

Markov-Modulated Linear Regression Parameter Estimation Using a Convolution of Exponential Densities

Nadezda Spiridovska
Transport and Telecommunication Institute
Lomonosov 1, LV-1019, Riga
Latvia

Received: April 24, 2020. Revised: June 2, 2020. Accepted: June 6, 2020. Published: June 8, 2020

Abstract— Markov-modulated linear regression model is a special case of the Markov-additive process $(Y, J) = \{(Y(t), J(t)), t \geq 0\}$, where component J is called Markov, and component Y is additive and described by a linear regression. The component J is a continuous-time homogeneous irreducible Markov chain with the known transition intensities between the states. Usually this Markov component is called the external environment or background process. Unknown regression coefficients depend on external environment state, but regressors remain constant. This research considers the case, when the Markov property is not satisfied, namely, the sojourn time in each state is not exponentially distributed. Estimation procedure for unknown model parameters is described when it's possible to represent transition intensities as a convolution of exponential densities. An efficiency of such an approach is evaluated by a simulation.

Keywords—Stochastic processes, density convolution, hypoexponential distribution, Markov-modulated linear regression, parametrical estimation, simulation.

I. INTRODUCTION

Markov processes themselves and their derivatives Markov-modulated processes are the most popular and commonly used modelling tools for stochastic systems. They are widely applied in various areas of human life, both in science and in industry: physics, chemistry, information and communication technologies, economics and finance, social sciences, games, music and many others. The concept of a second hidden or visible layer, which can be described by the Markov chain (usually for a more realistic representation of the studied process) proved to be suitable and highly demanded.

This work was financially supported by the specific support objective activity 1.1.1.2. "Post-doctoral Research Aid" (Project id. N. 1.1.1.2/16/1/001) of the Republic of Latvia, funded by the European Regional Development Fund. Nadezda Spiridovska research project No. 1.1.1.2/VIAA/1/16/075 "Non-traditional regression models in transport modelling".

A two-component Markov process (X, J) is called a Markov-modulated process where component J is a Markov process as well and the next behavior of component X depends on the past behavior of the process at the current state of component J [1]. The first references in the literature to Markov-modulated processes related to Markov-additive process (MAPs) [2, 3]. MAPs form a rather general class of two-component stochastic processes, which include many important models such as Markov-modulated Brownian motion, Markov random walk, Markov renewal processes and some other. They seem particularly well suited in connection with matrix-exponential methods [4]. But in general Markov-modulated processes include such processes as MAPs, MAPs of arrivals and Markov-modulated birth-death processes. MAPs of arrivals are MAPs, the additive component X of which takes on values in non-negative (also multidimensional) integers, so that its increments can be interpreted as corresponding to arrivals, what is classically observed in the queuing systems as different classes of arrivals. MAPs of arrivals are important in terms of different issues and their significance lies in their application as parts of higher in complexity systems, such as queueing models and also data communication models.

Markov-modulated Poisson processes (MMPPs) are the brightest representatives of MAPs of arrivals and are one of the most commonly used and popular models in information communication systems and signal processing. The most cited source on MMPP is cookbook of Wolfgang Fisher and Kathleen Meier-Hellstern [5], published in 1993. The authors provide dozens of examples of the application of this type of model in the superposition of packetized voice processes and packet data, for example [6, 7]. Naturally, since 1993, the number of applications of MMPP has increased. For instance, the paper [8] presents a study on the use of MMPPs for characterizing multimedia traffic with short-term and long-term correlation. In [9] the Coupled Markov Modulated Poisson Processes (CMMPP) framework is presented and it demonstrates the feasibility of source traffic modeling for Machine-type Communication, and many such examples can be found in reputable scientific publications.

A detailed description and systematization of Markov-modulated processes in general, and Markov-additive

processes in particular can be found in [10].

Markov-modulated linear regression (MMLR) model, which is a core of this research, was firstly proposed by Alexander Andronov in [11]. This model is a special case of the Markov-additive process (MAP). According to [1] the Markov-additive process $(Y, J) = \{(Y(t), J(t)), t \geq 0\}$ is two-component Markov process defined on the state space $R \times N$, so that for $s, t \geq 0$ conditional increment $(Y(s+t) - Y(s), J(s+t))$ under the condition $(Y(s), J(s))$ depends only on $J(s)$, $R = (-\infty; \infty)$, $N = \{0, 1, \dots, m\}$. Component $J(s)$ is called Markov, and component Y is additive. For MMLR the increment of the additive component Y is described by a linear regression, and the component J is a continuous-time homogeneous irreducible Markov chain with the known transition rates $\lambda_{i,j}$ from state s_i to state s_j , and $\Lambda_i = \sum_{j \neq i} \lambda_{i,j}$.

Regime switching regression models share the same idea of varying the regression parameters randomly in accordance with external environment, but these models consider Markov chain as unobservable and estimation procedure involves estimation of transition matrix, while MMLR describes the external environment as a continuous-time homogeneous irreducible Markov chain with *known* parameters.

MMLR is a very young and "inexperienced" model and has not been widely used in various areas of life (currently only in transportation), but since this model refers to stochastic processes, it undoubtedly has a powerful potential and the prospect of widespread use.

Broad application of Markov processes is caused by exponential distribution properties or Markov property. The Markov property is a "forgetting property" suggesting memorylessness in the distribution of time a continuous-time Markov chain spends in any state. In other words, if T_j denotes the sojourn (holding) time in state j , for $j \neq i$, then T_j is exponentially distributed with transition rate λ_j and T_j is independent of T_i . Often there are situations in practice when data samples, describing the external environment, do not correspond to the exponential distribution. In the process of estimating the unknown parameters of the additive component $Y(t)$ a random environment creates a big randomness by itself, additionally, if memorylessness property is not satisfied, then the estimates and the results of modelling become absolutely unreliable. This research considers the case when it's possible to represent transition intensities of component $J(t)$ as a convolution of the exponential densities.

The paper is structured as follows. Section 2 describes the related work and outlines the novelty. Section 3 provides the theoretical background of the Markov-modulated linear regression model. Section 4 sheds light on the idea of convolution of the exponential densities and provides the necessary methodology. Section 5 describes all necessary formulas for an estimation of regression parameters taking into account convolution of the exponential densities. Section 6 covers the efficiency evaluation of such an approach by a simulation study. Finally, the paper discusses the findings and gives recommendations for future research and possible applications. The model and experiments are implemented in Mathcad 14.

II. RELATED WORK AND NOVELTY

The first mention of the MMLR in the literature contains only simulation-based validation of the model [11], which demonstrates that the small sample size leads to completely unsatisfactory results and the estimates are far from the true values of the unknown parameters. Nonetheless an increased sample size clearly improves the estimates. The subsequent representations of MMLR are related to the practical application of this model. Two cases studies are considered. In the first case, analysis of coaches' delay time at the Riga Coach Terminal for the period from 2012 to 2017 is mentioned [12]. The weather conditions in the city, from which coaches start, is chosen as the external environment with two states: "No precipitation" and "Precipitation". The data about weather conditions are obtained from the Latvian Environment, Geology and Meteorology Centre (LEGMC) database. When hypothesis about the sojourn time distribution (which supposed to be exponential) in each state is tested, it turns out that the distribution in some months does not live up to expectations (for, example see Fig.1). For that study, using approximation and aggregation, the decision is made to neglect the absence of exponentiality, but in general this problem requires a separate solution.

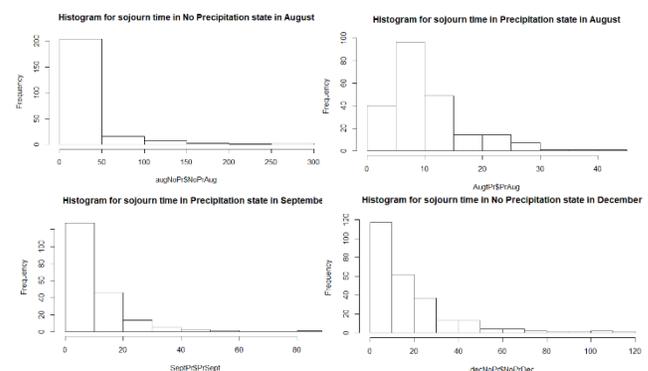


Fig.1. Histograms for distribution time in different months

The next approximation of the model is carried out on the data of trip validations provided by Rigas Satiksme, a provider of transport services in Riga [13, 14]. Initial data set covers time period of 4.5 months in 2017 and contains 1,048,001 observations. Weather conditions are also chosen as the external environment. Here the same situation arises when the sojourn time spent in each state does not obey the exponential distribution for some months.

What to do in the current situation of the absence of Markov property? There are several possible scenarios. Ignore this problem (and as a result potentially get unreliable model results), modify data (aggregate, re-group, etc.) if possible, or try to get an exponential distribution differently.

In connection with the results of the mentioned two cases, since the first scenario has already been completed, but data modification may be time- and resource-consuming, the idea arises to try to apply the procedure of approximation the inappropriate distribution by a convolution of the exponential densities, namely by hypoexponential distribution (or it is also called the generalized Erlang distribution), thereby preserving the memorylessness property. Then unknown MMLR model parameters are estimated using this kind of a convolution and

the proposed formulas, which is postulated as scientific novelty. In this study, the emphasis is made precisely on MMLR parameters estimation, assuming that the parameters of convolution were received earlier. The process of obtaining convolution parameters is omitted. This kind of research is completely original, firstly, because the model itself is new (a new analytical approach), and the estimation of its parameters using convolution, even more so.

In general, the usage of a convolution is a well-known approach in applied research. The simplest example is an Erlang distribution with the same parameter λ , followed by a hyperexponential and hypoeponential distributions [15, 16] with different parameters λ_i , and more general approach consists in application so called phase-type (PH) distribution [17]. Unfortunately, parameter's estimation of phase-type distribution is a difficult statistical problem. Therefore, the choice was reduced to a statistically simpler approach, namely the usage of the hypoeponential distribution, which is a special case of the phase-type distribution.

III. MARKOV-MODULATED LINEAR REGRESSION: THEORETICAL BACKGROUND

For a simpler presentation of the model description an example is considered with two states of the external environment ($m = 2$). Let $\Lambda_i = \sum_{j \neq i} \lambda_{i,j}$. For the observed case of Markov additive process, it is additionally supposed that, if on the interval $(s, s + t)$ Markov component state is j , then increment of additive component Y is described by linear regression:

$$Y(s + t) - Y(s) = xt\beta^{(j)} + Z\sqrt{t}, i = 1, \dots, n, \quad (1)$$

where $x = (x_1, x_2, \dots, x_k)$ is $1 \times k$ vector of regressors, $\beta^{(j)}$ is $k \times 1$ vector of unknown regression coefficients, which are different and depend on the state: $\beta^{(1)} = (\beta_{1,1}, \beta_{2,1}, \dots, \beta_{k,1})^T$ and $\beta^{(2)} = (\beta_{1,2}, \beta_{2,2}, \dots, \beta_{k,2})^T$ in case of alternating regression ($m = 2$), Z is disturbance scale with usual assumptions: Z doesn't depend on $Y(s)$ and j , has normal distribution with mean zero and constant variance σ^2 .

Further time interval $(0, \tau)$ is considered with initial values of Markov additive process $Y(0) = 0, J(0) = i$. Let denote by T_j total sojourn time of Markov component in state j during the interval $(0, \tau)$, $\vec{T} = (T_1, T_2)$. Obviously $T_1 + T_2 = \tau$. Let Z_1, Z_2 are disturbance terms which take place for the time moments T_1, T_2 , accordingly. Then additive component Y can be rewritten:

$$Y(\vec{T}) = x(\beta^{(1)}T_1 + \beta^{(2)}T_2) + (Z_1\sqrt{T_1} + Z_2\sqrt{T_2}). \quad (2)$$

Since Z_1, Z_2 and \vec{T} are independent and taking into account normal distribution properties:

$$E(Z_1\sqrt{T_1} + Z_2\sqrt{T_2}) = 0, D(Z_1\sqrt{T_1} + Z_2\sqrt{T_2} | \vec{T} = const) = (T_1D(Z_1) + T_2D(Z_2)) = \sigma^2\tau,$$

formula (2) can be rewritten:

$$Y(\vec{T}) = x(\beta^{(1)}T_1 + \beta^{(2)}T_2) + \sqrt{\tau}Z. \quad (3)$$

Using Kronecker product \otimes and vec operator it is possible to represent the model in standard form of linear regression:

$$Y(\vec{T}) = (\vec{T} \otimes x)vec\beta + \sqrt{\tau}Z, \quad (4)$$

where $\beta = (\beta^{(1)} \beta^{(2)})$ is a matrix composed of columns $\{\beta^{(j)}\}$.

It is necessary to estimate regression coefficients $\beta = (\beta^{(1)} \beta^{(2)})$ and variance of disturbance term σ^2 . While all parameters of Markov chain describing Markov component are known: number of states and transition intensities $\{\lambda_{i,j}\}$. Estimation is based on n independent observations of the process (4). The whole trajectory of the environment J is *unknown*, therefore sojourn times in each state are unknown for typical r -th observation $\vec{T}_r = (T_{r,1}, T_{r,2})$. However instead of actual sojourn times the estimated conditional average sojourn times $\vec{t}_r = E(\vec{T}_r) = (E(T_{r,1}), E(T_{r,2})) = (t_{r,1}, t_{r,2})$ will be used in the state j .

The model in matrix notation can be represented as follows:

$$Y = (Y_1, \dots, Y_n)^T = \begin{pmatrix} \vec{t}_1 \otimes x_1 \\ \vec{t}_2 \otimes x_2 \\ \dots \\ \vec{t}_n \otimes x_n \end{pmatrix} vec \beta + \\ + diag(\sqrt{t_1}, \sqrt{t_2}, \dots, \sqrt{t_n})Z \quad (5)$$

where n -vectors: $Y = (Y_1, \dots, Y_n)^T$ with scale responses, $Z = (Z_1, \dots, Z_n)^T$ with disturbance terms, $\sqrt{\tau} = (\sqrt{t_1}, \dots, \sqrt{t_n})$ with observation times; m -vector $\vec{t}_r = (t_{r,1}, t_{r,2})$ with estimated conditional average sojourn times; $n \times k$ matrix with regressors $X = (x_{r,k}) = (x_r: r = 1, \dots, n)$, the n -dimensional diagonal matrix $diag(\alpha)$ with the vector α on the main diagonal.

Further it is supposed that matrix $X = ((\vec{t}_1 \otimes x_1)^T (\vec{t}_2 \otimes x_2)^T \dots (\vec{t}_n \otimes x_n)^T)^T$ of size $n \times km$ has rank $r(X) = km$, so $(X^T X)^{-1}$ exists. In this case generalized least squares method (GLSM) gives the following estimates of regression coefficients [18, 19, 20]:

$$vec\tilde{\beta} = (X^T W^{-1} X)^{-1} X^T W^{-1} Y, \quad (6)$$

where $W = diag(w_1, w_2, \dots, w_n)$ is non-degenerate diagonal weight matrix. In case of Ordinary Least Squares (OLS) weight matrix is an identity matrix ($W = I$).

IV. CONVOLUTION OF THE EXPONENTIAL DENSITIES

It is supposed that data sample, describing the external environment J , does not correspond to the exponential distribution. If distribution type of the given sample is known, the following steps can be done: firstly, it is necessary to estimate the unknown parameters of the distribution; the second phase involves the approximation of corresponding density by a convolution of exponents.

Let Z_1, \dots, Z_p be independent random variables having exponential distributions with parameters $\lambda = (\lambda_1, \dots, \lambda_p)$, where all components $\{\lambda_i\}$ are different. A distribution of their sum $S = Z_1 + \dots + Z_p$ has the following density and cumulative distribution function for $z \geq 0$, see [21]:

$$f(z; \lambda) = \left(\prod_{i=1}^p \lambda_i \right) \sum_{i=1}^p \exp(-\lambda_i z) \cdot \left(\prod_{j=1, j \neq i}^p (\lambda_j - \lambda_i) \right)^{-1} \quad (7)$$

$$F(z; \lambda) = \left(\prod_{i=1}^p \lambda_i \right) \sum_{i=1}^p (1 - \exp(-\lambda_i z)) \cdot \left(\lambda_i \prod_{j=1, j \neq i}^p (\lambda_j - \lambda_i) \right)^{-1} \quad (8)$$

Moment of order r is calculated by formula

$$\mu_r(\lambda) = \left(\prod_{i=1}^p \lambda_i \right) \sum_{i=1}^p r! (\lambda_i^{r+1} \prod_{j=1, j \neq i}^p (\lambda_j - \lambda_i))^{-1}.$$

This distribution is known as hypoexponential.

Expectation of the random variable S is calculated by the following formula:

$$E(S) = \sum_{i=1}^p \frac{1}{\lambda_i} \quad (10)$$

If some (given) sample corresponds to some differentiable density $g(z)$ of a non-negative continuous random variable, then it is possible to approximate the density $g(z)$ by the density (7).

There are several possible solutions to this issue. One of them is to use the moments of the density $g(z)$ for approximation. Both the classical method of moments and its modification can be used, for example:

$$R_M(\lambda) = \sum_{r=1}^k |\mu_r(\lambda) - \mu_r^*|^{1/r}, \quad (11)$$

where μ_r^* are an empirical moment of the r -th order of density $g(z)$, $k \geq p$, $\mu^* = (\mu_1^*, \dots, \mu_k^*)$ and $\mu_r(\lambda)$ are moments of hypoexponential distribution, calculated by (9) (number of used moments must be equal to k). A value of λ , which minimizes this criterion, gives necessary estimate. This optimization problem can be solved using an enumeration approach.

Other way for the approximation is use as criterion of the estimation a square between curves $g(z)$ and $f(z; \lambda)$:

$$R_S(\lambda) = \int_0^\infty v(z) |f(z; \lambda) - g(z)| dz, \quad (12)$$

where $v(z) \geq 0$ is a known "weight" function. Here the multiplier $v(z)$ allows the deviations to get various weights $|f(z; \lambda) - g(z)|$ for various z . A minimization of integral (12/6) can be done through gradient method.

Important to underline that it is possible to use a mixture of various estimates of vector parameter λ . It is easier to apply equal weights of mixture's components. But at the same time, it is possible to select weights based on some assumptions or speculations. This procedure can be controlled by calculating the criteria (11) and (12).

The whole procedure of a parametrical estimation for a convolution of exponential densities is described in [22].

Next, returning back to the MMLR model, the situation is considered when the vector $\lambda = (\lambda_1, \dots, \lambda_p)$ for the sample is already matched, i.e. the components of this vector are already known.

Each random variable which corresponds to the state of the external environment is represented by a convolution of exponential densities with parameters $\lambda = (\lambda_1, \dots, \lambda_p)$, therefore the total number of states will be equal to mp . Moreover, the process is assumed to be cyclical with the possibility of transition to the next state in a row, and then the transition from the last state to the first state follows. The matrix with the intensities of such a process will have the following form:

$$\lambda = \begin{pmatrix} 0 & \lambda_1 & 0 & 0 & & \\ 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & 0 & \lambda_p & & \\ & \vdots & & & \ddots & \vdots \\ & & 0 & & \dots & \lambda_{(mp)} \end{pmatrix}$$

In order to calculate the average sojourn time spent in each state of the external environment, it is necessary to use (10).

V. MODULATING MARKOV CHAIN

Taking into account convolution of exponential densities let's consider transition probabilities, average sojourn time and covariance matrix calculation. For transition probabilities $p_{i,j}(t) = P\{J(t) = j | J(0) = i\}$ of described Markov chain a usual system of differential equations take place. Let $P(t) = (p_{i,j}(t))$ and $\lambda = (\lambda_{i,j})$ be the $m \times m$ matrices (in case of $m = 2$, so 2×2), Λ be m -dimensional diagonal matrix with vector $(\Lambda_1, \dots, \Lambda_m)$ on the main diagonal. Then, if p random variables make a convolution, then $P^{(*)}(t) = (p_{i,j}^{(*)}(t))$ and $\lambda^{(*)} = (\lambda_{i,j}^{(*)})$ be the $mp \times mp$ matrices (in case of $m = 2$, so $2p \times 2p$), Λ be mp -dimensional diagonal matrix with vector $(\Lambda_1, \dots, \Lambda_{mp})$ on the main diagonal. The matrix $P^{(*)}(t)$ can be represented by matrix exponent [Pacheco]:

$$P^{(*)}(t) = \exp(t(\lambda^{(*)} - \Lambda^{(*)})), \quad t \geq 0, \quad (8)$$

where $P^{(*)}(0) = I$.

If all eigenvalues of matrix $A = \lambda^{(*)} - \Lambda^{(*)}$ are different then solution (8) can be represented simpler. Let v_η and Z_η , $\eta = 1, \dots, mp$, be the eigenvalue and the corresponding eigenvector of A , $Z = (Z_1, \dots, Z_{mp})$ the matrix of the eigenvectors and by $\bar{Z} = Z^{-1} = (\bar{Z}_1^T, \dots, \bar{Z}_{mp}^T)^T$ the corresponding inverse matrix (here \bar{Z}_η is the η -th row of \bar{Z}). Then

$$P^{(*)}(t) = \exp(tA) = \text{diag}(\exp(v_1 t), \dots, \exp(v_{mp} t)) Z^{-1} = \sum_{\eta=1}^{mp} Z_\eta \exp(v_\eta t) \bar{Z}_\eta \quad (9)$$

For the conditional average sojourn time $t_{r,v}(\tau) = E(T_{r,v} | t_r = \tau, J_{r,0} = i, J_{r,\tau} = j)$ in the state $v \in S$ on the interval $(0, \tau)$ we have

$$t_{r,v}(\tau) = \frac{1}{p_{i,j}(\tau)} \int_0^\tau p_{i,v}(u) p_{v,j}(\tau - u) du. \quad (10)$$

$$\beta := \begin{pmatrix} 0 & 2.5 \\ 1 & 3 \\ 2 & 4 \end{pmatrix}$$

Since new number of states mp are equal to six the matrix with parameters of the regression model must be redefined:

$$\beta_{\text{multy}} = \begin{pmatrix} 0 & 0 & 0 & 2.5 & 2.5 & 2.5 \\ 1 & 1 & 1 & 3 & 3 & 3 \\ 2 & 2 & 2 & 4 & 4 & 4 \end{pmatrix}.$$

The following simulation experiment is organized. The last includes q independent batches. Each batch has the above described structure: the same random environment ($mp = 6$), number of observations $n = 30$, regressors' number $k = 3$. The matrix of regressors, the initial state of random environment and the observations' times are chosen at random according to the following distributions:

- the expectation of the matrix of regressors coincides with the above obtained matrix X , all elements of the 2nd and 3rd columns are independent and uniformly distributed on intervals $(-1, 1)$ and $(-0.5, 0.5)$ correspondingly;
- the expectation of observations' times coincides with previous values t , all times are independent and time of i -th observation t_i , has uniform distribution on $(0, 2t_i)$;
- the initial states I for various observations are independent and are chosen with respect to stationary distribution of the states for the random environment J .

Therefore, the total number of the observation for one experiment equals $n = 30q$.

External environment is presented by the following matrices: Generator (AA) and eigenvalues (χ) (Fig.4), also matrix of eigenvectors (M) (Fig.5.).

$$AA = \begin{pmatrix} -0.55 & 0 & 0 & 0 & 0 & 0.7 \\ 0.55 & -0.39 & 0 & 0 & 0 & 0 \\ 0 & 0.39 & -0.42 & 0 & 0 & 0 \\ 0 & 0 & 0.42 & -0.5 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & -0.6 & 0 \\ 0 & 0 & 0 & 0 & 0.6 & -0.7 \end{pmatrix} \quad \chi = \begin{pmatrix} 0 \\ -0.263 + 0.438i \\ -0.263 - 0.438i \\ -0.79 + 0.437i \\ -0.79 - 0.437i \\ -1.054 \end{pmatrix}$$

Fig.4. Generator (AA) and eigenvalues (χ)

$$M = \begin{pmatrix} 0.368 & 0.12 + 0.415i & 0.12 - 0.415i & 0.55 & 0.55 & -0.598 \\ 0.519 & 0.521 & 0.521 & -0.344 - 0.377i & -0.344 + 0.377i & 0.495 \\ 0.482 & 0.147 - 0.411i & 0.147 + 0.411i & -0.044 + 0.345i & -0.044 - 0.345i & -0.305 \\ 0.405 & -0.246 - 0.274i & -0.246 + 0.274i & 0.25 - 0.123i & 0.25 + 0.123i & 0.231 \\ 0.338 & -0.332 + 0.025i & -0.332 - 0.025i & -0.222 - 0.189i & -0.222 + 0.189i & -0.254 \\ 0.289 & -0.21 + 0.245i & -0.21 - 0.245i & -0.188 + 0.344i & -0.188 - 0.344i & 0.431 \end{pmatrix}$$

Fig.5. Matrix of eigenvectors

Based on vector of eigenvalues and matrix of eigenvectors matrix with transition probabilities ($t = 5$) is calculated using (9), see Fig.6.

$$Pr(5) = \begin{pmatrix} 0.093 & 0.282 & 0.285 & 0.186 & 0.102 & 0.052 \\ 0.066 & 0.175 & 0.267 & 0.235 & 0.16 & 0.097 \\ 0.147 & 0.093 & 0.155 & 0.222 & 0.215 & 0.169 \\ 0.236 & 0.199 & 0.086 & 0.113 & 0.172 & 0.194 \\ 0.248 & 0.304 & 0.169 & 0.072 & 0.078 & 0.129 \\ 0.171 & 0.334 & 0.248 & 0.13 & 0.06 & 0.057 \end{pmatrix}$$

Fig.6. Matrix with transition probabilities for the time period $t = 5$

For longer time period (for example $t = 100$) stationary state distribution is obtained (compare to (15)), see Fig.7.

$$Pr(100) = \begin{pmatrix} 0.153 & 0.216 & 0.201 & 0.169 & 0.141 & 0.12 \\ 0.153 & 0.216 & 0.201 & 0.169 & 0.141 & 0.12 \\ 0.153 & 0.216 & 0.201 & 0.169 & 0.141 & 0.12 \\ 0.153 & 0.216 & 0.201 & 0.169 & 0.141 & 0.12 \\ 0.153 & 0.216 & 0.201 & 0.169 & 0.141 & 0.12 \\ 0.153 & 0.216 & 0.201 & 0.169 & 0.141 & 0.12 \end{pmatrix}$$

Fig.7. Matrix with transition probabilities for the time period $t = 100$

As a calculation example of expectation of a sojourn time for six states and *united* expectation of a sojourn time for two states (if the initial state is 0 and total duration equals to 10) according to (10):

$$E\tau(0, 10)^T = (2.288 \quad 2.747 \quad 2.093 \quad 1.395 \quad 0.897 \quad 0.58);$$

$$E\tau(0, 10)^T = (7.128 \quad 2.872).$$

The covariance matrix obtained for six states and *united* covariance matrix for two states (if the initial state is 0 and duration equals to 10) according to (11):

$$Cov\mu_2(0, 10) = \begin{pmatrix} 3.265 & -1.008 & -0.974 & -0.692 & -0.405 & -0.187 \\ -1.008 & 4.562 & -1.286 & -1.055 & -0.74 & -0.474 \\ -0.974 & -1.286 & 3.415 & -0.457 & -0.399 & -0.299 \\ -0.692 & -1.055 & -0.457 & 2.224 & 0.018 & -0.039 \\ -0.405 & -0.74 & -0.399 & 0.018 & 1.377 & 0.149 \\ -0.187 & -0.474 & -0.299 & -0.039 & 0.149 & 0.849 \end{pmatrix}$$

$$UnionCov\mu_2(0, 10) = \begin{pmatrix} 4.707 & -4.707 \\ -4.707 & 4.707 \end{pmatrix}$$

Regression coefficients are estimated by formula (6). Different weight matrices are used: identity matrix, matrix with the inverse time observations and the matrix with inverse variance of the response variable (according to (14)). Initial β values are assumed to be equal to one (identity vector I).

Table 1. Convergence of estimates using the identity weight matrix $W = I, \sigma^2 = 1$

n	Real value	30	300	600	1500	3000	6000	9000
$\beta_{1,1}$	0	0.13	-0.79	-1.02	-0.07	-0.27	-0.03	-0.02
$\beta_{1,2}$	1	1.17	1.17	1.11	1.07	1.09	0.99	1.00
$\beta_{1,3}$	2	1.77	2.17	2.32	2.04	2.05	2.07	2.02
$\beta_{2,1}$	2.5	1.94	1.33	2.69	2.85	2.29	2.26	2.32
$\beta_{2,2}$	3	2.94	3.23	2.79	2.80	3.18	3.02	3.01
$\beta_{2,3}$	4	4.63	4.39	4.17	3.91	3.96	4.05	4.09

In case of the identity weight matrix, non-uniformity of observations is not taken into account, variance $\sigma^2 = 1$ and its estimation is not required. Table 1 demonstrates that even after hundreds of iterations, estimates of some coefficients are distinguished by some deviations from the true parameters.

Table 2. Convergence of estimates using the weight matrix $W = \text{diag}((\tau_1)^{-1}, \dots, (\tau_n)^{-1}), \sigma^2 = 1$

n	Real value	30	300	600	1500	3000	6000	9000
$\beta_{1,1}$	0	1.06	-0.42	0.23	-0.07	-0.01	-0.12	-0.06
$\beta_{1,2}$	1	0.80	1.01	0.89	1.01	1.06	1.02	1.01
$\beta_{1,3}$	2	1.69	2.29	1.95	1.99	1.99	2.07	2.01
$\beta_{2,1}$	2.5	3.18	2.28	2.33	2.25	2.35	2.25	2.27
$\beta_{2,2}$	3	2.85	3.13	3.21	3.09	3.03	3.03	3.04

$\beta_{2,3}$	4	3.68	4.07	3.82	4.08	4.05	4.06	4.05
---------------	---	------	------	------	------	------	------	------

Table 2. shows the results of MMLR parameters' estimation using weight matrix with the inverse time observations. This type of weight matrix helps to get rid of the problem of data unevenness, for balancing variables. As can be seen from Table 2 the results of estimation are comparable with the results from Table 1.

Table 3. Convergence of estimates using the weight matrix
 $W = \text{diag}(D(Y_1)^{-1}, \dots, D(Y_n)^{-1}), \sigma^2 = 1$

n	Real value	30	300	600	1500	3000	6000	9000
$\beta_{1,1}$	0	0.61	-0.18	-0.29	-0.02	-0.02	0.03	0.06
$\beta_{1,2}$	1	0.85	1.11	1	1.02	1.01	0.97	0.99
$\beta_{1,3}$	2	1.79	2.02	2.15	1.98	1.99	2.01	1.98
$\beta_{2,1}$	2.5	1.24	2.25	2.49	2.57	2.42	2.37	2.64
$\beta_{2,2}$	3	3.19	2.96	3.03	3.06	2.97	3.06	2.97
$\beta_{2,3}$	4	4.36	4.08	4.01	3.96	4.03	4.01	3.99

Table 3 shows the case of the diagonal weight matrix, in which the diagonal elements are the inverse values of the variances of the dependent variable. Such a weight matrix is applied to minimize the variance of the estimates and to improve the convergence. In practical settings the unknown parameters β and σ^2 are subject to estimation as well. However, this experiment assumes that their values are known, for example, obtained in a previous study. Table 3 also shows that in order to obtain satisfactory estimates the usage of such a weight matrix allows to reduce sample size, but the estimation time is significantly increasing at the same time.

As can be seen from all three tables the speed of estimates convergence is quite high. Based on these results, conclusion may be done about the effectiveness of suggested procedure.

VIII. CONCLUSION

This research considers the case, when the Markov property is not satisfied, namely, the sojourn time in each state is not exponentially distributed.

Case of a Markov-modulated linear regression (MMLR) with two states of external environment is described. Each random variable which corresponds to the state of the external environment is represented by a convolution of exponential densities with parameters $\lambda = (\lambda_1, \dots, \lambda_p)$. Moreover, the process is assumed to be cyclical with the possibility of transition only to the next state in a row.

Estimation procedure for unknown model parameters is described, the whole trajectory of the environment $J(.)$ is unknown and the estimated conditional average sojourn time is used instead of unknown sojourn times in the state s_j . Expressions for expectation and variance of response variable are represented.

Simulation study is considered when it's possible to represent transition intensities as a convolution of exponential densities. The proposed example considers a case with two states of the environment and three convolution parameters for each state. Regression coefficients are estimated. Different weight matrices are used: identity matrix, matrix with the inverse time observations and the matrix with inverse variance

of the response variable. All experiments show good results. Experimental study proves the effectiveness of suggested procedure.

ACKNOWLEDGMENT

The author would like to thank the Associate Editor for helping her in technical details and the Reviewers.

References

- [1] Pacheco A., Tang L.C., Prabhu N.U. (2009). Markov-Modulated Processes & Semiregenerative Phenomena. New Jersey – London; World Scientific.
- [2] Çinlar, Erhan: Markov additive processes. I. Probability Theory and Related Fields 24.2: 85-93 (1972).
- [3] Çinlar, Erhan: Markov additive processes. II. Probability Theory and Related Fields 24.2: 95-121 (1972).
- [4] Bladt M., Nielsen B.F. (2017) Markov Additive Processes. In: Matrix-Exponential Distributions in Applied Probability. Probability Theory and Stochastic Modelling, vol 81. Springer, Boston, MA.
- [5] Fischer, Wolfgang, and Kathleen Meier-Hellstern: The Markov-modulated Poisson process (MMPP) cookbook. Performance evaluation 18.2: 149-171 (1993).
- [6] H. Heffes and D.M. Lucantoni, A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance, IEEE J. Selected Areas Comm. 4(6) (1986) 856-868.
- [7] I. Ide, Superposition of interrupted Poisson processes and its application to packetized voice multiplexers, in: M. Bonatti, ed., Teletraffic Science for New Cost-Effective Systems, Networks and Services, Proc. 12th International Teletraffic Congress (ITC-12), Torino, Italy, 1-8 June 1988 (North-Holland, Amsterdam, 1989) 1399-1405.
- [8] Shah- Heydari, S., Le- Ngoc, T. (2000). MMPP models for multimedia traffic. Telecommunication Systems 15, 273–293. <https://doi.org/10.1023/A:1019199013546>
- [9] Laner, M., Svoboda, P., Nikaein, N., & Rupp, M. (2013, August). Traffic models for machine type communications. In ISWCS 2013; The Tenth International Symposium on Wireless Communication Systems (pp. 1-5). VDE.
- [10] Spiridovska N. (2020) Markov-Modulated Processes, Their Applications and Big Data Cases: State of the Art. In: Kabashkin I., Yatskiv I., Prentkovskis O. (eds) Reliability and Statistics in Transportation and Communication. RelStat 2019. Lecture Notes in Networks and Systems, vol 117. Springer, Cham
- [11] Alexander M.Andronov, Nadezda Spiridovska. (2012) Markov-Modulated Linear Regression. In proceedings' book: International conference on Statistical Models and Methods for Reliability and Survival Analysis and Their Validation (S2MRSA), Bordeaux, France, pp.24–28., arXiv:1901.09600v1.
- [12] Spiridovska N. (2018) A Quasi-Alternating Markov-Modulated Linear Regression: Model Implementation Using Data about Coaches' Delay Time, International

journal of circuits, systems and signal processing, Volume 12, pp.617-628.

- [13] Spiridovska N., Yatskiv (Jackiva) I. (2018) Public transport passenger flow analysis and prediction using alternating Markov-modulated linear regression, In 29th European Conference on Operational Research (Euro2018) handbook, p.208.
- [14] Irina Jackiva (Yatskiv), Nadezda Spiridovska (2019) Data Preparation Framework Development for Markov-Modulated Linear Regression Analysis. In: Kabashkin I., Yatskiv I., Prentkovskis O. (eds) Reliability and Statistics in Transportation and Communication. RelStat 2018. Lecture Notes in Networks and Systems, Springer, Cham (Scopus) DOI: 10.1007/978-3-030-12450-2_17
- [15] Nikita Ratanov (2015). Hypo-exponential distributions and compound Poisson processes with alternating parameters. *Statistics & Probability Letters*, ISSN: 0167-7152, Vol: 107, Page: 71-78, <https://doi.org/10.1016/j.spl.2015.08.006>
- [16] Liu, Xi, et al. "System reliability estimation of high power diode laser with hypo-exponential distribution." 2017 Annual Reliability and Maintainability Symposium (RAMS). IEEE, 2017.
- [17] Neuts M.F. 1981. *Matrix-geometric Solutions in Stochastic Models*. Baltimore: The Johns Hopkins University Press.
- [18] Rao C.R. *Linear statistical inference and its applications*. – New-York – London-Sydney: John Wiley&Sons Inc. – 1995.
- [19] Srivastava M.S. *Methods of Multivariate Statistics*. - New-York: Wiley-Interscience.- 2002.
- [20] Turkington D.A. *Matrix Calculus and Zero-One Matrices. Statistical and Econometric Applications*. – Cambridge: Cambridge University Press. – 2002. G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
- [21] Buslenko, N.P. 1968. *Complex System Modelling*. Moscow, Russia: "Nauka". (In Russian).
- [22] Andronov A., Spiridovska N., Santalova D. (2018) On a parametrical estimation for a convolution of exponential densities. In book of abstracts: The 27th Nordic conference in Mathematical Statistics (Nordstat2018), Tartu, Estonia, p.10.
- [23] Andronov, Alexander M., and Vladimir M. Vishnevsky.: *Reliability of Two Communication Channels in a Random Environment*. International Conference on Distributed Computer and Communication Networks. Springer, Cham, (2018).

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0

https://creativecommons.org/licenses/by/4.0/deed.en_US