PBRU SCIENCE JOURNAL
2023.20(1) : 85-96

A Simple Discrete-Weighted Exponential Distribution and Applications

Chadarat Tapan

Department of Mathematics, Faculty of Science and Technology, Pibulsongkram Rajabhat University

Corresponding author; E-mail: chadarat@psru.ac.th

Received: 02 September 2022 /Revised: 17 October 2022 /Accepted: 11 November 2022

Abstract

A discrete distribution has been widely used for count data. This study aimed to present a simple discrete-weighted exponential distribution and some statistical properties, such as the cumulative distribution function, probability generating function, expected, variance, moment generating function, survival function, index of dispersion and coefficient of variation. The method of estimation of the parameters was maximum likelihood estimation. Furthermore, the simple discrete-weighted exponential distribution was applied and compared to other distributions on over-dispersion data sets. The results indicated that the simple discrete-weighted exponential distribution could be used as an alternative to over-dispersion data.

Keywords: Discrete distribution, Probability mass function, Survival function, Exponential distribution

Introduction

Lifetime data is the time of occurrence based on a topic of interest for endurance, such as the period from epidemic beginning to illness recovery, mobile lifetime, the light bulb time period for use, and time of death in cancer patients. Lifetimes are sometimes referred to as "survival times" or "failure periods" depending on the event of interest. Popular distributions for lifetime data, such as Weibull, exponential, Pareto, and Lindley distributions, are commonly used to generate modeling data for lifetime data. In many cases, data contains the frequency of an event of interest at the time of occurrence or count data, such as the number of coronavirus infections in a month, accidents, and consumers. So, count data is a data type that describes the frequency of events or items that occur within a fixed period of time[1]. For modeling count data, the Poisson distribution is the most widely utilized. The conditional mean of occurrence is equal to the variance in the Poisson distribution assumptions, which is known as equi-dispersion. In most situations, however, the variance is larger than its mean, which is referred to as over-dispersion [2]. The discrete distributions are utilized for over-dispersion count data.

A variety of studies have been conducted to deal with the discrete distribution of count data. In 2015, Nekoukhou and Bidram[3] introduced the exponentiated discrete Weibull distribution and presented some basic distributional properties, moments, and order statistics. In addition, the exponentiated discrete Weibull distribution was applied to a real data set. In 2017, Borah[4] used the general approach of compounding for quasi-Poisson Lindley distribution. The result showed the properties of the quasi-Poisson Lindley distribution such as moment, skewness, and Kurtosis. In 2018, Mahdi et al.[5] suggested and evaluated parameters for a discrete weighted exponential distribution. The discrete exponential distribution, developed by Chakraborty[6], was a generalization of this distribution. The discrete weighted exponential distribution exceeded the exponentiated geometric and generalized geometric distributions. Khongthip et al.[7] proposed a discrete weighted exponential distribution. A discrete distribution was generated from a corresponding continuous distribution by applying Roy's method for discretization. In 2020, Diandarma et al.[8] presented a discrete Lindley distribution which was generated from the discretization method. The result found that a discrete Lindley distribution could be used as an alternative to the Poisson distribution in modeling over-dispersion data.

The aims of this paper were to present a simple discrete-weighted exponential distribution and to describe some properties of the simple discrete-weighted exponential distribution. For the rest of this paper, the second section presented methods and materials. In the third section, results such as the simple discrete-weighted exponential distribution, some statistical properties, parameter estimation, and the application of two data sets were shown. Finally, the last section provided a conclusion.

Materials and Methods

This section presents the discretization method for generating new discrete distributions and the weighted exponential distribution.

1. Discretization method

A discrete probability distribution function can be generated from a continuous probability distribution function using the discrete technique. Methods for producing discrete probability distribution functions from continuous probability distribution functions were demonstrated by Chakraborty[6]. In this paper, the method based on the continuous survival function was used.[6] The probability mass function (pmf) of discrete distributions is

$$f(y) = S_X(y) - S_X(y+1), \qquad (1)$$

where $y = 0,1,2,\ldots$ and $S_X(y)$ is the survival function.

2. Weighted exponential distribution

In 2016, Bashir and Naqvi[9] derived a weighted exponential distribution. The probability distribution function (pdf) of the weighted exponential distribution was given by

$$f(x;\lambda,\omega) = (\lambda-\omega)e^{-(\lambda-\omega)x}, \qquad (2)$$

where $x > 0, \lambda > 0, \lambda > \omega,$ and $0 < \omega < 1$.

The survival function (sf) of a weighted exponential distribution was given by

$$S(x) = e^{-(\lambda-\omega)x}. \qquad (3)$$

Results

In this section, the pmf, cumulative distribution function, and some statistical properties of a simple discrete weighted exponential distribution and parameter estimation are presented.

1. A simple discrete-weighted exponential distribution

Theorem 1: If Y be random variable of a simple discrete-weighted exponential distribution with parameters λ and ω are scale and shape parameters respectively. The pmf of the simple discrete-weighted exponential distribution is

$$f(y) = e^{-y(\lambda-\omega)}\left[1-e^{-(\lambda-\omega)}\right], \qquad (4)$$

where $y = 0,1,2,\ldots$, $\lambda > 0$, $\lambda > \omega$ and $0 < \omega < 1$.

Proof: Equation (2) shows the pdf of a weighted exponential distribution when Y is a random variable. The survival function is

$$S(y) = e^{-(\lambda-\omega)y} \qquad (5)$$

and $$S(y+1) = e^{-(\lambda-\omega)(y+1)} \qquad (6)$$

From the discretization method, the Equations (5) and (6) are plugged into the Equation (1). We will obtain

$$\begin{aligned} f(y) &= S(y) - S(y+1) \\ &= e^{-(\lambda-\omega)y} - e^{-(\lambda-\omega)(y+1)} \\ &= e^{-y(\lambda-\omega)}\left[1-e^{-(\lambda-\omega)}\right]. \end{aligned}$$

The properties of probability function are $f(y) \ge 0$ for all y and $\sum_{\forall y} f(y) = 1$.

$$\begin{aligned} \sum_{\forall y} f(y) &= \sum_{y=0}^{\infty} e^{-y(\lambda-\omega)}\left[1-e^{-(\lambda-\omega)}\right] \\ &= \left[1-e^{-(\lambda-\omega)}\right]\sum_{y=0}^{\infty} e^{-y(\lambda-\omega)} \\ &= \left[1-e^{-(\lambda-\omega)}\right]\left[\frac{1}{1-e^{-(\lambda-\omega)}}\right] \\ &= 1 \end{aligned}$$

Thus, the pmf has the properties of a probability function.

The cdf of the simple discrete-weighted exponential distribution is

$$F(y) = 1 - e^{-(\lambda-\omega)(y+1)}, \qquad (7)$$

where $y = 0,1,2,\ldots$, $\lambda > 0$, $\lambda > \omega$ and $0 < \omega < 1$.

The cdf of the simple discrete-weighted exponential distribution is the following:

$$
\begin{aligned}
F_Y(t) &= \sum_{t=0}^{y} f(t) \\
&= \sum_{t=0}^{y} e^{-t(\lambda-\omega)}\left[1-e^{-(\lambda-\omega)}\right] \\
&= \left[1-e^{-(\lambda-\omega)}\right]\left[\frac{1-e^{-(\lambda-\omega)(y+1)}}{1-e^{-(\lambda-\omega)}}\right] \\
&= 1-e^{-(\lambda-\omega)(y+1)}
\end{aligned}
$$

Figure 1 and Figure 2 provide some possible plots for the pmf and cdf of the simple discrete-weighted exponential distribution. Figure 1 presents the pmf plots of the simple discrete-weighted exponential distribution for different values of λ and ω, and the distribution is skewed to the right. Obviously, the scale parameter λ varies with the shape parameter ω. The pmf decreases as the shape parameter ω increases. Figure 2 presents the cdf of the simple discrete-weighted exponential distribution for different values of λ and ω, and it is increasing function.

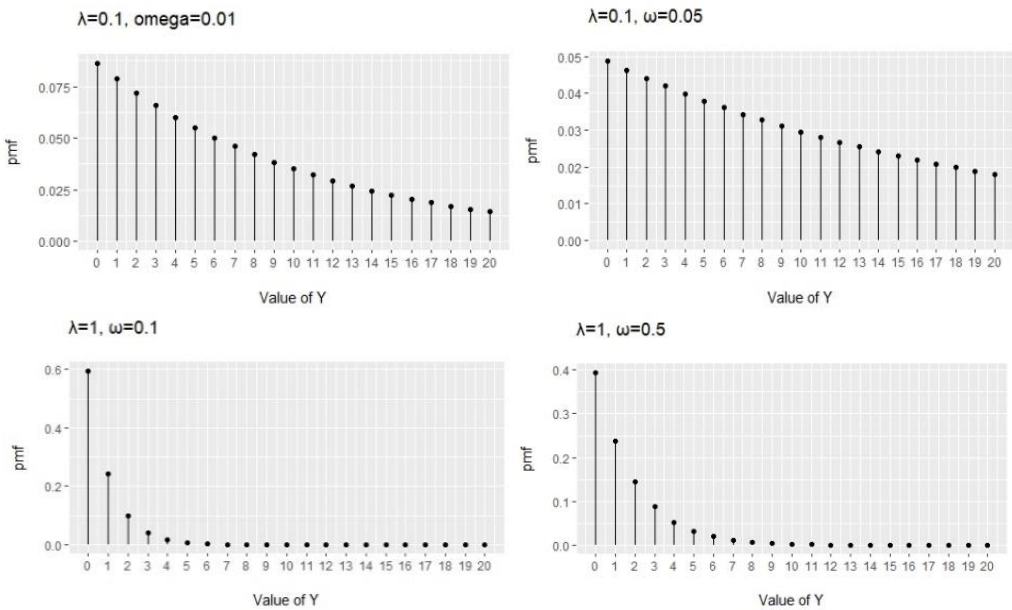


Figure1. The pmf plots of the simple discrete weighted-exponential distribution

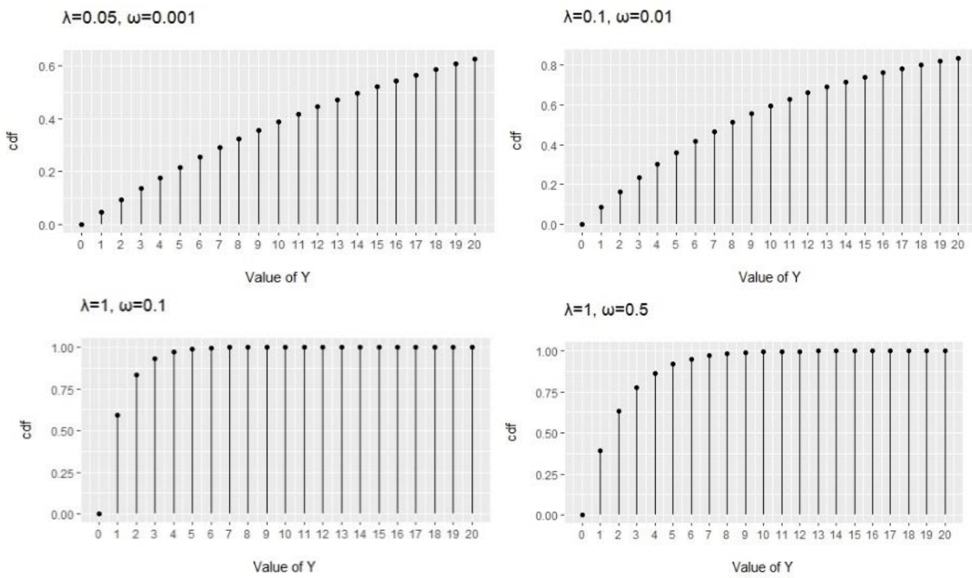


Figure2. The cdf plots of the simple discrete weighted-exponential distribution

2. Statistics properties

This section showed the probability generating function (pgf), expectation, variance, and moment generating function (mgf), survival function (sf), index of dispersion (ID), and coefficient of variation (CV).

2.1 Probability generating function

In statistical theory, the pgf is more important. This function can also find the expectation and variance of a random variable.

Theorem 2: Let Z be a random variable with a simple discrete-weighted exponential distribution, then the pgf is given by

$$G(z)=\frac{1-e^{-(\lambda-\omega)}}{1-ze^{-(\lambda-\omega)}}, \qquad (8)$$

where $y=0,1,2,\ldots$, $\lambda>0, \lambda>\omega, 0<\omega<1$ and z is a real number.

Proof: The pgf of the simple discrete-weighted exponential distribution can be shown as follows:

$$\begin{aligned} G(z) &= \sum_{y=0}^{\infty} f(y) z^y \\ &= \sum_{y=0}^{\infty} e^{-y(\lambda-\omega)}\left(1-e^{-(\lambda-\omega)}\right) z^y \\ &= \frac{1-e^{-(\lambda-\omega)}}{1-z\cdot e^{-(\lambda-\omega)}} \end{aligned}$$

2.2 Expectation, Variance and Moment Generating Function

From Theorem 2, the expectation of Y is

$$E(y)=G'(1)=\frac{e^{-(\lambda-\omega)}}{1-e^{-(\lambda-\omega)}}. \qquad (9)$$

The variance of Y can be derived from Theorem 2. We obtain

$$Var(Y) = G''(1) + G'(1) - (G'(1))^2$$
$$= \frac{2e^{-2(\lambda-\omega)}}{\left(1-e^{-(\lambda-\omega)}\right)^2} + \frac{e^{-(\lambda-\omega)}}{1-e^{-(\lambda-\omega)}} - \frac{e^{-2(\lambda-\omega)}}{\left(1-e^{-(\lambda-\omega)}\right)^2}$$
$$= \frac{e^{-(\lambda-\omega)}}{\left(1-e^{-(\lambda-\omega)}\right)^2} \qquad (10)$$

So, the variance of Y is $\frac{e^{-(\lambda-\omega)}}{\left(1-e^{-(\lambda-\omega)}\right)^2}$.

Set $z = e^t$ and we will obtain the mgf.

$$G(e^t) = E(e^{ty}) = M_Y t$$

Thus,
$$M_Y(t) = \frac{1-e^{-(\lambda-\omega)}}{1-e^{t-(\lambda-\omega)}} \qquad (11)$$

2.3 Survival function

The sf of the simple discrete-weighted exponential distribution is the following: $S(y) = 1 - F(y)$, then the sf [10] is

$$S(y) = e^{-(\lambda-\omega)(y+1)} \qquad (12)$$

Figure 3 shows some possible plots for the sf of the simple discrete weighted-exponential distribution. It can be seen that the sf function could be decreasing with respect to the value of Y based on the different parameter values.

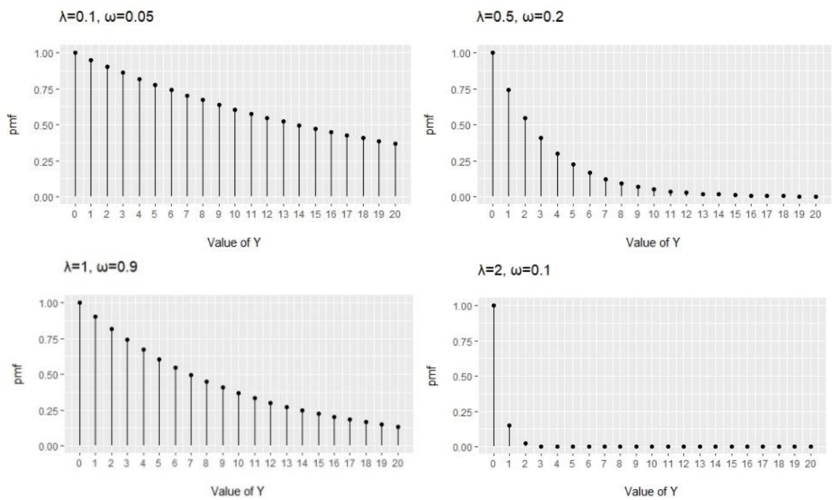


Figure 3. The sf of the simple discrete weighted exponential distribution

2.4 Index of dispersion and coefficient of variation

The ID is a measure of dispersion (over-dispersion, under-dispersion and equidispersion)[11]. The variance to mean ratio is the most basic definition. The ID of the simple discrete-weighted exponential distribution given as

$$ID = \frac{1}{1-e^{-(\lambda-\omega)}}. \tag{13}$$

$ID > 1$, for $\lambda > 0, \lambda > \omega, 0 < \omega < 1$.

Therefore, it follows that the simple discrete-weighted exponential distribution is over-dispersion.

The CV is expressed as the ratio of standard deviation to the mean. The CV of the simple discrete-weighted exponential distribution is given as

$$CV = \frac{1}{\sqrt{e^{-(\lambda-\omega)}}}. \tag{14}$$

3. Parameter estimation

This paper will discuss parameter estimation based on the maximum likelihood estimation (MLE) method. Let $Y_1, Y_2, Y_3, \ldots, Y_n$ denote random samples drawn from the simple discrete-weighted exponential distribution with parameters λ and ω. The log likelihood function $L(y;\lambda,\omega)$ is given by:

$$\ln L(y;\lambda,\omega) = -(\lambda-\omega)\sum_{i=1}^{n} y_i + n\ln\left(1-e^{-(\lambda-\omega)}\right) \tag{15}$$

We cannot obtain explicit expressions of estimations, so numerical analysis, such as the Newton-Raphson algorithm, can be used to estimate the value of parameters.

4. Applications

In this section, the simple discrete-weighted exponential (DWE) distribution will be examined for two real data sets and compared with the Poisson, negative binomial (NB), and discrete Lindley (DL) distributions. The analysis involved in this research was performed with the R program and estimated the parameters using the package's name "maxLik" [12]. The criteria for comparing models were the log-likelihood (LL) and the Akaike information criterion (AIC). The log-likelihood is calculated as; $LL = 2(\ln L_1 - \ln L_0)$ [13]. Where L_1 and L_0 are the log-likelihood under the respective hypothesis. The AIC is the most used fit statistic. Let L be the model likelihood, p is the number of parameters. The AIC is $AIC = -2\ln(L) + 2p$ [2]. In addition, we compute the expected frequencies for fitted DWE, Poisson, NB, DL distribution.

The StrikeNb dataset, which is included in the Ecdat R package (Website data: http://CRAN.R-project.org/package=Ecdat), is the first data set. The number of contract strikes in US industry was reported monthly from January 1968 to December 1976, which included 108 observations. The sample mean of the data is 6.75, the variance is 21.93, and the ID is 3.25. The variance is greater than the mean, indicating that there is over-dispersion. Table 1 shows the estimated parameters, LL, and AIC of the StrikeNb data set.

Table 1. The number of contracts strikes in the US manufacturing observed data set and its expected frequencies, estimated parameters, LL, and AIC

Count	Frequency	Expected frequencies			
		DWE	Poisson	NB	DL
0	5	14.1	0.6	5.0	5.6
1	12	12.3	3.2	9.3	8.7
2	14	10.7	8.3	11.5	10.1
3	11	9.3	14.3	11.8	10.4
4	9	8.1	18.4	10.9	10.0
5	14	7.0	18.9	9.4	9.2
6	9	6.1	16.2	7.8	8.3
7	4	5.3	11.9	6.2	7.3
8	7	4.6	7.7	4.7	6.4
9	10	4.0	4.4	3.6	5.5
10	6	3.5	2.3	2.6	4.6
11	1	3.0	1.1	1.9	3.9
13	3	2.3	0.5	1.0	2.7
15	1	1.7	0.2	0.5	1.8
16	1	1.5	0.1	0.3	1.5
18	1	1.1	0.0	0.2	1.0
Estimated Parameters		$\hat{\lambda} = 0.87$, $\hat{\omega} = 0.73$	$\hat{\lambda} = 5.15$	$\hat{r} = 3.20$, $\hat{p} = 0.28$	$\hat{\theta} = 0.26$
LL		-47.68	-309.26	-279.97	-59.20
AIC		99.36	620.58	563.94	120.40

Table 1 shows the first data set and the expected frequencies of DWE, Poisson, NB, and DL distributions. The comparison between the DWE distribution and other distributions based on AIC values for the StrikeNb data set. In the set of competing distributions, a distribution is selected as the best fitted distribution to the data that has the smallest AIC values. We observe that the AIC and LL values of DWE are 99.36 and -47.68, respectively, and the AIC value is the smallest. The DWE distribution is a better distribution. The MLE of the parameters of the DWE distribution fitted to the data are $\hat{\lambda} = 0.87$ and $\hat{\omega} = 0.73$.

The second set of data is the number of fires in Greece from 1 July 1998 to 31 August of the same year [3]. The sample mean of the data set is 8.20, the variance is 19.03, and the ID is 2.32. Since the sample variance is larger than the sample mean, this data set has over-dispersion. Table 2 shows the estimated parameters, LL, and AIC of the second data set.

Table 2. The number of fires in Greece data set and its expected frequencies, estimated parameters, LL, and AIC

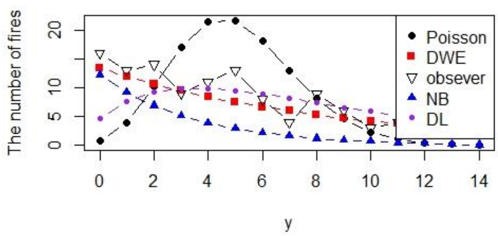
Count	Frequency	Expected frequencies			
		DWE	Poisson	NB	DL
0	16	13.4	0.8	12.2	4.7
1	13	11.9	4.0	9.2	7.6
2	14	10.6	10.2	6.9	9.2
3	9	9.5	17.0	5.2	9.8
4	11	8.4	21.4	3.9	9.9
5	13	7.5	21.6	2.9	9.5
6	8	6.7	18.1	2.2	8.9
7	4	6.0	13.0	1.7	8.2
8	9	5.3	8.2	1.2	7.4
9	6	4.7	4.6	0.9	6.6
10	3	4.2	2.3	0.7	5.9
11	4	3.8	1.1	0.5	5.1
12	6	3.4	0.4	0.4	4.5
15	4	2.4	0.2	0.2	2.9
≥ 16	3	2.1	0.1	0.1	2.4

Table 2: The number of fires in Greece data set and its expected frequencies, estimated parameters, LL, and AIC **(cont.)**

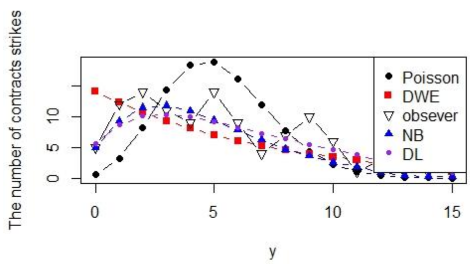
Count	Frequency	Expected frequencies			
		DWE	Poisson	NB	DL
Estimated Parameters		$\hat{\lambda}$ = 0.66	$\hat{\lambda}$ = 5.03	$\hat{r}$ =1.66	$\hat{\theta}$ = 0.22
		$\hat{\omega}$ = 0.54		$\hat{p}$ = 0.25	
LL		-47.44	- 401.79	-329.41	-51.95
AIC		98.88	804.57	662.82	105.90

Table 2 shows the second data set and the expected frequencies of DWE, Poisson, NB, and DL distributions. The comparison between the DWE distribution and other distributions based on AIC values. We found that the AIC and LL values of DWE are 98.88 and -47.44, respectively and the AIC value is the smallest. The DWE distribution is a better model. The MLE of the parameters of the DWE distribution fitted to the data are $\hat{\lambda}$ = 0.66 and $\hat{\omega}$ = 0.54.

Figure 4 displays a comparison of observed and expected frequencies in a fitted distribution; Figure (4a) shows the fitted distribution of the StrikeNb data set, while Figure (4b) shows the fitted distribution of the number of fires in the Greece data set.



(a)



(b)

Figure 4. Data and expected frequencies of fitted distribution with difference distribution

Conclusion

The discretization method is used in this paper to create a simple discrete-weighted exponential distribution. The pmf and cdf of the simple discrete-weighted exponential distribution are obtained, some statistical properties are presented, and the index of dispersion and coefficient of variation are included. Furthermore, the simple discrete-weighted exponential distribution is applied to and compared to the Poisson, NB, and DL based on maximum likelihood estimation for two over-dispersion data sets. The result, based on AIC and LL values, is that the AIC value of the DWE distribution is the smallest. Form this paper, we can conclude that the DWE distribution is better fitted to the data than the Poisson, NB, and DL distributions. So, the DWE distribution can be used as an alternative to the over-dispersion data.

References

1. Hilbe JM. Modeling count data. 3rd. Cambridge University Press; 2014.
2. Cameron AC, Trivedi PK. Regression analysis of count data. Cambridge. 2nd. New York: The United States of America; 2013.
3. Nekoukhou V, Bidram H. The exponentiated discrete Weibull distribution. SORT 2017;1:127-46.
4. Borah M. A new quasi poisson-lindley distribution: porperies and applications. J Stat Sci Appl 2017;16:576-88.
5. Rasekhi M, Chatrabgounand O, Daneshkhah A. Discrete weighted exponential distribution: properties and applications. Filomat 2018;32:3043-56. Doi:10.2298/FIL1808043R
6. Chakraborty S. Generating discrete analogues of continuous probability distribution-A survey of methods and constructions. J Stat Distrib 2015;2:1-30. Doi:10.1186/s40488-015-0028-6.
7. Khongthip P, Patummasut M, Bodhisuwan W. The discrete weighted exponential distribution and its applications. Songklanakarin J Sci Technol 2018;40:1105-14.
8. Diandarma R, Nurrohmah S, Fithrianai I. Discrete Lindley distribution. Proceedings of the 5th International Symposium on Current Progress in Mathematics and Sciences (ISCPMS2019); 2019 July 9-10; Depok, Indonesia. AIP Conference Proceedings 2242, 030024; 2020.
9. Bashir S, Naqvi IB. A new weighted exponential distribution and its applications on waiting time data. Int J Sci Res Publ 2016;6:698-702.
10. Rinne H. The Hazard Rete: Theory and Inference (with supplementary MATLAB-Programs). Department of Economics and Management Science, Justus-Liebig-University, Germany; 2014.

11. Norman JL, Kemp AW, Kotz S. Univariate Discrete Distribution. 3rd A John Wiley & Sons, INC.; 2005.

12. Henningsen A, Toomet O. maxLik: A package for maximum likelihood estimation in R. Comput Stat 2011;26:443-58.

13. Montgomery DC, Peck EA, Vining GG. Introduction linear regression analysis. John Wiley & Sons, Inc, Hoboken, New Jersey; 2012.