Chapter 3
Markov-Modulated Samples and Their Applications

Alexander Andronov

3.1 Problem Setting

The classical sample theory supposes that sample elements are identically distributed and independent (i.i.d.) random variables. Lately a great attention has been granted to dependence in probabilistic structures, for example, dependence between interarrival times of various flows, between service times, etc. Usually it is described by the so-called Markov-modulated processes. They are used widely in environmental, medical, industrial, and sociological researches. We restrict ourselves by a case when elements of the sample are positive random variables. It is convenient to consider them as lifetimes of unreliable elements.

Let us consider sample elements $\{X_i, i = 1, \ldots, n\}$, modulated by a finite continuous-time Markov chain (see [6]). For simplicity we say that the elements operate in the so-called random *environment*. The last is described by an "external" continuous-time ergodic Markov chain $J(t), t \geqslant 0$, with a final state space $E = \{1, 2, \ldots, k\}$. Let $\lambda_{i,j}$ be the transition rate from state i to state j.

Additionally, n binary identical elements are considered. Each component can be in two states: *up*(1) and *down*(0). The elements of system fail one by one, in random order. For a fixed state $i \in E$, all n elements have the same failure rate $\gamma_i(t)$ and are stochastically independent. When the external process changes its state from i to j at some random instant t, all elements, which are alive at time t, continue their life with new failure rate $\gamma_j(t)$. If on interval (t_0, t) the random environment has state $i \in E$, then the residual lifetime $\tau_r - t_0$ (*up*-state) of the rth component, $r = 1, 2, \ldots, n$, has a cumulative distribution function (CDF) with failure rate $\gamma_i(t)$ for time moment t, and the variables $\{\tau_r - t_0, r = 1, 2, \ldots, n\}$ are independent.

A. Andronov (✉)

Transport and Telecommunication institute, 1 Lomonosova street, Riga, LV-1019, Latvia
e-mail: lora@mailbox.riga.lv

We wish to get statistical estimates for the unknown parameters $\beta^{(i)} = (\beta_{1,i}, \beta_{2,i},$ $\ldots, \beta_{m,i})^T$, $i = 1, \ldots, k$. Note that in the above described process elements of the sample $\{X_i, i = 1, \ldots, n\}$ are no i.i.d. anymore, as it is assumed in the classical sampling theory.

Further we make the following suppositions. Firstly, parameters of the Markov-modulated processes $\{\lambda_{i,j}\}$ are known. Secondly, with respect to hazard rates $\gamma_i(t)$, a parametrical setting takes place: all $\gamma_i(t)$ are known, accurate to m parameters $\beta^{(i)} = (\beta_{1,i}, \beta_{2,i}, \ldots, \beta_{m,i})^T$, so we will write $\gamma_i(t; \beta^{(i)})$. Further, we use the $(m \times k)$-matrix $\beta = (\beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(k)})$ of unknown parameters. Thirdly, with respect to the available sample: sample elements are fixed corresponding to their appearance, so the order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are fixed. Finally, the states of the random environment $J(t)$ are known only for time moments $0, X_{(1)}, X_{(2)}, \ldots, X_{(n)}$.

The maximum likelihood estimates (see [5, 8, 9]) for the unknown parameters β are derived. Results of a simulation study illustrate the elaborated technique. Presented paper continues our previous investigations [1, 2].

3.2   Transition Probabilities

In this section we cite a result from the paper of Andronov and Gertsbakh [3]. Define $N(t)$ as the number of elements which are in the *up* state at time moment t. Obviously $P\{N(0) = n\} = 1$. We denote

$$p_{r,i,j}(t_0, t) = P\{N(t) = r,\ J(t) = j\ |\ N(t_0) = r,\ J(t_0) = i\}, r \in \{1, \ldots, n\},\ i, j \in E,$$

$$p_{r,i}(t_0, t) = (p_{r,i,1}(t_0, t), \ldots, p_{r,i,k}(t_0, t))^T,\ P_r(t_0, t) = (p_{r,1}(t_0, t), \ldots, p_{r,k}(t_0, t)),$$

$$\Gamma(t, \beta) = \mathrm{diag}\left(\gamma_1(t, \beta^{(1)}), \ldots, \gamma_1(t, \beta^{(k)})\right),\ \Lambda = \mathrm{diag}\left(-\sum_{i=1}^{k} \lambda_{i,1}, \ldots, -\sum_{i=1}^{k} \lambda_{i,k}\right).$$

$$(3.1)$$

It has been shown that

$$\dot{P}_r(t_0, t) = -\left(\Lambda + r\Gamma(t, \beta)\right) P_r(t_0, t) + \lambda^T P_r(t_0, t), 0 \leqslant t_0 \leqslant t. \qquad (3.2)$$

Below we consider a simple time-homogeneous case when $\gamma_i(t; \beta^{(i)}) = \gamma_i(\beta_i)\ \forall i$ $\Gamma(t, \beta) = \Gamma(\beta) = \mathrm{diag}(\gamma_1(\beta^{(1)}), \ldots, \gamma_k(\beta^{(k)}))$. Therefore,

$$\dot{P}_r(t_0, t) = \left(\lambda^T - (\Lambda + r\Gamma(\beta))\right) P_r(t_0, t),\ 0 \leqslant t_0 \leqslant t.$$

In this case a solution can be represented by matrix exponent (see [4, 7]):

$$P_r(t_0, t) = \exp\left((t - t_0)\left(\lambda^T - (\Lambda + r\Gamma(\beta))\right)\right),\ 0 \leqslant t_0 \leqslant t. \qquad (3.3)$$

3.3 Maximum Likelihood Estimates

In the considered case, besides the initial state $j(0)$ of $J(0)$, a sample of size n is given: $(x, j) = \{(x_{j(r)}, j(r)), r = 1, \ldots, n\}$, where $x_{j(r)}$ is the rth order statistic of the sample and $j(r) = J(x_{(r)})$ is a corresponding state of the random environment. Setting $x_{(0)} = 0$ we rewrite the log-likelihood function as

$$\mathrm{ll}(\beta; (x, j)) = \sum_{r=0}^{n-1} [\ln p_{n-r,j(r),j(r+1)}(x_{(r)}, x_{(r+1)}; \beta)$$

$$+ \ln(n - r) + \ln \gamma_{j(r+1)}(x_{(r+1)}; \beta^{(j(r+1))})]. \tag{3.4}$$

Considering gradients with respect to the column vectors $\beta^{(v)}$, $v = 1, \ldots, k$, we get maximum likelihood equations

$$\frac{\partial}{\partial \beta^{(v)}} \mathrm{ll}(\beta; (x, j)) = \sum_{r=0}^{n-1} \frac{1}{p_{n-r,j(r),j(r+1)}(x_{(r)}, x_{(r+1)}; \beta)}$$

$$\times \frac{\partial}{\partial \beta^{(v)}} p_{n-r,j(r),j(r+1)}(x_{(r)}, x_{(r+1)}; \beta)$$

$$+ \sum_{r=0}^{n-1} \frac{1}{\gamma_{j(r+1)}(x_{(r+1)}; \beta^{(j(r+1))})}$$

$$\times \frac{\partial}{\partial \beta^{(v)}} \gamma_{j(r+1)}(x_{(r+1)}; \beta^{(j(r+1))}) = 0, \ v = 1, \ldots, k. \tag{3.5}$$

Further, we consider a time-homogeneous case when all rate intensity $\gamma_i(\beta^{(i)})$ have one unknown scalar parameter β_i only, so $\gamma_i(\beta^{(i)}) = \beta_i$, $i = 1, \ldots, k$. We will write $p_{m,i,j}(t - t_0) = p_{m,i,j}(t_0, t)$. Then, the likelihood equations (3.5) have the following form:

$$\frac{\partial}{\partial \beta^{(v)}} \mathrm{ll}(\beta; (x, j)) = \sum_{r=0}^{n-1} \frac{1}{p_{n-r,j(r),j(r+1)}(x_{(r+1)} - x_{(r)}; \beta)}$$

$$\times \frac{\partial}{\partial \beta^{(v)}} p_{n-r,j(r),j(r+1)}(x_{(r+1)} - x_{(r)}; \beta)$$

$$+ \frac{1}{\beta_v} \sum_{r=0}^{n-1} \delta_{v,j(r+1)} = 0, \tag{3.6}$$

where $\delta_{v,j(r+1)}$ is the Kronecker symbol: $\delta_{v,j(r+1)} = 1$ if $v = j(r + 1)$, and $\delta_{v,j(r+1)} = 0$ otherwise.

Now we must get an expression for the derivative $\frac{\partial}{\partial \beta^{(v)}} p_{n-r,j(r),j(r+1)}(x_{(r+1)} - x_{(r)}; \beta)$. For that we use an expression for a derivative of a matrix exponent (see Lemma of Appendix). Let D_v be a square matrix from zero, where only one non-zero element equals 1 and takes the vth place of a main diagonal. Then, for the homogeneous case when $\Gamma(\beta) = \text{diag}(\gamma_1(\beta), \ldots, \gamma_k(\beta)) = \Gamma = \text{diag}(\beta_1, \ldots, \beta_k)$, according to (3.3), we have for $v = 1, \ldots, k$:

$$\frac{\partial}{\partial \beta_v} P_r(t_0, t) = \frac{\partial}{\partial \beta_v} \exp\{(t - t_0)(\lambda^T - \Lambda - r\Gamma)\} = \sum_{i=1}^{\infty} \frac{1}{i!}(t - t_0)^i$$

$$\sum_{j=0}^{i-1} (\lambda^T - \Lambda - r\Gamma)^j \left(-rD_v \frac{\partial}{\partial \beta_v}\beta_v\right)(\lambda^T - \Lambda - r\Gamma)^{i-1-j} =$$

$$-r \sum_{i=1}^{\infty} \frac{1}{i!}(t - t_0)^i \sum_{j=0}^{i-1}(\lambda^T - \Lambda - r\Gamma)^j D_v (\lambda^T - \Lambda - r\Gamma)^{i-1-j}.$$

Therefore

$$\frac{\partial}{\partial \beta_v} P_r(t_0, t) = -r \sum_{i=1}^{\infty} \frac{1}{i!}(t - t_0)^i \sum_{j=0}^{i-1} \left((\lambda^T - \Lambda - r\Gamma)^j\right)^{\langle v \rangle}$$

$$\left((\lambda^T - \Lambda - r\Gamma)^{i-1-j}\right)_{\langle v \rangle}, \tag{3.7}$$

where $M^{\langle v \rangle}$ and $M_{\langle v \rangle}$ mean γth column and γth row of matrix M.

Now we can use a numerical method for the solution of the likelihood equation (3.6). Note that parameter β_v can be non-trivially estimated if state v has been registered as some $j(r) = v, r = 0, 1, \ldots, n$.

3.4   Simulation Study

Below there are the results of a simulation study presented, they are performed for an analysis of the described estimating procedure efficiency. As initial data, data from the paper [3] have been used. Let us describe one. A random environment has three states ($k = 3$, $E = \{1, 2, 3\}$). Transition intensities $\{\lambda_{i,j}\}$ from state i to state j, $(i, j = 1, 2, 3)$ are given by a matrix

$$\lambda = (\lambda_{i,j}) = \begin{pmatrix} 0 & 0.2 & 0.3 \\ 0.1 & 0 & 0.2 \\ 0.4 & 0.2 & 0 \end{pmatrix}. \tag{3.8}$$

Let a number of the considered elements n equals 5. For the environment state $i \in E$, all elements have a constant failure rate $\gamma_i(t) = \beta_i$ and they fail independently. Therefore, a last time till a given element failure (for the same state i of the environment) has the exponential distribution with parameter β_i. These parameters must be estimated. For that purpose a sample is given. It contains a sequence of $n + 1$ pairs: $(x, j) = \{(x_{j(r)}, j(r)), r = 0, \ldots, n\}$, where $x_{j(r)}$ is the rth order statistic of n-sample, and $j(r) = J(x_{(r)})$ is an environment state in the instant $x_{(r)}$. The initial pair $(x_{(0)}, j(0))$ equals $(0, 1)$.

All the mentioned sampling data are given and are used in an estimating procedure. Own samples are simulated for the following parameter values: $\beta = (\beta_1 \ \beta_2 \ \beta_3)^T = (0.1 \ 0.2 \ 0.3)^T$. It is convenient to present these data as $3 \times (n + 1)$ matrix. An example of such matrix for $n = 5$ is the following:

$$Sample = \begin{pmatrix} r & 0 & 1 & 2 & 3 & 4 & 5 \\ x_{(r)} & 0 & 0.624 & 1.502 & 2.009 & 8.711 & 9.429 \\ J(x_{(r)}) & 1 & 1 & 2 & 1 & 3 & 3 \end{pmatrix}.$$

In the simulation process, samples are generated one by one. Various samples are independent. Each sample corresponds to the appointed initial state of the environment: a sample with number $3i + j$ corresponds to the initial state $J(0) = j$; $j = 1, 2, 3$; $i = 0, \ldots$. Further, q such three samples (with the initial states $j = 1, 2, 3$) form a block, containing $3q$ samples. A maximum log-likelihood estimate (MLE) $\tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3)^T$ is calculated for each block.

In broad outline, a procedure is as follows. For each sample, a changing of the environment $J(t)$, $t > 0$, and instants of element failure $x_{(r)}$, $r = 1, \ldots, n$, are simulated. Then, for the sample, a logarithm of likelihood function (3.4) and its gradient (3.5) or (3.6) are recorded. These expressions are used for MLE calculation. As an optimization method, the gradient method has been used.

The gradient method is given by the following parameters: n is a sample size (initial number of system elements); b_0 is an initial value of parameter estimate; d is a step of moving along the gradient; ε is a maximum module of a difference between sequential values of the parameter estimate β, for which a calculation is ended; L is a limit number of a gradient recalculation during moving from an initial point; K is a number of addends, appreciated in an expansion of the matrix exponent (3.7); $3q$ is a number of the samples in the block.

A set of such parameters numerical values $(n, b_0, d, \varepsilon, L, K, q)$ is called an experiment design. Below, the results of simulation study are presented.

In Table 3.1 the corresponding results are presented for the design experiment $n = 5$, $b_0 = (0.08 \ 0.22 \ 0.328)^T$, $d = 0.015$, $\varepsilon = 0.01$, $L = 20$, $K = 20$, $r = 5$, $q = 5$ and various values of total block number N. In the first column the initial value of the estimate $b_0 = (0.08 \ 0.22 \ 0.328)^T$ is written. The following columns there are the estimate values given by averaging over N blocks. For a big number of the blocks, the coefficients d and ε have been changed. Namely, for $N = 15, 17, 19, 21$ those values equal 0.002, and for $N = 21$, additionally, $L = 40$.

Table 3.1 Convergence of the estimates for initial value $b_0 = (0.08\ 0.22\ 0.328)^T$

N		1	2	3	4	5	6	7	8
$\tilde{\beta}_1$	0.080	0.076	0.103	0.119	0.111	0.097	0.088	0.098	0.101
$\tilde{\beta}_2$	0.220	0.217	0.208	0.211	0.219	0.219	0.213	0.211	0.213
$\tilde{\beta}_3$	0.328	0.219	0.303	0.333	0.351	0.316	0.280	0.256	0.292
N	9	10	11	12	13	15	17	19	21
$\tilde{\beta}_1$	0.094	0.109	0.103	0.098	0.103	0.100	0.101	0.102	0.102
$\tilde{\beta}_2$	0.209	0.210	0.207	0.209	0.209	0.208	0.208	0.208	0.208
$\tilde{\beta}_3$	0.267	0.298	0.279	0.280	0.298	0.283	0.290	0.295	0.298

An analysis of the Table 3.1 shows that a convergence to true values $(0.1\ 0.2\ 0.3)$ takes place but very slow.

In conclusion we would like to remark that considered approach allows improving probabilistic predictions for functioning of various complex technical and economical systems.

Appendix

Lemma 3.1. *If elements of matrix $G(t)$ are differentiable function of t, then*

$$\frac{\partial}{\partial t}G(t)^n = \sum_{i=0}^{n-1} G(t)^i \left[\frac{\partial}{\partial t}G(t)\right]G(t)^{n-1-i}, \ n = 1, 2, \ldots,$$

$$\frac{\partial}{\partial t}\exp(G(t)) = \frac{\partial}{\partial t}\sum_{i=0}^{\infty}\frac{1}{i!}G(t)^i = \sum_{i=1}^{\infty}\frac{1}{i!}\sum_{j=0}^{i-1}G(t)^j\left[\frac{\partial}{\partial t}G(t)\right]G(t)^{i-1-j}.$$

Proof. Lemma 3.1 is true for $n = 1$ and 2. If one is true for $n > 1$, then

$$\frac{\partial}{\partial t}G(t)^{n+1} = \left[\frac{\partial}{\partial t}G(t)\right]G(t)^n + G(t)\frac{\partial}{\partial t}G(t)^n$$

$$= \left[\frac{\partial}{\partial t}G(t)\right]G(t)^n + G(t)\sum_{i=0}^{n-1}G(t)^i\left[\frac{\partial}{\partial t}G(t)\right]G(t)^{n-1-i}$$

$$= \sum_{i=0}^{n}G(t)^i\left[\frac{\partial}{\partial t}G(t)\right]G(t)^{n-i}. \qquad \square$$

References

1. Andronov, A.M.: Parameter statistical estimates of Markov-modulated linear regression. In: Statistical Method of Parameter Estimation and Hypotheses Testing, vol. 24, pp. 163–180. Perm State University, Perm (2012) (in Russian)
2. Andronov, A.M.: Maximum Likelihood Estimates for Markov-additive Processes of Arrivals by Aggregated Data. In: Kollo, T. (ed.) Multivariate Statistics: Theory and Applications. Proceedings of IX Tartu Conference on Multivariate Statistics and XX International Workshop on Matrices and Statistics, pp. 17–33. World Scientific, Singapore (2013)
3. Andronov, A.M., Gertsbakh, I.B.: Signatures in Markov-modulated processes. Stoch. Models **30**, 1–15 (2014)
4. Bellman, R.: Introduction to Matrix Analysis. McGraw-Hill, New York (1969)
5. Kollo, T., von Rosen, D.: Advanced Multivariate Statistics with Matrices. Springer, Dordrecht (2005)
6. Pacheco A., Tang, L.C., Prabhu N.U.: Markov-Modulated Processes & Semiregenerative Phenomena. World Scientific, New Jersey (2009)
7. Pontryagin, L.S.: Ordinary Differential Equations. Nauka, Moscow (2011) (in Russian)
8. Rao, C.R.: Linear Statistical Inference and Its Application. Wiley, New York (1965)
9. Turkington, D.A.: Matrix calculus and zero-one matrices. Statistical and Econometric Applications. Cambridge University Press, Cambridge (2002)