Lower Bound of Variance for Multipartite Structure of Hidden Markov Model

Gohar Ayub* and Muhammad Iqbal

1Department of Mathematics and Statistics, University of Swat, Swat, KP, Pakistan;
goharayub@uswat.edu.pk
2Department of Statistics, University of Peshawar, Peshawar, Pakistan;
miqbal_70@uop.edu.pk

Abstract

Objective. This study derives an expression for obtaining lower bound of variance for estimates of hidden Markov model. Also, this study provides a parametric procedure for lower bound of variance. Methods/Statistical Analysis: In study a multipartite form of hidden Markov model considered. The states of model observed at time “2t-1” and “2t” respectively. The lower bound expression obtained by Louis methodology. The secondary data is used to study the validity of proposed lower bound. This dataset is a time series and named as “tree ring width”. Study also defines a parametric procedure for computing lower bound of variance. Findings: The study obtained the lower bound of variance for the maximum likelihood estimates of model using real-world data. The study compares the results of variance obtained by two procedures for various combinations of the states. It is found that for some combination of states, the lower bound of variance by two approaches found almost same. While for many combinations of states the lower bound of variance from proposed procedure found smaller than parametric procedure. The overall comparison lower bound of variance for maximum likelihood by proposed method explaining less variation in dataset then conventional i.e. parametric procedure. Conclusion: The study concluded that results for variance by derived expression found sharpen than those of parametric procedure.

Keywords: Hidden Markov Model, Lower Bound Variance, Maximum Likelihood, Multipartite Structure, Parametric Procedure

1. Introduction

The originally introduction of Hidden Markov Model (HMM) found in 1957 with applications in numerous areas of signal processing, medicine, engineering and management1. HMM defined and develop a procedure of estimating the maximum likelihoods (MLE’s) for its parameter’s convergence to actual value. An iteration method developed for computing MLE’s for HMM. A detail method of estimating convergence of HMM using EM-algorithm4. HMM used for forecasting financial time series data China from 1994 to 20044.

The lower bound of variance determined for wireless censor networks using Cramer Rao inequality. The study developed a lower bound of variance for model, based on the time of arrival, using fisher information matrix5. The variance covariance of the normal mixture model and normal HMM was studied using observed information matrix and compared on real-world data6.

*Author for correspondence
The 2-stage HMM studied in human-reboot interaction\textsuperscript{2}. Confidence interval was constructed for parameters of HMM using bootstrap approach for a real data regarding earthquakes\textsuperscript{3}.

A new form of the HMM was developed by dividing the states in to two partitions. Due to complex nature of transitions among the states, the model named as multipartite structure transition. The partitions of the model observed at time point 2t-1 and 2t respectively. The EM algorithm used to estimate the ML of the model\textsuperscript{8}. A methodology detailed out for obtaining variance of MLE’s for HMM using observed information matrix\textsuperscript{10}.

In this study Cramer Rao lower bound of variance derived for MLEs of multipartite structure of HMM. This proposed lower bound of variance differ from formal approach as it based on observed information matrix. Further, proposed technique been explained by a real-life example by considering a dataset “Tree Rings”, which is available in package R\textsuperscript{11}.

This manuscript organized as follow: Introduction followed by mathematical description of model in Section 2. The methodology of lower bound for variance with the derivation discussed in Section 3. The application of the proposed lower bounds of variance to real world data showed and discussed in section 4 and section 5 contains conclusion.

2. Multipartite Structure of Hidden Markov Model

In this section, the mathematical structure of multipartite graph transition of the HMM being described. The model considered times series data \(y_t\) observed at time \(t=1,2,\ldots,2n\). In this model, the states of the model classified into two partition. The first group of states contains \(m_1\) states i.e., \(\epsilon_{1}, \epsilon_{2}, \ldots, \epsilon_{m_1}\) observed at time 2t-1. The second group of states contains \(m_2\) states i.e. \(\epsilon_{m_1+1}, \epsilon_{m_1+2}, \ldots, \epsilon_{m_1+m_2}\) observed at time 2t respectively. The process takes one state at time 2t-1 i.e. \(S_{(2t-1)}\) in one partition and one state at time 2t i.e. \(S_{2t}\) in second partition of states follows a Markov property with homogeneous transition matrix, then

\[
y_{ij} = P(S_{t+1} = \epsilon_j | S_t = \epsilon_i) \; i,j = 1, 2, \ldots, m_1 + m_2
\]

The transition probabilities follow the following restrictions

\[
P(S_{t+1} = \epsilon_j | S_t = \epsilon_i) = \begin{cases} 0 & 1 \leq i, j \leq m_1, 1 \leq i, j \leq m_1 + m_2 \\ 1 \leq i, j \leq m_1, 1 \leq j \leq m_1 + m_2 \\ 1 \leq j \leq m_1, 1 \leq i \leq m_1 + m_2 \\ \end{cases}
\]

Moreover, Markov chain assumed to be at stationary state distribution, taking

\[
\Pi_1 = (\pi_1, \pi_2, \ldots, \pi_{m_1})
\]

\[
\Pi_2 = (\pi_{m_1+1}, \pi_{m_1+2}, \ldots, \pi_{m_1+m_2})
\]

as the stationary distribution for group 1 and group 2 set of states respectively. The observation \(y_t\) at time t, follows \(N(\mu_i, \sigma^2_i)\) in the state \(S_t\). Also, the observation \(y_t\) is conditionally independent and depends only on \(S_t\) at time t, i.e. the conditional distribution of \(y_t\) depends only on \(S_t\).

The stationary state distribution being computed from transition probabilities which omitted from the likelihood function.

The parameters of model are defined as follow:

\[
\Gamma_1 = (y_{ij}) \; 1 \leq i \leq m_1, m_1 + 1 \leq j \leq m_1 + m_2
\]

\[
\Gamma_2 = (y_{ij}) \; 1 \leq j \leq m_1, m_1 + 1 \leq i \leq m_1 + m_2
\]

\[
\Delta = (\mu_i, \sigma^2_i) 1 \leq i \leq m_1 and m_1 + 1 \leq i \leq m_1 + m_2
\]

Then the joint density function of \(X = (S,Y)\), is given as

\[
f(x|\theta) = f_S(x)\]

\[
= \prod_{i=1}^{m_1} \prod_{j=1}^{m_2} \prod_{k=1}^{m_1} \prod_{l=1}^{m_2} \left[ \prod_{g=1}^{n} \left( \frac{1}{\sigma_i g} \right)^{m_{(g-1)}} \frac{1}{\Gamma(g-1)} \right] \left[ \prod_{g=1}^{n} \left( \frac{1}{\sigma_l g} \right)^{m_{(g-1)}} \frac{1}{\Gamma(g-1)} \right] \left[ \prod_{g=1}^{n} \left( \frac{1}{\sigma_j g} \right)^{m_{(g-1)}} \frac{1}{\Gamma(g-1)} \right] \left[ \prod_{g=1}^{n} \left( \frac{1}{\sigma_l g} \right)^{m_{(g-1)}} \frac{1}{\Gamma(g-1)} \right]
\]
where \( s_{ni} \) is the \( i \)th element of \( s \) and \( g(\cdot; \mu_i, \sigma_i^2) \) are the density function.

3. Methodology

The following methodology used for the deriving the expression for lower bound of variance. To check the validity of the derived expression, a numerical example conducted. The secondary data considered regarding tree ring width from the statistical software. The methodology explained as

3.1 Cramer Rao Lower Bound for Variance

Let a random variable generated from function \( f(y; \theta) \), where \( \theta \) is a set of parameters estimated by EM-algorithm for the extended form of HMM. The lower bound of the variance for MLE’s can obtained from observed information matrix using Cauchy-Schwartz inequality. By definition

\[
\begin{align*}
\text{Var}(\hat{\theta}) & \geq \frac{m'(\hat{\theta})^2}{nI(\hat{\theta})} \\
\text{Var}(\hat{\theta}) & \geq \frac{m'(\hat{\theta})^2}{nI(\hat{\theta})}
\end{align*}
\]

3.2 Proposed Lower Bound of Variance for Estimated Parameters

To following expressions indicates the results of lower bound for the estimates of \( \mu_i \) and \( \sigma_i^2 \) observed at time point \( 2t-1 \) for \( 1 \leq i \leq m_1 \) and \( Y_{ij} \) for \( 1 \leq j \leq m_1, m_1 + 1 \leq i \leq m_1 + m_2 \)

\[
\begin{align*}
\text{Var}(\hat{\mu}_i) & \geq \frac{[n\hat{\mu}_i]^2}{n\hat{\sigma}_i^2}1 \leq i \leq m_1 \\
\text{Var}(\hat{\sigma}_i^2) & \geq \frac{[n\hat{\sigma}_i^2]^2}{n\hat{\sigma}_i^2}1 \leq i \leq m_1 + m_2
\end{align*}
\]

3.3 Parametric Bootstrap

The following bootstrap procedure is performed to obtain the lower bound of the parameters of the model.

1. Fit the model with the initial probabilities, mean and variance
2. Generate the sample data from the estimates obtained fitted model
3. Model again fitted by the sample data
4. Repeat the above steps for 5000 times
5. Compute the lower bound for the variance.
4. Results and Discussion

This section explains the application of the lower bound of variance for the parameter of HMM to real-world data for various combinations of states in two partitions. This data is a time series data about the tree rings, contains normalized tree-ring widths in dimensionless units. It’s a univariate time series with 7981 observations each tree ring.

The data set name as “tree rings”, which is free available in the statistical software R\textsuperscript{11}. In this example, the “tree-ring” width for the even year considered in one group and the odd year in 2nd group.

The Table 1 to 6 represents the comparison of variances obtained from two approaches i.e. proposed approach (OIM) and parametric bootstrap (BP) for the various combinations of states. The Table 1 compares

| $m_1=2$ | $m_2=2$ |
|---|---|
| $\gamma_{i1}$ | $\gamma_{i2}$ | $\gamma_{i1}$ | $\gamma_{i2}$ |
| $i$ | OIM | PB | OIM | PB | $j$ | OIM | PB | OIM | PB |
| 1 | 0.15 | 0.20 | 0.15 | 0.20 | 3 | 0.52 | 0.68 | 0.52 | 0.68 |
| 2 | 0.93 | 0.98 | 0.93 | 0.98 | 4 | 0.26 | 0.39 | 0.26 | 0.39 |

**Table 1.** Results of lower bound for $m_1 = 2$ and $m_2 = 2$

| $m_1=2$ | $m_2=3$ |
|---|---|
| $\gamma_{i1}$ | $\gamma_{i2}$ | $\gamma_{i3}$ | $\gamma_{i4}$ | $\gamma_{i5}$ | $\gamma_{i6}$ |
| $i$ | OIM | PB | OIM | PB | OIM | PB | OIM | PB | $j$ | OIM | PB | OIM | PB |
| 1 | 0.39 | 0.45 | 0.177 | 0.189 | 0.51 | 0.51 | 3 | 0.64 | 0.66 | 0.64 | 0.64 |
| 2 | 0.41 | 0.49 | 0.11 | 0.19 | 0.21 | 0.25 | 4 | 0.60 | 0.67 | 0.61 | 0.65 |

**Table 2.** Results of lower bound for $m_1 = 2$ and $m_2 = 3
the variances of two approaches for 2 states in one group and 2 states in second group. The results showed that the lower bound of variance by proposed approach is smaller than those of obtained by parametric approach.

The Table 2 indicates that comparison of variances for the 2 states in one group and 3 states in second group states. The comparison shows that variance of the estimates i.e. mean, variance and transition probability are sharpened than that of variance obtained from parametric approach. The Table 3 indicates that comparison of variances for the 3 states in one group and 2 states in second group states. The comparison shows that lower bound of variance by proposed method found smaller than from parametric approach.

**Table 3.** Results of lower bound for \( m_1 = 3 \) and \( m_2 = 2 \)

|       | \( y_{i_1} \) | \( y_{i_2} \) | \( y_{i_3} \) | \( y_{i_1} \) | \( y_{i_2} \) | \( y_{i_3} \) |
|-------|---------------|---------------|---------------|---------------|---------------|---------------|
| \( i \) |               |               |               | \( j \) |               |               |
| \( OIM \) | \( PB \) | \( OIM \) | \( PB \) | \( OIM \) | \( PB \) | \( OIM \) | \( PB \) | \( OIM \) | \( PB \) |
| 1     | 0.157         | 0.198         | 0.15          | 0.21         | 4             | 0.67          | 0.78          | 0.42          | 0.51          | 0.68          | 0.74          |
| 2     | 0.44          | 0.67          | 0.44          | 0.59         | 5             | 0.72          | 0.92          | 0.46          | 0.46          | 0.39          | 0.42          |
| 3     | 0.126         | 0.143         | 0.59          | 0.77         |               |               |               |               |               |               |               |

| Mean | Variance | Mean | Variance |
|------|----------|------|----------|
| 1    | 0.14     | 0.19 | 0.26     | 0.38     | 4             | 0.39          | 0.58          | 0.177        | 0.211        |
| 2    | 0.81     | 0.97 | 0.24     | 0.29     | 5             | 0.83          | 1.10          | 0.65         | 0.64         |
| 3    | 0.52     | 0.56 | 0.74     | 0.92     |               |               |               |               |               |

**Table 4.** Results of lower bound for \( m_1 = 3 \) and \( m_2 = 3 \)

|       | \( y_{i_1} \) | \( y_{i_2} \) | \( y_{i_3} \) | \( y_{i_1} \) | \( y_{i_2} \) | \( y_{i_3} \) | \( y_{i_1} \) | \( y_{i_2} \) | \( y_{i_3} \) |
|-------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| \( i \) |               |               |               | \( j \) |               |               | \( OIM \) | \( PB \) | \( OIM \) | \( PB \) | \( OIM \) | \( PB \) | \( OIM \) | \( PB \) |
| 1     | 0.41          | 0.49          | 0.19          | 0.18         | 0.45          | 0.49          | 4             | 0.42          | 0.54          | 0.21          | 0.35          | 0.28          | 0.35          |
| 2     | 0.46          | 0.53          | 0.24          | 0.27         | 0.11          | 0.18          | 5             | 0.22          | 0.28          | 0.25          | 0.36          | 0.28          | 0.31          |
| 3     | 0.41          | 0.49          | 0.28          | 0.31         | 0.44          | 0.48          | 6             | 0.42          | 0.51          | 0.24          | 0.24          | 0.33          | 0.26          |

| Mean | Variance | Mean | Variance | Mean | Variance |
|------|----------|------|----------|------|----------|
| 1    | 0.004    | 0.005| 0.14     | 0.19 | 4         | 0.169       | 0.178        | 0.37         | 0.423        |
| 2    | 1.44     | 1.53 | 0.314    | 0.309| 5         | 0.85        | 0.99         | 0.68         | 0.72         |
| 3    | 0.38     | 0.42 | 0.236    | 0.24 | 6         | 0.73        | 0.99         | 0.48         | 0.59         |
Table 5. Results of lower bound for $m_1 = 4$

|   | $y_{i5}$ | $y_{i6}$ | $y_{i7}$ | $y_{i8}$ |
|---|---------|---------|---------|---------|
| i | OIM     | PB      | OIM     | PB      |
| 1 | 0.15    | 0.23    | 0.75    | 0.81    |
| 2 | 0.31    | 0.39    | 0.42    | 0.53    |
| 3 | 0.84    | 0.86    | 0.63    | 0.68    |
| 4 | 0.24    | 0.38    | 0.44    | 0.50    |

| Mean | Variance |
|------|----------|
| 1    | 0.193    | 0.245    | 0.451    | 0.512    |
| 2    | 0.274    | 0.284    | 0.311    | 0.412    |
| 3    | 0.92     | 0.142    | 0.565    | 0.595    |
| 4    | 0.231    | 0.324    | 0.73     | 0.074    |

Table 6. Results of lower bound for $m_2 = 4$

|   | $y_{j1}$ | $y_{j2}$ | $y_{j3}$ | $y_{j4}$ |
|---|---------|---------|---------|---------|
| j | OIM     | PB      | OIM     | PB      |
| 5 | 0.21    | 0.28    | 0.35    | 0.39    |
| 6 | 0.53    | 0.68    | 0.62    | 0.74    |
| 7 | 0.85    | 0.67    | 0.34    | 0.29    |
| 8 | 0.12    | 0.19    | 0.36    | 0.49    |

| Mean | Variance |
|------|----------|
| 5    | 0.186    | 0.197    | 0.803    | 0.762    |
| 6    | 0.21     | 0.37     | 0.921    | 0.886    |
| 7    | 0.378    | 0.412    | 0.611    | 0.816    |
| 8    | 0.491    | 0.409    | 0.105    | 0.188    |

The Table 4 compares variance of two approaches for 3 states in one group and 3 states in second group. The overall comparison of table indicating that lower bound of variance by proposed approach observed smaller than the variance by parametric method. Similarly, the Tables 5
and 6 define the comparison of variance for 4 states in one group and 4 states in second group respectively.

5. Conclusion

We proposed an expression for the lower bound of variance of likelihood estimates of parameters i.e. mean $\mu_i$, variance $\sigma_i^2$ and transition probabilities $\gamma_{ij}$ of multipartite structure of HMM. The proposed lower bound computed for the various combinations of states in two partitions. The overall comparison indicated that lower bound of variance by proposed approach shows less variation i.e. found sharpen in the selected dataset and confirms the applicability.

6. References

1. MacDonald IL, Zucchini W. Hidden Markov and other models for discrete-valued time series. CRC Press; 1997. p. 1–5. PMCid:PMC1050837
2. Baum EL, Petrie T, Soules G, Weiss N. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chain. The Annals of Mathematical Statistics. 1970; 41(1):164–71. https://doi.org/10.1214/aoms/1177697196.
3. Kikuchi Y. Two modifications of hidden Markov model and their model comparison. Memoirs of the Faculty of Science Kochi University Series A Mathematics. 2002; 23:99–127.
4. Zhang Y. Prediction of financial time series with Hidden Markov Models. Applied Sciences: School of Computing Science; 2004. p. 102.
5. Jia T, Buehrer RM. A new Cramer-Rao lower bound for TOA-based localization. IEEE Military Communications Conference; 2008. p. 1–5. PMCid:PMC3496566
6. Iqbal M, Nishi A, Kikuchi Y, Nomakuchi, K. Estimation of the variance for the maximum likelihood estimates in normal mixture models and normal hidden Markov models. Journal of Japanese Society of Computational Statistics. 2011; 24:39–66. https://doi.org/10.5183/jjscs.1002001_183
7. Thanh NN, Lee S, Kim D. Two stage hidden Markov model in gesture recognition for human robot interaction. International Journal of Advanced Robotic Systems. 2012; 9(39):1–10.
8. Inaba Y, Miyaoka E. Bootstrapping Methods for constructing confidence intervals for estimated parameters of Hidden Markov model. Advances and Applications in Statistics. 2015; 46(2):97–105. https://doi.org/10.17654/ADASAug2015_097_105.
9. Kikuchi Y, Iqbal M, Nishi A, Nomakuchi K. Hidden Markov model with a multipartite graph structure Transition. Kochi Journal of Mathematics. 2015; 10:105–35.
10. Louis TA. Finding the observed information matrix when using the EM algorithm. Journal of the Royal Statistical Society. 1982; 44:226–33.
11. R: A language and Environment for Statistical Computing [Internet]. [cited 2015 Feb 10]. Available from: https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing.