STOCHASTIC GAMES WITH LIM SUP PAYOFF

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. Let u be a bounded real-valued function defined on the state space S and assume that the payoff from 2 to 1 along a play (or infinite history)

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

is

$$u^*(h) = \limsup_n u(z_n).$$

Suppose that the stochastic game begins at state $z \in S$, player 1 chooses the strategy σ, player 2 chooses the strategy τ, and $E_{z,\sigma,\tau}$ is the induced expectation on the space H_∞ of all plays h. Then the expected payoff from 2 to 1 is

$$E_{z,\sigma,\tau} u^*.$$

The main objective of this chapter is to sketch the proof that this lim sup game has a value and that there is an algorithm, albeit transfinite, for calculating the value. The chapter is based on our paper (Maitra and Sudderth [7]). A more detailed and more leisurely proof is in our book (Maitra and Sudderth [10]).

2. A Remark on the Lim Sup Payoff

The payoff function u^* is more general than it may first appear. To see this, make a change of coordinates

$$\tilde{z}_n = (z_0, a_1, b_1, z_1, \ldots, a_n, b_n, z_n),$$

and take

$$\tilde{u}(\tilde{z}_n) = \sum_{k=1}^{n} \beta^{k-1} r(z_{k-1}, a_k, b_k)$$

or

$$\tilde{u}(\tilde{z}_n) = \frac{1}{n} \sum_{k=1}^{n} r(z_{k-1}, a_k, b_k),$$

where $z_0 = z$ and r is a bounded daily payoff function. Thus lim sup games include discounted games and average reward games.

3. A Little History

Our work on lim sup games stems from two sources: the gambling theory of Dubins and Savage [4] and Blackwell's papers on G_δ games [1], [2]. If player 2 is a dummy with only one action, then the lim sup game becomes a nonleavable gambling problem in the sense of Dubins and Savage (cf. also [10]). The algorithm for the value first appeared in the context of gambling theory in Dubins, Maitra, Purves, and Sudderth [3]. If the function u is the indicator function of a set $W \subset S$, then the function u^* is the indicator of the set of plays

$$[W \text{ i.o.}] = \bigcap_n \bigcup_{k \geq n} \{h \,:\, z_n \in W\}.$$

This set is a G_δ, that is, a countable intersection of open sets, when S, A, and B are given the discrete topology and H_∞ has the product topology. Furthermore, every G_δ subset of H_∞ can be written this way if we make the change of coordinates as in Section 2 above. When u is an indicator function, the algorithm for the value of the lim sup game is essentially that of Blackwell for G_δ games.

4. The Algorithm

The algorithm for the value of the lim sup game uses a number of auxiliary games. Here is the first and most basic of them.

The one-day game $\mathcal{A}(u)(z)$: 1 chooses $a \in A$, 2 chooses $b \in B$, the next state z_1 has distribution $p(\cdot|z, a, b)$, and 2 pays 1 the expected value of $u(z_1)$. Let $(Gu)(z)$ be the value of $\mathcal{A}(u)(z)$, which exists by von Neumann's minmax theorem.

Here is the second type of auxiliary game.

The leavable game $\mathcal{L}(u)(z)$: 1 chooses a strategy σ in the stochastic game and a stop rule t with values in $\{0, 1, 2, \ldots\}$, 2 chooses a strategy τ in the stochastic game, and 2 pays 1 the quantity $E_{z,\sigma,\tau} u(z_t)$, the expected value of u evaluated at the state where the process is stopped. (Note: The stop rule t selected by player 1 is required to be finite on *every* history. It would not make sense to require that t be almost surely finite because player 1 does not know player 2's strategy and, therefore, does not know the probability measure on the space of histories.) This game also has a value and the value can be calculated by backward induction as follows. Define

$$U_0 = u, \qquad U_{n+1} = u \vee GU_n, \; n = 0, 1, \ldots$$

where $x \vee y$ denotes the maximum of x and y. Finally, set

$$U = \sup U_n.$$

Theorem 1 *For each $z \in S$, the leavable game $\mathcal{L}(u)(z)$ has value $U(x)$. Furthermore, $U = u \vee GU$.*

The proof of Theorem 1 uses standard techniques and is omitted here.

In the leavable game player 1 is allowed to stop the game at any time. In order to approach the lim sup game, we will force player 1 to stop later and later. As a first step, we alter the leavable game so that 1 cannot stop immediately.

The modified leavable game $\mathcal{L}^*(u)(z)$: This is the same as $\mathcal{L}(u)(z)$ except that player 1 must choose a stop rule $t \geq 1$.

After their first moves in this modified game, the players are faced with a leavable game as previously defined. Thus the modified game amounts to a one-day game with payoff U, and the following corollary to Theorem 1 is easy to prove.

Corollary 1 *For each $z \in S$, $\mathcal{L}^*(u)(z)$ has value $(GU)(z)$.*

Now define the operator T by the rule that $(Tu)(z)$ is the value of $\mathcal{L}^*(u)(z)$, or equivalently

$$(Tu)(z) = (GU)(z).$$

By iterating the operator T, we can, in effect, force player 1 to stop later and later. This is one of the key ideas for finding an algorithm for the lim sup game, to which we now turn.

The nonleavable (lim sup) game $\mathcal{N}(u)(z)$: Player 1 chooses a strategy σ, player 2 chooses a strategy τ, and 2 pays 1 $E_{z,\sigma,\tau}u^*$.

To obtain the value of this game, we iterate the operator T to force player 1 to stop later and later, but, in addition, we take the minimum with the function u at every step so that player 1 will also try to reach states where u is large.

Define

$$V_0 = Tu, \qquad V_n = T(u \wedge V_{n-1}),\ n = 1, 2, \ldots$$

where $x \wedge y$ denotes the minimum of x and y. Next, set

$$V_\omega = \inf V_n.$$

If S is finite, then $V_\omega(z)$ is the value of $\mathcal{N}(u)(z)$. However, in general, we must continue and define, for countable ordinal numbers α,

$$V_\alpha = \begin{cases} T(u \wedge V_{\alpha-1}), & \text{if } \alpha \text{ is a successor ordinal} \\ \inf_{\beta<\alpha} V_\beta, & \text{if } \alpha \text{ is a limit ordinal.} \end{cases}$$

Finally, let

$$V = V_{\omega_1} = \inf_{\alpha<\omega_1} V_\alpha,$$

where ω_1 is the first uncountable ordinal.

Theorem 2 *For each $z \in Z$, the nonleavable game $\mathcal{N}(u)(z)$ has value $V(z)$. Also, $V = T(u \wedge V)$.*

Before sketching the proof of Theorem 2, we will make a few digressions. We begin with a simple example to illustrate the algorithm for V.

Example 1 *A simple recursive game.* Recursive games are stochastic games in which payoff occurs only in absorbing states. They can be viewed as lim sup games such that, for every $z \in Z$, either $u(z) = 0$ or z is absorbing. Recursive games were introduced by Everett [5], and are treated in this volume in [12], [13]. Here is one of Everett's examples in suggestive notation.

$$\Gamma = \begin{pmatrix} 0 & 2^* \\ 1^* & 0^* \end{pmatrix}.$$

This means that there are four states $0, 0^*, 1^*$, and 2^*. The states with *'s are absorbing and $u(z^*) = z$ for $z = 0, 1$, and 2. The state 0 is not absorbing and $u(0) = 0$. Beginning at 0, the players face a matrix game with $A = \{a_1, a_2\}$ and $B = \{b_1, b_2\}$. If they play $a = a_1$ and $b = b_1$, then the game continues. If not, they move to one of the absorbing states and the game is over. Obviously, $V = u$ at any absorbing state, and we need only consider the state 0, as we go through the steps of the algorithm for V.

To calculate $V_0(0) = Tu(0) = GU(0)$, we need the values of $U_n(0)$. Now $U_0(0) = u(0) = 0$, and

$$U_{n+1}(0) = (GU_n)(0) = \text{value}\begin{pmatrix} U_n(0) & 2 \\ 1 & 0 \end{pmatrix} = \frac{2}{3 - U_n(0)}.$$

Passing to the limit, we see that $U(0) = \lim U_n(0) = 1$, and

$$V_0(0) = GU(0) = \text{value} \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix} = 1.$$

To calculate $V_1(0) = T(u \wedge V_0)(0)$, we need the value $\tilde{U}$ of the leavable game $\mathcal{L}(\tilde{u})(0)$, where $\tilde{u} = u \wedge V_0$. But $u \wedge V_0 = u$. So $\tilde{U} = U$, and $V_1 = G\tilde{U} = GU = V_0$. Thus the algorithm ends at V_0 and $V(0) = V_0(0) = 1$.

For general recursive games with S countable, the algorithm can continue without terminating all the way up to an arbitrary countable ordinal.

5. The Special Case $u = 1_W$

The algorithm for the value of the lim sup game is somewhat simpler in the special case when $u = 1_W$, the indicator function of a set $W \subset S$. If $z \notin W$, then it is not difficult to show that $U_n(z)$ is the value of a stochastic game in which the payoff function is the indicator of the set

$$[\text{reach } W \text{ by time } n] = \bigcup_{k=1}^{n} \{h : z_k \in W\},$$

and $U(z)$, the value of the leavable game, is also the value of a stochastic game with payoff function the indicator of the set

$$[\text{reach } W] = \bigcup_{k=1}^{\infty} \{h : z_k \in W\}.$$

(This is intuitively plausible because player 1 has no incentive to stop until a state in W is reached and then it is obviously optimal for player 1 to stop.) The set [reach W] is an open subset of H and, conversely, every open subset can be written in this form after a change of coordinates as in Section 2.

It can also be shown that, for $n = 0, 1, \ldots$, $V_n(z)$ is the value of a stochastic game with payoff function the indicator function of the set

$$[\text{reach } W \; n+1 \text{ times}] = \{h : z_k \in W \text{ for at least } n+1 \text{ indices } k\}$$

and, as was mentioned in Section 3, the nonleavable game has a payoff function equal to the indicator of [W *i.o.*]. So it is natural to hope that its value $V(z)$ will be equal to $V_\omega(z) = \inf V_n(z)$. This is true when S is finite, but not in general.

Example 2 Let S be the set of integers, $A = \{0, 1\}$, $B = \{0\}$, and $W = \{1, 2, \ldots\}$. At any negative integer z, player 1 can move to either $z - 1$ or

$-z$ by taking $a = 0$ or $a = 1$; at any strictly positive z, the motion is to $z - 1$ regardless of the action chosen; state 0 is absorbing. It is easy to see that $V_\omega(-1) = V_n(-1) = 1$ for $n = 0, 1, \ldots$, but $V(-1) = 0$.

6. Sketch of the Proof of Theorem 2

The proof will be given in four lemmas.

Lemma 1 $T(u \wedge V) = V$.

Proof. It is not difficult to show that the bounded function $V_\alpha(z)$ is nonincreasing in α for each fixed $z \in S$. By a property of the ordinals, there exists $\alpha(z) < \omega_1$ such that $V_\alpha(z) = V_{\alpha(z)}(z)$ for all $\alpha \geq \alpha(z)$. Define $\alpha^* = \sup \alpha(z)$. Then, by another property of the ordinals, $\alpha^* < \omega_1$. Also, we have $V_\alpha = V_{\alpha^*}$ for all $\alpha \geq \alpha^*$. Hence, $V = V_{\alpha^*} = V_{\alpha^*+1}$ and, in particular, $V = T(u \wedge V_{\alpha^*}) = T(u \wedge V)$. ∎

Lemma 1 says that the value of the modified leavable game $\mathcal{L}^*(u \wedge V)(z)$ has value $V(z)$ for each $z \in Z$. The corollary below thus follows immediately from the definition of this game.

Corollary 2 *For every $z \in Z$ and $\delta > 0$ there is a strategy $\sigma = \sigma(z, \delta)$ and a stop rule $t = t(z, \delta) \geq 1$ for player 1 such that*

$$E_{z,\sigma,\tau}(u \wedge V)(z_t) \geq V(z) - \delta$$

for all strategies τ of player 2.

Let $\underline{V}(z)$ be the lower value of the game $\mathcal{N}(u)(z)$.

Lemma 2 $\underline{V} \geq V$.

Proof. Fix $z \in Z$ and $\epsilon > 0$. It suffices to construct a strategy σ' for player 1 such that for all strategies τ for 2,

$$E_{z,\sigma',\tau} u^* \geq V(z) - \epsilon.$$

Referring to Corollary 2, let σ' be the strategy that follows

$$\sigma(z, \epsilon/2) \text{ up to time } s_1 = t(z, \epsilon/2)$$

and then follows

$$\sigma(z_{s_1}, \epsilon/4) \text{ up to time } s_2 = s_1 + t(z_{s_1}, \epsilon/4)$$

and then

$$\sigma(z_{s_2}, \epsilon/8) \text{ up to time } s_3 = s_2 + t(z_{s_2}, \epsilon/8)$$

and so on.

Fix a strategy τ for player 2 and write E for the expectation operator $E_{z,\sigma',\tau}$. Define random variables

$$Y_n = (u \wedge V)(z_{s_n}), \qquad n = 1, 2, \ldots.$$

It follows from our construction of σ' that

$$EY_1 \geq V(z) - \epsilon/2$$

and

$$E[Y_{n+1}|z_1, \ldots, z_{s_n}] \geq V(z_{s_n}) - \epsilon/2^{n+1} \geq Y_n - \epsilon/2^{n+1}.$$

Hence,

$$EY_n \geq V(z) - \epsilon$$

for all n, and

$$Eu^* = E[\limsup u(z_n)] \geq E[\limsup u(z_{s_n})] \geq E[\limsup Y_n] \geq \limsup EY_n$$
$$\geq V(z) - \epsilon.$$

■

The next lemma states a general property of the expectation of a lim sup.

Lemma 3 $Eu^* = \inf_s \sup_{t \geq s} Eu(z_t)$.

For a proof, see Maitra and Sudderth [10], Theorem 4.2.2.

Now let $\bar{V}(z)$ be the upper value of the game $\mathcal{N}(u)(z)$. The next lemma, together with Lemma 2, will complete our sketch of the proof of Theorem 2.

Lemma 4 $\bar{V} \leq V$.

Proof. Since $V = \inf_{\alpha<\omega_1} V_\alpha$, it suffices to show that $\bar{V} \leq V_\alpha$ for each $\alpha < \omega_1$. This we do by induction on α.

First $\bar{V} \leq V_0$ because

$$\bar{V}(z) = \inf_\tau \sup_\sigma E_{z,\sigma,\tau} u^* \leq \inf_\tau \sup_\sigma \sup_{t \geq 1} E_{z,\sigma,\tau} u(z_t) = (Tu)(z) = V_0(z),$$

where the inequality is by Lemma 3.

For the inductive step, suppose that $\bar{V} \leq V_\beta$ for all $\beta < \alpha$. If α is a limit ordinal, then

$$V_\alpha = \inf_{\beta<\alpha} V_\beta \geq \bar{V}.$$

So suppose that α is a successor. Then $V_\alpha = T(u \wedge V_{\alpha-1})$. Fix z and $\epsilon > 0$. It suffices to show that there is a strategy τ for player 2 such that, for all strategies σ of player 1,

$$E_{z,\sigma,\tau} u^* \le V_\alpha(z) + \epsilon.$$

Define τ to be a strategy that follows a strategy τ_1, which is $\epsilon/2$-optimal in $\mathcal{L}^*(u \wedge V_{\alpha-1})(z)$ up to time

$$\theta = \inf\{n \,:\, u(z_n) > V_{\alpha-1}(z_n)\,\}.$$

If the stopping time θ is finite, the strategy τ then switches to $\tau_2(z_\theta)$, a strategy that is $\epsilon/2$-optimal in $\mathcal{N}(u)(z_\theta)$. It can be shown, with some effort and again with the aid of Lemma 3, that this strategy τ does the job. For the details of this argument, see Maitra and Sudderth [7]. A somewhat different proof is given in Maitra and Sudderth [10]. ■

7. Approximation Theorems

Consider a stochastic game in which the payoff is the indicator function of an arbitrary Borel subset E of H. Let $\bar{V}(E)(z)$ be the upper value starting from the initial state z.

Theorem 3 (Maitra, Purves, and Sudderth [6])

$$\bar{V}(E)(z) = \inf\{V(O)(z) \,:\, O \textit{ is open},\ E \subseteq O\}.$$

As was remarked in Section 5, a stochastic game with payoff the indicator function of an open set corresponds to a leavable game after a change of coordinates. Thus Theorem 3 says that games with Borel set payoffs can, in a sense, be approximated by leavable games.

Next consider a stochastic game with payoff a bounded Borel measurable function f on H. Let $\bar{V}(f)(z)$ be the upper value. As in Section 2, let $\tilde{u}$ denote a bounded function defined on finite sequences $\tilde{z}_n = (z, a_1, b_1, z_1, \ldots, a_n, b_n, z_n)$ and set $\tilde{u}^*(h) = \limsup \tilde{u}(\tilde{z}_n)$, for $h = (z, a_1, b_1, z_1, \ldots)$.

Theorem 4 (Maitra and Sudderth [10])

$$\bar{V}(f)(z) = \inf\{V(\tilde{u}^*)(z) \,:\, \tilde{u}^* \ge f\}.$$

Thus we can approximate games with bounded Borel functions as payoffs by lim sup games. (It will be shown in the next chapter that all these Borel games do in fact have values - so the bars can be removed in Theorems 3 and 4. Also, a proof will be given for Theorem 4.)

It follows from the form of the algorithm in Section 4 together with Theorems 3 and 4 that the value of any game with a Borel payoff is completely determined by the one-day operator G. To see this, observe that each step in the algorithm for a leavable game $\mathcal{L}(u)(z)$ is determined by G, and that each step in the algorithm for a nonleavable game $\mathcal{N}(u)(z)$ amounts to another leavable game, which is also determined by G.

8. Generalizations

Theorems 1 and 2 can be extended to arbitrary sets S, A,and B in the framework of finitely additive probability theory. There is a drawback, however. Fubini's theorem fails for finitely additive measures and the value of the game may depend on the order of integration [8].

There is also a countably additive generalization in which, as in [11], S, A, and B are Borel subsets of Polish spaces with some additional requirements of compactness and continuity [9].

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. Dubins, L., Maitra, A., Purves, R., and Sudderth, W. (1989) Measurable, nonleavable gambling problems, *Israel Journal of Mathematics* **67**, 257–271.
4. Dubins, L. E. and Savage, L. J. (1965) *How to Gamble If You Must: Inequalities for Stochastic Processes*, McGraw-Hill, New York.
5. Everett, H. (1957) Recursive games, in M. Dresher, A.W. Tucker, and P. Wolfe (eds.), *Contributions to the Theory of Games, III*, Annals of Mathematics Studies 39, Princeton University Press, Princeton, NJ, pp. 47–78.
6. Maitra, A., Purves, R., and Sudderth, W. (1992) Approximation theorems for gambling problems and stochastic games, *Game Theory and Economic Applications*, Lecture Notes in Economics and Mathematical Systems 389, Springer-Verlag, Berlin, 114–132.
7. Maitra, A. and Sudderth, W. (1992) An operator solution of stochastic games, *Israel Journal of Mathematics* **78**, 33–49.
8. Maitra, A. and Sudderth, W. (1993a) Finitely additive and measurable stochastic games, *International Journal of Game Theory* **22**, 201–223.
9. Maitra, A. and Sudderth, W. (1993b) Borel stochastic games with limsup payoff, *Annals of Probability* **21**, 861–885.
10. Maitra, A. and Sudderth, W. (1996) *Discrete Gambling and Stochastic Games*, Springer-Verlag, New York.
11. Nowak, A.S. (1985) Universally measurable strategies in zero-sum stochastic games, *Annals of Probability* **13**, 269–287.
12. Thuijsman, F. (2003) Recursive games, in A. Neyman and S. Sorin (eds.), *Stochastic Games and Applications*, NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Kluwer Academic Publishers, Dordrecht, Chapter 16, pp. 253–264.

13. Vieille, N. (2003) On a class of recursive games, in A. Neyman and S. Sorin (eds.), *Stochastic Games and Applications*, NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Kluwer Academic Publishers, Dordrecht, Chapter 19, pp. 293–307.