ON THE EXISTENCE OF STATIONARY OPTIMAL STRATEGIES

DONALD ORNSTEIN

The question with which this paper is concerned is roughly speaking: In a gambling situation or dynamical programming situation, are strategies that take the past into account any better than those that are based only on the present situation? Let us now state precisely what situations we will be dealing with.

We have a set X, the states of our system (e.g. how much money you have). For each state $x \in X$ we have a collection V_x of gambles available to you when in state x. Each gamble $v \in V_x$ will be a measure on X with support on a countable number of points. (If you chose v you go from x to y with probability $v(y)$.) A strategy tells you how to choose a gamble, v_n, on the nth day as a function of the previous history of the system $(x_1, x_2, \cdots, x_n)$ (v_n must be in V_{x_n}).[1] A stationary strategy is one where the choice of v_n depends only on x_n.

Suppose there is a special state g(our goal). Then for each strategy s we have a function on X, $F_s(x)$ the probability of reaching g if we start at x and use s. Let $F(x) = \sup F_s(x)$, where sup is taken over all strategies s.

Our main result is

THEOREM A. *There is a system where X is uncountable such that $F(x) = 1$ for all x but if s is any stationary strategy, then for some x_0 in X, $F_s(x_0) < \frac{1}{2}$. (Each collection V_x will be countable and each v will have only 2 points in its support.)*

Theorem A should be compared to

THEOREM B. *If X is countable, then for each $\epsilon > 0$ there is a stationary s such that $F_s(x) \geqq (1-\epsilon)F(x)$ for all $x \in X$.*

Theorem B holds for more general systems. Instead of trying to reach a goal we can have a system where we get paid $p(x, v, x') \geqq 0$ if we are in state x, chose v in V_x and go from x to x'. We now define $F_s(x)$ as the expected value of $\sum_{i=1}^{\infty} p(x_i v_i x_{i+1})$ where $x_1 = x$, and we use s. Let $F(x) = \sup F_s(x)$, where sup is taken over all strategies s.

Received by the editors January 29, 1968.

[1] All the results in this paper would also hold if we changed the definition of strategy so that v_n is a function of $(x_1, v_1, x_2, v_2, \cdots, v_{n-1}, x_n)$

THEOREM C. *If we assume that* $F(x) < \infty$ *for all* x *then Theorem* B *is still true in this more general context.*

At first glance one might expect that if $F(x) = \infty$ for some x then for each n we would be able to find a stationary strategy s such that $F_s(x) > n$ for all x where $F(x) = \infty$. This is not so as the following simple example shows. Let X be the integers. If we are at $u > 0$, the only gamble available is to stay there and get paid nothing. If we are at 0 we can go to any $n > 0$ and get paid n. If we are at $-n$, $n > 0$, we go to $-(n-1)$ with probability $\frac{1}{2}$ and to 1 with probability $\frac{1}{2}$ getting paid nothing in either case. $F(n) = \infty$ for all $n \leqq 0$ but once we determine what to do at 0, our expected winnings are known if we are at $-n$. For example if we choose to go to k from 0, the expected winnings from $-n$ are $k/2^n$.

Theorem C comes close to being false. Blackwell has an example [2] in which $F(x)$ is finite for each x (but is unbounded) and for each stationary strategy s there is an x_0 such that $F(x_0) - F_s(x_0) > \frac{1}{2}$.

Dubins and Savage have a theorem [3, p. 60] that shows that a stochastic element is essential in Theorem A.

Theorem A gives us some information about finite or countable X because it limits the kind of algorithms we can have for choosing a stationary s with the property that $F_s(x) \geqq (1-\epsilon)F(x)$ for all $x \in X$.

§1 contains a proof of Theorem A and §2 contains a proof of Theorem C (which although more general is no harder to prove than Theorem B). In §3 we will give a discussion on extending Theorem C to the case where $p(x, v, x')$ may be negative. We will also discuss the case when there happens to be an optimal strategy.

Before concluding this introduction, I would like to mention that this paper is an outgrowth of many discussions with Lester Dubins.

1. **Proof of Theorem A.** Before actually giving the construction we will give a brief description of what it will look like. There will be a point $b \in X$, (distinct from g) from which you can never leave. For each ordinal $\alpha < \Omega$ (Ω = the first uncountable ordinal) there will correspond a collection of points C_α. X will be $\bigcup_{\alpha<\Omega} C_\alpha$ with b and g added. C_1 will consist of one point, y_1, for each n we will have the following gamble in V_{y_1}: go to g with probability $1-1/2^n$ and go to b with probability $1/2^n$.

Each V_x, $x \in C_\alpha$ will consist of a countable number of gambles each of which will have support on 2 points, a point in $\bigcup_{\beta<\alpha} C_\beta$ and b.

We will now define the C_α and the possible gambles at each point of C_α inductively. Suppose that we have done this for all $\beta < \alpha$, and that $F(x) = 1$ for all $x \in \bigcup_{\beta<\alpha} C_\beta$. For each stationary strategy s, on $\bigcup_{\beta<\alpha} C_\beta$ such that

$$\inf_{y \in \bigcup_{\beta<\alpha} C_\beta} F_s(y) > \tfrac{1}{2}$$

we will introduce one point x_s in C_α. To determine the possible gambles at x_s do the following: let $l_s = \inf F_s(x)$ where inf is taken over all $x \in \bigcup_{\beta<\alpha} C_\beta$. We now distinguish 2 cases:

(1) There is a point, say x_0, such that $F_s(x_0) = l_s$. In this case we introduce for each integer n the gamble: go to x_0 with probability $1 - 1/2^n$ and go to b with probability $1/2^n$.

(2) There is no x such that $F_s(x) = l_s$. In this case we pick a sequence x_n such that $\lim_{n\to\infty} F_s(x_n) = l_s$. For each n we introduce the gamble: go to x_n with probability $1 - 2(F_s(x_n) - l_s)$ and go to b with probability $2(F_s(x_n) - l_s)$.

We have now described the construction and we will check some of its properties.

(A) $F(x) = 1$ for all $x \in X$. To check (A) it is enough to check that $F(x) = 1$ for all $x \in C_\alpha$ assuming that $F(x) = 1$ for all $x \in C_\beta$ if $\beta < \alpha$. This is clear because for each $x \in C_\alpha$ and $\epsilon > 0$ we can reach some y, $y \in C_\beta$ $\beta < \alpha$, with probability $> 1 - \epsilon$ and by our induction hypotheses we can reach g from y with probability $> 1 - \epsilon$.

(B) For any stationary strategy t on

$$\bigcup_{\beta \le \alpha} C_\beta \quad \left(\text{with } \inf_{y \in \bigcup_{\beta<\alpha} C_\beta} F_t(y) > \frac{1}{2} \right)$$

there will be a point $x^t \in C_\alpha$ such that $F_t(x^t) < \inf F_t(y)$ where inf is taken over all $y \in \bigcup_{\beta<\alpha} C_\beta$.

To see (B): Let s be the restriction of t to $\bigcup_{\beta<\alpha} C_\beta$. We will take x^t to be x_s which was defined in the construction. As before let $l_s = \inf F_s(x)$ where inf is taken over all $x \in \bigcup_{\beta<\alpha} C_\beta$. If we are in case (1), i.e. there is an x_0 such that $F_s(x_0) = l_s$ and $x_0 \in \bigcup_{\beta<\alpha} C_\beta$, then it is clear that no matter what possible gamble t chooses at x_s we have $F_t(x_s) < F_s(x_0) = \inf F_s(y) = \inf F_t(y)$, where both infs are taken over all $y \in \bigcup_{\beta<\alpha} C_\beta$.

If we are in case (2), then t will assign to x_s the gamble: go to x_n with probability $1 - 2(F_s(x_n) - l_s)$ and to b with probability $2(F_s(x_n) - l_s)$ for some n. Then

$$\begin{aligned} F_t(x_s) &= [1 - 2(F_s(x_n) - l_s)]F_s(x_n) < F_s(x_n) - (F_s(x_n) - l_s) \\ &= l_s = \inf F_t(y) \end{aligned}$$

(inf taken over all $y \in \bigcup_{\beta<\alpha} C_\beta$). (We get the inequality because we assumed that $l_s > \frac{1}{2}$ and since $F_s(x_n) > l_s$ we have $2F_s(x_n) > 1$.)

(Note that we used the stationarity of t in first determining s, independently of x_s, and then picking our starting point to be x_s.)

We will now use ⊛ to show that no stationary strategy can be everywhere better than $\frac{1}{2}$.

Let s be a stationary strategy for X. For each ordinal α define $r(\alpha)=\inf F_s(x)$ taking inf over all $x\in\bigcup_{\beta\leq\alpha} C_\beta$. ⊛ implies that $r(\alpha)<\inf_{\beta<\alpha} r(\beta)$ or that

$$\inf_{x\in \bigcup_{\beta<\alpha} C_\beta} F_s(y) < \tfrac{1}{2}$$

in which case we are finished. However, it is easy to see that if we have a function $h(\alpha)$ from the ordinal up to Ω, to the reals with the property that $h(\alpha)<\inf_{\beta<\alpha} h(\beta)$, then $h(\alpha)=0$ for some α. (To see this last statement: there must be some n such that $\inf_{\beta<\alpha} h(\beta)-h(\alpha)>1/2^n$ for infinitely many α. Let α_1 be the first of these α's, α_2 the second, α_k the kth etc. Then $\alpha_k-\alpha_{k+1}>1/2^n$ for all integers k giving us a contradiction.)

2. Proof of Theorem C.

L1. *Theorem* C *is true if X is finite.*

A proof of this can be found in [1, p. 58]. To keep this paper self-contained we will give a proof, but this will be postponed until later.

We will start our proof by picking a state $y\in X$, $\epsilon>0$. Then pick $\epsilon_1>0$ and $\epsilon_2>0$ such that

$$1/(1+\epsilon_2) > 1-\tfrac{1}{2}\epsilon, \qquad 4\epsilon_1<\epsilon \quad \text{and} \quad 8(\epsilon_1/\epsilon_2)<\epsilon.$$

(1) We can find a finite set A (let $B=X-A$) and a strategy s such that under s we stop when we hit B and such that $F_s(y)\geqq(1-\epsilon_1)F(y)$.

To see (1): Let s_1 be such that $F_{s_1}(y)\geqq(1-\frac{1}{4}\epsilon_1)F(y)$. Let $F_s^N(x)$ be the expected amount won before time N starting at x under s. Pick N such that $F_{s_1}^N(y)\geqq(1-\frac{1}{2}\epsilon_1)F(y)$. Next, let v_n be the measure on X such that $v_n(x)=$probability that you are in x at time n, starting at y and using s. Pick a finite set A such that

$$\sum_{x\notin A} v_n(x)F(x) \leqq \tfrac{1}{4}\epsilon_1 \frac{1}{N} F(y) \qquad \text{for all } n\leqq N.$$

Now modify s_1 by stopping when we are outside of A. This gives us s and (1) is now clear.

(2) By L1 we can pick a stationary strategy t such that, using t we stop when we hit B, and $F_t(y)\geqq(1-2\epsilon_1)F(y)$.

(3) Let E be the set of $x\in A$ such that (a) $(1+\epsilon_2)F_t(x)\leqq F(x)$ and

let t' be the stationary strategy that stops when we are in E and that agrees with t when we are not in E. We then have (b) $F_{t'}(y) \geqq (1-\epsilon)F(y)$.

To see this: let $F_t(y) = a+b$ where a and b are the expected amounts won before and after hitting E respectively (starting at y, using t). Then $F(y) \geqq a+(1+\epsilon_2)b$ by (a). But by (2) we have $(a+b) \geqq (1-2\epsilon_1)F(y)$. We thus get

$$a + b \geqq (1 - 2\epsilon_1)[a + (1 + \epsilon_2)b] \geqq a + (1 + \epsilon_2)b - 4\epsilon_1(a + b);$$

hence $4\epsilon_1(a+b) \geqq \epsilon_2 b$ we therefore get that $b < \frac{1}{2}\epsilon F_t(y)$. This and (2) gives (3).

(4) Let s be any stationary strategy that agrees with t (and t') whenever we are in $A-E$. Then (a) $F_s(y) \geqq (1-\epsilon)F(y)$. This follows from (3)(b). Let ${}_1F(x) = \sup_{s \in C} F_s(x)$ where C is the collection of all strategies that agree with t whenever we are in $A-E$. Then (b) ${}_1F(x) \geqq (1-\epsilon)F(x)$ for all x. (To see this note that if $x \in A-E$, t will do because of (3a). If $x \notin A-E$ we can use an s such that $F_s(x) \geqq (1-\frac{1}{2}\epsilon)F(x)$ until we hit $A-E$ set, then continue using t. As before this works because of (3a).

(5) If we change each V_x, $x \in A-E$ to include only one gamble, the one designated by t we get a new system whose F function is ${}_1F$ (${}_1F(x) \geqq (1-\epsilon)F(x)$ for all $x \in X$). If s is any stationary strategy for this new system we will win more than $(1-\epsilon)F(y)$ if we start at y and use s by (4a). Now suppose that we ordered all the states in X, y_1, $y_2 \cdots$ and that $y = y_1$. Repeat the same procedure using y_2, our new system, and $\frac{1}{2}\epsilon$. We get a third system and repeat the procedure with this system y_3 and $\epsilon/2^2$. This gives Theorem C.

We will now give a proof of L1. The most straightforward way is to show first that for any x say x_0 and any $\epsilon > 0$ we can replace V_{x_0} by a collection consisting of only one gamble $v_0 \in V_{x_0}$ in such a way that if we call the F function for the new system ${}_1F$, ${}_1F(x) \geqq (1-\epsilon)F(x)$ for all x. To do this note that there must be some s with the property that $a > (1-\epsilon)(1-b)F(x_0)$ where b is the probability of returning to x_0 starting at x_0 and using s and a is the amount we win before returning to x_0. We can let v_0 be the first gamble we use when we start at x_0 and use s. We therefore have ${}_1F(x_0) \geqq (1-\epsilon)F(x_0)$ and it is easy to see that ${}_1F(x) \geqq (1-\epsilon)F(x)$ for all x (stop when we hit x_0 and continue by s until the next time we hit x_0 then start with s again, etc.).

We could also try this in the countable case getting ${}_nF(x) > (1 - \sum_{i=1}^{n} \epsilon/2^i)\ F(x)$ but there is trouble when we go to the limit.

3. We will now discuss what happens if $p(x, v, x')$ may be negative. We will have to be more careful in defining $F_s(x)$ (and $F(x)$) and we will have to make more assumptions.

(1) We will have to assume that we can stop whenever we want to (i.e. V_x contains a gamble, v, with support on x and $p(x, v, x) = 0$). To see why consider the following example: X consists of three state a, b, c. If we are at b we must go to a with probability 1 and win nothing. If we are at c we stay there and win nothing. If we are at a go to b with probability $1\text{-}1/n$ and win nothing and go to c with probability $1/n$ and lose 1 dollar. For any reasonable definition of F_s we have $F(a) = 0$ and $F_s(a) = -1$ for any stationary s.

(2) It is no longer reasonable to look for a stationary strategy s that is good in a percentage sense since $F(x)$ may be 0. We will therefore aim for an s such that $F(x) - F_s(x) \geqq \epsilon$.

(3) Because of 2 and Blackwell's example we will have to make some boundedness assumptions.

We shall restrict ourselves to strategies that stop with probability 1 (this seems reasonable because of (1)). We will then define $F_s(x)$ and $F(x)$ as before. We must however change the definition of stationary to be stationary in the old sense except that we are allowed to stop and the decision to stop will depend only on the state we are in and the sum of the p's.

PROPOSITION A. *If we assume that X is countable and both* $\sup_s F_s(x)$ *and* $\inf_s F_s(x)$ *are bounded* (sup *and* inf *are taken over all strategies that terminate with probability* 1), *then given $\epsilon > 0$ we can find a stationary s such that $F(x) - F_s(x) < \epsilon$ for all x.*

The proof is so similar to that of Theorem C that it will be omitted.

We will now consider the case where an optimal strategy happens to exist.

PROPOSITION B. *Suppose we have a system of the same kind as that discussed in Theorem* C *except that X may be uncountable. Suppose also that there is an s such that* (1) $F_s(x) = F(x)$. *Then there is a stationary s satisfying* (1).

Before starting the proof of Proposition B, I would like to point out that the techniques that we will use have considerable overlap with those used by William Sudderth in his thesis (Berkeley 1967), and in some unpublished work.

A good way to introduce some of the relevent ideas in proving Proposition B will be to give another proof of L1 of §2.

ALTERNATIVE PROOF OF L1. There is a theorem of Blackwell that (specialized to our case) says that if we discount our winnings by multiplying the amount won on the nth day by β^n $(0 < \beta < 1)$ and if we define $F_s^\beta(x)$ as the expected discounted amount won starting at x

and using s and define $F^\beta(x)$ as $\sup_s F_s^\beta(x)$ then we can find a stationary strategy s such that $F_s^\beta(x) \geqq (1-\epsilon)F^\beta(x)$. L1 follows by choosing β close enough to 1 so that $F^\beta(x) \geqq (1-\epsilon)F(x)$ for all x. [To prove Blackwell's theorem (in our special case) we chose a large number M depending on β.

We then chose a gamble v_x at each x such that for some strategy s

$$\sum_{y \in X} F_s^\beta(y)v_x(y) + \sum_{y \in X} p(x, v_x, y)v_x(y) > \left(1 - \frac{\epsilon}{M}\right)F^\beta(x).$$

Call this strategy t. There is another strategy t' which consists of using t, M times and then some other s such that $F_{t'}^\beta(x) \geqq (1-2\epsilon)F^\beta(x)$. However if M were chosen large enough the amount won after time M will be very small and hence, t is our desired stationary strategy.]

PROOF OF PROPOSITION B.

(1) We can assume that each V_x contains only gambles that are part of some optimal strategy.

(2) It is easy to see that if there is an optimal strategy and a stationary strategy s such that, $F_s(x) \geqq (1-\epsilon)F(x)$, then s is also optimal. (We are assuming (1).)

(3) We can use Blackwell's theorem to show that we can replace each V_x by a countable subcollection of V_x without changing F or the fact that there exists an optimal strategy.

(4) By (3) each x is contained in a countable closed set (i.e. you can never leave the set) and by (2) we can pick an optimal stationary strategy for this set.

(5) If we have a simply ordered family of closed sets and an optimal stationary strategy on each agreeing on intersections then the union has an optimal stationary strategy. We therefore have a maximal closed set E that has an optimal stationary strategy. If $E \neq X$ then pick an E' that is closed $E' \supset E$ and $E' - E$ is countable. We can then apply our countable result to E' getting a contradiction.

REFERENCES

1. David Blackwell, *Positive dynamic programming*, Proc. Fifth Berkeley Sympos. Math. Stat. and Prob., Vol. 1, 1967, pp. 415–418.

2. ———, *D is counted dynamic programming*, Ann. Math. Statist. **36** (1965), 226–235.

3. L. E. Dubins and L. J. Savage, *How to gamble if you must*, McGraw-Hill, New York, 1965.

STANFORD UNIVERSITY