Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2025 (43)** : 1–12.
ISSN-0037-8712
doi:10.5269/bspm.68835

An Integer-Valued AR(1) Process with Poisson-Modified Lindley Distributed Innovation and Modelling Practical Time Series Data

Veena G[1]*, Lishamol Tomy[2], Hassan S. Bakouch[3,4], Christophe Chesneau[5]

ABSTRACT: In this paper, a first-order, non-negative, integer-valued autoregressive model with Poisson-modified Lindley distributed innovation is developed. In contrast to the standard Poisson model, it can accommodate count time series data with over-dispersion. Statistical properties of the proposed model, including the mean, variance, conditional mean, conditional variance, and multi-step forecast conditional measures, are studied. To estimate the parameters involved, the conditional maximum likelihood estimation is used. The model's utility is demonstrated using two real-world data sets: the number of times a software is downloaded and the number of sudden death submissions of animals in a certain hospital in New Zealand.

Key Words: Binomial thinning operator, INAR(1) process, Lindley distribution, conditional maximum likelihood estimation, compounding distributions.

Contents

1. Introduction

The most common approach to interpreting time-dependent data is by conducting time series analysis. Time-dependent data of an integer nature can include the count of day-to-day stock market transactions, the counts of hurricanes on a perennial basis, the number of rainy days in the ensuing weeks, the head count of patients administered every day in an emergency service, and the daily counts of COVID-19 cases. This kind of data is therefore highly valued in many fields, including economics, medicine, and other real-world anomalies.

In general practice, a Gaussian distribution is used to model time series data. In some cases, however, this distribution cannot be used due to the inherent discreteness of the data. To overcome this problem, marginal distributions that are discrete in nature have been recommended for use in time series models. Integer-valued autoregressive processes with thinning operators were developed to this aim. The first-order integer-valued autoregressive (INAR(1)) model was proposed by [12] and [2], which is based on the binomial thinning operator.

* Corresponding author.
2010 *Mathematics Subject Classification:* 62M10.
Submitted July 06, 2023. Published December 13, 2025

Typeset by $\mathcal{B}^{S}\mathcal{P}_{\mathcal{M}}$ style.
© Soc. Paran. de Mat.

Some mathematical foundations of such a model are now recalled. Consider a positive integer-valued random process, (Y_t), and a constant $\beta \in (0,1)$. Based on it, the binomial thinning operator, denoted by $' \circ '$, is indicated as

$$\beta \circ Y_{t-1} = \sum_{i=1}^{Y_{t-1}} \eta_i, \tag{1.1}$$

where (η_i) is a sequence of independent and identically distributed (iid) Bernoulli random variables with parameter β, that is

$$\eta_i = \begin{cases} 1 \text{ with probability } \beta, \\ 0 \text{ with probability } 1-\beta. \end{cases}$$

For the standard properties of the binomial thinning operator, we redirect the readers to [19]. The INAR(1) model associated with (Y_t) is defined as follows:

$$Y_t = \beta \circ Y_{t-1} + \epsilon_t, \quad t = 0, 1, 2, \ldots \tag{1.2}$$

where (ϵ_t) is a sequence of iid discrete random variables, with mean μ_ϵ and finite variance, σ_ϵ^2. Here, for a given t, ϵ_t is independent of Y_{t-q} for $q \geq 1$. According to [6], the INAR(1) process is a homogeneous Markov chain with a one-step transition probability given by

$$\Pr\left(Y_t = k | Y_{t-1} = l\right) = \sum_{i=0}^{\min(k,l)} \binom{l}{i} \beta^i \left(1-\beta\right)^{l-i} \Pr(\epsilon_t = k-i),$$

where $k, l \geq 0$. They also shown that the INAR(1) model has a mean and variance indicated as, respectively,

$$\mu_Y = \frac{\mu_\epsilon}{1-\beta}, \quad \sigma_Y^2 = \frac{\sigma_\epsilon^2 + \beta\mu_\epsilon}{1-\beta^2},$$

where μ_ϵ and σ_ϵ^2 are the mean and variance of ϵ_t, respectively. The Poisson distribution was initially assumed to be the distribution of the innovations in the INAR(1) model. One limitation of this distribution is that its mean and variance are equivalent, known as "equi-dispersion". Real-world data sets may not invariably be equi-dispersed in nature; they can be under-dispersed or over-dispersed. As the Poisson model could not model data exhibiting over-dispersion and under-dispersion, the need for alternative models having more flexible innovations was met. As a result, the statistical literature has subsequently produced many models that capitalise on the shortcomings of the Poisson model.

This paper aims to address the weakness of Poisson models in dealing with over-dispersion and under-dispersion of time series data. In recent years, distinct models with more than one parameter in the distribution of innovation have been developed. The most recent ones include the zero-and-one inflated Poisson-Lindley INAR(1) model developed by [14], the INAR(1)-Poisson transmuted exponential model studied by [4], and the INAR(1) model with innovation based on the Bilal distribution studied by [5]. This paper utilizes a one parameter compound Poisson distribution for the innovation, named the Poisson-modified Lindley (P-ML) distribution established by [8]. The P-ML distribution is unimodal in nature, and the corresponding distribution assigns a larger probability to lower values of the variable, which is identical to the Poisson-Lindley distribution of [11]. It is useful for modelling uncommon events and has tractable moment measures. When the authors in [8] applied it to real-world biological and reliability data sets, the model outperformed the Poisson, Poisson Lindley, and Poisson Bilal distributions in terms of fit.

As a result, the P-ML distribution has the potential to be useful for modelling over-dispersed count data. Accordingly, we utilize it for modeling count time series and propose an INAR(1) model with P-ML innovations (INAR(1) P-ML), which can account for over-dispersion in an INAR(1) scheme. To accentuate the benefits of the proposed INAR(1) model, we compare it to other competitive INAR(1) models, such as the P-INAR(1) model studied by [12] and the Bell-INAR(1) model developed by [10].

Classical information criteria are used,, including the Akaike information criterion (AIC), as seen in [1], and the Bayesian information criterion (BIC), as studied in [16]. When comparing the outcomes of distinct information criteria, it can be seen that the INAR(1) P-ML model is a strong competitor for analyzing over-dispersed integer-valued time series data.

The rest of this work is organized in the following manner: The P-ML distribution, with its definition and other features, is succinctly discussed in Section 2. In Section 3, the P-ML-INAR(1) model is introduced, and its imperative features are developed, in conjunction with the conditional mean and variance. Section 4 gives the k-step-ahead conditional mean and varaiance for this model. The conditional maximum likelihood (CML) estimation method is used to estimate the model parameters in Section 5. Section 6 compares the developed framework to the previous two INAR(1)-type models by modeling two real-world data sets, demonstrating the proposed model's competitive dominance. The paper is concluded in Section 7.

2. A Retrospective on the P-ML Distribution

The P-ML distribution introduced by [8] is a compounding mixture of the conventional Poisson distribution and the single parameter modified Lindley distribution. A precise definition is given below.

Definition 1. *A discrete random variable Z is said to have the P-ML distribution with parameter $\theta > 0$ if it has the following probability mass function:*

$$\Pr(Z = m) = \frac{\theta[(2\theta+1)^{m+2} + (2\theta m - 1)(\theta+1)^m]}{(\theta+1)^{m+1}(2\theta+1)^{m+2}}, \quad m \in \mathbb{N}, \tag{2.1}$$

where $\mathbb{N}$ represents the sets of the natural numbers. In this case, one may write $Z \sim P-ML(\theta)$.

Thus, we can show that Equation (2.1) is a Poisson mixture of the modified Lindley distribution developed by [8]. The mean, variance and the probability generating function of the P-ML distribution are, respectively, given by

$$\mu = \mathrm{E}(Z) = \frac{4\theta+5}{4\theta(\theta+1)}, \tag{2.2}$$

$$\sigma^2 = \mathrm{V}(Z) = \frac{(4\theta+5)(2\theta+1)(2\theta+3)}{16\theta^2(\theta+1)^2} = \frac{\mu(2\theta+1)(2\theta+3)}{4\theta(\theta+1)} \tag{2.3}$$

and

$$G(s) = \mathrm{E}(e^{sZ}) = \frac{\theta}{\theta+1-s} + \frac{(s-1)\theta}{(1+\theta)(2\theta+1-s)^2}$$

for $s < \theta + 1$. Here, it is worth noting that the dispersion index (DI) is given by

$$DI = \frac{\mathrm{V}(Z)}{\mathrm{E}(Z)} = \frac{(2\theta+1)\,(2\theta+3)}{4\theta\,(\theta+1)}.$$

Since $DI > 1$, the P-ML distribution is over-dispersed; this advocates that the P-ML distribution may be pertinent for count data with over-dispersion in some scenarios. The applicability of the P-ML distribution over the Poisson distribution, Poisson Lindley distribution studied by [18] and Poisson Bilal distribution examined by [3], is shown using some real-world data sets, see [8].

Let the innovation ϵ_t in Equation (1.2) follow the P-ML distribution with parameter θ. As a result, Equation (1.2) is converted into the INAR(1) P-ML model. For the analysis of over-dispersed count time series, the model is developed.

3. The INAR(1) P-ML Process: Definition and Properties

In this part, we define the INAR(1) P-ML process and deduce its essential statistical features.

Definition 2. *Assume that Z_t has the following expression:*

$$\begin{cases} Z_t = \beta \circ Z_{t-1} + \epsilon_t, t \geq 1, \\ \epsilon_t \sim P - ML(\theta), \end{cases} \tag{3.1}$$

where (ϵ_t) is a sequence of iid P-ML random variables given by Equation (2.1), independent of η_i (counting series defined by Equation (1.1)), $\beta \in (0,1)$ and $\theta > 0$, and Z_{t-k} for $k \geq 1$. Thus defined, Z_t is called a INAR(1) P-ML process.

It is observed from Equations (2.2) and (2.3) that, the mean and variance of (ϵ_t) are finite, then based on the study of [9], the process Z_t defined by Equation (3.1) is an ergodic stationary Markov chain with

$$\begin{aligned} \delta_{lk} &= \Pr\left(Z_t = k | Z_{t-1} = l\right) = \Pr\left(\beta \circ Z_{t-1} + \epsilon_t = k | Z_{t-1} = l\right) \\ &= \sum_{i=0}^{\min(k,l)} \Pr\left(\beta \circ Z_{t-1} = i | Z_{t-1} = l\right) \Pr\left(\epsilon_t = k - i\right). \end{aligned}$$

Hence, we have

$$\delta_{lk} = \sum_{i=0}^{\min(k,l)} \binom{l}{i} \beta^i (1-\beta)^{l-i} \frac{\theta[(2\theta+1)^{k-i+2} + (2\theta(k-i)-1)(\theta+1)^{k-i}]}{(\theta+1)^{k-i+1}(2\theta+1)^{k-i+2}}, \tag{3.2}$$

where $l, k = 0, 1, \ldots$. In fact, the INAR(1) P-ML process is a Markov process. It represents the one-step transition probabilities of the process from state l to state k. With the transition probabilities in Equation (3.2), the marginal probability function of Z_t of the INAR(1) P-ML process is obtained as

$$\begin{aligned} \delta_k &= \Pr\left(Z_t = k\right) = \sum_{l=0}^{\infty} \Pr\left(Z_t = k | Z_{t-1} = l\right) \Pr(Z_{t-1} = l) \\ &= \sum_{l=0}^{\infty} \sum_{i=0}^{\min(k,l)} \Pr\left(\beta \circ Z_{t-1} = i | Z_{t-1} = l\right) \Pr\left(\epsilon_t = k - i\right). \end{aligned} \tag{3.3}$$

In the following lemma, the mean, variance, conditional mean and conditional variance of the INAR(1) P-ML process are obtained.

Lemma 3.1. *Let (ϵ_t) be an iid sequence of random variables with the P-ML distribution and parameter θ, $\beta \in (0,1)$ and Z_t represent a stationary INAR(1) P-ML process. Then, the corresponding unconditional mean and unconditional variance are, respectively,*

$$\mathrm{E}(Z_t) = \frac{4\theta + 5}{4\theta\left(\theta + 1\right)\left(1 - \beta\right)},$$

and

$$\mathrm{V}(Z_t) = \frac{4\theta + 5}{\left(1 - \beta^2\right) 4\theta(\theta + 1)} \left[\beta + \frac{(2\theta+1)(2\theta+3)}{4\theta(\theta+1)}\right].$$

Proof. Let the innovations in Equation (2.1) follow a P-ML distribution with parameter θ, so that the mean and variance of ϵ_t are as mentioned in Equations (2.2) and (2.3), respectively. Following that, the conditional mean and variance of Z_t are given by

$$\begin{aligned} \mathrm{E}\left(Z_t | Z_{t-1}\right) &= \mathrm{E}((\beta \circ Z_{t-1} + \epsilon_t) | Z_{t-1}) \\ &= \mathrm{E}\left(\beta \circ Z_{t-1} | Z_{t-1}\right) + \mathrm{E}\left(\epsilon_t | Z_{t-1}\right) \\ &= \beta Z_{t-1} + \frac{4\theta + 5}{4\theta(\theta+1)} \end{aligned}$$

and

$$\begin{aligned}\mathrm{V}\left(Z_t|Z_{t-1}\right) &= \mathrm{V}((\beta \circ Z_{t-1} + \epsilon_t)\,|Z_{t-1}) \\ &= \mathrm{V}\left(\beta \circ Z_{t-1}|Z_{t-1}\right) + \mathrm{V}\left(\epsilon_t|Z_{t-1}\right) \\ &= \beta\left(1-\beta\right)Z_{t-1} + \frac{(4\theta+5)(2\theta+1)(2\theta+3)}{16\theta^2(\theta+1)^2},\end{aligned}$$

respectively. The marginal mean and variance of Z_t, assuming stationarity, are determined below. We have

$$\begin{aligned}\mathrm{E}(Z_t) = \mathrm{E}(\mathrm{E}\left(Z_t|Z_{t-1}\right)) &= \mathrm{E}\left(\beta Z_{t-1} + \frac{4\theta+5}{4\theta(\theta+1)}\right) = \beta\mathrm{E}(Z_{t-1}) + \frac{4\theta+5}{4\theta(\theta+1)} \\ &= \beta\mathrm{E}(Z_t) + \frac{4\theta+5}{4\theta(\theta+1)} = \frac{4\theta+5}{4\theta\left(\theta+1\right)\left(1-\beta\right)}\end{aligned}$$

and

$$\begin{aligned}\mathrm{V}(Z_t) &= \mathrm{E}(\mathrm{V}(Z_t|Z_{t-1}) + \mathrm{V}(\mathrm{E}(Z_t|Z_{t-1}))) \\ &= \mathrm{E}\left(\beta\left(1-\beta\right)Z_{t-1} + \frac{(4\theta+5)(2\theta+1)(2\theta+3)}{16\theta^2(\theta+1)^2}\right) + \mathrm{V}\left(\beta Z_{t-1} + \frac{4\theta+5}{4\theta(\theta+1)}\right) \\ &= \beta\left(1-\beta\right)\mathrm{E}(Z_{t-1}) + \frac{(4\theta+5)(2\theta+1)(2\theta+3)}{16\theta^2(\theta+1)^2(1-\beta^2)} + \beta^2\mathrm{V}(Z_{t-1}),\end{aligned}$$

which implies that

$$\left(1-\beta^2\right)\mathrm{V}(Z_t) = \beta\left[\frac{4\theta+5}{4\theta(\theta+1)}\right] + \frac{4\theta+5}{4\theta(\theta+1)}\left(\frac{(2\theta+1)(2\theta+3)}{4\theta(\theta+1)}\right)$$

and we get

$$\mathrm{V}(Z_t) = \frac{4\theta+5}{\left(1-\beta^2\right)4\theta(\theta+1)}\left[\beta + \frac{(2\theta+1)(2\theta+3)}{4\theta(\theta+1)}\right].$$

This ends the proof. □

4. Features of k^{th} Step Ahead Forecast

In this Section, we present the k^{th} step ahead forecast of Z_{t+k}, with Z_t being a stationary INAR(1) P-ML process.

Proposition 4.1. *Assume that Z_t is a stationary INAR(1) P-ML process. Then, the k^{th} step ahead forecast, Z_{t+k} with the minimum mean square error (MMSE) has the following features:*

$$\mathrm{E}(Z_{t+k}|Z_t) = \beta^k Z_t + \frac{4\theta+5}{4\theta(\theta+1)}\sum_{j=0}^{k-1}\beta^j = \beta^k Z_t + \frac{4\theta+5}{4\theta(\theta+1)}\left(\frac{1-\beta^k}{1-\beta}\right)$$

and

$$\mathrm{V}(Z_{t+k}|Z_t) = \beta^k\left(1-\beta^k\right)Z_t + \frac{4\theta+5}{4\theta(\theta+1)}\left(\frac{1-\beta^k}{1-\beta}\right) + \frac{4\theta+5}{4\theta(\theta+1)}\left(\frac{(2\theta+1)(2\theta+3)}{4\theta(\theta+1)}\right)\left(\frac{1-\beta^{2k}}{1-\beta^2}\right).$$

The following limit results hold:

$$\begin{aligned}\lim_{k\to\infty}\mathrm{E}(Z_{t+k}|Z_{t-1}) &= \mathrm{E}(Z_t) \\ \lim_{k\to\infty}\mathrm{V}(Z_{t+k}|Z_{t-1}) &= \mathrm{V}(Z_t).\end{aligned}$$

Proof. The k^{th} step ahead forecast equation is given as follows:

$$Z_{t+k} = \beta \circ Z_{t+k-1} + \epsilon_{t+k},$$

which can be written recursively as

$$Z_{t+k} \stackrel{d}{=} \beta^k \circ Z_t + \sum_{j=0}^{k-1} \beta^j \epsilon_{t+k-j}.$$

When it comes to a INAR(1) P-ML process, $\beta^k \circ Z_{t+k-1}$ follows a binomial distribution with the parameters Z_t and β^k, while ϵ_{t+k-j} follows a P-ML distribution with the parameter θ. In accordance with the MMSE definition, the k^{th} step ahead forecast of Z_t is the conditional expectation of Z_{t+k} given Z_t. As a result, the conditional mean of k^{th} step ahead forecast function is

$$\begin{aligned}
\mathrm{E}(Z_{t+k}|Z_t) &= \mathrm{E}\left[\left(\beta^k \circ Z_t + \sum_{j=0}^{k-1} \beta^j \circ \epsilon_{t+k-j}\right)|Z_t\right] = \mathrm{E}\big(\beta^k \circ Z_t|Z_t\big) + \mathrm{E}\bigg(\sum_{j=0}^{k-1} \beta^j \circ \epsilon_{t+k-j}|Z_t\bigg) \\
&= \beta^k Z_t + \frac{4\theta+5}{4\theta(\theta+1)} \sum_{j=0}^{k-1} \beta^j = \beta^k Z_t + \frac{4\theta+5}{4\theta(\theta+1)} \left(\frac{1-\beta^k}{1-\beta}\right).
\end{aligned}$$

Now, taking the limit, we get

$$\lim_{k\to\infty} \mathrm{E}(Z_{t+k}|Z_t) = \lim_{k\to\infty} \left[\beta^k Z_t + \frac{4\theta+5}{4\theta(\theta+1)} \left(\frac{1-\beta^k}{1-\beta}\right)\right] = \frac{4\theta+5}{4\theta(\theta+1)\,(1-\beta)} = \mathrm{E}(Z_t).$$

On the other hand, the corresponding condition variance of k^{th} step ahead forecast function is as follows:

$$\begin{aligned}
\mathrm{V}(Z_{t+k}|Z_t) &= \mathrm{V}\left[\left(\beta^k \circ Z_t + \sum_{j=0}^{k-1} \beta^j \circ \epsilon_{t+k-j}\right)|Z_t\right] = \mathrm{V}(\beta^k \circ Z_t|Z_t) + \mathrm{V}\left(\sum_{j=0}^{k-1} \beta^j \circ \epsilon_{t+k-j}|Z_t\right) \\
&= \beta^k \left(1-\beta^k\right) Z_t + \sum_{j=0}^{k-1} \left[\beta^j(1-\beta^j)\frac{4\theta+5}{4\theta(\theta+1)}\right] + \sum_{j=0}^{k-1} \beta^{2j} \left[\frac{4\theta+5}{4\theta(\theta+1)}\left(1 + \frac{(2\theta+1)(2\theta+3)}{4\theta(\theta+1)}\right)\right] \\
&= \beta^k \left(1-\beta^k\right) Z_t + \frac{4\theta+5}{4\theta(\theta+1)} \left(\sum_{j=0}^{k-1} \beta^j + \left(\frac{(2\theta+1)(2\theta+3)}{4\theta(\theta+1)}\right) \sum_{j=0}^{k-1} \beta^{2j}\right) \\
&= \beta^k \left(1-\beta^k\right) Z_t + \frac{4\theta+5}{4\theta(\theta+1)} \left(\frac{1-\beta^k}{1-\beta}\right) + \frac{4\theta+5}{4\theta(\theta+1)} \left(\frac{(2\theta+1)(2\theta+3)}{4\theta(\theta+1)}\right) \left(\frac{1-\beta^{2k}}{1-\beta^2}\right).
\end{aligned}$$

Hence, by taking the limits we get

$$\begin{aligned}
\lim_{k\to\infty} \mathrm{V}(Z_{t+k}|Z_t) = \lim_{k\to\infty} \Bigg[&\beta^k \left(1-\beta^k\right) Z_t + \frac{4\theta+5}{4\theta(\theta+1)} \left(\frac{1-\beta^k}{1-\beta}\right) + \frac{4\theta+5}{4\theta(\theta+1)} \times \\
&\left(\frac{(2\theta+1)(2\theta+3)}{4\theta(\theta+1)}\right) \left(\frac{1-\beta^{2k}}{1-\beta^2}\right)\Bigg] \\
= \frac{4\theta+5}{(1-\beta^2)\,4\theta(\theta+1)} &\left[\beta + \frac{(2\theta+1)(2\theta+3)}{4\theta(\theta+1)}\right] = \mathrm{V}(Z_t).
\end{aligned}$$

This ends the proof. □

5. Parameter Estimation

In general, the explicit values of parameters β and θ are unknown; subsequently, one must estimate the value of the parameter vector $\gamma = (\beta, \theta)$. Three estimation approaches are widely used to estimate the

unknown parameter of the INAR(1) process, which include conditional least squares (CLS), Yule-Walker (YW), and conditional maximum likelihood (CML) methods. Previous studies on the INAR(1) process by [11] and [7] have shown that the CML estimation approach surpasses the YW and CLS approaches in terms of estimated biases and mean square errors. Simulation tests revealed that the CML estimation approach has been proven to outperform the other two estimating approaches for both large and small sample sizes. Considering these aspects, we adopt the CML technique for estimating the unknown parameters of the INAR(1) P-ML process. A conditional log-likelihood function is used in the INAR(1) P-ML process.

5.1. CML estimation

Let $Z_1, Z_2, \ldots, Z_n$ be realization from the INAR(1) P-ML process as defined in Equation (3.3), and $\gamma = (\beta, \theta)$ with $0 < \beta < 1$ and $\theta > 0$, denotes the parameter vector. The conditional probability distribution of Z_t given z_{t-1} for the INAR(1) P-ML model according to Equation (3.2) is a convolution of the binomial distribution $B(z_{t-1}, \beta)$, and the P-ML distribution with parameter θ. The conditional likelihood (L) function is given by

$$L(\gamma) = \prod_{t=1}^{n} \Pr(Z_{t+1} = z_{t+1} | Z_t = z_t) = \prod_{t=1}^{n} \left[\sum_{i=0}^{\min(z_{t+1}, z_t)} \binom{z_t}{i} \beta^i (1-\beta)^{z_t - i} \Pr(\epsilon_t = z_{t+1} - i) \right],$$

$$= \prod_{t=1}^{n} \left[\sum_{i=0}^{\min(z_{t+1}, z_t)} \binom{z_t}{i} \beta^i (1-\beta)^{z_t - i} \frac{\theta[(2\theta+1)^{k-i+2} + (2\theta(k-i)-1)(\theta+1)^{k-i}]}{(\theta+1)^{k-i+1}(2\theta+1)^{k-i+2}} \right]$$

and the log-likelihood ($\log L(\gamma)$) function is given by

$$\log L(\gamma) = \sum_{t=1}^{n} \log \left[\sum_{i=0}^{\min(z_{t+1}, z_t)} \binom{z_t}{i} \beta^i (1-\beta)^{z_t - i} \frac{\theta[(2\theta+1)^{k-i+2} + (2\theta(k-i)-1)(\theta+1)^{k-i}]}{(\theta+1)^{k-i+1}(2\theta+1)^{k-i+2}} \right]. \quad (5.1)$$

The formal formulations of the CML estimate of the parameter vector of the INAR(1) P-ML process, say, $\hat{\gamma} = (\hat{\beta}, \hat{\theta})$, are not available. As a result, using statistical tools such as R, Equation (5.1) must be maximized. To minimize the negative of the conditional log-likelihood function, we use the R software's optim function. Note that the studies presented by [7] showed that the CML estimates are asymptotically normal and consistent under the standard regularity conditions.

6. Real-world Data Analysis

This section is devoted to the application of the INAR(1) P-ML process to concrete data analysis scenarios.

6.1. Competitors

In this section, we evaluate the INAR(1) P-ML model against two competitive INAR models, the Poisson INAR(1) model and the Bell INAR(1) model. The transition probabilities of the Poisson INAR(1) model and the Bell INAR(1) model are, respectively, defined as

$$\delta_{kj} = \sum_{m=0}^{\min(k,j)} \binom{k}{m} \beta^m (1-\beta)^{k-m} \frac{e^{-\theta} \theta^{k-m}}{(k-m)!}$$

and

$$\delta_{ji} = \sum_{m=0}^{\min(j,i)} \binom{j}{m} \beta^m (1-\beta)^{j-m} \frac{\theta^{i-m} e^{-e^{\theta}+1} B_{i-m}}{(j-m)!},$$

where B_{i-m} is the Bell number, defined by $B_{i-m} = \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^{i-m}}{k!}$.

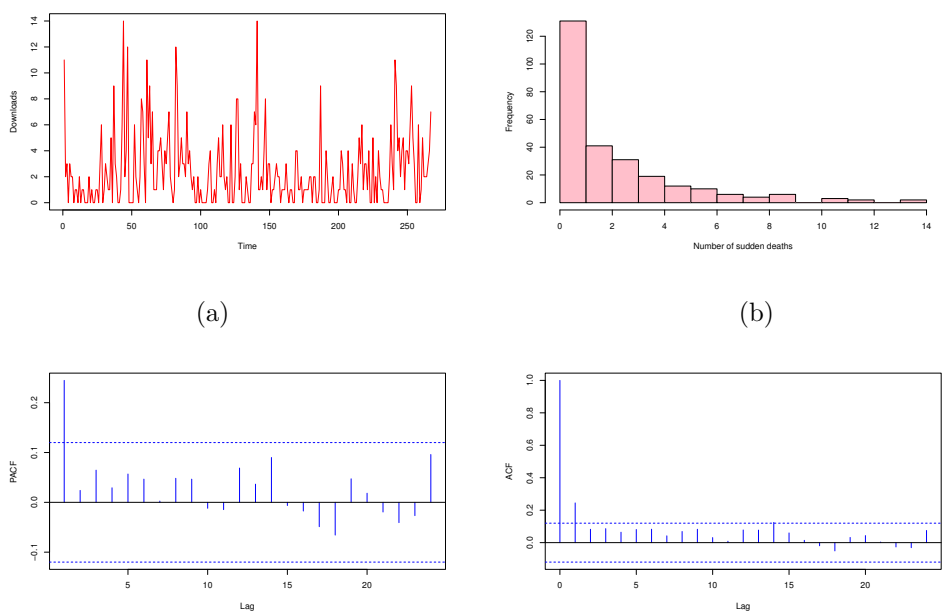
6.2. Methodology

- Initially, basic statistical measures of the data-sets are obtained.
- The over-dispersion test developed by [17] is used to identify whether the data is statistically over-dispersed or not.
- For the visualization of data, we include the line plot, histogram, partial autocorrelation function (PACF) and autocorrelation function (ACF) of the data sets.
- Finally, goodness-of-fit measures, including AIC and BIC, are used to evaluate the model against its counterparts.

For comparison purposes, we consider two real-world data sets, namely the downloads data and animal health care data, to check the significance of the INAR(1) P-ML model against its counterpart models. The descriptions of both data sets are outlined in the next two sections.

6.3. Downloads Data

We consider time series data representing the number of downloads of a TEX editor on a daily basis during the period from June 2006 to February 2007, with the number of observations, $n = 267$. This data set was previously used by [20]. The data clearly display serial dependencies, and the assumption of stationarity is not contradicted. The basic statistical measures, such as the mean and variance of the data, are estimated as 2.40075 and 7.53429, respectively. The DI of the data is obtained as 3.1383. In order to validate if the equi-dispersed Poisson INAR(1) model can be used in this situation, we use the over-dispersion test. This test, developed by [17], states that if the DI exceeds the critical value, then the data appears to be over-dispersed and the Poisson INAR(1) model is not appropriate to use. For the downloads data, the test gives a critical value (CV) as 1.144988 and hence, the Poisson INAR(1) model cannot be used here.
Figure 1 illustrates a visual presentation of the number of times the software was downloaded using a line plot, histogram, ACF, and PACF.



(a) (b)

(c) (d)

Figure 1: (a) Lineplot, (b) Histogram, (c) PACF and (d) ACF of downloads of the downloads data set

It can be observed from these illustrations that a stationary first-order autoregressive model is an appropriate choice for this data. As the method by which the data is generated behind the download counts is unknown, the dependence of the first-order can be explained in such a way that:

- There might have been users who downloaded the software, due to the usefulness of the software, and hence recommended them to their friends and colleagues, or
- Users may have installed the same software on different workstations, or
- Users could have downloaded the software on their own interest.

The parameters of the models are estimated using the CML method and also including the scenario of serial dependence. Comparing the values of the information criteria, AIC and BIC, is one of the selection criteria for the appropriate model of the downloads data. [15] states that the significant difference between the BIC of the models should be greater than 2, and the INAR(1) P-ML model justifies this when compared to the other models. As a result, we can see that the proposed model is better than the alternatives.

Table 1: Estimated parameters, along with the AIC and BIC values of the downloads data

Model	β	θ	AIC	BIC
Poisson INAR(1)	0.174	1.990	1292.8485	1300.0230
Bell INAR(1)	0.116	0.880	1124.1961	1131.3706
INAR(1) P-ML	0.1006	0.544	1100.348	1107.522

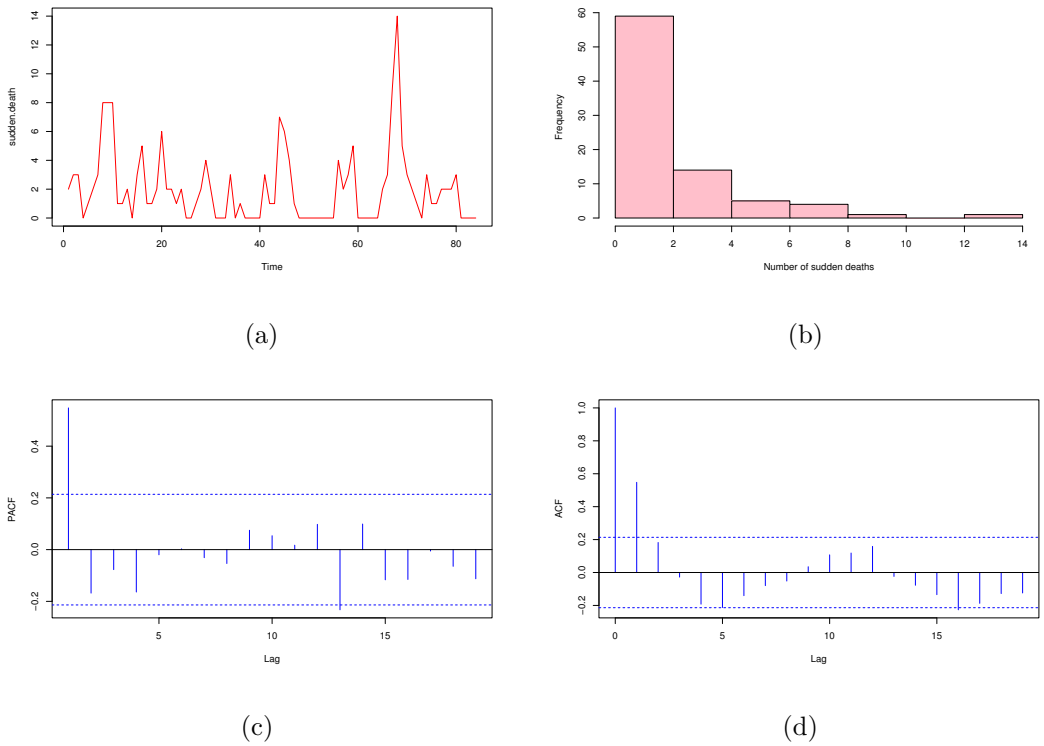
From Table 1, we can see that the INAR(1) P-ML model has the minimum values for the AIC and BIC, compared to the other INAR models. It can thus be considered as the best model.

6.4. Animal Health Data

The second data set consists of the numbers of submissions to animal health laboratories, on a monthly basis during the period 2003-2009, from a region in New Zealand. The submissions are categorized in various ways. Here, we consider one series, which gives the total number of sudden death cases by developing symptoms.

The mean and variance of the data are obtained as 0.959 and 0.044, respectively. The DI for the data is 3.187. We can say that the data is over-dispersed, and the Poisson INAR(1) model is not recommended to model this data set, as the CV is 1.305 which is lesser compared to the DI.

Figure 2 illustrates a visual representation of the number of sudden death cases admitted through a line plot, histogram, ACF, and PACF.



(a) (b)

(c) (d)

Figure 2: (a) lineplot, (b) histogram, (c) PACF and (d) ACF of the animal health data

The PACF and ACF suggest a first-order dependence. We would expect some positive correlation in such a series because the underlying processes causing disease will change smoothly over time. The parameters of the models are estimated using the CML method, and also including the scenario of serial dependence.

Table 2: Estimated parameters, along with the AIC and BIC values of the animal health data

Data	Model	β	θ	AIC	BIC
Sudden death	Poisson INAR(1)	0.382	1.240	347.4463	352.308
	Bell INAR(1)	0.328	0.678	317.466	322.327
	INAR(1) P-ML	0.309	0.824	313.415	318.277

From Table 2, we can see that the INAR(1) P-ML model has the lowest values for the AIC and BIC, compared to the other INAR models.

By modelling both the data sets, we can observe that the integer-valued autoregressive P-ML model of order one wins the goodness-of-fit comparisons against other INAR(1) models.

7. Conclusion

In this paper, the Poisson-modified Lindley distribution is used to define a stationary first-order integer-valued autoregressive model. One of the most significant advantages of this INAR model is the over-dispersion feature. This model outperforms the Poisson INAR model, which does not exhibit over-dispersion. The one-step transition probability, basic statistical measures, along with conditional

expectation, variance, and the autocorrelation function, are obtained. The unknown parameters are estimated using the conditional maximum likelihood approach. The estimated model is then applied to two real-world count time series datasets. These sets include the downloads of a particular software by the users and the data associated with the number of deaths of animals with symptoms in New Zealand animal health laboratories. It is demonstrated the effectiveness of the proposed INAR model on analyzing count data based on certain goodness-of-fit statistics, such as the Akaike and Bayesian information criteria, respectively. It is proven that the model is the best fit among the studied INAR(1) models. Finally, we anticipate that it can be applied to count time series data with a larger range of applications. In the future, we intend to use our model to analyze the integer-valued moving average (INMA), with additional estimation methods and predictions.

References

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *In Proceedings of the Second International Symposium on Information Theory*, 267–281.

[2] Al-Osh, M. A. and Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, 8, 261–275. doi: 10.1111/j.1467-9892.1987.tb00438.x

[3] Altun, E. (2020). A new one-parameter discrete distribution with associated regression and integer-valued autoregressive models, *Math. Slovaca*, 70, 979-994. doi: 10.1515/ms-2017-0407

[4] Altun, E., and Khan, N. M. (2022). Modelling with the novel INAR (1)-PTE process. *Methodology and Computing in Applied Probability*, 24(3), 1735-1751. doi: 10.1007/s11009-021-09878-2

[5] Altun, E., El-Morshedy, M., & Eliwa, M. S. (2022). A Study on Discrete Bilal Distribution with Properties and Applications on Integervalued Autoregressive Process. *REVSTAT-Statistical Journal*, 20(4), 501-528.

[6] Alzaid, A. A. and Al-Osh, M. A. (1988). First-order integer-valued autoregressive (INAR(1)) process: Distributional and regression properties. *Statistica Neerlandica*, 42, 53–61. doi: 10.1111/j.1467-9574.1988.tb01521.x

[7] Bourguignon, M., Rodrigues, J., and Santos-Neto, M. (2019). Extended Poisson INAR (1) processes with equidispersion, underdispersion and overdispersion. *Journal of Applied Statistics*, 46, 101–118. doi: 10.1080/02664763.2018.1458216

[8] Chesneau, C., Tomy, L., and Veena, G. (2022). The Poisson-Modified Lindley Distribution. *Applied Mathematics E-Notes*, 22, 18–31.

[9] Du, J. G., Li, Y. (1991). The integer valued autoregressive (INAR(p)) model. *Journal of Time seies analysis*, 12, 129–142. doi: 10.1111/j.1467-9892.1991.tb00073.x

[10] Huang, J., and Zhu, F. (2021). A new first-order integer-valued autoregressive model with Bell innovations. *Entropy*, 23(6), 713. doi: 10.3390/e23060713

[11] Lívio, T., Khan, N. M., Bourguignon, M., & Bakouch, H. S. (2018). An INAR (1) model with Poisson-Lindley innovations. Econ. Bull, 38(3), 1505-1513.

[12] McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin*, 21, 645–650. doi: 10.1111/j.1752-1688.1985.tb05379.x

[13] McKenzie, E. (2003). Discrete variate time series. In D. N. Shanbhag, & C. R. Rao (Eds.), Stochastic processes: modelling and simulation (pp. 573-606). *Handbook of statistics*, 21, 573-606. doi: 10.1016/S0169-7161(03)21018-X

[14] Mohammadi, Z., Sajjadnia, Z., Bakouch, H. S., & Sharafi, M. (2022). Zero-and-one inflated Poisson-Lindley INAR (1) process for modelling count time series with extra zeros and ones. *Journal of Statistical Computation and Simulation*, 92(10), 2018-2040. doi: 10.1080/00949655.2021.2019255

[15] Raftery, A.E. (1995). *Bayesian model selection in social research.* Sociological Methodology, 25, 111-163. doi: 10.2307/271063

[16] Schwarz G. (1978), Estimating the Dimension of a Model. *The Annals of Statistics.*, 6, 461–464. http://www.jstor.org/stable/2958889

[17] Schweer, S. and Weiß, C. H. (2014). Compound Poisson INAR(1) processes: Stochastic properties and testing for overdispersion. *Computational Statistics and Data Analysis*, 77, 267–284. doi: 10.1016/j.csda.2014.03.005

[18] Shanker, R., Sharma, S. and Shanker, R. (2012). A Discrete two-Parameter Poisson Lindley distribution, *Journal of Ethiopian Statistical Association*, 21, 15-22.

[19] Silva, M. E., and Oliveira, V. L. (2004). Difference equations for the higher-order moments and cumulants of the INAR (1) model. Journal of Time Series Analysis 25:317-333. doi:10.1111/j.1467-9892.2004.01685.x

[20] Weiß, C. H. (2008). Thining operations for modeling time series of count - a survey. *AStA Advances in Statistical Analysis*, 92, 319–341. doi: 10.1007/s10182-008-0072-3

[1] *Veena G,*
Department of Mathematics & Statistics,
GITAM (Deemed to be) University, Bengaluru, Karnataka-561203,
India.
E-mail address: veenagpillai@hotmail.com

and

[2] *Lishamol Tomy,*
Department of Statistics,
Deva Matha College, Kuravilangad, Kerala-686633,
India.
E-mail address: lishatomy@gmail.com

and

[3] *Hassan S. Bakouch,*
Department of Mathematics,
College of Science, Qassim University, Buraydah,
Saudi Arabia.
E-mail address: h.bakouch@qu.edu.sa

and

[4] *Hassan S. Bakouch,*
Department of Mathematics,
Faculty of Science, Tanta University, Tanta,
Egypt.
E-mail address: hassan.bakouch@science.tanta.edu.eg

and

[5] *Christophe Chesneau,*
Université de Caen,
LMNO, Campus II, Science 3, 14032, Caen,
France.
E-mail address: christophe.chesneau@gmail.com