

Article

# New One-Parameter Over-Dispersed Discrete Distribution and Its Application to the Nonnegative Integer-Valued Autoregressive Model of Order One

Muhammed Rasheed Irshad <sup>1,\*</sup>, Sreedeviamma Aswathy <sup>1</sup>, Radhakumari Maya <sup>2</sup> and Saralees Nadarajah <sup>3</sup><sup>1</sup> Department of Statistics, Cochin University of Science and Technology, Cochin 682022, India<sup>2</sup> Department of Statistics, University College, Thiruvananthapuram 695034, India; publicationsofmaya@gmail.com<sup>3</sup> Department of Mathematics, University of Manchester, Manchester M13 9PL, UK; mbbssn2@manchester.ac.uk

\* Correspondence: irshadmr@cusat.ac.in

**Abstract:** Count data arise in inference, modeling, prediction, anomaly detection, monitoring, resource allocation, evaluation, and performance measurement. This paper focuses on a one-parameter discrete distribution obtained by compounding the Poisson and new X-Lindley distributions. The probability-generating function, moments, skewness, kurtosis, and other properties are derived in the closed form. The maximum likelihood method, method of moments, least squares method, and weighted least squares method are used for parameter estimation. A simulation study is carried out. The proposed distribution is applied as the innovation in an INAR(1) process. The importance of the proposed model is confirmed through the analysis of two real datasets.

**Keywords:** discrete statistical model; dispersion index; hazard rate function; parameter estimation; simulation; INAR(1)

**MSC:** 60E05; 62E15; 62F10

**Citation:** Irshad, M.R.; Aswathy, S.; Maya, R.; Nadarajah, S. New One-Parameter Over-Dispersed Discrete Distribution and Its Application to the Nonnegative Integer-Valued Autoregressive Model of Order One. *Mathematics* **2024**, *12*, 81. <https://doi.org/10.3390/math12010081>

Academic Editor: Manuel Franco

Received: 10 November 2023

Revised: 17 December 2023

Accepted: 19 December 2023

Published: 26 December 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Count data find diverse applications across various fields, such as the frequency of typing errors on a page or the quantity of lice present on the heads of Hindu male prisoners in Cannanore [1]. Count data modeling provides a powerful framework for understanding and analyzing discrete events or occurrences. It allows researchers, policymakers, and organizations to quantify and interpret patterns, identify influential factors, make predictions, and inform evidence-based decision-making. The most typical models for count data are the Poisson and negative binomial distributions. Because of its equi-dispersive character, the Poisson distribution should not be applied when an over-dispersion issue arises. Note that count data commonly exhibit either over-dispersion or under-dispersion, and this has driven the development of more versatile models over the past few decades.

Recall that real data are often over-dispersed. Many researchers have developed mixed Poisson distributions such as the Poisson Weibull distribution [2], Conway–Maxwell–Poisson distribution [3], Poisson transmuted Lindley distribution [4], Poisson transmuted exponential distribution [5], Poisson quasi-Lindley distribution [6], Poisson Bilal distribution [7], Poisson Xgamma distribution [8], Poisson extended exponential distribution [9] and the Poisson generalized Lindley distribution [10].

Moreover, count data are prevalent in numerous applied research domains. Examples include the number of hospital admissions over time, monitoring the number of stock trades per minute or daily transaction volumes in financial markets, and analyzing the number of reported crimes per month in different regions. A nonnegative integer-valued

autoregressive process of order one (INAR(1)) is a discrete-time autoregressive model where the current value of the process depends on its previous value and is restricted to take nonnegative integer values. The INAR(1) process with Poisson innovations due to [11] was the pioneering work of INAR(1) processes. But Poisson distribution assumes that the variance is equal to the mean (equi-dispersion). In over-dispersed count data, this assumption is violated, as the variance is larger than the mean. Since [11], many researchers have suggested INAR(1) processes under non-Poisson innovations. Some examples of other innovations are geometric innovations [12], discrete three-parameter Lindley innovations [13], Bell innovations [14], and discrete Bilal innovations [15]. We list also some mixed Poisson innovations, as follows: Poisson–Lindley innovations [16], new Poisson weighted exponential innovations [17], Poisson quasi-Xgamma innovations [18], discrete pseudo-Lindley innovations [19], Poisson transmuted exponential innovations [20], and Poisson generalized Lindley innovations [10].

Lindley distribution has found applications in various fields such as finance, environmental studies, and medical research, among many others. Due to its ability to handle various types of data, Lindley distribution has become a valuable tool in statistical modeling, particularly in situations where traditional distributions may not provide an adequate fit. Researchers have often used Lindley distribution to gain insights into different datasets and make more accurate predictions and inferences. Here, we consider the continuous new X-Lindley (NXL) distribution [21]. It is a novel one-parameter distribution that incorporates the advantages of both Lindley and exponential distributions. It has potential applications in diverse fields such as biology, engineering, astronomy, actuarial science, and medicine. Moreover, this distribution exhibits an elevated risk rate and a diminishing average residual life function.

In this paper, we compound the Poisson and new X-Lindley distributions, resulting in a new one-parameter distribution, which is referred to as the Poisson new X-Lindley (PNXL) distribution. This new one-parameter distribution can handle over-dispersed count data.

The remainder of the paper is structured as follows. In Section 2, the one-parameter PNXL distribution is introduced and its statistical properties are derived. Estimation techniques utilized to estimate the unknown parameter are described in Section 3, and their finite sample performance is evaluated through a simulation study. A new INAR(1)PNXL process is described in Section 4. Two real datasets are analyzed in Section 5 to demonstrate the effectiveness of the suggested distribution. Conclusions are provided in Section 6.

## 2. Poisson New X-Lindley Distribution

### 2.1. The Poisson New X-Lindley Distribution and Its Statistical Properties

The NXL distribution is a special case of one-parameter polynomial exponential distribution (NPED) proposed in [22]. The probability density function (pdf) and cumulative distribution function (cdf) of the NXL distribution are given, respectively, by

$$p(x; \theta) = \frac{\theta(1 + \theta x)e^{-\theta x}}{2} \quad \text{and} \quad F(x; \theta) = 1 - \left(\frac{1}{2}\theta x + 1\right)e^{-\theta x},$$

respectively, for  $x > 0$  and  $\theta > 0$ . Our suggested one-parameter discrete compound distribution is built on the basis of the NXL distribution. That is, the PNXL distribution is a mixed-Poisson distribution obtained by compounding the Poisson and NXL distributions. Its probability mass function (pmf) is formulated as follows:

**Definition 1.** Let  $X$  denote a random variable having the PNXL distribution such that  $X|\lambda \sim P(\lambda)$  and  $\lambda|\theta \sim NXL(\theta)$ , where  $\lambda > 0$  and  $\theta > 0$ . The unconditional pmf of  $X$  is

$$p(x; \theta) = \int_0^\infty \frac{e^{-\lambda} \lambda^x}{x!} \frac{\theta e^{-\theta \lambda} (1 + \theta \lambda)}{2} d\lambda = \frac{\theta(2\theta + \theta x + 1)}{2(\theta + 1)^{x+2}} \tag{1}$$

for  $x = 0, 1, 2, \dots$  and  $\theta > 0$ .

The corresponding cdf is

$$F(x; \theta) = \frac{2\theta^2(1 + \theta)^x + 2[(1 + \theta)^x - 1] + \theta[4(1 + \theta)^x - x - 3]}{2(1 + \theta)^{x+2}}.$$

The pmf (1) is log concave since

$$\frac{p(x + 1; \theta)}{p(x; \theta)} = \frac{1 + (3 + x)\theta}{(1 + \theta)(1 + (2 + x)\theta)}$$

is a decreasing function in  $x$  for all parameter values. Furthermore,  $\frac{p(x + 1; \theta)}{p(x; \theta)} < 1$  for all  $x = 0, 1, \dots$  and  $\theta > 0$ , so the pmf is unimodal.

Figure 1 plots the pmf of the PNXL distribution.

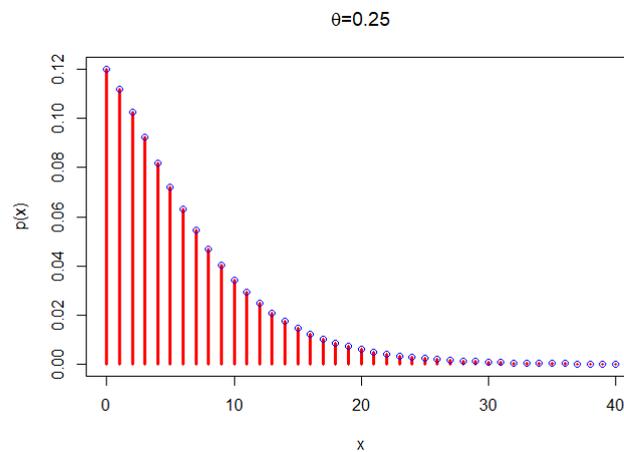


Figure 1. Pmf of the PNXL distribution for  $\theta = 0.25$ .

The survival function (sf) and hazard rate function (hrf) of  $X$  are

$$S(x; \theta) = \frac{2 + 3\theta + \theta x}{2(1 + \theta)^{x+2}}$$

and

$$H(x; \theta) = \frac{\theta(1 + 2\theta + \theta x)}{2 + 3\theta + x\theta},$$

respectively.

### 2.2. Moments, Skewness, and Kurtosis

The probability generating function (pgf) of  $X$  is

$$p(s; \theta) = \frac{\theta(1 - s + 2\theta)}{2(1 - s + \theta)^2}. \tag{2}$$

By replacing  $s$  in (2) with  $e^t$ , the moment-generating function (mgf) of  $X$  is

$$M(t) = \frac{\theta(1 - e^t + 2\theta)}{2(1 - e^t + \theta)^2}. \tag{3}$$

Using (3), we obtain the mean, variance, skewness, and kurtosis of  $X$  as

$$E(X) = \frac{3}{2\theta}$$

and

$$V(X) = \frac{7 + 6\theta}{4\theta^2},$$

$$skew(X) = \frac{36(2\theta^2 + 13\theta - 4)^2}{(7 + 6\theta)^3}$$

and

$$kurt(X) = \frac{333 + 612\theta + 304\theta^2 + 24\theta^3}{(7 + 6\theta)^2},$$

respectively. The dispersion index (DI) is  $1 + \frac{7}{6\theta}$ , which implies that the PNXL distribution is over-dispersed.

We see that moments, mean, variance, skewness, kurtosis, and generating functions are all in closed form. The mean and variance decrease as  $\theta$  increases. The PNXL distribution has positive skewness, which increases as  $\theta$  increases. Kurtosis decreases as  $\theta$  approaches 1, and thereafter, it increases.

### 3. Estimation of Parameters

Various techniques are employed to estimate unknown parameters. We consider the maximum likelihood (ML) method, method of moments (MM), least squares (LS) method, and weighted least squares (WLS) method. We suppose that  $\{x_1, x_2, \dots, x_n\}$  is a random sample of size  $n$  from the PNXL distribution with ordered values  $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ .

#### 3.1. Maximum Likelihood Estimation

The likelihood function is given by

$$L(\theta) = \left(\frac{\theta}{2}\right)^n \left\{ \prod_{i=1}^n \left[ \frac{2\theta + \theta x_i + 1}{(\theta + 1)^{x_i + 2}} \right] \right\}$$

and the log-likelihood function is given by

$$\log L(\theta) = n \log \theta - n \log 2 + \sum_{i=1}^n \log \left\{ \frac{2\theta + \theta x_i + 1}{(\theta + 1)^{x_i + 2}} \right\}.$$

The ML estimate (MLE) of  $\theta$  is obtained by maximizing  $L(\theta)$  or  $\log L(\theta)$  with respect to  $\theta$ . The first derivative of  $\log L(\theta)$  with respect to  $\theta$  is

$$\frac{\partial}{\partial \theta} \log L(\theta) = \frac{n(1 - \theta \bar{x} - \theta)}{\theta(1 + \theta)} + \sum_{i=1}^n \left\{ \frac{x_i + 2}{2\theta + \theta x_i + 1} \right\},$$

where  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ . The MLE of  $\theta$ , denoted by  $\hat{\theta}_{MLE}$ , can be obtained by solving  $\frac{\partial}{\partial \theta} \log L(\theta) = 0$ , provided that the root corresponds to a maximum. We can use the *optim* function in the R software (R 4.2.1) to obtain  $\hat{\theta}_{MLE}$  numerically.

#### 3.2. Method of Moments

The MM estimate (MME) can be obtained by equating theoretical and empirical moments. The MME of  $\theta$ , denoted by  $\hat{\theta}_{MME}$ , is

$$\hat{\theta}_{MME} = \frac{3}{2\bar{x}}.$$

**Proposition 1.** The MME  $\hat{\theta}_{MME}$  has positive bias.

**Proof.** Note that  $\hat{\theta}_{MME} = g(\bar{x})$ , where  $g(t) = \frac{3}{2t}$ ,  $t > 0$ , is strictly convex. Using Jensen’s inequality,  $E(g(\bar{X})) > g(E(\bar{X}))$ , where  $g(E(\bar{X})) = g\left(\frac{3}{2\theta}\right) = \theta$ . Hence,  $\hat{\theta}_{MME}$  has positive bias.  $\square$

3.3. Least Squares and Weighted Least Squares Estimation

The LS estimate (LSE) of  $\theta$ , denoted by  $\hat{\theta}_{LSE}$ , is obtained by minimizing

$$Q(\theta) = \sum_{i=1}^n \left[ F(x_{(i)}) - \frac{i}{n+1} \right]^2.$$

The WLS estimate (WLSE) of  $\theta$ , denoted by  $\hat{\theta}_{WLSE}$ , is obtained by minimizing

$$Q_w(\theta) = \sum_{i=1}^n \frac{(n+1)^2(n+2)}{i(n-i+1)} \left[ F(x_{(i)}) - \frac{i}{n+1} \right]^2.$$

The LSE and WLSE can be evaluated numerically using the *optim* function in the R software (R 4.2.1).

3.4. Simulation Study

This section compares various estimates of  $\theta$  using simulation. The average absolute biases (biases) and mean square errors (MSEs) were calculated for  $\theta = 0.3, 0.5, 1.2$  and  $n = 50, 100, 200, 250, 500$  with replicates  $N = 1000$ :

$$\text{Bias} = \frac{1}{N} \sum_{j=1}^N |\hat{\theta}_j - \theta| \text{ and } \text{MSE} = \frac{1}{N} \sum_{j=1}^N (\hat{\theta}_j - \theta)^2,$$

where  $\hat{\theta}_j$  denotes either the MLEs, MMEs, LSEs, or the WLSEs of  $\theta$ , computed from the  $j$ th sample. Table 1 gives the values of biases and MSEs.

**Table 1.** Simulation results for the PNXL distribution.

$n$	MLE		MME		LSE		WLSE	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\theta = 0.5$								
50	0.065	0.004	0.067	0.004	0.141	0.020	0.158	0.025
100	0.055	0.003	0.052	0.003	0.127	0.016	0.161	0.026
200	0.044	0.002	0.044	0.002	0.126	0.016	0.165	0.027
250	0.008	0.000	0.005	0.000	0.085	0.007	0.145	0.021
500	0.004	0.000	0.002	0.000	0.103	0.011	0.052	0.020
$\theta = 0.3$								
50	0.034	0.001	0.037	0.001	0.087	0.008	0.074	0.006
100	0.013	0.000	0.015	0.000	0.058	0.003	0.059	0.004
200	0.008	0.000	0.008	0.000	0.050	0.003	0.060	0.004
250	0.007	0.000	0.008	0.000	0.032	0.001	0.045	0.002
500	0.001	0.000	0.001	0.000	0.003	0.001	0.035	0.001

**Table 1.** Cont.

n	MLE		MME		LSE		WLSE	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\theta = 1.2$								
50	0.107	0.011	0.108	0.012	0.511	0.261	0.635	0.403
100	0.048	0.002	0.046	0.002	0.485	0.235	0.611	0.373
200	0.046	0.002	0.046	0.002	0.485	0.235	0.642	0.412
250	0.007	0.005	0.006	0.000	0.471	0.222	0.645	0.416
500	0.004	0.000	0.006	0.001	0.483	0.234	0.560	0.314
$\theta = 1.5$								
50	0.055	0.003	0.052	0.003	0.684	0.468	0.897	0.804
100	0.030	0.001	0.029	0.001	0.677	0.458	0.868	0.754
200	0.025	0.001	0.026	0.001	0.689	0.475	0.880	0.775
250	0.021	0.000	0.025	0.001	0.699	0.489	0.864	0.747
500	0.020	0.000	0.021	0.002	0.666	0.444	0.889	0.790

We can see that MLE and MME perform almost equally well. For large values of  $\theta$ , LSE and WLSE do not perform well. For MLEs, there is a noticeable decline in both absolute bias and MSE as the sample size increases. Consequently, the performance of MLE proves to be consistently reliable.

**4. The INAR(1) Process with PNXL Innovations**

According to [11], as an innovation for INAR(1) processes for over-dispersed count data, we employ the PNXL distribution, which is suitable for over-dispersed data. The INAR(1) process is given by

$$X_t = \alpha \circ X_{t-1} + \epsilon_t, \quad t \in \mathbb{Z},$$

where  $\alpha \in [0, 1)$ ,  $\{\epsilon_t\}_{t \in \mathbb{Z}}$  is a sequence of iid nonnegative integer-valued random variables from the PNXL distribution with mean  $E(\epsilon_t) = \mu_\epsilon$  and variance  $V(\epsilon_t) = \sigma_\epsilon^2$ . The binomial thinning operator denoted by ‘ $\circ$ ’ is defined as

$$\alpha \circ X_{t-1} = \sum_{j=1}^{X_{t-1}} W_j,$$

where  $\{W_j\}_{j \geq 1}$  is a sequence of iid Bernoulli random variables with probability of success  $p$ . The one-step transition probability matrix for the INAR(1) process is defined by

$$\Pr(X_t = k | X_{t-1} = l) = \sum_{i=1}^{\min(k,l)} \binom{l}{i} \alpha^i (1 - \alpha)^{l-i} \Pr(\epsilon_t = k - i), \quad k, l \geq 0.$$

PNXL innovations are used to propose a new INAR(1) process for over-dispersed data. Let  $\{\epsilon_t\}_{t \in \mathbb{Z}}$  follow the PNXL distribution. Then, the one-step transition probability matrix of the corresponding process is

$$\Pr(X_t = k | X_{t-1} = l) = \sum_{i=1}^{\min(k,l)} \binom{l}{i} \alpha^i (1 - \alpha)^{l-i} \left\{ \frac{\theta[2\theta + \theta(k - i) + 1]}{2(\theta + 1)^{k-i+2}} \right\}, \quad k, l \geq 0.$$

This new process is denoted by INAR(1)PNXL. We can obtain the joint probability function as

$$\begin{aligned} f(i_1, i_2, \dots, i_n) &= \Pr(X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) \\ &= \Pr(X_1 = i_1) \Pr(X_2 = i_2 | X_1 = i_1) \cdots \Pr(X_n = i_n | X_{n-1} = i_{n-1}) \\ &= \Pr(X_1 = i_1) \prod_{k=1}^{n-1} \left[ \sum_{m=0}^{\min(i_k, i_{k+1})} \binom{i_k}{m} \alpha^m (1 - \alpha)^{i_k - m} \Pr(\epsilon_{k+1} = i_{k+1} - m) \right]. \end{aligned}$$

The (conditional or unconditional) mean/variance, DI, and autocovariance/autocorrelation (ACF/PACF) at lag  $k$  of  $\{X_t\}_{t \in \mathbb{Z}}$  [23] are

$$E(X_t | X_{t-1}) = \alpha X_{t-1} + \mu_\epsilon = \alpha X_{t-1} + \frac{3}{2\theta}, \tag{4}$$

$$V(X_t | X_{t-1}) = \alpha(1 - \alpha)X_{t-1} + \sigma_\epsilon^2 = \alpha(1 - \alpha)X_{t-1} + \frac{7 + 6\theta}{4\theta^2}, \tag{5}$$

$$E(X_t) = \frac{\mu_\epsilon}{1 - \alpha} = \frac{3}{2\theta(1 - \alpha)},$$

$$V(X_t) = \frac{\sigma_\epsilon^2 + \alpha\mu_\epsilon}{1 - \alpha^2} = \frac{7 + 6(1 + \alpha)\theta}{4(1 - \alpha^2)\theta^2},$$

$$DI(X_t) = \frac{DI_\epsilon + \alpha}{1 + \alpha} = \frac{1 + \frac{7}{6\theta} + \alpha}{1 - \alpha},$$

$$\gamma_k = Cov(X_k, X_{k+1}) = \alpha^k V(X_t)$$

and

$$\rho_k = Corr(X_k, X_{k+1}) = \alpha^k,$$

respectively.

#### 4.1. Estimation of INAR(1)PNXL Process

We utilize the conditional maximum likelihood (CML), conditional least squares (CLS), and Yule–Walker (YW) methods. Let  $\{x_1, \dots, x_T\}$  be the observed count time series of length  $T$ .

##### 4.1.1. Conditional Maximum Likelihood

The conditional log likelihood function of the INAR(1) process is

$$\begin{aligned} l(\alpha, \theta) &= \sum_{t=2}^T \log[\Pr(X_t = k | X_{t-1} = l)] \\ &= \sum_{t=2}^T \log \left\{ \sum_{i=1}^{\min(x_t, x_{t-1})} \binom{x_{t-1}}{i} \alpha^i (1 - \alpha)^{x_{t-1} - i} \frac{\theta[2\theta + \theta(k - i) + 1]}{2(\theta + 1)^{k - i + 2}} \right\}. \end{aligned} \tag{6}$$

The CML estimates of  $\alpha$  and  $\theta$ , denoted by  $\hat{\alpha}_{CML}$  and  $\hat{\theta}_{CML}$ , respectively, can be obtained numerically by maximizing (6) with respect to  $\alpha$  and  $\theta$ .

##### 4.1.2. Yule–Walker

The YW estimates of  $\alpha$  and  $\theta$ , denoted by  $\hat{\alpha}_{YW}$  and  $\hat{\theta}_{YW}$ , respectively, can be computed by equating theoretical and empirical moments of the INAR(1)PNXL process, as follows:

$$\widehat{\alpha}_{YW} = \frac{\sum_{t=2}^T (x_t - \bar{x})(x_{t-1} - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2}$$

and

$$\widehat{\theta}_{YW} = \frac{3}{2(1 - \widehat{\alpha}_{YW})\bar{x}}, \tag{7}$$

where  $\bar{x} = \frac{1}{N} \sum_{t=1}^T x_t$ .

### 4.1.3. Conditional Least Squares

The CLS estimates of  $\alpha$  and  $\theta$ , denoted by  $\widehat{\alpha}_{CLS}$  and  $\widehat{\theta}_{CLS}$ , respectively, can be obtained by minimizing

$$\begin{aligned} Q(\eta) &= \sum_{t=2}^T [X_t - E(X_t|X_{t-1})]^2 \\ &= \sum_{t=2}^T \left( X_t - \alpha X_{t-1} - \frac{3}{2\theta} \right)^2, \end{aligned} \tag{8}$$

as follows

$$\widehat{\alpha}_{CLS} = \frac{(T-1) \sum_{t=2}^T X_t X_{t-1} - \sum_{t=2}^T X_t \sum_{t=2}^T X_{t-1}}{(T-1) \sum_{t=2}^T X_{t-1}^2 - \left( \sum_{t=2}^T X_{t-1} \right)^2}$$

and

$$\widehat{\theta}_{CLS} = \frac{3(T-1)}{2 \left( \sum_{t=2}^T X_t - \widehat{\alpha}_{CLS} \sum_{t=2}^T X_{t-1} \right)}.$$

### 4.2. Simulation of INAR(1)PNXL Process

A simulation study was carried out to assess the performances of CML, CLS, and YW estimates. The biases and MSEs were calculated for the three estimates for  $\alpha = 0.4, 0.8$ ,  $\theta = 0.8, 3$ , and  $n = 50, 100, 200, 250, 500$  with replication  $N = 1000$ . The results are given in Table 2.

**Table 2.** Simulation results for the INAR(1)PNXL process.

Parameter	n	$\alpha = 0.4$ and $\theta = 0.8$					
		CML		CLS		YW	
		Bias	MSE	Bias	MSE	Bias	MSE
$\alpha$	50	0.063	0.006	0.109	0.019	0.110	0.020
	100	0.044	0.003	0.080	0.010	0.081	0.010
	200	0.032	0.002	0.054	0.005	0.053	0.005
	250	0.029	0.001	0.049	0.004	0.049	0.004
	500	0.019	0.001	0.035	0.002	0.035	0.002
$\theta$	50	0.130	0.029	0.164	0.044	0.162	0.043
	100	0.094	0.015	0.122	0.025	0.122	0.025
	200	0.063	0.007	0.084	0.012	0.083	0.012
	250	0.058	0.005	0.078	0.010	0.078	0.010
	500	0.041	0.003	0.056	0.005	0.056	0.005

Table 2. Cont.

Parameter	n	$\alpha = 0.4$ and $\theta = 0.8$					
		CML		CLS		YW	
		Bias	MSE	Bias	MSE	Bias	MSE
$\alpha = 0.8$ and $\theta = 3$							
$\alpha$	50	0.041	0.003	0.098	0.017	0.105	0.019
	100	0.028	0.001	0.061	0.007	0.065	0.008
	200	0.022	0.001	0.047	0.004	0.049	0.004
	250	0.018	0.001	0.036	0.002	0.036	0.002
	500	0.012	0.000	0.025	0.001	0.025	0.001
$\theta$	50	0.745	0.978	1.024	1.722	1.008	1.691
	100	0.512	0.455	0.764	0.923	0.761	0.925
	200	0.391	0.241	0.648	0.652	0.655	0.665
	250	0.299	0.148	0.499	0.405	0.500	0.407
	500	0.212	0.070	0.377	0.223	0.377	0.222

Biases and MSEs of the CML estimate tend to zero more quickly than those of YW and CLS estimates, making them effective for both small and large sample sizes.

### 5. Data Analysis

In this section, two real datasets are analyzed using the PNXL distribution.

#### 5.1. Corn Borer Data

Corn borer data are biological experiment data representing the number of European corn borer larvae pyrausta in a field (see [24]). This dataset is taken to compare the performance of the PNXL distribution with the discrete Burr (DB) distribution [25], the discrete Pareto (DP) distribution [25], the discrete inverse Weibull distribution [26], the COM-Poisson (CMP) distribution [3], the discrete Gumbel (DG) distribution [27], the discrete inverse Rayleigh (DIR) distribution [28], the discrete log-logistic (DLL) distribution [29], and the discrete Bilal (DBL) distribution [15].

These distributions were compared using the Akaike information criterion (AIC) and Bayesian information criterion (BIC). Moreover, a  $\chi^2$  test and its  $p$ -value were used to determine the goodness of fit of each fitted distribution. The MLEs with their corresponding standard errors (SEs) and confidence intervals (CIs) (lower bound of CI, upper bound of CI) are provided in Table 3.

Table 3. Corn borer data: MLEs, SEs, and CIs.

Statistic	PNXL	DIW	DG	DLL	DB	DIR	DBL	DP	CMP	
$MLE_{\theta}$	1.012	0.345	3.106	1.943	2.357	0.320	0.657	0.329	0.672	
$SE_{\theta}$	0.111	0.043	0.367	0.188	0.366	0.042	0.019	0.034	0.090	
95% CI	lower	0.794	0.261	2.388	1.575	1.641	0.237	0.620	0.263	0.496
	upper	1.230	0.429	3.825	2.311	3.073	0.402	0.693	0.395	0.847
$MLE_{\beta}$	-	1.541	0.407	1.401	0.519	-	-	-	0.107	
$SE_{\beta}$	-	0.156	0.029	0.121	0.051	-	-	-	0.116	
95% CI	lower	-	1.235	0.349	1.163	0.419	-	-	-	0.121
	upper	-	1.847	0.464	1.638	0.619	-	-	-	0.334

Table 4 shows that the PNXL distribution gives the best fit as it gives the lowest AIC, the lowest BIC, and the highest  $p$ -value along with observed frequencies (of).

**Table 4.** Corn borer data:  $\log L$ ,  $\chi^2$ -value,  $p$ -value, AIC, and BIC for the competitive models.

X	Of	PNXL	DIW	DG	DLL	DB	DIR	DBL	DP	CMP
0	43	45.355	41.370	28.553	41.032	43.836	38.352	32.734	64.447	44.995
1	35	30.088	41.850	37.861	38.938	39.601	51.874	39.586	20.149	30.221
2	17	18.705	15.420	25.585	17.775	15.622	15.489	24.277	9.686	18.855
3	11	11.161	7.170	12.852	8.432	7.206	6.028	12.508	5.647	11.266
4	5	6.474	3.940	5.700	4.485	3.910	2.905	5.970	3.681	6.529
5	4	3.678	2.420	2.402	2.630	2.376	1.610	2.738	2.580	3.695
6	1	2.057	1.610	0.991	1.663	1.563	0.981	1.227	1.904	2.051
7	2	1.136	1.130	0.405	1.115	1.089	0.641	0.542	1.461	1.120
8	2	1.347	5.090	5.651	3.930	4.798	2.120	0.420	10.446	1.271
Total	120	120	120	120	120	120	120	120	120	120
$\log L$	-	200.432	204.810	231.191	202.630	204.293	208.440	204.675	220.618	200.415
AIC	-	402.863	413.621	430.382	409.261	412.587	418.881	411.351	443.236	404.830
BIC	-	405.651	419.195	435.957	414.836	418.162	421.668	414.138	446.024	410.405
$\chi^2$	-	1.115	5.511	7.615	1.311	2.674	14.295	6.996	30.518	1.063
df	-	3	3	2	2	2	3	3	3	2
$p$ -value	-	0.774	0.138	0.022	0.519	0.263	0.003	0.072	0.000	0.588

5.2. Weekly Number of Syphilis Cases Data

Weekly number of syphilis cases data, available in the *tsinteger* package of the R software, were fitted to the INAR(1)PNXL process. The effectiveness of this process was evaluated against the INAR(1)P process [30], the INAR(1)G process [12], the INAR(1)ZIP process [31], and the INAR(1)PWE process [17]. The dataset has a mean of 24.632 and a variance of 105.676, which shows significant over-dispersion.

The Pearson residuals are employed in residual analysis to assess statistical precision of the fitted INAR(1)PNXL process. These were calculated using

$$r_t = \frac{x_t - E(x_t|X_{t-1} = x_{t-1})}{V(x_t|X_{t-1} = x_{t-1})^{\frac{1}{2}}}$$

where  $E(x_t|X_{t-1} = x_{t-1})$  and  $V(x_t|X_{t-1} = x_{t-1})$  are given in (4) and (5), respectively. When the fitted INAR(1) process was statistically valid, the Pearson residuals had to be uncorrelated and should have zero mean and unit variance. The Pearson residuals are evaluated for correlation by generating a plot of their ACF. The randomness of the INAR(1)PNXL process can be examined by plotting cumulative periodograms (cpgrams) of the Pearson residuals for the series under consideration.

The ACF plot, partial ACF (PACF) plot, histogram, and time series plot of the data are shown in Figure 2. Only the first lag is noticeable in the PACF plot. So, the INAR(1) process could be a viable process for these data. The results of the INAR(1) process fitted to the data are shown in Table 5, together with parameter estimates, SEs, AICs, BICs, theoretical means, variances, and DIs.

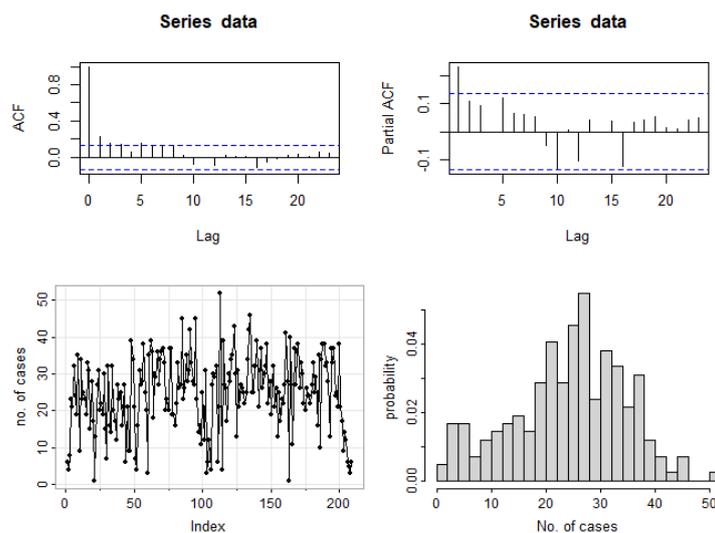


Figure 2. ACF plot, PACF plot, time series plot, and histogram of weekly number of syphilis cases data.

Table 5. Estimates and model adequacy statistics of the fitted models for the number of syphilis cases data.

Model	Parameters	Estimate	S.E.	AIC	BIC	$\mu$	$\sigma^2$	DI
INAR(1)PNXL	$\alpha$	0.316	0.034	1660.869	1667.554	23.943	255.917	10.689
	$\theta$	0.092	0.007					
INAR(1)P	$\alpha$	0.148	0.026	2016.534	2023.224	25.349	25.349	1.000
	$\lambda$	21.063	0.709					
INAR(1)G	$\alpha$	0.347	0.032	1686.428	1693.112	23.895	252.431	10.564
	$\lambda$	0.058	0.005					
INAR(1)PWE	$\alpha$	0.058	0.159	1688.428	1698.455	24.990	369.211	14.774
	$\lambda$	0.060	2.883					
	$\beta$	0.347	0.032					
INAR(1)ZIP	$\alpha$	20.552	0.595	1732.296	1742.323	25.332	58.543	2.307
	$\lambda$	0.113	0.024					
	$\beta$	0.262	0.024					

The INAR(1)PNXL process offers a better fit than other INAR(1) processes as it gives the lowest AIC and lowest BIC values. The accuracy of the fitted INAR(1)PNXL process was assessed using standardized Pearson residuals. Figure 3 presents the ACF for the Pearson residuals, revealing the absence of autocorrelation. To confirm this, a Ljung–Box test was conducted with 10 degrees of freedom, resulting in a  $p$ -value of 0.1119, which is greater than 0.05. This test unequivocally establishes the lack of correlation among the residuals, providing strong evidence for the accuracy and excellent fit of the INAR(1)PNXL process to the weekly number of syphilis cases dataset. Figure 4 shows that the INAR(1)PNXL process is random for the weekly number of syphilis cases data.

The INAR(1)PNXL model for the weekly number of syphilis cases data is given by

$$X_t = 0.316X_{t-1} + \epsilon_t,$$

where  $\epsilon_t \sim PNXL(0.092)$ . The predicted values for the weekly number of syphilis cases data obtained by the INAR(1)PNXL process are the following:

$$\widehat{X}_1 = E(X_t)_{\widehat{\theta}_{cml}} = 23.943,$$

$$\widehat{X}_t = E(X_t|X_{t-1})_{\widehat{\theta}_{cml}} = 0.316X_{t-1} + 16.388, \quad t = 2, 3, \dots, n.$$

Figure 5 plots the predicted versus the original values of the weekly number of syphilis cases.

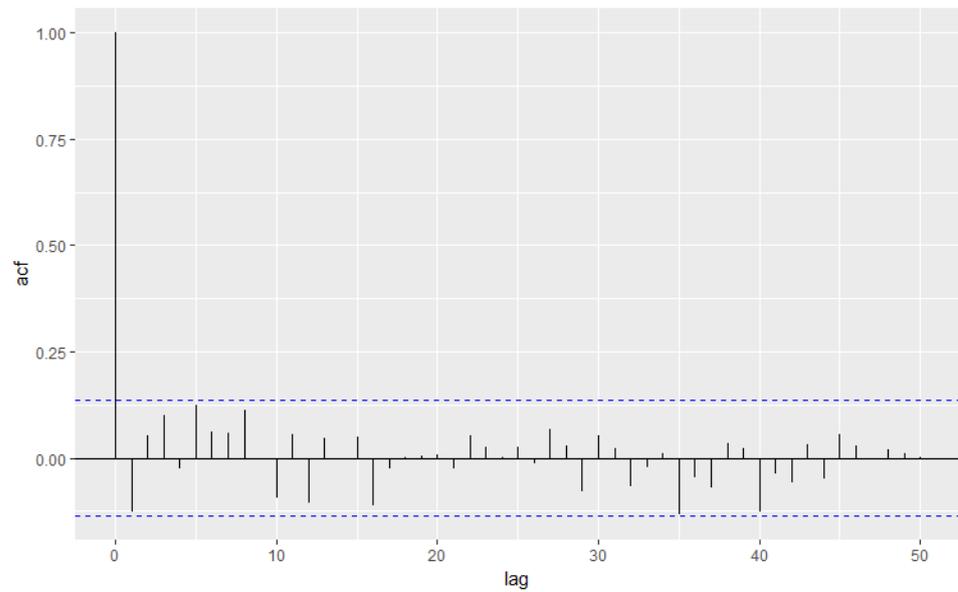


Figure 3. The ACF plot of the Pearson residuals.

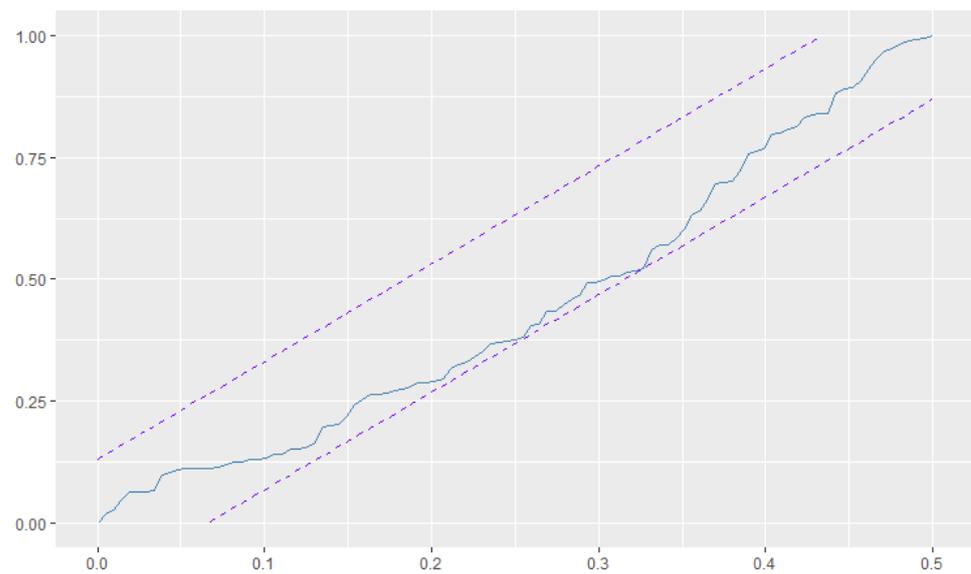
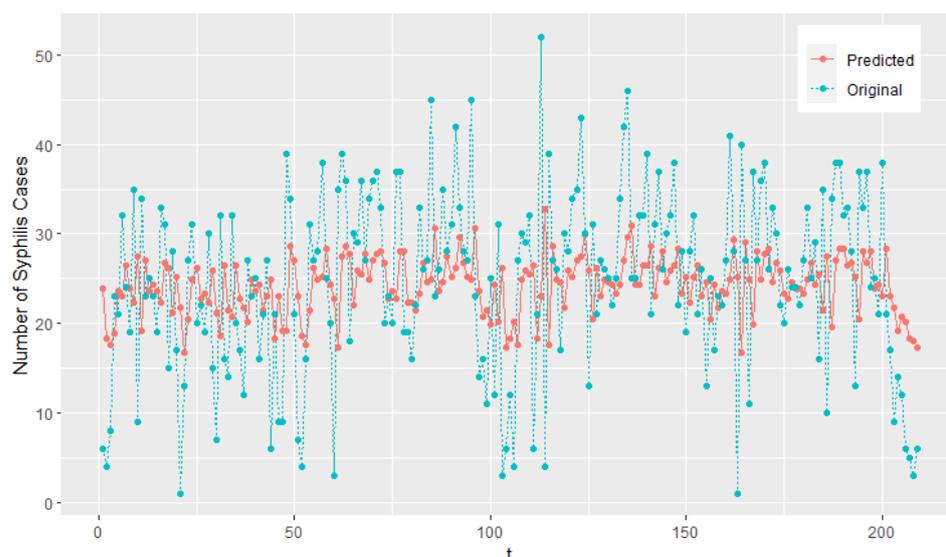


Figure 4. The cpgrams of the Pearson residuals of the weekly number of syphilis cases data.



**Figure 5.** The predicted versus the original values of the weekly number of syphilis cases data.

## 6. Conclusions

The PNXL distribution, a one-parameter discrete compound distribution capable of modeling data with over-dispersion, was proposed in this paper. Various probabilistic and statistical aspects, almost all of which have closed forms, show how adaptable and straightforward the one-parameter distribution is. Various methods were used to estimate its parameter. Simulation studies showed that ML and MM methods performed equally well in finite samples. Also, a new INAR(1)PNXL model was proposed. The better performance of the PNXL distribution or the INAR(1)PNXL model was illustrated using two real datasets, which was superior to several existing two-parameter models.

**Author Contributions:** Methodology, M.R.I., S.A., R.M. and S.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data can be obtained from the corresponding author.

**Acknowledgments:** The authors would like to thank the editor and the three referees for careful reading and comments which greatly improved the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Bliss, C.I.; Fisher, R.A. Fitting the negative binomial distribution to biological data. *Biometrics* **1953**, *9*, 176–200. [[CrossRef](#)]
2. Bereta, E.M.; Louzanda, F.; Franco, M.A. The Poisson-Weibull distribution. *Adv. Appl. Stat.* **2011**, *22*, 107–118.
3. Sellers, K.F.; Borle, S.; Shmueli, G. The COM-Poisson model for count data: A survey of methods and applications. *Appl. Stoch. Model. Bus. Ind.* **2012**, *28*, 104–116. [[CrossRef](#)]
4. Abd El-Monsef, M.; Sohsah, N. Poisson-transmuted Lindley distribution. *J. Adv. Math.* **2016**, *11*, 5631–5638. [[CrossRef](#)]
5. Bhati, D.; Kumawat, P.; Gómez-Déniz, E. A new count model generated from mixed Poisson transmuted exponential family with an application to health care data. *Commun. Stat. Theory Methods* **2017**, *46*, 11060–11076. [[CrossRef](#)]
6. Grine, R.; Zeghdoudi, H. On Poisson quasi-Lindley distribution and its applications. *J. Mod. Appl. Stat. Methods* **2017**, *16*, 21. [[CrossRef](#)]
7. Altun, E. A new one-parameter discrete distribution with associated regression and integer-valued autoregressive models. *Math. Slovaca* **2020**, *70*, 979–994. [[CrossRef](#)]
8. Altun, E.; Cordeiro, G.M.; Ristić, M.M. An one-parameter compounding discrete distribution. *J. Appl. Stat.* **2022**, *49*, 1935–1956. [[CrossRef](#)]
9. Maya, R.; Chesneau, C.; Krishna, A.; Irshad, M.R. Poisson Extended Exponential Distribution with Associated INAR(1) Process and Applications. *Stats* **2022**, *5*, 755–772. [[CrossRef](#)]

10. Irshad, M.; D’cruz, V.; Maya, R.; Mamode Khan, N. Inferential properties with a novel two parameter Poisson generalized Lindley distribution with regression and application to INAR(1) process. *J. Biopharm. Stat.* **2023**, *33*, 335–356. [[CrossRef](#)]
11. Al-Osh, M.A.; Alzaid, A.A. First-order integer-valued autoregressive (INAR(1)) process. *J. Time Ser. Anal.* **1987**, *8*, 261–275. [[CrossRef](#)]
12. Aghababaei Jazi, M.; Jones, G.; Lai, C.D. Integer valued AR(1) with geometric innovations. *J. Iran. Stat. Soc.* **2012**, *11*, 173–190.
13. Eliwa, M.S.; Altun, E.; El-Dawoody, M.; El-Morshedy, M. A new three-parameter discrete distribution with associated INAR(1) process and applications. *IEEE Access* **2020**, *8*, 91150–91162. [[CrossRef](#)]
14. Huang, J.; Zhu, F. A new first-order integer-valued autoregressive model with Bell innovations. *Entropy* **2021**, *23*, 713. [[CrossRef](#)] [[PubMed](#)]
15. Altun, E.; El-Morshedy, M.; Eliwa, M. A study on discrete Bilal distribution with properties and applications on integer valued autoregressive process. *REVSTAT-Stat. J.* **2022**, *20*, 501–528.
16. Lívio, T.; Khan, N.M.; Bourguignon, M.; Bakouch, H.S. An INAR(1) model with Poisson–Lindley innovations. *Econ. Bull.* **2018**, *38*, 1505–1513.
17. Altun, E. A new generalization of geometric distribution with properties and applications. *Commun. Stat. Simul. Comput.* **2020**, *49*, 793–807. [[CrossRef](#)]
18. Altun, E.; Bhati, D.; Khan, N.M. A new approach to model the counts of earthquakes: INARPQX(1) process. *SN Appl. Sci.* **2021**, *3*, 1–17. [[CrossRef](#)]
19. Irshad, M.R.; Chesneau, C.; D’cruz, V.; Maya, R. Discrete pseudo Lindley distribution: Properties, estimation and application on INAR(1) process. *Math. Comput. Appl.* **2021**, *26*, 76. [[CrossRef](#)]
20. Altun, E.; Khan, N.M. Modelling with the novel INAR(1)-PTE process. *Methodol. Comput. Appl. Probab.* **2022**, *24*, 1–17. [[CrossRef](#)]
21. Nawel, K.; Gemeay, A.M.; Zeghdoudi, H.; Karakaya, K.; Alshangiti, A.M.; Bakr, M.; Balogun, O.S.; Muse, A.H.; Hussam, E. Modelling Voltage Real Dataset by a New Version of Lindley Distribution. *IEEE Access* **2023**, *11*, 67220–67229.
22. Beghriche, A.; Zeghdoudi, H.; Raman, V.; Chouia, S. New polynomial exponential distribution: Properties and applications. *Stat. Transit. New Ser.* **2022**, *23*, 95–112. [[CrossRef](#)]
23. Weiß, C.H. *An Introduction to Discrete-Valued Time Series*; John Wiley & Sons: Hoboken, NJ, USA, 2018.
24. Bodhisuwan, W.; Sangpoom, S. The discrete weighted Lindley distribution. In Proceedings of the 2016 12th International Conference on Mathematics, Statistics, and Their Applications, ICMSA, Banda Aceh, Indonesia, 4–6 October 2016.
25. Krishna, H.; Pundir, P.S. Discrete Burr and discrete Pareto distributions. *Stat. Methodol.* **2009**, *6*, 177–188. [[CrossRef](#)]
26. Jazi, M.A.; Lai, C.D.; Alamatsaz, M.H. A discrete inverse Weibull distribution and estimation of its parameters. *Stat. Methodol.* **2010**, *7*, 121–132. [[CrossRef](#)]
27. Chakraborty, S.; Chakravarty, D. A Discrete Gumbel Distribution. *arXiv* **2014**, arXiv:1410.7568.
28. Hussain, T.; Ahmad, M. Discrete inverse Rayleigh distribution. *Pak. J. Stat.* **2014**, *30*, 203–222.
29. Para, B.A.; Jan, T.R. Discrete version of log-logistic distribution and its applications in genetics. *Int. J. Mod. Math. Sci.* **2016**, *14*, 407–422.
30. McKenzie, E. Some simple models for discrete variate time series 1. *J. Am. Water Resour. Assoc.* **1985**, *21*, 645–650. [[CrossRef](#)]
31. Jazi, M.A.; Jones, G.; Lai, C.D. First-order integer valued AR processes with zero inflated Poisson innovations. *J. Time Ser. Anal.* **2012**, *33*, 954–963. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.