REVSTAT – Statistical Journal
Volume 20, Number 4, July 2022, 501–528
https://doi.org/10.57805/revstat.v20i4.384

A Study on Discrete Bilal Distribution with Properties and Applications on Integer-Valued Autoregressive Process

Authors: Emrah Altun ✉
– Department of Mathematics, Bartin University, Bartin 74100, Turkey
emrahaltun@bartin.edu.tr

M. El-Morshedy
– Department of Mathematics, College of Science and Humanities, in Al-Kharj Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia
m.elmorshedy@psau.edu.sa
– Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt
mah_elmorshedy@mans.edu.eg

M. S. Eliwa
– Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt
mseliwa@mans.edu.eg
– Department of Statistics and Operation Research, College of Science, Qassim University, P.O. Box 6644, Buraydah 51482, Saudi Arabia
m.eliwa@qu.edu.sa

Received: December 2019 Revised: November 2020 Accepted: November 2020

Abstract:

- This study proposes a new one-parameter discrete distribution, called a discrete Bilal distribution. The structural properties of the proposed distribution, including the shape of the probability mass function, mode, moments, skewness, kurtosis, index of dispersion, mean deviation, stress-strength reliability, and order statistics, are derived. These properties are expressed in closed-forms. The maximum likelihood and method of moments estimation methods are considered to estimate the unknown model parameter. An extensive simulation study is carried out to examine the finite sample performance of estimation methods. The usefulness of the proposed model is illustrated in the first-order integer-valued autoregressive process. The empirical importance of the proposed models is proved through three real data applications.

Keywords:

- *Bilal distribution; INAR(1) process; method of moments; maximum likelihood; simulation.*

AMS Subject Classification:

- 60E05, 62E10, 62E15, 62F10, 62N05.

✉ Corresponding author.

1. INTRODUCTION

The count data sets arise in different fields such as yearly number destructive earthquakes, monthly traffic accidents and hourly bacterial growth and among others. These kind of data sets are modeled with discrete probability distributions. Poisson and negative-binomial distributions are the most popular distributions and are widely used to model these kind data sets. In recent years, researchers have shown great interest to introduce new discrete distributions by discretizing a continuous failure time model. Let the continuous random variable X has the survival function (sf) $S(x) = \Pr(X > x)$. The probability mass function (pmf) dealing with the continuous random variable X is given by

$$\Pr(X = x) = S(x) - S(x+1), \quad x = 0, 1, 2, ...$$

Many researchers have introduced sophisticated discrete distributions by applying the discretization method to the continuous failure time models. For instance, discrete Lindley distribution by Gómez-Déniz and Calderín-Ojeda (2011) [12], discrete Rayleigh distribution by Roy (2004) [28], discrete inverse Rayleigh distribution by Hussain and Ahmad (2014) [13], discrete Pareto distribution by Buddana and Kozubowski (2014) [6], discrete Weibull distribution by Nakagawa and Osaki (1975) [21], discrete Lomax distribution by Para and Jan (2016a) [24], discrete generalized Weibull distribution by Para and Jan (2017) [26] and exponentiated discrete Lindley by El-Morshedy *et al.* (2019) [10], discrete flexible one parameter distribution by Eliwa and El-Morshedy (2020) [7] and discrete gompertz-G by Eliwa *et al.* (2020a) [8] and among others. The discrete analogue of the Burr-Hatke distribution was introduced by El-Morshedy *et al.* (2020) [11] with its regression model and residual analysis. More recently, Eliwa *et al.* (2020b) [9] introduced the discrete analogue of the three-parameter Lindley distribution and demonstrated its performance in modeling the time series of counts.

In this paper, we introduce a new one-parameter discrete distribution by applying the discretization method to the Bilal distribution, proposed by Abd-Elrahman (2013) [4]. The arising distribution is called as the discrete Bilal (DBL) distribution. The DBL distribution has simple probability mass and cumulative distribution functions and statistical properties such as mean, mode, skewness, kurtosis measures, mean deviation and also stress-strength reliability are obtained in explicit forms. The DBL distribution provides an opportunity to model different types of the count data sets such over and under-dispersed. We illustrate the importance of DBL distribution in first-order integer-valued autoregressive (INAR(1)) process by applying the DBL distribution as an innovation process of INAR(1) process, introduced by McKenzie (1985) [20] and Al-Osh and Alzaid (1987) [1]. INAR(1) process is widely used to model time series of counts. Several researchers have done important studies on the INAR(1) processes with more flexible innovation distributions. For instance, Jazi *et al.* (2012) [14] introduced the INAR(1) process with geometric innovations (INAR(1)G) to model the over-dispersed time series of counts. Similarly, Lívio *et al.* (2018) [19] introduced the INAR(1) process with Poisson-Lindley innovations (INAR(1)PL) for over-dispersed time series of counts. More recently, Altun (2020a) [2] introduced a new generalization of the geometric and demonstrated its performance in INAR(1) process. More recently, Altun (2020b) [3] introduced a mixed Poisson distribution and defined a new INAR(1) process for over-dispersed time series of counts.

The remaining parts of the presented study is organized as follows. The statistical properties of the DBL distribution are obtained in Section 2. The parameter estimation of the DBL distribution is discussed in Section 3. The INAR(1) process with DBL innovations is introduced in Section 4 with its parameter estimation. In Section 5, we discuss the finite-sample performance of the parameter estimation methods via two simulation studies. In Section 6, three data sets are analyzed with DBL and other competitive models to prove the importance of the DBL distribution practically. Section 7 deals with the concluding remarks of the study.

2. THE DISCRETE-BILAL DISTRIBUTION

Recently, Abd-Elrahman (2013) [4] proposed a new flexible model, called Bilal (BL) distribution. The cumulative distribution function (cdf) of the BL distribution is

$$\Pi(x;\beta) = 1 - \left(3 - 2e^{-\frac{x}{\beta}}\right) e^{-\frac{2x}{\beta}}, \; x \geq 0, \beta > 0. \tag{2.1}$$

The sf and probability density function (pdf) of (2.1) are given, respectively, by

$$S(x;\beta) = \left(3 - 2e^{-\frac{x}{\beta}}\right) e^{-\frac{2x}{\beta}}, \; x \geq 0, \beta > 0, \tag{2.2}$$

$$\pi(x;\beta) = \frac{6}{\beta}\left(1 - e^{-\frac{x}{\beta}}\right) e^{-\frac{2x}{\beta}}, \; x \geq 0, \beta > 0. \tag{2.3}$$

Now, we introduce a DBL distribution by discretizing the sf of the BL distribution. Let the parameter $p = e^{-\frac{1}{\beta}}$, the cdf of DBL distribution is given by

$$F(x;p) := F(X \leq x) = 1 - \left(3 - 2p^{x+1}\right) p^{2(x+1)}, \; x = 0, 1, 2, 3, ... \tag{2.4}$$

The corresponding sf and pmf to (2.4) are given, respectively, by

$$S(x;p) = \left(3 - 2p^{x+1}\right) p^{2(x+1)}, \tag{2.5}$$

and

$$f(x;p) := P(X = x) = 2(p^3 - 1)p^{3x} - 3(p^2 - 1)p^{2x}, \;\; x = 0, 1, 2, 3, ... \tag{2.6}$$

The pmf in (2.6) is log-concave for all values of p, where

$$\frac{f(x+1;p)}{f(x;p)} = \frac{2p^{x+6} - 2p^{x+3} - 3p^4 + 3p^2}{2p^{x+3} - 3p^2 - 2p^x + 3} \tag{2.7}$$

is a decreasing function in x for all value of p. The possible pmf shapes of the DBL distribution are displayed in Figure 1. These figures show that the DBL distribution has right-skewed shapes and it has long right-tails.

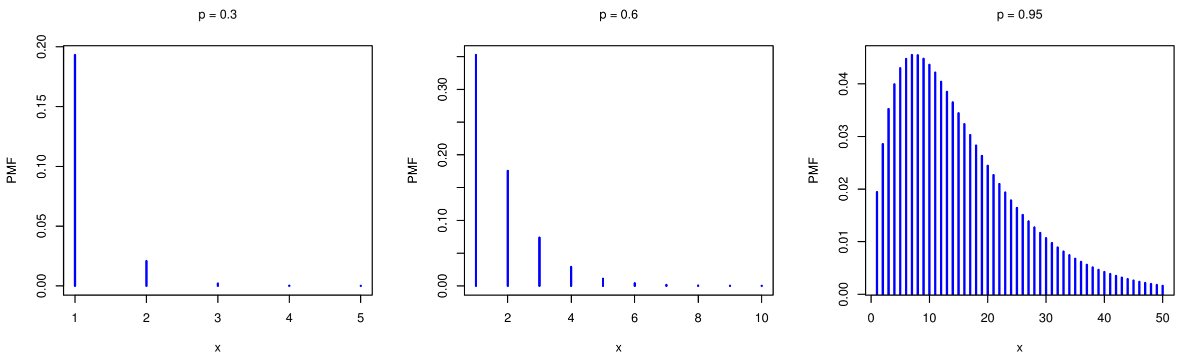


Figure 1: The pmf plots of the DBL distribution.

The hazard rate function (hrf) is

$$h(x;p) = \frac{2(p^3-1)p^x - 3(p^2-1)}{3-2p^x}, \quad x \in \mathbb{N}_0, \tag{2.8}$$

where $h(x;p) = \frac{f_x(x;p)}{R(x-1;p)}$. The reversed hazard rate function (rhrf) is

$$r(x;p) = \frac{2(p^3-1)p^{3x} - 3(p^2-1)p^{2x}}{1-(3-2p^{x+1})\,p^{2(x+1)}}, \quad x \in \mathbb{N}_0, \tag{2.9}$$

where $r(x;p) = \frac{f_x(x;p)}{F(x;p)}$. Figure 2 shows the hrf and rhrf plots for different values of the parameter p.

It is clear that the hrf of the DBL distribution increases up to time t where $0<t<x<\infty$, whereas the hrf is constant after time t. Regarding to the rhrf, it is seen that it always decreases for all x.

Suppose X_1 and X_2 are two independent random variables following the DBL distribution with the parameters p_1 and p_2, respectively. Let $W = \min(X_1, X_2)$ be a random variable which has a hrf

$$\begin{aligned} h_W(x;p_1,p_2) &= \frac{P(\min(X_1,X_2)\geq x) - P(\min(X_1,X_2)\geq x+1)}{P(\min(X_1,X_2)\geq x)} \\ &= \frac{2(p_1^3-1)p_1^x - 3(p_1^2-1)}{3-2p_1^x} + \frac{2(p_2^3-1)p_2^x - 3(p_2^2-1)}{3-2p_2^x} \\ &\quad - \frac{\left\{2(p_1^3-1)p_1^x - 3(p_1^2-1)\right\}\left\{2(p_2^3-1)p_2^x - 3(p_2^2-1)\right\}}{(3-2p_1^x)(3-2p_2^x)}. \end{aligned} \tag{2.10}$$

The extra term $h_1(x;p_1)h_2(x;p_2)$ arises because in the discrete case $P(X_1 = x, X_2 = x) \neq 0$, where $h_1(x;p_1)$ and $h_2(x;p_2)$ are the hrfs of X_1 and X_2, respectively. The rest of this section contains the statistical properties of the DBL distribution.

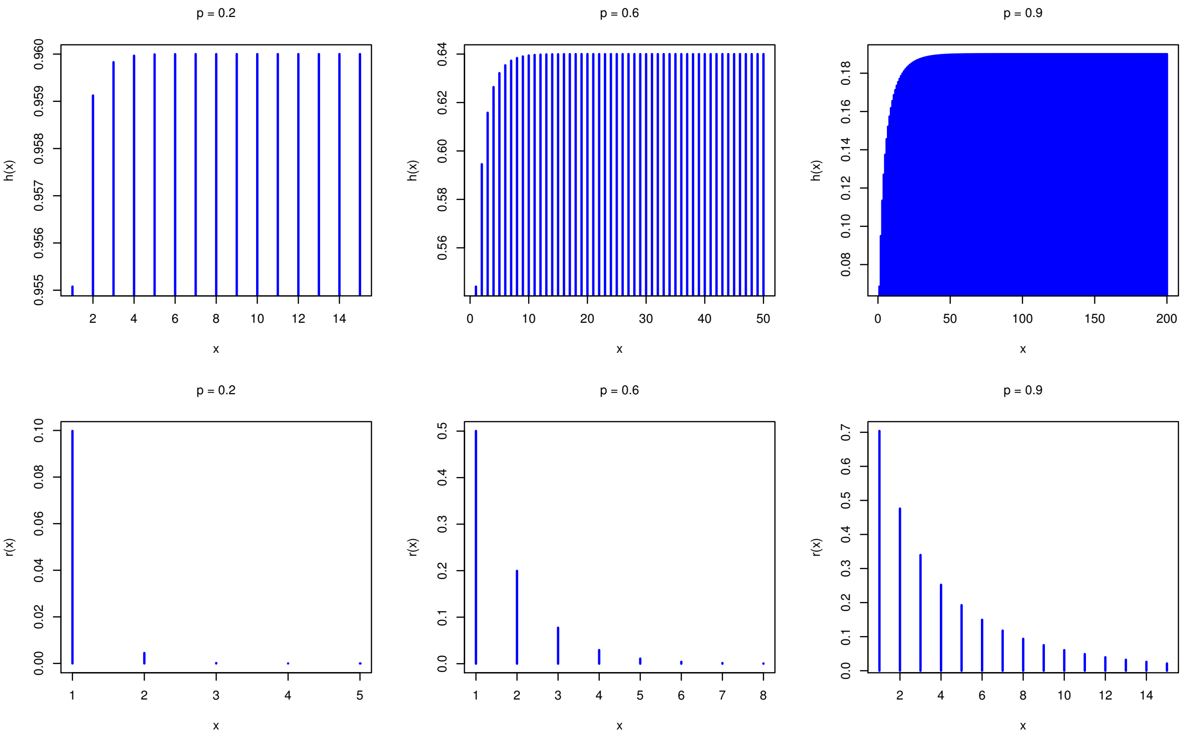


Figure 2: The hrf and rhrf of the DBL distribution.

2.1. Mode

The mode of the DBL distribution is obtained by solving (2.11):

$$6(p^3-1)p^{3x}\ln(p) - 6(p^2-1)p^{2x}\ln(p) = 0. \tag{2.11}$$

By solving (2.11), we have

$$\text{Mode}(X) = \frac{\ln(p+1) - \ln(p^2+p+1)}{\ln(p)}. \tag{2.12}$$

As seen from (2.12), mode of the DBL distribution is an increasing function of the parameter p.

2.2. Moments, skewness and kurtosis

The probability generating function (pgf) of the DBL distribution is obtained as follows:

$$\begin{aligned} G_X(s) &= \sum_{x=0}^{\infty} s^x f_x(x;p) \\ &= 2\sum_{x=0}^{\infty}(p^3-1)\left(p^3 s\right)^x - 3\sum_{x=0}^{\infty}(p^2-1)\left(p^2 s\right)^x \\ &= \frac{2(p^3-1)}{1-p^3 s} - \frac{3(p^2-1)}{1-p^2 s}, \end{aligned} \tag{2.13}$$

where $\sum_{x=0}^{\infty} aq^x = \frac{a}{1-q}$. Replacing s with e^s, the moment generating function (mgf) of the DBL distribution is

$$M_X(s) = \frac{2(p^3-1)}{1-p^3e^s} - \frac{3(p^2-1)}{1-p^2e^s}. \tag{2.14}$$

Using the mgf, given in (2.14), we obtain the mean, variance, skewness and kurtosis of the DBL distribution, given, respectively, by

$$\mathrm{E}(X) = \frac{p^2(p^2+p+3)}{(p^2+p+1)(1-p^2)}, \tag{2.15}$$

$$\mathrm{Var}(X) = \frac{p^2(3p^4+4p^3-p^2+4p+3)}{(p^2+p+1)^2(p^2-1)^2}, \tag{2.16}$$

$$\mathrm{Sk}(X) = -\frac{3p^8+7p^7-3p^6+6p^5+44p^4+6p^3-3p^2+7p+3}{p(3p^4+4p^3-p^2+4p+3)^{3/2}}, \tag{2.17}$$

and

$$\mathrm{Ku}(X) = \frac{\begin{array}{r}3p^{12}+10p^{11}+19p^{10}+72p^9+224p^8+206p^7+\\ +21p^6+206p^5+224p^4+72p^3+19p^2+10p+3\end{array}}{[p(3p^4+4p^3-p^2+4p+3)]^2}. \tag{2.18}$$

The behavior of the mean, variance, skewness and kurtosis are displayed in Figures 3.

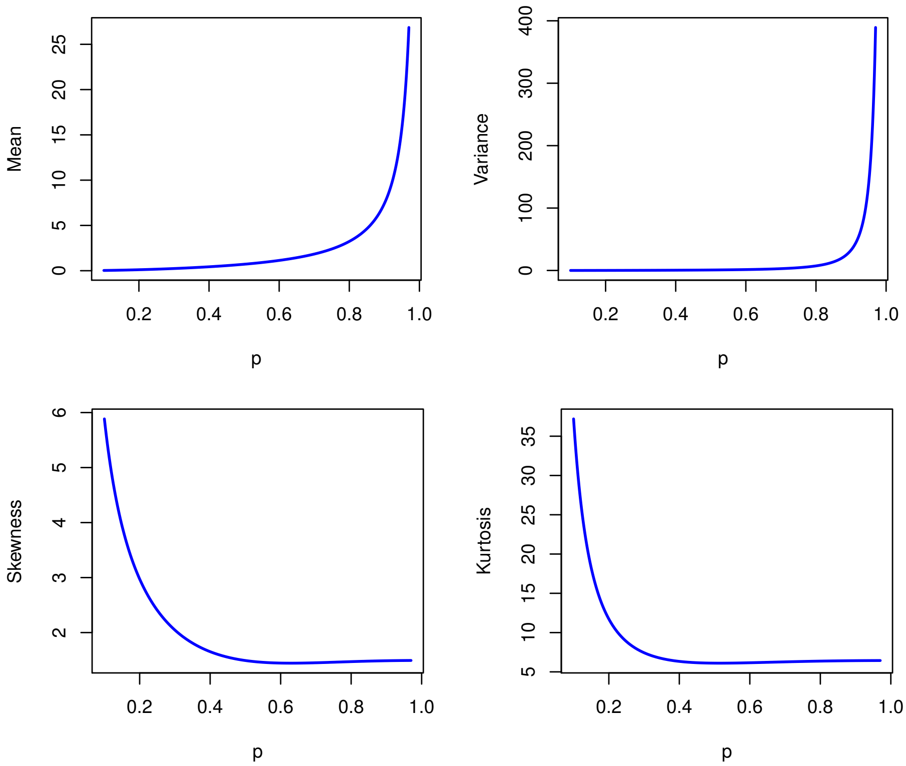


Figure 3: The mean, variance, skewness and kurtosis values of the DBL distribution.

According to results in Figure 3, the following observations are obtained:

1. The mean and variance increase as $p \to 1$;

2. The skewness and kurtosis decrease as $p \to 1$;

3. The proposed distribution is suitable model for the positively skewed count data sets;

4. The proposed distribution is leptokurtic since its kurtosis is always greater than 3.

2.3. Dispersion index and coefficient of variation

The dispersion index (DI) is calculated as variance to mean ratio. When DI is greater than 1, the distribution is over-dispersed, opposite case shows the under-dispersion. When DI is equal to 1, the distribution is equi-dispersed. The coefficient of variation (CV) is also very similar measure to DI. It is calculated as a ratio of the standard deviation to the mean. The DI and CV measures of the DBL distribution are given, respectively, by

$$\mathrm{DI}(X) = \frac{(3p^4 + 4p^3 - p^2 + 4p + 3)}{(p^2 + p + 1)(p^2 + p + 3)(1 - p^2)}, \tag{2.19}$$

$$\mathrm{CV}(X) = \frac{\sqrt{3p^4 + 4p^3 - p^2 + 4p + 3}}{p(p^2 + p + 3)}. \tag{2.20}$$

Figure 4 shows the DI and CV plots of the DBL distribution for various values of the model parameter. It is observed that DI can be either smaller or larger than one.

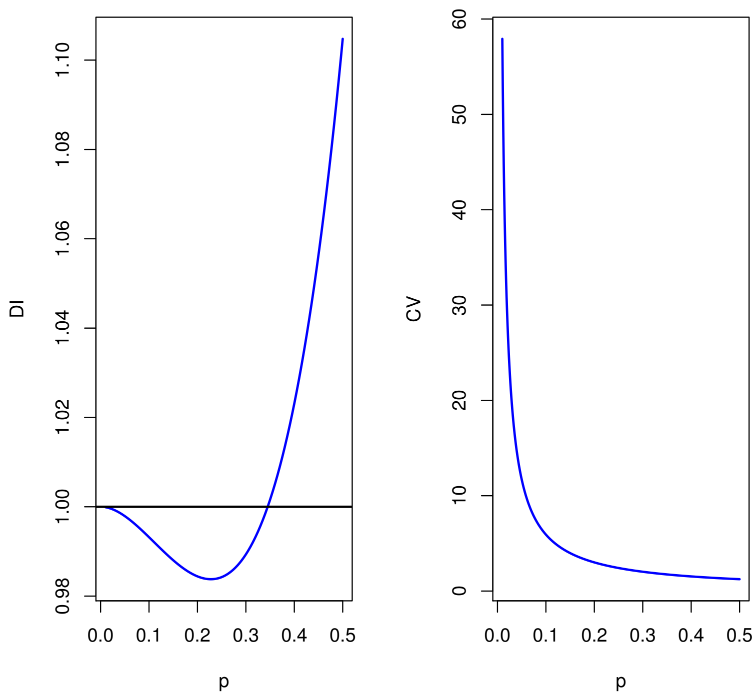


Figure 4: The DI and CV plots of the DBL distribution.

2.4. Mean deviation

The mean deviation (MD) about the mean measures the amount of scatter in a population. For a random variable X having a DBL distribution, the MD is defined as

$$\begin{aligned}
\mathrm{MD}(X) &= \sum_{x=0}^{\infty} |x - \mathrm{E}(X)|\, f(x;p) \\
&= \sum_{x=0}^{\mathrm{E}(X)} (\mathrm{E}(X) - x) f(x;p) \; + \sum_{x=\mathrm{E}(X)+1}^{\infty} (x - \mathrm{E}(X)) f(x;p) \\
&= 2\mathrm{E}(X) F(\mathrm{E}(X);p) \; - \; 2 \sum_{x=0}^{\mathrm{E}(X)} x f(x;p) \\
&= \frac{-2}{p^8+2p^7+p^6-2p^5-4p^4-2p^3+p^2+2p+1} \left\{ \begin{array}{l}
-6p^{\frac{p^4+p^3-6p^2-3p-3}{(p^2+p+1)(p^2-1)}} + 4p^{\frac{p^4+p^3-9p^2-4p-4}{(p^2+p+1)(p^2-1)}} \\
+2p^{\frac{2p^4+2p^3-9p^2-5p-5}{(p^2+p+1)(p^2-1)}} + 6p^{\frac{2(2p^4+2p^3-3p^2-3p-3)}{(p^2+p+1)(p^2-1)}} \\
-4p^{\frac{4p^4+4p^3-9p^2-7p-7}{(p^2+p+1)(p^2-1)}} + 6p^{\frac{5p^4+5p^3-6p^2-7p-7}{(p^2+p+1)(p^2-1)}} \\
-2p^{\frac{5p^4+5p^3-9p^2-8p-8}{(p^2+p+1)(p^2-1)}} + 3p^{\frac{2(3p^4+3p^3-3p^2-4p-4)}{(p^2+p+1)(p^2-1)}} \\
-6p^{\frac{2(p^4+p^3-3p^2-2p-2)}{(p^2+p+1)(p^2-1)}} - 2p^{\frac{3(p^4+p^3-3p^2-2p-2)}{(p^2+p+1)(p^2-1)}} \\
-3p^{\frac{-2(3p^2+p+1)}{(p^2+p+1)(p^2-1)}} + 2p^{\frac{-3(3p^2+p+1)}{(p^2+p+1)(p^2-1)}}
\end{array} \right\}.
\end{aligned}$$

The MD increases with $p \to 1$.

2.5. Stress-strength reliability

Stress-strength reliability (SSR) analysis is widely used in reliability engineering. Assume that both stress and strength are in the positive domain. Let $X_{\text{stress}} \sim \mathrm{DBL}(p)$ and $X_{\text{strength}} \sim \mathrm{DBL}(q)$. Then, the expected SSR can be expressed in a closed form as

$$\mathrm{SSR} := P\left[X_{\text{stress}} \le X_{\text{strength}}\right] = \sum_{x=0}^{\infty} f_{X_{\text{stress}}}(x;p) R_{X_{\text{strength}}}(x;q). \tag{2.21}$$

Using (2.5) and (2.6), we get

$$\mathrm{SSR} = \frac{4q^3(p^3-1)}{p^3q^3-1} + \frac{6q^2(1-p^3)}{p^3q^2-1} + \frac{6q^3(1-p^2)}{p^2q^3-1} + \frac{9q^2(p^2-1)}{p^2q^2-1}. \tag{2.22}$$

Figure 5 shows the SSR for various values of the parameters p and q. According to Figure 5, we concluded that:

- **(i)** The SSR increases for $q \to 1$ with fixed value of p;
- **(ii)** The SSR decreases for $p \to 1$ with fixed value of q.

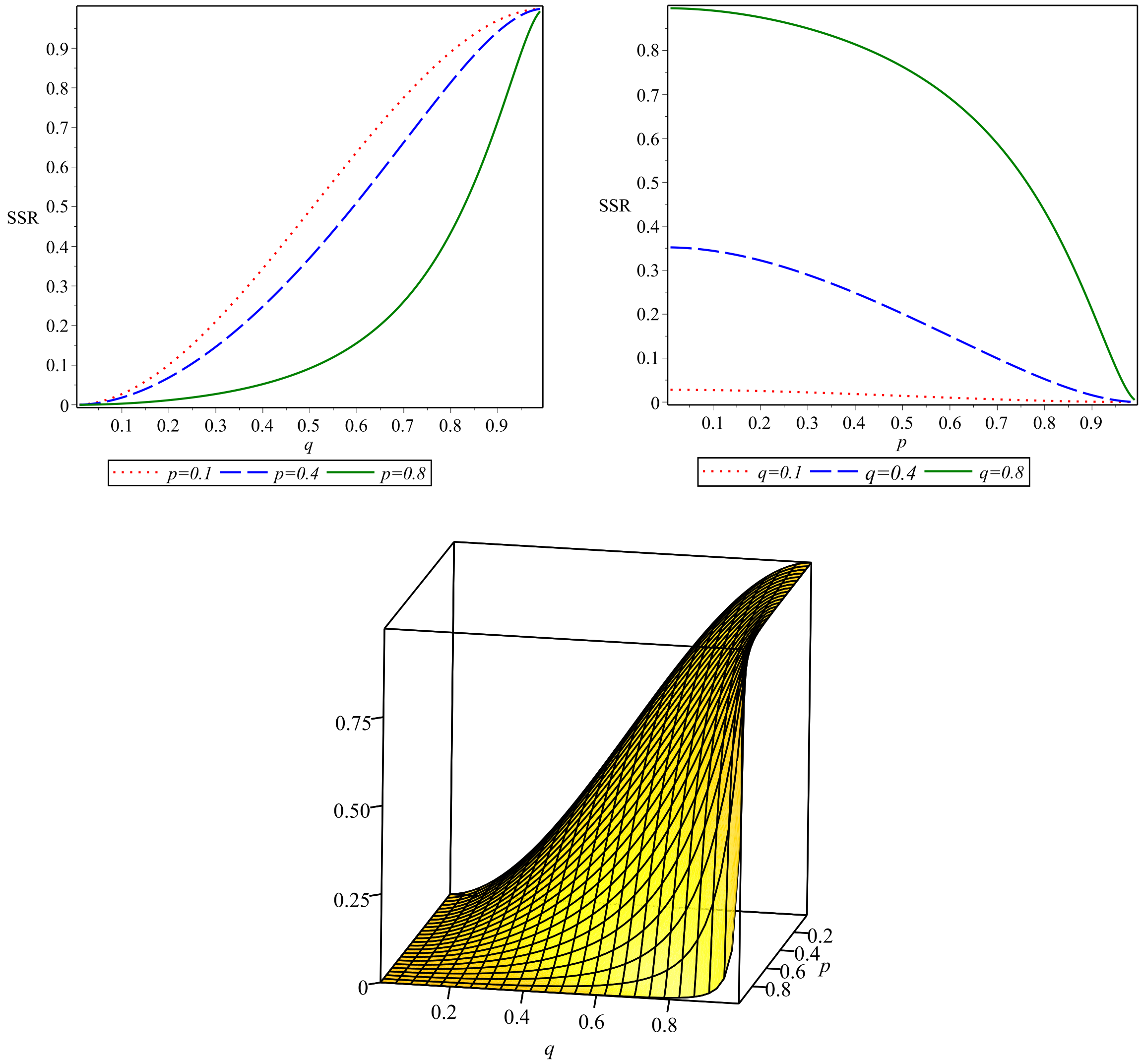


Figure 5: The SSR utilizing the DBL distribution.

2.6. Order statistics

Let $x_{1:n}, x_{2:n}, ..., x_{n:n}$ be the order statistics of a random sample from the DBL distribution. The cdf of i-th order statistics for an integer value of x is given by

$$
\begin{aligned}
F_{i:n}(x;p) &= \sum_{k=i}^{n} \binom{n}{k} \left[F_i(x;p\right]^k \left[1-F_i(x;p)\right]^{n-k} \\
&= \sum_{k=i}^{n} \sum_{j=0}^{n-k} \Upsilon_{(m)}^{(n,k)} \left[F_i(x;p)\right]^{k+j} \\
&= \sum_{k=i}^{n} \sum_{j=0}^{n-k} \Upsilon_{(m)}^{(n,k)} F_i(x;p,k+j),
\end{aligned}
\tag{2.23}
$$

where $\Upsilon_{(m)}^{(n,k)} := (-1)^j \binom{n}{k} \binom{n-k}{j}$ and $F_i(x;p,k+j) = \left[1-\left(3-2p^{x+1}\right)p^{2(x+1)}\right]^{k+j}$ represents the cdf of the exponentiated DBL distribution with power parameter $k+j$.

The corresponding pmf to (2.23) is given by

$$f_{i:n}(x;p) = F_{i:n}(x;p) - F_{i:n}(x-1;p)$$

$$= \sum_{k=i}^{n}\sum_{j=0}^{n-k} \Upsilon_{(m)}^{(n,k)} f_i(x;p,k+j), \tag{2.24}$$

where $f_i(x;p,k+j)$ represents the pmf of the exponentiated DBL distribution with power parameter $k+j$. Thus, the b-th moments of $X_{i:n}$ can be written as

$$\mathrm{E}(X_{i:n}^b) = \sum_{x=0}^{\infty}\sum_{k=i}^{n}\sum_{j=0}^{n-k} \Upsilon_{(m)}^{(n,k)} x^b f_i(x;p,k+j). \tag{2.25}$$

3. ESTIMATION METHODS

We use two estimation methods to estimate the unknown parameter of the DBL distribution. These methods are maximum likelihood estimation (MLE) and method of moments (MM).

3.1. Maximum likelihood estimation

Let $X_1, X_2, ..., X_n$ be random variables from the DBL distribution. The log-likelihood function (L) of the DBL distribution is

$$L(x;p) = n\ln(p-1) + 2\ln p \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \ln\left[2p^{x_i}\left(p^2+p+1\right) - 3p - 3\right]. \tag{3.1}$$

By differentiating (3.1) with respect to the parameter p, we have the following equation:

$$\frac{n}{p-1} + \frac{2}{p}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \frac{2p^{x_i}\left(2p+1\right) + 2x_i p^{x_i-1}\left(p^2+p+1\right) - 3}{2p^{x_i}\left(p^2+p+1\right) - 3p - 3} = 0. \tag{3.2}$$

The solution of the above equation gives MLE of the parameter p. However, it is not possible to obtain the exact form of the MLE of the parameter p since the equation has non-linear functions. For this reason, it has to be solved numerically. The other possible way to obtain the MLE of the parameter p is to direct minimization of the negative log-likelihood function. To do this, we use the `constrOptim` function of `R` software.

3.2. Moment estimation

The MM estimator of the parameter p is obtained by solving

$$\frac{p^2(p^2+p+3)}{(p^2+p+1)(p^2-1)} - \bar{x} = 0, \tag{3.3}$$

where $\bar{x} = \sum_{i=1}^{n} x_i/n$. We use `nleqslv` to solve (3.3).

4. INAR(1) PROCESS WITH DBL INNOVATIONS

Time series of counts arise in different fields such as econometrics, actuarial and medical sciences. For instance, yearly incidents of terrorism, daily number of doctor visits, yearly number of traffic accidents and among others. McKenzie (1985) [20] and Al-Osh and Alzaid (1987) [1] introduced the INAR(1) process with Poisson innovations to analyze these kind of data sets. It is said that $\{X_t\}_{t\in\mathbb{Z}}$ follows a stable INAR(1) process if

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{4.1}$$

where $0 \le \alpha < 1$. The innovation process, $\{\varepsilon_t\}_{t\in\mathbb{Z}}$, constitutes a sequence of the independent and identically distributed (iid) discrete random variables. The mean and variance of the innovation process are $E(\varepsilon_t) = \mu_\varepsilon$ and $\operatorname{Var}(\varepsilon_t) = \sigma_\varepsilon^2$, respectively. This model was shortly denoted as INAR(1)P process. Note that the innovations, $\{\varepsilon_t\}_{t\in\mathbb{Z}}$, are independent from X_{t-k}, $k \ge 1$. The binomial thinning operator, $\circ$, is defined by

$$\alpha \circ X_{t-1} := \sum_{j=1}^{X_{t-1}} W_j, \tag{4.2}$$

where $\{W_j\}_{j\ge1}$ is a sequence of iid Bernoulli random variables with probabilities $\Pr(W_j = 1) = 1 - \Pr(W_j = 0) = \alpha$. The one-step transition probability of the INAR(1) process is

$$\Pr(X_t = k | X_{t-1} = l) = \sum_{i=1}^{\min(k,l)} \Pr(B_l^\alpha = i)\Pr(\varepsilon_t = k - i), k, l \ge 0, \tag{4.3}$$

where $B_n^\alpha \sim \text{Binomial}(\alpha, n)$ and $\alpha \in [0, 1)$. According to the works of Al-Osh and Alzaid (1987) [1] and McKenzie (1985) [20], we introduce a new INAR(1) model with a more flexible innovation distribution. We assume that the innovations follow a DBL distribution with parameter p. We call this process as INAR(1)DBL. Since the dispersion of the DBL can be under or over the value 1, the INAR(1)DBL can be used to model both under-dispersed and over-dispersed time series of counts. Using (4.3), the one-step transition probability of INAR(1)DBL process is given by

$$\begin{aligned}\gamma_{i,j} &= \Pr(X_t = k | X_{t-1} = l) \\ &= \sum_{i=1}^{\min(k,l)} \binom{l}{i} \alpha^i (1-\alpha)^{l-i} \left[2\left(p^3 - 1\right) p^{3(k-i)} - 3\left(p^2 - 1\right) p^{2(k-i)}\right].\end{aligned} \tag{4.4}$$

The equation in (4.4) represents the one-step transition probability of the process from state l to state k. The marginal probability function of the INAR(1)DBL process is

$$\begin{aligned}\gamma_j &= \Pr(X_t = k) \\ &= \sum_{l=0}^{\infty} \gamma_{ij} \Pr(X_{t-1} = l) \\ &= \sum_{l=0}^{\infty} \sum_{i=1}^{\min(k,l)} \binom{l}{i} \alpha^i (1-\alpha)^{l-i} \left[2\left(p^3 - 1\right) p^{3(k-i)} - 3\left(p^2 - 1\right) p^{2(k-i)}\right] \gamma_i,\end{aligned} \tag{4.5}$$

where $k = 0, 1, 2, \ldots,$ (see Jazi *et al.*, 2012 [14]). Following the results given in Al-Osh and Alzaid (1987) [1], we obtain the mean, variance and DI of the INAR(1)DBL process and given, respectively, by

$$\mu_X = \frac{p^2\left(p^2+p+3\right)}{\left(p^2+p+1\right)\left(1-p^2\right)(1-\alpha)}, \tag{4.6}$$

$$\sigma_X^2 = \frac{\alpha}{\alpha^2-1}\left(\frac{3p^2\left(p^2-1\right)}{\left(p^2-1\right)^2} - \frac{2p^2\left(p^2-1\right)}{\left(p^3-1\right)^2}\right) - \frac{p^2\left(3p^4+4p^3-p^2+4p+3\right)}{\left(\alpha^2-1\right)\left(p^4+p^3-p-1\right)^2}, \tag{4.7}$$

$$\mathrm{DI}_X = \left(\alpha - \frac{3\,p^4+4\,p^3-p^2+4\,p+3}{p^6+2\,p^5+4\,p^4+2\,p^3-2\,p^2-4\,p-3}\right)(\alpha+1)^{-1}. \tag{4.8}$$

According to Al-Osh and Alzaid (1987) [1], the conditional expectation and variance of INAR(1)DBL process are given, respectively, by

$$\mathrm{E}\left(X_t \mid X_{t-1}\right) = \alpha X_{t-1} + \frac{p^2(p^2+p+3)}{(p^2+p+1)(1-p^2)}, \tag{4.9}$$

$$\mathrm{Var}\left(X_t \mid X_{t-1}\right) = \alpha\left(1-\alpha\right)X_{t-1} + \frac{p^2(3p^4+4p^3-p^2+4p+3)}{(p^2+p+1)^2(p^2-1)^2}. \tag{4.10}$$

4.1. Estimation of INAR(1)DBL process

Bourguignon *et al.* (2019) [5] and Lívio *et al.* (2018) [19] used three estimation methods to obtain the parameters of INAR(1) process defined under different innovation distributions. These methods are conditional least squares (CLS), Yule-Walker (YW) and the conditional maximum likelihood (CML) estimation methods. They compared the finite sample performance of these estimation methods for different sample sizes and parameter settings and concluded that CML estimation method provides better results than CLS and YW estimation methods. Here, we use these three estimation methods to obtain the unknown parameters of the INAR(1)DBL process. However, there are no explicit forms for the CLS and YW estimators of the INAR(1)DBL process because of the non-linearity of the equations.

Conditional maximum likelihood

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

$$\begin{aligned}\ell\left(\boldsymbol{\Theta}\right) &= \sum_{t=2}^{T} \ln\left[\Pr\left(X_t = k \mid X_{t-1} = l\right)\right] \\ &= \sum_{t=2}^{T} \ln\left[\begin{array}{l}\displaystyle\sum_{i=0}^{\min(x_t,x_{t-1})} \binom{x_{t-1}}{i} \alpha^i (1-\alpha)^{x_{t-1}-i} \\ \times\left\{2\left(p^3-1\right)p^{3(x_t-i)} - 3\left(p^2-1\right)p^{2(x_t-i)}\right\}\end{array}\right],\end{aligned} \tag{4.11}$$

where $\boldsymbol{\Theta} = (\alpha_{cml}, p_{cml})$ is the unknown parameter vector. The CML estimator of $\boldsymbol{\Theta}$, say $\hat{\boldsymbol{\Theta}}$ can be obtained by maximizing the equation (4.11). It is well-known that the maximization of (4.11) is equivalent to minimization of the negative of (4.11). Minimization of the negative of (4.11) could be done by using different software such as `R`, `MATLAB`, `C++` or `S-Plus`. Here, we prefer `constrOptim` function of `R` software to minimize the negative of (4.11). Note that the CML estimators are asymptotically normal and consistent under the regularity conditions (Bourguignon *et al.*, 2019 [5]).

Yule-Walker

The YW estimators are obtained by simultaneous solution of the equations for the theoretical and empirical moments of the INAR(1)DBL process. The autocorrelation function (ACF) of the INAR(1) process at lag h is $\rho_X(h) = \alpha^h$, and $\rho_X(1) = \alpha$ for $h = 1$. Therefore, the YW estimator of the parameter α is

$$\hat{\alpha}_{YW} = \frac{\sum\limits_{t=2}^{T} \left(X_t - \bar{X}\right)\left(X_{t-1} - \bar{X}\right)}{\sum\limits_{t=1}^{T} \left(X_t - \bar{X}\right)^2}. \tag{4.12}$$

The YW estimator of the parameter p, say $\hat{p}_{YW}$, can be obtained by solving

$$\frac{p^2\left(p^2 + p + 3\right)}{\left(p^2 + p + 1\right)\left(1 - p^2\right)\left(1 - \hat{\alpha}_{YW}\right)} = \bar{X}, \tag{4.13}$$

where $\bar{X} = \sum\limits_{t=1}^{T} X_t \Big/ T$. However, it is not possible to obtain the explicit forms of the YW estimators of the parameter p. Therefore, (4.13) has to solved numerically by using the software such as `R` or `MATLAB`. We use the `uniroot` function of the `R` software to obtain $\hat{p}_{YW}$.

Conditional least squares

The CLS estimators of the parameters α and p can be obtained by minimizing

$$S(\boldsymbol{\eta}) = \sum_{t=2}^{T} \left(X_t - E\left(X_t | X_{t-1}\right)\right)^2, \tag{4.14}$$

where $\boldsymbol{\eta} = (\alpha_{cls}, p_{cls})$ and $E(X_t | X_{t-1})$ is given in (4.9). Replacing $E(X_t | X_{t-1})$ with (4.9) in (4.14), we have

$$S(\boldsymbol{\eta}) = \sum_{t=2}^{T} \left(X_t - \alpha X_{t-1} - \frac{p^2\left(p^2 + p + 3\right)}{\left(p^2 + p + 1\right)\left(1 - p^2\right)}\right)^2. \tag{4.15}$$

The derivatives of (4.15) with respect to the parameters α and p and equating them to zero, we have

$$\frac{\partial S(\boldsymbol{\eta})}{\partial p} = \sum_{t=2}^{T} -\frac{12\,p\left(X_t - \alpha\,X_{t-1} + \frac{p^2\left(p^2+p+3\right)}{\left(p^2-1\right)\left(p^2+p+1\right)}\right)\left(p^4+p^3+p^2+p+1\right)}{\left(-p^4-p^3+p+1\right)^2} = 0, \tag{4.16}$$

$$\frac{\partial S(\boldsymbol{\eta})}{\partial \alpha} = \sum_{t=2}^{T} -2\,X_{t-1}\left(X_t - \alpha\,X_{t-1} + \frac{p^2\,\left(p^2+p+3\right)}{\left(p^2-1\right)\,\left(p^2+p+1\right)}\right) = 0. \tag{4.17}$$

The simultaneous solutions of (4.16) and (4.17) give the CLS estimators of the parameter α and p. However, since the mean of the DBL distribution has non-linear functions, it is not possible to obtain the p_{cls} in explicit form. However, when the parameter p is known, the CLS estimator of the parameter α is

$$\hat{\alpha}_{cls} = \sum_{t=2}^{T} \frac{\left(X_t+1\right)\left(p^4+p^3\right) - X_t\left(p+1\right) + 3p^2}{\left(p^4+p^3-p-1\right)X_{t-1}}, \tag{4.18}$$

where p can be replaced with $\hat{p}_{cml}$ (see, Bourguignon *et al.*, 2019 [5]).

5. SIMULATION STUDIES

Here, two simulation studies are given to evaluate the parameter estimation performance of proposed models.

5.1. Simulation of DBL model

The finite-sample performances of the MLE and MM methods are compared for small and large sample sizes based on the simulation study. The below simulation steps are used for this goal:

1. Generate $N = 10{,}000$ samples of size $n = 20, 50, 100, 200$ and 500 from DBL(0.1), DBL(0.5) and DBL(0.7), respectively.

2. Using each generated sample, compute the MLE and MM estimator of the parameter p, say $\widehat{p}_j$ where $j = 1, 2, ..., 10{,}000$.

3. Compute the biases, mean-squared errors (MSEs) and mean relative errors (MREs) using the following equations:

$$\text{Bias}(p) = \frac{1}{N}\sum_{j=1}^{N}\left(\widehat{p}_j - p\right), \quad \text{MSE}(p) = \frac{1}{N}\sum_{j=1}^{N}\left(\widehat{p}_j - p\right)^2 \quad \text{and} \quad \text{MRE} = \frac{1}{N}\sum_{j=1}^{N}\frac{\hat{p_j}}{p_j}.$$

The simulation results are reported in Table 1. The following remarks are obtained according to the results in Table 1:

1. The estimated biases always decrease and near the zero when $n \to \infty$.
2. The estimated MSEs decrease and near the zero when $n \to \infty$.
3. The estimated MREs are near the desired value, 1, especially for large sample sizes.
4. Both estimation methods work well for estimating the parameter p and produce similar results.

Similar results can be obtained for different values of the parameter p.

Table 1: The simulation results of DBL distribution.

Parameter	Sample size	Bias		MSE		MRE	
		MLE	MM	MLE	MM	MLE	MM
$p = 0.1$	20	−0.036585	−0.036506	0.006924	0.006926	0.634153	0.634937
	50	−0.016756	−0.016717	0.003237	0.003238	0.832445	0.832833
	100	−0.006808	−0.006800	0.001424	0.001424	0.931918	0.932002
	200	−0.002833	−0.002825	0.000523	0.000523	0.971665	0.971747
	500	−0.002338	−0.002341	0.000196	0.000196	0.976622	0.976590
$p = 0.5$	20	−0.008975	−0.008716	0.003892	0.003873	0.982050	0.982567
	50	−0.002803	−0.002843	0.001605	0.001610	0.994394	0.994314
	100	−0.001900	−0.001884	0.000682	0.000682	0.996201	0.996231
	200	−0.000803	−0.000765	0.000317	0.000317	0.998394	0.998470
	500	−0.000145	−0.000146	0.000150	0.000151	0.999101	0.999089
$p = 0.7$	20	−0.004901	−0.004959	0.001647	0.001647	0.992999	0.992915
	50	−0.001908	−0.001971	0.000700	0.000702	0.997275	0.997184
	100	−0.000833	−0.000854	0.000330	0.000329	0.998810	0.998780
	200	−0.000734	−0.000764	0.000170	0.000170	0.998952	0.998909
	500	−0.000856	−0.000859	0.000075	0.000075	0.998777	0.998773

5.2. Simulation of INAR(1)DBL process

We carry out a simulation study to evaluate the asymptotic behaviours of the CML, YW and CLS estimators of INAR(1)DBL process for small and sufficiently large sample sizes. The number of simulation replications is $N = 10,000$ and three sample sizes are used: $n = 25$, 50 and 100. Four parameter vectors are also used. These are ($\alpha = 0.3$, $p = 0.9$), ($\alpha = 0.5$, $p = 0.5$), ($\alpha = 0.2$, $p = 0.3$) and ($\alpha = 0.7$, $p = 0.6$). The biases, MSEs and MREs are used to evaluate the simulation results.

We expect that when the sample size is sufficiently large, the biases and MSEs near the zero and MREs are near the one. The simulation results are summarized in Table 2. As seen from the simulation results, the results of the CML and YW estimation methods are very near each other. However, the CML estimation method approaches to the desired values of the biases, MSEs and MREs more faster than those of the CLS and YW estimation methods.

The performance of the CML method is better than the CLS and YW estimation methods for both small and sufficiently large sample sizes. Therefore, we suggest to use the CML estimation to obtain the unknown parameters of the INAR(1)DBL process.

Table 2: Simulation results of INAR(1)DBL process.

Sample size	Parameters	CML			YW			CLS		
		Bias	MSE	MRE	Bias	MSE	MRE	Bias	MSE	MRE
$\alpha=0.3,\ p=0.9$										
$n=25$	α	−0.0020	0.0092	0.9959	−0.1239	0.0469	0.7620	−0.1198	0.0494	0.7768
	p	−0.0055	0.0016	0.9931	0.0209	0.0028	1.0261	0.0187	0.0030	1.0234
$n=50$	α	−0.0032	0.0042	0.9936	−0.0686	0.0195	0.8629	−0.0685	0.0203	0.8631
	p	−0.0010	0.0007	0.9987	0.0140	0.0014	1.0175	0.0132	0.0016	1.0165
$n=100$	α	−0.0003	0.0023	0.9993	−0.0270	0.0088	0.9461	−0.0257	0.0090	0.9487
	p	−0.0016	0.0004	0.9980	0.0038	0.0008	1.0048	0.0035	0.0009	1.0043
$\alpha=0.5,\ p=0.5$										
$n=25$	α	−0.0295	0.0257	0.9409	−0.1326	0.0533	0.7444	−0.1286	0.0579	0.7524
	p	0.0002	0.0049	1.0003	0.0356	0.0079	1.0713	0.0333	0.0087	1.0665
$n=50$	α	−0.0122	0.0112	0.9756	−0.0623	0.0207	0.8754	−0.0616	0.0218	0.8768
	p	0.0014	0.0024	1.0029	0.0196	0.0035	1.0392	0.0193	0.0039	1.0386
$n=100$	α	−0.0025	0.0054	0.9950	−0.0310	0.0095	0.9380	−0.0310	0.0100	0.9381
	p	−0.0010	0.0013	0.9979	0.0096	0.0020	1.0192	0.0098	0.0021	1.0197
$\alpha=0.2,\ p=0.3$										
$n=25$	α	−0.0285	0.0304	0.9661	−0.0910	0.0513	0.9550	−0.0814	0.0599	1.0285
	p	−0.0076	0.0046	0.9924	0.0033	0.0047	1.0111	−0.0007	0.0064	1.0053
$n=50$	α	−0.0276	0.0222	0.9762	−0.0502	0.0290	0.8860	−0.0493	0.0298	0.8980
	p	−0.0049	0.0023	0.9838	−0.0009	0.0023	0.9969	−0.0014	0.0024	0.9954
$n=100$	α	−0.0141	0.0116	0.9896	−0.0206	0.0135	0.9221	−0.0198	0.0134	0.9253
	p	−0.0017	0.0011	0.9944	−0.0006	0.0012	0.9980	−0.0007	0.0011	0.9976
$\alpha=0.7,\ p=0.6$										
$n=25$	α	−0.0134	0.0082	0.9808	−0.1689	0.0591	0.7590	−0.1657	0.0637	0.7649
	p	−0.0073	0.0047	0.9878	0.0667	0.0119	1.1112	0.0610	0.0141	1.1017
$n=50$	α	−0.0058	0.0036	0.9917	−0.0855	0.0216	0.8779	−0.0856	0.0227	0.8777
	p	−0.0014	0.0021	0.9977	0.0401	0.0063	1.0669	0.0396	0.0069	1.0660
$n=100$	α	−0.0051	0.0019	0.9928	−0.0433	0.0079	0.9381	−0.0434	0.0083	0.9380
	p	−0.0006	0.0011	0.9990	0.0208	0.0029	1.0346	0.0207	0.0031	1.0345

6. EMPIRICAL STUDIES

This section is devoted to illustrate the importance of the DBL distribution by analyzing the three real data sets with proposed and competitive models. The performance of fitted models are compared using goodness-of-fit criteria, Kolmogorov-Smirnov (K-S) test with its corresponding p-value.

6.1. Number of fires in Greece

The first data set deals with the number of fires in Greece for the period from 1 July 1998 to 31 August 1998. This data set was reported by Karlis and Xekalaki (2001) [16] and also is given in the Appendix. The performance of the DBL distribution is compared with competitive models listed in Table 3.

Table 3: The competitive models of the DBL distribution.

Distribution	Abbreviation	Author(s)
Geometric	Geo	—
Discrete Lindley	DLi	Gómez-Déniz and Calderín-Ojeda (2011) [12]
Discrete Rayleigh	DR	Roy (2004) [28]
Discrete inverse Rayleigh	DIR	Hussain and Ahmad (2014) [13]
Discrete Pareto	DPa	Krishna and Pundir (2009) [17]
Poisson	Poi	Poisson (1837) [27]
Discrete generalized exponential type II	DGE-II	Nekoukhou *et al.* (2013) [22]
Discrete Weibull	DW	Nakagawa and Osaki (1975) [21]
Discrete inverse Weibull	DIW	Jazi *et al.* (2010) [15]
Discrete Burr type II	DB-XII	Para and Jan (2016a) [24]
Exponentiated discrete Lindley	EDLi	El-morshedy *et al.* (2019) [10]
Discrete log-logistic	DLog-L	Para and Jan (2016b) [25]
Exponentiated discrete Weibull	EDW	Nekoukhou and Bidram (2015) [23]

Tables 4 and 5 contain the MLEs of the parameters for each fitted distribution with their standard errors (std-er). The asymptotic confidence intervals (CI) and the results of the goodness-of-fit test are also reported in these tables.

Table 4: The MLEs, CIs, K-S and p-values of fitted models with one-parameter for the number of fires in Greece.

Statistic		Model						
		DBL	Geo	DLi	DR	DIR	DPa	Poi
MLE_p		0.867	0.844	0.741	0.980	0.018	0.546	5.398
Std-er_p		0.008	0.013	0.014	0.023	0.007	0.029	0.209
95% CI	Lower_p	0.852	0.818	0.712	0.935	0.004	0.488	4.988
	Upper_p	0.883	0.869	0.769	1.00	0.033	0.605	5.809
K-S		0.096	0.164	0.097	0.183	0.429	0.355	0.854
p-value		0.202	0.003	0.198	<0.001	0	<0.001	0

According to Tables 4 and 5, two model provide the sufficient results for analyzing the number of fires in Greece since the p-values of these models are greater than 0.05. These are DBL and DLi distributions. However, DBL distribution has the smallest value of K-S statistic and the largest p-value among all competitive models as well as DLi distribution.

Table 5: The MLEs, CIs, K-S and p-values of fitted models with two and more parameters for the number of fires in Greece.

Statistic		DGE-II	DW	DIW	DB-XII	EDLi	DLog-L	EDW
MLE_p		0.822	0.879	0.079	0.761	0.766	4.226	0.860
Std-er_p		0.019	0.023	0.022	0.043	0.021	0.389	0.099
95% CI	Lower_p	0.785	0.835	0.035	0.677	0.725	3.462	0.665
	Upper_p	0.859	0.924	0.123	0.845	0.808	4.989	1.055
MLE_α		1.255	1.131	1.035	2.503	0.797	1.717	1.081
Std-er_α		0.175	0.082	0.079	0.487	0.113	0.138	0.238
95% CI	Lower_α	0.912	0.969	0.881	1.548	0.575	1.446	0.615
	Upper_α	1.598	1.292	1.189	3.457	1.018	1.988	1.549
MLE_θ		—	—	—	—	—	—	1.092
Std-er_θ		—	—	—	—	—	—	0.448
95% CI	Lower_θ	—	—	—	—	—	—	0.214
	Upper_θ	—	—	—	—	—	—	1.969
K-S		0.130	0.123	0.208	0.299	0.124	0.149	0.125
p-value		0.031	0.047	< 0.001	< 0.001	0.046	0.009	0.042

Figures 6 and 7 show the estimated cdfs and probability-probability (PP) plots. These figures support the results reported in Tables 4 and 5.

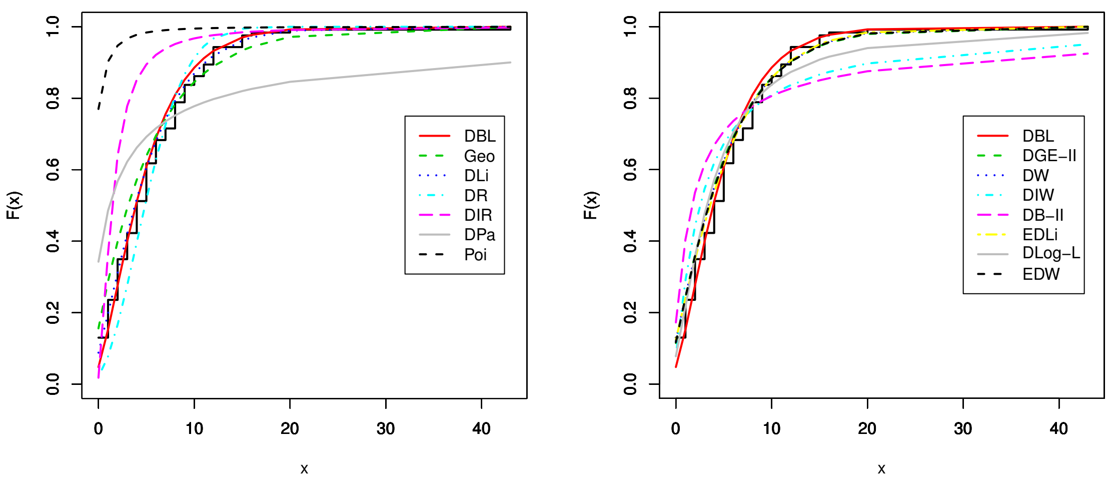


Figure 6: The estimated CDFs of fitted models.

Figure 8 shows the log-likelihood profile of $\widehat{p}$ where $L = -346.902$. It is found that the log-likelihood profile of $\widehat{p}$ is unimodal-shaped. Thus, this estimator is a unique and considered the best for the used data set.

Table 6 shows the results of MM method for the DBL parameter. It is clear that MM method works well for estimating the parameter p.

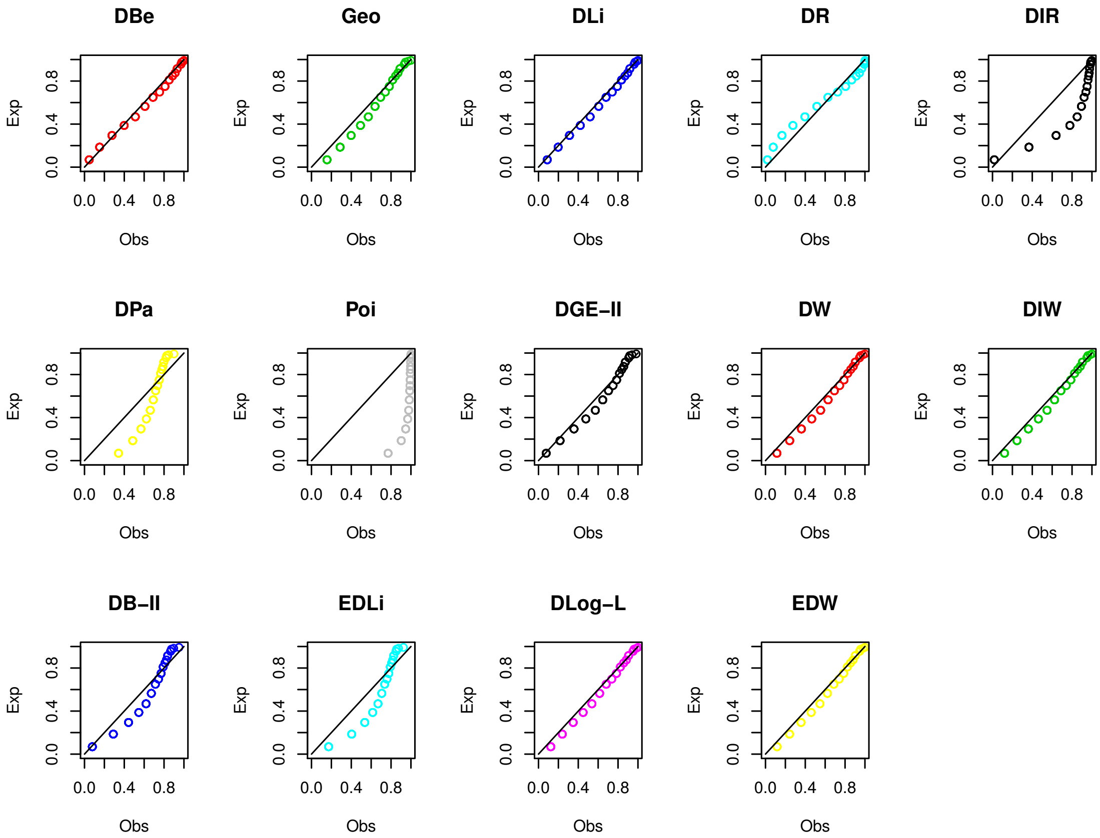


Figure 7: The PP plots of fitted models.

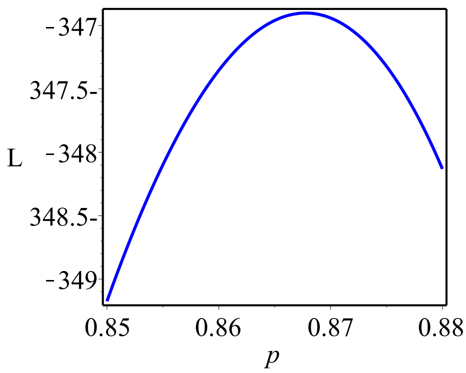


Figure 8: The log-likelihood profile of $\widehat{p}$ for the number of fires in Greece data set.

Table 6: The estimated parameter of DBL distribution with MM method.

Method	Measure		
	$\widehat{p}$	K-S	p-value
MM	0.868	0.095	0.220

Using the MM estimator of the parameter of p, the statistical properties of DBL distribution such as mean, mode, variance, DI, MD, CV, skewness and kurtosis values are listed in Table 7.

Table 7: The statistical properties of DBL distribution for the number of fires in Greece.

Method	Measure							
	Mean	Mode	Variance	DI	MD	CV	Skewness	Kurtosis
MM	5.3867	2.3936	18.1002	3.3601	3.2218	0.7897	1.4837	6.4127

6.2. Failure times

The data used represents the failure times for a sample of 15 electronic components in an acceleration life test (see Lawless, 2003 [18]). The performance of the DBL distribution is compared with discrete flexible model with one parameter (DFx-I), Geo, DR, DIR, DPa, DGE-II, DIW, DLog-L, DB-XII and discrete Lomax (DLo) distributions. The results of the fitted models with goodness-of-fit test are given in Tables 8 and 9.

Table 8: The MLEs, CIs, K-S and p-values of fitted models with one-parameter for the failure times data.

Statistic		Model					
		DBL	DFx-I	Geo	DR	DIR	DPa
MLE_p		0.971	0.973	0.965	0.999	1.8×10^{-7}	0.720
Std-er_p		0.005	0.006	0.009	2.58×10^{-4}	0.055	0.061
95% CI	Lower_p	0.960	0.961	0.948	0.998	0	0.600
	Upper_p	0.981	0.985	0.982	0.999	0.107	0.839
K-S		0.114	0.146	0.177	0.216	0.698	0.405
p-value		0.978	0.864	0.673	0.433	9.1×10^{-7}	0.009

Table 9: The MLEs, CIs, K-S and p-values of fitted models with two-parameters for the failure times data.

Statistic		Model				
		DGE-II	DIW	DLog-L	DB-XII	DLo
MLE_p		0.956	2.2×10^{-4}	21.463	0.975	0.012
Std-er_p		0.013	7.75×10^{-4}	5.387	0.051	0.039
95% CI	Lower_p	0.930	0	10.904	0.874	0
	Upper_p	0.981	0.001	32.021	1	0.088
MLE_α		1.491	0.875	1.791	13.367	104.506
Std-er_α		0.535	0.164	0.388	27.785	84.409
95% CI	Lower_α	0.441	0.554	1.031	0	0
	Upper_α	2.540	1.196	2.551	67.824	269.947
K-S		0.129	0.209	0.136	0.388	0.205
p-value		0.937	0.482	0.913	0.015	0.491

It is found that the DFx-I, Geo, DR, DGE-II, DIW, DLog-L and DLo distributions work quite well besides the DBL distribution. But the DBL distribution is the best among all tested models because it has the smallest value of K-S as well as it has the highest p-value. Figures 9 and 10 show the estimated cdfs and PP plots for the failure times data.

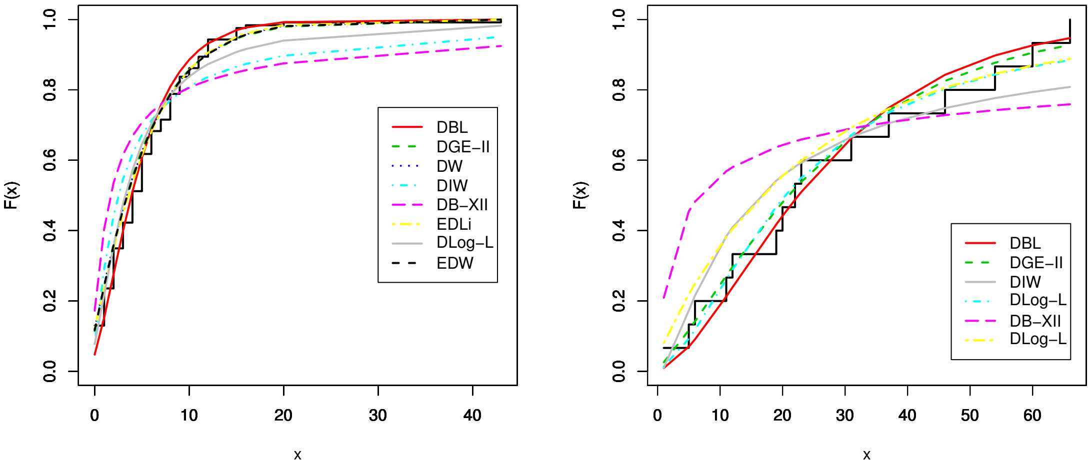


Figure 9: The estimated cdfs for the failure times data.

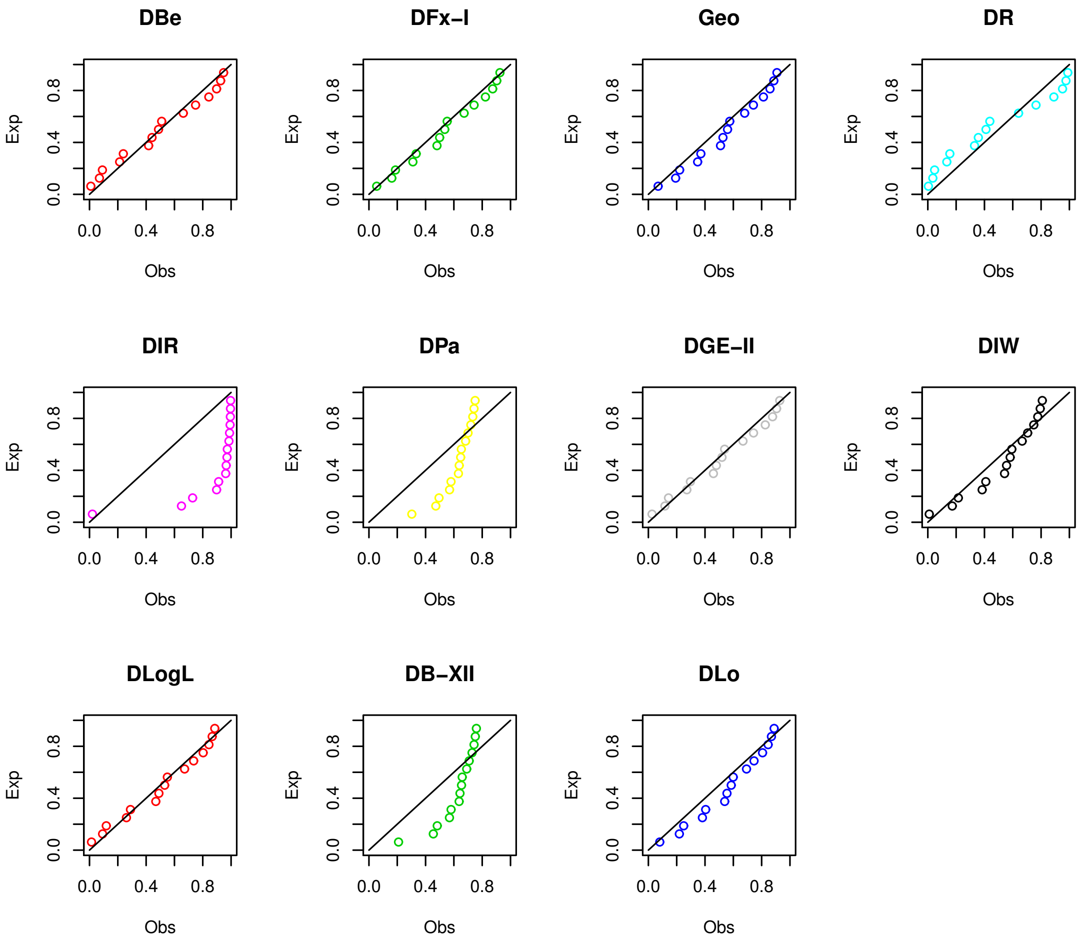
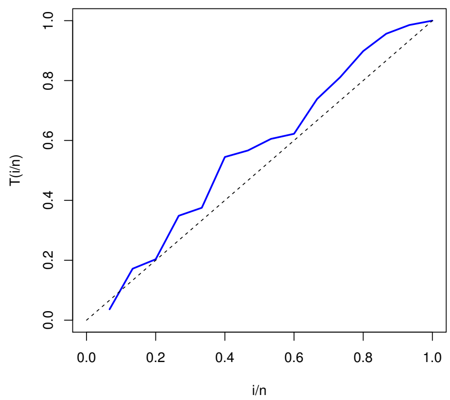


Figure 10: The PP plots for the failure times data.

It is clear that the DBL, DFx-I, Geo, DR, DGE-II, DIW, DLog-L and DLo distributions are suitable choices for this data set. However, the DBL distribution is the best choice since it has lowest value of the K-S test statistic. Figure 11 shows the TTT plot and log-likelihood profile of $\widehat{p}$, where $L = -64.784$.



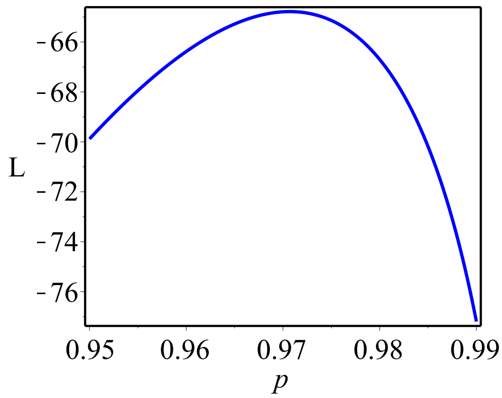


Figure 11: The TTT plot (left panel) and log-likelihood profile of $\widehat{p}$ (right panel) for the failure times data.

Regarding Figure 11, it is clear that the shape of the hrf can be increasing and the log-likelihood profile of $\widehat{p}$ is unimodal-shaped. Table 10 shows the estimation of the proposed model using the MM for the failure times data.

Table 10: Estimation and goodness of fit test for the failure times data.

Method	Statistic		
	p	K-S	p-value
MM	0.971	0.109	0.994

According to the p-value of the K-S test, MM method works quite well besides the MLE method for estimating the unknown parameter. But the MM is the best. Using the MM estimator of the parameter p, some statistics of the DBL distribution are reported in Table 11.

Table 11: Some descriptive statistics for data set II.

Method	Statistic							
	Mean	Mode	Variance	DI	MD	CV	Skewness	Kurtosis
MM	27.816	13.284	417.044	14.992	15.533	0.734	1.493	6.442

The data herein is suffering from over dispersion phenomena as DI > 1. Furthermore, it is moderately skewed right with leptokurtic.

6.3. Burglary crimes

The performance of the INAR(1)DBL process is compared with the INAR(1)P, INAR(1)PL and INAR(1)G processes. The one-step translation probabilities of the competitive INAR(1) models are given below:

1. INAR(1)P

$$\Pr\left(X_t = k \middle| X_{t-1} = l\right) = \sum_{i=0}^{\min(k,l)} \binom{l}{i} \alpha^i (1-\alpha)^{l-i} \frac{\exp(-\lambda)\,\lambda^{k-i}}{(k-i)!}, \quad \lambda > 0.$$

2. INAR(1)PL

$$\Pr\left(X_t = k \middle| X_{t-1} = l\right) = \sum_{i=0}^{\min(k,l)} \binom{l}{i} \alpha^i (1-\alpha)^{l-i} \frac{\theta^2\,(k-i+\theta+2)}{(\theta+1)^{k-i+3}}, \quad \theta > 0,$$

3. INAR(1)G

$$\Pr\left(X_t = k \middle| X_{t-1} = l\right) = \sum_{i=1}^{\min(k,l)} \binom{l}{i} \alpha^i (1-\alpha)^{l-i} \left[p(1-p)^{k-i}\right], \quad 0 < p < 1.$$

The CML estimation method is used to obtain unknown parameters of INAR(1)DBL, INAR(1)PL, INAR(1)G and INAR(1)P models. To decide the best model, two information criteria are used: Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). The smallest values of AIC and BIC and indicate the best fitted model on the data set.

The series of monthly counts of burglary crimes in the 22th police car beat in Pittsburgh is used to compare the performance of INAR(1)DBL, INAR(1)PL, INAR(1)G and INAR(1)P processes. The data set consists of 144 monthly observations between the date of January 1990 and December 2001 and is given in the Appendix. The data set can be also found in `http://www.forecastingprinciples.com/index.php/crimedata`. The mean, variance and DI values of the used data set are 6.111, 13.372 and 2.188, respectively. It is clear that monthly counts of burglary crimes exhibit over-dispersion. So, the innovation distribution of INAR(1) process should be able to model over-dispersion. Therefore, INAR(1) process with DBL innovations could be a good choice to model these data set.

The autocorrelation function (ACF) and partial ACF plots of the used data set are displayed in Figure 12. As seen from these plots, ACF has clear cut-off after the first lag. Therefore, AR(1) process could be a good choice for analyzing these data set.

The estimated parameters of the fitted INAR(1) process and model selection criteria are listed in Table 12. Since the INAR(1)DBL model has the smaller values of AIC and BIC statistics than those of INAR(1)P, INAR(1)PL and INAR(1)G processes, the INAR(1)DBL process provides better fits than other competitive INAR(1) processes. More importantly, the obtained DI value of INAR(1)DBL process is very near the empirical one. It is obvious that INAR(1)DBL astoundingly explains the characteristics of the data set.

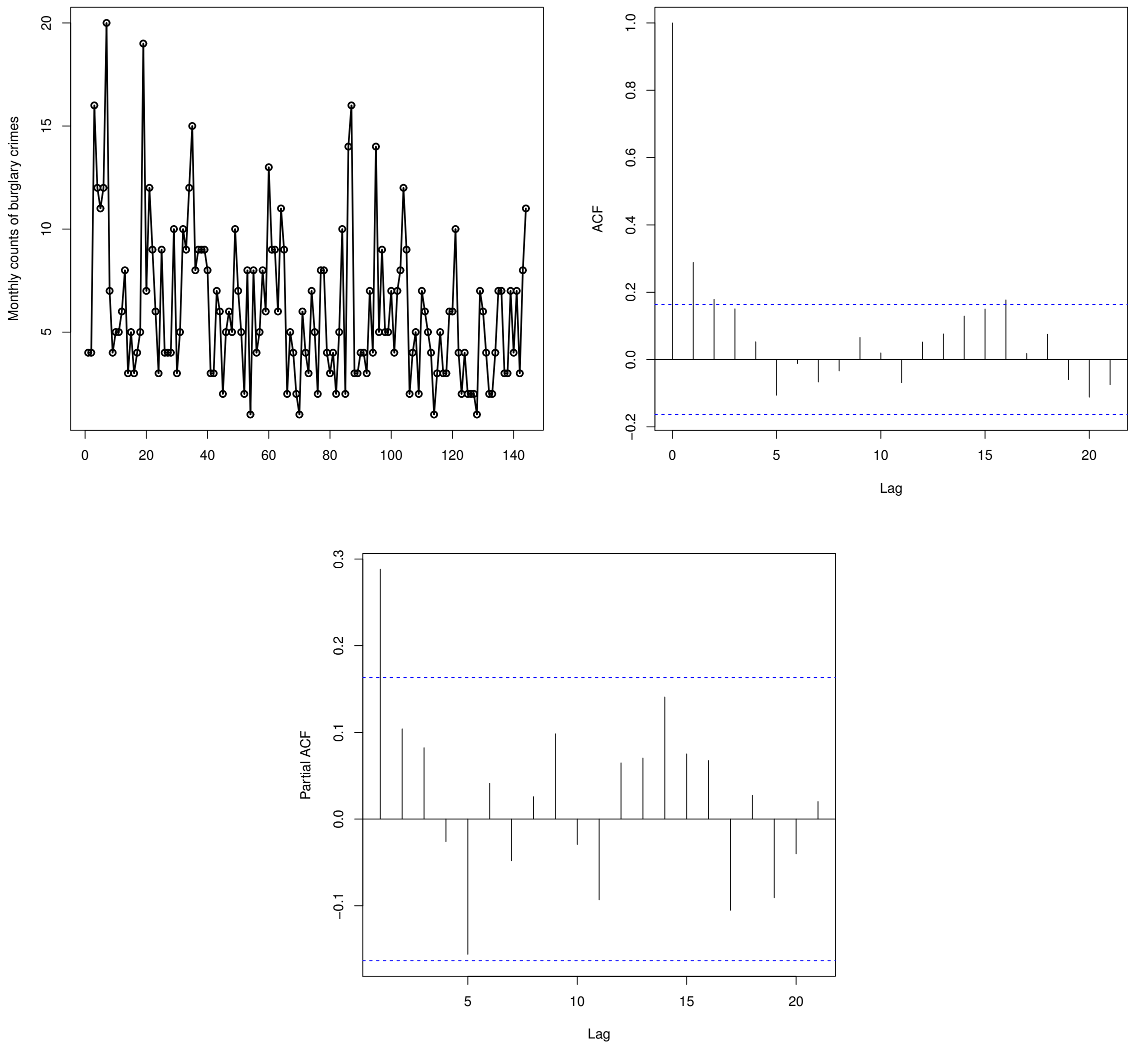


Figure 12: The plots of monthly counts of burglary crimes and its corresponding ACF and PACF plots.

Table 12: The CML estimates of INAR(1)DBL and INAR(1)P process and goodness-of-fit statistics.

Model	Parameters	Estimate	Std-er	AIC	BIC	μ_X	σ_X^2	DI
INAR(1)DBL	α p	0.3032 0.8402	0.0467 0.0121	733.1232	739.0628	6.1505	14.6336	2.3792
INAR(1)PL	α θ	0.3842 0.4451	0.0365 0.0147	739.8960	745.8356	6.1731	17.4559	2.8277
INAR(1)G	α p	0.4319 0.2221	0.0376 0.0192	747.7226	753.6622	6.1649	21.2445	3.4460
INAR(1)P	α λ	0.1952 4.9402	0.0194 0.0537	778.3730	784.3126	6.1381	6.1381	1
Empirical						6.1111	13.3722	2.1882

Additionally, the residual analysis is conducted to evaluate the accuracy of the fitted INAR(1)DBL model for the data used. The Pearson residuals of the INAR(1)DBL process are given by

$$r_t = \frac{X_t - \mathrm{E}\,(\,X_t|\,X_{t-1})}{\mathrm{Var}(\,X_t|\,X_{t-1})^{1/2}} \tag{6.1}$$

where $\mathrm{E}\,(\,X_t|\,X_{t-1})$ and $\mathrm{Var}\,(\,X_t|\,X_{t-1})$ are defined in (4.9) and (4.10), respectively. When the fitted INAR(1) process is valid for the modeled data, the Pearson residuals should have zero mean and unit variance as well as uncorrelated. The Pearson residuals of the INAR(1)DBL process are calculated by using (6.1). The mean and variance of these residuals are obtained as 0.0005 and 0.9917, respectively. The obtained values of the mean and variance of the Pearson residuals are very closed to the desired values. Moreover, the predicted values of the burglary crimes and the ACF plot of the Pearson residuals are displayed in Figure 13 which ensures that the residuals are uncorrelated.

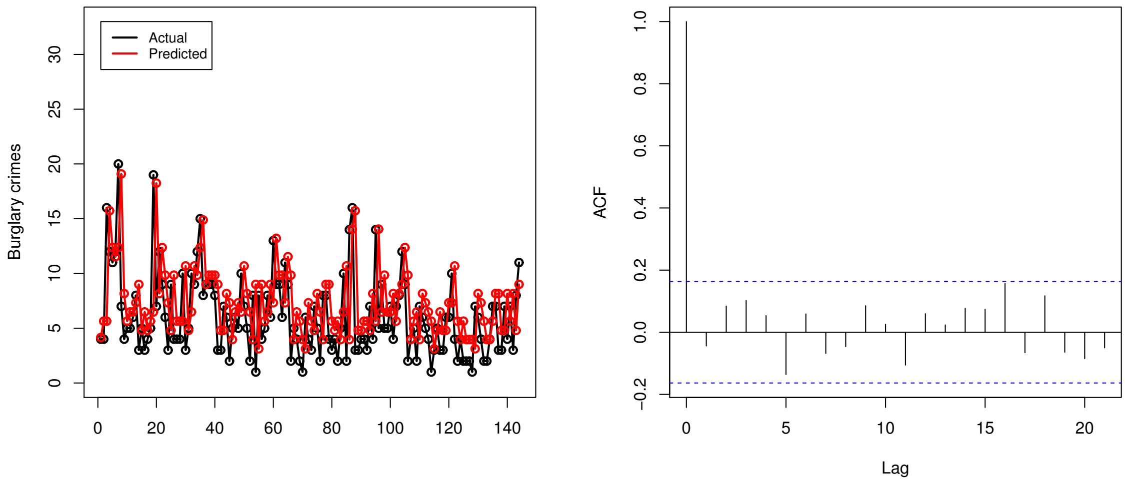


Figure 13: The predicted values of the burglary crimes (left) and the ACF plot of the Pearson residuals (right).

7. CONCLUSIONS

A new one-parameter discrete model is introduced. The statistical properties of proposed model are studied extensively. Two parameter estimation method are used. These are the maximum likelihood and method of moments estimation methods. The relative efficiency of parameter estimation methods are discussed via simulation study. Three applications to three real data sets are given to convince the readers in favour of DBL model. Empirical findings show that the DBL model is an attractive model and produce more reliable results than other its counterparts. More importantly, INAR(1) process with DBL innovations produce better results than INAR(1)P model in case of over-dispersion. We hope that DBL distribution gains much more attention and is applied to wider range of application fields.

A. APPENDIX

The data set used in Section 6.1:

Number of fires:	0	1	2	3	4	5	6	7	8	9	10	11	12	15	16	20	43
Observed values:	16	13	14	9	11	13	8	4	9	6	3	4	6	4	1	1	1

The data set used in Section 6.2:

1.0, 5.0, 6.0, 11.0, 12.0, 19.0, 20.0, 22.0, 23.0, 31.0, 37.0, 46.0, 54.0, 60.0, 66.0

The data set used in Section 6.3:

4	4	16	12	11	12	20	7	4	5	5	6	8	3	5	3	4	5	19	7
12	9	6	3	9	4	4	4	10	3	5	10	9	12	15	8	9	9	9	8
3	3	7	6	2	5	6	5	10	7	5	2	8	1	8	4	5	8	6	13
9	9	6	11	9	2	5	4	2	1	6	4	3	7	5	2	8	8	4	3
4	2	5	10	2	14	16	3	3	4	4	3	7	4	14	5	9	5	5	7
4	7	8	12	9	2	4	5	2	7	6	5	4	1	3	5	3	3	6	6
10	4	2	4	2	2	2	1	7	6	4	2	2	4	7	7	3	3	7	4
7	3	8	11																

ACKNOWLEDGMENTS

The author El-Morshedy would like to acknowledge that this publication was supported by the Deanship of Scientific Research at Prince Sattam bin Abdulaziz University, Alkharj, Saudi Arabia.

REFERENCES

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

[2] Altun, E. (2020a). A new generalization of geometric distribution with properties and applications, *Communications in Statistics – Simulation and Computation*, **49**(3), 793–807.

[3] Altun, E. (2020b). A new one-parameter discrete distribution with associated regression and integer-valued autoregressive models, *Mathematica Slovaca*, **70**(4), 979–994.

[4] Abd-Elrahman, A.M. (2013). Utilizing ordered statistics in lifetime distributions production: a new lifetime distribution and applications, *Journal of Probability and Statistical Science*, **11**(2), 153–164.

[5] 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.

[6] Buddana, A. and Kozubowski, T.J. (2014). Discrete Pareto distributions, *Economic Quality Control*, **29**(2), 143–156.

[7] Eliwa, M.S. and El-Morshedy, M. (2020). A one-parameter discrete distribution for over-dispersed data: statistical and reliability properties with applications, *Journal of Applied Statistics*, forthcoming.

[8] Eliwa, M.S.; Alhussain, Z.A. and El-Morshedy, M. (2020a). Discrete Gompertz-G family of distributions for over-and under-dispersed data with properties, estimation, and applications, *Mathematics*, **8**(3), 358.

[9] Eliwa, M.S.; Altun, E.; El-Dawoody, M. and El-Morshedy, M. (2020b). A new three-parameter discrete distribution with associated INAR(1) process and applications, *IEEE Access*, **8**, 91150–91162.

[10] El-Morshedy, M.; Eliwa, M.S. and Nagy, H. (2019). A new two-parameter exponentiated discrete Lindley distribution: properties, estimation and applications, *Journal of Applied Statistics*, `https://doi.org/10.1080/02664763.2019.1638893`.

[11] El-Morshedy, M.; Eliwa, M.S. and Altun, E. (2020). Discrete Burr-Hatke distribution with properties, estimation methods and regression model, *IEEE Access*, **8**, 74359–74370.

[12] Gómez-Déniz, E. and Calderín-Ojeda, E. (2011). The discrete Lindley distribution: properties and applications, *Journal of Statistical Computation and Simulation*, **81**(11), 1405–1416.

[13] Hussain, T. and Ahmad, M. (2014). Discrete inverse Rayleigh distribution, *Pakistan Journal of Statistics*, **30**(2), 203–222.

[14] Jazi, A.M.; Jones, G. and Lai, C.D. (2012). Integer valued AR(1) with geometric innovations, *Journal of the Iranian Statistical Society*, **11**(2), 173–190.

[15] Jazi, A.M.; Lai, D.C. and Alamatsaz, H.M. (2010). Inverse Weibull distribution and estimation of its parameters, *Statistical Methodology*, **7**(2), 121–132.

[16] Karlis, D. and Xekalaki, E. (2001). *On some discrete valued time series models based on mixtures and thinning.* In "Proceedings of the Fifth Hellenic-European Conference on Computer Mathematics and its Applications", pp. 872–877.

[17] Krishna, H. and Pundir, P.S. (2009). Discrete Burr and discrete Pareto distributions, *Statistical Methodology*, **6**, 177–188.

[18] Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data*, Wiley, New York.

[19] L'ivio, T.; Khan, N.M.; Bourguignon, M. and Bakouch, H.S. (2018). An INAR(1) model with Poisson-Lindley innovations, *Economics Bulletin*, **38**(3), 1505–1513.

[20] McKenzie, E. (1985). Some simple models for discrete variate time series, *Journal of the American Water Resources Association*, **21**, 645–650.

[21] Nakagawa, T. and Osaki, S. (1975). The discrete Weibull distribution, *IEEE Transactions on Reliability*, **24**(5), 300–301.

[22] Nekoukhou, N.; Alamatsaz, M.H. and Bidram, H. (2013). Discrete generalized exponential distribution of a second type, *Statistics*, **47**(4), 876–887.

[23] Nekoukhou, V. and Bidram, H. (2015). The exponentiated discrete Weibull distribution, *SORT*, **39**(1), 127–146.

[24] Para, B.A. and Jan, T.R. (2016a). On discrete three-parameter Burr type XII and discrete Lomax distributions and their applications to model count data from medical science, *Biometrics and Biostatistics International Journal*, **4**(2), 1–15.

[25] Para, B.A. and Jan, T.R. (2016b). Discrete version of log-logistic distribution and its applications in genetics, *International Journal of Modern Mathematical Sciences*, **14**(4), 407–422.

[26] Para, B.A. and Jan, T.R. (2017). Discrete generalized Weibull distribution: properties and applications in medical sciences, *Pakistan Journal of Statistics*, **33**, 337–354.

[27] Poisson, S.D. (1837). *Recherches sur la Probabilité des Jugements en Matiére Criminelle et en Matiére Civile, Précédées des Régles Générales du Calcul des Probabilitiés*, Paris, France: Bachelier, pp. 206–207.

[28] Roy, D. (2004). Discrete Rayleigh distribution, *IEEE Transactions on Reliability*, **53**(2), 255–260.