

Modeling and Applications to Circular Data with a Wrapped Poisson-Lindley Model

Christophe Chesneau^{1,*}, Hassan S. Bakouch² and Tassaddaq Hussain³

¹ *Université de Caen, LMNO, Campus II, Science 3, Caen, France.*

² *Department of Mathematics, Faculty of Science, Tanta University, Tanta, Egypt.*

³ *Department of Mathematics, Mirpur University of Science and Technology (MUST), Mirpur, Pakistan.*

Received April 12, 2021; Accepted July 7, 2021;

Published online February 10, 2022.

Abstract. In the recent years, researchers in many fields, including meteorology, economics, sociology, psychology, and epidemiology, have shown a keen interest in the analysis and modeling of wrapped data. This motivates the development of new wrapped distributions and related models. In this paper, a simple and new wrapped model based on the Poisson-Lindley distribution is developed. Many of its properties are obtained, such as probability mass and cumulative distribution functions, survival and hazard rate functions, and probability generating function. The estimation of the model parameter is investigated by the maximum likelihood method. Test and evaluation statistics are also considered to assess the performance of the distribution among the most frequently wrapped discrete probability models using three different circular practical data sets.

AMS subject classifications: 60E05, 62E15, 62F10

Key words: Wrapped distributions, Poisson-Lindley distribution, estimation, circular data, evaluating tests.

1 Introduction

In the current era, many diverse statistical fields are emerging, such as biostatistics, environmental statistics, geostatistics, statistical mechanics, statistical computing and management statistics. However, very little attention is paid to the directional statistics that deal with direction axes or rotations. Such directional statistics include the direction of the earth's magnetic pole, the direction of flight of a bird or the orientation of an animal,

*Corresponding author. *Email addresses:* christophe.chesneau@gmail.com (Chesneau C), hassan.bakouch@science.tanta.edu.eg (Bakouch H S), taskho2000@yahoo.com (Hussain T)

the arrival times of patients in an emergency clinic, incidences of a disease throughout the year, and the number of tourists (daily or monthly) in a city within a year, where the calendar is regarded as a one-year clock, the daily wind directions, the ocean current directions, departure directions of animals, direction of bone-fracture planes, and orientation of bees in a beehive after stimuli, etc.

The directional statistics mentioned above have many innovative landscapes, both in terms of modeling and statistical treatment. For this purpose, researchers have to adopt different ways to obtain circular distributions. One of them is wrapping a linear distribution around the unit circle. In this regard, [11] is the pioneer of wrapped distributions. Since this innovation, a lot of work has been done in this field. Some of them include the wrapped normal (WN) distribution, wrapped Cauchy (WC) distribution, wrapped Poisson distribution by [7], wrapped exponential and Laplace distributions by [6], wrapped lognormal, logistic, Weibull, and extreme-value distributions by [14], wrapped skew Laplace distribution by [8], wrapped geometric distribution, wrapped gamma distribution by [13], wrapped variance gamma distribution by [1], wrapped binomial distribution by [4], wrapped log Kumaraswamy distribution by [9], discrete wrapped exponential and negative binomial distributions by [17, 18] and, recently, [10] introduced the wrapped Lindley distribution. As a matter of fact, the existing wrapped distributions literature lacks the discrete wrapped distributions. Moreover, probability functions of the presented wrapped models are also not in the closed form and, ultimately, mathematically, are not amenable.

Now, let us briefly recall the methodology for the construction of a new wrapped distribution. Let X be a discrete random variable with integer values. Then, the corresponding wrapped distribution is the distribution of the following random variable: $X_w = 2\pi X \pmod{2\pi m}$, which has the support $\{2\pi r/m, r=0,1,\dots,m-1\}$. We now suppose that X follows the Poisson-Lindley distribution with parameter $\theta > 0$ introduced by [15] and characterized by the following probability mass function (pmf):

$$f(x;\theta) = \frac{\theta^2(x+\theta+2)}{(1+\theta)^{x+3}}, \quad x=0,1,2,\dots \quad (1.1)$$

It is of particular interest because of the following notable properties: tractable probability functions, unimodality, overdispersion, infinite divisibility and increasing hazard rate function (hrf). Also, in some circumstances, it provides a suitable alternative to the Poisson, geometric and negative binomial distributions (in particular, it is known to have smaller skewness and kurtosis than the negative binomial distribution). Also, the model in Eq. (1.1) was derived based on an interesting property of the linear exponential family of single parameter distributions (see [15]). It contains some interesting properties, including approximation to the negative binomial and Hermite distributions, and has a wide range of application in many fields, such as medicine, engineering, ecology and genetics (see [16]). Further, this distribution can be represented as a mixture of geometric with parameter $1/(1+\theta)$ and negative binomial with parameters 2 and $1/(1+\theta)$, where the mixing proportions are $\theta/(1+\theta)$ and $1/(1+\theta)$, respectively. More details on

the Poisson-Lindley distribution can be found in [15] and [3]. Hence, by the use of the Poisson-Lindley distribution, we aim to construct a new wrapped discrete distribution with attractive properties. We thus define the wrapped Poisson-Lindley (WPL) distribution as the distribution of the related random variable X_w and we set $X_w \sim \text{WPL}(m, \theta)$. As a starting point, the pmf of X_w is given by

$$P_{X_w}(r) = P\left(X_w = \frac{2\pi r}{m}\right) = \sum_{k=-\infty}^{+\infty} f(r+km; \theta), \quad r=0, 1, \dots, m-1.$$

The present paper investigates mathematical and practical properties of the WPL distribution. Other motivational factors include the following ones. First, unlike the other wrapped probability models, the WPL distribution has a closed form structure of pmf, cumulative distribution function (cdf), hrf and survival function (sf). Secondly, so far none of the researchers have studied applications of discrete wrapped probability models. Last but not least, parameter estimation is straightforward and efficient. A final motivation is that the pmf of the WPL distribution can be represented as a mixture of wrapped geometric model and wrapped negative binomial model, where the latter is not proposed yet in literature.

The paper is outlined as follows. Section 2 provides the main functions of the WPL distribution. Section 3 is devoted to estimation of the parameters via the maximum likelihood method, and a simulation study to show the numerical efficiency of the obtained estimates. Section 4 deals with the application on three lifetime data sets. Section 5 concludes the results obtained.

2 Main functions

2.1 Probability mass function

Here, we express the main functions related to the WPL distribution. We begin with the corresponding pmf. By incorporating Eq. (1.1) in the above expression, we have

$$P_{X_w}(r) = \sum_{k=-\infty}^{\infty} f(r+km; \theta) = \sum_{k=0}^{\infty} \frac{\theta^2(r+km+\theta+2)}{(1+\theta)^{r+km+3}}, \quad r=0, 1, \dots, m-1.$$

After some algebra, we obtain

$$\begin{aligned} P_{X_w}(r) &= \sum_{k=-\infty}^{+\infty} f(r+km; \theta) = \sum_{k=0}^{+\infty} \frac{\theta^2(r+km+\theta+2)}{(1+\theta)^{r+km+3}} \\ &= \frac{\theta^2}{(1+\theta)^{r+3}} \left[(r+\theta+2) \sum_{k=0}^{+\infty} \left\{ \frac{1}{(1+\theta)^m} \right\}^k + \frac{m}{(1+\theta)^m} \sum_{k=0}^{+\infty} k \left\{ \frac{1}{(1+\theta)^m} \right\}^{k-1} \right] \\ &= \frac{\theta^2}{(1+\theta)^{r+3}} \left[(r+\theta+2) \frac{(1+\theta)^m}{(1+\theta)^m - 1} + \frac{m}{(1+\theta)^m} \frac{(1+\theta)^{2m}}{[(1+\theta)^m - 1]^2} \right] \end{aligned}$$

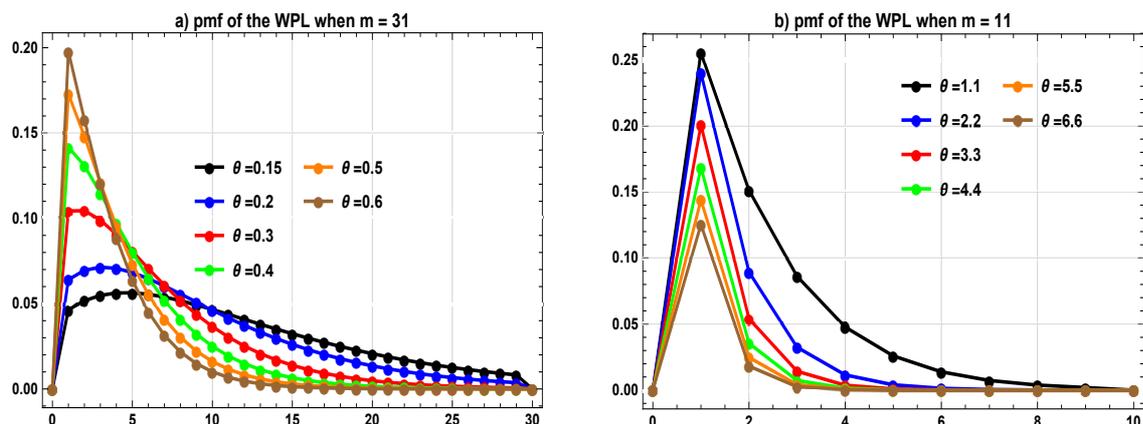


Figure 1: Linear representation for the pmf of the WPL distribution for the indicated values.

$$= \frac{\theta^2(1+\theta)^m}{(1+\theta)^{r+3}[(1+\theta)^m - 1]^2} \left\{ (r+\theta+2)[(1+\theta)^m - 1] + m \right\}.$$

Also, the pmf of the WPL distribution can be represented as a mixture of wrapped geometric (WG) distribution with parameter $1/(1+\theta)$ and wrapped negative binomial (WNB) distribution with parameters 2 and $1/(1+\theta)$, where the mixing proportions are $\theta/(1+\theta)$ and $1/(1+\theta)$, respectively. Noting that both $WG(1/(1+\theta))$ and $WNB(2,1/(1+\theta))$ are perfect pmfs and the sum of each is one. Further, one can check that

$$\sum_{r=0}^{m-1} \frac{\theta^2(1+\theta)^m}{(1+\theta)^{r+3}[(1+\theta)^m - 1]^2} \left\{ (r+\theta+2)[(1+\theta)^m - 1] + m \right\} = 1.$$

Also, by denoting $P_r = P_{X_w}(r)$, the following recursive relation between probabilities holds

$$P_{r+1}(1+\theta)(m+(2+r+\theta)((1+\theta)^m - 1)) = (m+(3+r+\theta)((1+\theta)^m - 1))P_r.$$

Now, let

$$a_m(\theta) = \frac{\theta^2(1+\theta)^{m-3}}{[(1+\theta)^m - 1]^2} \left\{ (\theta+2)[(1+\theta)^m - 1] + m \right\}, \quad b_m(\theta) = \frac{\theta^2(1+\theta)^{m-4}}{(1+\theta)^m - 1}.$$

Then, using those equalities, the pmf of the WPL distribution can be rewritten as

$$P_{X_w}(r) = \frac{1}{(1+\theta)^r} (a_m(\theta) + b_m(\theta)r(1+\theta)), \quad r = 0, 1, 2, \dots, m-1. \tag{2.1}$$

Figures 1 and 2 illustrate the pmf of the WPL distribution for various values of the parameters, and with linear and circular representations, respectively. Various modes, intensity of the peaks and increasing-decreasing structures can be observed.

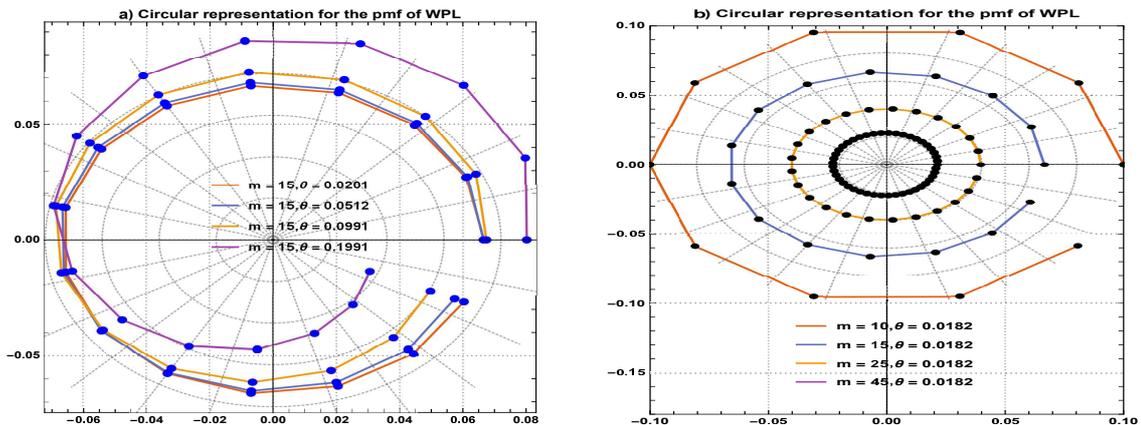


Figure 2: Circular representation for the pmf of the WPL distribution for the indicated values.

2.2 Cumulative distribution function

The above letting, as mentioned earlier, also helps us to calculate some of its important properties, such as the cdf which usually calculates the failure probabilities before a specific number, say r , and for $r=0,1, \dots, m-1$, it has the following form:

$$\begin{aligned}
 F(r) &= P(X_w \leq r) = \sum_{k=0}^r P_{X_w}(k) = \sum_{k=0}^r \left(a_m(\theta) \frac{1}{(1+\theta)^k} + b_m(\theta) \frac{k}{(1+\theta)^{k-1}} \right) \\
 &= a_m(\theta) \sum_{k=0}^r \frac{1}{(1+\theta)^k} + b_m(\theta) \sum_{k=0}^r \frac{k}{(1+\theta)^{k-1}} \\
 &= a_m(\theta) \left\{ \frac{1+\theta - (1+\theta)^{-r}}{\theta} \right\} \\
 &\quad + b_m(\theta) \left\{ \frac{(1+\theta)^{1-r} \{ (1+\theta)^r + \theta [(1+\theta)^r - 1 - r] - 1 \}}{\theta^2} \right\}. \tag{2.2}
 \end{aligned}$$

After some algebra, we obtain

$$\begin{aligned}
 F(r) &= \frac{1}{[(1+\theta)^m - 1]^2} (1+\theta)^{-r+m-3} \left\{ 1 - (1+\theta)^m + (1+\theta)^{1+r} \left(-1 - 2\theta + m\theta - \theta^2 \right. \right. \\
 &\quad \left. \left. + (1+\theta)^{2+m} \right) + \theta(-m - (3+r+\theta)[(1+\theta)^m - 1] \right\}.
 \end{aligned}$$

Figures 3 and 4 present some graphics of the cdf of the WPL distribution for various values of the parameters, and with linear and circular representations, respectively. We see that the cdf is concave with various inclinations depending on the values of x and the parameters. Unit spirals of various forms are observed in the circular representation.

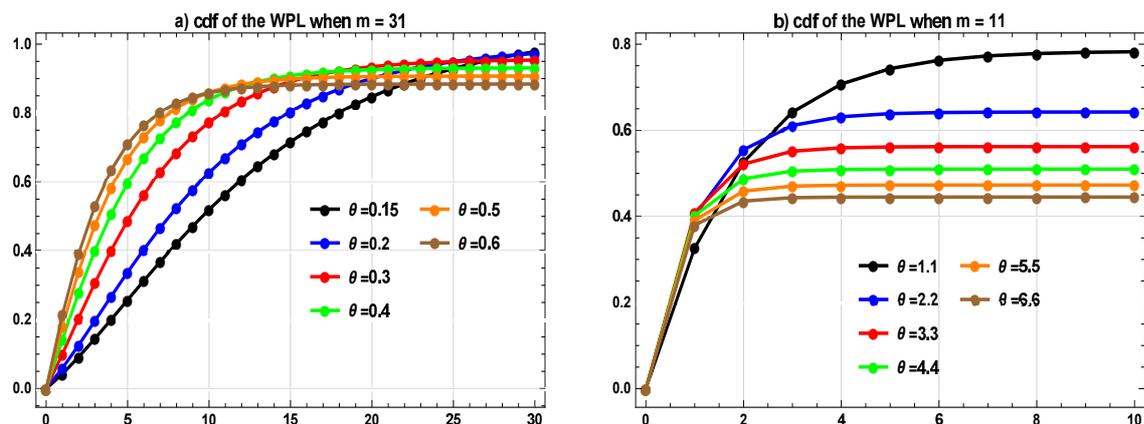


Figure 3: Linear representation for the cdf of the WPL distribution for the indicated values.

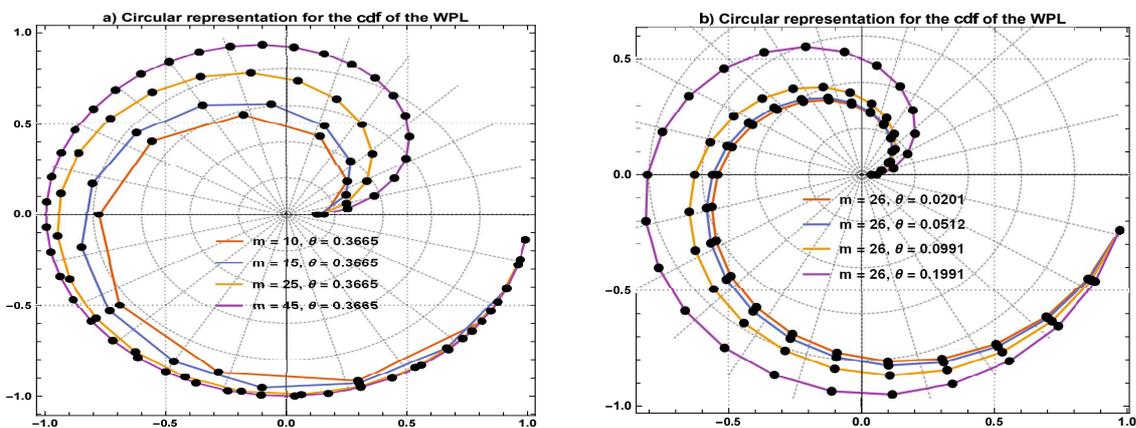


Figure 4: Circular representation for the cdf of the WPL distribution for the indicated values.

From both linear and circular representations of the cdf of the WPL distribution, it is obvious that the convergence is sharp to 1 for smaller values of θ , i.e., as $\theta \rightarrow 0$, as compared to larger values of θ . In the case of the linear representation, it is just linearly increasing. However, for polar form it moves in an anticlockwise direction and approaches to 1. Moreover, we see that $F(0) = 0$ and $F(m - 1) = 1$.

2.3 Survival and hazard rate functions

Similarly, another important reliability characteristic is the sf which usually calculates the probabilities beyond a certain number r and it is expressed as $S(r) = 1 - F(r)$, hence we

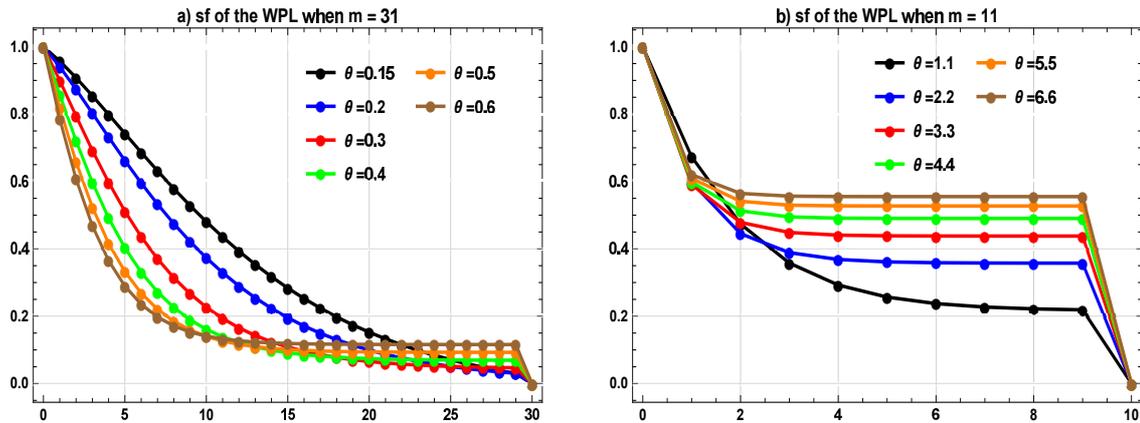


Figure 5: Linear representation of the sf of the WPL distribution for the indicated values.

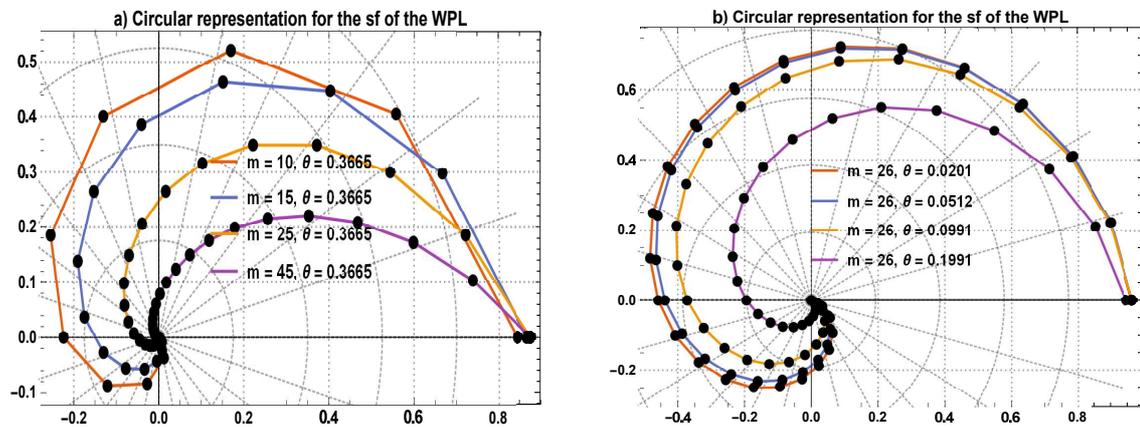


Figure 6: Circular representation of the sf of the WPL distribution for the indicated values.

have

$$\begin{aligned}
 S(r) = & 1 - \frac{1}{[(1+\theta)^m - 1]^2} (1+\theta)^{-r+m-3} \left\{ 1 - (1+\theta)^m \right. \\
 & + (1+\theta)^{1+r} (-1 - 2\theta + m\theta - \theta^2 + (1+\theta)^{2+m}) \\
 & \left. + \theta(-m - (3+r+\theta)[(1+\theta)^m - 1]) \right\}. \tag{2.3}
 \end{aligned}$$

Other important characteristics also include the hrf. It is usually defined as

$$h(r) = \frac{P_{X_w}(r)}{S(r)}.$$

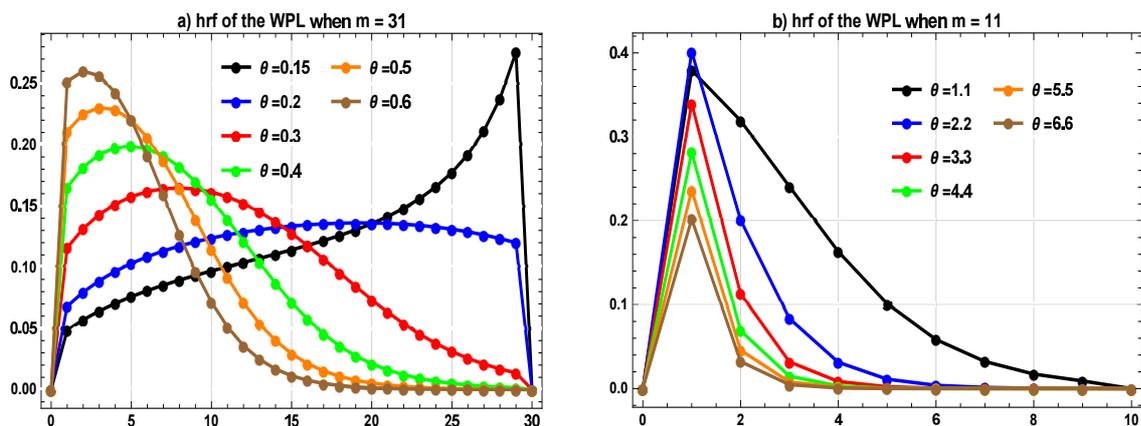


Figure 7: Linear representation of the hrf of the WPL distribution for the indicated values.

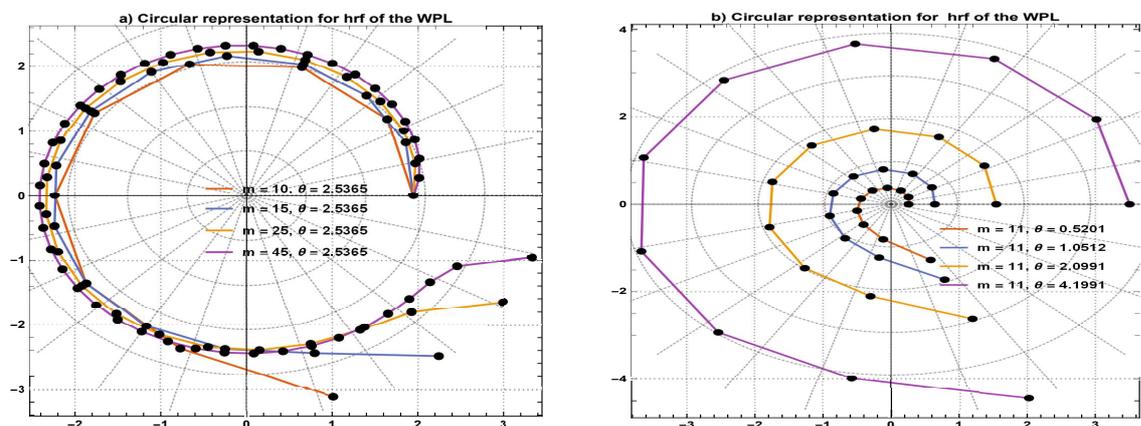


Figure 8: Circular representation of the hrf of the WPL distribution for the indicated values.

So, for the WPL distribution, it can be expressed by combining the already obtained expressions for $P_{X_w}(r)$ and $S(r)$. Figures 5 and 6 present some graphics of the sf of the WPL distribution for various values of the parameters, and with linear and circular representations, respectively. We can see that the sf converges sharply to 0 for smaller values of θ , i.e., as $\theta \rightarrow 0$, whereas for larger values of θ , the sf decreases linearly. However, for the polar form, it moves in a clockwise direction and approaches to 0.

Figures 7 and 8 display graphics for the hrf of the WPL distribution for various values of the parameters, and with linear and circular representations, respectively. Various forms of shapes are observed (decreasing, increasing, convex and concave shapes), ensuring the great flexibility of the related model and motivating the use of the WPL model

in a data analysis setting.

2.4 On the probability generating function

If $X_w \sim \text{WPL}(m, \theta)$, then the probability generating function of the random variable X_w is expressed as

$$\begin{aligned} G_{X_w}(t) &= E(t^{X_w}) = \sum_{r=0}^{m-1} t^{\frac{2\pi r}{m}} P\left(X_w = \frac{2\pi r}{m}\right) \\ &= \sum_{r=0}^{m-1} t^{\frac{2\pi r}{m}} \left[a_m(\theta) \frac{1}{(1+\theta)^r} + b_m(\theta) \frac{r}{(1+\theta)^{r-1}} \right] \\ &= a_m(\theta) \sum_{r=0}^{m-1} \left(\frac{t^{\frac{2\pi}{m}}}{1+\theta} \right)^r + b_m(\theta) t^{\frac{2\pi}{m}} \sum_{r=0}^{m-1} r \left(\frac{t^{\frac{2\pi}{m}}}{1+\theta} \right)^{r-1}. \end{aligned}$$

Thus, after some algebra, we get

$$\begin{aligned} G_{X_w}(t) &= a_m(\theta) \frac{1+\theta - (1+\theta)^{1-m} t^{2\pi}}{1+\theta - t^{\frac{2\pi}{m}}} \\ &\quad + b_m(\theta) t^{\frac{2\pi}{m}} \frac{(m-1)(1+\theta)^{2-m} t^{2\pi} - m(1+\theta)^{3-m} t^{2\pi} t^{-\frac{2\pi}{m}} + (1+\theta)^2}{(1+\theta - t^{\frac{2\pi}{m}})^2}. \end{aligned} \quad (2.4)$$

From $G_{X_w}(t)$, one can deduce several important measures on X_w , as the r^{th} descending moment given by $\mu_r^* = E[X_w(X_w-1)\dots(X_w-r+1)] = G_{X_w}(t)^{(r)}|_{t=1}$. For instance, by taking $r=1$, we get the mean of X_w as

$$\begin{aligned} E(X_w) &= \frac{1}{m\theta(1+\theta)^4 [(1+\theta)^m - 1]^2} \\ &\quad \times 2\pi \left\{ -m^2\theta^2 + (1+\theta)(2+\theta)[(1+\theta)^m - 1]^2 - m\theta[(1+\theta)^m - 1](1+\theta(2+m+\theta)) \right\}. \end{aligned}$$

The variance and other related measures can be expressed in a similar way and they remain available upon reader request.

3 Parameter estimation with inference

3.1 Maximum likelihood method

Let $X_{w_1}, X_{w_2}, \dots, X_{w_n}$ be a random sample drawn identically and independently from the WPL distribution with observed values $2\pi r_1/m, 2\pi r_2/m, \dots, 2\pi r_n/m$, assuming that m is known in advance (this point will be discussed later from the practical point of view).

Table 1: Simulation study for the MLE $\hat{\theta}$ of the WPL distribution, assuming $m=5$.

	$\theta = 0.5025$			$\theta = 0.9856$		
n	Bias	Variance	MSE	Bias	Variance	MSE
15	0.0400	0.1463	0.1464	0.0632	0.2393	0.2417
25	0.0165	0.0350	0.0349	0.0619	0.1041	0.1074
50	0.0153	0.0134	0.0172	-0.0097	0.0435	0.0432
150	0.0133	0.0048	0.0172	0.0134	0.0161	0.0165
350	0.0115	0.0047	0.0098	0.0230	0.0071	0.0076
500	0.0106	0.0024	0.0062	0.0205	0.0057	0.0064
	$\theta = 1.0132$			$\theta = 11.2315$		
n	Bias	Variance	MSE	Bias	Variance	MSE
15	0.0729	0.2187	0.2219	1.12643×10^7	7.9302×10^{14}	7.9302×10^{14}
25	0.0591	0.1377	0.1398	76431.9	1.4161×10^{14}	1.4603×10^{14}
50	0.0518	0.0626	0.0647	2.68391	62.9337	69.5078
150	0.0297	0.0199	0.0205	1.6266	12.8332	15.3509
350	0.0288	0.0075	0.0067	0.5632	4.2654	4.5400
500	0.0212	0.0052	0.0056	0.5137	2.6486	2.8860

Then, by using the representation in Eq. (2.1), the corresponding log-likelihood function is given by

$$\ln[L_m(\theta)] = -\ln(1+\theta) \sum_{i=1}^n r_i + \sum_{i=1}^n \ln(a_m(\theta) + b_m(\theta)(1+\theta)r_i). \quad (3.1)$$

As shown below, the maximum likelihood estimate (MLE) $\hat{\theta}$ of θ can then be uniquely determined by solving the partial differentiate of Eq. (3.1) equal to 0 with respect to θ :

$$-\frac{1}{1+\theta} \sum_{i=1}^n r_i + \sum_{i=1}^n \frac{a_m(\theta)^{\theta} + [b_m(\theta)^{\theta}(1+\theta) + b_m(\theta)]r_i}{a_m(\theta) + b_m(\theta)(1+\theta)r_i} = 0,$$

where $(\phi_m(\theta))^{\theta} = \partial \phi_m(\theta) / \partial \theta$. Since $\hat{\theta}$ has no closed form, it will be computed by using the NMaximized command of Mathematica 11.0. For a large enough n , the distribution of $\hat{\theta}$ can be approximated by an univariate normal distribution with mean θ and variance $I(\theta)^{-1}$, where $I(\theta) = -E(\partial^2 \ln[L_m(\theta)] / \partial \theta^2 |_{\theta=\hat{\theta}})$.

3.2 Simulation study

In this section, we investigate the behavior and performance of the MLE discussed in the section above for a finite sample size of n . Random observations are generated by using the computational package Mathematica 10.0. We consider the following different model parameters: $\theta = 0.5025, 0.9856, 1.0132$ and 11.2315 assuming $m=5$ to carry out the simulation study and the process was repeated 1000 times by going from small to large sample sizes $n = 15, 25, 50, 150, 350$ and 500 . The simulated results are given in Table

1. We observe from Table 1 that the agreement between theory and practice improves as the sample size n increases. Mean square error (MSE) and variance of the $\hat{\theta}$ suggest that the estimate is consistent as the MSE and variance are decreasing when the sample size increases. Thus, the maximum likelihood method performs quite well in estimating the model parameters of the WPL distribution by using the NMaximized command of Mathematica 11.0.

4 Competing models, evaluating tests and applications

In this section, we study three data sets ranging from lifetime perspective to participation aspects. The sources of data sets will be mentioned.

4.1 Competing models

Due to count data modeling, we have decided to compare some wrapped discrete models along with non wrapped discrete models. For this purpose, we examine the three competing models, which include the wrapped geometric (WGD) model studied by [5], the Poisson-Lindley (PL) model proposed and investigated by [15, 16] and the well known Poisson model.

4.2 Evaluating tests

We use various discrimination criterion to model the circular data, which are especially highly recommended for testing the significance of goodness-of-fit in circular data scenarios. These statistics range from the log-likelihood function to information criterion coupled with some goodness-of-fit tests. For a brief introduction of these test statistics, let k denote the number of parameters to be fitted, $\hat{\Theta}$ be the MLEs of Θ , n denote the number of observations, m denote the number of classes, $z_i = F(x_i)$ and $x_i, i = 1, 2, \dots, n$, be the ordered observations. Hence, the considered criteria are as follow: Akaike information criterion $AIC = 2k - 2\ell(\hat{\Theta})$, Bayesian information criterion $BIC = k \log(m) - 2\ell(\hat{\Theta})$, corrected Akaike information criterion $AIC_c = AIC + 2k(k+1) / (m-k-1)$, Hannan-Quinn Information criterion $HQIC = -2\ell(\hat{\Theta}) + 2k \log(\log(m))$ and consistent Akaike information criterion $CAIC = -2\ell(\hat{\Theta}) + 2km / (m-k-1)$. The best model is recognized with the minimum values of these criteria. Moreover, let r_1, r_2, \dots, r_n be i.i.d. random variables with a cdf $F(r)$ on $r=0, 1, \dots, m-1$. Similar to fit issue on the real line, we wish to test the assumed model for the given data, i.e., $H_0: F(r) = F_0(r)$, where $F_0(r)$ is a specified cdf. However, one can broadly classify the goodness-of-fit tests into three categories: i) the empirical cdf method ii) the successive points or "spacings" gaps method and iii) the χ^2 statistic. The spacings tests are the only general class of goodness-of-fit tests that can be used as an alternative to the χ^2 test. For this reason, the χ^2 along with Kolmogorov-Smirnov or the Cramér-von Mises's tests are not appropriate tests for the circular data since their values depend on the choice of the origin. So, in order to select an appropriate model, we use

Kuiper's (V^*), modified version of Kolmogorov-Smirnov (KS), with its p -value, and Watson (U^{*2}) tests which are particularly for circular data sets. The mathematical expression of such tests are given by

$$KS = \max \left\{ \frac{i}{m} - z_i, z_i - \frac{i}{m} \right\},$$

$$V^* = \sqrt{m}V \left(1 + \frac{0.155}{\sqrt{m}} + \frac{0.24}{m} \right),$$

where V is expressed as

$$V = \max \left(z_i - \frac{i}{m} \right) - \min \left(z_i - \frac{i-1}{m} \right) + \frac{1}{m}.$$

Similarly Watson's test statistic is given by

$$U^{2*} = \left(U^2 - \frac{0.1}{m} + \frac{0.1}{m^2} \right) \left(1 + \frac{0.8}{m} \right),$$

where

$$U^2 = \sum_{i=1}^m \left(z_i - \frac{\sum_{i=1}^m z_i}{m} - \frac{2i-1}{2m} + \frac{1}{2} \right)^2 + \frac{1}{12m}.$$

Vuong Test: Chi-square approximation in the regard of the likelihood ratio test statistic is valid only for testing restrictions on the parameters of a statistical model (i.e., H_0 and H_1 are nested hypotheses). However, when models are non-nested, we can not use the likelihood ratio tests for model comparison. In this scenario, the Vuong test is useful for non-nested models. [19] proposed a likelihood ratio-based statistic for testing the null hypothesis for competing models that are equally close to the true data generating process against the alternative model that is closer. For the details of the above mentioned tests, the readers may be directed to [7] and [19].

4.3 Practical applications

In order to assess the suitability of the proposed model, we consider different competing models and apply them to three real life data sets that describe the direction in degrees around the unit circle. These data sets are, first of all, converted into floor radians. While doing so, the maximum value of floor radians is 6. In this regard, the first example deals with the turtle egg laying procedure, which is perfectly at random and in any direction around the unit circle. In this procedure, the maximum angle that turtles attain is 350 degrees in a certain direction. The same is true with wind data direction showing a maximum value of 350 degrees. However, the data set that represents the arrival directions of low mu (as measurement unit) showers of cosmic rays displays a maximum angle of 356 degrees. These factors clearly aid in determining a maximum floor radian, i.e., 6 radians,

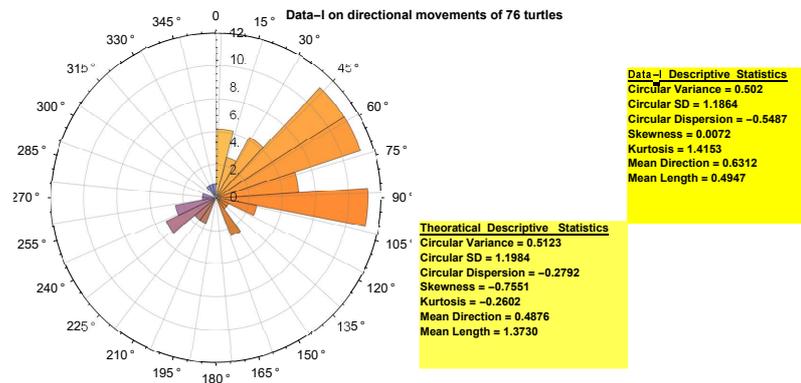


Figure 9: Circular graphic for data set I.

Table 2: MLEs and goodness-of-fit statistics of data set I.

Distribution	$\hat{\theta}$	U^{2*}	V^{2*}	KS	p -value
WPL	0.9709	0.1142	1.4627	0.3630	0.3148
WGD	0.3786	0.1147	1.4531	0.3611	0.3144
PL	1.0683	0.1357	1.5428	0.3737	0.2822
PD	1.3815	0.1538	1.5983	0.4093	0.1913

Table 3: Vuong test statistics of data set I.

WPL- WGD(p -value)	WPL- PL (p -value)	WPL- PD
5.6176(0.000)	12.7710 (0.000)	0.9688(indecisive)

implying that $m = 7$ according to the distribution domain. So the turtle egg laying/wind direction/cosmic rays showering data in the converted floor radians data indicate that the above data sets are not affected by any external factors, so the movements are all around the unit circle. However, if some external factors affect the above phenomenon, the movements of the turtles, wind directions and arrival directions of mu showering of cosmic rays become interrupted and these movements become more focused on certain directions, which seems to be impossible when nature works. Because these data sets are typically available in terms of measurement angles around the unit circle, we must first convert them from degrees to floor radians using the formula $\pi \text{ data} / 180$ before finding the MLEs and conducting the analysis. Again, it is necessary to mention that MLEs are determined by using NMaximized command of Mathematica 11.0.

Example 4.1. data set I. It is observed that directional data in two and/or three dimensions, usually arise quite frequently in many natural and physical sciences, such as biol-

Table 4: Log likelihood (l) and information criteria for data set I.

Distribution	$-l$	AIC	AICC	BIC	HQIC	CAIC
WPL	120.732	243.464	243.518	245.795	244.396	243.518
WGD	121.005	244.01	244.064	246.341	244.941	244.064
PL	122.242	246.484	246.538	248.815	247.416	246.538
PD	126.451	254.901	254.955	257.232	255.833	254.955

Table 5: MLEs and goodness-of-fit statistics of data set II.

Distribution	$\hat{\theta}$	U^{2*}	V^{2*}	KS	p -value
WPL	0.2899	0.0056	0.6544	0.0835	0.9999
WGD	0.0931	0.0056	0.6568	0.0843	0.9998
PL	0.6157	1.1614	0.0572	0.1666	0.9901
PD	2.6122	1.0821	0.0572	0.1619	0.9929

ogy, medicine, ecology, geology, etc. The following measures are the directional movements or orientations of 76 turtles after laying eggs. The data contain the measurements in degrees clockwise from north with observations 8, 38, 50, 64, 83, 98, 204, 257, 9, 38, 53, 65, 88, 100, 215, 268, 13, 40, 56, 65, 88, 103, 223, 285, 13, 44, 57, 68, 88, 106, 226, 319, 14, 45, 58, 70, 90, 113, 237, 343, 18, 47, 58, 73, 92, 118, 238, 350, 22, 48, 61, 78, 92, 138, 243, 27, 48, 63, 78, 93, 153, 244, 30, 48, 64, 78, 95, 153, 250, 34, 48, 64, 83, 96, 155, 251. This data set was reported by [7]. Analysis of the converted floor radians is given in Figure 9.

Analysis of data set I: Table 2 portrays that the proposed model is the strong competitor of the first data set, with least values of Kupier, Watson and KS test statistics. According to the Vuong test statistics given in Table 3, the proposed model is the only suitable choice for such data set. However, in the WPL-PD case, decision can not be made on the basis of the Vuong test, but the other goodness-of-fit statistics strongly suggest the suitability of the proposed model. Moreover, our findings are consolidated by the information criteria statistics which also indicate that the proposed model is a suitable model for such data sets (see Table 4). The circular representations of the estimated cdfs of the compared distributions are given in Figure 10 that shows agreement of the obtained results.

Example 4.2. data set II. The second data measure the wind directions in degrees at Gorleston England, between 11 a.m. and 12 noon on Sundays in 1968, for all of the four seasons with observations: 50, 120, 190, 210, 220, 250, 260, 290, 290, 320, 320, 340, 0, 20, 40, 60, 160, 170, 200, 220, 270, 290, 340, 350, 10, 10, 20, 20, 30, 30, 40, 150, 150, 150, 170, 190, 290, 30, 70, 110, 170, 180, 190, 240, 250, 260, 260, 290, 350. This data set was first reported by [12]. Analysis of the converted floor radians is given by Figure 11.

Analysis of data set II: Table 5 portrays that the proposed model is the strong competitor

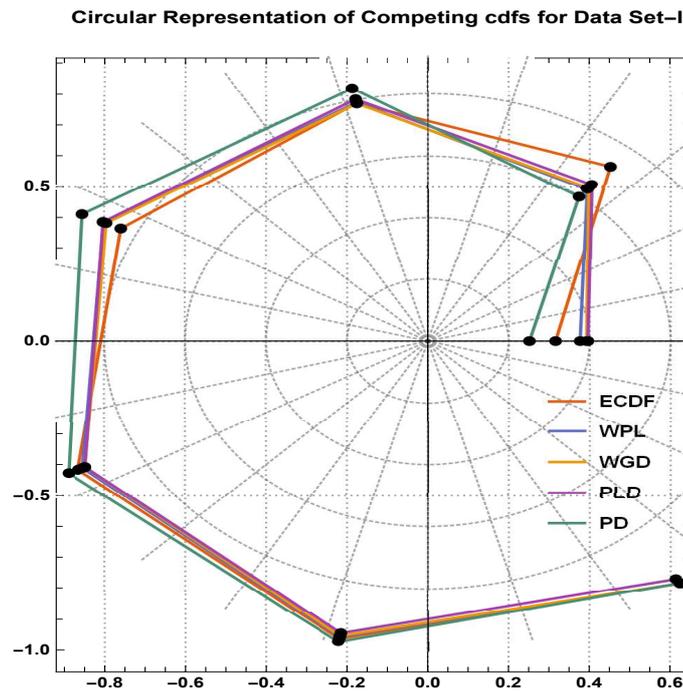


Figure 10: Circular representation of competing cdfs for data set I.

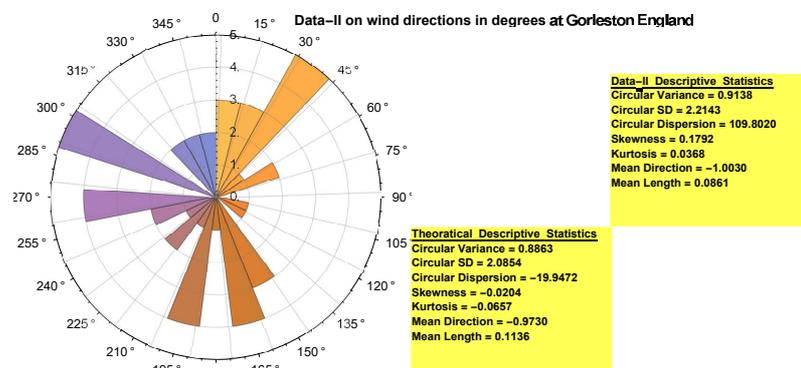


Figure 11: Circular graphic for data set II.

of the WGD model for this data set, by exhibiting equal values of Kupier, Watson and KS test statistics. On the other side, the Vuong test statistics given in Table 6 indicate that the WGD model is one of the suitable choices for such data set, while the proposed model has the least values of the information criteria statistics and thus depicts that the proposed model is the least loss of information model (see Table 7). The circular representations of the estimated cdfs of the compared distributions are given in Figure 12 which shows a

Table 6: Vuong test statistics of data set II.

WPL- WGD (p -value)	WPL- PL(p -value)	WPL- PD(p -value)
-3.9567 (0.3425)	7.6960(0.0000)	3.9415(0.0000)

Table 7: Log likelihood (l) and information criteria for data set II.

Distribution	$-l$	AIC	AICC	BIC	HQIC	CAIC
WPL	94.4784	190.957	191.042	192.849	191.675	191.042
WGD	94.425	190.85	190.935	192.742	191.568	190.935
PL	102.226	206.452	206.537	208.344	207.17	206.537
PD	103.465	208.929	209.014	210.821	209.647	209.014

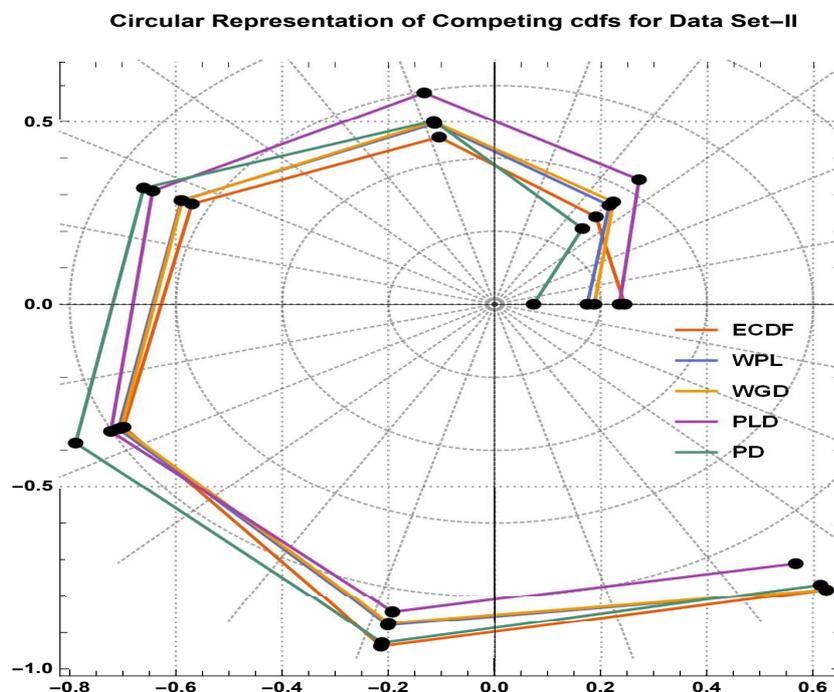


Figure 12: Circular representation of competing cdfs for data set II.

confirmation of the numerical results given above.

Example 4.3. data set III. The third data measure the arrival directions of low showers of cosmic rays and the coordinate system is declination and right Ascension. This data set was reported by [2]. The data consist of the following observations in degree: 315, 198, 99, 50, 86, 221, 0, 14, 63, 176, 185, 207, 221, 230, 243, 347, 342, 311, 293, 284, 216, 207, 144,

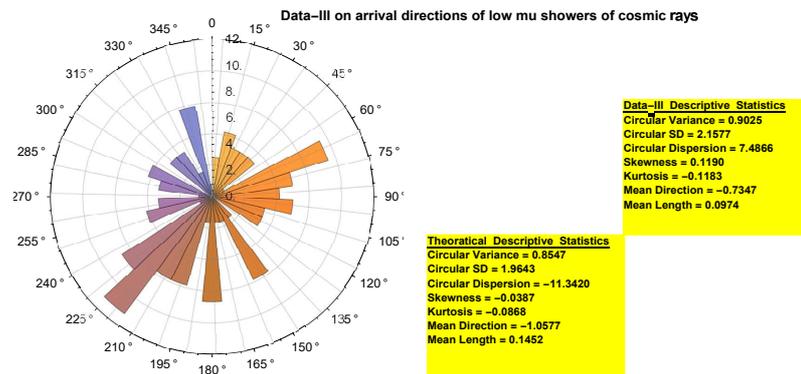


Figure 13: Circular graphic for data set III.

Table 8: MLEs and goodness-of-fit statistics of data set III.

Distribution	$\hat{\theta}$	U^{2*}	V^{2*}	KS	p-value
WPL	0.3409	0.0094	0.7211	0.1065	0.9998
WGD	0.1058	0.0074	0.6920	0.1104	0.9998
PL	0.6244	0.0583	1.1692	0.1726	0.9852
PD	2.5570	0.0599	1.1047	0.1738	0.9839

149, 77, 41, 27, 14, 5, 41, 36, 50, 63, 77, 77, 113, 153, 176, 203, 216, 252, 279, 288, 311, 320, 342, 356, 347, 347, 342, 338, 329, 293, 284, 261, 252, 257, 261, 261, 252, 230, 230, 216, 212, 207, 194, 158, 117, 99, 95, 14, 18, 27, 23, 50, 63, 77, 167, 176, 185, 194, 221, 216, 230, 234, 248, 279, 297, 324, 234, 185, 162, 144, 144, 108, 104, 86, 77, 68, 63, 27, 95, 99, 122, 140, 144, 171, 176, 185, 198, 216, 216, 234, 324, 311, 293, 275, 266, 207, 203, 212, 198, 140, 117, 117, 86, 68, 32, 14, 59, 68, 68, 72, 86, 104, 212, 153, 216, 342, 338, 9, 230, 221, 203, 198, 99, 86, 68, 45.

Figure 13 presents the circular graphics of the data.

Analysis of data set III: Table 8 shows that the proposed model is the strong competitor of the WGD model for this data set. Moreover, the Vuong test statistics in Table 9 indicate that the proposed model is the only suitable choice for such data set, with strong suitability to the WGD model. The proposed model is much better than the PL and PD models, and this result is confirmed by Table 10. The circular representations of the estimated cdfs of the compared distributions are given in Figure 14 that shows agreement with the preceding numerical results.

Table 9: Vuong test statistics for data set III.

WPL- WGD (p -value)	WPL- PL (p -value)	WPL- PD (p -value)
46.1951(0.0000)	15.7814 (0.0000)	0.9702(indecisive)

Table 10: Log likelihood (l) and information criterion for data set III.

Distribution	$-l$	AIC	AICC	BIC	HQIC	CAIC
WPL	284.437	572.874	572.956	578.882	572.094	572.956
WGD	286.2673	576.535	576.617	582.542	575.755	576.617
PL	305.6431	613.286	613.313	616.29	614.507	613.313
PD	288.1666	578.333	578.36	581.337	579.554	578.36

Circular Representation of Competing cdfs for Data Set-III

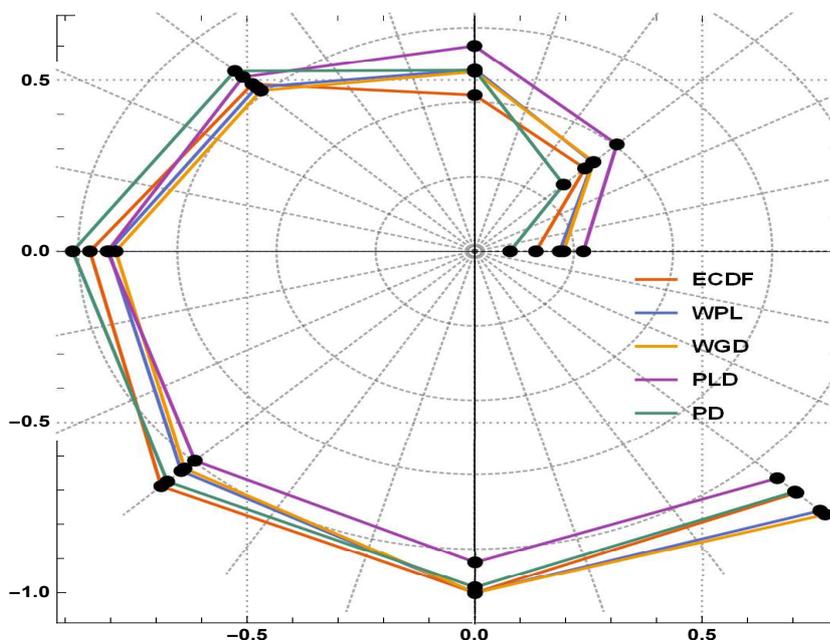


Figure 14: Circular representation of competing cdfs for data set III.

5 Conclusion

In this paper, we have discussed a new circular distribution resulting from wrapping the Poisson-Lindley distribution on the nonnegative integers around the circle. It is called the wrapped Poisson-Lindley (WPL) distribution. The probability mass, cumulative dis-

tribution, survival and probability generating functions of the WPL distribution admit explicit forms. Applications of the WPL model have been performed to three different circular data sets, with favorable results for the proposed WPL model.

Acknowledgments

The author would like to thank the two reviewers for thorough comments on the paper.

References

- [1] Adnan M A S and Roy S. Wrapped variance gamma distribution with an application to wind direction. *J. Environ. Stat.*, 2014, 6: 1-10.
- [2] Fisher N I, Lewis T, Embleton, B J J. *Statistical analysis of spherical data*. New York, Cambridge University Press, 1987.
- [3] Ghitany M E, Al-Mutairi D K. Estimation methods for the discrete Poisson-Lindley distribution. *J. Stat. Comput. Simul.*, 2009, 79: 1-9.
- [4] Girija S V S, Roy A V D, Srihari G V L N. On wrapped binomial model characteristics. *Math. Sta.*, 2014, 2: 231-234.
- [5] Jacob S, Jayakumar K. Wrapped geometric distribution: A new probability model for circular data. *J. Stat. Theory. Appl.*, 2013, 12: 348-355.
- [6] Jammalamadaka S R, Kozubowski T J. New families of wrapped distributions for modeling skew – circular data. *Commun. Stat. Theory Methods*, 2004, 33: 2059-2074.
- [7] Jammalamadaka S R, Sengupta A. *Topics in Circular Statistics*. World Scientific Publishing Co. Pte. Ltd., Singapore, 2001.
- [8] Jayakumar K, Jacob S. Wrapped skew Laplace distribution on integers: A new probability model for circular data. *Open J. Stat.*, 2012, 2: 106-114.
- [9] Jose K K, Varghese J. Wrapped log Kumaraswamy distribution and its applications. *Int. J. Comput. Math.*, 2018, 6: 1924-1930.
- [10] Joshi S, Jose K K, Bhati D. Estimation of a change point in the hazard rate of Lindley model under right censoring. *Commun. Stat. Simul. Comput.*, 2017, 46: 3563-3574.
- [11] Lévy P L. Addition des variable saléatoire esdéfinies sur une circonférence. *Bull. Soc. Math. Fr.*, 1939, 67: 1-41.
- [12] Mardia K V, Jupp P E. *Directional statistics 2nd Ed*. New-York, Wiley, 2000.
- [13] Rao A V D, Girija S V S, Devaraaj V J. On characteristics of wrapped gamma distribution. *Eng. Sci. Technol. an Int. J.*, 2013, 3(2).
- [14] Rao A V D, Sarma I R, Girija S V S. On wrapped version of some life testing models. *Commun. Stat. Theory Methods*, 2007, 36: 2027-2035.
- [15] Sankaran M. The discrete Poisson Lindley distribution. *Biometrics*, 1970, 26: 145-149.
- [16] Shanker R., Feshaye H. On Poisson-Lindley distribution and its applications to biological sciences. *Biom. Biostat. Int. J.*, 2015, 2: 103-107.
- [17] Srihari G V L N, Girija S V S, Rao A V D. On discrete wrapped exponential distribution-characteristics. *Int. J. Sci. Res. Math. Stat. Sci.*, 2018, 5: 57-64.
- [18] Srihari G V L N, Girija S V S, Rao A V D. On wrapped negative binomial model. *Int. J. Math. Archive*, 2018, 9: 148-153.
- [19] Vuong Q H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 1989, 57: 307-333.