Journal of Environmental Statistics
September 2018, Volume 8, Issue 9. http://www.jenvstat.org

Modeling of Dengue Fever Death Counts Using Hidden Markov Model

Joshni George
Department of Statistics,
St.Thomas College, Palai,
Mahatma Gandhi University,
Kerala-686574 India.

Seemon Thomas
Department of Statistics,
St.Thomas College, Palai,
Mahatma Gandhi University,
Kerala-686574 India.

Abstract

We explore the use of Poisson-hidden Markov model to describe an overdispersed data on monthly death counts due to Dengue fever. Independent Poisson mixture models of various components and stationary Poisson hidden Markov models of different states are fitted and the performance of each model is judged using model selection criteria. The sequence of hidden states are estimated based on the best fitted model. The method can be applied in identifying environmental factors affecting a stochastic process.

Keywords: Mixture model, Poisson-hidden Markov model, Viterbi algorithm, Transition probability, Stationary distribution.

1. Introduction

Dengue fever is one of the widely spreading diseases mostly in tropical and sub-tropical areas. Global incidence of Dengue is drastically increasing in the recent years. The year 2015 was characterized by large Dengue outbreaks worldwide. It is widespread throughout the tropics with local variations in risk influenced by environmental factors such as rainfall, temperature and unplanned rapid urbanization. India turned out to be a country with the world's highest Dengue burden with about 34 per cent of the total Dengue fever cases reported in the world. The number of deaths occurred due to Dengue fever is highest in Kerala when compared to other states in India. Studies reveal that this is mainly beacuse of the favourable environmental conditions in Kerala for the growth of Aedes aegypti mosquitoes which carry the Dengue virus. We make use of the time series data of monthly deaths occurred due to Dengue fever in Kerala from January 2006 to December 2015. The data is collected from the official website of the Directorate of Health Services, Kerala.

Usually for unbounded counts, Poisson distribution is a natural choice to describe them. But it is not suitable for situations where the variance is much larger than the mean. The data considered here is an overdispersed one and hence a simple Poisson distribution is not a suitable model. A commonly used alternative model is negative binomial distribution. But the structure of the present data does not allow a negative binomial distribution. For overdispersed count data set, an independent mixture of Poisson models can be used as it can accommodate unobserved heterogeneity in the population. But independent mixture of Poisson models also insufficient as the observations in this study are serially dependent. Hidden Markov models, that allow the probability distribution of each count to depend on the hidden state of a Markov chain can accommodate both overdispersion and serial dependence. In this context we considered Poisson-hidden Markov model to describe the situation more precisely.

Section 2 contains preliminary concepts and results required for the study. In Section 3, an independent mixture of Poisson models and different hidden Markov models are fitted to the Dengue fever death count data and the performance of each model is judged. Finally, a discussion of the results and inferences are given in Section 4.

2. Preliminaries

2.1. Mixture Model

An independent mixture distribution consists of m component distributions and a mixing distribution which selects from these components. We consider a sequence of observations, y_t, made at times t = 1, 2,..., n, where y_t represents the observed number of counts and let Y_t be the corresponding random variable. Let $\delta_1, \ldots, \delta_m$ such that $\sum_{i=1}^{m} \delta_i = 1$ be the respective assigned probabilities of m component distributions having probability functions $p_1, \ldots, p_m$. If Y_t is a random variable having the mixture distribution, then it has the following probability function:

$$P(Y_t = y) = \sum_{i=1}^{m} \delta_i p_i(y). \tag{1}$$

That is, an observation is generated by one of m distributions where the choice of the distribution p_i is made by a second random mechanism, called the parameter process. Here $\delta_1, \ldots, \delta_m$ are the mixing parameters such that $\sum_{i=1}^{m} \delta_i = 1$, and p_i the probability function of the component random variable X_i. The k^{th} moment about the origin is the following:

$$E(Y_t^k) = \sum_{i=1}^{m} \delta_i E(X_i^k),\ k = 1, 2, \ldots$$

The parameter estimation of a mixture distribution is performed by the method of maximum likelihood estimation. The likelihood of the sequence of observation $y_1, \ldots, y_n$ for an m

component mixture model is the following:

$$L(\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_m;\delta_1,\ldots,\delta_m|y_1,\ldots,y_n)=\prod_{j=1}^{n}\sum_{i=1}^{m}\delta_i p_i(y_j,\boldsymbol{\theta}_i) \tag{2}$$

where $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_m$ are the parameter vectors of component distributions. We assume that each Dengue fever death count is generated by one of m Poisson distributions, with means $\lambda_1,\ldots,\lambda_m$, where the choice of mean λ_i is selected with probability δ_i, where $i=1,2,\ldots,m$ and $\sum\limits_{i=1}^{m}\delta_i=1$. Thus $2m-1$ independent parameters are to be estimated. Here the problem of unbounded likelihood does not arise as the observations are discrete.

An independent mixture model does not allow for the serial dependence in the observations. The data considered in this study is a serially dependent one. One easy way of allowing serial dependence in the observations is to relax the assumption that the parameter process is serially independent. A most elementary and simple form of dependence is Markov dependence. So we assume that the parameter process is a Markov chain. The resulting model for the observations is called a Markov-switching model or Markov-dependent model or Poisson-hidden Markov model.

2.2. Hidden Markov Model

Hidden Markov models (HMMs) are models in which the distribution that changes through time according to the states of a hidden Markov chain. The theory of HMMs were introduced by Baum and Petrie (1966) and Baum et al. (1970). The underlying systems are directly observable in Markov models. But in real-world applications, the system may be unobservable but observable through another stochastic process. In HMMs, the state of an underlying system at any time t is unobservable. The observations are the outputs of another stochastic process under the influence of a hidden process. Therefore, the system follows a hidden process having the Markov property. A detailed description of the properties of HMMs can be found in Rabiner (1989) and Zucchini and MacDonald (2009). More accounts of the history of HMMs can be found in Cappe et al. (2005) and MacDonald and Zucchini (1997). HMMs were applied in the analysis of biological sequences and in signal processing applications. HMMs provide a powerful and flexible mathematical structure to make statistical inferences on partially observed stochastic processes more accurate. A wide spectrum of applications of HMMs can be seen from Juang and Rabiner (1991), Durbin et al. (1998) and Guttorp (1995).

Let $\{C_t : t=1,2,...\}$ represents the unobserved parameter process satisfying Markov property and $\{Y_t : t=1,2,...\}$ is the state-dependent process such that, when C_t is known, the distribution of Y_t depends only on current state C_t and not on previous states or observations. If $\mathbf{Y}^{(t)}$ and $\mathbf{C}^{(t)}$ representing the histories from time 1 to time t, then HMM $\{Y_t : t=1,2,...\}$ is a particular kind of dependent mixture, in such a way that:

$$\begin{aligned} P(C_t \mid \mathbf{C^{(t-1)}}) &= P(C_t \mid C_{t-1}),\ t=2,3,... \\ P(Y_t \mid \mathbf{Y^{(t-1)}},\mathbf{C^{(t)}}) &= P(Y_t \mid C_t), t \in \mathbb{N}. \end{aligned}$$

In the case of discrete observations, the probability mass function of Y_t if the Markov Chain is in state i at time t is the following:

$$p_i(y)=P(Y_t=y \mid C_t=i).$$

Now define $u_i(t) = P(C_t = i)$ for $t = 1, \ldots, n$. Then

$$\begin{aligned}
P(Y_t = y) &= \sum_{i=1}^{m} P(C_t = i) P(Y_t | C_t = i) \\
&= \sum_{i=1}^{m} u_i(t) p_i(y) \\
&= (u_1(t), \ldots, u_m(t)) \begin{pmatrix} p_1(y) & & 0 \\ & \ddots & \\ 0 & & p_m(y) \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \\
&= \boldsymbol{u}(t) \boldsymbol{P}(y) \mathbf{1}'.
\end{aligned}$$

where $\boldsymbol{P}(y)$ is the diagonal matrix with i^{th} diagonal element $p_i(y)$, $\boldsymbol{u}(t) = (u_1(t), \ldots, u_m(t))$ is the initial distribution of the Markov chain. If $\boldsymbol{u}(t)$ is the initial distribution and $\boldsymbol{\Gamma} = (\gamma_{ij})$ is the transition probability matrix of the Markov chain, then the distribution at time $t+1$ can be given as follows:

$$\boldsymbol{u}(t+1) = \boldsymbol{u}(t) \boldsymbol{\Gamma}.$$

Hence

$$P(Y_t = y) = \boldsymbol{u}_1 \boldsymbol{\Gamma}^{t-1} \boldsymbol{P}(y) \mathbf{1}'.$$

Therefore

$$E(Y_t) = \sum_{i=1}^{m} E(Y_t | C_t = i) P(C_t = i) = \sum_{i=1}^{m} u_i(t) E(Y_t | C_t = i).$$

If the Markov chain is stationary and has stationary distribution $\boldsymbol{\delta}$, then $\boldsymbol{\delta} \boldsymbol{\Gamma}^{t-1} = \boldsymbol{\delta}$ for all $t \in \mathbb{N}$ and hence

$$\boldsymbol{P}(Y_t = y) = \boldsymbol{\delta} \boldsymbol{P}(y) \mathbf{1}'.$$

Consider a stationary m-state Poisson-HMM $\{Y_t : t = 1, 2, \ldots\}$ with transition probability matrix $\boldsymbol{\Gamma}$ and state-dependent means $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$. Let $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)$ be the stationary distribution of the Markov chain and $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\lambda})$. Then

$$\begin{aligned}
E(Y_t) &= \boldsymbol{\delta} \boldsymbol{\lambda}'. && (3) \\
E({Y_t}^2) &= \sum_{i=1}^{m} (\lambda_i^2 + \lambda_i) \delta_i = \boldsymbol{\delta} \boldsymbol{\Lambda} \boldsymbol{\lambda}' + \boldsymbol{\delta} \boldsymbol{\lambda}'. \\
\mathrm{Var}(\mathrm{Y_t}) &= \boldsymbol{\delta} \boldsymbol{\Lambda} \boldsymbol{\lambda}' + \boldsymbol{\delta} \boldsymbol{\lambda}' - (\boldsymbol{\delta} \boldsymbol{\lambda}')^2 \geq E(Y_t). && (4) \\
E(Y_t Y_{t+k}) &= \boldsymbol{\delta} \boldsymbol{\Lambda} \boldsymbol{\Gamma}^k \boldsymbol{\lambda}'; \ k \in \mathbb{N}. \\
\rho(k) &= \mathrm{Corr}(\mathrm{Y_t}, \mathrm{Y_{t+k}}) = \frac{\boldsymbol{\delta} \Lambda \Gamma^{\mathrm{k}} \boldsymbol{\lambda}' - (\boldsymbol{\delta} \boldsymbol{\lambda}')^2}{\boldsymbol{\delta} \Lambda \boldsymbol{\lambda}' + \boldsymbol{\delta} \boldsymbol{\lambda}' - (\boldsymbol{\delta} \boldsymbol{\lambda}')^2}; \ \mathrm{k} \in \mathbb{N}. && (5)
\end{aligned}$$

If $y_1, y_2, \ldots, y_n$ is the observation sequence generated by a HMM, and $\boldsymbol{\delta}$ be the initial distribution which is assumed to be same as the stationary distribution implied by the transition probability matrix $\boldsymbol{\Gamma}$, then the likelihood function is the following:

$$L = \boldsymbol{\delta} \boldsymbol{\Gamma} \mathbf{P}(y_1) \boldsymbol{\Gamma} \mathbf{P}(y_2) \ldots \boldsymbol{\Gamma} \mathbf{P}(y_n) \mathbf{1}'. \tag{6}$$

We use numerical maximization of the likelihood for estimating the parameters of the HMM. Given some observed sequence $\boldsymbol{O} = (O_1, O_2, \ldots, O_r)$ of outputs, we wish to obtain the state sequence $\boldsymbol{S} = (S_1, S_2, ..., S_r)$ that has the highest conditional probability. Prediction of most likely sequence of states for each time point t by maximizing the joint probability is known as global decoding. Global decoding can be done using Viterbi algorithm and the details of this algorithm can be found in Viterbi (1967) and Zucchini and MacDonald(2009).

3. Modeling of Dengue Fever Death Counts

The monthly series of death counts due to Dengue fever in Kerala from 2006 January to 2015 December is displayed in Table 3.1 and in Figure 3.1.

Table 3.1: Monthly series of the number of deaths due to Dengue fever in Kerala from 2006 January to 2015 December

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Jan	0	0	0	0	0	0	1	1	0	0
Feb	0	0	0	0	0	0	1	2	0	1
Mar	1	0	1	1	1	1	1	0	0	1
Apr	1	0	1	0	1	0	0	0	2	1
May	1	0	1	1	0	0	1	0	7	2
Jun	0	3	6	6	1	6	1	0	6	10
Jul	1	0	4	2	3	0	0	0	8	3
Aug	0	0	1	0	3	2	0	1	1	2
Sep	0	0	1	3	3	0	0	3	3	1
Oct	1	0	2	0	1	2	0	3	0	3
Nov	0	0	0	1	0	0	1	0	1	5
Dec	0	0	0	2	0	0	0	0	1	0

For the data given in Table 3.1, the variance = 3.3277, which is much larger than the mean = 1.1583. Hence a single Poisson distribution will not be a good choice to represent the data. Since the unimodality of the data is in doubt it is tested using Hartigans' Dip Test. The value of the Dip test statistic obtained is 0.1375 with a p value 2.2×10^{-16}. As the data is multimodal let us try independent mixture of Poisson models and Poisson hidden Markov models to describe the data.

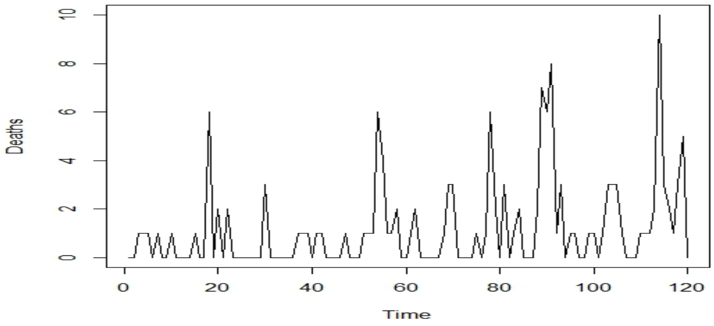

Figure 3.1: The number of deaths occurred in Kerala due to Dengue fever in 120 consecutive months (from 2006 January-2015 December).

3.1. Fitting of Independent Mixture of Poisson Models

The proposed mixture model is (1) where

$$p_i(y) = \frac{e^{\lambda_i}\lambda_i^y}{y!}; \lambda_i \ > \ 0, \ y \ = \ 0, 1, \ldots$$

and 0, elsewhere. Then the likelihood given in (2) takes the form:

$$L(\lambda_1, \ldots, \lambda_m; \ \delta_1, \ldots, \delta_m | y_1, \ldots, y_n) = \prod_{t=1}^{n} \sum_{i=1}^{m} \delta_i \frac{e^{\lambda_i}\lambda_i^{y_t}}{y_t!} \ ; \ y_t = 0, 1, \ldots. \tag{7}$$

Using R code one can maximize logarithm of (7) and estimate the parameters $\boldsymbol{\delta}$ and $\boldsymbol{\lambda}$ constrained by $\sum \delta_i = 1$, $\delta_i, \lambda_i > 0$ for $i = 1, 2, \ \ldots, \ m$.

Independent mixtures of two, three and four Poisson models are fitted and their Akaike Information criterion (AIC) and Bayesian Information Criterion (BIC) values are also computed for model selection. AIC and BIC are respectively defined as follows:

$$\begin{aligned} AIC &= -2 \ \text{log L} + 2 \ \text{p} \\ BIC &= -2 \ \text{log L} + \text{p log n} \end{aligned}$$

where log L is the log-likelihood of the fitted model, p denotes the number of parameters of the model and n, the number of observations. More details regarding model selection can be seen from Zucchini (2000). The different models fitted and their AIC and BIC values are shown in Table 3.2.

Table 3.2: Independent mixture of Poisson models fitted to the series of the number of deaths due to Dengue fever.

Number of components (m) of independendent Poisson mixture model	i	δ_i	λ_i	- log L	AIC	BIC	Mean	Variance
m=2	1	0.8542	0.5786					
	2	0.1458	4.5560	201.6819	409.3638	403.7463	1.1583	1.9696
m=3	1	0.6287	0.3728					
	2	0.2836	1.5018	195.5715	401.143	415.0805	1.1583	2.2136
	3	0.0877	5.6784					
m=4	1	0.2837	0.3728					
	2	0.3450	0.3728	194.6919	403.3838	422.8962	1.1583	2.2136
	3	0.2836	1.5018					
	4	0.0877	5.6784					

From Table 3.2 one can see that AIC selects three-component whereas BIC selects two-component independent Poisson mixture models.

Since independent mixture model ignores the possibility of serial dependence of observations, HMMs are now considered. We shall next fit Poisson-hidden Markov models to the data given in Table 3.1. We shall fit Poisson-HMM with one, two, three and four states to the data.

3.2. Fitting of Poisson-Hidden Markov Model

To fit a stationary Poisson-HMM we have to estimate $\boldsymbol{\delta}$, $\boldsymbol{\lambda}$ and $\boldsymbol{\Gamma}$ by maximizing the likelihood given in (7). Obviously, maximization of (7) and estimation of parameters are to be done subject to the following constraints:

$$\lambda_i \ \ \delta_i > \ 0, \ i \ = \ 1, \ldots, m, \sum_{i=1}^{m} \delta_i = 1, \sum_{j=1}^{m} \gamma_{ij} = 1 \text{ and } \gamma_{\mathrm{ij}} \ \geq 0 \ \forall \text{ i, j.}$$

By suitable reparameterization, unconstrained maximization of the likelihood is possible and the estimates can be obtained by using numerical optimizer ***nlm*** available in ***R***.

The estimates of the transition probability matrix $\boldsymbol{\Gamma}$ obtained by ***R*** code for different states are given below.

For two-state Poisson-HMM, the estimate of $\boldsymbol{\Gamma}$ obtained is the following:

$$\boldsymbol{\Gamma} = \begin{pmatrix} 0.9051 & 0.0949 \\ 0.4122 & 0.5878 \end{pmatrix}$$

.

The three-state HMM has

$$\boldsymbol{\Gamma} = \begin{pmatrix} 0.8504 & 0.1496 & 0.0000 \\ 0.4148 & 0.4584 & 0.1268 \\ 0.0000 & 0.6216 & 0.3784 \end{pmatrix}$$

and the four-state model has

$$\boldsymbol{\Gamma} = \begin{pmatrix} 0.9157 & 0.0843 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 1.0000 \\ 0.3613 & 0.0000 & 0.6387 & 0.0000 \\ 0.0000 & 0.0000 & 0.6893 & 0.3107 \end{pmatrix}.$$

The corresponding estimates of $\boldsymbol{\delta}$ and $\boldsymbol{\lambda}$ are shown in Table 3.3. AIC and BIC values for each fitted model are also shown in Table 3.3. On comparing AIC and BIC values of all the fitted models in Table 3.2 and Table 3.3, two-state Poisson-HMM has the least values. Hence AIC and BIC both select two state HMM.

Table 3.3: Poisson-hidden Markov model(P-HMM) fitted to the number of deaths occurred due to Dengue fever.

Model	i	δ_i	λ_i	- log L	AIC	BIC	Mean	Variance
1-state P-HMM	1	1	1.1583	211.2421	424.4842	426.4842	1.1583	1.1583
2-state P-HMM	1	0.8129	0.5007	171.4762	350.9524	362.1024	1.1485	2.9754
	2	0.1871	3.9658					
3-state P-HMM	1	0.6972	0.3802	168.6712	355.3424	380.4299	1.1472	3.327
	2	0.2515	2.1674					
	3	0.0513	6.5690					
4-state P-HMM	1	0.6944	0.4222	165.0656	362.1311	406.731	1.1447	2.5603
	2	0.0586	1.2675					
	3	0.1621	1.8244					
	4	0.0849	5.6718					

The autocorrelation function (ACF) of the data and the fitted two-state and three-state HMMs are given in Figure 3.2. It is clear from the figure that the ACF of two-state and three-state HMMs correspond well to the sample ACF.

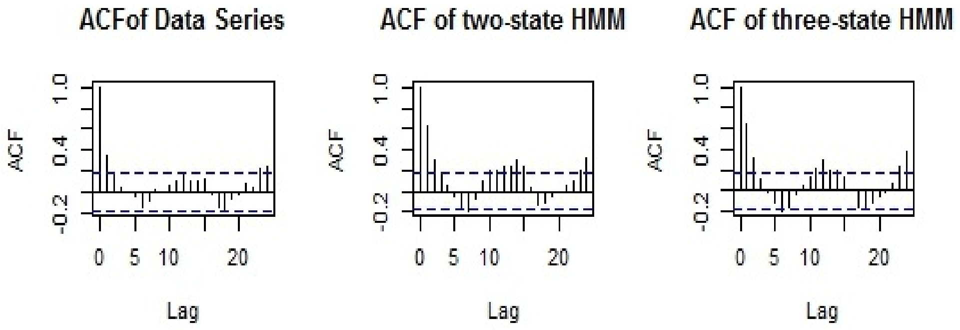

Figure 3.2: The ACF of data and fitted HMMs.

The autocorrelation of fitted two-state and three-state HMMs for $k = 1$ are computed using (5) and the values are 0.412 and 0.369 respectively. The autocorrelation of the data series for $k = 1$ is 0.345.

Prediction of the most likely sequence of Markov states given the observed data set (decoding) of two-state Poisson-HMM is done using Viterbi algorithm and is given in Table 3.4.

Table 3.4: The most likely sequence of hidden states of two-state Poisson-HMM given observed data series.

1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1
1	1	1	1	1	2	2	1	1	1	1	1	1	1	1	2	2	2	2	2	2	1	1	1
1	1	1	1	1	1	2	2	2	1	1	1	1	1	1	1	2	2	2	2	2	2	2	1

Using ***R*** the state predictions of death counts in the months of the year 2016 based on two-state Poisson-HMM is given in Table 3.5.

Table 3.5: State prediction using two-state HMM: the probability that the Markov chain will be in a given state in the specified month of 2016.

Year 2016	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
State=1	0.8838	0.8479	0.8308	0.8229	0.8198	0.8190	0.8194	0.8204	0.8217	0.8230	0.8245	0.8260
2	0.1167	0.1536	0.1723	0.1818	0.1868	0.1895	0.1909	0.1919	0.1926	0.1931	0.1935	0.1939

The actually observed deaths in the months from January 2016 to December 2016 were $1, 1, 0, 0, 2, 4, 3, 4, 1, 3, 0, 2$ respectively. The respective states of the Markov chain to which these observations assigned are $1, 1, 1, 1, 1, 2, 1, 2, 1, 1, 1, 1$; shows the consonance of the observed values with the fitted model.

4. Discussion

Standard model for unbounded counts, the Poisson distribution is inappropriate for the Dengue fever death count data considered in this study as the data is an overdispersed one. The sample autocorrelation function of the Dengue fever death count data displayed in Figure 3.2 shows a clear indication that the observations are serially dependent. As the Dengue fever death counts are serially dependent, an independent mixture of Poisson models found to be inappropriate. So we considered independent mixture models with one, two, three and four components and Poisson-hidden Markov models with two, three and four states. The independent mixture models do not perform well relative to the Poisson-HMM for the death count data. Of the seven models fitted, two-state Poisson-HMM has the lowest AIC and BIC values. Thus the best fitted model is the two-state Poisson-HMM having stationary distribution $\boldsymbol{\delta} = (0.8129, 0.1871)$ and state dependent mean vector $\boldsymbol{\lambda} = (0.5007, 3.9658)$. On studying the Viterbi path of states of two-state HMM in relation with the number of deaths it is found that state 1 corresponds to less than 3 deaths and state 2 corresponds to more than 3 deaths. If there are 3 deaths then it may be labeled as either state 1 or state 2. Analyzing the Viterbi path of hidden states of the process it is clear that the process is in state 2 when Kerala has rainy season, an environmental factor which causes Dengue fever. Normally Kerala has an average rainfall of more than 250mm during May to November. The temperature of Kerala ranges from 18^oC to 40^oC, except in hill stations, which is another favorable environmental factor that helps the growth of Dengue virus carrying mosquitoes. The probability that the Markov chain will be in a given state in the next twelve months shown in Table 3.6 is in concordance with the sample observations. The present study reveals that HMMs can be successfully used in identifying the environmental factors which influence a stochastic process. The algorithms required for numerical computations are quite simple and availability of necessary R codes make the parameter estimation easy. When we fit models with three or more states to relatively short series of observations the estimates of one or more transition probabilities turnout to be very close to zero.

References

1 Baum LE, Petrie T (1966). "Statistical inference on probabilistic functions of finite state Markov chains", *Annals of Mathematical Statistics*, **37**, 1554-1563.

2 Baum LE, Petrie T, Soules G, Weiss N (1970). "A maximization technique occurring in statistical analysis of probabilistic functions of Markov chains", *Annals of Mathematical Statistics*, **41(1)**, 164-171.

3 Cappe O, Moulines E, Ryden T (2005), *Inference in Hidden Markov Models*, Springer, New York.

4 Durbin R, Eddy S, Krogh A, Mitchison G (1998).*Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*, Cambridge University Press.

5 Guttorp P (1995). *Stochastic Modeling of Scientific Data*, Chapman and Hall, London.

6 Juang BH, Rabiner LR (1991). "Hidden Markov models for speech recognition", *Technometrics*, **33**, 251-272.

7 MacDonald IL, Zucchini W (1997). *Hidden Markov and Other Models for Discrete-Valued Time Series*, Chapman and Hall, New York.

8 Rabiner LR (1989). "A tutorial on hidden Markov models and selected applications in speech recognition", *Proceedings of the IEEE*, **77**, 257-286.

9 Viterbi AJ (1967). "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm", *IEEE Transactions on Information Theory*, **13**, 260-269.

10 Zucchini W (2000). "An introduction to model selection", *Journal of Mathematical Psychology* **44**, 41-61.

11 Zucchini W, MacDonald IL (2009). *Hidden Markov Models for time series: An Introduction Using R*, Monographs on statistics and applied probability ; 110.

12 http://dhs.kerala.gov.in

Affiliation:

Joshni George
Department of Statistics
St. Thomas College, Palai
Kerala-686574, India,
Email: (joshni.tony@gmail.com)

Journal of Environmental Statistics
Volume 8, Issue 9
September 2018

http://www.jenvstat.org
Submitted: 2018-04-18
Accepted: 2018-09-07