Scientific
Research

Modeling of Imperfect Data in Medical Sciences by Markov Chain with Numerical Computation

Mahmoud Afshari[1*], Anoshirvan Ghaffaripour[2]

[1]Department of Statistics, College of Science, Persian Gulf University, Bushehr, Iran
[2]Department of Statistics, College of Science, Yasouj University, Yasuj, Iran
Email: [*]afshar@pgu.ac.ir

Abstract

In this paper we consider sequences of observations that irregularly space at infrequent time intervals. We will discuss about one of the most important issues of stochastic processes, named Markov chains. We would reconstruct the collected imperfect data as a Markov chain and obtain an algorithm for finding maximum likelihood estimate of transition matrix. This approach is known as EM algorithm, which includes main optimum advantages among other approaches, and consists of two phases: phase (maximization of target function). Continue the phase E and M to achieve the sequence convergence of matrix. Its limit is the optimal estimator. This algorithm, in contrast with other optimum algorithms which could be used for this purpose, is practicable in maximum likelihood estimate, and unlike to the methods which involve mathematical, is executable by computer. At the end we will survey the theoretical outcomes with numerical computation by using R software.

Keywords

Markov Chain, Matrix Transition, Maximum Likelihood, EM Algorithm, Missing Data

1. Introduction

Today, we humans are faced with many of the random variables and processes. One of the most important issues

---

[*]Corresponding author.

is the random process of Markov chain. Markov chain is often used to describe how the system changes over time. If we are able to model the system sufficiently to form a Markov chain model, we can then evaluate its various theoretical results through system analysis. In this article, the Matrix transition of Markov chain needed for imperfect observation within a given time is obtained through EM algorithm.

The EM algorithm has two stages: E stage and M stage. This algorithm basically works in this way: We first take into consideration the initial value for the parameter which in this case is the transition matrix.

The full set of data (X) is then reconstructed from missing data by estimation of these parameters in E stage. In the M stage, the maximum likelihood of reconstructed data is maximized to get a new estimation of the parameter. Then, based on this new estimation, the E stage is repeated and we get the second estimation in the M stage. Finally, we continue this process as far as to get convergence estimates.

Dempstr [1], Raj [2], Johnson *et al*. [3] have obtained the value of maximum likelihood of the transition matrix for a sequence of observations.

Among the researchers who examined the issues of statistical inference of Markov chains and published articles in this area are Melichson [4], Capee [5], Chip [6] and Robert [7].

2. Modeling of Missing Data Based on Markov Chain

Following observations in **Table 1** are series of real data obtained about a kind of disease related to 3 patients at different times during 6 months in a hospital. We have missed some of the data because the patients have not attended the hospital in some cases. Number 1 indicates that the patient's condition is satisfactory and number 2 shows the patients are in a very bad condition. The asterisk indicates that in a particular time point, the data has not been observed.

First, the patient's condition is reported as: 1) Next month, no data are is available. Two months later the condition of that patient is reported as; 2) One month later his condition is reported as (1) again, and so on. We call such a set of data incomplete and it is shown as Y.

The incomplete data are collected against complete data (X) which are gathered in all intervals between the first and the last observation.

Such a system can be modeled as a Markov chain with E condition space and transition matrix P.

Generally, we will have $X = N_{ij}$ if we consider O_{ijt} as the number of condition change from i to j at the time interval t. P_{ijk} as the transition probability from i to j condition at t time, N_{ij} as the number of condition change from i to j in a time unit for the whole X data, and $\left(P^r\right)_{ij}$ as the element of matrix P^r, *i.e.* condition change probability r a stage from i to j condition. Having this fact in mind that with the incomplete data Y and the nonzero suitable initial transition matrix, there will be different complete data X with their own probabilities, we estimate the transition matrix by the EM algorithm as follows:

1—First, we change the specified incomplete data (Y) to complete data (X) by considering a suitable amount for those intervals we have no observation and data. In other words, we reconstruct the complete data (X) from the incomplete data.

2—We consider an initial value for an arbitrary non-zero element of the transfer matrix as $P^{(0)}$.

3—We compute maximum likelihood of X observations with the initial transmission matrix $P^{(0)}$ as follows:

Table 1. Missing data for patient.

Month	First patient	Second patient	Third patient
0	1	1	1
1	*	*	2
2	2	*	*
3	1	1	1
4	*	1	*
5	*	*	2
6	1	*	*

$$f\left(x\middle|P^{(0)}\right) = \prod_i \prod_j \left(P_{ij}\right)^{N_{ij}} \Rightarrow \log f\left(x\middle|P\right) = \log \prod_i \prod_j \left(P_{ij}\right)^{N_{ij}} = \sum_i \sum_j N_{ij} \log\left(P_{ij}\right)$$

$$L\left(P\middle|P^{(0)}\right) = E\left[\log f\left(x\middle|P\right)\middle|Y,P^{(0)}\right] = E\left[\sum_i \sum_j N_{ij} \log P_{ij}\middle|Y,P^{(0)}\right] = \sum_i \sum_j \log P_{ij} E\left(N_{ij}\middle|Y \cdot P^{(0)}\right) \qquad (1)$$

4—We calculate the relative maximum (1) and show it as $P^{(1)}$.

5—We repeat the steps 3 and 4 so much that the obtained matrix sequence $\left\{P^{(r)}, r \geq 0\right\}$ gets to convergence

P matrix, in other words, our optimal estimate is $\left\{P^{(r)}\right\}$.

The $E\left(N_{ij}\middle|Y,P^{(0)}\right)$ calculation and unique maximum p are stated through the following theorem.

Theorem 1: Suppose that the matrix transition of the imperfect data in **Table 1** is P, then

$$E\left(N_{ij}\middle|Y,P^{(0)}\right) = \sum_m \sum_n \sum_t O_{mnt} \sum_{k=0}^{t-1} P_{ijk\middle|mnt}$$

Proof: We assume the system at the time t_0 is in the m condition and at the next time interval it is in the n condition. The probability that a change occurs between conditions i and j is done in the k time interval is computed according to the **Figure 1** provided that the system gets to n condition at $t_0 + t$ from the starting point condition m at t_0.

First the system starting from m condition changes to i condition with the probability of $\left(P^k\right)_{mi}$ at the time unit of k. Then the system is likely to change to n condition at $t_0 + t$ with the probability of $\left(P^{l-k-1}\right)_{jn}$.

Suppose that we define A and B as follows:
A : Cases in which changes from i to j condition occurs at $t_0 + t$.
B : Cases in which system starting from m condition at t_0 gets to n condition at $t_0 + t$.
Then we have:

$$0 \leq k \leq t-1, \ P^0 = 1, \ P_{ijk.mnt} = P\left(A\middle|B\right) = \frac{P\left(A \cap B\right)}{P\left(B\right)} = \frac{\left(P^k\right)_{mi} P_{ij}\left(P^{t-k-1}\right)_{jn}}{\left(P^{(t)}\right)_{mn}}$$

It is clear that the probability of the number of condition changes between the i and j conditions along with a change from m to n condition at the t time interval is equal to: $\sum_{k=0}^{t-1} P_{ijk\middle|mnt}$.

So, the expected number of such condition changes in all time intervals is: $O_{mnt} \sum_{k=0}^{t-1} P_{ijk\middle|mnt}$ because m and n conditions repeatedly occurs at t intervals. Therefore, the expected number of transitions from i to j condition on the condition of observed data Y and Matrix transition P accumulating on all observed data is obtained as follows: $E\left(N_{ij}\middle|Y,P^{(0)}\right) = \sum_m \sum_n \sum_t O_{mnt} \sum_{k=0}^{t-1} P_{ijk\middle|mnt}$.

Theorem 2: Suppose that the matrix transition of the imperfect data in **Table 1** is P, then there is a unique relative maximum with its entry is as follows:

$$P_{ij} = \frac{E\left(N_{ij}\middle|Y,P^{(r)}\right)}{\sum_j E\left(N_{ij}\middle|Y,P^{(r)}\right) \neq 0}$$

Figure 1. System gets to *n* condition.

Proof:

$$Q_P = E\left[\sum_i \sum_j \left(N_{ij}\middle|Y,P^{(r)}\right)\log P_{ij}\right]$$

$$= E\left[\sum_{i=1}\sum_{j=2}\left(N_{ij}\middle|Y,P^{(r)}\right)\log P_{ij} + \sum_i\left(N_{i1}\middle|Y,P^{(r)}\right)\log P_{i1}\right] \tag{2}$$

$$= E\left[\sum_{i=1}\sum_{j=2}\left(N_{ij}\middle|Y,P^{(r)}\right)\log P_{ij} + \sum_i\left(N_{i1}\middle|Y,P^{(r)}\right)\log\left(1-\sum_{i=2}P_{i1}\right)\right]$$

Equation (2) has been obtained because the sum of entries in each line is one.

$$\frac{\partial Q_P}{\partial P_{ij}} = E\left[\frac{\left(N_{ij}\middle|Y,P^{(r)}\right)}{P_{ij}} + \left[N_{i1}\middle|Y,P^{(r)}\right]\frac{(-1)}{1-\sum_{i=2}^{\infty}P_{i1}}\right] = E\left[\frac{N_{ij}\middle|Y,P^{(r)}}{P_{ij}} - \frac{N_{i1}\middle|Y,P^{(r)}}{P_{i1}}\right]$$

$$\frac{\partial Q_P}{\partial P_{ij}} = 0 \Rightarrow \frac{E\left[N_{ij}\middle|Y,P^{(r)}\right]}{P_{ij}} = \frac{E\left[N_{i1}\middle|Y,P^{(r)}\right]}{P_{i1}}$$

$$\Rightarrow \frac{P_{ij}}{E\left[N_{ij}\middle|Y,P^{(r)}\right]} = \frac{P_{i1}}{E\left[N_{i1}\middle|Y,P^{(r)}\right]} \tag{3}$$

Adding Equation (3) on i, we have:

$$P_{i1} = \frac{E\left[N_{i1}\middle|Y,P^{(r)}\right]}{\sum_j E\left[N_{ij}\middle|Y,P^{(r)}\right]} \tag{4}$$

By replacing Equation (4) in Equation (3), the proof is completed.

Remark: the fourth stage in algorithm is computed and algorithm is completed as the following:

1) We compute $E\left(N_{ij}\middle|Y,P^{(0)}\right) = \sum_m\sum_n\sum_t O_{mnt}\sum_{k=0}^{t-1}P_{ijk|mnt}$.

2) We compute matrix transition of first stage through this formula:

$$P_{ij} = \frac{E\left(N_{ij}\middle|Y,P^{(0)}\right)}{\sum_k E\left(N_{ik}\middle|Y,P^{(0)}\right) \neq 0}$$

The obtained P_{ij} s for all i and j forms another matrix in the form of $P^{(1)}$ for the next iteration.

3. Numerical Computation

In this section, all the theories mentioned are surveyed on the **Table 1** through a program written in R software.

For actual observations we have **Table 2** which is number of change from i to j at the time interval t.

Conducting computer program written on the bases of the EM algorithm to estimate the matrix transition of the system and the original transition $P_{ij}^{(0)}$, we will have the following outputs:

$$P_{ij}^{(2)} = \begin{bmatrix} 0.2483 & 0.7517 \\ 0.5512 & 0.4488 \end{bmatrix}, \quad P_{ij}^{(1)} = \begin{bmatrix} 0.2509 & 0.7491 \\ 0.5206 & 0.4794 \end{bmatrix}, \quad P_{ij}^{(0)} = \begin{bmatrix} 0.333 & 0.667 \\ 0.667 & 0.333 \end{bmatrix}$$

$$P_{ij}^{(3)} = \begin{bmatrix} 0.2534 & 0.7466 \\ 0.5464 & 0.4536 \end{bmatrix}, \quad P_{ij}^{(4)} = \begin{bmatrix} 0.2515 & 0.7485 \\ 0.5472 & 0.4528 \end{bmatrix}, \quad P_{ij}^{(5)} = \begin{bmatrix} 0.2521 & 0.7479 \\ 0.5471 & 0.4529 \end{bmatrix}$$

$$P_{ij}^{(6)} = \begin{bmatrix} 0.2519 & 0.7481 \\ 0.5471 & 0.4529 \end{bmatrix}, \quad P_{ij}^{(7)} = \begin{bmatrix} 0.2520 & 0.7480 \\ 0.5471 & 0.4529 \end{bmatrix}, \quad P_{ij}^{(8)} = \begin{bmatrix} 0.2520 & 0.7480 \\ 0.5471 & 0.4529 \end{bmatrix}$$

$$P_{ij}^{(9)} = \begin{bmatrix} 0.2520 & 0.7480 \\ 0.5471 & 0.4529 \end{bmatrix}, \quad P_{ij}^{(10)} = \begin{bmatrix} 0.2520 & 0.7480 \\ 0.5471 & 0.4529 \end{bmatrix}, \quad P_{ij}^{(11)} = \begin{bmatrix} 0.2520 & 0.7480 \\ 0.5471 & 0.4529 \end{bmatrix}$$

$$P_{ij}^{(12)} = \begin{bmatrix} 0.2520 & 0.7480 \\ 0.5471 & 0.4529 \end{bmatrix},$$

$$\sum_i \sum_j \log P_{ij} E\left(N_{ij} \middle| Y \cdot P^{(12)}\right) = -14.106 \tag{5}$$

EM Algorithm with another initial matrix will be:

$$P_{ij}^{(2)} = \begin{bmatrix} 0.5650 & 0.4350 \\ 0.1591 & 0.8409 \end{bmatrix}, \quad P_{ij}^{(1)} = \begin{bmatrix} 0.4018 & 0.5982 \\ 0.1094 & 0.8906 \end{bmatrix}, \quad P_{ij}^{(0)} = \begin{bmatrix} 0.75 & 0.25 \\ 0.75 & 0.25 \end{bmatrix}$$

$$P_{ij}^{(3)} = \begin{bmatrix} 0.5059 & 0.4941 \\ 0.1397 & 0.8603 \end{bmatrix}, \quad P_{ij}^{(4)} = \begin{bmatrix} 0.5299 & 0.4701 \\ 0.1475 & 0.8525 \end{bmatrix}, \quad P_{ij}^{(5)} = \begin{bmatrix} 0.2521 & 0.7479 \\ 0.5471 & 0.4529 \end{bmatrix}$$

$$P_{ij}^{(6)} = \begin{bmatrix} 0.2519 & 0.7481 \\ 0.5471 & 0.4529 \end{bmatrix}, \quad P_{ij}^{(7)} = \begin{bmatrix} 0.5228 & 0.4772 \\ 0.1452 & 0.8548 \end{bmatrix}, \quad P_{ij}^{(8)} = \begin{bmatrix} 0.9489 & 0.8642 \\ 2.8409 & 16.7045 \end{bmatrix}$$

$$P_{ij}^{(9)} = \begin{bmatrix} 0.5231 & 0.4769 \\ 0.1453 & 0.8547 \end{bmatrix}, \quad P_{ij}^{(10)} = \begin{bmatrix} 0.5232 & 0.4768 \\ 0.1453 & 0.8547 \end{bmatrix}, \quad P_{ij}^{(11)} = \begin{bmatrix} 0.5232 & 0.4768 \\ 0.1453 & 0.8547 \end{bmatrix}$$

$$P_{ij}^{(12)} = \begin{bmatrix} 0.5232 & 0.4768 \\ 0.1453 & 0.8547 \end{bmatrix},$$

$$\sum_i \sum_j \log P_{ij} E\left(N_{ij} \middle| Y \cdot P^{(12)}\right) = -14.106 \tag{6}$$

As it is seen, the results (5) and (6) are the same in either way.

4. Conclusions

In using the *EM* algorithm, we are practically faced with some issues which should be focused to accelerate our work and save time.

1) If the initial transition matrix $P^{(0)}$, which is arbitrarily chosen, is of a zero entry or the number of transitions from one condition to another one is zero in a unit of time, the corresponding entries of all produced matrices are zero, too. The initial matrix entries must be non-zero to avoid such condition.

2) Increase in the number of time units between successive observations increases the power of transition matrix. If the number of the observed condition changes is relatively low, the time interval between observations is high. Accordingly, with respect to time needed for computation, it is not economical to integrate all these to-

Table 2. Condition change from i to j at the time interval t for missing data.

O_{ij} : number of condition change from i to j at the time t						
Time	1	2	3	4	5	6
O_{11}	1	*	4	1	*	1
O_{12}	1	2	*	*	1	*
O_{21}	1	4	*	1	*	*
O_{22}	*	*	*	*	*	*

gether in the analysis. Actually it is better to keep the maximum time interval at the level in which the probability of condition change is ignored. If the maximum time interval is too large, the increase in the time unit can be effective.

3) If the maximum time intervals are relatively low, the computation time of each stage is short, although different performances may be necessary before achieving convergence.

4) In practical computation, we have accepted that when the maximum difference between successive entries of transition matrix is smaller than a predetermined value, e.g. $\varepsilon = 10^{-4}$. In other words, convergence is achieved when for all values of i, j, and the given value for r, we have: $\left| P_{ij}^{(r+1)} - P_{ij}^{(r)} \right| < \varepsilon$.

Acknowledgements

References

[1] Dempstr, A.P., Larid, N.M. and Rubin, D.B. (1997) Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, **39**, 1-38.

[2] Raj, B. (2002) Asymmetry of Business Cycles: The Markov-Switching Approach, Soft-Tissue Material Properties Under Large Deformation: Strain Rate Effect. *Hand Book of Applied Econometrics and Statistical Inference*, **3**, 687-710.

[3] Johnson, C. and Gallivan, S. (1995) Estimating a Markov Transition Matrix from Observational Data. *Journal of the Operational Researches Society*, **46**, 405-410. http://dx.doi.org/10.1057/jors.1995.55

[4] Melichson, I. (1999) A Fast Improvement to the EM Algorithm on the Own Terms. *Journal of the Royal Statistical Society—Series B*, **51**, 127-138.

[5] Cappee, O., Moulines, E. and Ryden, T. (2005) Inference in Hidden Markov Models. Springer, New York.

[6] Chib, S. (1996) Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models. *Journal of Econometrics*, **75**, 79-97. http://dx.doi.org/10.1016/0304-4076(95)01770-4

[7] Roberts, W. and Ephraim, Y. (2008) An EM algorithm for Ion Chanel Current Estimation. *IEEE Transactions on Signal Processing*, **56**, 26-35. http://dx.doi.org/10.1109/TSP.2007.906743

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.