PHYSCON 2011, León, Spain, September, 5–September, 8 2011

STATE ESTIMATION PROBLEM FOR MARKOV CHAIN MODEL

Galina Timofeeva
Ural Federal University
Ural State University of Railway Transport
Russia
Gtimofeeva@math.usurt.ru

Nikolay Timofeev
Department of Mathematics
Ural State University of Railway Transport
Russia
NTimofeev@inbox.ru

Abstract
The state estimation problem for a dynamic system described by the Markov chain model with discrete time is considered. It is assumed that a matrix of the transition probabilities is incompletely known. Proposed approach is based on the confidence estimates.

Key words
State estimation, incomplete information, Markov chain.

1 Introduction

Markov chain models [Markov, 1906] are widely used to explain the dynamics of state changes for different systems. Often they are used as a mathematical model for some random physical process. If parameters of the chain are known, quantitative predictions can be made. Markovian systems appear extensively in thermodynamics and statistical mechanics, whenever probabilities are used to represent unknown or unmodelled details of the system, if it can be assumed that the dynamics are time–invariant, and that no relevant history need be considered which is not already included in the state description.

Chemistry is often a place where Markov chains and continuous–time Markov processes are especially useful because these simple physical systems tend to satisfy the Markov property quite well. The classical model of enzyme activity, Michaelis–Menten kinetics, can be viewed as a Markov chain, where at each time step the reaction proceeds in some direction. An algorithm based on a Markov chain was also used to focus the fragment–based growth of chemicals towards a desired class of compounds such as drugs or natural products.

Markov chains are used in finance and economics to model a variety of different phenomena, including asset prices, market crashes and credit portfolio dynamics [Jones, 2005; Thyagarajan, Saiful, 2005].

If the transition probability matrix of Markov chain is known then dynamics of the system states probabilities is completely described by a system of difference equations. But in most cases the transition probabilities are unknown and estimated during the system evolution.

2 Problem Statement

Let us consider a system with k states, the probability that the system is in i-th state at moment t denote by $x_i(t)$, $i = 1, \ldots, k$. Thus the following conditions hold:

$$0 \le x_i(t) \le 1, \quad x_1(t) + \ldots + x_k(t) = 1. \tag{1}$$

Let the dynamics of the system states probabilities is described by the discrete Markov chain model:

$$x_j(t+1) = \sum_{i=1}^{k} p_{ij} x_i(t), \;\; t = 0, 1, \ldots, T, \tag{2}$$

where p_{ij} is the probability of going from state i to state j on one step.

We impose a simple Markov structure on the transition probabilities and restrict our attention to first–order stationary Markov processes for simplicity.

Let denote by $x(t)$ a vector of states probabilities $x(t) = \{x_1(t), \ldots, x_k(t)\}^\top$, by P a matrix of the transition probabilities $P = \{p_{ij}\}$ and rewrite equation (2) in the vector form:

$$x(t+1) = P^\top x(t), \quad t = 0, 1, \ldots, T. \tag{3}$$

The problem is to estimate $x(T)$ if the transition probability matrix P is incompletely known.

It is assumed that we have information about the number of transitions from i-th state to j-th on t step, $t = 1, \ldots, m$. Different approaches to the estimation of the transition probabilities matrix are investigated.

3 Estimation of Transition Probabilities Matrix

For the estimation of the probability p_{ij} one usually use the statistical data about transitions from one state to another. Let n_{ij} denote the number of individuals who were in state i in period $t-1$ and are in state j in period t. We can estimate the probability p_{ij} of an individual being in state j in period t given that they were in state i in period $t-1$.

The probability of transition from any given state i is approximated by a proportion of individuals that started in state i and ended in state j as a proportion of all individuals in that started in state i:

$$w_{ij} = \frac{n_{ij}}{\sum\limits_{j} n_{ij}}. \tag{4}$$

Using the methods described above, it is possible to estimate a transition matrix using count data.

Anderson and Goodman [Anderson, Goodman, 1957] showed that the estimator w_{ij} given by equation (4) is a maximum-likelihood estimator that is consistent but biased, with the bias tending toward zero as the sample size increases.

Suppose that instead of observing the actual count of transitions from the different states, we only observe the aggregate proportions $y_i(t)$, which represent the proportion of observations with the state i. The aggregate proportions $y_i(t)$ estimate the system state probabilities:

$$y_i(t) \approx N x_i(t), \quad i = 1, \dots, N, \quad t = 1, \dots, T,$$

where N is a number of all individuals.

If the time series of observations T is sufficiently long, it is possible to estimate a transition matrix P from aggregate data using quadratic programming methods.

Let consider the relation:

$$y_j(t) = \sum_{i=1}^{k} y_i(t-1) p_{ij} + u_j(t-1), \tag{5}$$

where $y_j(t)$ is observed date, p_{ij} are unknown elements of the transition probability matrix, $u_j(t)$ are deviations which should to be minimized.

The equation (5) may be rewritten in the vector form:

$$Y = GP + U,$$

here

$$U = \{u_1(1), \dots, u_1(T), u_2(1), \dots, u_k(T)\}^\top \in \mathbb{R}^{kT},$$
$$Y = \{y_1(1), \dots, y_1(T), y_2(1), \dots, y_k(T)\}^\top \in \mathbb{R}^{kT},$$

G is $kT \times kT$ matrix,

Lee, Judge and Zellner [Lee, Judge, Zellner, 1970] suggest minimizing the sum of squared errors in equation (5) subject to linear constraints on the transition probabilities p_{ij}:

$$(Y - GP)^\top (Y - GP) \to min,$$

$$p_{ij} \geq 0, \quad \sum_j p_{ij} = 1. \tag{6}$$

This approach was continued by Kalbfleisch, Lawless and Vollmer [Kalbfleisch, Lawless, 1984; Kalbfleisch, Lawless, Vollmer, 1984].

MacRae [MacRae, 1977] noted that the variance of the error term U depends on the magnitude of y_{t-1}, so using ordinary least squares estimation techniques is not efficient estimates. She produced a more efficient estimator using an iterative generalized least squares method for calculating the matrix of transition probabilities P. The first step in the procedure is to estimate the transition matrix, and then use this to calculate a consistent estimate of the conditional covariance matrix. The estimated covariance matrix is used to obtain a subsequent estimate of the transition probabilities, and the procedure is repeated. The convergence of the procedure is investigated in the paper [MacRae, 1977].

Jones [Jones, 2005] studied the maximum-likelihood estimates for the transition probabilities matrix for use in credit risk modeling with a decades-old methodology that uses aggregate proportions data.

4 Confidence Approach

Let us denote by $\xi(t)$ a random vector described system state. It equals to i-th basic vectors if the system is in the i-th state. Thus

$$x_i(t) = Pr\{\xi(t) = e_i\}, \quad i = 1, \dots, k, \tag{7}$$

here $e_1, \dots, e_k$ are the base vectors in $\mathbb{R}^k$.

Denote a random loss function on m-th step by

$$\eta(m, T) = \sum_{t=m}^{T} [l(t)^\top \xi(t)],$$

where $l(t)$ is vector function $[0, \infty) \mapsto \mathbb{R}^k$.

In the loan portfolio model the vector function equals a loss (profitability) of the portfolio on the time interval $[0; T]$. In this case $l(t)$ has the form $l(t) = r(t)l$, where $l = \{l_1, \dots, l_k\}^\top$, l_i is a loss of loans in i-th group, $r(t)$ is a discount factor.

The purpose is to estimate a probability that the loss function exceed a given level γ

$$d_\gamma = Pr\{\eta(m, T) > \gamma\}. \tag{8}$$

If the transitions probability matrix P is known then using equation (3) we get the distribution of the random value $\eta(m,T)$ which is completely defined by vector of the states probability $x(m)$, thus $d_\gamma = d(x(m), P, \gamma)$.

Another approach is to estimate a quantile for a given probability β:

$$q_\beta = \max\{q : Pr\{\eta(T) > q\} \le \beta\}. \tag{9}$$

The quantile and probability functions are closely connected and their properties are studied by Precopa [Precopa, 1995], Kibzun and Kan [Kibzun, Kan, 1996].

In many cases a consideration of the estimates of the random value reduces to the estimation of the fist and second statistical moments of a loss function. In our problem $E(\eta(m,T))$ and $\sigma^2(\eta(m,T))$ are linear and quadric functions functions of $x(m)$ respectively in case of a known matrix P of the transition probabilities.

But in the consideration the transition probabilities matrix P is incompletely known and should to be estimated. The state estimation problem for multistage systems with gaussian perturbation and uncertain matrix was considered by Anan'ev [Anan'ev, 2010].

One of the general approach to quantile and probability estimation problems is to reduce the quantile optimization [Kibzun, Kan, 1996] to the optimal choice of the confidence regions for unknown parameters. Let us use this general approach to the estimation problem (8).

For simplicity of the consideration it is proposed further that the loss function has a form

$$\eta(T) = l^\top \xi(T) = l_1\xi_1(T) + \ldots + l_k\xi_k(T).$$

5 Method of Probability Estimation

The first step is to estimate the elements of the transition probability matrix P. Denote the confidence region for $\{p_{ij},\ i = 1,\ldots,k,\ j \ne i,\ j = 1,\ldots,k\}$ on m-th step by $Z_\alpha \subset \mathbb{R}^K,\ K = k(k-1)$.

Thus p_{ij} are the transition probabilities then conditions (6) hold and $Z_\alpha \subset Z_+ \subset \mathbb{R}^K$, where

$$Z_+ = \{p_{ij} :\ \ 0 \le p_{ij} \le 1,\ \ \sum_{j\ne i} p_{ij} \le 1\} \subset \mathbb{R}^K.$$

Estimation for p_{ii} follows from the equalities

$$p_{i1} + \ldots + p_{ik} = 1, \quad i = 1,\ldots,k. \tag{10}$$

In the considered model the differences $p_{ij} - w_{ij}$ is approximately normal distributed and we may use the confidence set Z_α defined by joint restrictions:

$$Z_\alpha = \{p_{ij} \in Z_+ : (p_{ij} - w_{ij})^\top G(p_{ij} - w_{ij}) \le b^{(\alpha)}\} \tag{11}$$

The next step is to solve the state estimation problem for a multistage deterministic system with uncertain matrix:

$$\begin{aligned} &x(t+1) = P^\top x(t), \quad t = m,\ldots,T,\\ &x(m) = x^*,\ \ P \subset Z. \end{aligned} \tag{12}$$

and to find an information set

$$X(t,Z) = \{x \in R^k_+ : x = (P^\top)^{T-m} x^*,\ \ P \in Z\}. \tag{13}$$

We may construct information sets for system (12) using approaches proposed by Kurzanski, Tanaka and Matasov [Kurzanski, Tanaka, 1989; Matasov, 1999].

The third step is to estimate the probability

$$d_\gamma = Pr\{\eta(T) = l^\top \xi(T) > \gamma\}$$

for a given level γ, where a distribution of the random vector $\xi(T)$ is defined by a vector of states probabilities $x(T) = \{x_1(t),\ldots,x_k(T)\}$, which is incompletely known. This vector is given by its confidence region $X(T, Z_\alpha)$,

$$P\{x(T) \in X(T, Z_\alpha)\} = \alpha, \tag{14}$$

where the set $X(T, Z_\alpha)$ is defined by (13).

The estimation of the probability d_γ is reduced to estimation of the probability

$$b_\gamma(z) = Pr\{l^\top \xi(x(T,z)) > \gamma\}, \tag{15}$$

where $x(T,z) \in X(T,Z)$.

Obtaining problem is a probability optimization problem with uncertainty which properties and methods of solving are studied in [Timofeeva, 2007; Timofeeva, 2010].

Let us denote lower and upper bounds of the probability $b_\gamma(z)$ by $b^{(1)}_\alpha$ and $b^{(2)}_\alpha$:

$$b^{(1)}_\alpha = \inf_{z\in Z_\alpha} b_\gamma(z), \quad b^{(2)}_\alpha = \sup_{z\in Z_\alpha} b_\gamma(z). \tag{16}$$

The complete probability formula for d_γ has the form:

$$\begin{aligned} &d_\gamma = Pr\{\eta(T,z) > \gamma \mid z \in Z_\alpha\} \cdot Pr\{z \in Z_\alpha\} +\\ &+ \Pr\{\eta(T,z) > \gamma \mid z \notin Z_\alpha\} \cdot Pr\{z \notin Z_\alpha\}, \end{aligned} \tag{17}$$

here $Pr\{A \mid B\}$ is a conditional probability of A in condition of B. Using equalities (14) and (16) we get from (17) the estimation for probability d_γ

$$\alpha b^{(1)}_\alpha \le d_\gamma \le \alpha b^{(2)}_\alpha + (1-\alpha). \tag{18}$$

Thus the method of solving the probability estimation problem consists of following stages:

1. take confidence probability α close to 1, (e.g. $\alpha = 0.99$, $\alpha = 0.995$) and calculate a confidence set Z_α for elements $\{p_{ij},\ \ j \neq i\}$ of the matrix P;
2. describe the information set $X(T, Z_\alpha)$ for the multistage system (12);
3. calculate lower and upper bounds (16) of the probability $b_\gamma(z)$, $z \in Z_\alpha$ and estimate the probability d_γ using inequalities (18).

6 Application to Credit Portfolio Dynamics

In particular the Markov process describes the credit portfolio dynamics [Jones, 2005; Thyagarajan, Saiful, 2005]. A change of shares of credits portfolio is described by Markov chain with discrete time.

In this case the credit state is determined on an accessory to this or that group of credits depending on presence of indebtedness and its terms. We use a model with discrete time and fix the system state through identical time intervals - once a month [Timofeev, 2011].

Usually the matrix of the transitive probabilities is incompletely known therefore its values are constantly specified according to the vintage analysis of a portfolio.

We considered a scheme with 6 groups of loans: a group of loans without delay, a group of loans with less than 31 days delay, a group of loans with 31 – 65 days delay, ect. and a group of problematic loans. Transition probabilities were estimated by using the confidence approach.

Confidence regions (11) for p_{ij} were constructed on the base of statistical data about transitions in relatively small portfolio of homogeneous loans ($N = 1000$). We estimated the probability $d_\gamma(T)$ that a share of problematic loans $y_6(T)$ exceed a critical level γ.

Estimations for $d_\gamma(T)$ were obtained using formula (18) for $T = 1, \ldots, 12$ months. Appropriate results were obtained only for $T = 1, \ldots, 4$. In the case of $T > 6$ months the confidence interval $[d_\gamma^-(T); d_\gamma^+(T)]$ was too large.

7 Conclusion

It is proposed in the paper that a system dynamics is described by the discrete Markov chain. The probability of an exceeding a given level of a linear loss function is estimated by using confidence sets for the transition probabilities.Obtained results apply to modeling of credit portfolio dynamics and to estimate the probability of future losses.

Acknowledgements

The investigations are supported by the Russian Foundation for Basic Researches, project 10-01-00672a.

References

Anderson, T. W. and Goodman, L. A. (1957) Statistical Inference About Markov Chains. *Annals of Mathematical Statistics*. **28**. pp. 89–110.

Anan'ev, B.I. (2010) Multiestimates for linear-gaussian continuous systems under communication constraints *From Physics to Control Through an Emergent View*. World Scientific Ser. on Nonlinear Science, Editor: L. O. Chua, Univ. of California, Berkeley, Ser. B, **15**. pp. 265–270.

Jones, M.T. (2005) Estimating Markov Transition Matrices Using Proportions Data: An Application to Credit Risk. *IMF Working Paper*. **WP-05-219**.

Kalbfleisch, J. D., Lawless J. F.(1984) Least-Squares Estimation of Transition Probabilities From Aggregate Data, *Canadian J. of Statistics*, **12**(3). pp. 169–182.

Kalbfleisch, J. D., Lawless J. F., Vollmer W.M. (1983) Estimation in Markov Models from Aggregate Data, *Biometrics*. **39**. pp. 907–919.

Kibzun A.I., Kan Yu.S. (1996) *Stochastic Programming Problem with Probability and Quantile Functions.* Chichester etc.: John Wiley & Sons.

Kurzanski A.B., Tanaka M. (1989) *On a unified framework for deterministic and stochastic treatment of identification problem.* Laxenburg, IIASA.

Lee T.C., Judge, G.G. and Zellner, A. (1970) *Estimating the Parameters of the Markov Probability Model From Aggregate Time Series Data*. Amsterdam: North Holland.

MacRae E. Ch. (1977) Estimation of Time-Varying Markov Processes with Aggregate Data, *Econometrica*, **45**(1). pp. 183–98.

Markov, A.A. (1906) Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga. *Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete*. Ser. 2, **15**, pp. 135-156. Reprinted in: Appendix B of: Howard, R. *Dynamic Probabilistic Systems*. **1**: Markov Chains. John Wiley and Sons, 1971.

Matasov A. (1999) *Estimators for uncertain dynamic systems*. Dordrecht etc.: Kluwer Acad. Publ.

Precopa, A. (1995) *Stochastic Programming.* Dordrecht etc.: Kluwer Acad. Publ.

Thyagarajan, V. and Saiful, M. (2005) Retail Banking Loan Portfolio Equilibrium Mix: A Markov Chain Model Analysis. *Amer. J. of Applied Sciences*. **2**(1). pp. 410–419.

Timofeev, N.A. (2011) Mathematical Model of the Vintage Analysis of Banking Credit Portfolio. *Gerald of USURT*, **9**(1). pp. 61–69.

Timofeeva, G.A. (2007) Optimal and Suboptimal Solutions to Stochastically Uncertain Problem of Quantile Optimization. *Automation and Remote Control.* **68**(3). pp. 1145–1157.

Timofeeva, G.A. (2010) Quntile optimization problem with incomplete information. *From Physics to Control Through an Emergent View*. World Scientific Ser. on Nonlinear Science, Ed.: L. O. Chua, Univ. of California, Berkeley, Ser. B, **15**, pp. 283–288.