

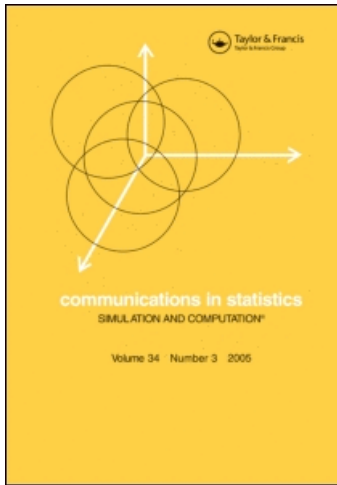
This article was downloaded by: [South Dakota State University]

On: 26 February 2009

Access details: Access Details: [subscription number 786945906]

Publisher Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Communications in Statistics - Simulation and Computation

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713597237>

A Novel Estimation Approach for Mixture Transition Distribution Model in High-Order Markov Chains

D. G. Chen ^{ab}; Y. L. Lio ^c

^a Department of Mathematics and Statistics, Agricultural Experiment Station, South Dakota State University, Brookings, South Dakota, USA ^b Department of Surgery, Sanford School of Medicine, University of South Dakota, Sioux Falls, South Dakota, USA ^c Department of Mathematical Sciences, University of South Dakota, Vermillion, South Dakota, USA

Online Publication Date: 01 May 2009

To cite this Article Chen, D. G. and Lio, Y. L.(2009)'A Novel Estimation Approach for Mixture Transition Distribution Model in High-Order Markov Chains', Communications in Statistics - Simulation and Computation, 38:5,990 — 1003

To link to this Article: DOI: 10.1080/03610910802715009

URL: <http://dx.doi.org/10.1080/03610910802715009>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

A Novel Estimation Approach for Mixture Transition Distribution Model in High-Order Markov Chains

D. G. CHEN¹ AND Y. L. LIO²

¹Department of Mathematics and Statistics, Agricultural Experiment Station, South Dakota State University, Brookings, South Dakota, USA and Department of Surgery, Sanford School of Medicine, University of South Dakota, Sioux Falls, South Dakota, USA

²Department of Mathematical Sciences, University of South Dakota, Vermillion, South Dakota, USA

A transformation is proposed to convert the nonlinear constraints of the parameters in the mixture transition distribution (MTD) model into box-constraints. The proposed transformation removes the difficulties associated with the maximum likelihood estimation (MLE) process in the MTD modeling so that the MLEs of the parameters can be easily obtained via a hybrid algorithm from the evolutionary algorithms and/or quasi-Newton algorithms for global optimization. Simulation studies are conducted to demonstrate MTD modeling by the proposed novel approach through a global search algorithm in R environment. Finally, the proposed approach is used for the MTD modelings of three real data sets.

Keywords Genetic algorithms; High-order temporal dependence; Markov chains; Maximum likelihood estimation; Mixture transition distribution.

Mathematics Subject Classification 65C40; 62M05.

1. Introduction

A Markov chain, which is a probabilistic model used to describe the temporal dependence among observed data, was proposed by Andrej Andreevic Markov at the beginning of the 20th century and subsequently developed for applications in many scientific fields beyond mathematics including biology, chemistry, physics, social sciences, and music.

In the modeling of high-order discrete-time sequence Markov chains, which need to estimate a large number of parameters for the transition probability, Raftery (1985) proposed a mixture transition distribution (MTD) as a mixture of

Received August 20, 2008; Accepted December 23, 2008

Address correspondence to Y. L. Lio, Department of Mathematical Sciences, University of South Dakota, Vermillion, SD 57069, USA; E-mail: Yuhlong.lio@usd.edu

the first-order Markov chains. This MTD model significantly reduces the number of parameters to be estimated but still poses flexible structure to model the high-order temporal dependence in Markov chains. Since then, a variety of MTD models have been proposed and successfully applied to analyze many different types of data sequences such as wind speeds, the twilight song of the wood pewee and DNA.

The most commonly used procedures to estimate the parameters in modeling the MTD are the MLE and minimum χ^2 estimation which have been outlined by Raftery and Tavaré (1994). The main difficulties for both estimations are the nonlinearity of objective functions, the storage management for counting the patterns of data sequences in the processes of MLE and the minimum χ^2 estimation, and the large number of constraints imposed on the parameters in the MTD models, although an effective procedure to reduce the number of constraints has been proposed by Raftery and Tavaré (1994). Berchtold (2001) used the idea of balancing an increase in one of the parameters with an equal decrease in another using the boundary adjustment in the MLE process that leads to a modification of the Newton's method. L ebre and Bourguignon (2008) introduced a hidden process for the coefficients of the MTD models and derived an Expectation-Maximization (EM) algorithm for the MLE of the MTD parameters. However, the complexity from the counts of the patterns of sequences is still unsolved for the EM algorithm.

In this article, a transformation method for the MTD parameters is introduced to convert the nonlinear constraints of the parameters to box-constraints such that the MLE procedure for the parameters in the MTD models can be easily implemented through a global search algorithm under R environment (R Development Core Team, 2006). This approach is different from the MLE processes proposed up to date for MTD models.

The outline of the article is as follows. In Sec. 2, MTD models and likelihood functions are reviewed briefly. A transformation for parameters of the MTD models and a hybrid global search algorithms will be addressed. In Sec. 3, a detailed description of the simulations for different modeling processes is described. The proposed estimation approach used for the MTD modelings of three real data sets (song of the wood pewee data, epileptic data, and α A-Crystallin Gene) is presented in Sec. 4 followed by conclusions in Sec. 5.

2. Transformation and Hybrid Optimization for MTD

2.1. High-Order Markov Chain

Let $\{Y_t : t = 1, 2, \dots\}$ be a time homogeneous l th-order Markov chain (MC) on a finite set of m states labeled as $1, 2, \dots, m$. The general objective in modeling a Markov chain is to model the temporal dependence for Y_t and to predict Y_{t+l} as a function of the observations from the previous observations of Y_τ , $\tau < t + l$. The general transition probabilities for l th-order Markov chains can then be denoted as:

$$p(i_0 | i_1, \dots, i_l) = P(Y_{t+l} = i_0 | Y_{t+l-1} = i_1, \dots, Y_t = i_l), \quad t = 1, 2, \dots \quad (2.1)$$

In the special case of first-order MC, the observation at time $t + 1$ is only dependent on the previous observation at time t , then Eq. (2.1) will be simplified as:

$$p(i_0 | i_1, \dots, i_l) = P(Y_{t+1} = i_0 | Y_t = i_1) = q_{i_1 i_0} \quad (2.2)$$

where $i_0, i_1, \dots \in \{1, \dots, m\}$. This first-order MC can then be presented by a transition matrix Q_{MC}^1 as:

$$Q_{MC}^1 = \begin{matrix} & & & & Y_t \\ & & & & 1 & \cdots & m \\ Y_{t-1} & & & & 1 & \cdots & m \\ & 1 & & & q_{11} & \cdots & q_{1m} \\ & \vdots & & & \vdots & \ddots & \vdots \\ & m & & & q_{m1} & \cdots & q_{mm} \end{matrix} \quad (2.3)$$

As a probability distribution, each row of Q_{MC}^1 should sum to one. Hence, only $(m - 1) \times m$ parameters should be estimated in this first-order MC defined in (2.2).

In the l th-order MC as defined in (2.1), the transition matrix becomes complicated and a classical way to manage the transition matrix is using collapsed or reduced form as discussed in Berchtold and Raftery (2002). As an example of order $l = 2$, the transition matrix Q_{MC}^2 can be rewritten as:

$$Q_{MC}^2 = \begin{matrix} & & & & & & & & Y_t \\ & & & & & & & & 1 & \cdots & m \\ & & & & & & & & 1 & \cdots & m \\ Y_{t-2} & & & & & & & & q_{111} & \cdots & q_{11m} \\ & & & & & & & & \vdots & \ddots & \vdots \\ & & & & & & & & m & 1 & \cdots & m \\ & & & & & & & & q_{m11} & \cdots & q_{m1m} \\ & & & & & & & & \vdots & \vdots & \vdots \\ & & & & & & & & 1 & m & \cdots & m \\ & & & & & & & & \vdots & \vdots & \ddots & \vdots \\ & & & & & & & & m & m & \cdots & m \end{matrix} \quad (2.4)$$

Therefore, there are $(m - 1)m^l$ independent parameters to be specified and estimated in the l th order MC defined in (2.1). The number of parameters increases exponentially as the order l , hence there is a definite need of parsimonious model for high-order MC.

2.2. MTD Model

Raftery (1985) introduced the mixture transition distribution (MTD_l) model of order l defined as:

$$p(i_0 | i_1, \dots, i_l) = \sum_{j=1}^l \lambda_j q(i_0 | i_j), \quad (2.5)$$

where $Q^T = \{q(i | j)\}$ is the common first transition matrix satisfying

$$q(i | j) \geq 0 \quad \text{and} \quad \sum_{r=1}^m q(r | j) = 1, \quad j = 1, \dots, m \quad (2.6)$$

and

$$\lambda_j \geq 0, \quad j = 1, \dots, l \quad \text{and} \quad \lambda_1 + \dots + \lambda_l = 1. \quad (2.7)$$

Downloaded By: [South Dakota State University] At: 15:12 26 February 2009

It should be noticed that the transition matrix Q of the MTD model is not the same as the first-order transition matrix of a first-order Markov chain except that both have the same size. To ensure the left side of (2.5) is a proper transition probability, we would also need:

$$\sum_{j=1}^l \lambda_j q(i_0 | i_j) \geq 0 \quad \text{for all } i_0, \dots, i_l. \tag{2.8}$$

With constraints defined in (2.6) and (2.7), constraints in (2.8) are automatically satisfied.

The MTD_l in (2.5) is a mixture of the first-order MC defined in (2.2) which has the same transition matrix Q for each time lag. The number of parameters $(m(m - 1) + l - 1)$ to be estimated in the MTD_l (2.5) is substantially smaller than the number of parameters in the MC (2.1) model of order l . For example, if $m = 4$ (such as in DNA studies) and $l = 2$, there are 13 parameters in MTD_2 model and 48 in the MC model of order 2. If $l = 4$, then there are 15 parameters in MTD_4 model but 768 parameters in the MC (2.1) model of order 4. The number of parameters in MTD_l model (2.5) increases linearly as the order l . However, it increases exponentially as the order, l , of the MC in (2.1).

Since then, an interesting development of the MTD model has included the use of a different transition matrix for each lag; see, for example, in Berchtold (2001). The l th order MTD model with different transition matrix for each lag is denoted by MTD_l^g which is less parsimonious than the MTD (2.5) for order of $l \geq 2$ because it requires $lm(m - 1) + l - 1$ parameters estimated. Moreover, the extended MTD_l^g induces confounding in the model parameters which hampers the statistical inference for the parameters of MTD.

2.3. Maximum Likelihood Estimation

For a given data sequence with length n $\{Y_t : t = 1, 2, \dots, n\}$ from the MTD_l^g model of order l , the likelihood function can be constructed as:

$$\begin{aligned} L(\theta) &= P_\theta(Y_1^n) = P(Y_1^l) P_\theta(Y_{l+1}^n | Y_1^l) \\ &= P(Y_1^l) \times \prod_{t=l+1}^n \left[\sum_{j=1}^l \lambda_j q_j(Y_t | Y_{t-j}) \right] \propto \prod_{t=l+1}^n \left[\sum_{j=1}^l \lambda_j q_j(Y_t | Y_{t-j}) \right], \end{aligned} \tag{2.9}$$

where $Y_{t_1}^{t_2} = (Y_{t_1}, \dots, Y_{t_2})$ is the sequence fragment from t_1 to t_2 for simplicity, $P(Y_1^l)$ is the joint distribution of Y_1^l , and $Q_g^T = \{q_g(i | j)\}$ is the first-order transition matrix given $g = 1, 2, \dots, l$.

To estimate the parameters of MTD_l^g model, we have excluded $P(Y_1^l)$ from the likelihood function. Therefore, the log-likelihood function is

$$l(\theta) = \sum_{t=l+1}^n \log \left[\sum_{j=1}^l \lambda_j q_j(Y_t | Y_{t-g}) \right] = \sum_{i_0, i_1, \dots, i_l} n_{i_0, i_1, \dots, i_l} \log \left[\sum_{j=1}^l \lambda_j q_j(i_0 | i_j) \right] \tag{2.10}$$

where $\theta_g = \{\lambda_g, [q_g(i | j)]_{i,j=1, \dots, m}\}_{1 \leq g \leq l}$ satisfy all nonlinear constraints (2.6) and (2.7) with the replacement of Q by Q_g , $g = 1, 2, \dots, l$. Hence, the MLEs of the parameters θ is the maximizer of $l(\theta)$ subject to all nonlinear constraints (2.6) and (2.7).

It is obvious that the biggest obstacle to find maximizer of $l(\theta)$ is the nonlinear constraints. Berchtold (2001) proposed an efficient iterative process with the boundary adjustment in the MLE process which leads to a modification of the Newton's method. With the constraints (2.7), Lébre and Bourguignon (2008) introduced a hidden process for the coefficients of the MTD models and derived an Expectation-Maximization (EM) algorithm for the MLE of the MTD parameters. However, the complexity from the counts of the patterns of sequences is still unsolved for the search of the maximizer globally.

2.4. Transformation

The most general-purpose optimization methods are based on the simplex algorithm by Nelder and Mead (1965), and quasi-Newton and conjugate-gradient algorithms from Fletcher and Reeves (1964) and Byrd et al. (1995). These methods are generally implemented in most of the softwares with the option for box-constrained optimization in that each parameter is given a lower and/or upper bound. More information about box-constrained optimization can be seen from Nocedal and Wright (1999). It is then a common practice to use an appropriate mathematical transformation changing the other type of constraint to the box-constraint for the implementation of optimization. In this section, we will develop a transformation for the constraint imbedding in the MTD modeling.

Given $g = 1, 2, \dots, l$, there are m sets of constraints in (2.6) and one set of constraint in (2.7) for a MTD $_l^g$. All these $lm + 1$ set of constraints have the same structure that each element is bounded from 0 to 1 and sum to be 1. For each set of constraint, there are k ($= m$ or l) parameters to be estimated where in fact only $k - 1$ independent parameters. That is:

$$0 \leq u_j \leq 1, \quad j = 1, \dots, k \quad \text{and} \quad u_1 + \dots + u_k = 1. \quad (2.11)$$

In order to estimate these $k - 1$ independent parameters, one can use the equation to solve for one of the k parameters, for example, $u_k = 1 - (u_1 + \dots + u_{k-1})$. However, there is no guarantee that u_k will be between 0 and 1. It is highly possible that u_k will be out of this range.

In addition, commonly used statistical software, such as *R* function `optim`, `nlimb`, `nlregb`, are customized to box-constraint and cannot be directly applied here. To confirm this constraint, we introduce a transformation to (2.11) such that the k parameters of u_j be box-constrained with $k - 1$ parameters of v_j , $j = 1, \dots, k - 1$, constraint between 0 and 1 satisfying:

$$\begin{aligned} u_1 &= v_1 \\ u_2 &= v_2(1 - v_1) \\ u_3 &= v_3(1 - v_2)(1 - v_1) \\ &\vdots \\ u_{k-1} &= v_{k-1}(1 - v_{k-2}) \cdots (1 - v_1) \\ u_k &= (1 - v_{k-1})(1 - v_{k-2}) \cdots (1 - v_1). \end{aligned} \quad (2.12)$$

It is easily verifying that the new box-constraint (2.12) satisfies the constraint (2.11) and is easy to be implemented in R.

2.5. Hybrid Global Search Algorithms

In the implementation to simulate and estimate parameters from MTD, we found that the Newton-type of search algorithms such as `optim` implemented in R could yield local minima. For example, modeling the Wood Pewee Song data via the MTD model of order 2 using `optim` with equal probability as the initial parameter input applying the transformation method described above, the resulting MTD model of order 2 has log likelihood value of -572.6874 and BIC 1195.709. However, the maximum log likelihood for Wood Pewee Song data assuming modeling MTD model of order 2 is -567.9955 and BIC for the model selection is 1186.326.

Our experience in statistical modeling and applications (Chen et al., 2000, 2002) led us to pursue the global search algorithms, such as genetic algorithms, etc. It is known that a genetic algorithm is a searching algorithm based on the mechanics of natural selection and genetics to combine the “survival of the fittest” concept yet randomized information exchange to form a search algorithm with some of the innovative flair of human search. A genetic algorithm differs from more conventional search techniques because it considers many points in the search space simultaneously and therefore has a reduced chance of converging to some local optima.

However, it is known that the genetic algorithms do not perform well for its search time when the objective function surface is relatively flat, where the Newton-type search algorithms do better. A hybrid algorithm of genetic algorithm and Newton-type search would take the advantages of these two. This is implemented in the R library `RGENOUD` which stands for “R-GENetic Optimization Using Derivatives” by Mebane and Sekhon which can be found on <http://sekhon.berkeley.edu/papers/rgenoudJSS.pdf>. The approach combining genetic algorithms and standard optimization algorithms is also used in the `MARCH` software implementing Berchtold’s estimation approach for MTD. `Genoud` combines the evolutionary algorithm methods with a derivative-based (quasi-Newton) method to solve difficult optimization problems.

As indicated by the authors, when a statistical model’s estimating function (for example, a log-likelihood) is nonlinear in the model’s parameters, the function to be optimized will generally not be globally concave and may have irregularities such as saddle points or discontinuities. Optimization methods that rely on derivatives of the objective function may be unable to find any optimum at all. Multiple local optima may exist, so that there is no guarantee that a derivative-based method will converge to the global optimum. On the other hand, algorithms that do not use derivative information (such as pure genetic algorithms) are for many problems needlessly poor at local hill climbing. Most statistical problems are regular in a neighborhood of the solution. Therefore, for some portion of the search space, derivative information is useful. `RGENOUD`, via the `cluster` option, supports the use of multiple computers, CPUs, or cores to perform parallel computations.

Switching from `optim` to `genoud` in modeling MTD model for real data, we could possibly select a MTD model with a higher log likelihood value and a smaller BIC.

3. Simulation Studies

A transformation method has been proposed to convert the nonlinear constraints imposed in the MTD models into box-constraints such that a global optimization algorithms in *R* environment such as `optim` and `genoud` will be easily used to find the MLEs of the parameters in the process of MTD modeling. The proposed approach, which allows us remove both the obstacle of constraints and the complexity of counting the patterns of observation sequences that usually occur during the MTD modeling procedure mentioned in Sec. 2, is different from all MLE processes developed up to date for the MTD modeling.

In this section, simulations will be conducted to implement this novel estimation approach for the MTD modeling via MLE procedure. The resulting MLE is called box-constraint MLE.

3.1. Simulation Procedure

Assume that a l th-order Markov chain on a finite set of m states labeled as $1, 2, \dots, m$ has the m^l by m transition probability matrix M and the r th row of M can be specified as follows:

$$p(i_0 | i_1, \dots, i_l) = \sum_{j=1}^l \lambda_j q_j(i_0 | i_j), \quad i_0 = 1, 2, \dots, m \quad (3.1)$$

where $r = 1 + \sum_{k=1}^l (i_k - 1) \times m^{k-1}$ and $Q_g^T = \{q_g(i | j)\}$ is the column stochastic matrix satisfying that for each given $g = 1, 2, \dots, l$,

$$q_g(i | j) \geq 0 \quad \text{and} \quad \sum_{r=1}^m q_g(r | j) = 1, \quad j = 1, \dots, m, \quad (3.2)$$

and

$$\lambda_1 + \dots + \lambda_l = 1. \quad (3.3)$$

Therefore, the l th-order Markov chains with the transition probability matrix M is called the mixture of the first order MTD_1 using $\{\lambda_g, Q_g; g = 1, 2, \dots, l\}$. Given $l \geq 1$ and $\{\lambda_g, Q_g; g = 1, 2, \dots, l\}$, the following processes are carried to generate a sequence of observations $\{Y_t : t = 1, 2, \dots, n\}$ of size n from a l th-order Markov chain process that has the transition probability matrix M . And then the box-constraint MLE is applied for MTD_L^g model building with order of $L \geq 1$:

1. Input the transition matrix M by the following way: given $r = 1, \dots, m^l$, the r th row of M for l th-order Markov chains is generated by linear combination (3.1). In this simulation study, l is chosen to be either 2 or 4 for simplification.
2. Randomly generate an initial status, $\{i_1, \dots, i_l\}$ of l components from $\{1, 2, \dots, m\}$ with replacement.
3. Find the corresponding r th row from general transition probability matrix M where $r = 1 + \sum_{g=1}^l (i_g - 1) \times m^{g-1}$.
4. Find the next state through the probability simulation using the conditional probability distribution which is the selected r th row of M in Step 3.

Table 1
Simulated transition matrices for MTD

Q_1				Q_2				Q_3			
0.100	0.200	0.300	0.400	0.100	0.100	0.100	0.700	0.225	0.140	0.506	0.129
0.400	0.300	0.200	0.100	0.200	0.200	0.400	0.200	0.354	0.300	0.008	0.338
0.200	0.200	0.200	0.400	0.300	0.300	0.300	0.100	0.271	0.123	0.456	0.150
0.400	0.200	0.200	0.200	0.300	0.200	0.300	0.200	0.166	0.191	0.430	0.213

5. Modify the initial status by leaving the first component out and entering the next state as the last component.
6. Go to Step 3 until n observations $\{Y_t : t = 1, 2, \dots, n\}$ are obtained.
7. The box-constraint MLEs of all parameters in the L th-order Markov model MTD_L^g are derived via the application of the optim procedure in R to the likelihood function (2.10), using n observations $\{Y_t : t = 1, 2, \dots, n\}$ obtained in Step 6, with the box-constraints (2.12) defined in Sec. 2.

3.2. Simulation Results and Comparisons

Since the extended MTD_l^g model induces confounding in the model parameters, only the MTD model which has the same first order transition matrix Q for each lag will be investigated in the simulation study. Given a fixed order l and an input $\{Q, \lambda_g; g = 1, 2, \dots, l\}$ for a true MTD model of order l , we generate 1,000 sequences of observations of size $n = 1, 800$ from the true MTD model of order l by the steps 1–6 described in Sec. 3.1. Each of 1,000 samples is used for modeling the MTD of order L . Hence, there are 1,000 box-constraint MLEs for each parameter in $Q^T = \{[q(i | j)]_{i,j=1,\dots,m}\}$. The mean squared error (MSE), variance, and bias for

Table 2

Summary for the simulations. Column of “True λ s” is the number of components which has been simulated as true MTD-2 and MTD-4 followed by MTD-2 (third column), MTD-3 (fourth column), and MTD-4 (fifth column) to model the simulated data. The estimated $\hat{\lambda}$ s are the average of 1,000 simulations

Case	True λ s	$\hat{\lambda}$ s in MTD-2	$\hat{\lambda}$ s in MTD-3	$\hat{\lambda}$ s in MTD-4
1	(0.5, 0.5)	(0.498, 0.502)	(0.486, 0.489, 0.024)	(0.474, 0.479, 0.023, 0.023)
2	(0.3, 0.7)	(0.300, 0.700)	(0.295, 0.681, 0.024)	(0.288, 0.663, 0.025, 0.025)
3	(0.25, 0.25, 0.25, 0.25)	(0.503, 0.497)	(0.336, 0.335, 0.329)	(0.250, 0.251, 0.248, 0.251)
4	(0.10, 0.20, 0.40, 0.30)	(0.321, 0.679)	(0.117, 0.284, 0.599)	(0.096, 0.200, 0.400, 0.304)
5	(0.5, 0.5)	(0.498, 0.502)	(0.489, 0.494, 0.017)	(0.481, 0.488, 0.016, 0.015)
6	(0.3, 0.7)	(0.300, 0.700)	(0.297, 0.686, 0.017)	(0.293, 0.676, 0.016, 0.014)
7	(0.25, 0.25, 0.25, 0.25)	(0.508, 0.492)	(0.337, 0.333, 0.330)	(0.249, 0.250, 0.251, 0.249)
8	(0.10, 0.20, 0.40, 0.30)	(0.351, 0.649)	(0.130, 0.287, 0.583)	(0.100, 0.200, 0.400, 0.300)
9	(0.5, 0.5)	(0.500, 0.502)	(0.492, 0.497, 0.011)	(0.490, 0.495, 0.008, 0.007)
10	(0.3, 0.7)	(0.300, 0.700)	(0.296, 0.694, 0.011)	(0.295, 0.691, 0.007, 0.007)
11	(0.25, 0.25, 0.25, 0.25)	(0.500, 0.500)	(0.333, 0.335, 0.332)	(0.249, 0.252, 0.252, 0.248)
12	(0.10, 0.20, 0.40, 0.30)	(0.395, 0.605)	(0.162, 0.282, 0.556)	(0.097, 0.197, 0.407, 0.300)

each parameter in the common transition matrix Q for the MTD are calculated from these 1,000 box-constraint MLEs. The mean of 1,000 box-constraint MLEs for each parameter in $\lambda_g; g = 1, 2, \dots, L$ is also calculated as the estimate for the corresponding λ_g where $g = 1, 2, \dots, L$.

In order to compare the performances of box-constraint MLE among different modeling processes, a transition matrix Q from Table 1 combining with an order, l , from $\{2, 4\}$ is used as a true model for simulating each sequence of observations. Then each sequence of observations is used for modeling a MTD of order L from $\{2, 3, 4\}$, respectively. The other input parameters $\lambda_g; g = 1, 2, \dots, l$ for the simulation are listed in Table 2 under the column labeled by True λ s. Table 2 summarizes the simulation setups as well as the estimates of λ s for three corresponding modeling MTDs. There are total 12 simulation cases in the study. Cases 1–4 use Q_1 , Cases 5–8 use Q_2 , and Cases 9–12 use Q_3 as the true common transition matrix, respectively. It should be noticed that the three input transition matrices, Q_1, Q_2 , and Q_3 are from L ebre and Bourguignon (2008).

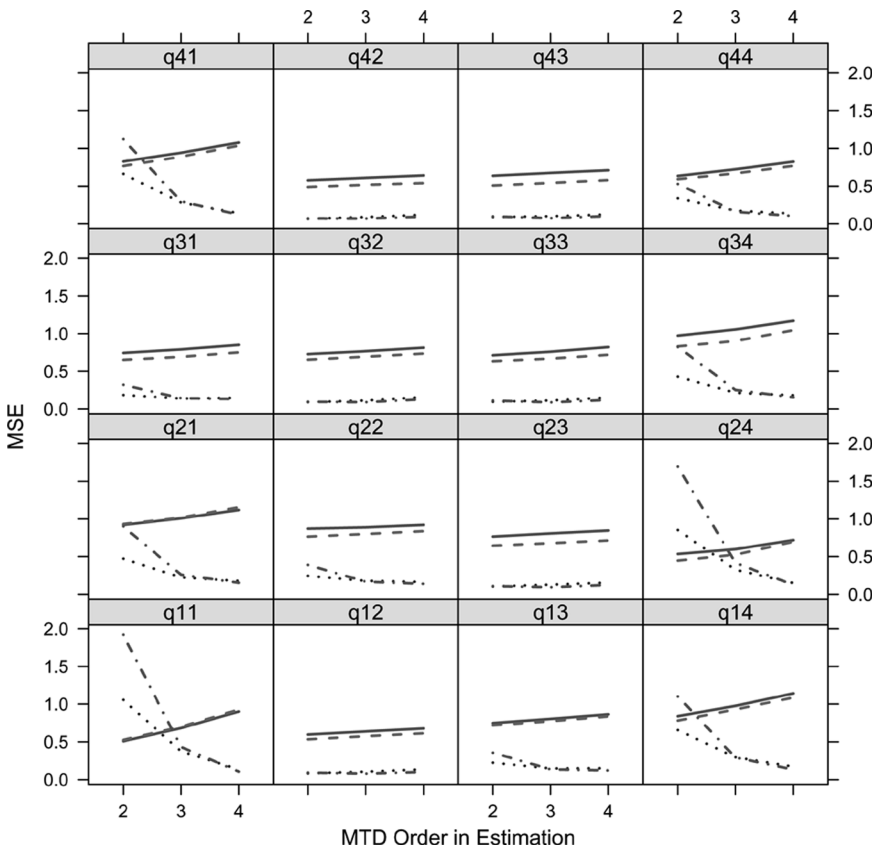


Figure 1. MSE for the estimated transition matrix $\hat{Q} = (\hat{q}_{ij}, i, j = 1, 2, 3, 4)$ as the MTD model order from 2 to 4 for simulation with Q_1 defined in Table 1. Four line types of solid, dashed, dashed with dots, dotted lines correspond to Cases 1–4 in Table 2. Similar figures can be produced for Cases 5–12.

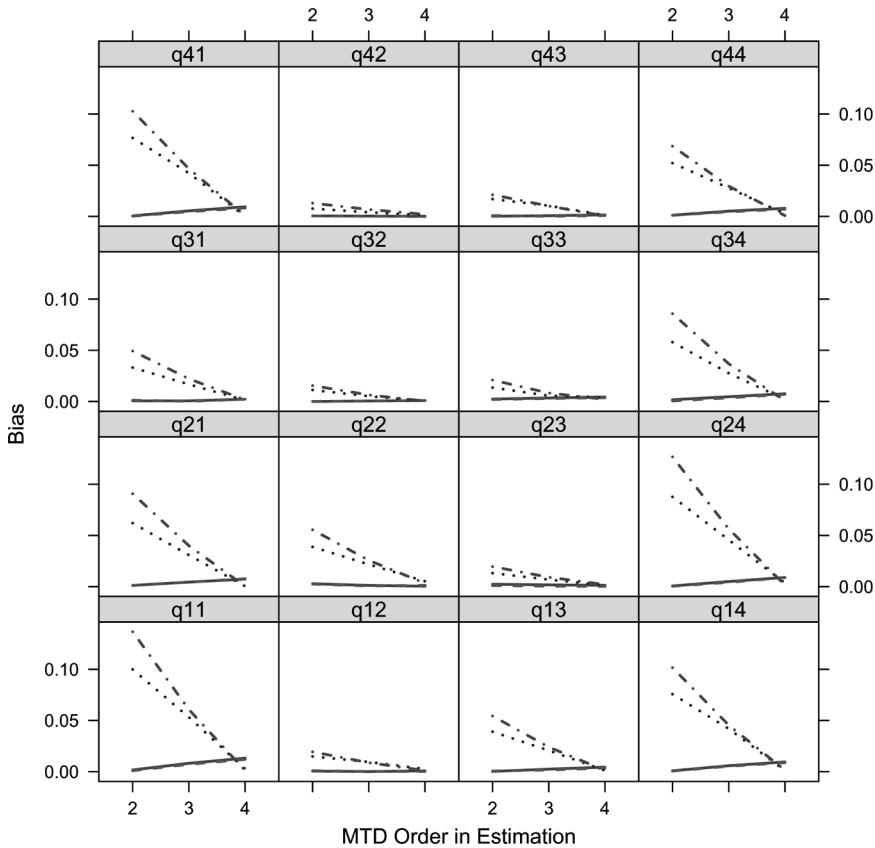


Figure 2. Bias for the estimated transition matrix $\hat{Q} = (\hat{q}_{ij}, i, j = 1, 2, 3, 4)$ as the MTD model order from 2 to 4 for simulation with Q_1 defined in Table 1. Four line types of solid, dashed, dashed with dots, dotted lines correspond to Cases 1–4 in Table 2. Similar figures can be produced for Cases 5–12.

We have done extensive simulations. Part of the simulation results are illustrated in Figs. 1–3 to show the trends on mean squared errors, biases, and variances, respectively, for box-constraint MLEs of the components in the common transition matrix Q_1 . Similar patterns are found for using Q_2 and Q_3 and therefore not produced for brevity. These extra results from simulation are summarized in tables and posted in the authors webpages.

It can be seen that the box-constraint MLE can more accurately estimate the parameters $\{[q(i | j)]_{i,j=1,\dots,m}\}$ as seen in Fig. 1 and produces the smallest bias as seen in Fig. 2 when the order of simulation MTD model is the same as the order of the modeling MTD model. Figure 3 shows that the variance of the box-constraint MLE is increasing when the order of modeling increases. In addition, when the true model is MTD₂, the estimated $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are very close to the true values, and the $\hat{\lambda}_3$ in modeling MTD₃ and $\hat{\lambda}_3$ and $\hat{\lambda}_4$ in modeling MTD₄ are very small. Whereas, when the true MTD model is of order 4, the modeling MTD₂ and the modeling MTD₃ give compromised estimates of λ s as seen in Table 2. This pattern also holds for true MTD models using Q_2 and Q_3 as seen for Cases 5–12 in Table 2.

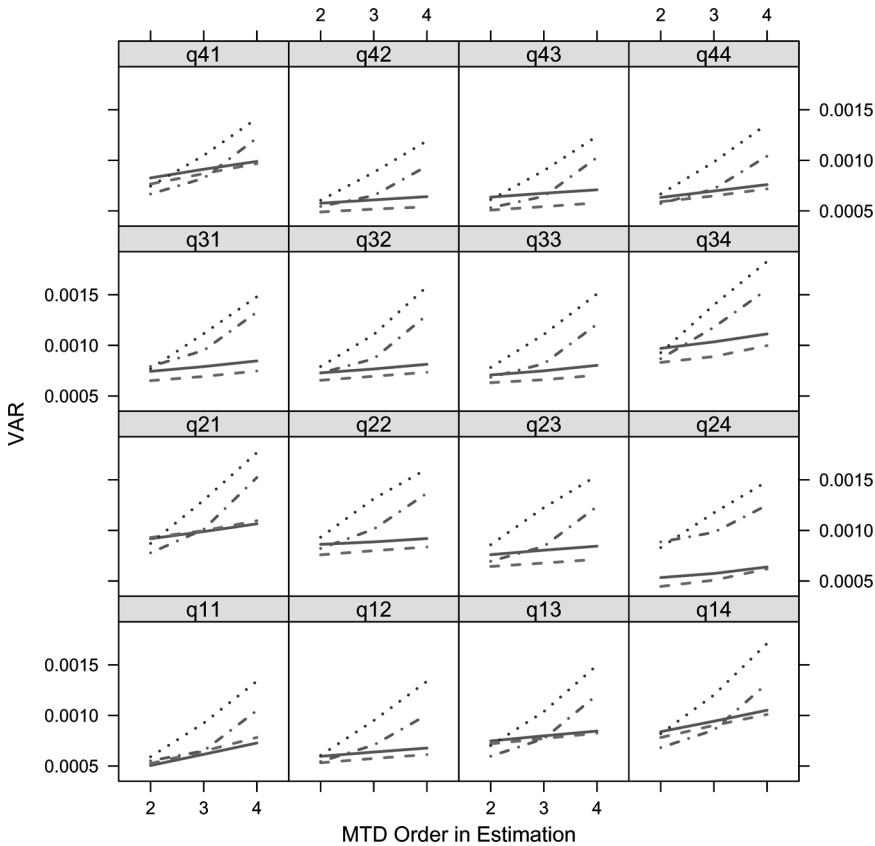


Figure 3. VAR for the estimated transition matrix $\hat{Q} = (\hat{q}_{ij}, i, j = 1, 2, 3, 4)$ as the MTD model order from 2 to 4 for simulation with Q_1 defined in Table 1. Four line types of solid, dashed, dashed with dots, dotted lines correspond to Cases 1–4 in Table 2. Similar figures can be produced for Cases 5–12.

4. Three Real Data Sets

In this section, the proposed transformation method to find MLE of MTD model is applied to analyze three data sets. The first one is the epileptic data; the second one is a time series of the twilight song of the wood pewee; and the third one is the Mouse α A-Crystallin Gene. These data sets had been used by Chatfield and Lemon (1970), Bishop et al. (1975), Raftery and Tavaré (1994), Berchtold (2001), and Lébre and Bourguignon (2008).

For the epileptic data set, the optimal model, obtained by Berchtold and Raftery (2002), was MTD₈ model which had log-likelihood -102.3 and BIC 251.9 . We apply the proposed process to fit the data with MTD models using the common transition matrix Q . The resulting optimal MTD model with common transition matrix is MTD₉, which has log-likelihood -99.12 and BIC 251.17 .

Using the wood pewee song data, Raftery and Tavaré (1994) fitted with general Markov chains in (2.1) of different orders and found that the data set best modeled with an order-two Markov chain which has log-likelihood -368.6 and BIC 866.6 . They applied MTD model of order 2 and 3, resulting log-likelihood -644.3 and

BIC 1338.9 for order 2, and log-likelihood -643.1 and BIC 1343.6 for order 3. They also applied other modelings using the pattern structure of the time series and resulting the final model which had log-likelihood -399.1 and BIC of 827. Berchtold (2001) tried MTD modeling with common transition matrix resulting the optimal MTD model of order 2 which had log-likelihood -568.0 and BIC 1186.3. Berchtold (2001) also fitted with MTD_l^g models which had different transition matrix for each time lag possibly and found the optimal model was MTD_2^g which had log-likelihood -486.4 and BIC 1066.2. We apply the proposed estimation to fit the data with MTD models of order 2 and 3 using the common transition matrix Q . The resulting optimal MTD model is MTD_2 which has loglikelihood -567.996 and has BIC 1186.33. With the common transition matrix, our optimal MTD model is the same as the optimal model by Berchtold (2001) and better than the one in Raftery and Tavaré (1994).

The modeling for the Mouse α A-Crystallin Gene data set, Raftery and Tavaré (1994) fitted with MTD_l model and found the optimal model is order 1 with BIC 3559.7. Their MTD_2 had BIC 3566.1. Again, Berchtold (2001) fitted with MTD_2 and MTD_3 and resulting log-likelihood -1734.0 and BIC 3561.2 for both models. Here, we apply the proposed estimation for MTD model of order 2 and 3 using the common transition matrix Q . The resulting optimal is MTD_2 which has log-likelihood -1733.902 and BIC 3561.085. However, the estimate of λ_2 is 0. Basically, we can see that our resulting model is the same as the optimal model obtained by Berchtold (2001) and the same as the optimal model in Raftery and Tavaré (1994). The common matrix estimated through our process is the same as the matrix in Raftery and Tavaré for order 1.

Our model fittings for these three data sets are summarized in Table 3. The optimal model for each data set shown in Table 3 is comparative to the optimal result of MTD with common transition matrix published on the articles mentioned in this section.

For the song of the wood pewee, the best MTD model is of order 2 with BIC of 1186.326. The estimated parameters are $\hat{\lambda}_1 = 0.0$ and $\hat{\lambda}_2 = 1.0$ with the transition

Table 3
Summary of model fit for three real datasets

Data	Model	No. Parameters	-LL	BIC
Epileptic	MTD2	3	119.50	254.96
	MTD3	4	117.68	256.64
	MTD4	5	113.25	253.08
	MTD5	6	113.25	258.40
	MTD6	7	111.78	260.79
	MTD7	8	109.25	261.05
	MTD8	9	103.54	254.94
	MTD9	10	99.12	251.41
	MTD10	11	98.33	255.17
	Wood Pewee	MTD2	7	567.996
Song	MTD3	8	567.996	1193.52
Mouse α A-Crystallin Gene	MTD2	13	1733.902	3561.085
	MTD3	14	1733.834	3568.125
	MTD4	15	1730.645	3568.923

matrix estimated as

$$\hat{Q} = \begin{pmatrix} 0.904 & 0.077 & 0.019 \\ 0.164 & 0.093 & 0.743 \\ 0.032 & 0.957 & 0.011 \end{pmatrix}. \quad (4.1)$$

For Mouse α A-Crystallin Gene, the best MTD model is of order 2 with BIC of 3561.085. The estimated parameters are $\lambda_1 = 1.0$ and $\lambda_2 = 0.0$ with estimated transition matrix

$$\hat{Q} = \begin{pmatrix} 0.225 & 0.344 & 0.251 & 0.180 \\ 0.228 & 0.316 & 0.269 & 0.187 \\ 0.299 & 0.059 & 0.342 & 0.299 \\ 0.186 & 0.305 & 0.278 & 0.231 \end{pmatrix}. \quad (4.2)$$

For epileptic data, the best MTD model is of order 10 with BIC as 251.4. The estimated parameters are $\hat{\lambda} = (0.03, 0.04, 0.085, 0.182, 0, 0.099, 0.115, 0.243, 0.206)$ with estimated transition matrix

$$\hat{Q} = \begin{pmatrix} 1.000 & 0.000 \\ 0.068 & 0.932 \end{pmatrix}. \quad (4.3)$$

5. Conclusions

A novel approach of MLE for the MTD model has been proposed in which the nonlinear-constraints imbedded in the MTD model are converted into box-constraints through a transformation. The transformation method simplifies the MLE process for the MTD model such that the MLEs of its parameters can be easily obtained by the global search algorithm in *R* environment. An extensive simulation study has demonstrated different advantages of this novel estimation approach in the MTD modeling as well as the applications to three data sets—the epileptic data, the wood pewee, and the Mouse α A-Crystallin Gene—produce the optimal models which are compatible to the optimal MTD models achieved by many authors up to date.

Although only a common transition matrix for the MTD model is used here for brevity, the extension dealing with the MTD_{*j*}^{*g*} model is very straightforward. Hence, this novel procedure provides an efficient way to find the MLE for the MTD model. We believe that it adds to the MTD modeling.

Acknowledgments

The authors would like to thank anonymous referees and the Editor for the comments and suggestions which significantly improved this article.

References

- Berchtold, A. (2001). Estimation in the mixture transition distribution model. *Journal of Time Series Analysis* 22(4):379–397.
- Berchtold, A., Raftery, A. E. (2002). The mixture transition distribution model for high-order Markov chains and non-Gaussian time series. *Statistical Science* 17:328–356.

- Bishop, Y. M. M., Fienberg, S. E., Holland, P. W. (1975). *Discrete Multivariate Analysis*. Cambridge, MA: MIT Press.
- Byrd, R. H., Lu, P., Nocedal, J., Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal of Scientific Computing* 16:1190–1208.
- Chatfield, C., Lemon, R. E. (1970). Analysing sequences of behavioral events. *Journal of Theoretical Biology* 29:427–445.
- Chen, D. G., Hargreaves, B., Ware, D. M., Liu, Y. (2000). A fuzzy logic model with genetic algorithms for analyzing fish stock-recruitment relationship. *Canadian Journal of Fishery and Aquatic Sciences* 57:1878–1887.
- Chen, D. G., Irvine, J. R., Cass, A. (2002). Incorporating Allee effects in fish stock-recruitment models and applications for determining reference points. *Canadian Journal of Fishery and Aquatic Sciences* 59(2):242–249.
- Fletcher, R., Reeves, C. M. (1964). Function minimization by conjugate gradients. *Computer Journal* 7:148–154.
- Lébre, S., Bourguignon, P.-Y. (2008). An EM algorithm for estimation in the mixture transition distribution model. *Journal of Statistical Computation and Simulation* 78:713–729.
- Nelder, J. A., Mead, R. (1965). A simplex algorithm for function minimization. *Computer Journal* 7:308–313.
- Nocedal, J., Wright, S. J. (1999). *Numerical Optimization*. New York: Springer-Verlag.
- R Development Core Team (2006). *A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*. Vienna, Austria.
- Raftery, A. E. (1985). A model for high-order Markov chains. *Journal of Royal Statistical Society Series B* 47:528–539.
- Raftery, A. E., Tavaré, S. (1994). Estimation and modeling repeated patterns in high order Markov chains with the mixture transition distribution model. *Applied Statistics* 43(1):179–199.