Филиал Кемеровского государственного университета
в г. Анжеро-Судженске

Национальный исследовательский
Томский государственный университет

Кемеровский государственный университет

Институт проблем управления им. В. А. Трапезникова РАН

Институт вычислительных технологий СО РАН

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ (ИТММ–2015)

Материалы XIV Международной конференции имени А. Ф. Терпугова

18–22 ноября 2015 г.

Часть 1

Издательство Томского университета

2015

ВЕРОЯТНОСТНЫЕ И ЭКОНОМЕТРИЧЕСКИЕ МЕТОДЫ И МОДЕЛИ

CUSUM ALGORITHMS FOR PARAMETER ESTIMATION IN QUEUEING SYSTEMS WITH JUMP INTENSITY OF THE ARRIVAL PROCESS

Yu. Burkatovskaya, T. Kabanova, S. Vorobeychikov

National Research Tomsk Polytechnic University, Tomsk, Russia

National Research Tomsk State University, Tomsk, Russia

1. Introduction

Markovian arrival processes form a powerful class of stochastic processes introduced in [1] and [2] and thereafter they are widely used now as models for input flows to queueing systems where the rate of the arrival of customers depends on some external factors. MAP is a counting process whose arrival rate is governed by a continuous-time Markov chain. One of the problems connected with MAP is the estimation of intensity parameters by observing flow of events. A survey of estimation methods is given in [3]. Its emphasis is on maximum likelihood estimation and its implementation via the EM (expectation-maximization) algorithm. This approach is developed for different conditions in [4, 5], etc. The survey [6] with a huge bibliography is focused on matching moment method which is also widely used for parameter estimation in MAP because of its simplicity. This method is used, for example, in [7]. Bayesian approach based on the a posteriori probability of the controlling chain state is developed in [8]. Quality of those methods is typically examined via simulation.

In this paper we propose a different approach to MAP parameter estimation using the sequential analysis methods described in [9] and [10]. The key idea is to consider time intervals between arrivals as a stochastic process which parameters change in random instants. First we detect these points using sequential change point detection methods. Then we estimate the intensity parameters under the assumption that the intensity is constant between detected change points.

2. Problem statement

We consider a Markov-modulated poisson process, i.e. a flow of events, controlled by a Markovian chain with a continuous time. The chain has two states, transition between the states happens at random instants. The time of sojourn of the chain in the l-th state is exponentially distributed with the parameter α_l, $l=1,2$.

The flow of events has the exponential distribution with the intensity parameter λ_1 or λ_2 subject to the state of the Marcovian chain. The parameters of the system λ_1, λ_2 and the instants of switching between the states are supposed to be unknown. We also suppose that $\lambda_l << \alpha_l$, i.e. changes of the controlling chain states occur more rarely than observed events. The sequence of instants of arriving events is observed. The problem is to estimate the parameters λ_1, λ_2, α_1, α_2.

3. Algorithm 1

Consider the process $\{\tau_i\}_{i\geq 1}$, where $\tau_i = t_i - t_{i-1}$ is the length of the i-th interval between arriving events in the observed flow. If the controlling chain is in the l-th state then the mean length between events is equal to $1/\lambda_l$. So at the first stage of our procedure we try to detect the instants of the chain transition from one state to another as the instants of change in the mean of the process $\{\tau_i\}_{i\geq 1}$ using CUSUM procedures.

Let the parameters λ_1, λ_2 satisfy the condition

$$0 < \lambda_2 < \lambda_1; \quad 1/\lambda_2 - 1/\lambda_1 \geq d, \tag{1}$$

where d is a certain positive parameter. If the constant d is unknown then it can be chosen as the minimal difference between the mean lengths of the intervals τ_i when the controlling chain is in different states that should be detected. Let n be a lower bound of the mean number of events between switchings of the controlling chain states. For the model under consideration it means that $n\alpha_l < \lambda_l$. We suppose that n is rather large (for example, $n \geq 10$), hence changes of the controlling chain states occur more rarely than observed events. This situation is typical for real processes such as call-center or http-server because one of the states can be interpreted as a "usual" state of the system and another state as a "peak-time" state and during each of these states several customers are supposed to arrive. Besides processes having this property are often used for simulation study of algorithms for processes with jump intensity of customer arrivals (for example, see [5, 8]).

Choose then an integer parameter $k > 1$ describing the memory depth According to our analysis, a good choice of the parameter k is $k \approx n/2$. The idea is to compare the values τ_i and τ_{i-k}, $i > k$. If there are no changes of the controlling chain state within the interval $[t_{i-k-1}, t_i]$ then the values τ_i and τ_{i-k} have the identical exponential distribution with the mean $1/\lambda_1$ or $1/\lambda_2$. If the chain state changes within the interval $[t_{i-k-1}, t_i]$ then the expectations of the values τ_i and τ_{i-k} are different. On one hand the parameter k should allow us to detect changes with minimal delay, on the other hand it should not be too large to contain more than one chain state change within the interval $[t_{i-k-1}, t_i]$.

As the initial state of the chain is unknown, we shall consider two CUSUM procedures simultaneously. The first procedure is set up to detect increase in the mean of the process and hence, decrease of the intensity, and the second procedure is set up to detect decrease in the mean and hence, increase of the intensity. For the first procedure we choose the positive parameter $\Delta < d$ and introduce the sequence of the statistics

$$z_i^{(1)} = \tau_i - \tau_{i-k} - \Delta, \quad i > k. \tag{2}$$

For the second procedure we introduce the sequence of the statistics

$$z_i^{(2)} = \tau_{i-k} - \tau_i - \Delta, \quad i > k. \tag{3}$$

Consider then four hypothesis concerning the state of the controlling chain:

– $H_l(t_{i-k-1}, t_i)$ – the intensity of the arrival process on the interval $[t_{i-k-1}, t_i]$ is constant and equal to λ_l, $l = 1,2$;

– $H_{l,m}(t_{i-k}, t_{i-1})$ – the intensity of the arrival process on the interval $[t_{i-k}, t_{i-1}]$ changed once from λ_l to λ_m, $l = 1$, $m = 2$ or $l = 2$, $m = 1$;

The statistics $z_i^{(j)}$, $j \in \{1,2\}$ (2), (3) have the following properties:

$$\begin{aligned} &E\left[z_i^{(1)} \middle| H_l\right] < 0, \quad l = 1,2; \quad E\left[z_i^{(1)} \middle| H_{1,2}\right] > 0; \\ &E\left[z_i^{(2)} \middle| H_l\right] < 0, \quad l = 1,2; \quad E\left[z_i^{(2)} \middle| H_{2,1}\right] > 0. \end{aligned} \tag{4}$$

So the means of statistics (2), (3) change from negative to positive values when the intensity of the process changes. These properties determine the construction of the procedures. We introduce positive constants h_1 and h_2 as the procedures thresholds and construct the cumulative sums $S_i^{(1)}$ and $S_i^{(2)}$ which are recalculated at the instants t_i. It is defined as follows

$$\begin{gathered} S_0^{(l)} = \Delta; \quad l = 1,2 \\ S_i^{(l)} = \max\{0, S_{i-1}^{(l)} + z_i^{(l)}\}, \quad i > k; \\ S_i^{(l)} = 0, \quad if \quad S_i^{(l)} \ge h_l. \end{gathered} \tag{5}$$

Reaching the threshold h_l by the sum $S_i^{(l)}$ results in a decision considering the parameters changes; $l = 1$ indicates decision on increase of the mean length of the interval between the events and decrease of the process intensity, $l = 2$ indicates the opposite.

Let the sequence $\left\{\sigma_m^{(l)}\right\}_{m \ge 0}$ be the sequence of the instants when the cumulative sum in the l-th procedure reaches the threshold h_l, i.e.

$$\sigma_0^{(l)} = 0; \;\; \sigma_m^{(l)} = \min\left\{t_j > \sigma_{m-1}^{(l)} : S_j^{(l)} \ge h_l\right\}. \tag{6}$$

Consider a sequence $\left\{n_i^{(l)}\right\}_{i \ge 0}$ associated with the sequence $\left\{\sigma_m^{(l)}\right\}_{m \ge 0}$ as follows

$$n_0^{(l)} = 0; \;\; n_m^{(l)} = \max\left\{t_j \le \sigma_m^{(l)} : S_j^{(l)} > 0, S_{j-1}^{(l)} = 0\right\}. \tag{7}$$

Thus the instant $n_m^{(l)}$ is the first instant when the cumulative sum becomes positive to reach then the threshold. We consider the instants $n_i^{(1)}$ ($n_i^{(2)}$) as the estimators for the instants when the mean length of the interval between the events increases (decreases).

When implementing the procedure it is possible to encounter false alarm situations. We shall record all the exceeding the thresholds by either first or the second cumulative sum. If the same sum reaches threshold several times in a row, we only record the first occurrence.

Thus the procedure for estimation of instants of intensity switching is described as follows. Calculate two cumulative sums given by equations (5). Then construct the sequences $\{\sigma_m^{(l)}\}$, $\{n_m^{(l)}\}$ defined by equations (6), (7). Let $n_1^{(1)} < n_1^{(2)}$, then the initial value of the intensity is equal to λ_1. Define the sequence

$$q_0 = 0;$$
$$q_{2l+1} = \min\{n_i^{(1)} : n_i^{(1)} > q_{2l}\}, \quad l \geq 0; \tag{8}$$
$$q_{2l+2} = \min\{n_i^{(2)} : n_i^{(2)} > q_{2l+1}\}, \quad l \geq 0.$$

The values $q_1, q_2, \ldots$ are calculated using formula (8) while it is possible. If

$$\{n_i^{(2)} : n_i^{(1)} > q_{2l}\} = \varnothing \quad \left(\{n_i^{(1)} : n_i^{(2)} > q_{2l+1}\} = \varnothing\right)$$

then we set $q_{2l+1} = N$ ($q_{2l+2} = N$), where N is the instant of the last occurrence. Here the odd instants q_{2l+1} are the estimators of the instants when the intensity changes from λ_1 to λ_2, and the even instants q_{2l+2} are the estimators of the instants when the intensity changes from λ_2 to λ_1.

Define estimators for the parameters λ_1, λ_2

$$\hat{\lambda}_1 = N_1/T_1, \quad \hat{\lambda}_2 = N_2/T_2, \tag{9}$$

where N_1 is the total number of events occurred at the intervals $[q_{2l}, q_{2l+1}]$, $q_{2l+1} \leq N$ and T_1 is the total length of these intervals; N_2 is the total number of events occurred at the intervals $[q_{2l+1}, q_{2l+2}]$, $q_{2l+2} \leq N$ and T_2 is the total length of these intervals; $l \geq 0$.

Define estimators for the parameters α_1, α_2

$$\hat{\alpha}_1 = L_1/T_1, \quad \hat{\alpha}_2 = L_2/T_2, \tag{10}$$

where L_1 is the total number of the switching points $q_{2l+1} \leq N$, L_2 is the total number of the switching points $q_{2l+2} \leq N$, $l \geq 0$.

The parameters Δ and h_i affect the characteristics of the CUSUM procedure, i.e., the mean delay and the mean time between false alarms (see [10]). If there are no additional conditions then the procedure is considered to be optimal when the probabilities of the false detection and the skip of the change are equal. It can be guaranteed by the choise of the parameter Δ as $\Delta \approx d/2$. If the memory depth is equal to k then the sum $S_i^{(l)}$ has to reach the threshold h_l in not more then k steps (while $Ez_l^{(1)} > 0$). It can be provided by the choise of h_l as $h_l < k\Delta \approx nd/4$. Note that the parameter h_l should not be significantly less than its upper bound because it can increase the number of false alarms.

4. Algorithm 2

Let we have a certain period of observation $[0, T]$ and N is the number of occurrences at the interval. First, we calculate the mean of the length between occurrences using the usual formula

$$\hat{\tau} = T/N. \qquad (11)$$

The value $\hat{\tau}$ exceeds the mean length of the interval τ_i when the controlling chain is in the first state, and vice versa, the mean length of the interval τ_i exceeds the value $\hat{\tau}$ when the controlling chain is in the second state, i.e.

$$1/\lambda_1 < E\hat{\tau} < 1/\lambda_2. \qquad (12)$$

We introduce the sequence of the statistics

$$z_i^{(1)} = \tau_i - \hat{\tau}.\ z_i^{(2)} = -\tau_i + \hat{\tau}. \qquad (13)$$

Using in Algorithm 1 statistics (13) instead of (2), (3) we obtain Algorithm 2. The parameter h_l can be chosen as $h_l < nd/4$, i.e. as at the first algorithm.

5. Numerical simulation

The model for the considered flow and the suggested algorithms was implemented with varying parameters. The results of the simulation for Algorithm 2 are presented in the table below.

T	λ_1	λ_2	α_1	α_2	h_1	h_2	$\hat{\tau}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
1000	5	1	0,3	0,2	0,5	0,5	0,3883	5,2336	1,2276	0,3220	0,1732
1000	5	1	0,3	0,2	0,8	0,8	0,3929	5,0591	1,2370	0,2573	0,1322
1000	5	1	0,3	0,2	1	1	0,4355	4,7475	1,1972	0,2587	0,1144
1000	5	1	0,1	0,2	0,5	0,5	0,2668	5,6804	2,0053	0,2912	0,2604
1000	5	1	0,1	0,2	0,8	0,8	0,2924	5,1544	1,6825	0,1180	0,1880
1000	5	1	0,1	0,2	1	1	0,2501	5,2498	2,6283	0,1207	0,1297
1000	5	2	0,1	0,2	0,5	0,5	0,2351	6,1085	2,8854	0,3632	0,2656
1000	5	2	0,1	0,2	0,8	0,8	0,2564	5,1785	2,8092	0,1652	0,1289
1000	5	2	0,1	0,2	1	1	0,2486	5,1949	2,8831	0,1219	0,1162
10000	5	1	0,3	0,2	0,8	0,8	0,3830	4,8379	1,3439	0,2316	0,1318
10000	5	1	0,3	0,2	1	1	0,3766	4,6783	1,4326	0,1917	0,1157

First, the quality of the proposed algorithms on the threshold parameters h_l was studied. Increasing of h_l leads to decreasing of probability for the cumulative sums to reach the thresholds and hence an intensity change can be undetected. It causes increasing of error of the estimators $\hat{\lambda}_i$ because of not correct estimation of the controlling chain current state. On the other hand, increasing of h_l leads to decreasing the total number of false alarms. These theoretical conclusions are supported by the simulation results. As the thresholds increase the estimators of the switching parameters $\hat{\alpha}_l$ decrease because less switching points are detected on the first stage of the procedures. In the Table for $h_1 = h_2 = 1$ one can see that the estimators $\hat{\alpha}_l$ considerably less the real values of the parameters α_l. The best results are obtained for $h_1 = h_2 = 0{,}8$ for all intensity parameter values. Thus, choice of the algorithm parameters is a rather difficult problem requiring further theoretical investigations.

Increasing of the simulation time from 1000 to 10000 does not influence significantly the estimators quality. This result stress the fact that the proposed algorithms can be used for a small sample size.

Conclusion

MAPs are used as models for real processes, particularly, for call-centers or http-server customers (see [3], [4]), healthcare systems (see [5]), etc. Input flow intensity estimation and pertinent model setup is necessary to develop dispatching rule, to calculate optimal number of servers, etc. The suggested algorithms do not need the distribution function of the observing flow and, hence, can be applied to parameter estimation of other types of flows.

This paper is supported by The National Research Tomsk State University Academic D.I. Mendeleev Fund Program (NU 8.1.55.2015 L) in 2014–2015.

References

1. Neuts, M. F. A versatile Markovian point process. // J. Appl. Probab. 1979. No. 16. P. 764–774.

2. Lucantoni D. M., Meier-Hellstern K. S., Neuts M. F. A single server queue with server vacations and a class of non-renewal arrival processes // Adv. Appl. Prob., 1990. Vol. 22. P. 676–705.

3. Asmussen S. Phase-type distributions and related point processes: Fitting and recent advances // Chakravarthy and Alfa (eds.), Matrix-analytic methods in stochastic models, NY: Marcel Dekker. Lect. Notes Pure Appl. Math. 1997. Vol. 183. P. 137–149.

4. Breuer, L., Kume, A. An EM algorithm for Markovian arrival processes observed at discrete times // Measurement, Modelling, and Evaluation of Computing Systems and Dependability and Fault Tolerance Lecture Notes in Computer Science. 2010. Vol. 5987. P. 242–258.

5. Okamura H., Dohi T., Trivedi K. S. Markovian Arrival Process Parameter Estimation With Group Data // IEEE/ACM Transaction on Networking, 2009. Vol. 17, No. 4, P. 1326–1340.

6. Gerhardt I., Nelson B. L. On capturing dependence in point processes: Matching moments and other techniques // Technical report, Northwestern univ. 2009.

7. Duffie D., Glynn P. Estimation of Continuous-Time Markov Processes Sampled at Random Time Intervals // Econometrica. 2004. Vol. 72. P. 1773–1808.

8. Gortsev A. M., Zuevich V. L. Optimal estimation of parameters of an asynchronous doubly stochastic flow of events with arbitrary number of the states. // Tomsk State University Journal of Control and Computer Science. 2011. Vol. 4(17). P. 25–40.

9. Page E. S. Continuous inspection schemes // Biometrica. 1956. Vol. 42, No. 1. P. 100–115.

10. Lorden G. Procedures for reacting to a change in distribution // Annals. Math. Statist. 1971. No. 42. P. 1897–1971.