Sborník prací Přírodovědecké fakulty University Palackého v Olomouci. Matematika

Pavla Kunderová

On a mean reward from a common Markov replacement process

Sborník prací Přírodovědecké fakulty University Palackého v Olomouci. Matematika, Vol. 19 (1980), No. 1, 39--50

Persistent URL: `http://dml.cz/dmlcz/120089`

1980 — ACTA UNIVERSITATIS PALACKIANAE OLOMUCENSIS
FACULTAS RERUM NATURALIUM — TOM 65

Katedra matematické analýzy a numerické matematiky přírodovědecké fakulty University Palackého v Olomouci

Vedoucí katedry: prof. RNDr. Miroslav Laitoch, CSc.

ON A MEAN REWARD FROM A COMMON MARKOV REPLACEMENT PROCESS

PAVLA KUNDEROVÁ

(*Received March 31, 1979*)

Summary

The object of investigation in this paper is a Markov replacement process with rewards under a common stationary replacement policy as described in [5]. The quality of the replacement policy is characterized by the expected mean reward from the process $\Theta(i)$, $i \in I$, defined in paragraph 2. In Theorem 1 we derive a system of equations (11) for establishing the mean rewards $\Theta(i)$ and there is proved the uniqueness of its solution. A common Howard's iteration method is constructed (see [1]) for finding the optimal stationary replacement policy under which the maximal reward is reached. This paper refers to paragraph 10 in [4], which deals with a mean reward from the controlled Markov chain.

1. Basic definitions and notations

Let a homogeneous Markov process with rewards $\{X_t, t \geqq 0\}$ (see [5]) describing the evolution of a system in state space $I = \{1, 2, \dots, r\}$ be defined by exit intensities $(\mu(1), \dots, \mu(r))$, $0 < \mu(j) \leqq \infty, j = 1, \dots, r$ and by a stochastic matrix $\mathbf{P} = \| p(i,j) \|_{i,j=1}^{r}$, $p(i, i) = 0$, of transition probabilities in the moment of the exit. We constitute a matrix of the so-called transition intensities $\mathbf{M} = \| \mu(i,j) \|_{i,j=1}^{r}$, where $\mu(i,j) = \mu(i)\, p(i,j)$ for $i \neq j$, $\mu(i, i) = -\mu(i)$,

$$-\mu(i, i) = \sum_{j \neq i} \mu(i, j). \tag{1}$$

The system being in state i at time t passes through the infinitesimal interval $(t, t + dt)$ into state j with the probability $\mu(i, j)\, dt$.

Consider a situation, where the development of the process can be influenced by an action called replacement (see [5]). Under a replacement of type $(i, +j)$ we mean the instantaneous shift of the system from state i into state j. The information on the development of the process up to the n-th state change is given by the sequence of states visited

$$i_0, i_1, i_2, \ldots, i_{n-1}, i_n = j, \tag{2}$$

by the corresponding sojourn times

$$t_0, t_1, t_2, \ldots, t_{n-1}, \tag{3}$$

and by the sequence

$$\delta_0, \delta_1, \delta_2, \ldots, \delta_{n-1}, \tag{4}$$

where $\delta_m = 0$ if the system was left i_m without interference and $\delta_m = 1$ if the passage from i_m into i_{m+1} was the result of a replacement.

For the history of the process up to the n-th state change we use the notation

$$\omega_n = [i_0, t_0, \delta_0; i_1, t_1, \delta_1; \ldots; i_{n-1}, t_{n-1}, \delta_{n-1}; i_n],$$

and the complete history of the process is given by a sequence

$$\omega = [i_0, t_0, \delta_0; i_1, t_1, \delta_1; \ldots].$$

A replacement policy (see [5]) is a decision, for all possible sequences (2)–(4) and all states j, on how long the system will be left in j without shifting (maximal sojourn time) and in what state it is to be shifted.

We denote by D the set of couples $(i, +j)$ meaning admissible replacements, $D_i = \{j: (i, +j) \in D\}$.

A stationary replacement policy f is given by a function $f(j)$ defined on a subset $I_f \subset I$ and taking values in I such that $f(j) \in D_j$ for $j \in I_f$, $f(j) \neq j$. The replacement policy f is the prescription to realize instantaneously the replacement $j \to f(j)$ whenever the transition in state j occurs. No replacements are made in states $j \notin I_f$.

For stationary replacement policies we make

Assumption 1.

$$f(j) \notin I_f \qquad \text{for every } j \in I_f .$$

According to the assumption there is assigned to nearly every ω the trajectory of the replacement process $\{Y_t, t \geqq 0\}$, not being left continuous at time of the transition and not right continuous at time of the replacement.

In what follows we denote by E_j^f the mathematical expectation in a replacement process under the stationary replacement policy f and under the condition $i_0 = j$, $\varrho(i)$, $i \in I$, the reward per a time unit in state i, $r(i, j)$, $i, j \in I$, the reward from the transition (i, j); we set $r(i, i) = 0$, $v(i, j)$, $i, j \in I$, the reward from the replacement $(i, +j)$; we set $v(i, i) = 0$. Let us make besides

Assumption 2.

$$(i, +j) \in D,\ (j, +k) \in D \Rightarrow (i, +k) \in D \ \text{or}\ i = k,$$
$$v(i, j) + v(j, k) \leqq v(i, k).$$

2. The mean reward per a time unit from the common process

Let us have the Markov process under the stationary replacement policy f. Let the matrix $\boldsymbol{P}$ of transition probabilities under this policy define isolated recurrent classes $I_1, \ldots, I_m$ and the transient class I'.

(A case with the state space of the process under the stationary policy f containing just one recurrent class see in [2].) Let π_{ij} denote the probability that the first recurrent state reached with the initial state i is the state j, $\pi_{ii} = 1$ for $i \in I - I'$.

The quality of the policy f is characterized by the mean reward per a time unit $\Theta(i)$, $i \in I$, defined as follows: we choose in every isolated recurrent class one state $j_i \in I_i$, $i = 1, \ldots, m$. Let

$$T^i = \inf \{t\colon Y_t = j_i,\ Y_t^- \neq j_i\}$$

be the time of the first transition into the state j_i. We define

$$\Theta(j) = \frac{E^f_{j_i}(R_{T^i})}{E^f_{j_i}(T^i)} \qquad \text{for } j \in I_i,$$

$$\Theta(j) = \sum_{k \in I - I'} \pi_{jk} \Theta(k) \qquad \text{for } j \in I',$$

where R_T is the mean reward from the process up to the time T (see [2]).

Let us denote for $j \in I_i$, $i = 1, \ldots, m$,

$$w(j) = E^f_j(R_{T^i}) - \Theta(j)\, E^f_j(T^i).$$

For $j \notin I_f$ holds

$$w(j) = \frac{\varrho(j)}{\mu(j)} + \sum_{k \neq j} p(j, k) \left[r(j, k) + E^f_k(R_{T^i})\right] - \Theta(j) \left[\frac{1}{\mu(j)} + \sum_{k \neq j} p(j, k)\, E^f_k(T^i)\right] =$$
$$= \frac{\varrho(j)}{\mu(j)} + \sum_{k \neq j} \frac{\mu(j, k)}{\mu(j)} \left[r(j, k) + E^f_k(R_{T^i}) - \Theta(j)\, E^f_k(T^i)\right] - \frac{\Theta(j)}{\mu(j)}\,. \tag{6}$$

Let $j \in I_i$, $i = 1, \ldots, m$. If $\mu(j, k) > 0$, then also $k \in I_i$ and thus $\Theta(j) = \Theta(k)$, which after a modification of (6) gives

$$\varrho(j) + \sum_{k \neq j} \mu(j, k) \left[r(j, k) + w(k) - w(j)\right] - \Theta(j) = 0, \qquad j \notin I_f\,. \tag{7}$$

For $j \in I_f$, $j \in I_i$, we have from the first line (6) in using $\mu(j) \gtrsim \infty$

$$v(j, f(j)) + w(f(j)) - w(j) = 0, \qquad j \in I_f. \tag{8}$$

So we obtain for $j \in I - I'$ the following system of equations

$$\begin{aligned} v(j, f(j)) + w(f(j)) - w(j) = 0, \qquad j \in I_f, \\ \varrho(j) + \sum_{k \neq j} \mu(j, k) [r(j, k) + w(k) - w(j)] - \Theta(j) = 0, \qquad j \notin I_f. \end{aligned} \tag{9}$$

Solving (9) for every isolated recurrent class I_i particularly, then $\Theta(j), j \in I_i$, is independent of j and uniquely determined by system (9), $w(j), j \in I_i$, uniquely up to the additive constant (see [3]). From the definition $\Theta(j)$ for $j \in I'$ it follows that $\Theta(j)$ are uniquely determined by (9) for all $j \in I$. For $j \in I'$ (9) may be regarded as a system of equations for establishing $w(j)$: for $j \in I_f$

$$w(j) = v(j, f(j)) + w(f(j))$$

and since $f(j) \notin I_f$, it suffices to confine to states $j \notin I_f$. From (9) for $j \in I'$, $j \notin I_f$ follows

$$w(j) - \sum_{k \in I'} p(j, k)\, w(k) = \frac{\varrho(j)}{\mu(j)} - \frac{\Theta(j)}{\mu(j)} + \sum_{k \in I} p(j, k)\, r(j, k) + \sum_{k \in I - I'} p(j, k)\, w(k).$$

If we use the symbol $s(j)$ to denote the right side of the equality, we get the solution see the derivation in Theorem 3, paragraph 2 in [4])

$$w(j) = \sum_{n=0}^{\infty} \sum_{k \in I'} p^{(n)}(j, k)\, s(k), \qquad j \in I', j \notin I_f.$$

Theorem 1

$\Theta(1), \Theta(2), \ldots, \Theta(r)$ *are the single possible numbers such that*

$$\begin{aligned} \Theta(f(j)) - \Theta(j) = 0 \qquad \text{for } j \in I_f, \\ \sum_k \mu(j, k)\, \Theta(k) = 0 \qquad \text{for } j \notin I_f, \end{aligned} \tag{10}$$

holds and to which $w(1), \ldots, w(r)$ *are to find so that*

$$\begin{aligned} v(j, f(j)) + w(f(j)) - w(j) = 0 \qquad \text{for } j \in I_f, \\ \varrho(j) + \sum_{k \neq j} \mu(j, k) [r(j, k) + w(k) - w(j)] - \Theta(j) = 0 \qquad \text{for } j \notin I_f. \end{aligned} \tag{11}$$

Proof. We have just proved the existence of the numbers $w(1), \ldots, w(r)$. From the definition π_{ij} and from the definition $\Theta(j)$ for $j \in I'$ follows that

$$\Theta(j) = \sum_{k \in I} \pi_{jk} \Theta(k), \qquad j \in I. \tag{12}$$

The quantities π_{ij} satisfy the relations

$$\pi_{jk} = \pi_{f(j)k}, \qquad j \in I_f,$$
$$\sum_k \mu(j, k)\, \pi_{ki} = 0, \qquad j \notin I_f.$$

(10) follows from here and from (12).

The uniqueness of the solution $\Theta(1), \ldots, \Theta(r)$ was shown in the foregoing considerations on system (9).

Now we describe the Howard's iteration procedure for determining the maximal reward and the optimal stationary replacement policy. Let us $\mathbf{M}_n = \| \mu_n(j, k) \|_{j,k=1}^r$ denote the matrix of the transition intensities of the process under the stationary policy f_n, where $\mu_n(j, k) = \mu(j, k)$ for $j \notin I_{f_n}$.

Choosing an arbitrary stationary replacement policy f_0 we successively determine the stationary replacement policy f_{n+1} on the basis f_n for $n = 0, 1, 2, \ldots$ as follows:

1. We determine the solution $\Theta_n(1), \ldots, \Theta_n(r)$ and $w_n(1), \ldots, w_n(r)$ from equations

$$v(j, f_n(j)) + w_n(f_n(i)) - w_n(j) = 0, \qquad j \in I_{f_n}, \tag{13}$$
$$\varrho(j) + \sum_{k \neq j} \mu(j, k) \left[r(j, k) + w_n(k) - w_n(j) \right] - \Theta_n(j) = 0, \qquad j \notin I_{f_n};$$

$$\Theta_n(f_n(j)) - \Theta_n(j) = 0, \qquad j \in I_{f_n}, \tag{14}$$
$$\sum_k \mu(j, k)\, \Theta_n(k) = 0, \qquad j \notin I_{f_n}.$$

If here $n \neq 0$, we choose one state k in every isolated recurrent class $I_{1n}, \ldots, I_{m^n}$ with respect to the matrix $\mathbf{M}_n$, for which we put $w_n(k) = w_{n-1}(k)$. We proceed in such way that we first solve (13) for every isolated recurrent class with $\Theta_n(j)$ being an unknown independent of j. Inserting the above values in (14) we obtain the system of equations for $\Theta_n(j)$, $j \in I'_n$. Finally inserting all calculed variables in (13), we obtain the system of equations for $w_n(j)$, $j \in I'_n$.

2. We determine f_{n+1} as follows:

We seek step by step for all $j \in I$

(A) $$\max \{\Theta_n(k) - \Theta_n(j),\ k \in D_j;\ \sum_k \mu(j, k)\, \Theta_n(k)\}.$$

If the maximum for a given $j \in I$ is reached by a single expression in the compound racket, we proceed as follows

a) if the maximum is reached by the expression $\Theta_n(i) - \Theta_n(j)$, then $j \in I_{f_{n+1}}$, $f_{n+1}(j) = i$;

b) if the maximum is reached by means of $\sum_k \mu(j, k)\, \Theta_n(k)$, then $j \notin I_{f_{n+1}}$.

he maximum in (A) for a given $j \in I$ is reached by more than only one expression, we use an auxiliary criterion to determine the policy f_{n+1}: we search for

(B) $$\max \{v(j,k) + w_n(k) - w_n(j), \quad k \in D_j; \\ \varrho(j) + \sum_{k \neq j} \mu(j,k)[r(j,k) + w_n(k) - w_n(j)] - \Theta_n(j)\}.$$

If the maximum assumes the expression

$$\varrho(j) + \sum_{k \neq j} \mu(j,k)[r(j,k) + w_n(k) - w_n(j)] - \Theta_n(j),$$

we prefer then not to perform any replacements, i.e. $j \notin I_{f_{n+1}}$. Otherwise, if the maximum in (B) is obtained by the expression

$$v(j,i) + w_n(i) - w_n(j),$$

we choose $j \in I_{f_{n+1}}, f_{n+1}(j) = i$. Hereby preference is given to $f_{n+1}(j) = f_n(j)$, if this choice is in agreement with the criterion (B).

3. If such a policy f_{n+1} does not posses Assumption 1, we change it to the policy f'_{n+1} as follows: in states $j \in I_{f_{n+1}}$, where $f_{n+1}(j) \in I_{f_{n+1}}$ we take $f'_{n+1}(j) = f_{n+1}(f_{n+1}(j))$; in others $j \in I_{f_{n+1}}$ we have $f'_{n+1}(j) = f_{n+1}(j)$.

We now demonstrate the correctness of the procedure in 3. Let us suppose $f_n(j) \notin \notin I_{f_n}$, $j \in I_{f_n}$, and the policy f_{n+1} to be constructed as described above. Further let

$$j \in I_{f_{n+1}}, \qquad f_{n+1}(j) = i \in I_{f_{n+1}}, \qquad f_{n+1}(i) = i',$$

which according to criterion (A), with respect to (14) and to the construction of the replacement policy f_{n+1} implies that

$$\Theta_n(i) - \Theta_l(j) \geqq 0, \qquad \Theta_n(i') - \Theta_n(i) \geqq 0,$$

therefrom

$$\Theta_n(i') - \Theta_n(j) \geqq \Theta_n(i) - \Theta_n(j).$$

There must hold the equality in the last relation (because $j \in I_{f_{n+1}}$) i.e.

$$\Theta_n(i') - \Theta_n(i) = 0,$$

consequently, there was either $i' = f_n(i)$ or there was also used the criterion (B) for the state i.

In either case

$$v(i,i') + w_n(i') - w_n(i) \geqq 0.$$

Therefrom $v(j,i) + w_n(i) - w_n(j) \leqq v(j,i) + v(i,i') + w_n(i') - w_n(j) \leqq v(j,i') + + w_n(i') - w_n(j)$. Again, we see that the equality must hold here (in applying criterion (B) in the state j).

We are thus led to the conclusion that i' is equivalent to i for the state j by the criterions (A), (B). Moreover

$$\Theta_n(i') - \Theta_n(i) = 0, \tag{15}$$

$$v(i,i') + w_n(i') - w_n(i) = 0. \tag{16}$$

We can argue by contradiction that also

$$i \in I_{f_n}, \qquad i' = f_n(i).$$

Hence, there cannot occur the situation

$$f_{n+1}(j) = i, \ \ f_{n+1}(i) = i', \ \ f_{n+1}(i') = i'',$$

since otherwise there would be also

$$f_n(i) = i', \ \ f_n(i') = i'',$$

which contradicts the assumption of the replacement policy f_n. Thus it suffices to change the constructed policy as described in 3. So, we have described the iteration procedure for the construction of f_n, $n = 0, 1, 2, \ldots$

If for any n

$$\Theta_n(j) = \Theta_{n+1}(j), \qquad w_n(j) = w_{n+1}(j), \qquad j \in I, \tag{17}$$

we stop the iteration procedure. Then f_n is the optimal stationary replacement policy, i.e.

$$\Theta_n(j) = \max \{\Theta_f(j) : f \text{ stationary replacement policy}\}, \qquad j \in I. \tag{18}$$

We now verify, that (17) must truly hold.

Let us denote $\Theta_{n+1}(j) - \Theta_n(j) = \bar{\Theta}(j), j \in I$. Again we assume the matrix $\mathbf{M}_{n+1} = \| \mu_{n+1}(j, k) \|_{j,k=1}^{r}$ of the transition intensities under the policy f_{n+1} to define the isolated recurrent classes $I_1, \ldots, I_m$ and the transient class I'.

First, we prove that $\Theta_n(j)$, $n = 0, 1, 2, \ldots$ constitute a not decreasing succession. By (14) and by the construction of f_{n+1} there is

$$\begin{aligned} &\Theta_n(f_{n+1}(j)) - \Theta_n(j) - d_j = 0, \qquad && j \in I_{f_{n+1}}, \\ &\sum_k \mu_{n+1}(j, k)\, \Theta_n(k) - d_j = 0, && j \notin I_{f_{n+1}}, \end{aligned} \tag{19}$$

where $d_j \geqq 0, j \in I$.

Subtracting (19) from the corresponding equations in (10), Theorem 1, for f_{n+1} we obtain

$$\begin{aligned} &\bar{\Theta}(f_{n+1}(j)) - \bar{\Theta}(j) + d_j = 0, \qquad && j \in I_{f_{n+1}}, d_j \geqq 0, \\ &\sum_k \mu_{n+1}(j, k)\, \bar{\Theta}(k) + d_j = 0, && j \notin I_{f_{n+1}}, d_j \geqq 0. \end{aligned} \tag{20}$$

Let $\bar{\mathbf{M}}_{n+1} = \| \bar{\mu}_{n+1}(j, k) \|_{j,k=1}^{r}$ denote the (quasistochastic) matrix of the system in (20) with respect to the variables $\bar{\Theta}(1), \ldots, \bar{\Theta}(r)$ and $\mathbf{x}' = (x_1, \ldots, x_r)$ the stationary distribution, which is the solution of the system

$$\mathbf{x}' \bar{\mathbf{M}}_{n+1} = \mathbf{0}.$$

On multiplying the s-th equation in (20) by the number x_s, $s = 1, \ldots, r$, and on adding all equations we obtain

$$\sum_{j=1}^{r} d_j x_j = 0.$$

Since $x_j = 0$ for $j \in I'$, $x_j \neq 0$ for $j \in I - I'$, this means with respect to $d_j \geqq 0$ that

$$d_j = 0 \qquad \text{for } j \in I - I'.$$

For $j \in I - I'$ is thus the main criterion (A) maximized by the expression $\sum_k \mu(j, k)\, \Theta_n(k) = 0$ or by the expression $\Theta_n(f_{n+1}(j)) - \Theta_n(j) = 0$, if the maximal value is one and only one, or the auxiliary criterion (B) was applied.

In either case we may write for $j \in I - I'$ with respect to (13)

$$v(j, f_{n+1}(j)) + w_n(f_{n+1}(j)) - w_n(j) - e_j = 0, \qquad j \in I_{f_{n+1}} \tag{21}$$
$$\varrho(j) + \sum_{k \neq j} \mu_{n+1}(j, k)\left[r(j, k) + w_n(k) - w_n(j)\right] - \Theta_n(j) - e_j = 0, \qquad j \notin I_{f_{n+1}},$$

where $e_j \geqq 0$.

Subtracting for j mentioned (21) from the corresponding equations in (11) for f_{n+1}, we obtain for $j \in I - I'$ with the notation $w'(j) = w_{n+1}(j) - w_n(j)$

$$w'(f_{n+1}(j)) - w'(j) + e_j = 0, \qquad j \in I_{f_{n+1}}, \tag{22}$$
$$\sum_{k \neq j} \mu_{n+1}(j, k)\left[w'(k) - w'(j)\right] - \bar{\Theta}(j) + e_j = 0, \qquad j \notin I_{f_{n+1}},$$

where $e_j \geqq 0$.

$\bar{\Theta}(j)$ is expressed in (22) and (20) for $j \in I - I'$ as a mean reward. Since $e_j \geqq 0$, we have from Theorem 1 (in choosing $\bar{v}(j, f_{n+1}(j)) = e_j$ for $j \in I_{f_{n+1}}$; $\bar{r}(j, k) = 0$, $\bar{\varrho}(j) = e_j$ for $j \notin I_{f_{n+1}}$)

$$\bar{\Theta}(j) \geqq 0, \qquad j \in I - I'.$$

For $j \in I'$ we obtain from (20)

$$-\sum_{k \in I'} \bar{\mu}_{n+1}(j, k)\, \bar{\Theta}(k) = d_j + \sum_{k \in I - I'} \bar{\mu}_{n+1}(j, k)\, \bar{\Theta}(k), \tag{23}$$

where for the elements $\bar{\mu}_{n+1}(j, k)$ of the matrix $\bar{\mathbf{M}}_{n+1}$

$$\bar{\mu}_{n+1}(j, k) \geqq 0 \quad \text{for } j \neq k; \qquad \bar{\mu}_{n+1}(j, j) = -1 \quad \text{for } j \in I_{f_{n+1}};$$
$$\bar{\mu}_{n+1}(j, j) = -\mu(j) \quad \text{for } j \notin I_{f_{n+1}}, \qquad 0 < \mu(j) < \infty.$$

Let $d'(j)$ denote the right side of (23), which according to the foregoing always a non-negative expression is; then

$$-\bar{\mu}_{n+1}(j, j)\, \bar{\Theta}(j) - \sum_{\substack{k \in I' \\ k \neq j}} \bar{\mu}_{n+1}(j, k)\, \bar{\Theta}(k) = d'_j \geqq 0,$$

whence

$$\bar{\Theta}(j) - \sum_{k \in I'} p_{n+1}(j, k)\, \bar{\Theta}(k) = d''_j \geqq 0, \qquad j \in I',$$

where

$$d''_j = d'_j \quad \text{for } j \in I_{f_{n+1}}, \qquad d''_j = \frac{d'_j}{\mu(j)} \quad \text{for } j \notin I_{f_{n+1}}.$$

On successive substituting we come to

$$\bar{\Theta}(j) = \sum_{m=0}^{N} (\sum_{k \in I'} p_{n+1}^{(m)}(j, k)\, d_k'') + \sum_{k \in I'} p_{n+1}^{(N+1)}(j, k)\, \bar{\Theta}(k), \qquad j \in I'.$$

Because of $k \in I'$ the serie $\sum_{m=0}^{\infty} p_{n+1}^{(m)}(j, k)$ converges for $j \in I$ (see [4], page 8) and thus passing to the limit for $N \to \infty$

$$\bar{\Theta}(j) = \sum_{m=0}^{\infty} \sum_{k \in I'} p_{n+1}^{(m)}(j, k)\, d_k'' \geqq 0, \qquad j \in I'.$$

Thus we have proved that

$$\bar{\Theta}(j) = \Theta_{n+1}(j) - \Theta_n(j) \geqq 0, \qquad \text{i.e.} \qquad \Theta_n(j) \leqq \Theta_{n+1}(j), \qquad j \in I.$$

We conclude from the finiteness of the set of the stationary replacement policies that there exists a q such that

$$\Theta_{n+1}(j) = \Theta_n(j) \qquad \text{for } j \in I, n = q, q+1, \dots \tag{24}$$

If (24) holds, then from (23) $d_j = 0$ for $j \in I'$ and by an analogous consideration as above it can be proved, that the system (22) for $j \in I'$ holds as well.

Under the validity of (24) i.e. from (22) with some modification

$$w'(j) = e'(j) + \sum_k p_{n+1}(j, k)\, w'(k), \qquad j \in I, \tag{25}$$

$e'(j) = e_j$, for $j \in I_{f_{n+1}}$, $e'(j) = \dfrac{e_j}{\mu(j)}$ for $j \notin I_{f_{n+1}}$.

Analogous to the proof of $d_j = 0$ for $j \in I - I'$ in (20) we can verify that (25) yields

$$e'(j) = 0 \qquad \text{for } j \in I - I'.$$

Then

$$w'(j) = \sum_{k \in I_i} p_{n+1}(j, k)\, w'(k), \qquad j \in I_i,\ i = 1, \dots, m,$$

hence $w'(j) =$ constant for $j \in I_i$. Since in every isolated recurrent class there exists one state k for which $w_{n+1}(k) = w_n(k)$ was chosen, it turns out that

$$w'(j) = w_{n+1}(j) - w_n(j) = 0, \qquad j \in I - I'. \tag{26}$$

From (25) and (26) we can write for $j \in I'$

$$w'(j) = e'(j) + \sum_{k \in I'} p_{n+1}(j, k)\, w'(k)$$

and proceeding similarly as in deriving $\bar{\Theta}(j) \geqq 0$, $j \in I'$, we come to the conclusion that $w'(j) \geqq 0$, $j \in I'$, that is for all $j \in I$, $n = q, q+1, \dots$

$$w'(j) = w_{n+1}(j) - w_n(j) \geqq 0,$$

hence

$$w_n(j) \leqq w_{n+1}(j), \qquad j \in I, n = q, q + 1, \ldots \tag{27}$$

Let us remark that the equality in (27) holds for all j whenever the stationary policies f_n and f_{n+1} are equal to each other. A finite number of the stationary replacement policies leads to a conclusion that $n \geqq q$ can be found so that (17) holds.

We have now to prove that in stopping the common iteration procedure we obtain the optimal stationary policy. We apply a similar consideration to that used in proving that $\Theta_n(j)$, $n = 0, 1, 2, \ldots$ form a non-decreasing succession.

Let (17) hold, we want to prove (18). Let f be an arbitrary stationary policy, $\mathbf{M} = \| \mu(i, j) \|_{i,j=1}^{r}$ the matrix of transition intensities determined by the policy f, $I_1, \ldots, I_m$ the recurrent classes with respect to the matrix $\mathbf{M}$, and I' the transient class.

By (17) and by the construction of f_{n+1} the maximum in (A) is reached either by the expression

$$\Theta_n(f_{n+1}(j)) - \Theta_n(j) = \Theta_{n+1}(f_{n+1}(j)) - \Theta_{n+1}(j) = 0, \qquad j \in I_{f_{n+1}},$$

or by the expression

$$\sum_k \mu(j, k)\, \Theta_n(k) = \sum_k \mu(j, k)\, \Theta_{n+1}(k) = 0, \qquad j \notin I_{f_{n+1}},$$

fromwhere for $j \in I$

$$\begin{aligned} &\Theta_n(k) - \Theta_n(j) + d_{jk} = 0, \qquad \text{where} \qquad k \in D_j,\ d_{jk} \geqq 0, \\ &\sum_k \mu(j, k)\, \Theta_n(k) + d_j = 0, \text{ where } d_j \geqq 0. \end{aligned} \tag{28}$$

Subtracting (10) from (28) for $k = f(j)$ we come to

$$\begin{aligned} &\Theta_n(f(j)) - \Theta(f(j)) + \Theta(j) - \Theta_n(j) + d_{jf(j)} = 0, \qquad j \in I_f, \\ &\sum_k \mu(j, k) [\Theta_n(k) - \Theta(k)] + d_j = 0, \qquad j \notin I_f. \end{aligned} \tag{29}$$

Let us introduce for simplification $\Theta_n(k) - \Theta(k) = \bar{\Theta}(k)$, $d_{jf(j)} = d_j$, $j \in I_f$. Then (29) has the form

$$\begin{aligned} &\bar{\Theta}(f(j)) - \bar{\Theta}(j) + d_j = 0, \qquad j \in I_f, \\ &\sum_k \mu(j, k)\, \bar{\Theta}(k) + d_j = 0, \qquad j \notin I_f. \end{aligned} \tag{30}$$

a) In the same manner as we have deduced from (20) that $d_j = 0$ for $j \in I - I'$ we obtain from (30)

$$d_j = 0 \qquad \text{for } j \in I - I'.$$

We can see from (28) that the criterion (A) reaches its maximum for $j \in I - I'$ either by the expression $\Theta_n(f(j)) - \Theta_n(j) = 0$ or by the expression $\sum_k \mu(j, k)\, \Theta_n(k) = 0$.

It means for $j \in I - I'$.

1. if the maximum was reached by only one expression, it was either $j \notin I_f$ and at the same time $j \notin I_{f_{n+1}}$ or $j \in I_f$ and at the same time $j \in I_{f_{n+1}}$, $f(j) = f_{n+1}(j)$;

2. or the policy f_{n+1} was obtained in the states $j \in I - I'$ by the maximalization of the criterion (B).

Thus it holds for $j \in I - I'$

$$v(j, k) + w_n(k) - w_n(j) + e_{jk} = 0, \qquad k \in D_j, e_{jk} \geqq 0, \tag{31}$$
$$\varrho(j) + \sum_{k \neq j} \mu(j, k) [r(j, k) + w_n(k) - w_n(j)] - \Theta_n(j) + e_j = 0, \qquad e_j \geqq 0.$$

Subtracting from (31) the corresponding equations from (11) (in the first row we choose $k = f(j) \in D_j$), we obtain with the notation

$$w_n(k) - w(k) = w'(k), \qquad e_{jf(j)} = e_j, \qquad j \in I_f,$$
$$\Theta_n(k) - \Theta(k) = \bar{\Theta}(k),$$

the following equations

$$w'(f(j)) - w'(j) + e_j = 0, \qquad j \in I_f, \qquad e_j \geqq 0, \tag{32}$$
$$\sum_{k \neq j} \mu(j, k) [w'(k) - w'(j)] - \bar{\Theta}(j) + e_j = 0, \qquad j \notin I_f, e_j \geqq 0.$$

(30) and (32) analogously to (20) and (22) yield

$$\bar{\Theta}(j) \geqq 0 \qquad \text{for } j \in I - I',$$

that is

$$\Theta_n(j) \geqq \Theta(j), \qquad j \in I - I'.$$

b) For $j \in I'$ we get from (30)

$$-\sum_{k \in I'} \bar{\mu}(j, k) \bar{\Theta}(k) = d_j + \sum_{k \in I - I'} \bar{\mu}(j, k) \bar{\Theta}(k),$$

where $\bar{\boldsymbol{M}} = \| \bar{\mu}(j, k) \|_{j,k=1}^{r}$ is the matrix of the system in (30). From this we deduce in the same manner as from (23)

$$\bar{\Theta}(j) \geqq 0, \qquad j \in I',$$

i.e.

$$\Theta_n(j) \geqq \Theta(j), \qquad j \in I'.$$

The proof of relation (18) is thus complete.

Finally I should like to express my gratitude to dr. P. Mandl, DrSc., for providing me with valuable expert advice and helpful criticism in writing this article.

REFERENCES

[1] Howard, R. A.: *Dynamic programming and Markov processes.* New York—London 1960.

[2] Kunderová, P.: *On a mean reward from a Markov replacement process with only one isolated class of recurrent states.* Acta Universitatis Palackianae Olomucensis, Facultas rerum naturalium. 1979, Tom 61.

[3] Mandl, P.: *An identity for Markovian replacement processes.* Journal of Applied Probability, Vol. 6, No. 2, pp. 348—354.
[4] Mandl, P.: *Řízené Markovovy řetězce.* Příloha časopisu Kybernetika, roč. 5, 1969, Academia Praha.
[5] Mandl, P.: *Some results on Markovian replacement processes.* Journal of Applied Probability, Vol. 8, pp. 357—365.

Souhrn

PRŮMĚRNÝ VÝNOS Z OBECNÉHO MARKOVOVA PROCESU S OBNOVAMI

PAVLA KUNDEROVÁ

Uvažuje se Markovův proces s obnovami popsaný v [5] s obecnou stacionární strategií obnovy. Za charakteristiku kvality strategie se považuje očekávaný průměrný výnos na jednotku času $\Theta(i)$, $i \in I$, definovaný v odstavci 2. Ve větě 1 je odvozena soustava rovnic (11) pro výpočet výnosů $\Theta(i)$ a ukázána jednoznačnost jejího řešení. Je zkonstruován obecný Howardův iterační postup (viz [1]) k nacházení optimální stacionární strategie, při níž se dosahuje optimálního výnosu. Článek navazuje na par. 10 práce [4], který se zabývá průměrným výnosem z řízeného Markovova řetězce.

Резюме

СРЕДНИЙ ДОХОД ИЗ ОБЩЕГО ПРОЦЕССА МАРКОВА С ВОССТАНОВЛЕНИЯМИ

ПАВЛА КУНДЕРОВА

В работе рассмотрен процесс Маркова с восстановлениями (определённый в [5]) при использовании общей стационарной стратегии восстановления. Характеристикой качества стратегии является ожидаемый средний доход на единицу времени $\Theta(i)$, $i \in I$, определённый в пар. 2. В теореме 1 введена система уравнений (11) для доходов $\Theta(i)$ и показана единственность решения этой системы. Описан итерационный метод Ховарда для нахождения оптимальной стационарной стратегии при которой достигается максимального дохода. Статья относится к пар. 10 работы [4], которая занимается средним доходом из управляемой цепи Маркова.