LIETUVOS MATEMATIKOS RINKINYS. LMD DARBAI
52 tomas, 2011, 214–219

ISSN 0132-2818
www.mii.lt/LMR/

MCMC based modelling of queuing systems from empirical data

Mantas Landauskas, Eimutis Valakevičius

Faculty of Fundamental Sciences, Kaunas University of Technology
K. Donelaičio 73, LT-44029 Kaunas

E-mail: mantas.landauskas@ktu.lt; eimval@ktu.lt

Abstract. Markov chain Monte Carlo (MCMC) modelling technique requires one to be able to construct a proposal density. There is no universal way to achieve this. This paper considers the universal proposal selection technique based on the kernel density estimate. Two channel queuing system with a priority was modelled using this technique. Empirical data (the observed service times) and the rates of arrival processes are all the information used for simulating the system.

Keywords: Markov chain Monte Carlo, the kernel density estimate, queuing system.

Introduction

This paper considers the MCMC approach to modelling service times in the $M/G/2/\infty$ system with a priority. The system is depicted in Fig. 1. The probability distribution of the service times is evaluated from the data observed. The kernel density estimate (KDE) was used for this purpose.

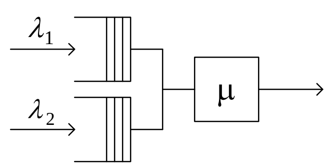


Fig. 1. The $M/G/2/\infty$ queuing system with a priority.

It could be difficult to draw samples from a distribution of a complicated or unknown analytical form. Having the empirical data it is possible, to construct a non-parametrical probability density estimate for it. Custom approach of MCMC method helps to sample from such a distribution.

By performing the modelling of the system the mean number of customers $L^{(i)}$ and the mean waiting time $W^{(i)}$, $i \in \{1;2\}$, are obtained. Fig. 2 illustrates a sample run of a two channel queuing system. The results are compared to the theoretical

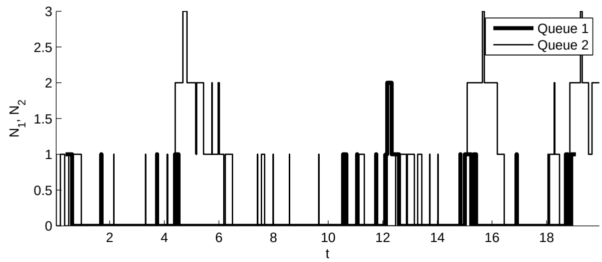


Fig. 2. The run of the $M/G/2/\infty$ queuing system.

system characteristics according to the formulas (1) and (2) [2].

$$W^{(1)} = \frac{(\lambda_1 + \lambda_2)E^2(x)(1 + v_x^2)}{2(1 - \lambda_1 E(x))}, \qquad L^{(1)} = \lambda_1 \cdot W^{(1)}, \qquad v_x = \frac{\sigma(x)}{E(x)}, \tag{1}$$

$$W^{(2)} = \frac{(\lambda_1 + \lambda_2)E^2(x)(1 + v_x^2)}{2(1 - \lambda_1 E(x))(1 - (\lambda_1 + \lambda_2)E(x))}, \qquad L^{(2)} = \lambda_2 \cdot W^{(2)}. \tag{2}$$

1 Markov Chain Monte Carlo (MCMC)

Suppose a researcher must generate random service times x_i, $i = \overline{1, n}$ which are distributed according to the $\pi(\cdot)$. The idea of MCMC is to construct a Markov chain $\{x_i\}_{i=0}^{\infty}$ such that $\lim_{i\to\infty} P(x_i = x) = \pi(x)$.

$$P(X_0 = x), \qquad P(y|x) = P(X_{i+1} = y|X_i = x). \tag{3}$$

Every Markov chain can be determined through an initial state and a transition kernel (3). It is known that the stationary distribution is unique if Markov chain is ergodic:

$$\pi(y) = \sum_{x\in\Omega} \pi(x)P(y|x), \quad \forall y \in \Omega. \tag{4}$$

Having ergodic and discrete Markov chain with the discrete stationary probabilities $\pi(x_i)$, the equation (4) holds. Total number of $(n-1)$ equations and $n(n-1)$ unknown transition kernel probabilities are apparent.

Thus there exist an infinite number of transition kernels representing the stationary distribution $\pi(x)$. Any of those transition kernels can be constructed and used for generating x_i. One of the most widely used methods for constructing such Markov chain is Metropolis–Hastings algorithm [1]. It is implemented as follows. At first an optional transition kernel $Q(.y|x)$ is chosen. Then there exists a probability α for chosen kernel Q being equal to transition kernel P:

$$P(y|x) = Q(y|x)\alpha(y|x), \quad y \neq x. \tag{5}$$

By substituting (5) into (4), we have:

$$\pi(x)Q(y|x)\alpha(y|x) = \pi(y)Q(x|y)\alpha(x|y), \quad \forall x \neq y. \tag{6}$$

By solving (6) and taking in mind the higher acceptance ratio when sampling random numbers [3], it is shown that:

$$\alpha(y|x) = \min\left(1, \frac{\pi(y)Q(x|y)}{\pi(x)Q(y|x)}\right). \tag{7}$$

Sampling of each x_i is performed in 4 steps. Firstly a candidate point x_i is drawn from the proposal distribution. Then the probability α_i, indicating that this point is also distributed by the target density, is calculated. The next step is to draw $u_i \sim U(0;1)$ and compare it to α_i. Finally, x_i is accepted to the sample if $u_i < \alpha_i$. Otherwise $x_i = x_{i-1}$.

From (7) it is evident that $\pi(x)$ can be determined up to a multiplicative constant c, i.e. $\pi(x) = c \cdot h(x)$, where $h(x)$ is a probability density function. An MCMC independence sampler is implemented if $Q(x|y) \equiv Q(x)$.

2 Nonparametric Probability Density Function (PDF) estimation

Consider a sample consisting of random independent and identically distributed values x_i. The kernel density estimate was chosen in order to evaluate the probability density of x_i.

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - x_i}{h}\right), \tag{8}$$

where $k(\cdot)$ is the kernel function, h is its width [4]. The triangular kernel function is useful if the data has sharp edged distribution. Gaussian kernel (9) makes the plot of the estimate's PDF very smooth.

$$k(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{s}}. \tag{9}$$

Such probability density estimation could be interpreted as the assignment of kernel density function to each x_i plus the weighted sum of all the other assignations. The contribution of any other x_j to the probability value at x_i decreases if $x_i - x_j$ increases. The only drawback of such estimation is the necessity of using all the points from the sample while evaluating the probability at a particular point.

Further research requires the cumulative density function of KDE. By integrating the (8) we get (10) where $K(\cdot)$ is the cumulative of $k(\cdot)$.

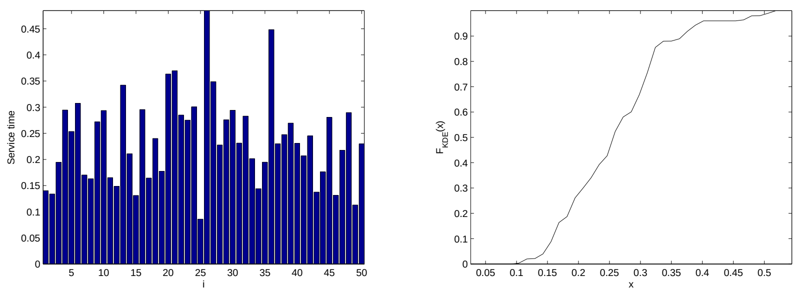
$$\hat{F}(x) = \frac{1}{nh}\int_0^x \sum_{i=1}^{n} k\left(\frac{t - x_i}{h}\right) dt = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{t - x_i}{h}\right). \tag{10}$$

3 Custom scheme for modelling service times

The first step of modelling a $M/G/2/\infty$ queue is to observe the real system and pick the empirical service times x_i (Fig. 3a)). Here both of the arrival processes are considered to have exponential distributions. The empirical data will be used to calculate both: KDE and cumulative KDE functions. So the second step is to evaluate (8) and (10) by finding optimal h according to the empirical data. (8) will serve as the target distribution in Metropolis–Hastings method.

The key feature of the scheme presented is the proposal density $q(\cdot)$ for KDE. There are a number of techniques to construct it, because the problem is to find probability distribution similar in shape to the target distribution. In this paper the proposal density is considered to be a piecewise linear distribution, which acts as a histogram of KDE.

The researcher can divide the possible service times into intervals and set probabilities for them according to the areas below the KDE in those intervals. This is a fast KDE approximation, but we propose a different approach instead. Calculating the area below KDE in any interval involves approximating it by an angled line. We bypass this by calculating the cumulative KDE values at the endpoints of the intervals (Fig. 3b)) and thus knowing the values for proposal density. By doing this, the approximation of cumulative KDE is obtained and is used for sampling from the proposal distribution.



(a) the observed service times

(b) the cumulative KDE function

Fig. 3. Evaluation of the service time distribution function.

Next follows the sampling process itself. It is performed by simply drawing a number u_i and reflecting it to the domain using obtained approximation of the cumulative KDE (which is also an angled line). The resulting value x_i is accepted to the sample according to the (7).

The intervals between arrival times for each queue are generated by inverse CDF of exponential distribution. Now the implementation of the queuing system model in a chosen programming language remains.

4 The results of $M/G/2/\infty$ system simulation

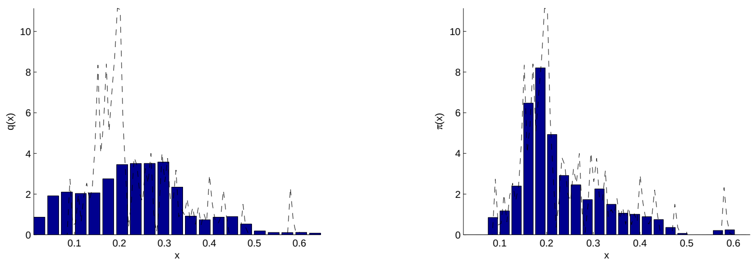
The queuing system with parameters $\lambda_1 = 0.7$, $\lambda_2 = 2.2$ and log-normally distributed service times has been modelled with the technique discussed above. The parameters of the service time distribution were $\mu_{logn} = -1.5$ and $\sigma_{logn} = 0.4$. 5 evenly spaced values of service time x were used for constructing the approximation of the cumulative service time distribution function. By using the inverse log-normal distribution function, 100 empirical service times were generated.

According to the table above, the relative error of the modelling is unacceptable. This is due to the constant kernel width and the fact that KDE was performed on the

Table 1. The results of the modelling.

Characteristics	$L^{(1)}$	$W^{(1)}$	$L^{(2)}$	$W^{(2)}$
Theoretical	0.052	0.075	0.550	0.250
Empirical	0.587	0.108	1.280	0.350
Relative error	0.911	0.443	1.326	0.398

support $(0; \infty)$. Using variable KDE width would decrease relative errors to a certain degree. On the other hand, randomly sampled service times do not follow log-normal distribution identically. In other words, our approach simulates the queuing system as is, i.e. simulation is completely based upon empirical data.



(a) using the proposal density (b) applying MCMC technique

Fig. 4. The histograms of the generated random service times.

Fig. 4 shows the histograms of random service times which were generated by the proposed approach. Here the process of generating x_i is divided into 2 parts. Firstly the proposal density is sampled. Then each x_i is accepted to the sample or not. It is noticeable that MCMC is similar to well known rejection sampling.

5 Conclusions

1. For better results of simulating a probability distribution with domain $(0; \infty)$ variable KDE width estimation should be used.
2. The more μ are closer to $\lambda_1 - \lambda_2$ the higher the relative errors of the system characteristics will be.
3. The proposed approach is universal and could be used for modelling any real stochastic system having only empirical data.
4. The sample $x_i \sim \pi(x)$ must be scrambled for better quality of x_i.
5. The experiments performed showed best results if the number of intervals used for constructing piecewise-uniform probability density is greater than 15.

References

[1] J. S. Dagpunar. *Simulation and Monte Carlo.* John Willey & Sons Ltd., Great Britain, Chippenham, Willtshire, 2007.

[2] G. Mickevičius and E. Valakevičius. Modelling of non-Markovian queuing systems. *Tech. Econ. Develop. Econ.*, **12**(4):295–300, 2006.

[3] V. Prokaj. Proposal selection for MCMC simulation. *Appl. Stoch. Mod. Data Anal.*, pp. 61–65, 2009.

[4] B. W. Silverman. Kernel density estimation using the fast Fourier transform. *Roy. Stat. Soc. Ser. C* (*Appl. Stat.*), **31**(1):93–99, 1982.

REZIUMĖ

MCMC metodu paremtas aptarnavimo sistemos modeliavimas iš empirinių duomenų

M. Landauskas, E. Valakevičius

Markovo grandinių Monte Karlo (MCMC) metodas reikalauja sukonstruoti generuojamo atsitiktinio dydžio tankio funkcijos aproksimaciją. Universalaus būdo, kaip tai padaryti, nėra. Straipsnyje pateikta universali tankio funkcijos aproksimacijos MCMC metodui schema, paremta branduoliniu tankio įverčiu. Dvikanalė aptarnavimo sistema su prioritetu buvo sumodeliuota šia technika. Empiriniai duomenys (stebėti aptarnavimo laikai) ir ateinančių paraiškų srauto intensyvumai yra visi sistemos modeliavimui reikalingi duomenys.

Raktiniai žodžiai: Markovo grandinių Monte Karlo metodas, branduolinis tankio įvertis, aptarnavimo sistema.