

## Markov decision processes in minimization of expected costs

Marija Rukav<sup>1,\*</sup>, Kruno Stražanac<sup>2</sup>, Nenad Šuvak<sup>1</sup> and Zoran Tomljanović<sup>1</sup>

<sup>1</sup> *Department of Mathematics, Josip Juraj Strossmayer University of Osijek  
Trg Ljudevita Gaja 6, 31 000 Osijek, Croatia  
E-mail: {mrukav, nsuvak, ztomljan}@mathos.hr*

<sup>2</sup> *Farmeron d.o.o.  
Ružina 7, 31 000 Osijek, Croatia  
E-mail: {kruno@farmeron.com}*

**Abstract.** Basics of Markov decision processes will be introduced in order to obtain the optimization goal function for minimizing the long-run expected cost. We focus on minimization of such cost of the farmer's policy consisting of different decisions in specific states regarding both milk quality and quantity (lactation states) produced by a dairy cow. The transition probability matrix of the Markov process, used here for modeling of transitions of a dairy cow from one state to another, will be estimated from the data simulated from the lactation model that is often used in practice. We want to choose optimal actions in the states of this Markov process regarding the farmer's costs. This problem can be solved by exhaustive enumeration of all possible cases in order to obtain the optimal policy. However, this is feasible only for a small number of states. Generally, this problem can be approached in the linear programming setting which yields an efficient solution. In order to demonstrate and compare these two approaches, we present an example based on the simulated data regarding milk quality and quantity.

**Key words:** Linear programming methods, long-run expected cost, Markov decision process, stationary distribution, optimal policy.

Received: September 17, 2014; accepted: December 7, 2014; available online: December 30, 2014

---

## 1. Introduction

The development of Markov decision processes started in the '50s of the last century with the evolution of operational research. Mathematicians wanted to define models that will make the best decisions in some sense. At the beginning, these processes were studied by Dick Bellman and Ron Howard [2, 9], who provided an entirely new perspective on the theory of control. With further research and development, decision models became more generalized. Today, Markov decision processes (see [1, 8]) are a mathematical framework for making decisions in situations where outcomes are partly random and partly controlled by decision makers.

---

\*Corresponding author.

In the last decade, the development of Markov decision processes became very important. These processes are used not only in solving optimization problems, but also in dynamic programming, and therefore respective problems are often referred to as stochastic dynamic programs or stochastic control problems. They also find application in many real situations and different fields such as economics, ecology, biology, robotics and planning and operations management. The most famous examples of Markov decision processes are certainly games where an objective is to find a strategy that will maximize some profit. These processes are also used for finding optimal strategies in problems of optimal stopping, replacement problems and repair problems, respectively. The main objective of the decision process is to find strategies that provide optimal solutions considering the states, actions (decisions), costs and transition probabilities.

Nowadays, optimal replacement problems are also investigated within the framework of the dairy cow replacement, see, e.g., [3, 7], what is the main subject of this paper. First, we will define the above stated terms and explain their basic properties. After the theoretical part in Section 2, we will define the problem in which the number of states of the Markov process and respective decisions are based on the simulated data regarding milk quality and quantity produced by the dairy cow (Subsection 2.1). This is the basis for determining the optimal policy (strategy) of decisions, i.e., the policy that has the minimal cost. This will be done by two approaches: probabilistic properties of Markov decision processes (Subsection 2.2) and linear programming techniques (Section 3). For numerical calculations we used programming package Matlab. All numerical results are written in Matlab's short format (format with 4 digits after the decimal point).

## 2. Stochastic model for optimal policy

In order to apply Markov decision processes to the problem of finding the policy that minimizes farmer's long-run expected cost, we will introduce some basics of both Markov processes and Markov decision processes.

### 2.1. Markov chain model

A particular stochastic process for which the future behavior of the process depends only on its present state and not on its past behavior is called the Markov process. The basic model used in this paper is the discrete time Markov chain, i.e., the discrete time stochastic process  $X = (X_n, n \in \mathbb{N}_0)^\ddagger$  with the discrete state space  $S$ , for which the previously described "memoryless" property could be written in the following form:

$$P(X_n = j | X_{n-1} = i, X_{n-2} = i_{n-2}, \dots, X_0 = i_0) = P(X_n = j | X_{n-1} = i). \quad (1)$$

This property is called the Markov property. Furthermore, if probability (1) is independent of time  $n$ , we say that the Markov chain is time homogeneous and we denote the transition probabilities (1) simply by  $p_{ij}$ ,  $i, j \in S$ . In addition to

---

<sup>‡</sup>Without loss of generality, we take the index set to be  $\mathbb{N}_0$ .

the state space and the matrix of transition probabilities  $\Pi = [p_{ij}]_{i,j \in S}$  (i.e., a one-step transition matrix), we also need to specify an initial distribution  $\lambda$  of the homogeneous Markov chain, i.e., the distribution of a random variable  $X_0$ . Then, from the transition matrix  $\Pi$  and the initial distribution  $\lambda$  we can calculate arbitrary finite-dimensional Markov chain distributions. For more details regarding Markov chains we refer to [10].

In this paper, we focus on a problem of finding the optimal policy that minimizes farmer's long-run expected cost per a dairy cow which naturally changes the states defined regarding both milk quality and quantity. The discrete-time homogeneous Markov chain seems to be an adequate probabilistic model for describing transitions of a dairy cow from one state to another. For simplicity of this example, we define just eight states of such Markov chain, taking into account both quality and quantity of milk produced by the respective cow. In a more general case, one could consider a more general Markov chain model, e.g., including fat, lactose and other characteristics of milk quality.

To define these eight states, we first define special categories of milk quality and quantity in which the realized lactation yields will be classified. First, we define only two quality categories based on the average of optimal milk fat, lactose and proteins concentration: category 0 - milk of good quality (above average); category 1 - milk of poor quality (below or equal to average).

The definition of quantity categories is not so simple since we want to put it in a close relationship with the optimal lactation yield of the cow. For this purpose we observe the optimal lactation curve resulting from the Wood model which is often referred to as the best model for a description of lactation (see, e.g., [4]). The optimal lactation curve from the Wood model is estimated from the real lactation yields. In addition to the estimated model function, we observe weakly averages of the simulated daily lactation yields for one particular cow for three lactation periods (see Figure 1).<sup>§</sup>

Once we have the optimal lactation curve and the simulated lactation yields of the observed cow (see Figure 1), we define four categories referring to milk quantity:

- category 1 - if the quantity of produced milk is above the optimal value (with respect to the optimal value from the Wood model),
- category 2 - if the quantity of produced milk is at most 1% below the optimal value,
- category 3 - if the quantity of produced milk is between 1% and 4% at most below the optimal value,
- category 4 - if the quantity of produced milk is more than 4% below the optimal value.

After defining quality and quantity categories, we define states of the Markov chain. For example, if the milk is of good quality (category 0) and the lactation yield is above the optimal level (category 1), we define this to be state 1 of the

---

<sup>§</sup>The simulated data and the estimated Wood model function are obtained from the Farmeron company.

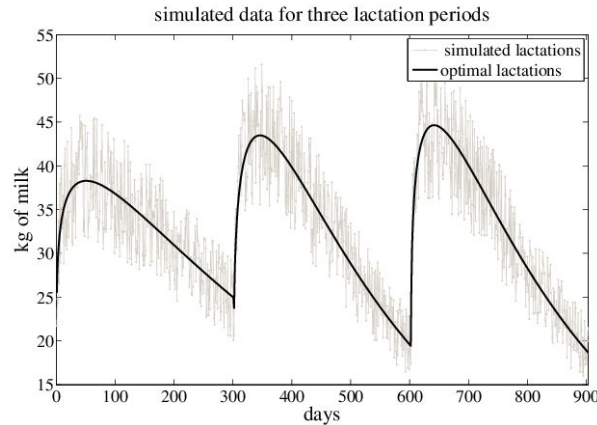


Figure 1: The Wood model function and the simulated data on daily lactations

Markov chain. Clearly, state 1 corresponds to pair (0, 1) of the quality and quantity categories. All states of this Markov chain are defined in Table 1.

	Quantity	1	2	3	4
Quality					
0		state 1	state 3	state 5	state 7
1		state 2	state 4	state 6	state 8

Table 1: States of the Markov chain regarding both quality and quantity categories of milk

Now we are dealing with the discrete-time homogeneous Markov chain with the finite state space  $S = \{1, 2, 3, 4, 5, 6, 7, 8\}$  and an unknown transition matrix  $\Pi$ . In this setting, the transition matrix must be estimated from the data describing both the quality and the quantity of milk of a particular cow over time, i.e., from the available simulated lactation data. The one-step transition probability from state  $i$  to state  $j$ , i.e., the probability  $p_{ij} = P(X_n = j | X_{n-1} = i)$  is estimated as the following quotient of relative frequencies (see [5]):

$$\hat{p}_{ij} = \frac{\frac{\text{number of consecutive pairs } (i,j) \text{ of states}}{\text{number of all consecutive pairs of states}}}{\frac{\text{number of weeks in state } i}{\text{number of all weeks}}}. \tag{2}$$

This procedure results in the following estimate of the transition matrix  $\Pi$ :

$$\hat{\Pi} = \begin{bmatrix} 0.3889 & 0.0556 & 0 & 0 & 0.2778 & 0 & 0.2778 & 0 \\ 0 & 0.4545 & 0.0227 & 0.0909 & 0 & 0.3864 & 0 & 0.0455 \\ 0.25 & 0 & 0.25 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0.25 & 0 & 0.25 \\ 0.4286 & 0.0714 & 0.0714 & 0 & 0.2857 & 0.0714 & 0.0714 & 0 \\ 0 & 0.56 & 0 & 0.16 & 0.04 & 0.16 & 0 & 0.08 \\ 0.375 & 0 & 0.125 & 0 & 0.25 & 0 & 0.25 & 0 \\ 0.1429 & 0.7143 & 0 & 0 & 0 & 0.1429 & 0 & 0 \end{bmatrix}. \tag{3}$$

Without loss of generality, we can assume that the observation starts when the cow is in its best state, i.e., in state 1. It means that we assume that the initial

distribution of this Markov chain is the distribution  $\lambda = (1, 0, 0, 0, 0, 0, 0, 0)$ , i.e., the distribution given by the rule  $P(X_0 = 1) = 1$ . Now we have a uniquely specified discrete-time homogeneous  $(\lambda, \hat{\Pi})$ -Markov chain that describes transitions of a dairy cow from one to another lactation state from the state space  $S$ .

## 2.2. Markov decision model and optimal policy

We introduce the possibility of taking some predefined action, i.e., making a specified decision from the set of  $K$  possible decisions in each state from  $S$ . By adding the set of  $K$  possible decisions to the discrete-time homogeneous  $(\lambda, \Pi)$ -Markov chain, we actually built a new stochastic model - a Markov decision process.

The decisions made in states of the observed Markov chain affect its transition probabilities, so at first glance it could seem that we lost the homogeneity of our Markov model. However, this is not the case, since making decisions affects the transition matrix on the following way:

- the state  $i \in S = \{1, \dots, M\}$  of the  $(\lambda, \Pi)$ -Markov chain is observed after each transition,
- after each observation, a decision (action)  $k$  is chosen from a set of  $K$  possible decisions (some of the  $K$  decisions may not be relevant for some of the states),
- if the decision  $d_i$  is made in state  $i \in S$ , an immediate cost  $C_{ik}$  is incurred,  $k \in \{1, \dots, K\}$ ,
- the decision  $d_i = k$ ,  $k \in \{1, \dots, K\}$ , in state  $i$  determines what the transition probabilities will be for the next transition from state  $i$  to another state  $j \in S$  - denote these transition probabilities by  $p_{ij}(k)$ ,
- a specification of the decisions  $R = (d_i, i \in S)$  in the respective states prescribes a policy for the Markov decision process with a new transition matrix  $\Pi^{(R)}$ .

Therefore, for each policy  $R = (d_i, i \in S)$  we have a new homogeneous  $(\lambda, \Pi^{(R)})$ -Markov chain. The objective is to find an optimal policy according to some cost criterion - one common criterion is to minimize the long-run expected cost.

In our example, dealing with lactation yields of the dairy cow, the purpose of making decisions is to improve either the quality or the quantity of milk. To be more precise, we introduce the following four decisions (i.e.,  $K = 4$ ):  $N$  - do nothing;  $L$  - improve lactation yield (i.e., the quantity of milk);  $Q$  - improve the quality of milk;  $Z$  - replace the dairy cow with a new one (i.e., switch to state 1).

However, not every decision could be made in each of the states from  $S = \{1, \dots, 8\}$  (i.e.,  $M = 8$ ):

- decision  $N$  could be taken in every state, except state 8,
- decision  $L$  improves the lactation yield and, when taken, puts a cow in the closest state in which the lactation yield is better, but the quality remains the same - it could be taken in states 3, 4, 5, 6, 7 and 8,

- decision  $Q$  improves the quality of milk and it could be taken only if the quality of milk is poor, i.e., in states 2, 4, 6 and 8,
- decision  $Z$  could be taken only in state 8.

For example,  $(N, N, N, N, N, N, N, Z)$  is the policy in which no action is taken in states 1 – 7, but when a cow is in state 8 (poor lactation yield and poor quality of milk), it is replaced. In the same manner, by taking exactly one of the possible decisions in each state from  $S$ , we can observe  $2^4 \cdot 3^3 = 432$  different policies.

Furthermore, every decision in a particular state generates some cost. The costs in HRK for all possible decisions, except the replacement decision  $Z$  that could be taken only in state 8 and that generates the cost of 5,000 HRK, are shown in Figure 2.

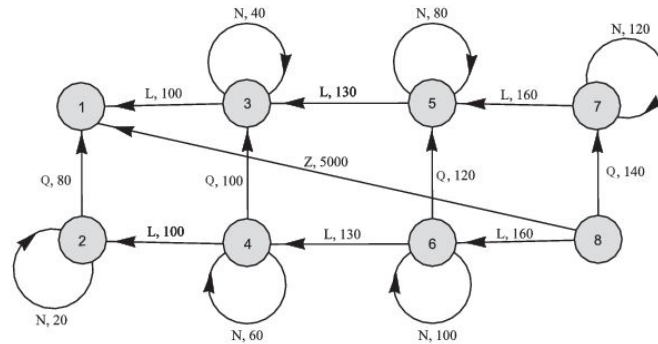


Figure 2: The scheme of decision costs

Our objective is to determine the policy the with minimal long-run expected cost - it means that we are interested in the long-run distribution of costs for each and every of 432 policies. Namely, this distribution is nothing but a transformation of the distribution that describes the long-run behavior of the  $(\lambda, \Pi^{(R)})$ -Markov chain, i.e., its limiting distribution

$$\left( \lim_{n \rightarrow \infty} p_{ij}^{(n)}(k), j \in S \right) = \left( \lim_{n \rightarrow \infty} P^{(k)}(X_n = j | X_0 = i), j \in S \right) \quad (4)$$

for some  $i \in S$ , where  $p_{ij}^{(n)}(k)$  are elements of the  $n$ -th power of the transition matrix  $\Pi^{(R)} = [p_{ij}(k)]_{i,j \in S}$ .

The  $(\lambda, \Pi^{(R)})$ -Markov chain has a finite number of states and it could be shown that its limiting distribution exists. According to [10, Theorem 1.7.2, p. 33], its limiting distribution coincides with its stationary distribution, which is significantly easier to determine. Namely, the stationary distribution  $\pi^{(R)} = (\pi_i^{(R)}, i \in S)$  is a distribution with the characteristics that, if it is taken to be the initial distribution of the Markov chain, its one-dimensional distributions do not change over time. The stationary distribution is a solution of a system of linear equations

$$\pi^{(R)} \cdot \Pi^{(R)} = \pi^{(R)}, \quad \sum_{i \in S} \pi_i^{(R)} = 1. \quad (5)$$

In the specific problem considered here, the transition matrix  $\Pi^{(R)}$  could be obtained from the estimated transition matrix (3) of the original  $(\lambda, \widehat{\Pi})$ -Markov chain. For example, for the strategy  $R = (N, Q, N, L, N, N, N, Q)$  the transition matrix  $\Pi^{(R)}$  is obtained from the transition matrix  $\widehat{\Pi}$  by the following rules:

- rows 1, 3, 5, 6 and 7 remain unchanged,
- according to this strategy, in state 4 we make the decision  $L$ , which means that when a cow is in state 4, its lactation yield is improved by some treatments and it moves to the 2 with probability 1, i.e.,  $\widehat{p}_{42}(L) = 1$  and  $\widehat{p}_{4j}(L) = 0$  for all  $j \in S \setminus \{2\}$ ,
- analogously, in states 2 and 8 we make the decision  $Q$ , which means that when a cow is in one of these states, its lactation yield is improved by some treatments and it moves to states 1 and 7 with probability 1, respectively. This means that  $\widehat{p}_{21}(Q) = 1$ ,  $\widehat{p}_{2j}(Q) = 0$  for all  $j \in S \setminus \{1\}$  and  $\widehat{p}_{87}(Q) = 1$ ,  $\widehat{p}_{8j}(Q) = 0$  for all  $j \in S \setminus \{7\}$ .

Now we are dealing with the new  $(\lambda, \Pi^{(R)})$ -Markov chain that satisfies all conditions from [10, Theorem 1.8.3, p. 41], so its stationary distribution exists and it coincides with its limiting distribution. By solving the system of equations (5) it follows that the stationary distribution of this Markov chain, i.e., its long-run distribution, is the distribution  $\pi^{(R)}$  given by the table

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 0.4138 & 0.0580 & 0.0553 & 0.0036 & 0.2642 & 0.0225 & 0.1808 & 0.0018 \end{pmatrix}. \tag{6}$$

The distribution of the long-run cost of the particular policy is just a transformation of the long-run distribution (6), e.g., for the policy  $R = (N, Q, N, L, N, N, N, Q)$  the long-run distribution of the policy cost is given by the following table:

$$C^{(R)} = \begin{pmatrix} 0 & 80 & 40 & 100 & 80 & 100 & 120 & 140 \\ 0.4138 & 0.0580 & 0.0553 & 0.0036 & 0.2642 & 0.0225 & 0.1808 & 0.0018 \end{pmatrix}.$$

Therefore, the expected long-run cost of the policy  $R = (N, Q, N, L, N, N, N, Q)$  is

$$E(C^{(R)}) = \sum_{i=1}^8 C_i \cdot \pi_i^{(R)} = 52.5449. \tag{7}$$

In the same way, we can calculate the expected long-run cost for each of 432 possible policies. The optimal policy will be the one with the minimal expected long-run cost. In this example, this is the observed policy  $R = (N, Q, N, L, N, N, N, Q)$ .

### 3. Linear programming and optimal policies

The approach introduced in the previous section is a direct approach that enumerates and examines all decisions that are relevant to our problem setting. This approach is feasible only for a small number of states, but for examples with a large number

of states we would like to consider an approach that calculates an optimal solution efficiently.

One approach that improves the previous direct approach is based on linear programming which will be introduced in this section.

In order to write decisions in matrix form we will describe decision policy  $R$  such that we prescribe decisions for each state  $i \in S = \{1, \dots, M\}$ . This can be written in matrix form by prescribing the values of matrix elements of matrix  $D \in \mathbb{R}^{M \times K}$ , where all elements of matrix  $D$  are zeros, except  $D_{ij} = 1$  if decision  $j \in \{1, \dots, K\}$  is made in state  $i \in S$ .

For example, if the  $i$ -th row of matrix  $D$  is equal to  $[0 \ 0 \ 1 \ 0]$ , this means that we will make the third decision in state  $i \in S$ .

In what follows we would like to introduce randomized policies that provide a linear programming setting of the considered problem. For that purpose we will consider a randomized version of the considered policy. We consider matrix form of prescribing the strategy  $R$ , where we consider probability distributions for a decision that has to be made in state  $i$ . More precisely, this means that  $D_{ik}$  is the probability of choosing the decision  $k$  if the system is in state  $i \in S$  or

$$D_{ik} = P\{\text{decision} = k \mid \text{state} = i\}. \quad (8)$$

In that case, the  $i$ -th row of the matrix  $D$  is the probability distribution for the decision to be made in state  $i$ , which is called the randomized policy. Directly from its definition, it yields that elements of the matrix  $D$  satisfy  $\sum_{k=1}^K D_{ik} = 1$  and  $0 \leq D_{ij} \leq 1, \forall i, j \in S$ .

For example, if the  $i$ -th row of the matrix  $D$  is equal to  $[\frac{1}{3} \ 0 \ \frac{1}{3} \ \frac{1}{3}]$ , this means that with equal probability  $\frac{1}{3}$  we will make the first, the third or the fourth decision in state  $i \in S$ .

A random strategy can be performed in real life problems, but it is more important that making  $D_{ij}, \forall i, j$  allow us to formulate a linear programming setting.

First, we introduce probabilities that the system is in state  $i$  and decision  $j$  is made, and denote them by  $y_{ik}$ , that is

$$y_{ik} = P\{\text{decision} = k, \text{state} = i\}. \quad (9)$$

Variables  $y_{ij}$  will correspond to variables of linear programming formulation that will be introduced in the sequel. In particular, from the decision problem setting described in Figure 2 we can observe that we can not consider all decisions in all states. In our formulation, we will assume that all  $y_{ij}$  not relevant for the considered situation are equal to zero.

From the definition of  $y_{ij}$  from (9) and long-run probabilities  $\pi_i^{(R)}$  solving system (5), by using standard results from probability theory, it can be shown that

$$D_{ij} = \frac{y_{ij}}{\pi_i^{(R)}} = \frac{y_{ij}}{\sum_{k=1}^K y_{ik}}, \quad (10)$$

$$\sum_{i=1}^M \sum_{k=1}^K y_{ik} = 1. \quad (11)$$



Moreover, by using the stationarity relation  $\pi_i^{(R)} = \sum_{k=1}^M \pi_i^{(R)} p_{ij}$ , it can be shown that

$$\sum_{k=1}^K y_{jk} = \sum_{i=1}^M \sum_{k=1}^K y_{ik} p_{ij}(k), \quad j = 1, \dots, M.$$

The long-run expected cost is given by

$$E(C^{(R)}) = \sum_{i=1}^M \sum_{k=1}^K \pi_i^{(R)} C_{ik}^{(R)} D_{ik} = \sum_{i=1}^M \sum_{k=1}^K C_{ik}^{(R)} y_{ik}. \tag{12}$$

Now, putting all needed equations together, linear programming formulation with variables  $y_{ij}$  is to minimize the expected cost from (12) subject to constraints

$$\sum_{i=1}^M \sum_{k=1}^K y_{ik} = 1, \quad y_{ik} \geq 0, \quad i = 1, \dots, M, \quad k = 1, \dots, K. \tag{13}$$

$$\sum_{k=1}^K y_{jk} = \sum_{i=1}^M \sum_{k=1}^K y_{ik} p_{ij}(k), \quad j = 1, \dots, M \tag{14}$$

As long as the model is not too large, it can be solved by a simplex method. In our example, we use Matlab’s implementation for solving linear programming problems implemented in function `linprog` and more details can be found in [6, 11].

By summing equations (14) for all  $j$  and using a standard result from probability theory, it can be shown that one equation is redundant, so one of these  $M$  equations can be deleted from linear programming formulation. Thus, the linear problem contains  $M$  constraints and  $M \cdot K$  variables with  $M$  basic variables.

Moreover, it can be shown that for each index  $i \in \{1, 2, \dots, M\}$  there exists at least one index  $j \in \{1, 2, \dots, K\}$  such that  $y_{ij} > 0$ . Since the linear problem has  $M$  basic variables, it yields that for each index  $i \in \{1, 2, \dots, M\}$  there exists exactly one index  $j \in \{1, 2, \dots, K\}$  such that  $y_{ij} > 0$ . Then directly from (10) we have that  $D_{ij}$  is equal to 0 or 1. Note that we have considered a randomized version of the considered problem by allowing that  $D_{ij}$  is a continuous variable, but we have just shown that this problem results in a solution that gives a deterministic solution.

We will illustrate performance of linear programming approach on the example introduced in Section 2.2. First, we should write relevant variables that should be considered in a linear problem setting. In Section 2.2, we have emphasized that we do not allow every decision in each of the considered states. Thus, from the notation introduced in (9) and Section 2.2, it is easy to see that we have 18 relevant variables. Variables  $y_{11}, y_{21}, y_{31}, y_{41}, y_{51}, y_{61}, y_{71}, y_{32}, y_{42}, y_{52}, y_{62}, y_{72}, y_{82}, y_{23}, y_{43}, y_{63}, y_{83}, y_{84}$  are particularly relevant for our example and in this order we will sort variables in linear programming formulation.

Now, equations of linear problem formulation can be written in matrix form. Equations from (13) can be directly written, while with direct calculation it can be shown that equation (14) is in matrix form equal to  $Ay = 0$ , where  $y$  contains relevant variables  $y_{ij}$  in the above mentioned order and matrix  $A$  is equal to:

$$\begin{bmatrix} 0.6111 & 0 & -0.25 & 0 & -0.4286 & 0 & -0.375 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 \\ -0.0556 & 0.5455 & 0 & -0.5 & -0.0714 & -0.56 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -0.0227 & 0.75 & 0 & -0.0714 & 0 & -0.125 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & -0.0909 & 0 & 1 & 0 & -0.16 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ -0.2778 & 0 & -0.5 & 0 & 0.7143 & -0.04 & -0.25 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & -0.3864 & 0 & -0.25 & -0.0714 & 0.84 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 1 & 0 & 0 \\ -0.2778 & 0 & 0 & 0 & -0.0714 & 0 & 0.75 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & -0.0455 & 0 & -0.25 & 0 & -0.08 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

In this example, using costs introduced in Figure 2 we have that the long-run expected cost from (12) is given by  $E(C^{(R)})$  equal to

$$20y_{21} + 40y_{31} + 60y_{41} + 80y_{51} + 100y_{61} + 120y_{71} + 100y_{32} + 100y_{42} + 130y_{52} \\ + 130y_{62} + 160y_{72} + 160y_{82} + 80y_{23} + 100y_{43} + 120y_{63} + 140y_{83} + 5000y_{84},$$

for an arbitrary policy  $R$  determined by relevant variables.

Using Matlab's function `linprog` we have calculated optimal solution  $y$  which is equal to

$$y = [0.4138 \ 0 \ 0.0553 \ 0 \ 0.2642 \ 0.0225 \ 0.1808 \ 0 \ 0.0036 \ 0 \ 0 \ 0 \ 0 \ 0.058 \ 0 \ 0 \ 0.0018 \ 0]^T.$$

That is, it holds that non zero elements of vector  $y$  correspond to the following relevant variables  $y_{11} = 0.4138$ ,  $y_{31} = 0.0553$ ,  $y_{51} = 0.2642$ ,  $y_{61} = 0.0225$ ,  $y_{71} = 0.1808$ ,  $y_{42} = 0.0036$ ,  $y_{23} = 0.058$ ,  $y_{83} = 0.0018$ .

Finally, by using (10) we can conclude that matrix  $D$  introduced in the beginning of this section is determined by

$$D^T = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus, the optimal strategy obtained by using linear programming formulation is equal to  $R = (N, Q, N, L, N, N, N, Q)$ , which is the same as the one obtained as optimal in the approach where we have enumerated all relevant policies.

**Remark 3.1.** Note that null elements in the long-run distribution  $\pi^{(R)}$  indicate that it is possible that the problem setting has more optimal solutions. This means that if we have the optimal strategy  $R$  that is not unique (regarding the expected long-run cost), there exists state  $i \in S$  such that  $\pi_i^{(R)} = 0$ . Similarly, in linear programming formulation we will obtain that for that fixed  $i$   $y_{ij} = 0$  for any  $j$ . This means that in both cases we obtain that in state  $i$  we can make any feasible decision.

#### 4. Discussion and conclusion

We have considered the problem of minimizing the expected cost of the farmer's policy consisting of different decisions in specific states of a dairy cow, which are defined according to milk quality and quantity. In problem formulation, for modeling transitions of a dairy cow from one state to another regarding the specific decisions, we used the decision Markov chain model. For minimization of the expected cost of

the policy we presented an approach based on exhaustive enumeration of all possible policies and a more efficient approach based on linear programming techniques. In a numerical example based on the simulated data, we have demonstrated both approaches. This example illustrates that a linear programming setting is more efficient, which makes it more appropriate for real life applications characterized by a large number of states and decisions. One benefit of these results is the potential application in dairy cow herd management, which opens up a possibility for the future cooperation with the real sector.

## Acknowledgement

We would like to thank Prof. Kristian Sabo for introducing this interesting problem to us as well as for many useful discussions and advice.

## References

- [1] Bauerle, N. and Rieder, U. (2011). *Markov Decision Processes with Applications to Finance*. Berlin: Springer.
- [2] Bellman, R.E. (1957). *Dynamic Programming*. Princeton: Princeton University Press.
- [3] Cabrera, V.E. (2012). A simple formulation and solution to the replacement problem: A practical tool to assess the economic cow value, the value of a new pregnancy, and the cost of a pregnancy loss. *J. Dairy Sci.*, 95, 4683-4698.
- [4] Cankaya, S., Unalan, A. and Soydan, E. (2011). Selection of a mathematical model to describe the lactation curves of Jersey cattle. *Archiv Tierzucht*, 54, 27–35.
- [5] Craig, B.A. and Sendi, P.P. (2002). Estimation of the transition matrix of a discrete-time Markov chain. *Health Economics*, 11, 33–42.
- [6] Dantzig, G.B., Orden, A., and Wolfe, P. (1955). Generalized simplex method for minimizing a linear form under linear inequality restraints. *Pacific Journal Math.*, 5, 183-195.
- [7] Grön, Y.T., Rajala-Schultz, P.J., Allore, H.G, DeLorenzo, M.A., Hertl, J.A. and Galligan D.T. (2003). Optimizing replacement of dairy cows: Modeling the effects of diseases. *Preventive Veterinary Medicine*, 61, 27-43.
- [8] Hiller, F.S. and Lieberman, G.J. (2001). *Introduction to Operations Research*. New York: McGraw-Hill.
- [9] Howard, R.A. (1960). *Dynamic Programming and Markov Processes*. Cambridge: The M.I.T. Press.
- [10] Norris, J.R. (1998). *Markov Chains*. Cambridge: Cambridge University Press.
- [11] Bertsimas, D. and Tsitsiklis J.N. (1997). *Introduction to Linear Optimization*. Nashua, USA: Athena Scientific.