Malaysian Journal of Mathematical Sciences 15(3): 447–456 (2021)

Malaysian Journal of Mathematical Sciences

Journal homepage: https://einspem.upm.edu.my/journal

A New Discrete-Time Series Model for Overdispersion

Khoo, W. C.*[1] and Ong, S. H.[2]

[1]*School of Mathematical Sciences, Department of Applied Statistics, Sunway University, Malaysia*
[2]*Department of Actuarial Science and Applied Statistics, UCSI University, Malaysia*

E-mail: wooichenk@sunway.edu.my
**Corresponding author*

Received: 26 Dec 2020
Accepted: 22 June 2021

Abstract

The classical Poisson first-order integer-valued autoregressive (INAR(1)) is a popular discrete time series model. Due to equality of mean and variance, it is inappropriate for data showing overdispersion (variance exceeding mean). Often, negative binomial innovation is used to tackle this case. The main contribution of this paper is to consider a few innovation processes in a mixture time series model, i.e. zero-inflated Poisson, geometric, negative binomial and new geometric distributions, which are well-known for overdispersion. The Expectation-Maximization (EM) algorithm is applied to estimate the parameters of the mixture model. An illustration with a real-life example is presented to show that the innovation processes are viable for discrete-time series analysis.

Keywords: Binomial thinning; mixture; overdispersion; zero-inflated; discrete-time series.

1 Introduction

Count data often exhibit over dispersion in time series modelling. The negative binomial (NB) innovation is widely used for such type of data when the Poisson model is no longer applicable. One of the factors for this over dispersion is the excessive zero counts in the data. For instance, the number of flight crashes contains very high zero count due to the good safety records of the aviation industry. Since the INAR(1) with NB innovations may not be able account for very high zeros, this paper considers a mixture model with innovation terms such as zero-inflated Poisson, geometric, negative binomial and new geometric innovations to cater for over dispersion and excess zeros in the modelling of time series of counts.

The authors in [3] introduced the conventional continuous autoregressive (AR) time series model defined by

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_n X_{t-n}.$$

This recursive model is well-known for its simple interpretation and attractive properties. However, it cannot accommodate count data due to the multiplication operator that yields the non-integer values. The binomial thinning operator replaces the scalar multiplication in order to retain the integer characteristic.

Modelling a sequence of correlated discrete-time random variables has been of growing interest recently. Over the years, researchers have been developing discrete-time series models to handle count data. To the best of our knowledge, [11] was the pioneer who introduced integer-valued first-order Autoregressive (INAR (1)) process for count data. [1] discussed some important findings in parameter estimation methods and the model properties. The binomial thinning operation is self-decomposable [18] and the definition is given as follows.

Definition 1.1. Let X be a nonnegative discrete random variable. The binomial thinning operator is defined by

$$\alpha \circ X = \sum_{i=1}^{n} \beta_i(\alpha), \tag{1}$$

where $\beta_i(\alpha)$ are independent and identically distributed Bernoulli random variables with probability $P\big(\beta_i(\alpha) = 1\big) = 1 - P\big(\beta_i(\alpha) = 0\big) = \alpha$. The said definition is valid with $\alpha \in [0, 1]$.

Definition 1.2. Let X_t be discrete-time dependent integer-valued time series follows an INAR (1) process, if

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t, t = 0, 1, 2, \dots. \tag{2}$$

where ε_t is the innovation term having mean μ_ε and finite variance σ_ε^2.

The time series in Definition 1.2 is stationary. Equation (2) comprises of two processes at time t; the first process defines the number of survivors at time $t-1$ abbreviated by X_{t-1} with each survival possibility of α, and the new element entered the system as innovation sequence ε_t during interval $(t-1, t]$. An interpretation in the context of the sex offence data is as follows. Assuming that at the particular year t, the number of criminal is represented by X_t, and ε_t is the new criminal count, then the number of criminal generated at year $t-1$ will be denoted as X_{t-1} with the generated probability of α.

Model (2) with a simple interpretation has many real-life applications. See, for example, a recent paper of [17]. An alternative discrete-time series model was proposed by [2], based on the mixture operator $(*)$ defined by [14]. The model is defined as follows.

Definition 1.3. Let X_t be a non-negative discrete-valued stochastic process such that

$$X_t = \Big(\phi_1, I\Big[X_{t-1}\Big]\Big) * \Big(1-\phi_1, \varepsilon_t\Big). \tag{3}$$

The operator $*$ is known as the Pegram's operator and (3) denotes a mixture discrete distribution, where $I[Y_{t-1}]$ is the indicator variable, and $\varepsilon_t \sim D(p_0, p_1, \dots)$ is a discrete random variable with the mixing weights of $\phi_1 \in (0,1)$ and $1-\phi_1$, respectively. The conditional probability mass function (pmf) has the form

$$P(X_t = j \mid X_{t-1}) = (1-\phi_1)p_j + \phi_1 I[X_{t-1} = j],$$

for every $t \in 0, \pm 1, \pm 2, \dots$. The innovation term ε_t can be any discrete random variables. This model is simple but has flexibility for modelling count data.

Section 2 presents some existing works in time series modeling with overdispersion. We aim to consider various overdispersed processess to fit in a mixture time series model based upon the mixture of both operators in Definitions 1.2 and 1.3, namely, mixture of Pegram and Thinning (MPT) processes. This mixture model is extensively discussed in [8]. Here, we consider zero-inflated Poisson, geometric, negative binomial and new geometric as the innovations for the mixture discrete-time series model. The theoretical definition will be presented in Section 3. Section 4 considers parameter estimation by Expectation-Maximization (EM) algorithm. Section 5 illustrates the model with the sex offence data. Section 6 concludes.

2 Literature Review

The Poisson INAR(1) process is the benchmark model in discrete time series modelling [4, 5]. For the past few decades, the model has received much attention. However, it is only appropriate for equidispersed data when over dispersion predominates in real life situations. There is an extensive study on models based on binomial thinning to handle overdispersed data [19]. Generalizations of the thinning operation are given in [19], but may lead to certain level of difficulty in analyzing the real data sets as the model is not tractable.

It is known that geometric and negative binomial distributions are usually used to accommodate overdispersed data. In the 1980s, [12] suggested the negative binomial INAR(1) model which is an analogue of the gamma distribution, which is the continuous counterpart. This model gives a complex expression for the innovation process. By considering the geometric distribution as the innovation process, some important studies have been done by [13], [6] and [15]. Subsequently, [20] and [16] considered the innovation process with the negative binomial distribution. Recently, [10, 9] discussed the zero-inflated count time series modelling. There is limited work that deals with overdispersed counts due to excess zero counts. This motivates us to propose a mixture model with zero-inflated Poisson and other potential distributions as innovation processes.

3 Model Construction with Overdispersion

The model construction is based upon the mixture of operators given in Definitions 1.2 and 1.3. This model possesses a simple interpretation and is able to handle multimodality which is often observed in data. Here, more model flexibility to handle overdispersed data is the aim for this paper. The mixture of Pegram and thinning (MPT(1)) model of [8] is given here for readers' convenience. The model construction and some simulation studies have been discussed earlier. This paper focuses on the fitting of various innovation terms in the MPT(1) model.

The MPT(1) model mixes the survival $\alpha \circ X_{t-1}$ and the new elements ε_t with mixture operator $(*)$, and the mixing parameters are ϕ and $1-\phi_1$ respectively. It is defined as follows.

Definition 3.1. (Mixture of Pegram and INAR(1)) Let X_t be a discrete random variable with the initial distribution $P(X_0 = i) = p_0$, then for every $t \in 0, \pm 1, \pm 2, \dots$,

$$X_t = (\phi, \alpha \circ X_{t-1}) * (1-\phi, \varepsilon_t). \tag{4}$$

The model is valid with $\phi \in (0,1)$. Eq. (4) is denoted as MPT (mixture of Pegram and thinning). The conditional probability function gives

$$P(X_t = i \mid X_{t-1} = j) = \phi \binom{j}{i} \alpha^i (1-\alpha)^{j-i} + (1-\phi) P(\varepsilon_t = i). \tag{5}$$

Now, we consider X_t be a Poisson process. The Poisson fitting is given here to ease the construction of Definition 3.1.1. Next section provides the fitting of various innovation marginal in the mixture time series process in Equation (5).

$$P(X_t = i \mid X_{t-1} = j) = \phi \binom{j}{i} \alpha^i (1-\alpha)^{j-i} + \frac{e^{-\lambda}\lambda^i}{i!} - \phi \frac{e^{-\lambda\alpha}(\lambda\alpha)^i}{i!}, i = 0, 1, \dots. \tag{6}$$

3.1 Zero-Inflated Poisson Distribution

The zero-inflated Poisson is a well-known distribution to model excessive zero-valued observations. The first-order zero-inflated Poisson MPT (ZIPMPT(1)) model employs two processes. The first process is a binary random variable that generate structural zeros, and the second process generates counts by Poisson MPT(1) process. The ZIPMPT(1) process is defined as follows.

Definition 3.1.1. (Zero-inflated Poisson MPT(1) Process) Let Z_t be a zero-inflated MPT(1) process which is

$$Z_t = \begin{cases} 0 & \text{with probability } 1-\tau, \\ X_t & \text{with probability } \tau, \end{cases} \tag{7}$$

where $\tau \in (0,1)$, and X_t is given in Equation (6). Therefore, the marginal probability function of Equation (7) is

$$P(Z_t = i) = (1-\tau) I_{\{0\}}(i) + \tau P(X_t = i),$$

where $I(\cdot)$ is the indicator random variable and $P(X_t = i) = \frac{e^{-\lambda}\lambda^i}{i!}$. The expected value of Z_t is $E[Z_t] = (1-\tau)P(Z_t = 0) + \tau\lambda$ and $Var[Z_t] = (1-\tau)^2[P(Z_t = 0)(1 - P(Z_t = 0))] + \tau^2\lambda$. Theorem 3.1.1 presents the conditional distribution of Z_t given Z_{t-1} as follows:

Theorem 3.1.1. (Conditional distributions of ZIPMPT(1) Process) Given Z_t is the process as defined in Definition 3.1.1, the conditional distributions are

$$P(Z_t = i \mid Z_{t-1} = j) = \begin{cases} (1-\tau) + \dfrac{\tau e^{-\lambda}}{1-\tau+\tau e^{-\lambda}}\{(1-\tau) \\ \quad + \tau[\phi(1-e^{-\alpha\lambda}) + e^{-\lambda}]\} & \text{if } i = j = 0, \\ (1-\tau) + \dfrac{j!}{\lambda^j}\Big\{\tau\Big[\phi\Big(\dfrac{\alpha}{1-\alpha}\Big)^j \\ \quad + \dfrac{e^{-\lambda}\lambda^j}{j!} - \phi\dfrac{e^{-\alpha\lambda}(\alpha\lambda)^j}{j!}\Big]\Big\} & \text{if } i = 0\,,\, j \neq 0, \\ \dfrac{\tau\frac{e^{-\lambda}\lambda^i}{i!}}{1-\tau+\tau e^{-\lambda}}\{(1-\tau) \\ \quad + \tau[\phi(1-\alpha)^i + e^{-\lambda} - \phi e^{-\lambda\alpha}]\} & \text{if } i \neq 0\,,\, j = 0, \\ \dfrac{j!}{i!}\lambda^{i-j}\Big\{\tau\Big[\phi\dbinom{i}{j}\alpha^j(1-\alpha)^{i-j} \\ \quad + \dfrac{e^{-\lambda}\lambda^j}{j!} - \phi\dfrac{e^{-\alpha\lambda}(\alpha\lambda)^j}{j!}\Big]\Big\} & \text{if } i \neq 0, j \neq 0. \end{cases} \tag{8}$$

Theorem 3.1.1 is important for real data analysis. It can be used to understand the pattern of the data. Next, some other marginal distributions have been fitted in MPT(1) process. Here, we consider also other distributions for overdispersion; for example geometric, negative binomial and new geometric marginal distributions. The models are given in the following sections.

3.2 Geometric Distribution

Definition 3.2.1. (Geometric marginal) Let X_t be a process of geometric marginal with parameter p, then the pmf of ε_t is

$$P(\varepsilon_t = i) = \frac{1}{1-\phi}[p(1-p)^i - \phi\alpha p(1-p)^i], \tag{9}$$

where the parameters $0 < p < 1$ and $0 < \phi < 1$. The marginal distribution must fulfill the condition $\alpha < \frac{1}{\phi}$ to ensure a valid innovation process. Consequently, the respective mean and the variance are

$$\mu_\varepsilon = \left(\frac{1-\alpha\phi}{1-\phi}\right)p,$$

and

$$\sigma_\varepsilon^2 = \frac{1}{1-\phi}[p^2 + p(1+p)] - \frac{\phi}{1-\phi}[\alpha p + (\alpha p)^2] - \left[\left(\frac{1-\alpha\phi}{1-\phi}\right)p\right]^2.$$

Then, the transition probabilities are

$$P(X_t = i \mid X_{t-1} = j) = \phi \binom{j}{i} \alpha^i (1-\alpha)^{j-i} + [p(1-p)^i - \phi\alpha p(1-p)^i], i = 0, 1, \ldots. \tag{10}$$

3.3 Negative Binomial Distribution

Definition 3.3.1. (Negative Binomial marginal) Given that X_t is negative binomial with the parameters of k and p, where $k > 0$, $p > 0$, then the pmf of ε_t is

$$P(\varepsilon_t = i) = \frac{1}{1-\phi}\left[\binom{k+x-1}{k-1} p^k (1-p)^i - \phi \binom{k+x-1}{k-1} (\alpha p)^k (1-\alpha p)^i\right].$$

The mean and the variance is

$$\mu_\varepsilon = \left(\frac{1-\alpha\phi}{1-\phi}\right) kp,$$

and

$$\sigma_\varepsilon^2 = \frac{1}{1-\phi}[(kp)^2 + kp(1+p)] - \frac{\phi}{1-\phi}[\alpha kp + (\alpha kp)^2] - \left[\left(\frac{1-\alpha\phi}{1-\phi}\right) kp\right]^2,$$

respectively. Subsequently, the conditional distribution is

$$P(X_t = i \mid X_{t-1} = j) = \phi \binom{j}{i} \alpha^i (1-\alpha)^{j-i} + \left[\binom{k+x-1}{k-1} p^k (1-p)^i - \phi \binom{k+x-1}{k-1} (\alpha p)^k (1-\alpha p)^i\right], i = 0, 1, \ldots. \tag{11}$$

3.4 New Geometric Distribution

Definition 3.4.1. (New Geometric marginal) Given that X_t is a new geometric process with $\left(1, \frac{p}{1+p}\right)$ and B_i in Definition 1.1, then the conditional pmf is

$$P(X_t = i | X_{t-1} = j) = \phi \binom{j}{i} \alpha^i (1-\alpha)^{j-i} + (1-\phi) P(\varepsilon_t = i), i = 0, 1, \ldots,$$

where ε_t is given by

$$P(\varepsilon_t = 0) = \frac{1}{1-\phi}\left[\frac{1}{1+p} - \phi \frac{1+\alpha}{1+\alpha(1+p)}\right],$$

$$P(\varepsilon_t = i) = \frac{1}{1-\phi}\left\{\frac{p^i}{(1+p)^i} - \alpha^i \phi p \left[\frac{(1+p)^{i-1}}{[1+\alpha(1+p)]^{i+1}}\right]\right\}, \tag{12}$$

for $i = 1, 2, 3, \ldots$. The respective mean and the variance can be easily obtained as follows:

$$\mu_\varepsilon = \left(\frac{1-\alpha\phi}{1-\phi}\right) p,$$

and
$$\sigma_\varepsilon^2 = \frac{2p^2}{1-\phi} - \frac{\phi}{1-\phi}\alpha^2 p(1+p) - \frac{(1-\alpha\phi)p}{1-\phi}\left(1 - \frac{(1-\alpha\phi)p}{1-\phi}\right).$$

4 Maximum Likelihood Estimation (MLE)

This section considers the EM algorithm for finite mixture distributions. [7, Section 2] implemented the EM algorithm which is described as follows. Let X be the finite mixture process with pmf $g(x)$ which is composed of the components of $f(x \mid \theta)$ with respective parameters of $\phi_j > 0$, and $\theta_j \in (0,1)$. The mixture distribution is valid with the sum of the weights equals to one. The pmf of X is
$$g(x) = \sum_{j=1}^{k} \phi_j f(x|\theta_j). \tag{13}$$

The EM algorithm is given as follows:

E-step: Given the current estimates p_j^{old} and $\mu(\theta_j^{old})$, calculate
$$w_{ij} = \frac{p_j^{old} f(x_i|\theta_j^{old})}{g(x_i)}, i = 1, \ldots, n, j = 1, \ldots, k. \tag{14}$$

M-step: New parameter estimates $\mu(\theta_j^{new})$ and p_j^{new} are obtained from
$$\mu(\theta_j^{new}) = \frac{\sum_{i=1}^{n} w_{ij} x_i}{\sum_{i=1}^{n} w_{ij}}, \tag{15}$$

and
$$p_j^{new} = \frac{\sum_{i=1}^{n} w_{ij}}{n}, j = 1, \ldots, k. \tag{16}$$
The iteration is terminated when convergence is reached within a margin of error of 0.001. Moment estimates are suggested as initial values for MLE.

5 Real Data Application

We consider a time series count of sex offences to illustrate the application. The one-month data were reported by the 21st police car beat in Pittsburgh. The period of the data starts from January 1990 until December 2001. The total observations are 144. One can find the data in the crime section of the forecasting principle website. Figure 1 displays the sex offence counts.

The partial autocorrelation indicates that the first order of MPT model is reasonable. The mean is 0.5903 and the variance is 1.0268, which yields the index of data dispersion as 1.7395. The

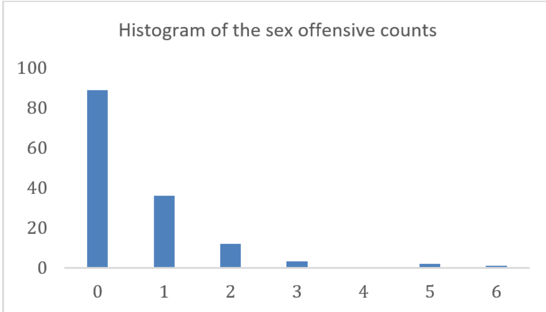

Figure 1: Sex offensive counts.

autocorrelation value is 0.2348. We are interested in comparing the fitting of MPT(1) with all the marginals as discussed in Section 3.

For each model we derive the MLE via EM algorithm. Here, the pmf $g(x)$ is considered with the respective conditional probability function in Section 3. The summary is shown in Table 1. In addition, we compared two models adopted from [15], and use the results therein. The superiority of the model is justified by Akaike Information Criterion (AIC) and we have tabulated the results in Table 1. Note that the lowest AIC value of 243.19 is from the new geometric MPT(1) model. Also, it can be observed that the zero-inflated Poisson appears as the competitive candidate to describe the excessive zero count data.

Table 1: Summary of the bus routes.

Marginal Distribution	Conditional Probability Function	Parameter Estimates	AIC
Zero-inflated Poisson	Equation 8	$\hat{\lambda} = 0.5749$, $\hat{\alpha} = 0.7936$, $\hat{\phi} = 0.1154$, $\hat{\tau} = 0.9618$	245.73
Geometric	Equation 10	$\hat{p} = 0.6288$, $\hat{\alpha} = 0.9388$, $\hat{\phi} = 0.0848$	304.70
Negative Binomial	Equation 11	$\hat{k} = 0.9741$, $\hat{p} = 0.6451$, $\hat{\alpha} = 0.9394$, $\hat{\phi} = 0.0807$	307.05
New Geometric	Equation 12	$\hat{p} = 0.6288$, $\hat{\alpha} = 0.9388$, $\hat{\phi} = 0.3304$	243.19
INAR with negative binomial	[15, Table 2]	$\hat{r} = 1.1167$, $\hat{p} = 0.9388$, $\hat{\alpha} = 0.1113$	305.67
INAR with new geometric	[15, Table 2]	$\hat{\mu} = 0.5872$, $\hat{\alpha} = 0.1650$	303.74

6 Concluding Remarks

Fitting the real data with an appropriate model is a concern in the field of data analysis. Reviews showed that there is relatively less work on the integer-valued time series models to deal with overdispersed data. This paper considers an integer time series model MPT(1) with overdispersion, in particular, those well-known marginal are considered in the model fitting. The zero-inflated Poisson, geometric, negative binomial and new geometric innovations have been compared and the results show that geometric MPT(1) outperformed the counterparts. The zero-inflated Poisson INAR(1) is also a viable process to handle the overdispersed data.

Acknowledgement The authors wish to thank the reviewers for their useful suggestions. The authors are supported by Sunway Individual Research Grant GRTIN-IRG-122-2021 and the Ministry of Higher Education grant FRGS/1/2020/STG06/SYUC/02/1.

Conflicts of Interest The authors declare no conflict of interest.

References

[1] M. A. Al-Osh & A. A. Alzaid (1987). First-order integer-valued autoregressive. *Journal of Time Series Analysis*, *8*(3), 261–275.

[2] A. Biswas & P. X. K. Song (2009). Discrete-valued ARMA processes. *Statistics and Probability Letter*, *79*(17), 1884–1889.

[3] G. E. P. Box & G. M. Jenkins (1976). *Time series analysis: Forecasting and control*. Holden-Day, University of Michigan.

[4] R. K. Freeland (1998). *Statistical Analysis of Discrete Time Series with Applications to The Analysis of Workers Compensation Claims Data*. PhD thesis, The University of British Columbia, Vancouver, Canada.

[5] R. K. Freeland & B. P. M. McCabe (2004). Analysis of low count time series data by Poisson autoregression. *Journal of Time Series Analysis*, *25*(5), 701–722.

[6] M. A. Jazi, G. Jones & C. D. Lai (2012). First-order integer-valued AR processes with zero inflated Poisson innovations. *Journal of Time Series Analysis*, *33*(6), 954–963.

[7] D. Karlis & E. Xekalaki (1999). Improving the EM algorithm for mixtures. *Statistics and Computing*, *9*, 303–307.

[8] W. C. Khoo, S. H. Ong & A. Biswas (2017). Modeling time series of counts with a new class of INAR(1) model. *Statistical Papers*, *58*(2), 393–416.

[9] R. Maiti, A. Biswas & S. Das (2015). Time series of zero-inflated counts and their coherent forecasting. *Journal of Forecasting*, *34*(8), 694–707.

[10] R. Maiti, A. Biswas, A. Guha & S. H. Ong (2014). Modelling and coherent forecasting of zero-inflated count time series. *Statistical Modelling*, *14*(5), 375–398.

[11] E. McKenzie (1985). Some simple models for discrete variate time series. *Water Resour Bull*, *21*(4), 645–650.

[12] E. McKenzie (1987). Innovation distributions for Gamma and negative binomial autoregressions. *Scandivanian Journal of Statistics*, *14*(1), 79–85.

[13] A. S. Nastić & M. M. Ristić (2012). Some geometric mixed integer-valued autoregressive (INAR) models. *Statistics and Probability Letters*, *82*(4), 805–811.

[14] G. G. S. Pegram (1980). An autoregressive model for multilag Markov chain. *Journal of Applied Probability*, *17*(2), 350–362.

[15] M. M. Ristić, H. S. Bakouch & A. S. Nastić (2009). A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *Journal of Statistical Planning and Inference*, *139*(7), 2218–2226.

[16] M. M. Ristić, A. S. Nastić & H. S. Bakouch (2012). Estimation in an integer-valued autoregressive process with negative binomial marginals (NBINAR(1)). *Communications in Statistics - Theory and Methods*, *41*(4), 606–618.

[17] M. G. Scotto, C. H. Weiss & S. Gouveia (2015). Thinning-based models in the analysis of integer-valued time series: A review. *Statistical modelling*, *15*(6), 590–618.

[18] F. W. Steutel & K. Van Harn (1979). Discrete analogues of self-decomposability and stability. *Annals of Probability*, *7*(5), 893–899.

[19] C. H. Weiβ (2008). Thinning operations for modeling time series of counts - A survey. *AStA Advances in Statistical Analysis*, *92*(3), 319–341.

[20] R. Zhu & H. Joe (2010). Negative binomial time series models based on expectation thinning operators. *Journal of Statistical Planning and Inference*, *140*, 1874–1888.