Generation and Statistical Properties for Lindley-Polynomial Distribution

Dariush Ghorbanzadeh

Volume 16 | Issue 1 Article 44

6-2021

Generation and Statistical Properties for Lindley-Polynomial Distribution

Dariush Ghorbanzadeh
CNAM

Available at
http://pvamu.edu/aam
Appl. Appl. Math.
ISSN: 1932-9466

Applications and Applied Mathematics:
An International Journal
(AAM)

Vol. 16, Issue 1 (June 2021), pp. 776 – 786

Generation and Statistical Properties for Lindley-Polynomial Distribution

Dariush Ghorbanzadeh

Département Mathématiques-Statistiques
CNAM
292, Rue Saint-Martin, 75141-Paris CEDEX 03, France
dariush.ghorbanzadeh@lecnam.net

Received: October 11, 2018; Accepted: September 26, 2020

Abstract

For the modeling of the wind speed, we propose a family of distributions in polynomial form generating the Lindley distribution. We call this distribution Lindley-Polynomial distribution. The estimation of parameters using the maximum product spacing estimation method. A real data set has been considered to illustrate the practical utility of the paper.

Keywords: Lindley distribution; Lindley-Polynomial distribution; Maximum Product Spacing estimation method; Quantiles

MSC 2010 No.: 65C99, 65C10, 62N02

1. Introduction

In 1958, Lindley (1958) suggested a one parameter distribution to illustrate the difference between fiducial distribution and posterior distribution. The Lindley distribution has been used for modeling lifetime data and studying some stress-strength problems and has been the subject of studies by several authors. For example, we mention the works of Sankaran (1970) introduced the Discrete Poisson-Lindley Distribution and Ghitany et al. (2008) presents a treatment of the mathematical properties for the Lindley distribution. Ghitany et al. (2012) investigated the Marshall-Olkin extended Lindley distribution and Rodrigues et al. (2015) study the beta exponentiated Lindley distribution. In this work, we propose a family of distributions in polynomial form generating the Lindley distribution. We call this distribution Lindley-Polynomial distribution. The estimation of parameters using the maximum product spacing estimation method. The maximum product of spacings (MPS) method for estimating parameters in continuous univariate distributions was proposed by Cheng and Amin (1983) and independently by Ranneby (1984). The MPS method produces consistent and asymptotically efficient estimators. These estimators are consistent see Ranneby (1984). In their work Heathcote et al. (2002) proposes a method that uses empirical quantiles. Knowledge of extreme wind speed distributions is essential for the design of wind turbines and structures. Statistical parameters to express wind distribution speeds are very useful, and considerable work has been carried on in recent years. Many authors have shown the fit of wind speed to the Weibull distribution step by step. However, these studies are performed through priori acceptance. Probability density function of wind speed is not always statistically accepted as Weibull pdf.

Celik (2003) proposes an empirical study and Whalen et al. (2004) proposes the method of self-determined probability-weighted moments. Carta and Ramírez (2007) present the analysis of two-component mixture Weibull statistics for estimation of wind speed distributions and Ghorbanzadeh et al. (2016) proposes a change-point model.

The paper examines the applicability of probability distributions commonly used to model wind speeds to data representing the average daily wind speed (km/h) at the Orly airport (Paris, France), from January 1, 2006 to December 31, 2015 (data is available at: https://www.wunderground.com/history/airport/).

2. Construction of the Lindley-Polynomial distribution

In this section we consider g a probability density function (pdf) on $(0,\infty)$ and m a positive integer. We assume that the moments of order k of g exists:

$$\mu_k = \int_0^{\infty} x^k \, g(x) \, dx < \infty, \forall k \in \{0, 1, \ldots, m\}.$$

In this paper we consider the densities family

$$f_m(x) = C_m \left(\sum_{k=0}^{m} x^k \right) g(x) \, \mathbb{1}_{(0,\infty)}(x), \tag{1}$$

where $\mathbb{1}$ is the indicator function and $C_m = \big(1 + \mu_1 + \ldots + \mu_m\big)^{-1}$.

For generating a random variable from density (1), we have the following algorithm.

Proposition 2.1.

Consider X and N two random variables such that:

(1) N is discrete values in the set $\{0, 1, \ldots, m\}$ with probability mass function (pmf) : $\mathbb{P}\big(N =$

$k) = p_k$ where p_k is defined by

$$p_k = \frac{\mu_k}{1 + \mu_1 + \ldots + \mu_m} = \mu_k C_m, \tag{2}$$

(2) given as N, X has the pdf

$$f_X(x|N = k) = \frac{1}{\mu_k} x^k g(x)\, \mathbb{1}_{(0,\infty)}(x). \tag{3}$$

Then, the random variable X has the pdf defined in (1).

Proof:

The unconditional density of the random variable X is given by

$$f_m(x) = \sum_{k=0}^{m} p_k\, f_X(x|N = k) = \left(\sum_{k=0}^{m} \frac{p_k}{\mu_k}\, x^k \right) g(x)\, \mathbb{1}_{(0,\infty)}(x).$$

Using (2) we get the result. ■

In the following, we will consider a special case for the pdf g ($g(x) = \theta\, e^{-\theta x}$), which corresponds to the exponential distribution with $\theta > 0$ parameter. In this case we have the following family of distributions (Lindley-Polynomial distribution):

$$f_m(x, \theta) = C_m(\theta) \left(\sum_{k=0}^{m} x^k \right) e^{-\theta x}\, \mathbb{1}_{(0,\infty)}(x), \tag{4}$$

where $C_m(\theta)$ is defined by

$$C_m(\theta) = \frac{\theta^{m+1}}{\sum\limits_{k=0}^{m} k!\, \theta^{m-k}}. \tag{5}$$

For generating a random variable from density (4), we have the following algorithm.

Proposition 2.2.

Consider X and N two random variables such that:

(1) N is discrete values in the set $\{0, 1, \ldots, m\}$ with probability mass function (pmf)

$$p_k = \frac{k!\, \theta^{m-k}}{\sum\limits_{k=0}^{m} k! \theta^{m-k}} = \frac{k!}{\theta^{k+1}}\, C_m(\theta), \tag{6}$$

(2) given as N, X has the gamma distribution with parameters $k + 1$ and θ ($\gamma(k + 1, \theta)$)

$$f_X(x|N = k) = \frac{\theta^{k+1}}{k!} x^k e^{-\theta x}\, \mathbb{1}_{(0,\infty)}(x). \tag{7}$$

Then, the random variable X has the pdf defined in (4).

Proof:

We have $\mu_k = \displaystyle\int_0^\infty x^k\, e^{-\theta x}\, dx = \frac{k!}{\theta^{k+1}}$, and using Proposition (2.1), we get the result. ■

For $m = 1$, we get the pdf $f_1(x, \theta) = \frac{\theta^2}{1+\theta}\,(1 + x)\, e^{-\theta x}\, \mathbb{1}_{(0,\infty)}(x)$, which is the Lindley distribution.

The following figures shows the pdf of Lindley-Polynomial distribution defined in (4) as a function of variation of θ and m.

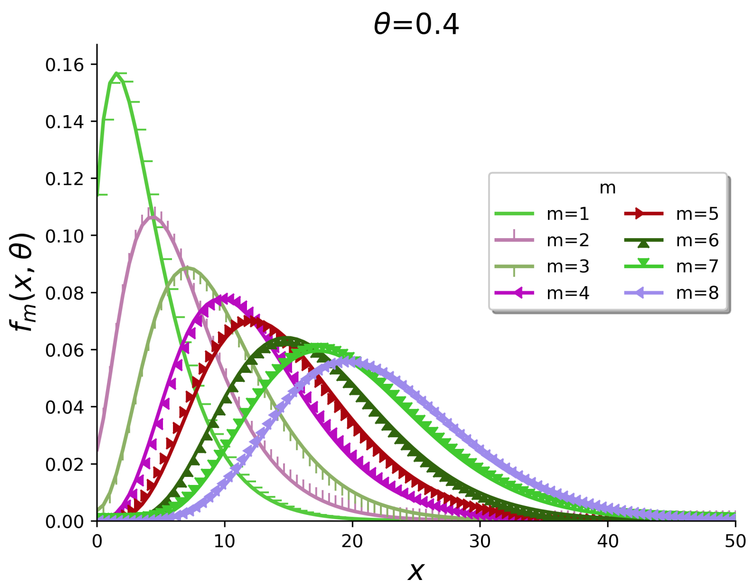


Figure 1. The pdf of Lindley-Polynomial distribution defined in (4) with $\theta = 0.4$

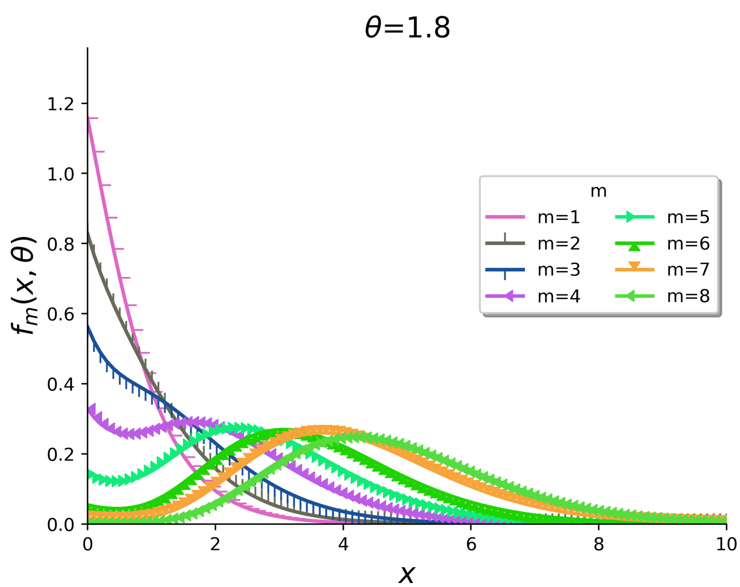


Figure 2. The pdf of Lindley-Polynomial distribution defined in (4) with $\theta = 1.8$

For the statistical properties of the Lindley-Polynomial distribution, we have the following proposition.

Proposition 2.3.

(1) The cumulative distribution function (cdf) is defined by

$$F_m(x,\theta) = 1 - C_m(\theta) \sum_{k=0}^{m} \sum_{j=0}^{k} \frac{k!}{j!} \theta^{j-k-1} x^j e^{-\theta x}. \tag{8}$$

(2) The characteristic function is defined by

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = C_m(\theta) \sum_{k=0}^{m} \frac{k!}{(\theta - it)^{k+1}}. \tag{9}$$

(3) The r^{th} moment is defined by

$$\mathbb{E}[X^r] = \frac{1}{\theta^r} \frac{\sum_{k=0}^{m} (k+r)!\, \theta^{m-k}}{\sum_{k=0}^{m} k!\, \theta^{m-k}}. \tag{10}$$

(4) The hazard rate function (failure rate function) is defined by

$$h(x) = \frac{f_m(x,\theta)}{1 - F_m(x,\theta)} = \frac{\sum_{k=0}^{m} x^k}{\sum_{k=0}^{m} \sum_{j=0}^{k} \frac{k!}{j!} \theta^{j-k-1} x^j}. \tag{11}$$

3. Estimation of Parameters

In this section we use the MPS estimates techniques to estimate the parameters of the distribution. The method is based on maximization of the geometric mean of probability spacings in the data where the spacings are defined as the differences between the values of the cumulative distribution function at sequential data indices. The MPS method, originally suggested by Cheng and Amin (1983), this method was also independently developed by Ranneby (1984) as approximation to the Kullback-Leibler measure of information. In their work Heathcote et al. (2002) proposes a method that uses empirical quantiles.

Let $X_1, \ldots, X_n$ be a random sample from the Lindley-Polynomial distribution (4). We suppose that a random sample of size n has been allocated in k classes: $C_1, \ldots, C_k$ with observed frequency $n_1, \ldots, n_k$. The probability of class C_i is

$$p_i(m,\theta) = \int_{C_i} f_m(x,\theta)\, dx. \tag{12}$$

The likelihood function of a multinomial distribution defined on the classes $C_1, \dots, C_k$, that is,

$$L_m(\theta) = \prod_{i=1}^{k} \Big(p_i(m,\theta) \Big)^{n_i}, \tag{13}$$

and the log-likelihood function is given as

$$\ell_m(\theta) = \log\big(L_m(\theta)\big) = \sum_{i=1}^{k} n_i \log\big(p_i(m,\theta)\big). \tag{14}$$

The estimator of θ is defined by

$$\hat{\theta}(m) = \underset{\theta>0}{\operatorname{argsup}}\ \ell_m(\theta), \tag{15}$$

and the estimator of m is defined by

$$\hat{m} = \underset{m\geq 1}{\operatorname{argmax}}\ \ell_m\big(\hat{\theta}(m)\big). \tag{16}$$

In this work, we deliberately consider the classes

$$C_1 = (0, q_1), C_2 = (q_1, q_2), \dots, C_k = (q_{k-1}, \infty), \tag{17}$$

with the corresponding probabilities

$$\begin{cases} p_1(m,\theta) = F_m(q_1,\theta), \\ p_i(m,\theta) = F_m(q_i,\theta) - F_m(q_{i-1},\theta), \quad i = 2, \dots, k-1, \\ p_k(m,\theta) = 1 - F_m(q_{k-1},\theta), \end{cases} \tag{18}$$

where q_i denotes the i^{th} quantile: $F_m(q_i,\theta) = i/100$.

4. Application

We apply the model used in this work to the data representing the average daily wind speed (km/h) at the Orly airport (Paris, France), from January 1, 2006 to December 31, 2015.

We consider two types of analysis. The first analysis is to do a monthly study of the data, studying the twelve months, the ten years, separately (the January data gather the data for the ten months January 2006, ..., January 2015 and the same for the other months) and the two analysis consists in doing a seasonal study considering the four seasons.

In this section, we consider the following nine classes:

$$\begin{aligned} &C_1 = (0, q_{10}), \quad C_2 = (q_{10}, q_{20}), \quad C_3 = (q_{20}, q_{35}), \quad C_4 = (q_{35}, q_{50}), \quad C_5 = (q_{50}, q_{65}), \\ &C_6 = (q_{65}, q_{75}), \quad C_7 = (q_{75}, q_{85}), \quad C_8 = (q_{85}, q_{90}), \quad C_9 = (q_{90}, \infty). \end{aligned} \tag{19}$$

4.1. Results for monthly analysis

The following table represents the results of estimation of the parameters obtained by the monthly analysis of the data.

Table 1. Monthly data analysis results

January		February		March	
$\hat{m} = 4$	$\hat{\theta} = 0.3748$	$\hat{m} = 4$	$\hat{\theta} = 0.3606$	$\hat{m} = 4$	$\hat{\theta} = 0.3699$
April		May		June	
$\hat{m} = 4$	$\hat{\theta} = 0.4270$	$\hat{m} = 6$	$\hat{\theta} = 0.5957$	$\hat{m} = 7$	$\hat{\theta} = 0.6645$
July		August		September	
$\hat{m} = 8$	$\hat{\theta} = 0.7618$	$\hat{m} = 8$	$\hat{\theta} = 0.8350$	$\hat{m} = 3$	$\hat{\theta} = 0.3595$
October		November		December	
$\hat{m} = 4$	$\hat{\theta} = 0.4807$	$\hat{m} = 4$	$\hat{\theta} = 0.4137$	$\hat{m} = 3$	$\hat{\theta} = 0.3331$

The results obtained by Table 1 show that, whatever the month, the estimator of θ is less than 1. On the other hand, the value of the estimators for the months May, June, July and August are well above 0.5 and for the other months the value being less than 0.5. For the estimator $\hat{\theta}$, we have the following statistics: mean of $\hat{\theta} = 0.4980$ and standard deviation of $\hat{\theta} = 0.1658$. For the estimator $\hat{m}$, we have the following statistics: mean of $\hat{m} = 4.9167$ and standard deviation of $\hat{m} = 1.7539$. We deduce that mean of $\hat{m} \in \{4, 5\}$ and standard deviation of $\hat{m} \in \{1, 2\}$. In order to visualize the graphs of the results obtained, we present the results for the months of January.

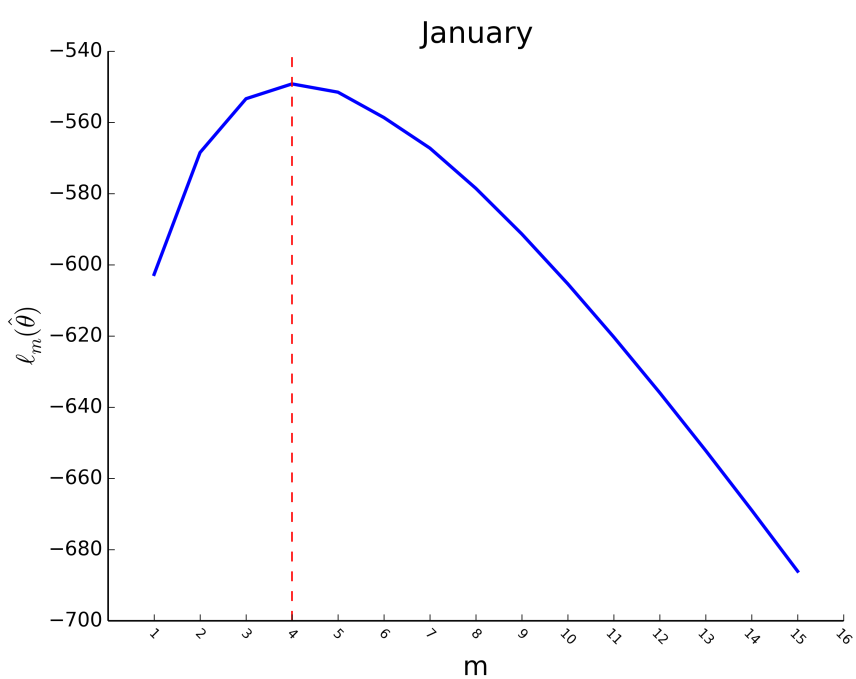


Figure 3. Graph of the log likelihood function, for the months of January, defined in (15) for different values of m

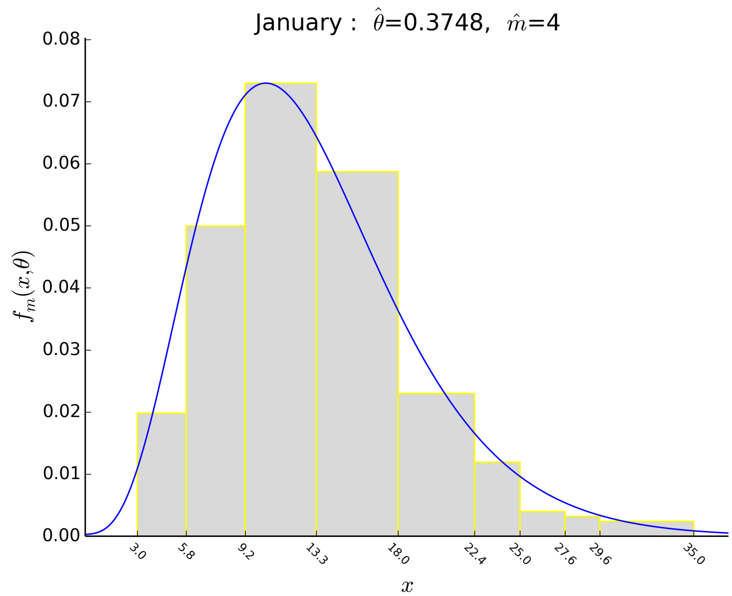


Figure 4. Graph of the pdf defined in (4) as a function of the estimated parameters and histogram of the January data taking into account the classes defined in (19)

4.2. Results for seasonal analysis

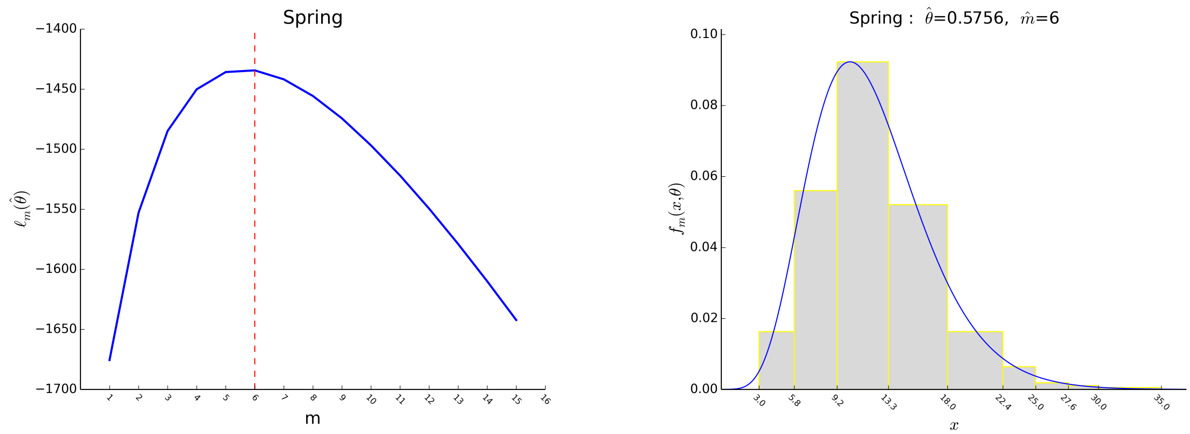
The following table represents the results of estimation of the parameters obtained by the seasonal analysis of the data.

Table 2. Seasonal data analysis results

Spring		Summer	
$\hat{m} = 6$	$\hat{\theta} = 0.5756$	$\hat{m} = 5$	$\hat{\theta} = 0.5330$
Autumn		Winter	
$\hat{m} = 3$	$\hat{\theta} = 0.3379$	$\hat{m} = 3$	$\hat{\theta} = 0.3186$

The results obtained show that, whatever the season, the estimator of θ is less than 1. In particular, for the autumn and winter seasons, the estimator of m is equal to 3.

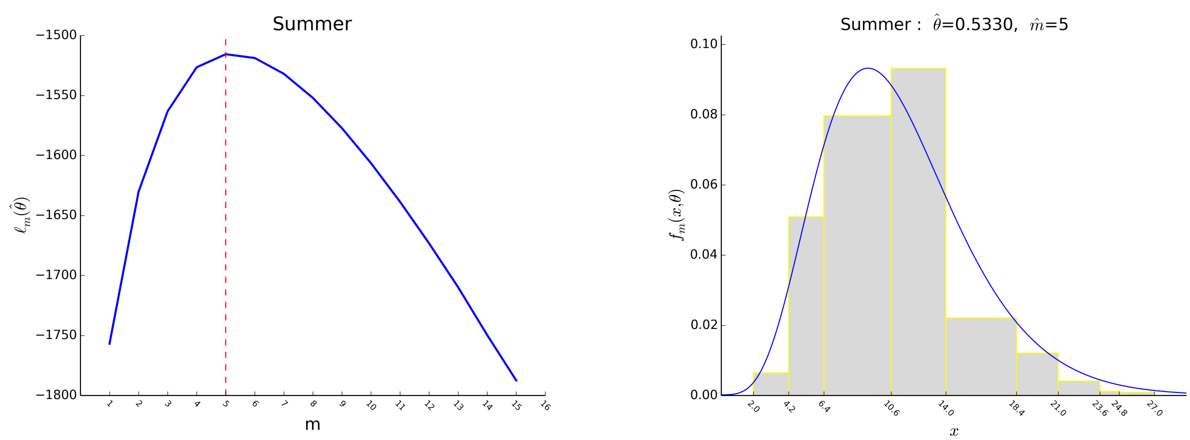
The following figures represent the results obtained for the four seasons.



(a) The log Likelihood function for spring data.

(b) The pdf and histogram for spring data.

Figure 5. Seasonal analysis results graph for spring data



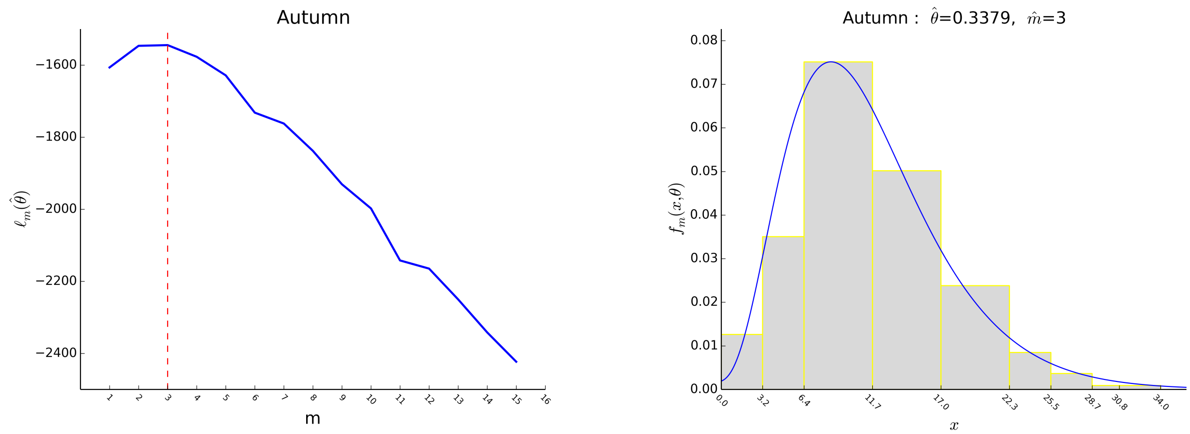
(a) The log Likelihood function for summer data.

(b) The pdf and histogram for summer data.

Figure 6. Seasonal analysis results graph for summer data

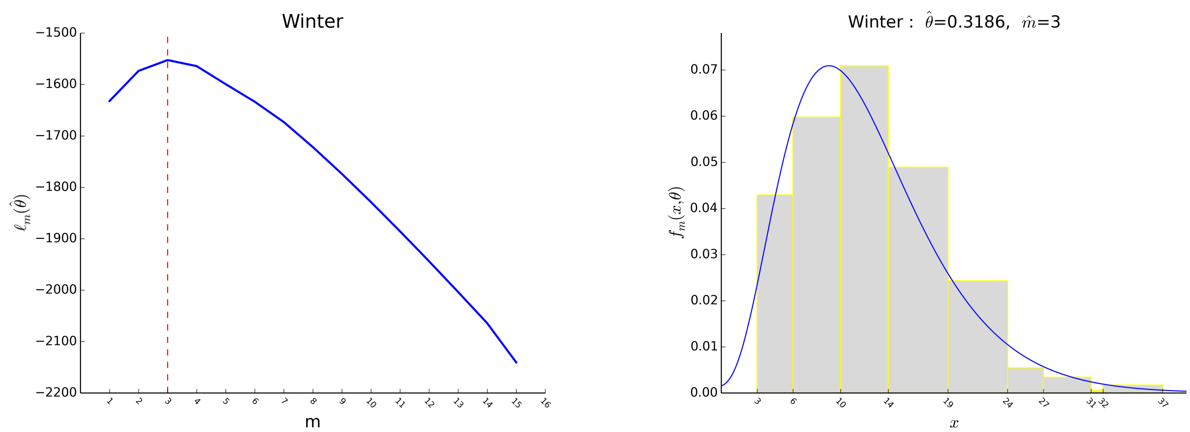
5. Conclusion

The wind speed distributions and energy intensity measured at the location reveals that the current technology does not provide economical electricity production from wind power and that the measurements should be evaluated in the long term in accordance with technological developments and reduction in the cost of turbines. The model studied in this work is very simple to use in practice. The proposed algorithms are very simple to implement for the simulation of the Lindley-Polynomial distribution. The results obtained show, by monthly and seasonal analysis, that overall the data used are well adjusted to the Lindley-Polynomial distribution with a parameter θ less than 1. The result of this study encourages the use of wind potential at the approximate location.



(a) The log Likelihood function for autumn data. (b) The pdf and histogram for autumn data.

Figure 7. Seasonal analysis results graph for autumn data



(a) The log Likelihood function for winter data. (b) The pdf and histogram for winter data.

Figure 8. Seasonal analysis results graph for winter data

REFERENCES

Carta, J.A. and Ramírez, P. (2007). Analysis of two-component mixture Weibull statistics for estimation of wind speed distributions, Renewable Energy, Vol. 32, pp. 518-531.

Celik, AN. (2003). Assessing the suitability of wind speed probability distribution functions based on wind power density, Renew Energy, Vol. 28, pp. 1563-74.

Cheng, R. C. H. and Amin, N. A. K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin, J. Roy. Statist. Soc. Ser. B., Vol. 45, pp. 394-403.

Ghitany, M.E., Al-Mutairi, D.K, Al-Awadhi, F.A. and Al-Burais, M.M. (2012). Marshall-Olkin extended Lindley distribution and its application, International Journal of Applied Mathematics, Vol. 25, pp. 709-721.

Ghitany, M.E., Atieh, B. and Nadarajah, S. (2008). Lindley distribution and its applications, Mathematics and Computers in Simulation, Vol. 78, No. 4, pp. 493-506.

Ghorbanzadeh, D., Durand, P. and Jaupi, L. (2016). An analytical method for detecting the change-point in simple linear regression model, Application at Weibull distribution, Journal of Applied Quantitative Methods, Vol. 11, Issue 1, pp. 1-10.

Heathcote, A., Brown, S. and Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions, Psychonomic Bulletin & Review, Vol. 9, pp. 394-401.

Lindley, D.V. (1958). Fiducial distributions and Bayes' Theorem, Journal of the Royal Statistical Society, Vol. 20, No. 1, pp. 102-107.

Ranneby, B. (1984). The maximum spacing method. An estimation method related to the maximum likelihood method, Scand, J, Statist, Vol. 11, pp. 93-112.

Rodrigues, J.A., Silva, A.P.C.M. and Hamedani, G.G. (2015). The beta exponentiated Lindley distribution, Journal of Statistical Theory and Applications, Vol. 14, No. 1, pp. 60-75.

Sankaran, M. (1970). The discrete Poisson-Lindley distribution, Biometrics, Vol. 26, No. 1, pp. 145-149.

Whalen, T.M., Savage, G.T. and Jeong, G.D. (2004). An evaluation of the self-determined probability-weighted moment method for estimating extreme wind speeds, Journal of Wind Engineering and Industrial Aerodynamics, Vol. 92, pp. 219-239.