THINK INDIA JOURNAL

ISSN:0971-1260
Vol-22- Issue-14-December-2019

Negative Binomial Hidden Markov Model for AES count data

Joshni George [1] **and Seemon Thomas** [2]
Department of Statistics, St.Thomas College Pala, Arunapuram, Kerala - 686574, India
joshni.tony@gmail.com [1], seemonpala@rediffmail.com [2]

Abstract: In this paper, we propose negative binomial hidden Markov model for the modeling of Acute encephalitis syndrome (AES) cases reported in Kerala. The parameters of the model are estimated using EM algorithm and the sequence of hidden states are obtained using the best fitted model. The transition probabilities of the hidden markov chain are obtained.
Keywords: Hidden Markov model, Transition probability, Parameter estimation, Decoding.

1. Introduction

The problem of overdispersion and serial dependence are usually present in count data. The canonical model for count data, the Poisson distribution, is not suitable for overdispersed series of counts. The data considered in our study, the monthly counts of Acute encephalitis syndrome (AES) in Kerala from January 2012 to December 2016, is an overdispersed one. Negative binomial distribution, that allows overdispersion is also inappropriate when the data is multimodal. In this situation, hidden Markov models (HMMs) are considered as they can accommodate both overdispersion and multimodality. A problem arises when one uses HMM is that of choosing appropriate state dependent distribution. In our study, we assume the conditional distribution of AES counts given the hidden states as negative binomial.

2. Materials and methods

A Hidden Markov model (HMM) is a stochastic model with an underlying Markov process that is hidden but can be observable through another stochastic process depend on hidden states that produce a sequence of observations. The theory of HMM was introduced by Baum and his colleagues [1], [2] as an extension to the first order stochastic Markov process. In the late 1980s and early 1990s, HMMs were introduced to computational sequence analysis [3]. HMMs have been widely used in modern continuous speech recognition systems [5], biological sequence analysis [4], speech recognition [8], computational finance [7], gene prediction [6] etc. The theory, application and computation of the HMMs are described in [10].
HMMs are stochastic models in which the distribution that generates an observation depends on the state of an unobserved Markov process. A hidden Markov model consists of two stochastic processes: an observed sequence of random variable and a hidden parameter process assumed to satisfy Markov property. Let $\{C_t : t = 1,2,...\}$ represents an unobservable finite-state homogeneous Markov chain referred as hidden process and $\{Z_t : t = 1,2,...\}$ is the observable state-dependent process such that, when C_t is known, the distribution of Z_t depends only on current state C_t and not on previous states or observations. If $\mathbf{Z}^{(t)}$ and $\mathbf{C}^{(t)}$ representing the histories from time 1 to time t, then HMM $\{Z_t : t = 1,2,...\}$ is a particular kind of dependent mixture, in such a way that:

$$P(C_t \mid C^{(t-1)}) = P(C_t \mid C_{t-1}), t = 2,3,...$$
$$P(Z_t \mid Z^{(t-1)}, C^{(t)}) = P(Z_t \mid C_t), t \in \mathbb{N}.$$

Then the marginal distribution of Z_t is a finite mixture. Observations can be either discrete or continuous. The probability mass function of Z_t if the Markov Chain is in state i at time t is the following:

$$p_i(z) = P(Z_t = z \mid C_t = i).$$

Negative Binomial Hidden Markov Models

In this study, we assume that Z_t is the sequence of monthly AES counts in Kerala from January 2012 to December 2016 and the distribution of Z_t is a finite mixture of negative binomial distributions. In negative

ISSN:0971-1260
Vol-22- Issue-14-December-2019

binomial regression, the distribution is specified in terms of its mean, $\mu = \frac{(1-p)^{\alpha}}{p}$ and one can derive $p = \frac{\alpha}{\mu+\alpha}$, $1-p = \frac{\mu}{\mu+\alpha}$. Then substituting these expressions in the probability mass function,

$$p(Z=z) = \frac{\Gamma(z+\alpha)}{\Gamma(z+1)\Gamma(\alpha)}\left(\frac{\alpha}{\alpha+\mu}\right)^{\alpha}\left(\frac{\mu}{\alpha+\mu}\right)^{z},$$

for z= 0, 1, 2, ... The variance can then be written as $\mu + \frac{\mu^2}{\alpha}$.

Let us consider a stationary m -state negative binomial hidden Markov model $\{Z_t : t = 1,2,\ldots\}$ with transition probability matrix $\Gamma = (\gamma_{ij})$. Let us assume that the conditional distribution of any observation is negative binomial. Then we have the following:

$$p(z_t \mid C_t = i, \mu, \alpha) = \frac{\Gamma(z_t+\alpha_i)}{\Gamma(z_t+1)\Gamma(\alpha_i)}\left(\frac{\alpha_i}{\alpha_i+\mu_i}\right)^{\alpha_i}\left(\frac{\mu_i}{\alpha_i+\mu_i}\right)^{z_t}$$

for t=1,2,..., where $\mu = (\mu_1, \ldots, \mu_m)$ and $\alpha = (\alpha_1, \ldots, \alpha_m)$ with $\mu_i > 0$ and $\alpha_i > 0$ for any i $\in$ 1, 2, ..., m. The conditional mean and conditional variance are respectively given by

$$E(Z_t \mid C_t = i) = \mu_i$$

$$V(Z_t \mid C_t = i) = \mu_i\left(1 + \frac{\mu_i}{\alpha_i}\right).$$

The marginal mean and variance are respectively given below:

$$E(Z_t) = \sum_i \mu_i \delta_i$$

$$V(Z_t) = \sum_i \left(\mu_i + \mu_i^2 + \frac{\mu_i^2}{\alpha_i}\right) - \left(\sum_i \mu_i \delta_i\right)^2$$

where $\delta = (\delta_1 \ldots \delta_m)'$ is the stationary distribution of the Markov chain so that $\delta' = \delta'\Gamma$ holds.
The parameters are estimated using EM algorithm. A detailed description of the iterative procedure involved in EM algorithm is available in [10]. The state dependent means μ_i must be non negative for $i = 1,2,\ldots,m$ and all parameters γ_{ij} must be non negative and the rows of the transition probability matrix Γ must add to 1.
We use two standard model selection criteria namely Akaike information criterion (AIC) and Bayesian information criterion (BIC) for choosing the appropriate number of states m.

$$AIC = -2\log L + 2k$$

$$BIC = -2\log L + k \log n$$

where $logL$ is the log-likelihood of the model, k is the number of parameters and n is the total number of observations. Prediction of most likely sequence of states for each time point t by maximizing the joint probability can be done using Viterbi algorithm and the details of this algorithm can be found in [9] and in [10].

3. Results and Discussion

Modeling of AES counts

The data given in Table 3.1 is the monthly AES counts for the period January 2012 to December 2016.

Table 3.1: Monthly AES counts in Kerala from January 2012 to December 2016.

Month &Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2012	9	5	0	5	1	11	6	0	1	0	1	2

ISSN:0971-1260
Vol-22- Issue-14-December-2019

2013	6	8	5	16	6	2	3	0	0	0	4	1
2014	1	1	0	1	0	0	0	1	3	0	1	1
2015	0	3	2	2	3	7	3	0	0	1	6	2
2016	1	5	2	1	0	1	1	2	0	0	5	0

Figure 3.1 is the graphical display of the data.

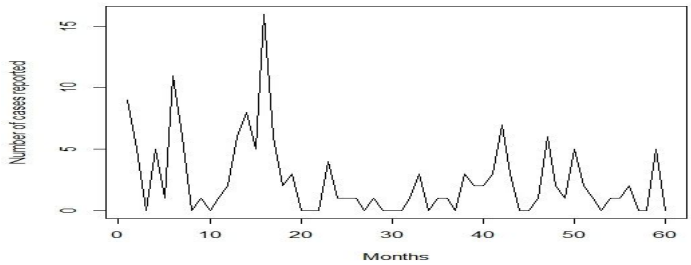

Figure 3.1: AES cases reported in Kerala in 60 consecutive months (from 2012 January-2016 December).

For the data given in Table 3.1, the mean and variance are 2.46 and 9.81 respectively so that the data is an overdispersed one. Negative binomial distribution, which allows for overdispersion is also inappropriate when the data is multimodal. 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 0.125 with a p value 2.2×10^{-16}. In this context we fit NB-HMM to the data. Fitting of a negative binomial hidden Markov model (NB-HMM) involves estimation of δ, μ, α and $\mathbf{\Gamma}$ by EM algorithm. AIC and BIC values of each fitted model are given in Table 3.2.

Table 3.2: Comparison of fitted models by AIC and BIC.

model	- log L	AIC	BIC
Negative Binomial	962.0852	1928.170	1932.359
2-state NB-HMM	120.6869	253.3738	265.9399
3-state NB-HMM	193.1755	410.351	435.4831
4-state NB-HMM	192.4790	424.9580	466.8449

On comparing AIC and BIC values one can see that 2-state NB-HMM is the best fitted model. For the fitted 2-state NB-HMM, the estimate of the transition probability matrix $\mathbf{\Gamma}$ obtained is the following:

$$\mathbf{\Gamma} = \begin{pmatrix} 0.9356 & 0.0644 \\ 0.3954 & 0.6046 \end{pmatrix}.$$

The corresponding estimates of δ, μ and σ are shown in Table 3.3.

Table 3.3: Stationary distribution and parameters of 2-state NB-HMM.

Parameter	state 1	state 2
δ	0.8599	0.1401
α	1.3184	7.6527
μ	1.1804	8.0728

The mean and variance of 2-state NB-HMM computed are 2.1459 and 9.9709 respectively. Note that these values are very close to the sample mean (2.4666) and sample variance (9.8124).

Prediction of the most likely sequence of Markov states given the observed data set (decoding) of 2-state NB-HMM is done using Viterbi algorithm and is given in Table 3.4.

Table 3.4: The most likely sequence of hidden states of 3-state Normal-HMM.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec

ISSN:0971-1260
Vol-22- Issue-14-December-2019

&Year												
2012	2	2	1	1	1	2	2	1	1	1	1	1
2013	2	2	2	2	2	1	1	1	1	1	1	1
2014	1	1	1	1	1	1	1	1	1	1	1	1
2015	1	1	1	1	1	2	1	1	1	1	2	1
2016	1	1	1	1	1	1	1	1	1	1	1	1

4. Conclusion

The best fitted model is found to be the 2-state NB-HMM having stationary distribution δ = (0.8599 0.1401), state dependent mean vector μ = (1.1804 8.0728). On studying the Viterbi path of states of 2-state HMM in relation with the disease counts, it is found that state 1 corresponds to less than 6 counts, state 2 corresponds to 6 or more counts. The present study reveals that HMMs can be effectively used to model and study the hidden factors (states) which affects the disease counts. The actually observed counts from January 2017 to December 2017 are respectively 2, 0, 0, 0, 0, 1, 2, 0, 0, 0, 0, 0.

5. References

[1] Baum, L.E.; Petrie, T. Statistical inference on probabilistic functions of finite state Markov chains, *Ann. Math.Stat* (1966), **37**, pp. 1554-1563.

[2] Baum, L.E.; Petrie, T.; Soules, G.; Weiss, N. A maximization technique occurring in statistical analysis of probabilistic functions of Markov chains, *Ann. Math. Stat* (1970), 41**(1)**, pp. 164-171.

[3] Churchill, G.A. *Stochastic Models for Heterogeneous DNA sequences*, Bull Math Biol, (1989), 51(**1**), pp. 79-94.

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

[5] Gales, M.; Young, S. The application of hidden Markov models in speech recognition, Found. Trends Signal Process., (2008), **1**(3), pp. 195-304.

[6] Munch, K.; Krogh, A. Automatic generation of gene finders for eukaryotic species, *BMC Bioinformatics*, (2006), **7**, p. 263.

[7] Petropoulos, A.; Chatzis, S.P.; Xanthopoulos, S. A Novel Corporate Credit Rating System Based on Student's-t Hidden Markov Models, *Expert Syst Appl*. (2016), **53**, pp. 87-105.

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

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

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

[11] www.dhs.gov.in

