INFORMATION SYSTEMS IN MANAGEMENT

Information Systems in Management (2017) Vol. 6 (4) 259−269
DOI: 10.22630/ISIM.2017.6.4.1

USING DISCRETE MARKOV CHAINS IN PREDICTION OF HEALTH ECONOMICS BEHAVIOUR

WALDEMAR BAUER [a)], ANNA WIECZÓR [b)]

[a)] *AGH University of Science and Technology, Department of Automatics and Biomedical Engineering, bauer@agh.edu.pl*
[b)] *Cracow University of Economics, Department of Microeconomics, wieczora@uek.krakow.pl*

The aim of this article is show the concept of using of the Discrete Markov Chains to predict economic phenomena. This subject is important for two reasons. The first of them are models based on Markov chains use the statistical informations obtained during the investigation processes. Another important reason is the fact that this way of modeling is highly flexible and can be used to simulation of economic phenomenas. In this paper authors describe the idea of modeling and present the example of simply model of patient population of primary health care and show preliminary simulation results.

Keywords: Economic Behaviour, Primary Health Care, Stochastic process modeling, Markov Chain Monte Carlo, MCMC, PHC

1. Introduction

This article focuses on the subject of using discrete Markov Chains in prediction a lot of processes but especially in economics in the management of health care. Markov Chains was discovered in 1906 by Andriej Markov, the Russian mathematician. Then this theory was developed by Andriej Kolmogorov. In the beginning of the article, authors present the method of finance the Primary Health Care System. Described inter alia: fee for service, capitation, fixed fee (salary) and fee for the case. There are systems which are used currently in the world. The next part of the article presents methodology – the stochastic process modeling (stochastic process, the Markov Chain Process, Markov Chain Monte Carlo Simulation and Metropolis-Hasting algorithm). After the theoretical parts, authors to conducted a simulation experiment about the behavior of patients of Primary Health Care System in Poland and final conclusion. Because the discrete Markov Chain can be used in many areas the article shows also other the field of economic where this simulation experiment have application.

Until now the Discrete Markov Chains were used in the financial risk management, in particular in banks and insurance companies. In the banking system were used in estimating the credit risk and in the study of changes in the behavior of users of credit cards. In the insurance systems inter alia in the system of Bonus-Malus. The Bonus-Malus system is a rating system which is used in the motor insurance, to the price adjustment communications products. In this system, the transition between classes depends on the number of the injuries suffered by the insured during the period of contributory (see [10]) for example in Poland it is a one year.

Other areas where the Discrete Markov Chains were used to predict in economics was the capital market. Investigated the trading on the Warsaw Stock Exchange and other processes occurring in the capital market and made predictions about probable, further investment of stock market investors (see [12]). In addition, portfolio analysis can be performed by using the Discrete Markov Chains. It was also checked and described in literature (see [13]).By using the Discrete Markov Chains were also carried out an analysis of regional convergence. Analyzed may be whether and how quickly the regions with lower than average income can move "upward" and to the long-term, invariable distribution (see [8]). Because other classical methods of analysis of convergence were often criticized, the analysis of the convergence of both internal and external using the Discrete Markov Chains is now becoming increasingly popular (see [15]).

In prediction in health economics by Discrete Markov Chains interesting is the model which allows to model the transition of a population of patients through a series of health states that are followed over time. It can be include for example: living with a particular disease; having a treatment; being cured; having complications; or becoming deceased. We have a probability that a patient staying in their existing state or moving to a different one. By this simulation it is possible to predict worsening or improvementing of patients health (see [11]).

The possibility of using the Discrete Markov Chains in the economics is virtually unlimited. The every process, which meets the basic assumptions discussed in the theoretical part of this article, it can be predicted by using this method.

2. Primary Health Care System

Primary Health Care (PHC) is a multidimensional and country-to-country diverse part of the health care system, thus it is a real challenge for modeling these systems. The international research team used existing data sets, reports, publications as well as gray data and opinions of key informants to provide innovative international comparisons of PHC systems in Europe. In Poland, as in other Central and Eastern European countries, the methods for reliable monitoring of care provided to patients has not been introduced in primary health care. Moreover, the data collected by the National Health Found are not be published. The above situation makes difficult take steps in family physician practices to improve the quality of care (see [1, 2]).

The world currently uses the following methods of financing PHC (see [1]):**Fee for service** (is the pay gap between a doctor service providers, or any other professional employee), **Capitation (**a system in which a physician or other medical worker or trader receives a fixed amount for each person covered by the care), **Fixed fee** (salary,) **Fee for the case (**the way in which the payment service provider receives a fixed salary for a comprehensive investigation in a particular case, or disease entity).

All the described methods of finance primary health care based on the number of patients served at the facility. It follows that an efficient way simulation of PHC will be a simulation number of patients in a time range.

3. Stochastic process modeling

Many stochastic processes used for the modeling of financial marks, biological systems, social systems and other systems in engineering are Markovian. To simulate the process can be used Markov Chain Monte Carlo (MCMC). In statistics, MCMC method is a class of algorithms for sampling from a probability distribution. Probability distributions can be found the context of Bayesian data analysis. The goal will be to find parameter values in a probabilistic model that the best explains the data. It is based on known information (a priori). The created mathematical models are then created some posterior. Such approach guarantees that solution is influenced by the known data, therefore the models are often more accurate than obtained with other methods (see [2, 8]). In this section, the author will describe basic definitions from stochastic theory, the Markov Chain process, Markov Chain Monte Carlo Simulation and the Metropolis-Hasting algorithm.

3.1. Stochastic process

A stochastic process X^n is a family of random variables indexed by parameter n (this parameter can by associated with time). Formally, a stochastic process for probability space (Ω, F, P) and measurable space $(S, \sum)$. The sample space Ω is a set of outcomes, where an outcome is the result of a single execution of a stochastic model. F is set of all events in the model and P is the probability measure. The probability measure is the function returning an event's probability $(P : F \to [0,1])$ this can by. The S-valued stochastic process is a collection of S-valued random variables on **Ω**. The stochastic process is indexed by a totally ordered set T. That is, a stochastic process X is the collection $\{X^n : n \in T\}$.

3.2. The Markov Chain process

Markov Chain is a stochastic process where we transition from one state to another using a sequential procedure. We start Markov Chain in the state x^0, and use a transition function $p(x^n \mid x^{n-1})$, to determine the next state, conditional on the last state. We can say a stochastic process $\{x^n : n \geq 0\}$ is a Markov chain if for all times $n \geq 0$ and all state $i_0, i_1, ..., i_n, j \in S$:

$$P(x^n = j \mid x^{n-1} = i_{n-1}, x^{n-2} = i_{n-2}, ..., x^0 = i_0) = P(x^n = j \mid x^{n-1} = i_{n-1}) = P_{ij}$$

P_{ij} denotes the probability that the chain moves from state x^{n-1} to state x^n. This value is referred to as a one-step transition probability. The square matrix $\boldsymbol{P} = P(i, j) \in S$ is called the one-step transition matrix and each row sum to one (see [6, 11]).

3.3. Markov Chain Monte Carlo Simulation

Markov chains are relatively easy to simulate from, they can be used to sample from an a priori unknown and probability distribution. Monte Carlo sampling allows one to estimate various characteristics of a distribution such as the mean, variance, kurtosis, or any other statistic of interest to a researcher. Markov chains involve a stochastic sequential process where we can sample states from some stationary distribution.

The Markov Chain Monte Carlo (MCMC) method is a general simulation method for sampling from posterior distributions and computing posterior quantities of interest. MCMC methods sample successively from a target distribution. Each sample depends on the previous one, hence the notion of the Markov chain.

Monte Carlo, as in Monte Carlo integration, is mainly used to approximate an expectation by using the Markov chain samples. In the simplest version

$$\int_S g(\theta)p(\theta)d\theta \cong \frac{1}{n}\sum_{t=1}^{n} g\left(\theta^t\right).$$

Where g is a function of interest and $\theta^t$are samples from $p(\theta)$ on its support S. This approximates the expected value of $g(\theta)$ .With the MCMC method, it is possible to generate samples from an arbitrary posterior density $p(\theta \mid y)$ and to use these samples to approximate expectations of quantities of interest (see [6],[10]).

3.4. Metropolis-Hasting algorithm

To illustrate the work of all MCMC methods the Metropolis-Hastings method has been described.

Suppose our goal is to sample from the target density $p(\theta)$. The Metropolis-Hastings method creates a Markov chain that produces sequences of state:

$$\theta^0 \to \theta^1 \to \ldots \to \theta^n$$

where $\theta^{(t)}$ is a state at iteration. The samples from the chain, after burning, reflect samples from the target distribution $p(\theta)$. In this algorithm, we initialize the first state from a random value. We then use a proposal distribution $p(\theta^n \mid \theta^{n-1})$ to generate a new candidate state θ^*, that is conditional on the previous state. The proposal distribution is chosen by the research and good choices for the distribution depend on the problem. To the choose proposed distribution, we can use e.g.: maximum entropy, nuclear estimators, and transformation groups.

The next step is to either accepted or reject proposal state. The probability of accepting the state θ^* is:

$$\alpha = \min\left(1, \frac{p(\theta^*)q(\theta^{i-1} \mid \theta^*)}{p(\theta^{i-1})q(\theta^* \mid \theta^{n-1})}\right)$$

To decide on whether to accept or reject the proposed state, we generate a uniform deviate u. If $u \le \alpha$ the proposal is accepted and the next state value is equal θ^*, else we reject the proposal and next state value is equal to the old state value. We continue generating new proposals conditional on the current state of the method, and either accept or reject the proposals. This procedure continues until the sample reaches convergence. At this point, samples $\theta^{(i)}$ the samples from the target distribution $p(\theta)$.

This can be converted into an algorithm as follows:

1. Generate initial value of u, and set $\theta^0 = u$ and $i = 0$;
2. Set max iteration number N;
3. Repeat:
    a. $i = i + 1$,
    b. Generate proposal θ^* from $p(\theta^n \mid \theta^{n-1})$,
    c. Calculate the accepted probability:
    $$\alpha = \min\left(1, \frac{p(\theta^*)q(\theta^{i-1} \mid \theta^*)}{p(\theta^{i-1})q(\theta^* \mid \theta^{n-1})}\right),$$
    d. Generate u from a uniform (0; 1) distribution,
    e. If $u \leq \alpha$, accept new state and $\theta^{i-1} = \theta^*$, else set $\theta^i = \theta^{i-1}$;
4. Until $i = N$.

The fact that asymmetric proposal distributions can be used allows the Metropolis-Hastings procedure to sample from target distributions that are defined on a limited range (see [9], [6]). With bounded variables, care should be taken in constructing a suitable proposal distribution. Therefore, the sample will move towards the regions of the state space where the target function has high density. However, note that if the new proposal is less likely than the current state, it is still possible to accept this "worse" proposal and move toward it. This process of always accepting a "good" proposal, and occasionally accepting a "bad" proposal ensures that the sampler explores the whole state space and samples from all parts of a distribution (including the tails). In the numerical experiments, all algorithms have been prepared by the authors.

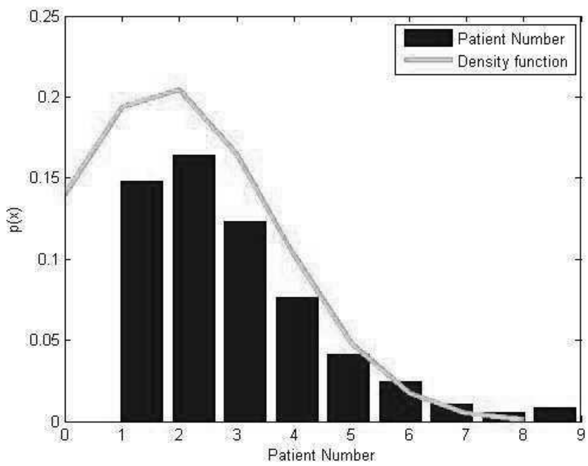

Figure 1. Real data histogram with Normal Density Function

4. Simulation Results

In this section, will be described an example of simulation. For the purpose of the classifiers verification – the data gathered during the research project "Optimisation of the Polish financing system for the Primary Care Units" (in Polish: 'Optymalizacja polskiego systemu finansowania podstawowej opieki zdrowotnej' – see [1]). The data was collected in years 2009-2011. This simulations are based on data from 2010 for the five years old patients from the PHC units.

In 2010, in this age group was 621 visits. The average value of the day for normalized value of data is advice about 1.7, while the standard deviation is about 0.0027. The maximum number of visits in a single day to 9 and the minimum 1. We used this information to estimated number of the visit in PHC units in next year.

How we can see in figure 1 density function describe this data have Normal form. In simulation we show result for native density function (Poisson density) and Normal density function.

4.1. Poisson density function example

For native case of density function we can define acceptation function in Metropolias-Hasting algorithm as:

$$\alpha = \min\left(1, \frac{\frac{\mu^{\theta^8} e^{-\mu}}{\theta^*!}}{\frac{\mu^{\theta^{(i-1)}} e^{-\mu}}{\theta^{(i-1)}!}}\right)$$

where μ is the expected value.

Simulations result for this acceptation function have been showed in figure 2. Figure 3 present density function for simulation result. Total number of visit in this simulation was 11296. Value of mean visit per day was 1.7014 and standard deviation have value 0.0025. As it is easy to see, all the parameters describing density function in five years old group have been preserved by simulation.

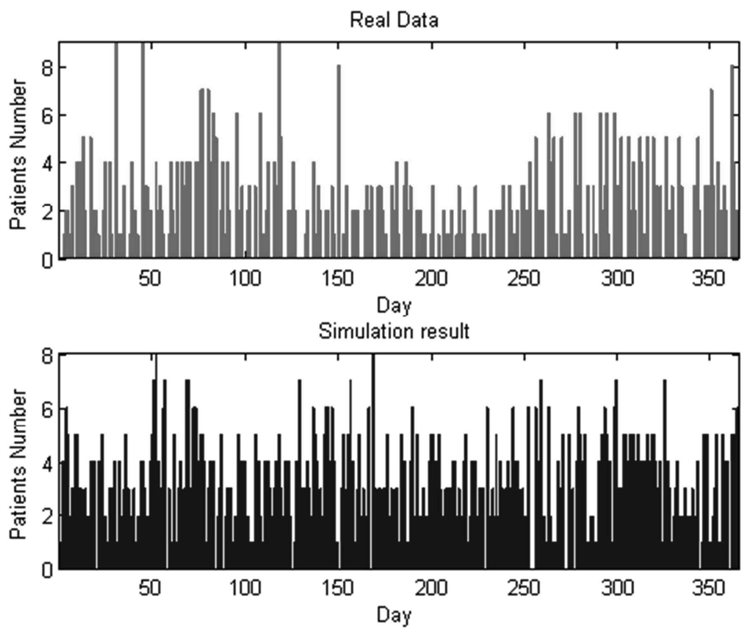

Figure 2. Comparison real data with simulation result for normal density function

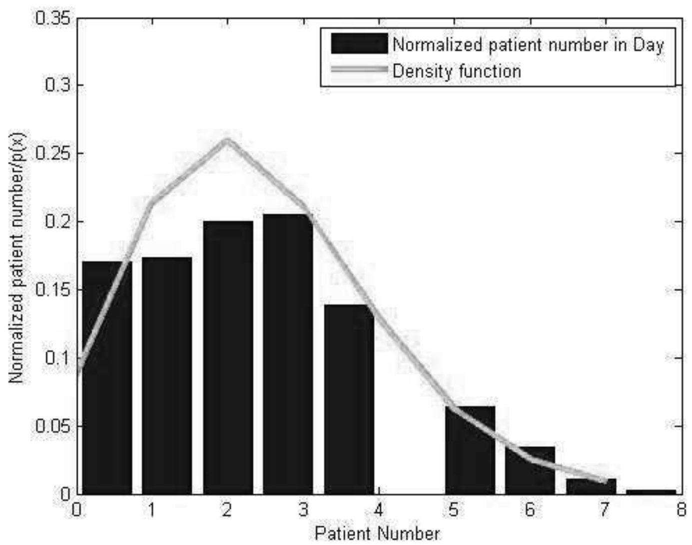

Figure 3. Simulated data histogram with Poisson Density Function

4.2. Normal density function example

In this case, the function of acceptance State the Metropolias-Hastings will take the form of:

$$\alpha = \min\left(1, \frac{p\left(\theta^{*}, \mu, \sigma\right)}{p\left(\theta^{(i-1)}, \mu, \sigma\right)}\right)$$

and

$$p(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Where μ is the expected value and standard deviation σ.

Simulations result have been showed in Figure 4. In the simulation have been 888 visits or expected value was approximately 1.8 and the standard deviation of about 0.0026. Figure 5 present density function for data form simulation.

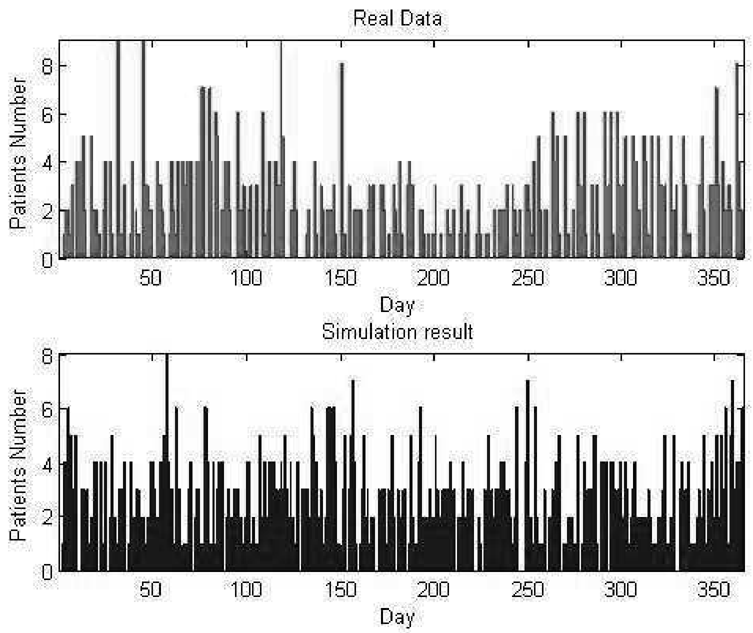

Figure 4. Comparison real data with simulation result for normal density function

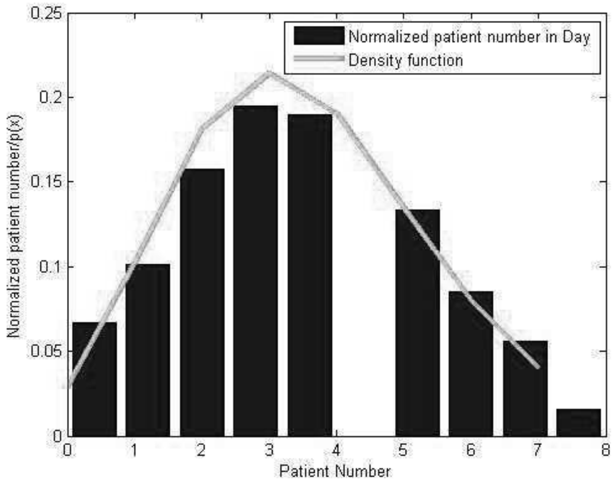

Figure 5. Simulated data histogram with normal density function

5. Conclusion

This paper present simple example using Markov Chain Monte Carlo method in simulation health economics of polish primary health care. As it is easy to see the results of the simulation for Poisson and Normal distribution (Figure 2 and Figure 4) are similar. In this case, it is shown that this simple model is good to simulation of behavior patient in different year groups. The study has proven the need for further research of complexity model of Primary Health Care unit. and will be the subject of further work.

Bibliography

[1] Baranowski J., Windak A. (2012), *Optymalizacja polskiego systemu finansowania podstawowej opieki zdrowotnej*, Technical Report, Ernst & Young Usługi Finansowe Audyt sp. z o.o. Polska sp.k., Warszawa.

[2] Bergstra J., Yamins D., Cox D.D. (2013), *Hyperopt: A Python Library for optimizing the hyperparameters of machine learning algorithms*, Python in Science Conference, , pp. 13–20.

[3] Garnett R., Osborne M.A., Roberts S.J. (2010), *Bayesian optimization for sensor set selection*, International Conference on Information Processing in Sensor Networks – IPSN, pp. 209–219.

[4] Grzybowska U., Karwański M. (2014), *Przykłady zastosowania macierzy migracji w zarządzaniu ryzykiem finansowym,* Zeszyty Naukowe Wydziałowe Uniwersytetu Ekonomicznego w Katowicach. Studia Ekonomiczne, Katowice, pp. 206-219.

[5] http://www.match.ac.uk/markovtool.html

[6] Jaynes E.T. (2003), *Probability Theory: The Logic of Science*, Cambridge University Press.

[7] Kringos D.S., Boerma W.G.W., Bourgueil Y., Cartier T., Hasvold T., Hutchinson A., Lember M., Oleszczyk M., Pavlic D.R., Svab I., Tedeschi P., Wilson A., Windak A., Dedeu T., Wilm S. (2010), *The European primary care monitor: structure, process and outcome indicators*, BMC Family Practice, 11, pp. 81.

[8] Łaźniewska E., Górecki T. (2012), *Analiza konwergencji podregionów za pomocą łańcuchów Markowa*, Wiadomości statystyczne, (5), Główny Urząd Statystyczny, Warszawa, pp. 1-9.

[9] Schäfer W.L., Boerma W.G.W., Kringos D.S., De Ryck E., Greß S., Heinemann S., Murante A.M., Rotar-Pavlic D., Schellevis F.G., Seghieri C., Van den Berg M.J., Westert G.P., Willems S., Groenewegen P.P. (2013), *Measures of quality, costs and equity in primary health care instruments developed to analyse and compare primary care in 35 countries*, Quality in Primary Care, 21, pp. 67–79.

[10] Sivia D., Skilling J. (2006), *Data Analysis: A Bayesian Tutorial*, Oxford University Press, Oxford.

[11] Srinivas N., Krause A., Kakade S.M., Seeger M.W. (2012), *Information-theoretic regret bounds for Gaussian process optimization in the bandit setting*, IEEE Transactions on Information Theory, 58(5), 3250–3265.

[12] Stawicki J. (2004), *Wykorzystanie łańcuchów Markowa w analizie rynku kapitałowego*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.

[13] Stawicki J. (2008), *Wykorzystanie łańcuchów decyzyjnych Markowa do analizy portfelowej,* Prace Naukowe, Akademia Ekonomiczna w Katowicach, Katowice, pp. 137-146.

[14] Thornton C., Hutter F., Hoos H.H., Leyton-Brown K. (2013), *Auto-WEKA: Combined Selection and Hyperparameter Optimization of Classification Algorithms*, in: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, New York, NY, USA, pp. 847–855.

[15] Wałęga A. (2014), *Spójność ekonomiczna regionów Polski przed i po przystąpieniu do Unii Europejskiej*, (203), Studia Ekonomiczne. Uniwersytet Ekonomiczny w Katowicach, Katowice, pp. 172-180.