stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE BW 7/71 FEBRUARY

A. HORDIJK and H.C. TIJMS

A COUNTEREXAMPLE IN DISCOUNTED DYNAMIC PROGRAMMING

Prepublication

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O), by the Municipality of Amsterdam, by the University of Amsterdam, by the Free University at Amsterdam, and by industries.

A COUNTEREXAMPLE IN DISCOUNTED DYNAMIC PROGRAMMING

A. HORDIJK and H.C. TIJMS *)

1. INTRODUCTION

We are concerned with a dynamic system which at times $t = 0,1,\ldots$ is observed to be in one of a possible number of states. Let I denote the space of all possible states. We assume I to be denumerable. If at time t the system is observed in state i then a decision k must be chosen from a given finite set K_i. Let Y_t and Δ_t, $t = 0,1,\ldots$, denote the sequences of states and decisions.

If the system is in state i at time t and decision k is chosen, then two things happen:

(i) We incur a known cost w_{ik} and

(ii) $P\{Y_{t+1} = j \mid Y_0, \Delta_0,\ldots, Y_t = i, \Delta_t = k\} = q_{ij}(k)$, where the $q_{ij}(k)$'s are known.

Finally there is specified a discount factor α, $0 < \alpha < 1$, so that a unit of value at time $t=n$ has a value of α^n at time $t=0$.

A rule R for controlling the system is a set of non-negative functions $D_k(Y_0,\Delta_0,\ldots,Y_t)$, $k \in K(Y_t)$; $t \geq 0$, where in every case $\sum_k D_k(\cdot) = 1$. As part of a controlling rule, $D_k(Y_0,\Delta_0,\ldots,Y_t)$ is the instruction at time t to make decision k with probability $D_k(Y_0,\Delta_0,\ldots,Y_t)$ if the particular history $Y_0,\Delta_0,\ldots,Y_t$ has occurred.

Let C denote the class of all possible rules. Let C^M denote the class of all memoryless rules, i.e. $D_k(Y_0,\Delta_0,\ldots,Y_t=i) = D_{ik}^{(t)}$ independent of the past history except for the present state. A nonrandomized stationary rule is a memoryless rule for which $D_{ik}^{(t)} = D_{ik}$ independent of t, and in addition $D_{ik} = 1$, or 0 for all i,k.

For any rule $R \in C$ and state $i \in I$, let

$$\psi(i,\alpha,R) = \sum_{t=0}^{\infty} \alpha^t \sum_{j,k} w_{jk}\, P_R\,(Y_t=j,\Delta_t=k \mid Y_0=i) \ ,$$

*) Report BW 7/71 of the Mathematical Centre, Amsterdam.

provided it exists. The quantity $\psi(i,\alpha,R)$ represents the expected total discounted cost when the initial state is i and rule R is used.

We say that a rule $R^* \in C$ is optimal if $\psi(i,\alpha,R^*) \leq \psi(i,\alpha,R)$ for all $R \in C$, $i \in I$.

It is known [1,2] that there exists an optimal nonrandomized stationary rule when the cost function w_{ik} is bounded. We shall show that an optimal rule may not exist if the boundedness condition on $\{w_{ik}\}$ is weakened. The counterexample given in [2] does not show this result, but proves only that an optimal nonrandomized stationary rule may not exist if the cost function w_{ik} is not bounded. In that counterexample the rule R, which makes with probability $1/(2+t)$ decision 2 when in state i_a at time t, is optimal, since $\psi(i_a,\alpha,R) = -\infty$ for all states i_a.

We shall now give our counterexample.

2. COUNTEREXAMPLE

$$I = \{1,1',2,2',\ldots\}\ ,\ K_{i'} = \{1\}\ ,\ K_i = \{1,2\}\ ,\quad i \geq 1,$$

$$q_{i'i'}(1) = q_{i,i+1}(1) = 1\ ,\ q_{ii'}(2) = 1,\quad i \geq 1\ ,$$

$$w_{i'1} = w_{i1} = 0\ ,\ w_{i2} = -(1-\tfrac{1}{i})\alpha^{-i}\ ,\quad i \geq 1\ .$$

Clearly, $\psi(i',\alpha,R) = 0$ for all $i \geq 1$, $R \in C$. Next we shall prove

$$\psi(i,\alpha,R) > -\alpha^{-i} \qquad \text{for all } i \geq 1,\ R \in C, \tag{1}$$

and

$$\inf_{R\in C} \psi(i,\alpha,R) = -\alpha^{-i} \qquad \text{for all } i \geq 1. \tag{2}$$

Since the proof of theorem 2 in [3] holds also for a denumerable state space, for every $i_0 \in I$ and $R_0 \in C$ there exists a $R \in C^M$ such that $P_R(Y_t=i,\Delta_t=k|Y_0=i_0) = P_{R_0}(Y_t=i,\Delta_t=k|Y_0=i_0)$ for every i,k and t. Hence it suffices to prove (1) for $R \in C^M$.

Let rule $R \in C^M$ and state $i \in I$ be fixed. Denote by $P_i(t)$ the probability that R makes decision 1 when in state $i+t$ at time t. If $P_i(t) = 1$ for all $t \geq 0$, then $\psi(i,\alpha,R) = 0 > -\alpha^{-i}$. Suppose now $P_i(t) < 1$ for at least one t. We have

$$\psi(i,\alpha,R) = \sum_{t=0}^{\infty} -\alpha^{t} \{1-P_i(t)\} \prod_{k=0}^{t-1} P_i(k) \quad (1-\frac{1}{i+t})\alpha^{-(i+t)} .$$

Using the identity $\sum_{t=0}^{\infty} \{1-P_i(t)\} \prod_{k=0}^{t-1} P_i(k) = 1 - \prod_{t=0}^{\infty} P_i(t)$, we obtain

$$\psi(i,\alpha,R) > -\alpha^{-i} \sum_{t=0}^{\infty} \{1-P_i(t)\} \prod_{k=0}^{t-1} P_i(k) \geq -\alpha^{-i} .$$

We have now proved relation (1).

If R_n denotes the rule: Make always decision 1 in the states $1,\ldots,n-1$, and make always decision 2 in the states $n,n+1,\ldots,$ then

$$\psi(i,\alpha,R_n) = -\alpha^{n-i} (1-\frac{1}{n})\alpha^{-n} = -\alpha^{-i}(1-\frac{1}{n}) \quad , \quad n \geq i, \ i \geq 1 .$$

This relation together with (1) proves (2). By (1) and (2), no optimal rule exists.

REFERENCES

1. D. BLACKWELL, Discounted dynamic programming, Ann. Math. Statist. 36 (1965), 226-235.

2. C. DERMAN, Markovian sequential control processes - Denumerable state space, J. Math. Anal. Appl. 10 (1965), 295-302.

3. C. DERMAN and R.E. STRAUCH, A note on memoryless rules for controlling sequential control processes, Ann. Math. Statist. 37 (1966), 276-278.