Gulf Journal of Mathematics ISSN: 2309-4966
Vol 19, Issue 2 (2025)156-167 https://doi.org/10.56947/gjom.v19i2.2661

FIRST-ORDER INTEGER-VALUED AUTO-REGRESSIVE PROCESS WITH DISCRETE LINDLEY DISTRIBUTION MARGINAL

TÉGAWENDÉ MARTIN KABORE[1*], O. JEAN-ETIENNE OUEDRAOGO[2] and MARCEL ILBOUDO[3]

Abstract. In this paper, we propose a stationary first-order integer-valued auto-regressive (INAR(1)) process with the DLD marginal distribution (DLD-INAR(1)) for dealing with overdispersed data. Several statistical properties of the model are established, such as conditional measures, autocorrelation, and non-reversibility over time. We consider the conditional maximum likelihood method to estimate the model's unknown parameters and study the properties of the estimators. The model is applied to a real-world time series dataset to evaluate its competitiveness against other INAR(1) models.

1. Introduction

In recent years, modeling stationary time series with discrete marginal distributions has attracted increasing interest, mainly due to their potential application areas. Count time series often exhibit correlation between observations, and a primary approach for their analysis is the use of integer-valued auto-regressive (INAR) models. First-order non-negative integer-valued auto-regressive processes (INAR(1)) were introduced by McKenzie, Al-Osh, and Alzaid [11]. Since then, several models based on the binomial thinning introduced by Steutel and van Harn [9] or its generalizations have emerged. Originally, the binomial thinning operator is generated by counting a series of independent random variables distributed according to Bernoulli's law and is defined as follows:

$$\alpha \star X = \sum_{i=1}^{X} W_i, \quad \alpha \in [0,1]\,; \tag{1.1}$$

where X is a non-negative integer-valued random variable and W_i is a sequence of independent and identically distributed (i.i.d.) random variables with a Bernoulli (α) distribution, independent of X.

Date: Received: Jan 30, 2025; Accepted: Feb 23, 2025.

* Corresponding author.

2020 *Mathematics Subject Classification.* Primary 46L55; Secondary 44B20.

Key words and phrases. Integer-valued Auto-Regressive process, Discrete Lindley Distribution, Over-dispersion, DLD-INAR(1), stationary process, Conditional Maximum Likelihood, Count series, time series, INAR..

Among the INAR(1) models based on the binomial thinning operator are the Poisson model (Al-Osh and Alzaid [11], Alzaid and Al-Osh [2], McKenzie [7]), negative geometric and binomial models (Alzaid and Al-Osh [2], McKenzie [8]) and the generalized Poisson model (Alzaid and Al-Osh [1]). Among those based on generalizations of the binomial thinning operator, we cite the negative binomial model (Al-Osh and Aly [4], Aly and Bouzar [5], Ristic, Bakouch, and Nastic [10], Ristic, Nastic, and Bakouch [13]) and the geometric model of Poisson by Aly and Bouzar [6]. The various integer-valued autoregressive models with discrete marginal distributions above have been proposed to improve the fit to real data, in order to provide a better analysis of these data. Sometimes, these models give an approximate analysis because of the nature and properties of the count data, which present limitations in their marginal distributions. For example, equi-dispersion has a variance equal to the mean, a situation that may not be consistent with the data, and under-dispersion occurs less frequently than over-dispersion.

This paper deals with over-dispersed count data. The Discrete Lindley Distribution (DLD) marginal distribution is used to construct an INAR(1) model using the binomial thinning operator with independent random variables from the Bernoulli counting series. Hereafter, the first-order integer-valued autoregressive model introduced with the DLD marginal distribution shall be denoted by DLD-INAR(1). The Discrete Lindley Distribution of a random variable X has the probability mass function (pmf):

$$P(X = x) = (e^{\theta} - 1)^2 (1 + x) e^{-\theta(2+x)}, \quad \theta > 0, \quad x = 0, 1, ...;$$

was introduced by Berhane Abebe and Rama Shanker [3]. The DLD(θ) distribution belongs to the compound Poisson family and has other common properties, such as uni-modality, over-dispersion, and infinite divisibility. Its generating function $G_X(s)$ is given by:

$$G_X(s) = \left(\frac{e^{\theta} - 1}{e^{\theta} - s} \right)^2, \quad \theta > 0, \quad s \neq e^{\theta}.$$

It is interesting to note that the unique parameter, θ, of the DLD distribution can be interpreted as follows: From a statistical perspective, θ determines the shape and dispersion of the distribution, and it is associated with the probability of rare events occurring. From a biological perspective, θ can be seen as a factor influencing the frequency of rare events in a biological system, such as the number of mutations or the prevalence of infections.

The paper is organized as follows. In Section 2, we construct a stationary first-order auto-regressive integer-valued process with the discrete Lindley distribution marginal (DLD-INAR (1)) and obtain the probability mass function of the innovation term. In Section 3, we study the properties of the proposed model. Section 4 deals with the estimation of unknown parameters of the model using the conditional maximum likelihood method. An application of the process to a real data set is presented in Section 5 to verify the competitiveness of the model. Section 6 presents the concluding remarks.

2. The stationary DLD-INAR(1) process.

A stationary DLD-INAR(1) process $\{X_t\}$ is defined by the recursive equation :

$$X_t = \alpha \star X_{t-1} + \varepsilon_t, \;\; t \geqslant 1; \tag{2.1}$$

where:

(1) the marginal distribution of X_t is DLD;
(2) the innovation sequence ε_t are independent and identically distributed according a non negative discrete distribution;
(3) X_{t-1} and ε_t are independent;
(4) the notation $\star$ denotes the binomial thinning operator, defined by the Equation 1.1, with $\alpha \in (0,1)$.

Furthermore, one can see that for $\alpha = 1$, the process is not stationary, and when $\alpha = 0$, the process denotes independent count data. Thus, in the following, we consider $\alpha \in (0,1)$.

Now, we aim to characterize the distribution of the innovation term for the stationary DLD-INAR(1) process. The following result provides its generating function.

Lemma 2.1. *Let $\{X_t\}$ be a stationary DLD-INAR(1) process and ε_t its innovation sequence. The generating function of ε_t is given by:*

$$G_\varepsilon(s) = \alpha^2 + (1-\alpha^2)\left[\frac{2\alpha}{1+\alpha}\frac{e^\theta - 1}{e^\theta - s} + \frac{1-\alpha}{1+\alpha}\frac{\left(e^\theta-1\right)^2}{(e^\theta - s)^2}\right] \quad \theta > 0, \quad s \neq e^\theta.$$

Proof. Using,

$G_X(s) = G_\varepsilon(s) G_X(1-\alpha+\alpha s)$, and $G_X(s) = \left(\dfrac{e^\theta - 1}{e^\theta - s}\right)^2$.

One has, $G_\varepsilon(s) = \left[\dfrac{e^\theta - 1 + \alpha - \alpha s}{e^\theta - s}\right]^2$. After rearrangement, the result hold. □

In the following, the results stated will allow us to determine the probability mass function of innovation sequence.

Lemma 2.2. *Let $\alpha \in (0,1)$ and $\theta > 0$, then the mixture*

$$h(x) \;\; = \;\; \frac{2\alpha}{1+\alpha}\left(1-e^{-\theta}\right)\left(e^{-\theta}\right)^x + \frac{1-\alpha}{1+\alpha}(1+x)\left(1-e^{-\theta}\right)^2\left(e^{-\theta}\right)^x, \quad x = 0,1,...;$$

is a probability mass function.

Proof. First, it is clear that $h(x) \geq 0$ for $x \geq 0$ and $\theta > 0$. Furthermore, we have $\sum\limits_{x \geqslant 0} h(x) = 1$, since $\sum\limits_{x\geqslant 0}\left(e^{-\theta}\right)^x = \dfrac{1}{1-e^{-\theta}}$ and $\sum\limits_{x\geqslant 0}(1+x)\left(e^{-\theta}\right)^x = \dfrac{e^{-\theta}}{(1-e^{-\theta})^2}$.
Hence $h(x)$ is a mass probability function. □

Proposition 2.3. *The probability mass function (pmf) defined in Lemma 2.2 is a mixture of* $\mathbf{Geo}(1-e^{-\theta})$ *and* $\mathbf{NB}(2, 1-e^{-\theta})$ *with respective weights* $\dfrac{2\alpha}{1+\alpha}$

and $\frac{1-\alpha}{1+\alpha}$, *where* **Geo** *and* **NB** *denote the Geometric and Binomial Negative distribution, respectively.*

Proof. The probability mass functions of the distributions $\mathbf{Geo}(1-e^{-\theta})$ and $\mathbf{NB}(2,1-e^{-\theta})$ are given respectively by: $P_1(X=x,\theta)=(1-e^{-\theta})(e^{-\theta})^x$, $P_2(X=x,\theta)=(1+x)(1-e^{-\theta})^2(e^{-\theta})^x$. We see that $h(x)=\frac{2\alpha}{1+\alpha}P_1(X=x,\theta)+\frac{1-\alpha}{1+\alpha}P_2(X=x,\theta)$. Therefore, $h(x)$ is a mixture of $\mathbf{Geo}(1-e^{-\theta})$ and $\mathbf{NB}(2,1-e^{-\theta})$ of respective weights $\frac{2\alpha}{1+\alpha}$ and $\frac{1-\alpha}{1+\alpha}$. $\square$

From Lemma 2.2, one can derive the result below, which stands for the pmf of innovations.

Proposition 2.4. *Consider the stationary DLD-INAR(1) process* $\{X_t\}$ *defined by Equation 2.1. The probability mass function of the innovation term* $\{\varepsilon_t\}$ *is given as:*

$$f_\varepsilon(x)=\alpha^2 I_0(x)+(1-\alpha^2)h(x),\quad x=0,1,...;$$

where $I_0(x)$ *is the degenerate distribution function at zero, and* $h(x)$ *is a probability mass function defined in Lemma 2.2.*

Proof. Considering,

$G_\varepsilon(s)=\alpha^2+(1-\alpha^2)\left[A\frac{e^\theta-1}{e^\theta-s}+B\frac{\left(e^\theta-1\right)^2}{(e^\theta-s)^2}\right]$, with $A=\frac{2\alpha}{1+\alpha}$ and $B=\frac{1-\alpha}{1+\alpha}$,

the pmf of innovation term ε_t, $f_\varepsilon(x,\theta)$ is given by:

$$\frac{G_\varepsilon^{(k)}(0)}{k!}=\begin{cases}\alpha^2+(1-\alpha^2)\Big[A\left(1-e^{-\theta}\right)+B\left(1-e^{-\theta}\right)^2\Big], & \text{if } k=0,\\ (1-\alpha^2)\Big[A\left(1-e^{-\theta}\right)\left(e^{-\theta}\right)^k+B(1+k)\left(1-e^{-\theta}\right)^2\left(e^{-\theta}\right)^k\Big], & \text{if } k\geq 1.\end{cases}$$

Then,

$f_\varepsilon(x,\theta)=\alpha^2 I_0(x)+(1-\alpha^2)\Big[A\left(1-e^{-\theta}\right)\left(e^{-\theta}\right)^x+B(1+x)\left(1-e^{-\theta}\right)^2\left(e^{-\theta}\right)^x\Big]$, $\forall\ x\geq 0$.

We see that, $h(x)=A\left(1-e^{-\theta}\right)\left(e^{-\theta}\right)^x+B(1+x)\left(1-e^{-\theta}\right)^2\left(e^{-\theta}\right)^x$.

Hence, $f_\varepsilon(x,\theta)=\alpha^2 I_0(x)+(1-\alpha^2)h(x)$, with $I_0(x)=\begin{cases}1 & \text{if } x=0,\\ 0 & \text{if } x\neq 0.\end{cases}$

$\square$

Based on Proposition 2.3 and 2.4, the probability mass function of the innovation term is a mixture of geometric, negative binomial, and a degenerate distribution. Consequently, the DLD distribution is discrete self-decomposable (see Steutel and van Harn [9]) for $\theta>0$. Furthermore, appealing to Proposition 2.4, the process defined by Equation 2.1 can be rewritten as:

$$X_t=\begin{cases}\alpha\star X_{t-1}, & \text{w.p. } \alpha^2,\\ \alpha\star X_{t-1}+\varepsilon_t, & \text{w.p. } 1-\alpha^2;\end{cases}$$

where w.p. stands for 'with probability'. Consequently, after rearrangement, it follows that:

$$X_t = \alpha \star X_{t-1} + I_t H_t, \quad t \geq 1, \quad \alpha \in (0,1);$$

where $P(I_t = 0) = 1 - P(I_t = 1) = \alpha^2$, H_t has the probability mass function h given by Lemma 2.2, and X_{t-k} independent of $I_t H_t$ for $k \geq 1$.

Based on the properties of the binomial thinning operator, the model given by the Equation 2.1 can be expressed in terms of the innovation sequence $\{\varepsilon_t\}$ as follows:

$$X_t \stackrel{d}{=} \sum_{j=0}^{\infty} \alpha^j \star \varepsilon_{t-j}, \quad t \geq 1, \quad \alpha \in (0,1). \tag{2.2}$$

3. Statistical properties of the model.

Statistical properties of the model, including moments of innovation sequence, the correlation structure, and conditional moments of the process, are highlighted in the section.

Lemma 3.1. *Let ε_t be the innovation term of the stationary DLD-INAR(1) process. Its first and second moments are given by:*

$$\begin{aligned} E(\varepsilon_t) &= \frac{2(1-\alpha)}{e^\theta - 1}; \\ Var(\varepsilon_t) &= \frac{2e^\theta(1-\alpha^2) - 2\alpha(e^\theta - 1)(1-\alpha)}{(e^\theta - 1)^2}. \end{aligned}$$

Proof. Due the stationarity of the process,
$E(X_t) = \alpha E(X_t) + E(\varepsilon_t) \Rightarrow E(\varepsilon_t) = (1-\alpha)E(X_t)$,
with $E(X_t) = \dfrac{2}{e^\theta - 1}$, hence $E(\varepsilon_t) = \dfrac{2(1-\alpha)}{e^\theta - 1}$.
Considering Equation 2.2, one has:

$$\begin{aligned} Var(X_t) &= Var(\sum_{j=0}^{\infty} \alpha^j \star \varepsilon_{t-j}) \\ &= \sum_{j=0}^{\infty} \alpha^{2j} Var(\varepsilon_{t-j}) + \sum_{j=0}^{\infty} \alpha^j(1-\alpha^j)E(\varepsilon_{t-j}) \\ Var(X_t) &= \frac{1}{1-\alpha^2}\big[Var(\varepsilon_t) + \alpha E(\varepsilon_t)\big]; \end{aligned}$$

then,

$$\begin{aligned} Var(\varepsilon_t) &= (1-\alpha^2)Var(X_t) - \alpha E(\varepsilon_t) \\ &= \frac{2e^\theta(1-\alpha^2) - 2\alpha(e^\theta - 1)(1-\alpha)}{(e^\theta - 1)^2}; \end{aligned}$$

where $Var(X_t) = \dfrac{2e^\theta}{(e^\theta - 1)^2}$.

$\square$

From now on, the following results can be derived.

Proposition 3.2. *Let X_t be a stationary DLD-INAR(1) process. The following properties hold:*

$$E(X_t \mid X_{t-1} = x) = \alpha x + \frac{2(1-\alpha)}{(e^{\theta}-1)};$$

$$Var(X_t \mid X_{t-1} = x) = \alpha(1-\alpha)x + \frac{2e^{\theta}(1-\alpha^2) - 2\alpha(e^{\theta}-1)(1-\alpha)}{(e^{\theta}-1)^2}.$$

The auto-covariance function at lag $k \geq 0$ is given by:

$$\gamma_k = \alpha^k \gamma_0, \quad \alpha \in (0,1);$$

where γ_0 denotes the variance of X_t.

Proof. The conditional mean is given by:

$$\begin{aligned} E(X_t \mid X_{t-1} = x) &= \alpha x + E(\varepsilon_t) \\ &= \alpha x + \frac{2(1-\alpha)}{e^{\theta}-1}; \end{aligned}$$

the conditional variance is given by:

$$\begin{aligned} Var(X_t \mid X_{t-1} = x) &= \alpha(1-\alpha)x + Var(\varepsilon_t) \\ &= \alpha(1-\alpha)x + \frac{2e^{\theta}(1-\alpha^2) - 2\alpha(e^{\theta}-1)(1-\alpha)}{(e^{\theta}-1)^2}; \end{aligned}$$

and, the auto-covariance function is given by:

$$\begin{aligned} \gamma_k &= Cov(X_t, X_{t-k}) \\ &= Cov(\alpha \star X_{t-1} + \varepsilon_t, X_{t-k}) \\ &= Cov(\alpha \star X_{t-1}, X_{t-k}) + Cov(\varepsilon_t, X_{t-k}) \\ \gamma_k &= \alpha^k \gamma_0, \quad k \geq 0. \end{aligned}$$

□

Since $\hat{I}_{X_t} = \frac{Var(X_t)}{E(X_t)} = \frac{e^{\theta}}{e^{\theta}-1} = 1 + \frac{1}{e^{\theta}-1}$, it follows that $\hat{I}_{X_t} > 1$ for $\theta > 0$. Therefore, note that $\hat{I}_{X_t}$ is a decreasing function of the unique parameter θ. Which shows that this parameter therefore controls the degree of overdispersion in the proposed model. Hence, the proposed DLD-INAR(1) model is particularly suitable for overdispersed count data.

Moreover, the above proposition shows that the stationary DLD-INAR(1) process has a first-order linear auto-regressive structure, and the auto-correlation function is $\gamma_k = \alpha^k$, where $\alpha \in (0,1)$ and $k \geq 0$. That is, the autocorrelation function decreases exponentially as $k \to +\infty$. The correlation structure of the proposed process is, therefore, analogous to the INAR(1) and AR(1) processes. Another important result is given below.

Proposition 3.3. *If X_t is generate according to the stationary DLD-INAR(1) process, then the joint probability generating function of the random variables X_t and X_{t-1} is given by:*

$$G_{X_t,X_{t-1}}(s_1,s_2) = \left(\frac{e^\theta-1}{e^\theta-s_2(1-\alpha+\alpha s_1)}\right)^2+(1-\alpha^2)\left(\frac{e^\theta-1+\alpha(1-s_1)}{e^\theta-s_1}\right)^2.$$

Proof. The joint probability generating function is given by:

$$\begin{aligned}
G_{X_t,X_{t-1}}(s_1,s_2) &= \alpha^2 G_{\alpha\star X_{t-1},X_{t-1}}(s_1,s_2)+(1-\alpha^2)G_{\alpha\star X_{t-1}+\varepsilon_t,X_{t-1}}(s_1,s_2)\\
&= \alpha^2 G_{X_t}(s_2(1-\alpha+\alpha s_1))+(1-\alpha^2)\left[G_{\varepsilon_t}(s_1)G_{X_t}(s_2(1-\alpha+\alpha s_1))\right]\\
&= \alpha^2\left(\frac{e^\theta-1}{e^\theta-s_2(1-\alpha+\alpha s_1)}\right)^2+(1-\alpha^2)\Bigg[\left(\frac{e^\theta-1+\alpha(1-s_1)}{e^\theta-s_1}\right)\\
&\times \left(\frac{e^\theta-1}{e^\theta-s_2(1-\alpha+\alpha s_1)}\right)\Bigg]^2\\
G_{X_t,X_{t-1}}(s_1,s_2) &= \left(\frac{e^\theta-1}{e^\theta-s_2(1-\alpha+\alpha s_1)}\right)^2+(1-\alpha^2)\left(\frac{e^\theta-1+\alpha(1-s_1)}{e^\theta-s_1}\right)^2.
\end{aligned}$$

□

By Proposition 3.3, $G_{X_t,X_{t-1}}(s_1,s_2)$ is not symmetric in s_1 and s_2, which means that the stationary DLD-INAR(1) process is not reversible in time.

The following result concerns the transition probability of the stationary DLD-INAR(1) process, which plays an important role in estimating the unknown parameters of the model.

Proposition 3.4. *Let X_t be a stationary DLD-INAR(1) process. Its transition probability is given by:*

$$\begin{aligned}
P(X_t=x_t\mid X_{t-1}=x_{t-1}) &= \sum_{k=0}^{min(x_{t-1},x_t)}\binom{x_{t-1}}{k}\alpha^k(1-\alpha)^{x_{t-1}-k}\Bigg[\alpha^2 I_0(x_t-k)\\
&+ (1-\alpha)\times e^{-\theta(x_t-k)}\Bigg(2\alpha(1-e^{-\theta})\\
&+ (1-\alpha)(1-e^{-\theta})^2(1+x_t-k)\Bigg)\Bigg].
\end{aligned}$$

Proof. From Equation 2.1, one has:

$$P(X_t=x_t\mid X_{t-1}=x_{t-1}) = \sum_{k=0}^{min(x_{t-1};x_t)}\binom{x_{t-1}}{k}\alpha^k(1-\alpha)^{x_{t-1}-k}P(\varepsilon_t=x_t-k);$$

however,

$$P(\varepsilon_t=x_t-k)=f_\varepsilon(x_t-k)=\alpha^2 I_0(x_t-k)+(1-\alpha^2)h(x_t-k);$$

and,

$$h(x_t - k) = \frac{e^{-\theta(x_t-k)}}{1+\alpha}\Big(2\alpha(1-e^{-\theta}) + (1-\alpha)(1-e^{-\theta})^2(1+x_t-k)\Big);$$

one has:

$$\begin{aligned} P(X_t = x_t \mid X_{t-1} = x_{t-1}) &= \sum_{k=0}^{min(x_{t-1};x_t)} \binom{x_{t-1}}{k} \alpha^k (1-\alpha)^{x_{t-1}-k}\Big[\alpha^2 I_0(x_t-k) \\ &+ (1-\alpha) \times e^{-\theta(x_t-k)}\Big(2\alpha(1-e^{-\theta}) \\ &+ (1-\alpha)(1-e^{-\theta})^2(1+x_t-k)\Big)\Big]. \end{aligned}$$

□

In the next section, we develop the conditional likelihood procedure for estimating the unknown parameters.

4. Estimation procedure and inference

In this section, we estimate the unknown parameters of the model using the conditional maximum likelihood method, and check the asymptotic properties of these estimators. To optimize the conditional log-likelihood function, we use the *nlminb* function in the Stats package of R software.

4.1. **Conditional maximum likelihood method.** The conditional log-Likelihood function $\mathrm{L}(x_t, \lambda)$ is defined by:

$$\mathrm{L}(x_t, \lambda) = \sum_{t=2}^{n} \ln P(X_t = x_t \mid X_{t-1} = x_{t-1}, \lambda), \quad x_t \in \mathbb{N}, \quad t \geq 1.$$

Considering Proposition 3.4, the conditional log-likelihood function $\mathrm{L}(x_t, \lambda)$ of the stationary DLD-INAR(1) process, X_t, with parameter $\lambda = (\alpha, \theta)$, is given by:

$$\begin{aligned} \mathrm{L}(x_t, \lambda) &= \sum_{t=2}^{n} \ln\Bigg[\sum_{k=0}^{min(x_{t-1},x_t)} \binom{x_{t-1}}{k} \alpha^k (1-\alpha)^{x_{t-1}-k}\Big[\alpha^2 I_0(x_t-k) + (1-\alpha) \\ &\times \; e^{-\theta(x_t-k)}\Big(2\alpha(1-e^{-\theta}) + (1-\alpha)(1-e^{-\theta})^2(1+x_t-k)\Big)\Big]\Bigg]. \end{aligned}$$

4.2. **Some simulation results.** A Monte-Carlo simulation study is conducted using R software, based on data from 1000 series generated by the stationary DLD-INAR(1) process. The objective is to determine the estimators of the parameters (α and θ), analyze their asymptotic properties through the evaluation of bias and RMSE, and finally, examine the execution time associated with the estimation.

$$Bias(\hat{\lambda}) = \frac{1}{N}\sum_{t=1}^{N} \hat{\lambda}_t - \lambda, \; RMSE(\hat{\lambda}) = \sqrt{\frac{1}{N}\sum_{t=1}^{N} (\hat{\lambda}_t - \lambda)^2}, \text{ with } N = 1000.$$

In order to obtain the estimator $\hat{\alpha}$, the simulation is performed with theoretical values of: $\alpha = \{0.2, 0.4, 0.5, 0.7, 0.9\}$ and $T = \{25, 50, 200, 500\}$, where θ is fixed at 1 (see Figure 1). Furthermore, for the estimator $\hat{\theta}$, the simulation is conducted with theoretical values of: $\theta = \{0.2, 0.4, 0.5, 0.7, 0.9\}$ and $T = \{25, 50, 200, 500\}$, where α is fixed at 0.5 (see Figure 2).

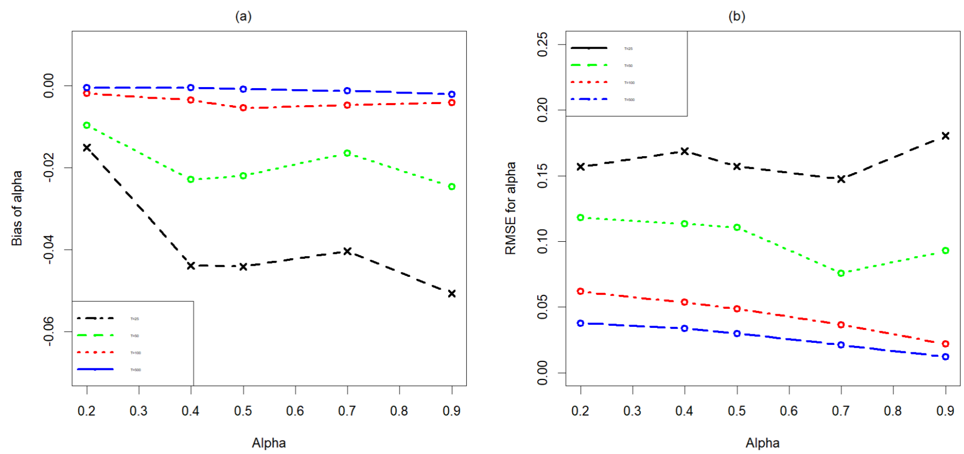

FIGURE 1. Bias and RMSE of the estimator $\hat{\alpha}$.

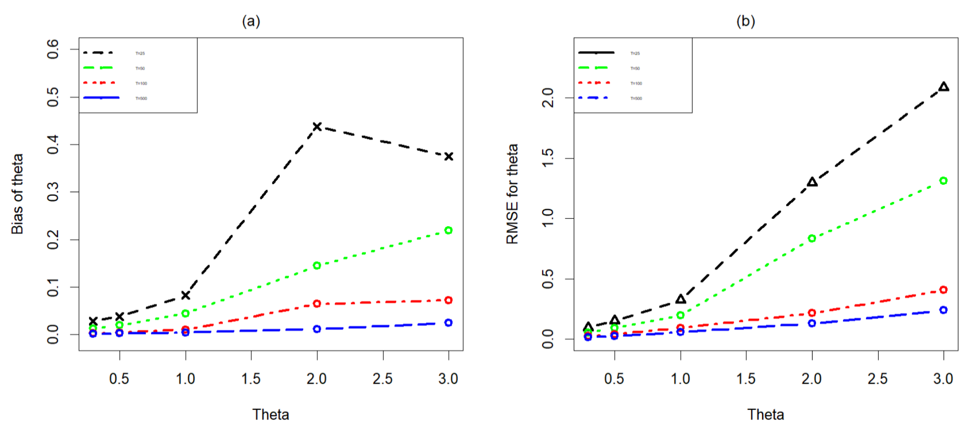

FIGURE 2. Bias and RMSE of the estimator $\hat{\theta}$.

In light of the graphs presented in Figures 1-(a) and 2-(a), it is clear that the biases of the estimators $\hat{\alpha}$ and $\hat{\theta}$ tend to zero as the sample size increases. This indicates that the estimators $\hat{\alpha}$ and $\hat{\theta}$ converge to their respective theoretical values α and θ as the sample size T grows. Consequently, the conditional maximum likelihood estimators $\hat{\alpha}$ and $\hat{\theta}$ are asymptotically unbiased. It is also worth noting that, for each given sample size, the bias of the estimator $\hat{\theta}$ (Figure 2-(a)) remains positive, while that of $\hat{\alpha}$ (Figure 1-(a)) remains negative. This means that the estimator $\hat{\theta}$ tends to overestimate its theoretical value, while the estimator $\hat{\alpha}$ tends to underestimate it. Moreover, it is observed that the estimator $\hat{\theta}$ converges more quickly to zero than $\hat{\alpha}$ as the sample size increases. This makes $\hat{\theta}$ more suitable for estimating smaller-sized series, while $\hat{\alpha}$ is better suited for larger-sized series.

Regarding the RMSE plots for the parameters $\hat{\alpha}$ (Figure 1-(b)) and $\hat{\theta}$ (Figure 2-(b)), it is clear that as the sample size increases, the RMSE values decrease. However, the RMSE of $\hat{\theta}$ decreases even more rapidly than that of $\hat{\alpha}$, suggesting that the estimator $\hat{\theta}$ is more consistent than $\hat{\alpha}$.

However, it should be emphasized that the execution time for a simulation with a sample of one thousand series, each of size one thousand, is 2,434.86 seconds. This simulation was run on an "ASUS Vivobook" computer with the following specifications: x64 processor, Intel(R) Core(TM) i3-1215U of the 12th generation at 1.20 GHz, and 8 GB of RAM.

Having built this new INAR(1) process, it was imperative to verify its competitiveness against its INAR(1) counterparts by applying it to real data.

5. Competitiveness of the DLD-INAR(1) model

In this section, we present an application of the DLD-INAR(1) model on a real dataset of count series. The data concerns the number of animals that presented skin lesions, recorded monthly from January 2003 to December 2009, in a region of New Zealand (cf. [12]). This dataset, related to skin lesions, is empirically overdispersed, with a dispersion index of 2.34. The objective of this application is to fit these data using the DLD-INAR(1) model in order to compute statistical criteria such as the log-likelihood function (LL), the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the consistent Akaike information criterion (CAIC), as well as the estimated values ($\hat{\alpha}$ and $\hat{\theta}$). These statistical criteria, which measure the performance of the model, will then be compared with those obtained from other reference INAR(1) models available in the literature. It is worth noting that no specific model selection criteria were applied. These models were chosen for this performance study due to the availability in the literature of their fit statistics (AIC, LL, BIC, and CAIC) on the same dataset (cf. [12]). The results are presented in Table 1.

After analyzing the results presented in Table 1, it is evident that the PLINAR(1) model outperforms the DLD-INAR(1) model in terms of optimal fit performance (AIC, LL, BIC, and CAIC). However, the DLD-INAR(1) model provides the second-best fit, slightly ahead of the NBIINAR(1) model. Consequently, the DLD-INAR(1) model constitutes a relevant alternative for the analysis and prediction of overdispersed count data.

6. Conclusions

A new stationary first-order integer-valued auto-regressive model with the DLD marginal distribution, named DLD-INAR(1), is introduced for dealing with overdispersed data. We derive the probability mass function of the innovation term of the process as a mixture of geometric, negative binomial, and degenerate distribution. Various statistical properties of the model are obtained, such as the conditional mean and variance, autocovariance function, autocorrelation function,

Table 1. MLE, AIC, LL, BIC, and CAIC for the skin lesions data.

Models	**MLE**	**AIC**	**LL**	**BIC**	**CAIC**
PINAR(1)	$\hat{\lambda} = 1.18$,				
	$\hat{\alpha} = 0.17$	306.22	-151.11	311.08	306.36
GINAR(1)	$\hat{p} = 0.58$,				
	$\hat{\alpha} = 0.12$	277.72	-136.86	282.58	277.86
NGINAR(1)	$\hat{p} = 1.41$,				
	$\hat{\alpha} = 0.17$	277.12	-136.56	281.98	277.26
NBRCINAR(1)	$\hat{n} = 1.2$,				
	$\hat{p} = 0.45$,				
	$\hat{\rho} = 0.17$	279.93	-136.96	284.79	280.07
NBIINAR(1)	$\hat{n} = 1.01$,				
	$\hat{p} = 1.03$,				
	$\hat{\rho} = 0.31$	276.14	-135.07	281.00	276.28
DLD-INAR(1)	$\hat{\theta} = \mathbf{0.87}$,				
	$\hat{\alpha} = \mathbf{0.15}$	**276.13**	**-135.06**	**280.99**	**276.25**
PLINAR(1)	$\hat{\theta} = 1.05$,				
	$\hat{\alpha} = 0.25$	223.81	-109.9	228.67	223.95

joint probability generating function, and transition probability. The unknown parameters of the model are estimated using the maximum likelihood method, and the performance of the estimates is assessed through simulations. The model is fitted to a dataset concerning the symptom of skin lesions from New Zealand animal health laboratories, and a brief analysis of these overdispersed data is conducted within this framework. Based on various obtained fit statistics, such as the Akaike information criterion (AIC), the log-likelihood (LL), the Bayesian information criterion (BIC), and the consistent Akaike information criterion (CAIC), the DLD-INAR(1) model provides the second-best fit, following the PLINAR(1) model. Consequently, the DLD-INAR(1) model constitutes a relevant alternative for the analysis and prediction of overdispersed count data.

An extension of the results to the DLD-INAR(p) could be an avenue for future research.

Acknowledgement

The authors would like to express their sincere gratitude to the editorial board and the two reviewers for their careful reading of the article and for the insightful comments they provided.

References

1. A. A. Alzaid; M. A. Al-Osh, "Some autoregressive moving average processes with generalized Poisson marginal distributions", Annals of the Institute of Statistical Mathematics. https://doi.org/10.1007/BF00775809
2. A. Alzaid; M. Al-Osh, "First-Order Integer-Valued Autoregressive (INAR (1)) Process: Distributional and Regression Properties", Statistica Neerlandica, 1988. https://doi.org/10.1111/j.1467-9574.1988.tb01521.x

3. Abebe B, Shanker R, "A discrete lindley distribution with applications in biological sciences", Biom Biostat Int J. 2018.
https://doi.org/10.15406/bbij.2018.07.00189
4. Al-Osh, Mohamed A.; Aly, Emad-Eldin A.A., "First order autoregressive time series with negative binomial and geometric marginals", Communication in Statistics- Theory and Methods, 1992.
https://doi.org/10.1080/03610929208830925
5. Aly, Emad-Eldin A.A, Bouzar, Nadjib, "Explicit stationary distributions for some galton-watson processes with immigration", Communications in Statistics Stochastic Models, 1994.
https://doi.org/10.1080/15326349408807305
6. E.E.A.A. Aly, N. Bouzar, "On Some Integer-Valued Autoregressive Moving Average Models, Journal of Multivariate Analysis", 1994.
https://doi.org/10.1006/jmva.1994.1038
7. Ed McKenzie, "Some ARMA Models for Dependent Sequences of Poisson Counts", Advances in Applied Probability, 1988.
https://doi.org/10.1017/S0001867800018395
8. Ed McKenzie, "Autoregressive Moving-Average Processes with Negative-Binomial and Geometric Marginal Distributions", Advances in Applied Probability, 1986.
https://doi.org/10.2307/1427183
9. F. W. Steutel and K. van Harn, "Discrete Analogues of Self-Decomposability and Stability", The Annals of Probability, 1979.
https://doi.org/10.1214/aop/1176994950
10. Hassan S. Bakouch, Miroslav M. Ristic, "Zero truncated Poisson integer-valued AR(1) model", Metrika, 2009.
https://doi.org/10.1007/s00184-009-0252-5
11. M. A. Al-Osh, A. A. Alzaid, "FIRST-ORDER INTEGER-VALUED AUTOREGRESSIVE (INAR(1)) PROCESS", Journal of Time Series Analysis, 1987.
https://doi.org/10.1111/j.1467-9892.1987.tb00438.x
12. Mansour Aghababaei Jazi, Geoff Jones, Chin-Diew Lai, "First-order integer valued AR processes with zero inflated poisson innovations", Journal of Time Series Analysis, 2012.
https://doi.org/10.1111/j.1467-9892.2012.00809.x
13. Ristic, Miroslav M. Nastic, Aleksandar S. Bakouch, Hassan S, "Estimation in an Integer-Valued Autoregressive Process with Negative Binomial Marginals (NBINAR(1))", Communication in Statistics- Theory and Methods, 2012.
https://doi.org/10.1080/03610926.2010.529528
14. Saidi, Abdelaziz ABBADI and Mohamed TAHRICHI, " Eestimating paramatric surface areas usingspline quasi-interpolants", Gulf Journal Mathematics, 2024.
https://doi.org/10.56947/gjom.v18i1.2370

[1] L@mia Laboratory, Norbert Zongo University, Koudougou, Burkina Faso.
Email address: tzorghoe@gmail.com

[2] L@mia Laboratory, Norbert Zongo University, Koudougou, Burkina Faso.
Email address: ouedraogoetienne@yahoo.fr

[3] L@mia Laboratory, Norbert Zongo University, Koudougou, Burkina Faso.
Email address: imarcelilboudo@yahoo.fr