Bull. Soc. Royale Sciences Liège, 33e année, nº 9-10, 1964, pp. 520-529

ON THE THEORY OF STOCHASTIC CONTROL PROCESSES (*)

par MARTIN J. BECKMANN
Professeur à l'Institut d'Économétrie et de Recherche opérationelle de l'Université de Bonn
Professeur à la Brown University — Providence, U.S.A.

In this paper we consider sequential decision processes of the following type : A system undergoes changes of state at random intervals. Immediately after such transitions a decision may be made to place the system instantaneously in another state, or to leave the state unchanged. Associated with decisions and transitions, as well as with the system's resting in the various states are payoffs. The object is to determine decision rules which will maximize the average payoff per unit time. Decision processes of this type may be called stochastic control processes. They arise for instance in connection with the maintenance and replacement of equipment and in the control of inventory and production. We shall consider one inventory model of fairly general structure which may be described as follows (**) : Suppose that demand for a product arises at time intervals whose length is a random variable subject to a given but arbitrary distribution. The problem of the optimal inventory policy has been solved [1].

Under the following additional assumptions :

1) Demands at different times are stochasticaly independent, but on each occasion demand may depend on the length of the time interval since the last demand.

2) Delivery times are fixed and different from zero. No demand is lost during stock-outs.

(*) Présenté par L. Derwidué, le 17 septembre 1964.

(**) An earlier version of this model in terms of discounted cost was studied in [1]. I am indebted to Prof. de Ghellinck for the suggestion to attempt an approach in average cost terms. (Cf. also [2].)

3) Costs of storage and shortage are proportional to storage and shortage respectively but with different coefficients of proportionallity.

4) The cost of ordering consists of a fixed part and a part proportional to the size of the order.

5) Future costs are discounted at a constant exponentional discount rate.

In this paper an alternative method of solution is developed which is moreover applicable to a larger class of sequential decision problems under risk. It is based on the minimization of average cost rather than of discounted expected costs and is thus more closely related to the approach of Howard [2]. However the following characteristic difference arises. In Howard's schema, the decision k influences the transition probabilities p_{ij}^k from i at the next transition. In the inventory problem it is more natural to say that the decision brings about an immediate change in the state i. The difference in the timing of the change — and of the cost — is not immaterial when either average cost per unit time or discounted total cost is considered — it is irrelevant only when the undiscounted sum of cost is under consideration. Formally, however it will turn out that the resulting equations are closely similar to those of Howard.

Section 1 presents the general model, section 2 its application to the inventory problem.

1. — Consider a finite set S of states i. The probabilities of a transition from i to j depends jointly on state j and the time t since the last transition

$$q_{ij}(t)$$

but is independent of all previous events. The transition times are thus « regeneration points » in which the process is Markovian.

Immediately after a transition the state i may be changed to a state in a set S_i. The following payoffs (or costs, when negative) arise

$a_i(t)\,dt$ payoff during a small time interval dt when the system is in state i at a time t after the last transition.

$a_{ij}(t)$ payoff due to a transition from i to j which takes place after an interval t since the last transition.

b_{ik} payoff due to a decision which changes the state instantaneously from i to k; of course $b_{ii} = 0$.

s_i a terminal payoff when the system is terminated in state i.

Let $k = \delta(i)$ denote a decision rule. We shall assume that the set of possible actions in the various states are such that every state can be reached from every other state with a positive probability if a suitable decision rule is assumed. This property shall be called ergodicity. It is a natural extension of the concept of an ergodic Markov chain.

In an ergodic system a decision rule now defines a recurrent chain of states. Consider now sequences of n decisions under a fixed decision rule. The total payoff $V_n(i)$ is then recursively defined by the following relationship

$$V_0(i) = s_i. \tag{1}$$

$$V_n(i) = b_{ik} + \sum_j \int_0^\infty p_{kj}(t) \cdot \left[\int_0^t a_k(\tau)\, d\tau + a_{kj} + V_{n-1}(j) \right] dt \tag{2}$$

$$n = 1, \ldots$$

where $k = \delta(i)$ is given by the decision rule.

Define

$$p_{ij} = \int_0^\infty p_{ij}(t)\, dt$$

$$g_{ik} = b_{ik} + \sum_j \int_0^\infty p_{kj}(t) \left[a_{kj} + \int_0^t a_k(\tau)\, d\tau \right] dt.$$

Then (2) assumes the form

$$V_n(i) = g_{ik} + \sum_j p_{kj}\, V_{n-1}(j). \tag{3}$$

In particular for optimal decision rules

$$V_n(i) = \max_\delta \left[g_{i\delta(i)} + \sum_j p_{\delta(i)j} \cdot V_{n-1}(j) \right]$$

$$= \max_k \left[g_{ik} + \sum_j p_{kj}\, V_{n-1}(j) \right]. \tag{4}$$

Equation (4) corresponds to the formula (3.3) of Howard [2] for ergodic Markov chains. The problem has thus been reduced to one in terms of an ordinary, ergodic Markov chain.

Let now $V_n(i)$ be decomposed into an average payoff $\bar{v}$ per decision times the number of decisions n plus a residual term $v_n(i)$.

$$V_n(i) = n \cdot \bar{v} + v_n(i). \tag{5}$$

Upon substitution in (3) and cancellation of terms we have

$$v_n(i) + \bar{v} = g_{ik} + \sum_j p_{kj}\, v_{n-1}(j) \tag{6}$$

where $k = \delta(i)$ is given by the decision rule.

For ergodic systems, as $n \to \infty$ the states and hence the transitions are realised with constant probabilities independent of the initial positions, and limits of $v_n(i)$ must exist which are also independent of the initial values. Then (5) becomes

$$v(i) + \bar{v} = g_{ik} + \sum_j p_{kj}\, v(j) \tag{7}$$

where $v(i)$ denotes the residual term v for an infinite horizon of decisions. An optimal decision rule is now defined as one which maximizes $\bar{v}$.

$$\begin{aligned} \bar{v} &= -v(i) + \operatorname*{Max}_{\delta(i)} \left[g_{i\delta(i)} + \sum_j p_{\delta(i)j} \cdot v(j) \right] \\ &= -v(i) + \operatorname*{Max}_{k} \left[g_{ik} + \sum_j p_{kj}\, v(j) \right]. \end{aligned} \tag{8}$$

Consider the parallel system of equations determining the state probabilities π_i

$$\pi_j = \sum_j \pi_i \cdot p_{\delta(i)j}. \tag{9}$$

Upon multiplication of (7) by π_i and summation one has

$$\sum_i \pi_i\, v(i) + \bar{v} \sum_i \pi_i = \sum_i \pi_i \cdot g_{ik} + \sum_i \pi_i\, p_{kj}\, v(j).$$

Due to equation (9) the outer terms cancel and one has

$$\bar{v} = \sum_i \pi_i \, g_{ik}.$$

Let $\pi_i(\delta)$ denote the state probabilities as functions of the decision rule. An alternative formulation of the minimization problem (8) is therefore

$$\bar{v} = \operatorname*{Min}_{\delta} \sum_i \pi_i(\delta) \, g_{i\delta(i)}. \tag{10}$$

The representation (10) is advantageous, whenever the state probabilities can be obtained conveniently from the decision rule. This will be illustrated in the case of the inventory problem.

2. — The inventory problem [cf. 1] is characterized by transition probabilities

$$p_{ij}(t) = p_{i-j}(t) \tag{1}$$

and costs

$$a_i(t) = a_i \qquad \text{independent of } t$$

$$b_{ik} = \begin{cases} 0 & k = i \\ c & \text{if } k \neq i. \end{cases} \tag{2}$$

Write

$$\int_0^\infty \sum_j p_{ij}(t) \, a_i \cdot t \, dt = f_i.$$

Here a_i represents the expected storage and shortage cost T units of time later conditional on a present inventory level of i ; T is the delivery time ; c is the fixed cost of ordering. It can be shown that the proportional cost of ordering which arises in any case, is irrelevant [1].

Substitute (1) and (2) in (1.2) and (1.3) and obtain

$$v(i) + \bar{v} = \operatorname*{Min}_k [c \cdot \delta_{ik} + f_k + \sum_j p_{k-j} \, v(j)]. \tag{3}$$

Equation (3) corresponds to the Arrow-Harris-Marschak equation of optimum inventory policy [3] ; the latter is, of course, in terms of discounted expected cost.

The well known theorem of Scarf [4] assures that under the assumptions made the optimum decision rule is of the following simple so-called s,S type :

$$k = \begin{cases} i & i > s \\ \mathrm{S} & i \leqq s \end{cases}$$

where s is the reordering point and $\mathrm{S} \geqq s$ is the optimum starting stock. (Under less stringent assumptions the optimum policy may involve multiple reordering points $s_1, s_2, \ldots s_m$. But these cases are not likely to arise in practice.)

In terms of this decision rule the inventory equation is

$$v(i) + \bar{v} = \begin{cases} c + f_{\mathrm{S}} + \sum_{j=1}^{\infty} p_j \, v(\mathrm{S} - j) & i \leqq s \\ f_i + \sum_{j=1}^{\infty} p_j \, v(i - j) & i > s \end{cases} \tag{4}$$

Since the $v(i)$ are arbitrary up to an additive constant we may set $v(\mathrm{S}) = 0$.

Apply now (4) to $v(\mathrm{S})$ repeatedly.

$$v(\mathrm{S}) + \bar{v} = f_{\mathrm{S}} + \sum_{j=0}^{\mathrm{S}-s-1} p_j \, . \, [f_{\mathrm{S}-j} - \bar{v} + \sum_{i=1}^{\mathrm{S}-s-1} p_j \, . \sum_{i=1}^{\infty} p_i \, . \, v(\mathrm{S} - i - j)] + \sum_{j=\mathrm{S}-s}^{\infty} p_j \, . \, [v(\mathrm{S}) + c]. \tag{5}$$

Write $\mathrm{S} - s - 1 = \mathrm{D}$.

Now

$$\sum_{j=1}^{\mathrm{D}} p_j \, . \sum_{i=1}^{\infty} p_i \, v(\mathrm{S} - i - j) =$$

$$\sum_{j=1}^{\mathrm{D}} p_j \sum_{i=1}^{j-s-1} p_i \, v(\mathrm{S} - i - j) + \sum_{i=1}^{\mathrm{D}} p_j \sum_{i=j-s}^{\infty} p_i \, . \, [v(\mathrm{S}) + c]$$

$$= \sum_{j=1}^{\mathrm{D}} p_j^{(2)} \, . \, v(\mathrm{S} - j) + q_2 \, . \, [v(\mathrm{S}) + c] \tag{6}$$

where $p_j^{(2)}$ is the compound distribution of a total demand for j on two occasions and q_2 denotes the probability that the reorder point is first reached after two occasions of demand.

Collecting terms in (6) we have

$$v(S) = f_S + \sum_{j=0}^{D} [p_j + p_j^{(2)}] \cdot f_{S-j} + (q_1 + q_2) \cdot [v(S) + c] - \bar{v} \cdot [1 + \sum_{j=0}^{D} p_j].$$

Continueing in this way we obtain

$$(7) \qquad v(S) = -\bar{v} \cdot \sum_{j=0}^{D} \cdot [p_j^{(0)} + p_j^{(1)} + \dots p_j^{(n-1)}j] + \sum_{j=0}^{D} [p_j^{(0)} + p_j^{(1)} + \dots p_j^{(n)}] \cdot f_{S-j} + (q_1 + q_2 + \dots q_n) \cdot [c + v(S)]$$

where we have put $p_j^{(0)} = \begin{cases} 1 & j = 0 \\ 0 & j > j. \end{cases}$

Now $q_n = \sum_{j=1}^{D} [p_j^{(n-1)} - p_j^{(n)}]$.

It follows that

$$\sum_{n=0}^{N} q_n = \sum_{j=0}^{D} p_j^{(0)} - \sum_{j=0}^{D} p_j^{(N)}.$$

But if on each occasion demand is for at least one unit, the total demand in one D + 1 or more occasions exceeds D so that $p_j^{(N)} = 0$ for N > D and $j = 0, \dots D$. Hence

$$\sum_{n=0}^{\infty} q_n = \sum_{j=0}^{D} p_j^{(0)} = p_0^{(0)} = 1.$$

Letting $n \to \infty$ in (7) the term $v(S)$ therefore cancels out and we have

$$0 = -\bar{v} \cdot \sum_{j=0}^{D} p_j^{(n)} + \sum_{j=0}^{D} p_j^{(n)} f_{S-j} + c$$

or

$$\bar{v} = \frac{c + \sum_{j=0}^{D} \sum_{n=0}^{\infty} p_j^{(n)} f_{S-j}}{\sum_{j=0}^{D} \sum_{n=0}^{\infty} p_j^{(n)}} \tag{8}$$

A straight forward calculation shows that

$$\frac{\sum_{n=0}^{\infty} p_j^{(n)}}{\sum_{j=0}^{D} \sum_{n=0}^{\infty} p_j^{(n)}} \qquad j > 0$$

is the state probability π_{S-j} i. e. the probability of finding the system in state $S - j$ at times immediately after a transition. For $j = 0$, π_S denotes the probability of the system being in state S after a transition followed, if necessary by an order.

From $$q_n = \sum_{p_j=0}^{D} p_j^{(n-1)} - p_j^{(n)}$$

one obtains through summation by parts

$$\sum_{n-1}^{\infty} n \cdot q_n = \sum_{n=0}^{\infty} \sum_{j=0}^{D} p_j^{(n)}.$$

Therefore (8) may also be written

$$\bar{v} = \frac{c + \sum_{n=0}^{\infty} \sum_{j=0}^{D} p_j^{(n)} f_{S-j}}{\sum_{n=0}^{\infty} n q_n}. \tag{9}$$

The numerator of this expression lists all costs during a reordering cycle, the denominator gives the average number of demand intervals between successive stock orders. $\bar{v}$ is therefore average cost per time if the unit of time is chosen as the average interval between demands. (The choice of this or any other time unit of course does not affect the decision problem.) In view of the definition of f_l

$$f_l = \int_0^\infty \sum_j p_j(t) \,.\, t \,.\, a_l \, dt$$

an alternative expression for $\bar{v}$ may be given. Write

$$p(i, t)$$

for the probability that at time t accumulated demand since $t = 0$ is i on the assumption that immediately previous to $t = 0$ there was a demand. Then the probability that the system is still in state zero after time t is the probability that the first demand occured at $t + 0$ or later

$$p(0, t) = \int_{t+0}^\infty \sum_j p_j(t) \, dt.$$

Now
$$\int_0^\infty \sum_j p_j(t) \,.\, t \, dt = -t \int_t^\infty \sum_j p_j(\tau) \, d\tau \Big|_0^\infty + \int_0^\infty \int_t^\infty \sum_j p_j(\tau) \, d\tau \, dt$$

through integration by parts.

$$= \int_0^\infty \int_t^\infty \sum_j p_j(t) \, d\tau \, dt$$

$$= \int_0^\infty p(0, t) \, dt$$

provided that $p_j(t) \to 0$ as $t \to \infty$ with sufficient rapidity.

Now let τ denote the time since the last demand and $t \geqq \tau$ the time since the system was last started in state S. We then have

$$\sum_{n=0}^\infty p_j^{(n)} \,.\, p(0, \tau) = p(j, t).$$

Expression (8) can therefore be rewritten as

$$\bar{v} = \frac{c + \sum_{j=0}^{D} a_j \int_0^{\infty} p(j,t)\,dt}{\sum_{j=0}^{D} \int_0^{\infty} p(j,t)\,dt} \tag{10}$$

This formula expresses average cost in terms of costs a_j associated with states j times their average duration. The cost c of an order is weighed with the reciprocal of the average time between such orders.

The form of this expression suggests that it applies to the even more general case of a stochastic process which is Markovian only at those points where active decisions are taken, i. e. where a state is changed through a decision, provided the costs a_j associated with a state per unit time are independent of time and the costs c of a change of state are independent of the change. The probabilities of finding the system in state j at a time t $p(j,t)$ are then well defined provided time is counted always from the last active decision.

REFERENCES

[1] Martin J. Beckmann, An Inventory Model for Arbitrary Interval and Quantity Distributions of Demand, *Management Science*, Vol. 8, No. 1, 1961, pp. 35-57.

[2] Guy de Ghellinck, Application de la Théorie des Graphes Matrices de Markov et Programmes Dynamiques, *Cahiers du Centre de Recherche Opérationnelle*, Vol. 3, No. 1, 1961.

[3] K. Arrow, T. Harris and J. Marschak, Optimal Inventory Policy, *Econometrica*, Vol. XIX, 1951, pp. 250-272.

[4] H. Scarf, The Optimality of (S,s) Policies in the Dynamic Inventory Problem, *Mathematical Methods in the Social Sciences*, 1959, Stanford University Press, 1960, pp. 196-202.