof of this remarkable theorem.

Martin entitled his paper "The determinacy of Blackwell games" and explained how Blackwell [1], [2] formulated the general question and solved the special case of G_δ games. Blackwell also conjectured that all Borel games would have a value, as Martin has now confirmed.

Theorem 1 is a vast generalization of the theorem on $\limsup$ games in the previous chapter. However, the proof that $\limsup$ games have a value also provided an algorithm for calculating it. The proof of Theorem 1 below does not reveal how the value of $\mathcal{S}(f)(z)$ can be found. Also, it will be shown in Section 6 that the more general Borel games can, in a sense, be approximated by $\limsup$ games.

2. Games of Perfect Information

The proof of Theorem 1 depends on another celebrated theorem of Martin [4] on games of perfect information. We first recall how these games are played.

Let the action sets A and B for players 1 and 2, respectively, be completely arbitrary nonempty sets. The players alternate turns and do not randomize. Thus player 1 chooses a_1; then, knowing a_1, player 2 chooses b_1. For $n \geq 2$, player 1, knowing $a_1, b_1, \ldots, b_{n-1}$, chooses a_n, and then player 2, knowing $a_1, b_1, \ldots, b_{n-1}, a_n$, chooses b_n. The choices of the players at each stage may be restricted to nonempty subsets of their respective action sets that depend on the actions played up to that point. There is a distinguished subset W of the product space of possible plays $(A \times B)^N = A \times B \times A \times B \times \ldots$ and player 1 wins the game if $(a_1, b_1, a_2, b_2, \ldots) \in W$ and loses otherwise. Assume that the action sets A and B are given their discrete topologies and that the product space $(A \times B)^N$ is given the product topology.

Theorem 2 (Martin [4]) *If W is a Borel subset of $(A \times B)^N$, then either player 1 has a winning strategy or player 2 does.*

We omit the proof of Theorem 2. A very special example may provide some insight.

Example. Suppose that $A = B = \{0, 1\}$. Then each infinite play $(a_1, b_1, a_2, b_2, \ldots)$ can be identified with a binary decimal $.a_1 b_1 a_2 b_2 \ldots$ in the unit interval [0,1]. Thus in every game of perfect information defined by a Borel set $W \subseteq [0, 1]$, either player 1 has a winning strategy or player 2 does.

3. The Games $\mathcal{G}(v, z)$

We will now define a family of perfect information games $\mathcal{G}(v, z)$ for $0 \leq v \leq 1, z \in S$ that are associated with the stochastic game $\mathcal{S}(f)(z)$. An analysis of these games will allow us to deduce Theorem 1 from Theorem 2.

As a first step, define $X = A \times B \times S$ and let ϕ be a bounded function defined on X. Here is a slight variant of the one-day game of the previous chapter.

The one-day game $\mathcal{A}(\phi)(z)$: 1 chooses $a_1 \in A$, 2 chooses $b_1 \in B$, the next state z_1 has distribution $p(\cdot|z, a_1, b_1)$, and 2 pays 1 the expected value of $\phi(a_1, b_1, z_1)$. Let $(G\phi)(z)$ be the value of $\mathcal{A}(\phi)(z)$.

Assume from now on, without loss of generality, that the payoff function f takes values in the unit interval [0,1].

The perfect information game $\mathcal{G}(v, z), 0 \leq v \leq 1, z \in S$:

Player 1 chooses $\phi_0 : X \to [0,1]$ such that $(G\phi_0)(z) \geq v$;
player 2 chooses $x_1 = (a_1, b_1, z_1) \in X$;
player 1 chooses $\phi_1 : X \to [0,1]$ such that $(G\phi_1)(z_1) \geq \phi_0(x_1)$;
player 2 chooses $x_2 = (a_2, b_2, z_2) \in X$;
player 1 chooses $\phi_2 : X \to [0,1]$ such that $(G\phi_2)(z_2) \geq \phi_1(x_2)$;
and so forth.
Player 1 wins the game if $f(x_1, x_2, \ldots) \geq \liminf_n \phi_{n-1}(x_n)$.

Several remarks are in order:

1. It is straightforward to check that the winning set for player 1 is Borel. Thus Theorem 2 applies to show that one of the players has a winning strategy.
2. Player 1 has a legal move at each stage of play, e.g., $\phi_n \equiv 1$.
3. In order to win the game, player 1 would like to choose the functions ϕ_n "small," but the rules require player 1 to choose them sufficiently "large." This creates tension in the game.
4. If player 1 has a winning play in the game $\mathcal{G}(v, z)$ and $0 \leq v' \leq v$, then the same play will win the game $\mathcal{G}(v', z)$.

Define $\tilde{v}$ to be the supremum of the set of all $v \in [0,1]$ such that player 1 has a winning strategy in $\mathcal{G}(v, z)$. It follows from the last remark that player 1 has a win for all $v \in [0, \tilde{v})$ and, by Theorem 2, player 2 has a winning strategy for all $v \in (\tilde{v}, 1]$. It turns out that the value of the stochastic game $\mathcal{S}(f)(z)$ is $\tilde{v}$, as will follow from two lemmas.

Let $\overline{V}(z)$ and $\underline{V}(z)$ be the upper and lower values of the game $\mathcal{S}(f)(z)$, respectively.

Lemma 1 *If player 1 has a winning strategy for $\mathcal{G}(v, z)$, then $\underline{V}(z) \geq v$. Hence, $\underline{V}(z) \geq \tilde{v}$.*

Lemma 2 *If player 2 has a winning strategy for $\mathcal{G}(v, z)$, then $\overline{V}(z) \leq v$. Hence, $\overline{V}(z) \leq \tilde{v}$.*

Theorem 1 is immediate from the two lemmas. The next two sections are devoted to their proofs.

4. The Proof of Lemma 1

Let $\phi_0, \phi_1, \ldots$ be a winning strategy for player 1 in $\mathcal{G}(v, z)$. (Note: Player 1's nth move ϕ_n will, in general, depend on the first n moves $x_1, x_2, \ldots, x_n$ of player 2, but we will usually suppress this dependence to simplify notation.) We will define a strategy $\sigma = (\sigma_0, \sigma_1, \ldots)$ for player 1 in the stochastic game $\mathcal{S}(f)(z)$ such that $E_{z,\sigma,\tau} f \geq v$ for all strategies τ for player 2 in $\mathcal{S}(f)(z)$. This is sufficient.

Let σ_0 be optimal for player 1 in the one-day game $\mathcal{A}(\phi_0)(z)$; thus

$$E_{z,\sigma_0,\nu}\phi_0 \geq (G\phi_0)(z) \geq v,$$

for all probability measures ν on B. For $n \geq 1$, let $\sigma_n = \sigma_n(x_1, x_2, \ldots, x_n)$ be optimal for 1 in $\mathcal{A}(\phi_n)(z_n)$; thus

$$E_{z_n,\sigma_n,\nu}\phi_n \geq (G\phi_n)(z_n) \geq \phi_{n-1}(x_n),$$

for all ν.

Now fix a strategy τ for player 2 in $\mathcal{S}(f)(z)$ and write E for the expectation operator $E_{z,\sigma,\tau}$. Define random variables

$$Y_0 = v, \qquad Y_n = \phi_{n-1}(x_n)\ , n \geq 1.$$

Then, for $n \geq 0$,

$$E[Y_{n+1}|x_1, \ldots, x_n] = E_{z_n,\sigma_n,\tau_n}\phi_n \geq \phi_{n-1}(x_n) = Y_n,$$

and we see that $\{Y_n\}$ is a bounded submartingale. Hence,

$$E[\lim_n Y_n] = \lim_n EY_n \geq v.$$

But

$$f \geq \liminf_n \phi_{n-1}(x_n) = \liminf_n Y_n$$

because $\{\phi_n\}$ is a winning strategy for 1 in $\mathcal{G}(v, z)$. Thus $Ef \geq v$, and the proof of Lemma 1 is complete. ∎

5. The Proof of Lemma 2

Let τ^* be a winning play for player 2 in $\mathcal{G}(v, z)$ and let $\delta > 0$. We will construct a strategy τ for player 2 in the stochastic game $\mathcal{S}(f)(z)$ such that $E_{z,\sigma,\tau}f \leq v + \delta$ for all strategies σ for player 1 in $\mathcal{S}(f)(z)$. This will be sufficient.

Our definition of τ uses a kind of "best response" ψ_n to τ^* in $\mathcal{G}(v, z)$. The ψ_n will be chosen to be as small as possible at each stage, but every response is ultimately futile against the winning play τ^*. Also, the ψ_n can fail to be legal, and we will also need to approximate ψ_n by a legal strategy ϕ_n^*. Finally, the proof will use the notion of a partial history $p_n = (x_1, x_2, \ldots, x_n)$ that is "consistent" with τ^*. The ψ_n, ϕ_n^*, and the notion of "consistency" will all be defined inductively on n.

Call $p_1 = (x_1)$ *consistent* (with τ^*) if there exists a legal first move ϕ_0 for player 1 in $\mathcal{G}(v, z)$ such that player 2 (using τ^*) plays x_1 in response. For (x_1) consistent, let

$$\psi_0(x_1) = \inf \phi_0(x_1),$$

where the infimum is taken over all such ϕ_0, and choose one of them $\phi_0^* = \phi_0^*(x_1)$ such that

$$\phi_0^*(x_1)(x_1) \leq \psi_0(x_1) + \delta/2.$$

Thus $\phi_0^*(x_1)$ is an approximate best (i.e., smallest) legal response when x_1 is consistent. For x_1 inconsistent, the choice is less critical and we set $\psi_0(x_1) = \phi_0^*(x_1) = 1$.

Claim 1 $(G\psi_0)(z) \leq v$.

Proof. Suppose not and let $\epsilon = (G\psi_0)(z) - v$. Set $\phi(x) = \max\{\psi_0(x) - \epsilon, 0\}$, $x \in X$. Then

$$(G\phi)(z) \geq (G\psi_0)(z) - \epsilon = v.$$

So ϕ is a legal first move for player 1 in $\mathcal{G}(v, z)$. Let x^* be the response of player 2 using τ^*. If $\phi(x^*) = 0$, then 1 can play $\phi_n = 0$, for all n and win $\mathcal{G}(v, z)$ - contradicting the assumption that τ^* is a winning play for 2. The other possibility is that $\phi(x^*) = \psi_0(x^*) - \epsilon < \psi_0(x^*)$, which contradicts the definition of ψ_0. This proves the claim.

Now assume that $p_n = (x_1, \ldots, x_n)$ is consistent and that

$$\alpha = \phi_0^*, x_1, \phi_1^*, \ldots, \phi_{n-1}^*, x_n$$

is a legal "partial run" in $\mathcal{G}(v, z)$. Call $p_{n+1} = (x_1, \ldots, x_n, x_{n+1})$ *consistent* if there exists a move ϕ_n for player 1 in $\mathcal{G}(v, z)$ such that player 2 (using τ^*) plays x_{n+1} in response to

$$\alpha, \phi_n = \phi_0^*, x_1, \phi_1^*, \ldots, \phi_{n-1}^*, x_n, \phi_n.$$

For p_{n+1} consistent, define

$$\psi_n(x_{n+1}) = \psi_n(p_n)(x_{n+1}) = \inf \phi_n(x_{n+1}),$$

where the infimum is over all such ϕ_n, and choose one of them $\phi_n^* = \phi_n^*(p_{n+1})$ such that

$$\phi_n^*(p_{n+1})(x_{n+1}) \leq \psi_n(x_{n+1}) + \delta/2^{n+1}.$$

For p_{n+1} inconsistent, let $\psi_n(x_{n+1}) = \phi_n^*(p_{n+1})(x_{n+1}) = 1$.

Claim 2 For $p_n = (x_1, \ldots, x_n)$ consistent, $(G\psi_n)(z_n) \leq \phi_{n-1}^*(p_n)(x_n)$.

Proof. Similar to that of Claim 1. ∎

We are finally ready to define the strategy τ for player 2 in the stochastic game $\mathcal{S}(f)(z)$. Let τ_0 be optimal for player 2 in the one-day game $\mathcal{A}(\psi_0)(z)$

and, for every partial history $p_n = (x_1, \ldots, x_n)$, let $\tau_n(p_n)$ be optimal for 2 in $\mathcal{A}(\psi_n(p_n))(z_n)$, where $x_n = (a_n, b_n, z_n)$. Now fix a strategy σ for player 1 in $\mathcal{S}(f)(z)$ and write E for $E_{z,\sigma,\tau}$.

Define random variables

$$W_0 = v, \qquad W_n = \phi^*_{n-1}(x_1, \ldots, x_n)(x_n) \ , n \geq 1.$$

Then

$$E[W_{n+1}|x_1, \ldots, x_n] \leq E[\psi_n(x_{n+1})|x_1, \ldots, x_n] + \delta/2^{n+1}.$$

Also,

$$E[\psi_n(x_{n+1})|x_1, \ldots, x_n] = E_{z_n,\sigma_n,\tau_n}\psi_n \leq (G\psi_n)(z_n) \leq$$
$$\leq \phi^*_{n-1}(x_1, \ldots, x_n)(x_n) = W_n,$$

where the first inequality is by definition of τ_n and the second is a consequence of the claims and the definition of ϕ^*_{n-1} on inconsistent partial histories $p_n = (x_1, \ldots, x_n)$. Hence,

$$EW_{n+1} \leq EW_n + \delta/2^{n+1}, \qquad n \geq 0,$$

and so

$$EW_n \leq EW_0 + \delta = v + \delta, \qquad n \geq 0.$$

By the Fatou inequality,

$$E[\liminf W_n] \leq \liminf EW_n \leq v + \delta.$$

But the strategy τ^* is a winning strategy for 2 in $\mathcal{G}(v, z)$; so

$$f \leq \liminf \phi^*_{n-1}(x_1, \ldots, x_n)(x_n) = \liminf W_n.$$

Thus $Ef \leq v + \delta$ and the proof is complete. ∎

6. An Approximation Theorem

For a function u with domain the set of all finite sequences $(x_1, \ldots, x_n)$ of elements of X, define functions u^* and u_* on each history $h = (x_1, \ldots, x_n, \ldots)$ by

$$u^*(h) = \limsup_n u(x_1, \ldots, x_n), \qquad u_*(h) = \liminf_n u(x_1, \ldots, x_n).$$

Let $V(f)(z)$ be the value of the stochastic game $\mathcal{S}(f)(z)$ for f a bounded Borel-measurable function as in Theorem 1.

Theorem 3

$$V(f)(z) = \sup\{V(u_*)(z) \,:\, u_* \le f\} = \inf\{V(u^*)(z)\,;\, u^* \ge f\}.$$

Proof. For the first equality, let $v < V(f)(z)$ and define

$$u(x_1, \ldots, x_n) = \phi_{n-1}(x_n) = Y_n,$$

where ϕ_{n-1} and Y_n are as in the proof of Lemma 1. Then $u_* \le f$ and $V(u_*) \ge v$.

The second equality follows from the first by considering $-f$. ∎

7. A Generalization and a Question

Theorem 1 can be extended to arbitrary sets S, A, and B in the framework of finitely additive probability theory [3]. Indeed, the same proof goes through with only minor changes. It is an open question whether the theorem can be generalized to a Borel-measurable setting like that of Nowak [6] in which S, A, B are Borel subsets of Polish spaces with some additional requirements of compactness and continuity. A serious technical difficulty arises in the proof because now the strategies in the stochastic game must be measurable. But since good strategies in the stochastic game are constructed from winning strategies in the game of perfect information, their measurability depends on the existence of measurable winning strategies in the game of perfect information. We do not know whether or not measurable winning strategies always exist for the game of perfect information in a Borel setting.

References

1. Blackwell, D. (1969) Infinite G_δ games with imperfect information, *Zastosowania Matematyki* **10**, 99–101.
2. Blackwell, D. (1989) Operator solution of infinite G_δ games of imperfect information, in T.W. Anderson, K. Athreya, and D.L. Iglehart (eds.), *Probability, Statistics, and Mathematics: Papers in Honor of Samuel Karlin*, Academic Press, New York, pp. 83–87.
3. Maitra, A. and Sudderth, W. (1998) Finitely additive stochastic games with Borel-measurable payoffs, *International Journal of Game Theory* **27**, 257–267.
4. Martin, D.A. (1975) Borel determinacy, *Annals of Mathematics* **102**, 363–371.
5. Martin, D.A. (1998) The determinacy of Blackwell games, *The Journal of Symbolic Logic* **63**, 1565–1581.
6. Nowak, A.S. (1985) Universally measurable strategies in zero-sum stochastic games, *Annals of Probability* **13**, 269–287.