Revista Colombiana de Estadística
January 2020, Volume 43, Issue 1, pp. 71 to 82
DOI: http://dx.doi.org/10.15446/rce.v43n1.77542

Generalized Poisson Hidden Markov Model for Overdispersed or Underdispersed Count Data

Modelo oculto de Markov de Poisson generalizado para datos de recuento sobredispersados o subdispersos

Sebastian George[a], Ambily Jose[b]

Department of Statistics, St. Thomas College, Pala, India

Abstract

The most suitable statistical method for explaining serial dependency in time series count data is that based on Hidden Markov Models (HMMs). These models assume that the observations are generated from a finite mixture of distributions governed by the principle of Markov chain (MC). Poisson-Hidden Markov Model (P-HMM) may be the most widely used method for modelling the above said situations. However, in real life scenario, this model cannot be considered as the best choice. Taking this fact into account, we, in this paper, go for Generalised Poisson Distribution (GPD) for modelling count data. This method can rectify the overdispersion and underdispersion in the Poisson model. Here, we develop Generalised Poisson Hidden Markov model (GP-HMM) by combining GPD with HMM for modelling such data. The results of the study on simulated data and an application of real data, monthly cases of Leptospirosis in the state of Kerala in South India, show good convergence properties, proving that the GP-HMM is a better method compared to P-HMM.

***Key words*:** EM algorithm; Generalized Poisson distribution; Hidden Markov Model; Overdispersion.

Resumen

El mtodo estadstico ms adecuado para explicar la dependencia serial en los datos de recuento de series de tiempo se basan en los modelos ocultos de Markov (HMM). Estos modelos suponen que las observaciones se generan a partir de un finito mezcla de distribuciones regidas por el principio de la cadena de Markov (MC). El modelo de Markov oculto de Poisson (P-HMM) puede ser el mtodo ms utilizado para modelar las situaciones mencionadas anteriormente. Sin embargo, en el escenario de la vida real, este

[a]PhD. E-mail: sthottom@gmail.com
[b]Research Scholar. E-mail: ambilystat06@gmail.com

modelo no puede considerarse como la mejor opcin. Teniendo en cuenta este hecho, nosotros, en este artculo, apostamos por la distribucin generalizada de Poisson (GPD) para modelar datos de conteo. Este mtodo puede rectificar la sobredispersin y subdispersin en el modelo de Poisson. Aqu desarrollamos Poisson generalizado Modelo de Markov oculto (GP-HMM) combinando GPD con HMM para modelando tales datos. Los resultados del estudio sobre datos simulados y una aplicacin de datos reales, casos mensuales de leptospirosis en el estado de Kerala en South India, muestra buenas propiedades de convergencia, lo que demuestra que el GP-HMM Es un mtodo mejor en comparacin con P-HMM.

***Palabras clave*:** Algoritmo EM; Distribucin generalizada de Poisson; Modelo oculto de Markov; Sobredispersin.

1. Introduction

Poisson model is the most commonly used method for modelling time series count data. Though equidispersion is the unique feature of Poisson distribution, in practical cases, either the mean will be greater than variance or vice-versa, making the Poisson assumption wrong. In many populations of Poisson nature, the probability of the occurrence of an event does not remain constant and is affected by previous occurrences, resulting in unequal mean and variance in the data (Kendall & Stuart 1963). To deal with such situations, modification and generalization of the Poisson distribution were considered by Greenwood & Yule (1920) and by Neyman (1931). Wang & Famoye (1997) introduced generalized Poisson regression for modelling household fertility decisions. Recently, Cepeda-Cuervo & Cifuentes-Amado (2017) also developed mean and dispersion regression models to fit overdispersed data based on beta binomial and negative binomial models.

An important generalization for the Poisson distribution was introduced by Consul & Jain (1973) with two parameters λ_1 and λ_2 and it can be considered as a limiting form of the generalized negative binomial distribution. Consul & Shoukri (1984) obtained maximum likelihood estimators of the parameters of GPD. The properties of the GPD are discussed by Consul (1989) and Tuenter (2000). The variance of GPD model is greater than, equal to, or less than the mean when the second parameter λ_2 is positive, zero or, negative respectively. Both the mean and variance tend to increase or decrease in value with respect to the change in λ_1. When λ_2 is positive, the mean and the variance increase in value. However, when λ_2 increases, the variance increases faster than mean which results in overdispersion or vice versa. The probability mass function of GPD is given by

$$Pr(X = x) = \frac{\lambda_1(\lambda_1 + x\lambda_2)^{x-1}}{x!} exp(-\lambda_1 - x\lambda_2), x = 0, 1, 2, \ldots \tag{1}$$

where $\lambda_1 > 0$ and $|\lambda_2| < 1$. The mean and variance of the GPD are

$$\mu = \frac{\lambda_1}{(1 - \lambda_2)}.$$

$$\sigma^2 = \frac{\lambda_1}{(1-\lambda_2)^3}.$$

The GPD is often used in researches and studies for modelling data in many situations. It can be used to adjust overdispersion in Poisson model, as in the case of negative binomial model. The GPD is also apt for modelling underdispersed Poisson data. Going by all these details, one can consider GP-HMM as a better option than P-HMM as shown in Sebastian, Jeyaseelan, Jeyaseelan, Anandan, George & Bangdiwala (2019) for count data modelling. However, the idea of using GPD in HMM is not new. In 2014, Witowski et.al made a simulation study for using HMM to improve quantifying physical activity in accelerometer data (Witowski, Foraita, Pitsiladis, Pigeot & Wirsik 2014). They used P-HMM, GP-HMM and normal-HMM for their comparative study. The following part of this paper has been categorized into four sections, detailing the methods and estimation of parameters of GP-HMM, a simulation study and a real data application of GP-HMM.

2. Generalized Poisson Hidden Markov model

In HMM, there is an underlying unobserved state of the system that changes with time in line with the Markov process. The distribution of observations at a given time is determined by the system's state at that time (Zucchini & MacDonald 2009). Let H_t, $t \in 1, 2, \ldots, T$ be an MC on a finite state space, $S = \{1, 2, \ldots, m\}$ with transition probability matrix $\boldsymbol{A} = (a_{ij})$, where $a_{ij} = Pr[H_{t+1} = j | H_t = i]$ for any i y $j \in S$ and with the initial distribution $\pi = (\pi_1, \pi_2, \ldots, \pi_m)'$, where $\pi_i = Pr[H_1 = i]$ for any $i \in S$. The MC H_t, defined on a finite state space, is homogenous and irreducible. So, the initial distribution π is stationary, which satisfies the condition $\pi' = \pi'\boldsymbol{A}$.

Let X_t, $t \in N$, an HMM, be a particular type of dependent mixture, with $X^{(t)} = (X_1, \ldots, X_t)$ and $H^{(t)} = (H_1, \ldots, H_t)$ having the following properties.

$$Pr[H_t \mid H^{(t-1)}] = Pr[H_t \mid H_{t-1}], \quad t = 2, 3, \ldots$$

$$Pr[X_t \mid X^{(t-1)}, H^{(t)}] = Pr[X_t \mid H_t], \quad t \in N.$$

The first property, 'parameter process', H_t, $t = 1, 2, \ldots$ satisfies Markov property, while the second one, 'state-dependent process', X_t, $t = 1, 2, \ldots$ such that the distribution of X_t, depends only on the current state H_t and not on previous states or observations. Now, we introduce some notations - the probability mass function (pmf) of X_t, given the Markov chain is in state i at time t, is denoted by p_i and it is called state-dependent distribution of the model.

$$p_i(x) = Pr(X_t = x \mid H_t = i) \qquad \text{for} \quad i = 1, 2, \ldots, m.$$

In the case of GPD, the pmf is given by

$$p_i(x) = \frac{\lambda_{1i}(\lambda_{1i} + x\lambda_{2i})^{x-1}}{x!} \exp(-\lambda_{1i} - x\lambda_{2i}).$$

Now, $\pi_i(t) = Pr(H_t = i)$ is the unconditional probabilities of MC being at state i at time t. Let X_t be a discrete valued random variable, such that

$$Pr(X_t = x) = \sum_{i=1}^{m} Pr(H_t = i) Pr(X_t = x \mid H_t = i) = \sum_{i=1}^{m} \pi_i(t) p_i(x).$$

This expression can be rewritten in matrix form as

$$Pr(X_t = x) = \boldsymbol{\pi}(t) \boldsymbol{P}(x) \mathbf{1}',$$

where $\boldsymbol{P}(x)$ is defined as the diagonal matrix with i^{th} diagonal element $p_i(x)$ and $1'$ is the m-dimensional vector of ones.

Now,

$$\pi(t) = \pi(1) \boldsymbol{A}^{t-1},$$

and, therefore,

$$Pr(X_t = x) = \pi(1) \boldsymbol{A}^{t-1} \boldsymbol{P}(x) \mathbf{1}'.$$

If the MC is stationary, then $\pi(1)\boldsymbol{A}^{t-1} = \pi$ and in this case,

$$Pr(X_t = x) = \pi \boldsymbol{P}(x) \mathbf{1}'.$$

In the case of a GP-HMM, the mean and variance of X_t are given by

$$E[X_t] = \sum_{i=1}^{m} \pi_i \frac{\lambda_{1i}}{(1 - \lambda_{2i})}.$$

$$V(X_t) = \sum_{i=1}^{m} \pi_i \left[\frac{\lambda_{1i}^2}{(1 - \lambda_{2i})^2} + \frac{\lambda_{1i}}{(1 - \lambda_{2i})^3} \right] - \left[\sum_{i=1}^{m} \frac{\pi_i \lambda_{1i}}{(1 - \lambda_{2i})} \right]^2.$$

So, the model $\{X_t, H_t\}$ $t = 1, 2, \ldots, T$ is characterized by (i) the stationary initial distribution π, (ii) the transition probability matrix $\boldsymbol{A}$ and (iii) the state-dependent pmfs $p_i(x)$. Let $\boldsymbol{\Phi = (\Pi, A, \Theta)}$ denote the set of parameter space. The parameters to be estimated are: the $(m^2 - m)$ transition probabilities (a_{ij}) for any $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, m - 1$; the m entries of the vector $\boldsymbol{\Pi}$ and the m parameters of λ_{1i} and λ_{2i} of the GP random variables X_t. Hence, the vector of unknown parameters is given by

$$\phi = (a_{11}, \ldots, a_{1m-1}, \ldots, a_{m1}, \ldots, a_{mm-1}, \lambda_{11}, \lambda_{12}, \ldots, \lambda_{1m}, \lambda_{21}, \lambda_{22}, \ldots, \lambda_{2m})'$$

which belongs to the parameter space $\boldsymbol{\Phi}$. The initial distribution π will be estimated by the equality $\pi' = \pi' \boldsymbol{A}$, after the estimation of the matrix $\boldsymbol{A}$. The estimators of the vector $\boldsymbol{\phi}$ will be obtained by the EM algorithm. The likelihood function is given by:

$$L_T = \pi \boldsymbol{P}(x_1) \boldsymbol{A} \boldsymbol{P}(x_2) \boldsymbol{A} \boldsymbol{P}(x_3) \cdots \boldsymbol{A} \boldsymbol{P}(x_T) \mathbf{1}'. \tag{2}$$

2.1. Joint Probability Mass Functions of the Model

In HMM $\{X_t; H_t\}$ with the sequence of observations, $x_1, x_2, \ldots, x_T$ and the sequence of states of the Markov chain $h_1, h_2, \ldots, h_T$, the joint pmf is given by

$$\begin{aligned} Pr[x_1, x_2, \ldots, x_T, h_1, h_2, \ldots, h_T] =& \pi_{h_1} a_{h_1 h_2} a_{h_2 h_3} \ldots a_{h_{T-1} h_T} p[x_1|h_1] p[x_2|h_2] \\ & \ldots p[x_t|h_t] = \pi_{h_1} p[x_1|h_1] \prod_{t=2}^{T} a_{h_{t-1} h_t} p[x_t|h_t]. \end{aligned}$$

Summing over $h_1, h_2, \ldots, h_T$, we have the joint pmf

$$Pr[x_1, x_2, \ldots, x_T] = \sum_{h_1} \sum_{h_2} \cdots \sum_{h_T} \{\pi_{h_1} p[x_1|h_1] \prod_{t=2}^{T} a_{h_{t-1} h_t} p[x_t|h_t]\}. \tag{3}$$

Define $\boldsymbol{P}(x)$ as the diagonal matrix with i^{th} diagonal element $p_i(x)$, and we have

$$Pr[x_1, x_2, \ldots, x_T] = \pi' \boldsymbol{P}(x_1) \boldsymbol{A} \boldsymbol{P}(x_2) \ldots \mathrm{AP}(x_T) \mathbf{1}'.$$

When π is stationary then it can be replaced by $\pi' \boldsymbol{A}$. Now, the joint pmf becomes

$$P[x_1, x_2, \ldots, x_T] = \pi' \boldsymbol{A} \boldsymbol{P}(x_1) \boldsymbol{A} \boldsymbol{P}(x_2) \cdots \boldsymbol{A} \boldsymbol{P}(x_T) \mathbf{1}'.$$

If π is not stationary, then the state of the Markov chain has to be estimated. So, the estimate of the initial distribution from one observation at time 1 is not useful. If it is stationary, clearly π is fully estimated by its transition probabilities (Zucchini & MacDonald 2009).

3. Estimation by EM Algorithm

A commonly used method of fitting HMMs is the EM algorithm introduced by Dempster, Laird & Rubin (1977) for the computation of maximum likelihood estimates based on incomplete data. Also Pereira, Marques & da Costa (2012) have studied the performance of the estimates produced by EM algorithm for mixture model. Here, we use this tool for the estimation of parameters of HMM with forward and backward probabilities (Baum 1972), which are also adopted for decoding and state prediction of HMM. For $t = 1, 2, \ldots, T$, the (row) vector α_t is as follows:

$$\alpha_t = \pi' \boldsymbol{P}(x_1) \boldsymbol{A} \boldsymbol{P}(x_2) \ldots \boldsymbol{A} \boldsymbol{P}(x_T) = \pi' \boldsymbol{P}(x_1) \prod_{s=2}^{T} \boldsymbol{A} \boldsymbol{P}(x_s),$$

with π denoting the initial distribution of the MC. The elements of α_t are called *forward probabilities*. For $t = 1, 2, \ldots, T$, and $j = 1, 2, \ldots, m$, we have the joint probability

$$\alpha_t(j) = Pr(X^{(t)} = x^{(t)}, H_t = j), \tag{4}$$

which is the j^{th} component of α_t. Now, the vector of backward probabilities β_t, for $t = 1, 2, \ldots, T$, is defined by

$$\beta_t = \pi' \boldsymbol{P}(x_{t+1}) \boldsymbol{A} \boldsymbol{P}(x_{t+2}) \cdots \boldsymbol{A} \boldsymbol{P}(x_T) \mathbf{1}' = \left(\prod_{s=t+1}^{T} \boldsymbol{A} \boldsymbol{P}(x_s) \right) \mathbf{1}'.$$

The j^{th} component of β_t is the conditional probability and the elements of β_t are called *backward probabilities*. For $t = 1, 2, \ldots, T-1$, and $j = 1, 2, \ldots, m$, we have the conditional probability

$$\beta_t(j) = Pr(X_{t+1}^{(T)} = x_{t+1}^{(T)} \Big| H_t = j), \tag{5}$$

where X_a^b denotes the vector $(X_a, X_{a+1}, \ldots, X_b)$. For $t = 1, 2, \ldots, T$, and $i = 1, 2, \ldots, m$, the probability

$$Pr(X^{(t)} = x^{(t)}, H_t = j) = \alpha_t(i)\beta_t(i), \tag{6}$$

and consequently $\alpha_t \beta_t' = Pr(X^{(T)} = x^{(T)}) = L_T$, for each t. In HMM, this property is used for applying the EM algorithm and in local decoding. For $t = 1, 2, \ldots, T$, firstly

$$Pr(H_t = j \mid X^{(t)} = x^{(t)}) = \frac{\alpha_t(i)\beta_t(i)}{L_T}, \tag{7}$$

and secondly,

$$Pr(H_{t-1} = j, H_t = k \mid X^{(t)} = x^{(t)}) = \frac{\alpha_{t-1}(j) a_{jk} p_k(x_t) \beta_t(k)}{L_T}. \tag{8}$$

The EM algorithm is an iterative method for performing maximum likelihood estimation when some data are missing. In this case, the complete-data log-likelihood (CDLL) –the log-likelihood of the parameters of interest θ, based on both the observed data and the missing data– is to be maximized. The algorithm is started by choosing values for the parameters $\boldsymbol{\Theta}$. Then, the following steps are repeated:

- E step: Compute the conditional expectations of those functions of the missing data that appear in the CDLL.
- M step: Maximize, with respect to Θ, CDLL with the functions of the missing data replaced by their conditional expectations.

The sequence of states $h_1, h_2, \ldots, h_T$ followed by the MC is defined by the zero-one random variables and is given by

$$u_j(t) = 1 \text{ if and only if } h_t = j \quad (t = 1, 2, \ldots, T)$$

$$v_{jk}(t) = 1 \text{ if and only if } h_{t-1} = j \text{ and } h_t = k \quad (t = 2, 3, \ldots, T).$$

Now, the log-likelihood of the observations $x_1, x_2, \ldots, x_T$ and the missing data $h_1, h_2, \ldots, h_T$ are given by

$$\begin{aligned}\log\left(Pr\left(X^{(T)}, h^{(T)}\right)\right) &= \log(\pi_{h_1}\prod_{t=2}^{T} a_{h_{t-1},h_t}\prod_{t=1}^{T} p_{h_t}(x_t)) \\ &= \log \pi_{h_1} + \sum_{t=2}^{T}\log\ (a_{h_{t-1},h_t}) + \sum_{t=1}^{T}\log\ p_{h_t}(x_t).\end{aligned}$$

Hence, the CDLL is

$$\begin{aligned}\log(Pr(X^{(T)}, h^{(T)})) = &\sum_{j=1}^{m} u_j(1)\log\ \pi_j + \sum_{j=1}^{m}\sum_{k=1}^{m}\left(\sum_{t=2}^{T} v_{jk}(t)\right)\log\ a_{jk} \\ &+ \sum_{j=1}^{m}\sum_{t=1}^{T} u_j(t)\log\ p_j(x_t)).\end{aligned}$$

Here, π is to be understood as the initial distribution of the MC (the distribution of H_1), not necessarily the stationary distribution. Of course, it is not reasonable to try to estimate the initial distribution from just one observation at time 1, especially as the state of the MC itself is not observed. The EM algorithm for HMMs proceeds as follows:

- E step: Replace the quantities $v_{jk}(t)$ and $u_j(t)$ by their conditional expectations given the observations $x^{(T)}$, then it is given by

$$\hat{u}_j(t) = Pr(H_t = j \mid x^{(T)}) = \frac{\alpha_t(j)\beta_t(j)}{L_T}$$

$$\hat{v}_{jk}(t) = Pr(H_{t-1} = j, H_t = k \mid x^{(T)}) = \frac{\alpha_{t-1}(j)a_{jk}p_k(x_t)\beta_t(k)}{L_T}.$$

- M step: Replace $v_{jk}(t)$ and $u_j(t)$ by $\hat{v}_{jk}(t)$ and $\hat{u}_j(t)$ and maximize the CDLL with respect to the three sets of parameters: the initial distribution π, the transition probability matrix $\boldsymbol{A}$ and the parameters of the state-dependent distributions $\lambda_{11}, \lambda_{12}, \ldots, \lambda_{1m}, \lambda_{21}, \lambda_{22}, \ldots, \lambda_{2m}$.

M step splits neatly into three separate maximizations: term 1 depends only on the initial distribution π, term 2 on the transition probability matrix $\boldsymbol{A}$ and term 3 on the state-dependent parameters of GPD and they are estimated using numerical maximization Zucchini & MacDonald (2009).

4. A Simulation Study

This section shows the results of a simulation study conducted on the performance of the GP-HMM model. We assessed the performance of the estimates

using the mean squared error (MSE) on the simulated data. The simulations were repeated on different sample sizes ranging from 500 to 10 000 and on different parameter values. Prior to checking the results obtained for the parameter estimates, we had run the EM algorithm with $m = 2, \ldots, 6$ states to select the appropriate number of states of the GP-HMM. For this purpose, we computed the AIC and BIC for each GP-HMM. Upon computation, both Akaike information criterion (AIC) and Bayesian information criterion (BIC) gave the lowest value for $m = 2$ state. So, we chose, two state GP-HMM for further computation. The initial values of the parameters are given below

$$a_{11} = 0.9,\ a_{12} = 0.1,\ \lambda_{11} = 10,\ \lambda_{12} = 30,\ \lambda_{21} = 0.7, \text{ and } \lambda_{22} = 0.6.$$

We had run 50 replications of the EM algorithm on the simulated data and then computed the MSE and biases (given in brackets), the results of which to shown in Table 1. It shows that MSE and biases of the estimate of the parameters tend to become zero as the sample size increases. Also, the mean estimates of the parameters come closer to the true parameter values.

5. Real Data Application

The GP-HMM model is applied to a real-life data set. A series of monthly Leptospirosis incidence counts in the state of Kerala in South India during the 2006-2017 period is considered for the analysis. The data are sourced from the official website of the Directorate of Health Services, Government of Kerala, India. A graphical representation of the said data, with 144 time points, is shown in figure 1. As many as 14,460 cases, having the mean value 100.42 and variance 3,342.41, are subjected to study. In this case, the sample variance is greater than its mean, which shows the data is clearly overdispersed. As per the data, February 2014 recorded the minimum Leptospirosis cases of 20, while the maximum of 292 was reported in September 2008. two models, HMM with Poisson distribution and GP-HMM, were estimated and compared on the basis of -LogL, AIC and BIC. Table 2 presents the relevant comparison. In this, table k represents the number of parameters. For P-HMM $k = m^2$ and GP-HMM $k = m(m + 1)$. The model is considered to be the most apt one, which can be identified using BIC. So the model with two-state GP-HMM is the most appropriate.

Table 1: MSE(Bias) of the GP-HMM parameters for 50 simulation runs of the EM algorithm with $m = 2$ states.

N	500	1000	5000	10000
a_{11}	0.0020(0.0374)	0.0005(0.0148)	$5.674e^{-05}(-0.0027)$	$2.5e^{-05}(0.0005)$
a_{12}	0.0020(−0.0374)	0.0005(−0.0148)	$5.674e^{-05}(0.0027)$	$2.5e^{-05}(0.0005)$
λ_{11}	0.2884(0.0118)	0.1547(−0.1520)	0.1588(−0.3633)	0.0102(0.0428)
λ_{12}	0.0003(−1.1530)	0.0006(1.6355)	0.0002(−0.7830)	$1.9e^{-05}(-0.1919)$
λ_{21}	0.0005(−0.0080)	0.0042(0.0003)	0.0001(0.0090)	$1.4e^{-05}(0.0013)$
λ_{22}	0.0003(0.0058)	0.0006(−0.0195)	0.0002(0.0125)	$1.9^{-05}(0.0019)$

TABLE 2: comparison of Poisson and GP HMMs by AIC and BIC.

Model	k	-LogL	AIC	BIC
Poisson-Hidden Markov Model				
Simple Poisson	1	2696.54	5395.08	5398.05
2-state P-HMM	4	1298.16	2610.31	2631.10
3-state P-HMM	9	889.58	1813.15	1863.64
4-state P-HMM	16	806.29	**1674.57**	**1766.64**
5-state P-HMM	25	802.92	1703.84	1849.34
Generalized Poisson-Hidden Markov Model				
Simple GP	2	768.76	1541.51	1547.45
2-state GP-HMM	6	741.06	1496.12	**1516.92**
3-state GP-HMM	12	727.18	**1482.36**	1523.93
4-state GP-HMM	20	721.87	1489.74	1558.05

Figure 1 shows the result of fitting four state P-HMM and two state GP-HMM to the leptospirosis series using EM estimates. The four-state model is fitted to the leptospirosis series by using HMM with Poisson distribution, while the two-state model is fitted to the data using GPD as state dependent distribution. However, on close analysis of the results, the GP-HMM model seems to fit the data better. The estimated transition probability matrices for four-state P-HMM and two-state GP-HMM are

$$\boldsymbol{A}^{P-HMM} = \begin{bmatrix} 0.7244 & 0.2756 & 0.0000 & 0.0000 \\ 0.2230 & 0.5452 & 0.1982 & 0.0337 \\ 0.0479 & 0.5435 & 0.2796 & 0.1290 \\ 0.0000 & 0.0000 & 0.6022 & 0.3978 \end{bmatrix}$$

and

$$\boldsymbol{A}^{GP-HMM} = \begin{bmatrix} 0.8335 & 0.1665 \\ 0.1322 & 0.8678 \end{bmatrix}$$

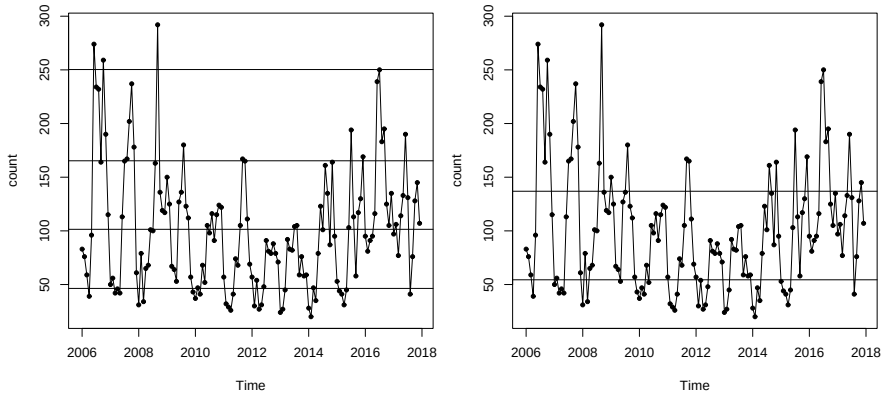

FIGURE 1: The state-dependent means (horizontal lines), left: 4 state P-HMM and right: 2 state GP-HMM, compared to the observations.

The Poisson parameters estimated by the HMM are

$$\lambda = (\ 46.3694 \quad 101.4253 \quad 165.2471 \quad 250.2855\ )$$

and the GP parameters are $\lambda_1 = (\ 18.9565 \quad 32.6556\ )$ and $\lambda_2 = (\ 0.6512 \quad 0.7615\ )$. The whole analysis is done using R software. The package $HMMpa$ is used for modelling Witowski & Foraita (2013).

6. Discussion and Conclusion

We propose to deal with overdispersion and underdispersion in time series of count data by introducing GPD in HMM. Here, the is EM algorithm used for the estimation of the parameters, by implementing the R-package (Witowski & Foraita 2013). Also, the validity of the estimates is verified by carrying out a simulation study. We take out an original data –monthly count of cases Leptospirosis in Kerala between 2006 and 2017– to show that GP-HMM performs much better than P-HMM. Leptospirosis, according to the World Health Organization, is a bacterial disease detected in places which witness excess rainfall and flooding. It transmits to humans through the cuts on the skin or through the mucous membranes of the eyes, nose and mouth with water contaminated with the urine of infected animals. In this study, the occurrences of the bacterial disease are modelled using the HMM because of the dependent nature of the leptospirosis data. The P-HMM is the most widely used method for modelling such data. In this case, we develop GP-HMM for fitting leptospirosis series and compare it with P-HMM using AIC and BIC. From all these findings, we prove that GP-HMM is much better compared to P-HMM.

For modelling overdispersed count data, a suitable model is a mixture of Poisson model. It can be seen that the negative binomial distribution (NBD) and GPD are in fact mixtures of Poisson models, where the mixing distributions are continuous distributions. The mixing distribution in the case of NBD is a gamma distribution and is suitable for modelling data with excess zeroes (Joe & Zhu 2005). However, GPD is better suited when more mass is concentrated on the tail of the distribution compared to NBD. Further, also GPD is suitable for both over-dispersed and under-dispersed data. This might be the reason for getting better fit on GP-HMM for the data in our example. When excess zero and heavy tail are present, we may use a zero-inflated GP-HMM for modelling serially correlated data. We will concentrate our attention on this aspect in our further study. The results based on our simulation study and example will help both theoreticians and practitioners for making inference on unequal mean-variance scenario in the case of serially correlated time series count data.

Acknowledgements

This paper is a part of the Ph.D. programme of the second author, under the Mahatma Gandhi University, Kerala, India. The authors are grateful to the

editors and referees for sending valuable feedback and comments, which helped us improve the manuscript.

[Received: January 2019 — Accepted: December 2019]

References

Baum, L. E. (1972), 'An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes', *Inequalities* **3**, 1–8.

Cepeda-Cuervo, E. & Cifuentes-Amado, M. V. (2017), 'Double Generalized Beta-Binomial and Negative Binomial Regression Models', *Revista Colombiana de Estadística* **40**(1), 141–163.

Consul, P. C. (1989), *Generalized Poisson Distributions: Properties and Applications*, Dekker, New York.

Consul, P. C. & Jain, G. C. (1973), 'A Generalization of Poisson Distribution', *Technometrics* **15**(4), 791–799.

Consul, P. C. & Shoukri, M. M. (1984), 'Maximum likelihood estimation for the generalized Poisson distribution', *Communication in Statistics - Theory and Methods* **13**(12), 1533–1547.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977), 'Maximum Likelihood from Incomplete Data via the EM Algorithm', *Journal of the Royal Statistical Society, Serie B* **39**(1), 1–38.

Greenwood, M. G. & Yule, G. U. (1920), 'An inquiry into the nature of frequency distributions representative of multiple happenings, with particular reference to the occurrence of multiple attacks of disease or of repeated accidence', *Journal Royal Statistical Society* **83**, 255–279.

Joe, H. & Zhu, R. (2005), 'Generalized Poisson Distribution: the Property of Mixture of Poisson and Comparison with Negative Binomial Distribution', *Biometrical Journal* **47**(2), 219–229.

Kendall, M. & Stuart, A. (1963), *The Advanced Theory of Statistics*, Vol. 1, Hafner Publishing Co., New York.

Neyman, J. (1931), 'On a new class of contagious distributions, applicable in entomology and bacteriology', *Technometrics* **10**, 35–57.

Pereira, J. R., Marques, L. A. & da Costa, J. M. (2012), 'An Empirical Comparison of EM Initialization Methods and Model Choice Criteria for Mixtures of Skew-Normal Distributions', *Revista Colombiana de Estadística* **35**(3), 457–478.

Sebastian, T., Jeyaseelan, V., Jeyaseelan, L., Anandan, S., George, S. & Bangdiwala, S. (2019), 'Decoding and modelling of time series count data using Poisson hidden Markov model and Markov ordinal logistic regression models', *Statistical Methods in Medical Research* **28**(5), 1552–1563.

Tuenter, H. J. H. (2000), 'On the generalized Poisson distribution', *Statistica Neerlandica* **54**, 374–376.

Wang, W. & Famoye, F. (1997), 'Modelling household fertility decisions with generalized Poisson regression', *Journal of Population Economics* **10**, 273–283.

Witowski, V. & Foraita, R. (2013), *HMMpa: Analysing accelerometer data using hidden markov models*, R package version 1.0.1.
*https://cran.r-project.org/package=HMMpa

Witowski, V., Foraita, R., Pitsiladis, Y., Pigeot, I. & Wirsik, N. (2014), 'Using hidden Markov models to improve quantifying physical activity in accelerometer data - A simulation study', *PLOS ONE* **9**(12), 77–92.

Zucchini, W. & MacDonald, I. L. (2009), *Hidden Markov Models for Time Series: An Introduction Using R*, Chapman and Hall, Boca Raton.