MDP: Monotone Optimal Policies

When solving an inventory replenishment problem using an MDP model,
knowing that the optimal policy is of the form (s,S) can reduce the computational burden.

That is, if it is optimal to replenish the inventory when the inventory level is n, then it is optimal to replenish when the inventory level is n-1.

Likewise, properties of the optimal policy for equipment replacement & maintenance problems can be used to reduce the computation.

Under certain reasonable assumptions, if it is optimal to replace a machine of age n, then it is optimal to replace an older machine.

Definition: Suppose the states and actions of a MDP are real variables, and let $a_n^*(\bullet)$ be the optimal decision rule as a function of the state when n periods remain.
Then $a_n^*(\bullet)$ is

monotone nondecreasing if

$$a_n^*(s) \le a_n^*(s+\gamma) \text{ for each } \gamma > 0 \text{ and } s \in S$$

monotone nonincreasing if

$$a_n^*(s) \ge a_n^*(s+\gamma) \text{ for each } \gamma > 0 \text{ and } s \in S$$

Example: *equipment replacement* problem, where
s = age of the equipment, and
$A_s = \{0, 1\}$ where *1* indicates "replace", *0* indicates "keep".
The optimal policy is *monotone nondecreasing*,
since
if $a_n^*(s)=1$ (the optimal decision is to replace equipment at age s),
the optimal decision is also "replace" for older pieces of equipment.

Example: *inventory replenishment* problem, where

- s = inventory level and
- $a \in A_s$ is the inventory level after replenishment, i.e., the order-up-to level.

The optimal policy is *monotone nonincreasing*, since if the optimal decision is to order up to level s' when the current level is s, then if the inventory level were greater, one would not order up to a larger quantity. Rather, if the inventory level s exceeds the reorder point, the order-up-to level is also s, while if the inventory level falls below the reorder point, the order-up-to level increases.

Definitions:

- A ***binary decision process*** (BDP) is a Markov decision process with a finite number of states and action set $A_s = \{0, 1\}$ for each $s \in S$.
- A ***control limit*** V_n for a BDP is a quantity such that the optimal action is $a_n(s)=0$ if and only if $s_n \geq V_n$.

Examples:

Equipment replacement, where actions $a_n(s)$ are 0 "keep" and 1 "replace".

Stopping problem, where the actions $a_n(s)$ are 0 "stop" and 1 "continue".

Processer with two rates, where the state is the number of customers waiting for processing and the two actions are "slow" and "fast" processing rates.

Consider the **B**inary **D**ecision **P**rocess where

$$f_n(s) = \min\left\{ c_s^a + \beta \sum_{j \in S} p_{sj}^a f_{n-1}(j) \mid a \in \{0,1\} \right\}$$

Define the function

$$\gamma_i(s,a) = \sum_{j \geq i} p_{sj}^a$$

i.e., the conditional probability that, given action a selected in state s, the next state of the system is i or greater.

Theorem: Suppose

$$0 \le c_{s+1}^1 - c_s^1 \le c_{s+1}^0 - c_s^0 \quad \text{for } s \in S$$

and for each i, the functions $\gamma_i(\bullet,0)$, $\gamma_i(\bullet,1)$ & the difference

$\gamma_i(\bullet,0) - \gamma_i(\bullet,1)$ are all nondecreasing.

Then for each n there is a *control limit* V_n, i.e., an optimal policy is given by

$$a_n = \begin{cases} 0 \text{ if } s_n < V_n \\ 1 \text{ if } s_n \ge V_n \end{cases}$$

Reference: Daniel P. Heyman & Matthew J. Sobel, *Stochastic Models in Operations Research, Volume II: Stochastic Optimization*, McGraw-Hill Book Company, 1984, page 387.

Interpretation:

The condition $0 \le c_{s+1}^1 - c_s^1 \le c_{s+1}^0 - c_s^0$ states that the single-period cost increases as s increases,

but that the increment in cost is greater if $a = 0$ than if $a = 1$.

The condition that $\gamma_i(\bullet,0)$ and $\gamma_i(\bullet,1)$ are nondecreasing means that, regardless of the action, large values of s_n tend to be followed by large values of s_{n+1},

while the condition that $\gamma_i(\bullet,0) - \gamma_i(\bullet,1)$ is nondecreasing means that this tendency is greater if $a = 0$ than if $a = 1$.

Equipment Replacement

Example:

Equipment replacement problem, where state s = *age* of the equipment, and actions are:

$a = 1$ denotes "replacement" and $a = 0$ denotes "keep".

Consider first the *deterministic* problem with no failures, i.e.,

$$p_{s,s+1}^0 = p_{s,0}^1 = 1,$$

so that

$\gamma_i(s,0) = 0$ for s ≤ $i - 1$ and $\gamma_i(s,0) = 1$ for s > $i - 1$.

Furthermore, $\gamma_0(s,1) = 1$ for all s, and $\gamma_i(s,1) = 0$ for i ≥ 1.

It follows that the *conditions upon the function γ in the theorem are satisfied.*

Equipment Replacement

Suppose that c_s^0 is the operating cost at age s and that c_s^1 has the form $(r - L_s)$ where r is the cost of the new piece of equipment and L_s is the salvage value of the replaced equipment.

Then the conditions of the theorem require that

$$0 \le L_s - L_{s+1} \le c_{s+1}^0 - c_s^0$$

that is, operating cost increases with age while salvage value decreases, with operating costs rising with age at a rate at least as great as the reduction in salvage value.

Given these reasonable assumptions, the theorem implies existence of an optimal control limit, i.e., it is optimal to replace the equipment when it exceeds a certain age.

Equipment Replacement

Consider now the more realistic case in which *random failures* may occur: b_s is the *probability of failure* of a machine of age s. Then

$$p^0_{s,s+1} = 1 - b_s \;\&\; p^0_{s,0} = b_s$$

since failing units are immediately replaced, so that

$$\gamma_i(s,0) = \begin{cases} 0 & \text{if } s+1 < i \\ 1-b_s & \text{if } 0 < i \le s+1 \\ 1 & \text{if } i = 0 \end{cases}$$

which is not necessarily *nondecreasing* in *s* as is required to apply the theorem.

Equipment Replacement

However, "*increasing failure rate*", i.e., $b_s \le b_{s+1}$ is often a valid assumption, and in order to apply the theorem we re-define the states:

Delete state 0 and let state ∞ denote the "highest" state such that breakdowns cause a transition into this state, from which replacement in mandatory, i.e., $A_\infty = \{1\}$.

Equipment Replacement

For all s, let $p^1_{s,\infty} = b_0$ denote the probability that a new item breaks down.

Then $p^1_{s,\infty} = b_s$ and $p^0_{s,s+1} = 1 - b_s$ for s≥1, s≠∞, so that

$$\gamma_i(s,0) = 1 \quad \text{if } i \le s+1$$

and

$$\gamma_i(s,0) = b_s \quad \text{if } i > s+1$$

Therefore, $\gamma_i(\bullet,0)$ is nondecreasing if failure rates are nondecreasing ($b_s \le b_{s+1}$ for all s ≠∞).

Similarly, $\gamma_i(s,1)$ is a constant with respect to s, and so the conditions of the theorem relating to γ are satisfied.

Equipment Replacement

Let c^1_s be defined as before for s≠∞ and $c^1_\infty = r - L_\infty$.

Also, we should increase expected costs under action 0 ("keep") to include the expected replacement costs, i.e.,

$$c^0_s = b_s\left[r(1-\beta) - K_s'\right] + (1-b_s)K_s$$

where K_s' is the salvage value of a broken-down item and K_s is the operating cost of one that has not broken down.

Equipment Replacement

Then $\{c_s^k\}$ is nondecreasing in s if

$$L_{s+1} \le L_s,$$
$$b_s \le b_{s+1},$$
$$(1-b_s)K_s \le (1-b_{s+1})K_{s+1},$$
$$b_s K'_s \ge b_{s+1} K'_{s+1}, \text{ and}$$
$$L_\infty \le L_s$$

for all s.

Under these conditions, the theorem implies an *optimal control limit policy*.

Equipment Replacement

Suppose that the states are defined not by chronological age, but by condition, so that $p_{sj}^0 > 0$ for many j (not merely for j = s+1 and j = ∞).

In order for the conditions of the theorem to hold, it is necessary that

$$\sum_{j \ge i} p_{sj}^0$$

be nondecreasing in s for each i.

If $c_s^1 = r - L_s$, then a ***control limit policy is optimal***.