Journal of Polish Safety and Reliability Association
Summer Safety and Reliability Seminars, Volume 6, Number 1, 2015

Grabski Franciszek
Naval University, Gdynia, Poland

Decision problem for a finite states change of semi-Markov process

Keywords

reliability, semi-Markov decision processes, optimization, Howard algorithm, linear programming

Abstract

In the paper there are presented basic concepts and some results of the theory of semi-Markov decision processes. The algorithm of optimization a SM decision process with a finite number of state changes is discussed here. The algorithm is based on a dynamic programming method. To clarify it the SM decision model for the maintenance operation is shown.

1. Introduction

Semi-Markov decision processes theory delivers methods which give the opportunity to control an operation processes of the systems. In such kind of problems we choose the most rewarding process among some alternatives available for the operation.
The problem is solved by the algorithm which is based on a dynamic programming method. Semi-Markov decision processes theory was developed by Jewell [8], Howard [5, 6, 7], Main and Osaki [11], Gercbakh [1, 2]. Those processes are also discussed in [3] and [4].

2. Semi-Markov decision processes

Semi-Markov (SM) decision process is a SM process with a finite states space $S = \{1,\dots, N\}$ such that its trajectory depends on decisions which are made at an initial instant and at the moments of the state changes. We assume that a set of decision in each state i, denoting by D_i, is finite. To take a decision $k \in D_i$, means to select k-th row among the alternating rows of the semi-Markov kernels.

$$\{Q_{ij}^{(k)}(t): t \geq 0, k \in D_i, i, j \in S\},$$

where

$$Q_{ij}^{(k)}(t) = p_{ij}^{(k)} F_{ij}^{(k)}(t). \tag{1}$$

If an initial state is i and a decision (alternative) $k \in D_i$ is chosen at initial moment then there is determined a probabilistic mechanism of a the first change of the state and the evolution of the system on the interval $[0, \tau_1^{(k)})$. The mechanism is defined by a transition probability (1). The decision $k \in D_i$ at some instant $\tau_n^{(k)}$ determines the evolution of the system on the interval $[\tau_n^{(k)}, \tau_{n+1}^{(k)})$. More precisely, the decision $\delta_i(n) = k \in D_i$ means, that according to the distribution $(p_{ij}^{(k)} : j \in S)$, there is selected a state j for which the process jumps at the moment $\tau_{n+1}^{(k)}$, and the length of the interval $[\tau_n^{(k)}, \tau_{n+1}^{(k)})$. is chosen according to distribution given by the CDF $F_{ij}^{(k)}(t)$.
A sequence of decision at the instant $\tau_n^{(k)}$

$$\delta(n) = (\delta_1(n),\dots,\delta_N(n)) \tag{2}$$

is said to be a *policy* for the stage n. A sequence of polices

$$d = \{\delta(n) : n = 0,1,2,\dots\} \tag{3}$$

is called a *strategy*.
We assume that the strategy has the Markov property - it means that for every state $i \in S$ a decision $\delta_i(n) \in D_i$ does not depend on the process evolution until the moment $\tau_n^{(k)}$. If $\delta_i(n) = \delta_i$, then it is called a *stationary decision*. This means that the decision does not depend on n. The policy consisting of stationary decisions is called a *stationary policy*. Hence a stationary policy is defined by the sequence

$\delta = (\delta_1, ..., \delta_N)$. Strategy that is a sequence of stationary policies is called a *stationary strategy*.
To formulate the optimization problem we have to introduce the reward structure for the process. We assume that the system which occupies the state *i* when a successor state is *j*, earns a gain (reward) at a rate

$$r_{ij}^{(k)}(x), i, j \in S, k \in D_i$$

at a moment x of the entering state *i* for a decision $k \in D_i$. The function $r_{ij}^{(k)}(x)$ is called the "yield rate" of state *i* at an instant *x* when the successor state is *j* and *k* is a chosen decision [9]. A negative reward at a rate $r_{ij}^{(k)}(x)$ denotes a loss or a cost of that one. A value of a function

$$R_{ij}^{(k)}(t) = \int_0^t r_{ij}^{(k)}(x)\, dx, i, j \in S, k \in D_i, \tag{4}$$

denotes the reward that the system earns by spending a time *t* in a state *i* before making a transition to state *j*, for the decision $k \in D_i$. When the transition from the state *i* to the state *j* for the decision *k* is actually made, the system earns a bonus as a fixed sum. The bonus is denotes by

$$b_{ij}^{(k)}(x), i, j \in S, k \in D_i$$

A number

$$u_i^{(k)} = \sum_{j \in S} \int_0^\infty \left(R_{ij}^{(k)}(t) + b_{ij}^{(k)}(t) \right) dQ_{ij}^{(k)}(t) \tag{5}$$

is an expected value of the gain that is generated by the process in the state *i* at one interval of its realization for the decision $k \in D_i$.
In this paper we suppose that

$$r_{ij}^{(k)}(x) = r_{ij}^{(k)}, k \in D_i, i, j \in S = \{1, ..., 6\}. \tag{6}$$

From (5) we obtain

$$R_{ij}^{(k)}(t) = r_{ij}^{(k)} t, i, j \in S, k \in D_i \tag{7}$$

and

$$u_i^{(k)} = \sum_{j \in S} \left(p_{ij}^{(k)} \left(r_{ij}^{(k)} m_{ij}^{(k)} + b_{ij}^{(k)} \right) \right), \tag{8}$$

where $m_{ij}^{(k)} = E(T_{ij}^{(k)})$ denotes the expectation of the holding time of the state *i* if the successor state is *j*. Moreower we suppose

$$m_{ij}^{(k)} = E(T_i^{(k)}) = m_i^{(k)}, i, j \in S, k \in D_i \tag{9}$$

and

$$b_{ij}^{(k)} = 0, i, j \in S, k \in D_i.$$

Now the equality (5) takes the form

$$u_i^{(k)} = m_i^{(k)} \sum_{j \in S} p_{ij}^{(k)} r_{ij}^{(k)} = m_i^{(k)} r_i^{(k)}. \tag{10}$$

3. Optimization for a finite states change

We formulate the optimization problem of a semi-Markov process for a finite states change. That kind of problem was investigated by Howard [9] .
We denote by $V_i(d_m), i \in S$ the expected value of the gain (reward) that is generated by the process during a time interval $[0, \tau_{m+1})$ under the condition that the initial state is $i \in S$ and a sequence of polices is

$$\{d_n = (\delta_1(\tau_n^{(k)}), ..., \delta_N(\tau_n^{(k)})), n = 0,1,...,m\}$$

$$m = 0,1,... \tag{11}$$

By $V_j(d_{m-1}),\ j \in S$ we denote the expected value of the gain that is generated by the process during a time interval $[\tau_1^{(k)}, \tau_{m+1})$ under the condition that the process has just entered the state $j \in S$ at the moment τ_1 and a sequence of polices

$$\{d_n = (\delta_1(\tau_n^{(k)}), ..., \delta_N(\tau_n^{(k)})), n = 1,...,m\}$$

$$m = 1,... \tag{12}$$

is chosen.
The expected value of the gain during a time interval $[0, \tau_{m+1})$ under the condition that the initial state is $i \in S$ is the sum of expectation of the gain that is generated by the process during an interval $[0, \tau_1^{(k)})$ and the gain that is generated by the process during the time $[\tau_1^{(k)}, \tau_{m+1})$.

$$V_i(d_m) = u_i^{(k)} + \sum_{j\in S} p_{ij}^{(k)} V_j(d_{m-1}),\ i \in S \qquad (13)$$

Substituting (5) in thi s equality we get

$$V_i(d_m) = \sum_{j\in S} \int_0^{\infty} \left(R_{ij}^{(k)}(t) + b_{ij}^{(k)}(t)\right) dQ_{ij}^{(k)}(t)$$

$$+ \sum_{j\in S} p_{ij}^{(k)} V_j(d_{m-1}),\ i \in S$$

The strategy (the sequence of polices) d_m^* is called optimal in gain maximum problem on interval $[\tau_0' \ \tau_{m+1})$ for the semi-Markov decision process which starts from state i, if

$$V_i(d_m^*) = \max_{k\in D_i}[V_i(d_m)], \quad i \in S \qquad (15)$$

It means that

$$V_i(d_m^*) \geq V_i(d_m), \quad i \in S \text{ for all strategies } d_m.$$

We can get the optimal strategy by using the dynamic programming technique which uses the Bellman principle of optimality. In our case the principle can be formulated as follows:

B e l l m a n principle of optimality.
Let for any initial state and adopted in this state strategy process move to a new state. If the initial strategy is optimal then its remaining part is also optimal for the process whose initial state is a new state that has been reached at the moment of the first state changes.
This principle allows to obtain an algorithm of computing the optimal strategy. The algorithm is defined by the following formulae

$$V_i(d_n^*) = \max_{k\in D_i}[u_i^{(k)} + \sum_{j\in S} p_{ij}^{(k)} V_j(d_{n-1}^*)], \quad i \in S$$
$$n = 1,2,\ldots,m \qquad (12)$$

$$V_i(d_0^*) = \max_{k\in D_i}[u_i^{(k)}], \quad i \in S \qquad (13)$$

4. Optimization of maintenance operation

To ilustrate and explain presented above problem a simple model of a maitanance operation will be construced. A similar model was discussed in [4]. Semi-Markov maintenance nets were presented by Silvestov in [12].
We start from determining the states of the maintenance operation:
1 – an object waiting for start of the operation
2 - main stage of the maintenance operation
3 - control of a maintenance quality
4 - checking of the object's technical condition
5 - waiting for reuse

The possible state changes of the maintenance process are shown in *Figure 1*.

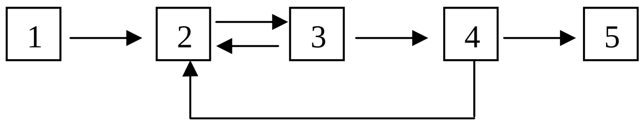

Figure 1. Flow graph of the possible state changes

A semi-Markov process with a set of states $S = \{1, 2, 3, 4, 5\}$ is an appropriate stochastic model of the maintenance operation. To construct semi-Markov decision model in this case, we have to determine the set of decision $D_i, \quad i = 1,\ldots,5$. Assume that

$$D_1 = \{1, 2\}, \quad D_2 = \{1,2\}, \quad D_3 = \{1,2\},$$

$$D_4 = \{1,2\}, \quad D_5 = \{1,2\},$$

where
D_1**:** 1 – long waiting**,**
2 – short waiting,
D_2**:** 1 – normal maintenance,
2 – expensive maintenance,
D_3**:** 1– normal control**,**
2 –expensive control,
D_4**:** 1 – normal checking**,**
2 – expensive checking,
D_5**:** 1 – long waiting for reuse
2 – short waiting for reuse**,**

Semi-Markov decision model is defined by a family of kernels

$$\mathbf{Q^{(k)}}(t) = \begin{bmatrix} 0 & Q_{12}^{(k)}(t) & 0 & 0 & 0 \\ 0 & 0 & Q_{23}^{(k)}(t) & 0 & 0 \\ 0 & Q_{32}^{(k)}(t) & 0 & Q_{34}^{(k)}(t) & 0 \\ 0 & Q_{42}^{(k)}(t) & 0 & 0 & Q_{45}^{(k)}(t) \\ 0 & 0 & 0 & 0 & Q_{55}^{(k)}(t) \end{bmatrix},$$

$$k \in D_i, \quad i = 1,\ldots,5. \qquad (14)$$

Assume that

$$Q_{ij}^{(k)}(t) = p_{ij}^{(k)} G_i^{(k)}(t), \quad i, j \in S, k \in D_i, \qquad (15)$$

where $G_i^{(k)}(t), \quad t \geq 0$ is the cumulative distribution function of a random variable $T_i^{(k)}$, which denotes a waiting time in state i under decision $k \in D_i$.

Moreover, we suppose $b_{ij}^{(k)} = 0, i, j \in S, k \in D_i$.

The transition probability matrix of embedded Markov chain of the SM decision processes is

$$\mathbf{P}^{(\mathbf{k})} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & p_{32}^{(k)} & 0 & p_{34}^{(k)} & 0 \\ 0 & p_{42}^{(k)} & 0 & 0 & p_{45}^{(k)} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \qquad (16)$$

From assumption (15), we have

$$m_{ij}^{(k)} = E(T_{ij}^{(k)}) = E(T_i^{(k)}) = m_i^{(k)},$$

$$i, j \in S, k \in D_i. \qquad (17)$$

Now the equality (8) takes the form

$$u_i^{(k)} = m_i^{(k)} \sum_{j \in S} p_{ij}^{(k)} r_{ij}^{(k)}, i \in S, k \in D_i. \qquad (18)$$

Finally, the algorithm of computing the optimal strategy takes the following form:

Algorithm

1. Compute
   $$u_i^{(k)} = m_i^{(k)} \sum_{j \in S} p_{ij}^{(k)} r_{ij}^{(k)}$$
   for $i \in S = \{1, 2, 3, 4, 5\}, k \in D_i$.

2. Find d_0^* such that
   $$V_i(d_0^*) = \max_{k \in D_i}[u_i^{(k)}], \quad i \in S$$

3. Find d_l^* such that
   $$V_i(d_l^*) = \max_{k \in D_i}[u_i^{(k)} + \sum_{j \in S} p_{ij}^{(k)} V_j(d_{l-1}^*)], \quad i \in S$$
   $$l = 1, 2, \ldots, m-1$$

5. Numerical illustrative example

We determine the numerical data (*Tables 1-3*).

Table 1. Transition probabilities of the maintenance

State i	Decision k	$p_{i1}^{(k)}$	$p_{i2}^{(k)}$	$p_{i3}^{(k)}$	$p_{i4}^{(k)}$	$p_{i5}^{(k)}$
1	1	0	1	0	0	0
	2	0	1	0	0	0
2	1	0	0	1	0	0
	2	0	0	1	0	0
3	1	0	0.08	0	0.92	0
	2	0	0.02	0	0.98	0
4	1	0	0.12	0	0	0.88
	2	0	0.06	0	0	0.94
5	1	0	0	0	0	1
	2	0	0	0	0	1

Table 2. Mean waiting times of maintenance

State i	Decision k	$m_i^{(k)}$ [h]
1	1	0.10
	2	0.02
2	1	1.52
	2	2.05
3	1	0.04
	2	0.08
4	1	0.12
	2	0.05
5	1	0.02
	2	0.01

Table 3. Gain rate of costumer for the maintenance process

State i	Decision k	$r_{i1}^{(k)}$	$r_{i2}^{(k)}$	$r_{i3}^{(k)}$	$r_{i4}^{(k)}$	$r_{i5}^{(k)}$
1	1	0	-20.	0	0	0
	2	0	-20.	0	0	0
2	1	0	0	-40.	0	0
	2	0	0	-60.	0	0
3	1	0	8.	0	-2.0	0
	2	0	20.	0	-10.	0
4	1	0	-4.	0	0	-5.
	2	0	-5.	0	0	-10.
5	1	0	0	0	0	30.
	2	0	0	0	0	50.

According to the formula 1 of the algorithm we compute a "gain" $u_i^{(k)}$, $i = 1,2,...,5$ for the 0 stage. For the data presented in *Tables 1-3* we get:

1. $u_1^{(1)} = -2,\quad u_2^{(1)} = -60.8,\quad u_3^{(1)} = -0.048$
$u_4^{(1)} = -0.4496 \quad u_5^{(1)} = 1.5,$
$u_1^{(2)} = -2,\quad u_2^{(2)} = -123.0,\quad u_3^{(2)} = -0.47.$

From the second point of the algorithm we obtain:

2. $V_1(d_0^*) = -2,\quad V_2(d_0^*) = -60.8,$
$V_3(d_0^*) = -0.048,\quad V_4(d_0^*) = -0.4496,$
$V_5(d_0^*) = 2.5.$

Using recurring formula 3 of the algorithm we get:

Step 1

$$V_1(d_1^*) = -62.8,\quad V_2(d_1^*) = -60.848,$$
$$V_3(d_1^*) = -0.048,\quad V_4(d_1^*) = -0.4496$$
$$V_5(d_1^*) = 2.5,$$
$$\delta_1^* = (2,\ 1,\ 2,\ 2\ \ 2);$$

Step 2

$$V_1(d_2^*) = -62.848,\quad V_2(d_2^*) = -62.9266,$$
$$V_3(d_2^*) = -4.0997,\ V_4(d_2^*) = -0.1148,$$
$$V_5(d_2^*) = 7.5,$$
$$\delta_2^* = (2,\ 1,\ 2,\ 2,\ 2);$$

Step 3

$$V_1(d_3^*) = -64.9266,\quad V_2(d_3^*) = -64.8997,$$
$$V_3(d_3^*) = -1.8411,\quad V_4(d_3^*) = 2.11,\quad V_5(d_3^*) = 10,$$
$$\delta_3^* = (2,\ 1,\ 2,\ 2,\ 2);$$

Step 4

$$V_1(d_4^*) = -66.8997,\quad V_2(d_4^*) = -62.6411,$$
$$V_3(d_4^*) = 0.3002,\quad V_4(d_4^*) = 4.342,$$
$$V_5(d_3^*) = 12.5,$$
$$\delta_3^* = (2,\ 1,\ 2,\ 2\ \ 2);$$

Step 5

$$V_1(d_5^*) = -64.6411,\quad V_2(d_5^*) = -60.4988,$$
$$V_3(d_5^*) = 2.5323,\quad V_4(d_5^*) = 6.8275,$$
$$V_5(d_5^*) = 15,$$
$$\delta_5^* = (2,\ 1,\ 2,\ 2\ \ 2);$$

Step 6

$$V_1(d_6^*) = -62.4998,\quad V_2(d_6^*) = -58.2676,$$
$$V_3(d_6^*) = 5.0199,\quad V_4(d_6^*) = 9.306,$$
$$V_5(d_6^*) = 17.5,$$
$$\delta_6^* = (2,\ 1,\ 2,\ 2\ \ 2);$$

Transition matrix for the best strategy of costumer for the Markov chain of the SM maintenance process is

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0.02 & 0 & 0.98 & 0 \\ 0 & 0.04 & 0 & 0 & 0.96 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

For homogenous Markov chains the *n*-step transition probabilities

$$p_{ij}(n) = P(X(\tau_n) = j \mid X(\tau_0) = i)$$

are the elements of the *n*-th powers of the matrix $\mathbf{P}$. The the *n*-step distribution of Markov chain is given by the rule

$$\mathbf{p}(n) = [P(X(\tau_n) = j);\quad j \in S] = \mathbf{p}(0)\mathbf{P}^{\mathbf{n}}$$

In our case for $n = 9$ we have

$$\mathbf{p}(9) = [0.,\ 0.000092,\ 0.000047,\ 0.001514,\ 0.998347]$$

It means that after 9 steps of the maintenance operations will be finished with probability 0.998347.

6. Conclusion

Semi-Markov decision processes theory provides the possibility to formulate and solve the optimization problems that can be modelled by SM processes. In such kind of problems we choose the process that

brings the most profit among some decisions available for the operation. Main concepts of the semi-Markov decion processes theory like: decision (alternative), policy, strategy, gain, criterion function are explained in the paper. The algorithm of optimization a SM decision process with a finite number of state changes is discussed here. The algorithm is based on a dynamic programming method. To clarify it the SM decision model for the maintenance operation is shown.

References

[1] Gertsbakh, I.B. (1969). *Models of preventive service.* Sovetskoe radio. Moscow (in Russian).
[2] Gertsbakh, I.B. (1977). *Models of Preventive Maintenance*. North-Holland, Oxford 1977.
[3] Grabski, F. (1982). *Theory of Semi-Markov Operation Processes* (in Polish). ZN AMW, 75A, Gdynia.
[4] Grabski, F. (2014). *Semi-Markov Processes: Applications in Systems Reliability and Maintenance*. Elsevier; Amsterdam, Boston, Heidelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Sydney.
[5] Howard, R. A. (1960). *Dynamic Programing and Markov Processes*. Cambridge, Massachusetts, MIT Press.
[6] Howard, R.A. (1964). Research of semi-Markovian decision structures. *Journal of Operations Research Society* Japan 6, 163-199.
[7] Howard, R.A. (1971). Dynamic probabilistic system. *Semi-Markow and decision processes*. New York – London – Sydney – Toronto, Wiley, Vol II.
[8] Jewell,W.S. (1963). Markov-renewal Programming. *Operation Research* 11, 938-971.
[9] Korolyuk, V.S. & Turbin, A.F. (1976). *Semi-Markov processes and their applications* (in Russian), Kiev, Naukova Dumka.
[10] Limnios, N & Oprisan, G. (2001) *Semi-Markov Processes and Reliability*. Boston, Birkhauser.
[11] Mine, H. & Osaki, S. (1970). *Markovian Decision Processes*. New York, AEPCI.
[12] Silvestrov, D. C. (1980). *Semi-Markov processes with a discrete state space* (in Russian), Moscow, Sovetskoe Radio.