Grabski Franciszek
Naval University, Gdynia, Poland

Decision problem for infinite duration semi-Markov process

Keywords

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

Abstract

In the paper there are presented basic concepts and some results of the theory of semi-Markov decision processes. The optimization problem for the infinite duration SM process is connsider in the paper. The Howard algoritm which enables to find the optimal stationary strategy is also discussed here. The algorithm is applied in a decision problem concerning the two components renewable series system is. It is also shown that this algorithm is equivalent to the some linear programing problem.

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. We investigate the infinite duration SM decision processes. The Howard algorithm modified by Main and Osaki is applied for finding an optimal stationary policy for the kind of processes. This algorithm is equivalent to the some problem of linear programing. Semi-Markov decision processes theory was developed by Jewell [10], Howard [7]-[9], Main and Osaki [13], Gercbakh [4]. Those processes are also discussed in [5] and [6] .

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

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

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 $\tau_n^{(k)}$

$$\delta(n) = (\delta_1(n),\dots,\delta_N(n)) \quad (2)$$

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

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

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

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

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

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

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

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

From (5) we obtain

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

and

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

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*. Moreover we suppose

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

and

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

Now the equality (5) takes the form

$$u_i^{(k)} = m_i^{(k)} \sum_{j \in S} p_{ij}^{(k)} r_{ij}^{(k)} = m_i^{(k)} r_i^{(k)}. \qquad (6)$$

3. Optimization problem for infinite duration process

We formulate the optimization problem of a semi-Markov process on infinite interval $[0, \infty)$. This problem was investigated by Howard [9] and by Mine and Osaki [13]. It is known as decision problem without discounting.

3.1. Hovard algorithm

We assume that considered semi-Markov decision process with a finite state space $S = \{1, \ldots, N\}$ satisfy assumption of the limiting theorem from [6], [11], [12].
The criterion function

$$g(\delta) = \frac{\sum_{i \in S} \pi_i(\delta) u_i^{(k)}}{\sum_{i \in S} \pi_i(\delta) m_i^{(k)}} = \frac{\sum_{i \in S} \pi_i(\delta) m_i^{(k)} r_i^{(k)}}{\sum_{i \in S} \pi_i(\delta) m_i^{(k)}} \qquad (7)$$

means the gain per unit of time as a result of a long operating system. The numbers $\pi_i(\delta),\ i \in S$, represent the stationary distribution of the embedded Markov chain of the semi-Markov process defined by the kernel

$$Q^{(\delta)}(t) = [Q_{ij}^{(k)}(t) : t \geq 0,\ i, j \in S,\ k \in D_i] \qquad (8)$$

It means that for every decision $k \in D_i$ those probabilities satisfy the following linear system of equations

$$\sum_{i \in S} \pi_i(\delta) p_{ij}^{(k)} = \pi_j(\delta), \ j \in S,$$

$$\sum_{i \in S} \pi_i(\delta) = 1, \quad \pi_j(\delta) > 0, \ j \in S, \tag{9}$$

where

$$p_{ij}^{(k)} = \lim_{t \to \infty} Q_{ij}^{(k)}(t), \ i, j \in S. \tag{10}$$

The number

$$m_i^{(k)} = E(T_i^{(k)}) = \lim_{t \to \infty} \int_0^\infty t \ dG_i^{(k)}(t), \ i \in S, k \in D_i \tag{11}$$

is an expected value of a waiting time in a state *i*, for an decision (alternative) $k \in D$.
A stationary policy δ^* is said to be optimal if it maximize the gain per unit of time:

$$g(\delta^*) = \max_d [g(\delta)]. \tag{12}$$

In [13] it is proved that the optimal stationary strategy there exists. In [9] and [13] there is presented Howard Algorithm which enables to find the optimal stationary strategy. Here we present the Howard Algorithm using our own notation.

The Algorithm

1. *Data*

- Sets of decisions (alternatives)

$$D_i, i \in S = \{1, 2, \ldots, N\},$$

- Set of functions definig the semi-Markov decision processes

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

- Sets of functions that define the unit gains

$$\{r_{ij}^{(k)}(x) : x \geq 0, d_i \in D_i, i, j \in S\},$$

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

2. *Initial calculation procedure*
Compute according to (7), (8) and (9)

$$p_{ij}^{(k)}, m_i^{(k)}, u_i^{(k)}$$

for each decision $k \in D_i$, $i, j \in S$.

3. *Policy Evaluation*
For the present policy $\delta = (\delta_1, \ldots, \delta_N)$, $\delta_i = k \in D_i$ *calculate the gain* $g = g(\delta)$ and *solve the system of linear equation*

$$g \ m_i^{(k)} + w_i = u_i^{(k)} + \sum_{j \in S} p_{ij}^{(k)} w_j, \ i \in S, \tag{13}$$

with $w_N = 0$ and the unknown weights $w_1, w_2, \ldots, w_{N-1}$.

4. *Policy Improvement*
For each state $i \in S$ find the set of decisions (alternatives)

$$\Delta_i = \left\{ k \in D_i : \Gamma_i^{(k)} = \frac{u_i^{(k)} + \sum_{j \in S} p_{ij}^{(k)} w_j - w_i}{m_i^{(k)}} > g(\delta) \right\}. \tag{14}$$

If for each $i \in S$ the set $\Delta_i = \emptyset$ then the policy $\delta = (\delta_1, \ldots, \delta_N)$, is optimal and the strategy corresponding to it is also optimal. If at least there is one state $i \in S$ such that $\Delta_i \neq \emptyset$ then the policy is not optimal and it must be improved. Therefore, substitute the policy $\delta = (\delta_1, \ldots, \delta_N)$, by the policy $\delta' = (\delta'_1, \ldots, \delta'_N)$, where $\delta'_i = \delta_i$ if $\Delta_i = \emptyset$ and $\delta'_i \in D_i$ is any other decision if $\Delta_i \neq \emptyset$. Repeat procedures 3 and 4.
It is proved [13], [5], that $g(\delta') > g(\delta)$ and the optimal decision is achieved after finite number of iterations.

3.2. Linear programing method

Mine and Osaki [13] presented linear programing method for solving the considered problem of optimization.
Let $a_j^{(k)}$ be a probability that in the state $j \in S$ has been taken decision $k \in D_j$. It is obvious that

$$\sum_{k \in D_i} a_j^{(k)} = 1, \ 0 \leq a_j^{(k)} \leq 1, \ \ j \in S. \tag{15}$$

The criterion function (7) can be written as

$$g(\delta)=\frac{\sum_{i\in S}\sum_{k\in D_i}a_j^{(k)}\pi_i(\delta)u_i^{(k)}}{\sum_{i\in S}\sum_{k\in D_i}a_j^{(k)}\pi_i(\delta)m_i^{(k)}} \tag{16}$$

The equations (6) are equivalent to the system of equtions

$$\sum_{i\in S}\sum_{k\in D_i}a_j^{(k)}\pi_i(\delta)p_{ij}^{(k)}=\pi_j(\delta),$$

$$\sum_{i\in S}\pi_i(\delta)=1,\ =\pi_j(\delta)>0,\ \ j\in S, k\in D_i \tag{17}$$

Substituting

$$a_j^{(k)}\pi_i(\delta)=x_j^{(k)}\geq 0,\ \ j\in S,\ \ k\in D_i, \tag{18}$$

and taking into account that

$$\pi_j(\delta)=\sum_{k\in D_i}x_j^{(k)},\ j\in S,, \tag{19}$$

from (16),(17) and (18) we get a following problem of a linear programing:

$$\max\left[\frac{\sum_{i\in S}\sum_{k\in D_i}u_i^{(k)}x_j^{(k)}}{\sum_{i\in S}\sum_{k\in D_i}m_i^{(k)}x_j^{(k)}}\right] \tag{20}$$

under constrains

$$\sum_{k\in D_i}x_j^{(k)}-\sum_{i\in S}\sum_{k\in D_i}p_{ij}^{(k)}x_j^{(k)}=0,\ j\in S,$$
$$\sum_{j\in S}\sum_{k\in D_i}x_j^{(k)}=1,\ x_j^{(k)}\geq 0, j\in S, k\in D_i. \tag{21}$$

We define new variable in the following way

$$y_j^{(k)}=\frac{x_j^{(k)}}{\sum_{i\in S}\sum_{k\in D_i}m_i^{(k)}x_j^{(k)}}, \tag{22}$$

$$y=\frac{1}{\sum_{i\in S}\sum_{k\in D_i}m_i^{(k)}x_j^{(k)}}. \tag{23}$$

These substitutions allow to formulate and and prove following theorem [13]:

Theorem. The problem: find δ^* such that

$$g(\delta^*)=\max_d[g(\delta)],$$

where the criterion function $g(\delta)$ is defined by (7), is equivalent to the following problem of linear programing: Let

$$y_j^{(k)}\geq 0,\ j\in S, k\in D_i$$

Find

$$\max_{y_j^{(k)}}\left[\sum_{j\in S}\sum_{k\in D_j}u_i^{(k)}y_j^{(k)}\right]$$

under constraints

$$\sum_{k\in D_j}y_j^{(k)}-\sum_{i\in S}\sum_{k\in D_i}p_{ij}^{(k)}y_j^{(k)}=0,\ j=1,...,N-1,$$

$$\sum_{j\in S}\sum_{k\in D_i}m_i^{(k)}y_j^{(k)}=1.$$

We obtain the optimal policy using the rule

$$a_j^{(k)}=\frac{y_j^{(k)}}{\sum_{k\in D_i}y_j^{(k)}},$$

where $a_j^{(k)}$ denotes a probability that in the state $j\in S$ a decision $k\in D_j$ has been taken.

The best decision in the state $j\in S$ is such alternative $k^*\in D_j$ for which the probability $a_j^{(k^*)}$ is the greatest. The best decisions in states $1, 2,..., N$ form the optimal policy.

4. Decision problem for renewable series system

We assume that a system consists of 2 components which form a series reliability structure. We assume that a lifetime of component k is represented by a random variable ζ_k with exponential PDF

$$f_k(t)=\lambda e^{-\lambda_k t}I_{[0,\infty)}(t).$$

From the structure of the system it follows that the damage of the system takes place if a failure of any component occurs. A damaged component is renewed. We assume that the renewal time of k-th component is a non-negative random variable γ_k

with a CDF

$$H_k(t) = P(\gamma_k \le t), \ k = 1, 2.$$

It is well known that the exponential probability distribution has memoryless property. Therefore the renewal of component means renewal of the whole system. We also assume that the random variables denoting successive times to failure of k-th component and random variables denoting their consecutive renewal times are independent copies of the random variables ζ_k and γ_k accordingly. We suppose that the random variables ζ_1, ζ_2, γ_1, γ_2 are mutually independent. Moreover we assume that γ_1, γ_2 have the positive, finite expected values and variances.

We determine states:

1 – renewal of a first component after its failure,
2 – renewal of a second component after its failure,
3 – work of the "up" system.

The "down" states are represented by a set $A = \{1, 2\}$ while the "up" state is represented by one element set $A' = \{3\}$. We assume that

$$D_1 = \{1, 2\}, \ D_2 = \{1, 2\}, \ D_3 = \{1, 2, 3, 4\}$$

are the sets of decisions (alternatives) for the states 1, 2, 3.

D_1: 1 – normal reneval of a first component
2 – expensive reneval of a first component

D_2: 1 – normal reneval of a first component
2 – expensive reneval of a first component

D_3: 1 – normal reliability of a first and second component
2 – normal reliability of a first component and higher of a second component
3 – higherer reliability of a first component and normal of a second component
4 – higher reliability of the both components

The semi-Markov decision model is determined by the family of kernels

$$Q^{(\delta)}(t) = \begin{bmatrix} 0 & 0 & H_1^{(k)}(t) \\ 0 & 0 & H_2^{(k)}(t) \\ \frac{\lambda_1^{(k)}}{\Lambda^{(k)}}(1 - e^{-\Lambda^{(k)}t}) & \frac{\lambda_2^{(k)}}{\Lambda^{(k)}}(1 - e^{-\Lambda^{(k)}t}) & 0 \end{bmatrix}$$

where

$$\Delta^{(k)} = \lambda_1^{(k_1)} + \lambda_2^{(k_2)}, \ k \in D_3, \ k_1 = 1, 2, \ k_2 = 1, 2.$$

Assume that

$$H_1^{(k)}(t) = 1 - \left(1 + \alpha_1^{(k)} t\right) e^{-\alpha_1^{(k)}(t)}, \ t \ge 0, \ k \in D_1,$$

$$H_2^{(k)}(t) = 1 - \left(1 + \alpha_2^{(k)} t\right) e^{-\alpha_2^{(k)}(t)}, \ t \ge 0, \ k \in D_2,$$

$$F_1(t) = 1 - e^{-\lambda_1^{(k_1)} t}, \ t \ge 0, \ k_1 = 1, 2,$$

$$F_2(t) = 1 - e^{-\lambda_2^{(k_2)} t}, \ t \ge 0, \ k_2 = 1, 2,$$

where

$$D_3 = \{1, 2, 3, 4\} = \{(1,1), (1, 2), (2, 1), (2, 2)\}.$$

is the set of alternative for the state 3.

For $k = 1 \in D_3$ $\lambda_1^{(k)} = \lambda_1^{(1)}$, $\lambda_2^{(k)} = \lambda_2^{(1)}$,
for $k = 2 \in D_3$ $\lambda_1^{(k)} = \lambda_1^{(1)}$, $\lambda_2^{(k)} = \lambda_2^{(2)}$,
for $k = 3 \in D_3$ $\lambda_1^{(k)} = \lambda_1^{(2)}$, $\lambda_2^{(k)} = \lambda_2^{(1)}$,
for $k = 4 \in D_3$ $\lambda_1^{(k)} = \lambda_1^{(2)}$, $\lambda_2^{(k)} = \lambda_2^{(2)}$.

The matrix of transition probabilities of embedded Markov chain corresponding to the kernel (8) has the form

$$P^{(\delta)} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ \frac{\lambda_1^{(k)}}{\Lambda^{(k)}} & \frac{\lambda_2^{(k_1)}}{\Lambda^{(k)}} & 0 \end{bmatrix}$$

In this case the solution of the system of equations (9) takes the form

$$\pi_1(\delta) = \frac{\lambda_1^{(k_1)}}{2\Lambda^{(k)}}, \ \pi_2(\delta) = \frac{\lambda_2^{(k_2)}}{2\Lambda^{(k)}}, \ \pi_3(\delta) = \frac{1}{2}.$$

The expected values of waiting times in states $i \in S$ for decisions $k \in D_i$ are

$$m_1^{(k)} = \frac{2}{\alpha_1^{(k)}} \ \text{for} \ k \in D_1 = \{1, 2\},$$

$$m_2^{(k)} = \frac{2}{\alpha_2^{(k)}} \ \text{for} \ k \in D_2 = \{1, 2\},$$

$$m_3^{(k)} = \frac{1}{\Lambda^{(k)}} \ \text{for}\ k \in D_3 = \{1, 2, 3, 4\}.$$

In this case the criterion function (7) takes the form

$$g(\delta) = \frac{\left(\lambda_1^{(k)} u_1^{(k)} + \lambda_2^{(k)} u_2^{(k)} + \Lambda^{(k)} u_3^{(k)}\right)\alpha_1^{(k)} \alpha_2^{(k)}}{2\lambda_1^{(k)} \alpha_2^{(k)} + 2\lambda_2^{(k)} \alpha_1^{(k)} + \alpha_1^{(k)} \alpha_2^{(k)}}$$

We suppose that

$$r_{ij}^{(k)}(x) = r_{ij}^{(k)},\ k \in D_i,\ i, j \in S = \{1, 2, 3\}$$

and

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

We determine the numerical data.
Parameters of CDF's for alternatives $k \in D_1$:

$$\alpha_1^{(1)} = 0.2,\ \alpha_1^{(2)} = 0.125.$$

Parameters of CDF's for alternatives $k \in D_2$:

$$\alpha_2^{(1)} = 0.25,\ \alpha_2^{(2)} = 0.2.$$

Parameters of CDF's for alternatives $k \in D_3$:

$$\lambda_1^{(1)} = 0.008,\ \lambda_2^{(1)} = 0.009,\ \Lambda^{(1)} = 0.017,$$

$$\lambda_1^{(2)} = 0.008,\ \lambda_2^{(2)} = 0.006,\ \Lambda^{(2)} = 0.014,$$

$$\lambda_1^{(3)} = 0.004,\ \lambda_2^{(3)} = 0.009,\ \Lambda^{(3)} = 0.013,$$

$$\lambda_1^{(4)} = 0.004,\ \lambda_2^{(4)} = 0.006,\ \Lambda^{(4)} = 0.010.$$

Table 1. The transition probabilities and the mean waiting times for the process

State i	Decision k	$p_{i_1}^{(k)}$	$p_{i_2}^{(k)}$	$p_{i_3}^{(k)}$	$m_i^{(k)}$
1	1	0	0	1	5
	2	0	0	1	8
2	1	0	0	1	4
	2	0	0	1	5
3	1	0.47	0.53	0	58.82
	2	0.57	0.43	0	71.43
	3	0.31	0.69	0	76.92
	4	0.40	0.60	0	100.0

Table 2. The gain rate for the process

State i	Decision k	$r_{i_1}^{(k)}$	$r_{i_2}^{(k)}$	$r_{i_3}^{(k)}$	$u_i^{(k)}$
1	1	0	0	-54	-270
	2	0	0	-62	-496
2	1	0	0	-58	-232
	2	0	0	-64	-320
3	1	21	24	0	1328.74
	2	21	28	0	1715.03
	3	25	24	0	1869.93
	4	25	28	0	2680.00

We have all data to start the iteration cycle of the Howard and Main & Osaki algorithm. Let $\delta = (1,1,1)$ be the initial policy. Now the rule has the form

$$g((1,1,1)) = \frac{\left(\lambda_1^{(1)} u_1^{(1)} + \lambda_2^{(1)} u_2^{(1)} + \Lambda^{(1)} u_3^{(1)}\right)\alpha_1^{(k)} \alpha_2^{(1)}}{2\lambda_1^{(1)} \alpha_2^{(k)} + 2\lambda_2^{(1)} \alpha_1^{(1)} + \alpha_1^{(1)} \alpha_2^{(1)}}$$

Using the equality we calculate the gain $g = g((1,1,1))$. For this gain we solve the system of equations(13). As the result we get the weights w_1 and w_2. The solution is determined by the rules

$$w_1 = \frac{g(p_{32}^{(k)} m_2^{(k)} + m_3^{(k)})}{p_{21}^{(k)} p_{32}^{(k)} + p_{31}^{(k)}},\ w_2 = \frac{g(p_{21}^{(k)} m_3^{(k)} - p_{31}^{(k)} m_2^{(k)})}{p_{21}^{(k)} p_{32}^{(k)} + p_{31}^{(k)}}.$$

Substituting the appropriate numerical values we obtain

$$g = -5.7064,\ w_1 = 688.367,\ w_2 = 22.8242,$$

$$\Gamma_1^{(1)} = -191.673,\ \Gamma_1^{(2)} = -148.046,$$

$$\Gamma_2^{(1)} = -63.706,\ \Gamma_2^{(2)} = -68.5648,$$

$$\Gamma_3^{(1)} = 28.296,\ \Gamma_3^{(2)} = 29.6404,$$

$$\Gamma_3^{(3)} = 27.289,\ \Gamma_3^{(4)} = 29.6904.$$

Hence

$$\Delta_1 = \emptyset,\ \Delta_2 = \emptyset,\ \Delta_3 = \{1, 2, 3, 4\}.$$

According to the algorithm we substitute the policy (1, 1, 1) by the policy (2, 1, 4) and we repeat procedures. Now we get

$$g = -27.3125,\ w_1 = -6664.25,\ w_2 = -109.25,$$

$\Gamma_1^{(1)} = 1278.85,\ \Gamma_1^{(2)} = 771.0312,$

$\Gamma_2^{(1)} = -30.6875,\ \Gamma_2^{(2)} = -42.15,$

$\Gamma_3^{(1)} = -31.645,\ \Gamma_3^{(2)} = -29.8274,$

$\Gamma_3^{(3)} = -3.52795,\ \Gamma_3^{(4)} = -0.5125.$

From here

$\Delta_1 = \{1, 2\},\ \Delta_2 = \varnothing,\ \Delta_3 = \varnothing.$

Therefore in a next step we substitute the policy (2, 1, 4) by the policy (2, 2, 4). In this case we have

$g = 27.7474,\ w_1 = -175.661,\ w_2 = -3.62188,$

$\Gamma_1^{(1)} = -18.8678,\ \Gamma_1^{(2)} = -40.0424,$

$\Gamma_2^{(1)} = -57.0945,\ \Gamma_2^{(2)} = -63.2756,$

$\Gamma_3^{(1)} = 21.1537,\ \Gamma_3^{(2)} = 22.5864,$

$\Gamma_3^{(3)} = 23.5696,\ \Gamma_3^{(4)} = 26.0756.$

and

$\Delta_1 = \varnothing,\ \Delta_2 = \varnothing,\ \Delta_3 = \varnothing.$

It means that the policy $\delta^* = (2, 2, 4)$ maximizes the criterion function $g(\delta),\ \delta \in D_1 \times D_2 \times D_3$.

5. 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. The problem requires the use of terms such as decision (alternative), policy, strategy, gain, criterion function. If the semi-Markov process describing the evolution of the real system in a long time satisfies the assumptions of the limit theorem, we can use the results of the infinite duration SM decision processes theory. We can apply the Howard algorithm for finding an optimal stationary policy. This algorithm is equivalent to the some problem of linear programing.

References

[1] Barlow, R.E. & Proschan, F. (1965). *Mathematical theory of reliability*. Wiley, New York, London, Sydney.

[2] Barlow, R.E. & Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing. Probability Models*. Holt Rinehart and Winston, Inc., New York.

[3] Feller, W. (1966). *An Introduction to Probability Theory and Its Applications*, New York, London, Sydney; John Wiley& Sons. Inc, 2.

[4] Gertsbakh, I.B. (1969). *Models of preventive service.* Sovetskoe radio. Moscow (in Russian).

[5] Grabski, F. (1982). *Theory of Semi-Markov Operation Processes* (in Polish). ZN AMW, 75A, Gdynia

[6] Grabski, F. (2014) *Semi-Markov Processes: Applications in Systems Reliability and Maintenance*, Elsevier (in preparation).

[7] Howard, R.A. (1960). *Dynamic Programing and Markov Processes*. Cambridge, Massachusetts, MIT Press.

[8] Howard, R.A. (1964). Research of semi-Markovian decision structures. *Journal of Operations Research Society* Japan 6, 163-199.

[9] Howard, R.A. (1971). Dynamic probabilistic system. *Semi-Markow and decision processes*. New York – London – Sydney – Toronto, Wiley, Vol II.

[10] Jewell, W.S. (1963). Markov-renewal Programming. *Operation Research* 11, 938-971.

[11] Korolyuk, V.S. & Turbin, A.F. (1976). *Semi-Markov processes and their applications* (in Russian), Kiev, Naukova Dumka.

[12] Limnios, N & Oprisan, G. (2001) *Semi-Markov Processes and Reliability*. Boston, Birkhauser.

[13] Mine, H. & Osaki, S. (1970). *Markovian Decision Processes*. New York, AEPCI.

[14] Silvestrov, D. C. (1980). *Semi-Markov processes with a discrete state space* (in Russian), Moscow, Sovetskoe Radio.