A note on the iterated expectation criterion for discrete dynamic programming

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Department of Mathematics

PROBABILITY THEORY, STATISTICS AND OPERATIONS RESEARCH GROUP

Memorandum COSOR 76-03

A note on the iterated expectation criterion for discrete dynamic programming

by

K.M. van Hee and J.A.E.E. van Nunen

Eindhoven, March 1976

The Netherlands

A note on the iterated expectation criterion for discrete dynamic programming

by

K.M. van Hee and J.A.E.E. van Nunen

1. Introduction

Recently several authors have investigated the discrete dynamic programming model with unbounded rewards. We refer to Harrison (1972), Lippman (1973), (1975), Wessels (1975), Hinderer (1975) and van Nunen and Wessels (1975). In this paper we consider the discrete dynamic programming model with unbounded rewards as treated by Harrison (1973). The aim of this note is to illustrate, first that the conditions as imposed by Harrison are insufficient and secondly how the imperfection can be repaired. We will give a counter example exhibiting the imperfection of Harrison's conditions, whereas we introduce an extended notion of expectation in order to create a framework in which the results of Harrison can be deduced in exactly the same way as in Harrison's paper. The iterated expectation criterion is defined by prescribing the order of summation and integration in computing expectations. The usual notion of expectation is included if absolute convergence is available.

We shall use the notations of Harrison (1972) with a slight modification: as can be done without loss of generality, we assume, for notational convenience, that there is only one action space for all $s \in S$, i.e. $A_s = A$ for all $s \in S$. We note by Π the set of all policies.

We recall here Harrison's conditions on the transition probabilities $p(\cdot \mid \cdot,\cdot)$ and the rewards $r(\cdot,\cdot)$.

1) $\sum_{t \in S} p(t \mid s,a) \mid r(t,f(t))| < \infty$, for all $s \in S$, $a \in A$, and for all (Markov) decision rules $f \in F$.

2) there exists a bound $d > 0$ such that for all $s \in S$, $a \in A$ and $f \in F$

$$\left| \sum_{t \in S} p(t \mid s,a) r(t,f(t)) - r(s,a) \right| \le d .$$

3) there exists a number M such that

$$U(s) - L(s) \le M, \quad \text{for all } s \in S .$$

We remind the definitions of U and L:

$$L(s) := \inf_{a\in A} r(s,a),\ U(s) := \sup_{a\in A} r(s,a), \quad s \in S .$$

Assumption 1) is not sufficient to guarantee (in the usual sense) the existence of the expected reward at time n, given the starting state. This is shown in the simple example below. Hence conditional expectations, such as

$$\mathbb{E}[r(\sigma_{n+1},\alpha_{n+1}) \mid \sigma_1]$$

(see lemma 1 in Harrison (1972)) are not defined property in general.

2. Counter example

Let $\mathbb{N}$ denote the set of positive integers.
The state space $S := (\{-1,0,1\} \times \mathbb{N}) \cup \{0\}$, the action space A consists of only one element, i.e. $A := \{a\}$ and the reward function r is defined by

$$r(0,a) := 0;\ r((i,j),a) := \begin{cases} 0 & \text{if } i = 0 \\ \operatorname{sgn}(i)2^j & \text{if } i \neq 0\ ((i,j) \in S) \end{cases}$$

and the transition probabilities are defined by

$$p((0,j) \mid 0,a) := 2^{-j} \quad \text{if} \quad j \in \mathbb{N},$$

$$p((i,j) \mid (i,j),a) := 1 \text{ if } i \neq 0,\ j \in \mathbb{N},$$

$$p((i,j) \mid (0,j),a) := \tfrac{1}{2} \text{ if } i \neq 0,\ j \in \mathbb{N}.$$

It is obvious that 3) holds and the verification of 1) and 2) is straightforward. But

$$\sum_{t\in S} p^{(2)}(t \mid 0,a)r(t,a)$$

is undefined, since

$$\sum_{t\in S} p^{(2)}(t \mid 0,a)|r(t,a)| = \infty .$$

(Note that $p^{(2)}(t \mid s,a) := \sum_{\ell\in S} p(t \mid \ell,a)p(\ell \mid s,a)$.)

3. The iterated expectation criterion

Throughout this section we fix an arbitrary policy $\pi = (q_1,q_2,q_3,\ldots)$ and some starting state $s \in S$. Note that p, π and s determine a probability ($\mathbb{P}_s^\pi$) for the random process $\{(\sigma_n,\alpha_n),\ n \in \mathbb{N}\}$. All concepts introduced from now are defined w.r.t. these π and s. We consider a slightly modified definition of expectation of a real function g on H_n, by defining first the conditional expectation with respect to h_{n-1}. Let η_n denote the random vector $(\sigma_1,\alpha_1,\ldots,\sigma_n)$.

Definition.

1) The *conditional expectation* of a real function g on H_n, given $\eta_{n-1} = h_{n-1}$ with $h_{n-1} = (s_1,a_1,\ldots,a_{n-2},s_{n-1})$ is

$$\mathbb{E}[g|\eta_{n-1}=h_{n-1}] := \sum_{t\in S}\sum_{a\in A} q_{n-1}(a|h_{n-1})p(t|s_{n-1},a)g(h_{n-1},a,t)$$

if the right hand side converges absolutely.

2) Let G_n be the set of real functions on H_n such that

$$\mathbb{E}[g \mid \eta_{n-1} = h_{n-1}]$$

is defined for all $h_{n-1} \in H_{n-1}$ with $\mathbb{P}_s^\pi[\eta_{n-1} = h_{n-1}] > 0$.

3) Let $g \in G_n$. For $k = n-2,n-3,\ldots,1$ we define recursively

$$\mathbb{E}[g \mid \eta_k = h_k] := \mathbb{E}[\{\mathbb{E}[g \mid \eta_{k+1} = h_{k+1}]\} \mid \eta_k = h_k]$$

if

$$\mathbb{E}[g \mid \eta_{k+1} = h_{k+1}] \in G_{k+1}\ .$$

4) The iterated expectation of $g \in G_n$ is defined by

$$\mathbb{E}[g] := \mathbb{E}[g \mid \sigma_1 = s]$$

if $\mathbb{E}[g \mid \sigma_1 = s]$ is defined.

Remarks

1) If $g(\sigma_1,\alpha_1,\ldots,\sigma_n)$ is integrable w.r.t. $\mathbb{P}_s^\pi$ the usual conditional expectation equals ours.
2) If for $g,\ell \in G_n$ the iterated expectation is defined, it holds that the iterated expectation of $g + \ell$ exists and

$$\mathbb{E}[g + \ell] = \mathbb{E}[g] + \mathbb{E}[\ell] \quad .$$

It is obvious that, for $g \in G_n$ with $g \geq 0$, it holds that $\mathbb{E}[g] \geq 0$ hence the iterated expectation is a positive and linear operator.

Finally we note that the assumptions 1), 2) and 3) guarantee the existence of

$$\mathbb{E}[\sum_{k=1}^{n} \beta^k r(\sigma_k,\alpha_k)] \ .$$

The discounted iterated expected value belonging to π and s is now defined by:

$$v(\pi)(s) := \lim_{n\to\infty} \mathbb{E}[\sum_{k=1}^{n} \beta^k r(\sigma_k,\alpha_k)] \ .$$

Remarks

3) If the state space and the action space are Polish the iterated expectation can be defined analogously.
4) It is easy to see that Harrison's paper is correct with our definition of the iterated expectation.

References

[1] Harrison, J.M. Discrete dynamic programming with unbounded rewards. Ann. Math. Statist. 43 (1972), 636-644.

[2] Hinderer, K. Bounds for stationary finite stage dynamic programs with unbounded reward functions.
Hamburg, Institut für Mathematische Stochastik der Universität Hamburg, June 1975 (report).

[3] Lippman, S.A. Semi-Markov decision processes with unbounded rewards.
Managament Sci. 19 (1973), 717-731.

[4] Lippman, S.A. On dynamic programming with unbounded rewards.
Management Sci. 21 (1975), 1225-1233.

[5] Van Nunen, J.A.E.E. and J. Wessels, A note on dynamic programming with unbounded rewards.
Eindhoven, University of Technology Eindhoven, Dept. of Math., 1975, Memorandum COSOR 75-13.

[6] Wessels, J. Markov programming by successive approximations with respect to weighted supremum norms.
J. Math. Anal. Appl. (to appear).