NOTES

MEMORYLESS STRATEGIES IN FINITE-STAGE DYNAMIC PROGRAMMING[1]

By David Blackwell

University of California, Berkeley

Given three sets X, Y, A and a bounded function u on $Y \times A$, suppose that we are to observe a point $(x, y) \varepsilon X \times Y$ and then select any point a we please from A, after which we receive an income $u(y, a)$. In trying to maximize our income, is there any point to letting our choice of a depend on x as well as on y? We shall give a formalization to this question in which sometimes there is a point. If (x, y) is selected according to a known distribution Q, however, we show that dependence on x is pointless, and apply the result to obtain memoryless strategies in finite-stage dynamic programming problems.

We suppose that X, Y, A are Borel sets in Euclidean spaces and that u is bounded and Borel measurable. A *strategy* σ is a Borel measurable map of $X \times Y$ into A: $\sigma(x, y)$ is the a selected by σ when (x, y) is observed. The *income from* σ is the function I_σ on $X \times Y$: $I_\sigma(x, y) = u(y, \sigma(x, y))$. A *memoryless strategy* τ is a Borel measurable function from Y into A; its income is $I_\tau(x, y) = u(y, \tau(y))$. I_τ is defined on $X \times Y$, but depends on y only.

Question 1. Given any σ, is there a τ with $I_\tau \geqq I_\sigma$ for all (x, y)?

If A is finite, the answer is clearly yes: define $v(y) = \max_a u(y, a)$ and choose τ so that $u(y, \tau(y)) = v(y)$. Then, for any σ, $I_\sigma(x, y) \leqq v(y) = I_\tau(x, y)$.

If A is countable, the answer is no, in an uninteresting ϵ sense. Here is an example: $X = \{1 - 1/n, n = 1, 2, \cdots\}$, $Y = \{0\}$, $A = X$, and $u(y, a) = a$. The σ with $\sigma(x, y) = x$ has $I_\sigma(x, 0) = x$, so that $\sup_x I_\sigma(x, 0) = 1$. For any τ, $I_\tau \equiv \tau(0) < 1$, so that there is an x with $I_\sigma(x, 0) > I_\tau(x, 0)$. But for countable A, given any $\epsilon > 0$ (where ϵ can even be a Borel measurable function of y), there is a τ such that, for any σ, $I_\tau > I_\sigma - \epsilon$ for all (x, y): put $v(y) = \sup_a u(y, a)$ and choose τ so that $u(y, \tau(y)) > v(y) - \epsilon$.

Question 2. Given any σ and any $\epsilon > 0$, is there a τ with $I_\tau > I_\sigma - \epsilon$ for all (x, y)? Section 2.16 of [2] implies an affirmative answer with certain additional not very restrictive hypotheses. But here is an example where the answer is no. X is a Borel subset of the unit square $R \times S$ whose projection D on R is not a Borel set. $Y = A =$ unit interval, and u is the indicator of X:

$$u(y, a) = 1, \qquad \text{if } (y, a) \varepsilon X,$$
$$= 0, \qquad \text{if } (y, a) \not\varepsilon X.$$

For the strategy σ: $\sigma(x, y) = s$ for $x = (r, s)$, we have $I_\sigma((r, s), r) = u(r, s) = 1$, so that I_σ is 1 on the subset F of $X \times Y$ consisting of all points $((r, s), y)$ with $y = r$. But for any τ, $I_\tau(x, y) = u(y, \tau(y))$. The projection of $G = \{(x, y): I_\tau(x, y) = 1\}$ on Y is just the y-set $\{u(y, \tau(y)) = 1\}$, which is a Borel subset

Received 18 September 1963; revised 17 December 1963.

[1] Prepared with the partial support of the National Science Foundation, Grant GP-10.

of D, while the projection of F on Y is D itself. Thus F contains points (x, y) not in G. For these points $I_\sigma = 1$ and $I_\tau = 0$.

Here are two ways to avoid the unnatural conclusion that σ's cannot be replaced by τ's.

(1) Do not insist that strategies be Borel measurable. Then with $v(y) = \sup_a u(y, a)$, there is for any $\epsilon > 0$ a τ with $u(y, \tau(y)) > v(y) - \epsilon$ for all y, so that, for any σ,

$$I_\sigma(x, y) - \epsilon \leqq v(y) - \epsilon < I_\tau(x, y) \quad \text{for all} \quad (x, y).$$

Dubins and Savage [2] have found it convenient to admit nonmeasurable strategies in their theory of gambling.

(2) Do not insist that $I_\tau \geqq I_\sigma - \epsilon$ everywhere, but only on a set of Q-probability 1, where Q is some given distribution on $X \times Y$. Part (b) of the theorem below asserts that this can be done.

THEOREM. *Given any σ and any probability distribution Q on $X \times Y$,*

(a) *there is a τ with*

$$I_2 = \int I_\tau(x, y)\, dQ(x, y) \geqq \int I_\sigma(x, y)\, dQ(x, y) = I_1 .$$

(b) *For any $\epsilon > 0$, there is a τ with $I_\tau > I_\sigma - \epsilon$ on a set of Q-probability* 1.

PROOF. (a) Denote by μ the marginal distribution on Y determined by Q, and by $m(\cdot \mid \cdot)$ a version of the conditional distribution on A given Y induced by Q and σ. Thus $m(\cdot \mid y)$ is for each y a probability measure on the Borel sets of A and $m(B \mid \cdot)$ is for each Borel set $B \subset A$ a Borel measurable function of y such that, for every bounded Borel measurable ϕ on $Y \times A$

$$\int \phi(y, \sigma(x, y))\, dQ(x, y) = [\int \phi(y, a)\, dm(a \mid y)]\, d\mu(y).$$

In particular, for $\phi = u$,

$$I_1 = \int [\int u(y, a)\, dm(a \mid y)]\, d\mu(y) = \int h(y)\, d\mu(y).$$

The set D of all (y, a) for which $u(y, a) \geqq h(y)$ has $m(D_y \mid y) > 0$ for all y, so that, from a known result [1], there is a Borel measurable function τ from Y to A whose graph is a subset of D: $u(y, \tau(y)) \geqq h(y)$ for all y. For this τ,

$$I_2 = \int u(y, \tau(y))\, dQ(x, y) = \int u(y, \tau(y))\, d\mu(y) \geqq \int h(y)\, d\mu(y) = I_1 .$$

For (b), we proceed as in (a), but use instead of D the set D_1 of all (y, a) for which $u(y, a) > S(y) - \epsilon$, where $S(y)$, the conditional essential supremum of $u(y, a)$ given y, is defined as the sup of all rational numbers r for which $m(\{a: u(y, a) > r\} \mid y) > 0$. Choosing τ whose graph is in D_1 makes

$$m(\{a: u(y, \tau(y)) > u(y, a) - \epsilon\} \mid y) = 1$$

for all y, which implies $I_\tau > I_\sigma - \epsilon$ with Q-probability 1.

We remark that the same method, using $D \cap D_1$, yields a τ satisfying both (a) and (b).

The theorem enables us, in finite-stage dynamic programming problems, to replace any strategy by a memoryless strategy without loss. We illustrate the idea for a two-stage problem. We are given four Borel sets X, A, Y, B, a function $q(\cdot|\cdot, \cdot)$ such that $q(\cdot | x, a)$ is for each $(x, a) \varepsilon X \times A$ a distribution on the Borel sets of Y and $q(F | \cdot, \cdot)$ is for each Borel subset of Y a Borel function on $X \times A$, two bounded Borel functions, u_1 on $X \times A \times Y$ and u_2 on $X \times B$, and a distribution P on the Borel sets of X. An initial state x of the system is selected according to P. We observe x and choose any $a \varepsilon A$. The system then moves to a state $y \varepsilon Y$, selected according to $q(\cdot | x, a)$. We observe y, then choose $b \varepsilon B$ and receive the income $u(x, a, y, b) = u_1(x, a, y) + u_2(y, b)$. A strategy σ is a pair σ_1, σ_2, where σ_1 maps X into A and σ_2 maps $X \times Y$ into B. σ(with P, q) determines a distribution P_σ on $X \times A \times Y \times B$, and our expected income is $I(\sigma) = \int u dP_\sigma$. A σ is memoryless if σ_2 depends on y only. To replace any $\sigma = (\sigma_1, \sigma_2)$ by a memoryless $\tau = (\tau_1, \tau_2)$ let Q be the distribution on $X \times Y$ determined by P_σ. Note that Q depends on σ_1 only. Our theorem, applied to Q, σ_2, u_2 yields a τ_2 mapping Y into B with

$$\int u_2(y, \tau(y))\, dQ(x, y) \geqq \int u_2(y, \sigma_2(x, y))\, dQ(x, y).$$

Thus, with $\tau = (\sigma_1, \tau_2)$, the above inequality asserts $\int u_2\, dP_\tau \geqq \int u_2\, dP_\sigma$. Since $\int u_1\, dP_\tau = \int u_1\, dP_\sigma$, we conclude $I(\tau) \geqq I(\sigma)$.

I am grateful to L. E. Dubins and L. J. Savage for several helpful comments.

REFERENCES

[1] Blackwell, D. and Ryll-Nardzewski, C. (1963). Non-existence of everywhere proper conditional distributions. *Ann. Math. Statist.* **34** 223–225.

[2] Dubins, L. E. and Savage, L. J. (1963). *How to Gamble If You Must.* (Unpublished.)