PROGRESS IN APPLIED SCIENCE AND TECHNOLOGY
Faculty of Science and Technology
Rajamangala University of Technology Thanyaburi
Prog Appl Sci Tech. 2023; 13(3):1-12
ISSN (Print): 2730-3012
ISSN (Online): 2730-3020
https://ph02.tci-thaijo.org/index.php/past

Research Article

Received: June 23, 2023
Revised: August 14, 2023
Accepted: August 25, 2023

DOI: 10.60101/past.2023.250004

Discrete Odd Inverse Pareto Exponential Distribution: Properties, Estimation and Applications

Pornpop Saengthong[1] and Palakorn Seenoi[2*]
[1] *Institute for Research and Academic Services, Assumption University, Samut Prakan 10570, Thailand*
[2] *Department of Statistics, Faculty of Science, Khon Kaen University, Khon Kaen 40002, Thailand*
*E-mail: palakorns@kku.ac.th

Abstract

In this paper, a new discrete distribution named the discrete odd inverse Pareto exponential (DOIPEx) distribution is introduced for modeling count data. The proposed distribution can be viewed as a generalization of the discrete Marshall-Olkin exponential, discrete exponentiate exponential, and discrete exponential distributions. Some basic distributional properties, quantile function, hazard and reversed hazard functions, and order statistics are derived. Estimation of the parameters is illustrated using the maximum likelihood estimation (MLE), and three real data sets are discussed to demonstrate the usefulness and applicability of the DOIPEx distribution.

Keywords: Discrete Odd Inverse Pareto Exponential Distribution, Discrete Lifetime Distribution, Quantile Function, Maximum Likelihood Estimation, Count Data

1. Introduction

The classical lifetime distribution, such as exponential, gamma, Lindley, and Weibull distributions, are commonly used for modeling lifetime data. However, they may not fit some datasets in reliability, engineering, biomedical, and economics due to their monotone hazard rate shapes. For that reason, several methods for extensions and modified forms of the lifetime distributions have been proposed to accommodate non-monotone hazard rates for the past two decades.

In addition, it has been observed that there are many situations where original variables may not be measured on a continuous scale in practice but may often be recorded as discrete for convenience, for example, the age of people, the number of hospitals stays, or life length of a device. In this case, generalization techniques for the traditional lifetime distribution are not capable of deriving an associated distribution. Therefore, discrete distributions have been developed by discretizing the continuous distributions to describe discrete lifetime data. Applications using the discretized distributions can be found in many areas, for instance, reliability and failure times (1, 2), economics (3), engineering (4), medical sciences (5, 6), and accident statistics (7, 8). Although many discrete distributions exist in the statistical literature, only some are suitable in some situations. We propose a new three-parameter discrete odd inverse Pareto exponential (DOIPEx) distribution to fill these gaps as a competitive alternative for modeling discrete lifetime data.

For the rest of this paper, the odd inverse Pareto exponential (OIPEx) distribution is introduced in the material and methods section. Next, a construction of a new discrete distribution is provided, and some characteristics, properties, and parameter estimation are derived.

Moreover, it deals with three real applications with underdispersed, equalizers, and overdispersed data. Finally, the conclusions of the paper are presented.

2. Materials and Experiment

2.1 The Odd Inverse Pareto Exponential Distribution

Recently, Aldahlan and Afify (9) have developed a new three-parameter lifetime distribution called the odd inverse Pareto exponential (OIPEx) distribution. Its failure rate function allows for constant, decreasing, increasing, and upside-down bathtub or bathtub-shaped. The probability density function (pdf), cumulative distribution function (cdf), along with survival function of the OIPEx distribution are respectively given by:

$$f_{\text{OIPEx}}(x;\alpha,\beta,\lambda) = \frac{\alpha\beta\lambda\, exp(-\lambda x)[1-exp(-\lambda x)]^{\alpha-1}}{[1-(1-\beta)\, exp(-\lambda x)]^{\alpha+1}}, \quad (2.1)$$

$$F_{\text{OIPEx}}(x;\alpha,\beta,\lambda) = \frac{[1-exp(-\lambda x)]^{\alpha}}{[1-(1-\beta)\, exp(-\lambda x)]^{\alpha}}, \quad (2.2)$$

$$S_{\text{OIPEx}}(x;\alpha,\beta,\lambda) = 1-\left(\frac{[1-exp(-\lambda x)]^{\alpha}}{[1-(1-\beta)\, exp(-\lambda x)]^{\alpha}}\right), \quad (2.3)$$

By $f_{\text{OIPEx}}(x;\alpha,\beta,\lambda)$ in (2.1) and $S_{\text{OIPEx}}(x;\alpha,\beta,\lambda)$ in (2.3), the hazard rate function is defined by:

$$h_{\text{OIPEx}}(x;\alpha,\beta,\lambda) = \frac{\alpha\beta\lambda\, exp(-\lambda x)[1-exp(-\lambda x)]^{\alpha-1}[1-(1-\beta)\, exp(-\lambda x)]^{-\alpha-1}}{1-[1-exp(-\lambda x)]^{\alpha}[1-(1-\beta)\, exp(-\lambda x)]^{-\alpha}}, \quad (2.4)$$

where $x>0$ and $\alpha,\beta,\lambda>0$. This distribution contains three lifetime distributions as special cases, namely the exponentiated exponential distribution (when $\beta=1$), the Marshall-Olkin exponential distribution (when $\alpha=1$) and the exponential distribution (when $\alpha=\beta=1$).

2.2 Discretized distributions

Let X be a continous random variable that has a survival function, $S_X(x)$, then the discrete random variable Y is defined as the floor value of X, denoted as $Y=\lfloor X \rfloor$, where $\lfloor X \rfloor$ represents the largest integer that is less than or equal to X which will have the probability mass function (pmf) is

$$\begin{aligned} P(Y=y) &= P(\lfloor X \rfloor = y) \\ &= P(y \le x < y+1) \\ &= S_X(y)-S_X(y+1), \end{aligned} \quad (2.5)$$

where $y=0,1,2,\ldots$ (10-11).

3. Results and Discussion

This section explains how to develop the DOIPEx distribution. Some structural properties of the distribution, including hazard and reversed hazard functions, quantile function, and order statistics, are discussed. In addition, the parameter estimation and the application study are illustrated.

3.1 A New Discretization of the OIPEx distribution

Theorem 1. Let Y be a random variable of the DOIPEx distribution with parameters α,β and λ, denoted as $Y \sim \text{DOIPEx}(\alpha,\beta,\lambda)$. Then, the pmf of Y is given by

$$f_{\text{DOIPEx}}(y;\alpha,\beta,\lambda) = \left(\frac{[1-exp(-\lambda y-\lambda)]^{\alpha}}{[1-(1-\beta)\, exp(-\lambda y-\lambda)]^{\alpha}}\right) - \left(\frac{[1-exp(-\lambda y)]^{\alpha}}{[1-(1-\beta)\, exp(-\lambda y)]^{\alpha}}\right), \quad (3.1)$$

where $y=0,1,2,\ldots,$ for α,β and $\lambda>0$.

Proof. It can be verified that the pmf of the $\text{DOIPEx}(\alpha,\beta,\lambda)$ is obtained by replacing the survival function of the OIPEx distribution in (2.3) into (2.5). Then, it is finally given as

$$\begin{aligned} f_{\text{DOIPEx}}(y;\alpha,\beta,\lambda) &= S_{\text{OIPEx}}(y)-S_{\text{OIPEx}}(y+1) \\ &= 1-\left(\frac{[1-exp(-\lambda y)]^{\alpha}}{[1-(1-\beta)\, exp(-\lambda y)]^{\alpha}}\right) \\ &\quad -\left(1-\left(\frac{[1-exp(-\lambda(y+1))]^{\alpha}}{[1-(1-\beta)\, exp(-\lambda(y+1))]^{\alpha}}\right)\right) \\ &= \left(\frac{[1-exp(-\lambda y-\lambda)]^{\alpha}}{[1-(1-\beta)\, exp(-\lambda y-\lambda)]^{\alpha}}\right) \\ &\quad -\left(\frac{[1-exp(-\lambda y)]^{\alpha}}{[1-(1-\beta)\, exp(-\lambda y)]^{\alpha}}\right). \end{aligned}$$

Some specified parameters of the DOIPEx distribution and their pmf are plotted in Figure 1.

It is evident that the distribution of the random variable Y can be monotonically decreasing (reverse J-shape), right skewed, or closed to a symmetric. Additionally, the shape of distribution changes when the parameters α and β changes.

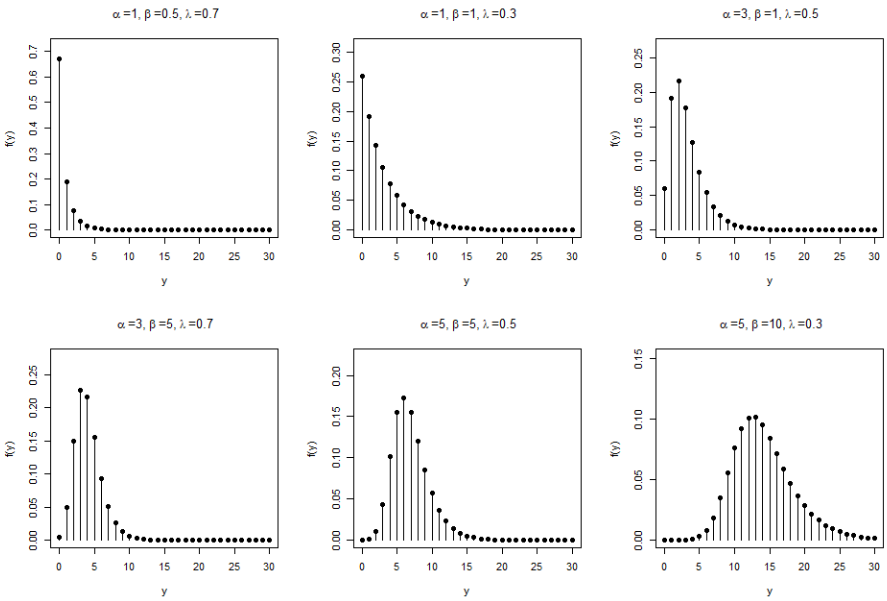

Figure 1 The probability mass function of a DOIPEx random variable with some specific values of α, β and λ

Theorem 2. Let Y be a random variable of the DOIPEx distribution with parameters α, β and λ, denoted as $Y \sim \text{DOIPEx}(\alpha, \beta, \lambda)$. Then, the cumulative distribution function (cdf) of Y is obtained as

$$F_{\text{DOIPEx}}(y; \alpha, \beta, \lambda) = \left(\frac{[1-exp(-\lambda y-\lambda)]^{\alpha}}{[1-(1-\beta)\, exp(-\lambda y-\lambda)]^{\alpha}}\right), \tag{3.2}$$

where $y = 0,1,2,...,$ for α, β and $\lambda > 0$.

Proof. From (6)

$$F(y) = P(Y \leq y)$$
$$= 1 - S_X(y) + P(Y = y)$$
$$= 1 - S_X(y+1),$$

we get

$$F_{\text{DOIPEx}}(y; \alpha, \beta, \lambda) = 1 - S_{\text{OIPEx}}(y+1)$$
$$= 1 - \left(1 - \left(\frac{[1-exp(-\lambda(y+1))]^{\alpha}}{[1-(1-\beta)\, exp(-\lambda(y+1))]^{\alpha}}\right)\right)$$
$$= \left(\frac{[1-exp(-\lambda y-\lambda)]^{\alpha}}{[1-(1-\beta)\, exp(-\lambda y-\lambda)]^{\alpha}}\right).$$

Plots of the cdf of DOIPEx distribution with some specific parameter values are shown in Figure 2.

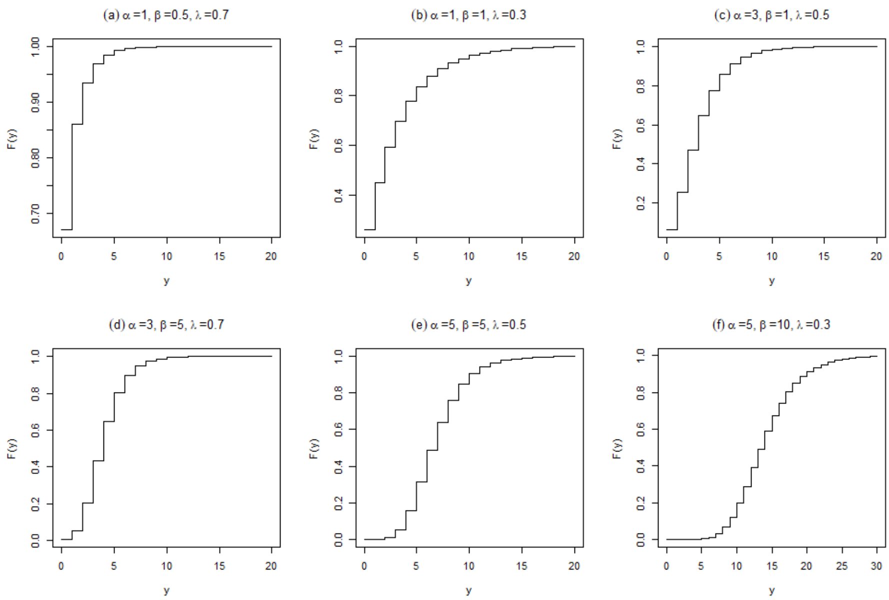

Figure 2 The cumulative distribution function of a DOIPEx random variable with some specific values of α, β and λ

3.2 Hazard and reversed hazard functions

The survival and hazard rate functions of $\mathrm{DOIPEx}(\alpha,\beta,\lambda)$ distribution are given by

$$S_{\mathrm{DOIPEx}}(y;\alpha,\beta,\lambda) = 1 - F_{\mathrm{DOIPEx}}(y;\alpha,\beta,\lambda)$$
$$= 1 - \left(\frac{[1-exp(-\lambda y-\lambda)]^{\alpha}}{[1-(1-\beta)\,exp(-\lambda y-\lambda)]^{\alpha}}\right),$$

and

$$h_{\mathrm{DOIPEx}}(y;\alpha,\beta,\lambda) = \frac{f_{\mathrm{DOIPEx}}(y;\alpha,\beta,\lambda)}{S_{\mathrm{DOIPEx}}(y-1;\alpha,\beta,\lambda)}$$
$$= \frac{\left(\frac{[1-exp(-\lambda y-\lambda)]^{\alpha}}{[1-(1-\beta)\,exp(-\lambda y-\lambda)]^{\alpha}}\right) - \left(\frac{[1-exp(-\lambda y)]^{\alpha}}{[1-(1-\beta)\,exp(-\lambda y)]^{\alpha}}\right)}{1-\left(\frac{[1-exp(-\lambda(y-1)-\lambda)]^{\alpha}}{[1-(1-\beta)\,exp(-\lambda(y-1)-\lambda)]^{\alpha}}\right)}$$
$$= \frac{\left(\frac{[1-exp(-\lambda y-\lambda)]^{\alpha}}{[1-(1-\beta)\,exp(-\lambda y-\lambda)]^{\alpha}}\right) - \left(\frac{[1-exp(-\lambda y)]^{\alpha}}{[1-(1-\beta)\,exp(-\lambda y)]^{\alpha}}\right)}{1-\left(\frac{[1-exp(-\lambda y)]^{\alpha}}{[1-(1-\beta)\,exp(-\lambda y)]^{\alpha}}\right)}.$$

In addition, the reversed hazard function is defined as

$$r_{\mathrm{DOIPEx}}(y;\alpha,\beta,\lambda) = \frac{f_{\mathrm{DOIPEx}}(y;\alpha,\beta,\lambda)}{F_{\mathrm{DOIPEx}}(y;\alpha,\beta,\lambda)}$$
$$= \frac{\left(\frac{[1-exp(-\lambda y-\lambda)]^{\alpha}}{[1-(1-\beta)\,exp(-\lambda y-\lambda)]^{\alpha}}\right) - \left(\frac{[1-exp(-\lambda y)]^{\alpha}}{[1-(1-\beta)\,exp(-\lambda y)]^{\alpha}}\right)}{\left(\frac{[1-exp(-\lambda y-\lambda)]^{\alpha}}{[1-(1-\beta)\,exp(-\lambda y-\lambda)]^{\alpha}}\right)}$$
$$= 1-\left(\frac{\frac{[1-exp(-\lambda y)]^{\alpha}}{[1-(1-\beta)\,exp(-\lambda y)]^{\alpha}}}{\frac{[1-exp(-\lambda y-\lambda)]^{\alpha}}{[1-(1-\beta)\,exp(-\lambda y-\lambda)]^{\alpha}}}\right).$$

Figure 3 and Figure 4 illustrate the hazard and reversed hazard functions of $\mathrm{DOIPEx}(\alpha,\beta,\lambda)$ distribution for different values of α,β and λ. In addition, we obtained the following discrete distributions as special cases.

1) The Discrete Marshall-Olkin Exponential (DMOEx) Distribution

If $\alpha = 1$, then the DOIPEx distribution reduces to the DMOEx distribution with the pmf given as

$$f_{\mathrm{DMOEx}}(y;\beta,\lambda) = \left(\frac{1-exp(-\lambda y-\lambda)}{1-(1-\beta)\,exp(-\lambda y-\lambda)}\right) - \left(\frac{1-exp(-\lambda y)}{1-(1-\beta)\,exp(-\lambda y)}\right), \quad (3.3)$$

where $y = 0,1,2,\ldots$, for β and $\lambda > 0$.

2) The Discrete Exponentiate Exponential (DEEx) Distribution

If $\beta = 1$, then the DOIPEx distribution reduces to the DEEx distribution with the pmf given by

$$f_{\text{DEEx}}(y;\alpha,\lambda) = [1 - exp(-\lambda y - \lambda)]^{\alpha} - [1 - exp(-\lambda y)]^{\alpha}, \quad (3.4)$$

where $y = 0,1,2,...,$ for α and $\lambda > 0$.

3) The Discrete Exponential (DEx) Distribution

If $\alpha = \beta = 1$, then the DOIPEx distribution reduces to the DEx distribution (12). The pmf of the DEx distribution is

$$f_{\text{DEx}}(y;\lambda) = exp(-\lambda y) - exp(-\lambda y - \lambda), \quad (3.5)$$

where $y = 0,1,2,...,$ for $\lambda > 0$.

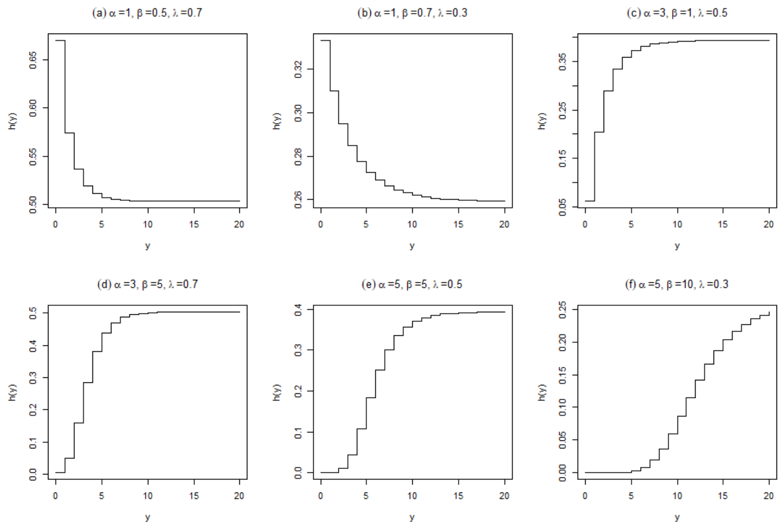

Figure 3 The hazard function of a DOIPEx random variable with some specific values of α, β and λ

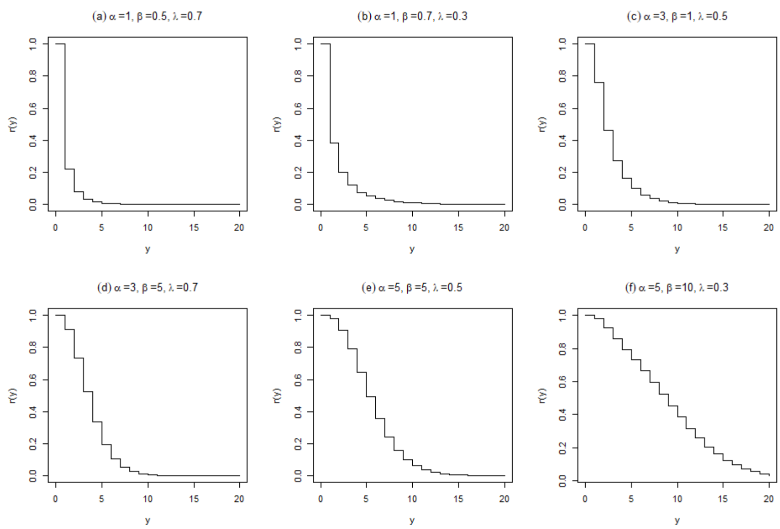

Figure 4 The reversed hazard function of a DOIPEx random variable with some specific values of α, β and λ

3.3 Quantile Function

From (9), the Quantile Function of OIPEx distribution can be expressed as

$$y_u = Q_{\text{OIPEx}}(u) = \frac{-1}{\lambda} log\left(1 - \frac{\beta u^{1/\alpha}}{1-(1-\beta)u^{1/\alpha}}\right).$$

Simultaneously, we found that $F_{\text{DOIPEx}}(y;\alpha,\beta,\lambda) = F_{\text{OIPEx}}(y+1;\alpha,\beta,\lambda)$, then the quantile function of the DOIPEx, $Q_{\text{DOIPEx}}(u)$, can be obtained as follows:

1) Let $u = F_{\text{DOIPEx}}(y)$ or $u = F_{\text{OIPEx}}(y+1)$, then $Q_{\text{OIPEx}}(u) = y_u + 1$.

2) Substituting $Q_{\text{OIPEx}}(u)$ into the above equation, we get $\frac{-1}{\lambda} log\left(1 - \frac{\beta u^{1/\alpha}}{1-(1-\beta)u^{1/\alpha}}\right) = y_u + 1$.

Accordingly,

$$y_u = \frac{-1}{\lambda} log\left(1 - \frac{\beta u^{1/\alpha}}{1-(1-\beta)u^{1/\alpha}}\right) - 1$$
$$= -\left[1 + \frac{1}{\lambda} log\left(1 - \frac{\beta u^{1/\alpha}}{1-(1-\beta)u^{1/\alpha}}\right)\right].$$

3) Since Y is a discrete random variable, the number is rounded to the nearest integer by using round function. Then, the quantile function of the DOIPEx distribution is

$$Q_{\text{DOIPEx}}(u) = y_u = \left\lfloor -\left[1 + \frac{1}{\lambda} log\left(1 - \frac{\beta u^{1/\alpha}}{1-(1-\beta)u^{1/\alpha}}\right)\right]\right\rfloor, \quad (3.6)$$

where $\lfloor . \rfloor$ is the floor function.

To generate a random variable Y from the DOIPEx(α,β,λ), which is based on generating random data by (3.6), one can use following algorithm:

1) Set the values of α, β and λ.

2) Set the sample size of n.

3) Generate U_i according to the uniform distribution on interval (0,1) where $i = 1,2,...,n$.

4) Make the random number of Y by

$$Y_i = \left\lfloor -\left[1 + \frac{1}{\lambda} log\left(1 - \frac{\beta u_i^{1/\alpha}}{1-(1-\beta)u_i^{1/\alpha}}\right)\right]\right\rfloor.$$

3.4 Order Statistics

Let $Y_1, Y_2, \ldots, Y_n$ be n independent and identically distributed (iid) random variables, i.e., $Y_i \sim$ DOIPEx(α,β,λ), and let $Y_{(1)} \le Y_{(2)} \le \cdots \le Y_{(n)}$ denote these random variables rearranged in non-descending order magnitude. Then, the pmf of the *r*th order statistic can always be expressed as follows

$$f_{\text{DOIPEx}_{(r)}}(y) = \frac{n!}{(r-1)!\,(n-r)!} f_{\text{DOIPEx}}(y) \times [F_{\text{DOIPEx}}(y)]^{r-1}[1 - F_{\text{DOIPEx}}(y)]^{n-r}$$
$$= \frac{n!}{(r-1)!\,(n-r)!}\left\{\left(\frac{[1-exp(-\lambda(y+1))]^{\alpha}}{[1-(1-\beta)\,exp(-\lambda(y+1))]^{\alpha}}\right) - \left(\frac{[1-exp(-\lambda y)]^{\alpha}}{[1-(1-\beta)\,exp(-\lambda y)]^{\alpha}}\right)\right\}$$
$$\times \left\{\frac{[1-exp(-\lambda(y+1))]^{\alpha}}{[1-(1-\beta)\,exp(-\lambda(y+1))]^{\alpha}}\right\}^{r-1}$$
$$\times \left\{1 - \frac{[1-exp(-\lambda(y+1))]^{\alpha}}{[1-(1-\beta)\,exp(-\lambda(y+1))]^{\alpha}}\right\}^{n-r}, \quad (3.7)$$

where $y = 0,1,2,\ldots$ and $r = 1,2,\ldots,n$.

3.5 Index of Dispersion

The index of dispersion (ID) is defined as the ratio of the variance to the mean. It indicates underdispersion for ID < 1, equidispersion for ID =1 and overdispersion for ID > 1. The mean and variance of DOIPEx distribution are respectively given by

$$E(Y) = \sum_{y} y f_{\text{DOIPEx}}(y;\alpha,\beta,\lambda)$$

and

$$V(Y) = \textstyle\sum_{Y} y^2 f_{\text{DOIPEx}}(y;\alpha,\beta,\lambda) - \left(E(Y)\right)^2,$$

where $y = 0,1,2,...,$ and $f_{\text{DOIPEx}}(y;\alpha,\beta,\lambda)$ is the pmf of the DOIPEx(α,β,λ) as in (3.1). However, these above equations cannot be expressed in closed forms. Table 1 presents the values of the mean, variance, and ID for case examples using specific parameter values α, β, and λ from Figure 1. Derived from calculations using the R program (13), when substituting parameter values α, β, and λ and $y=1,2,\ldots,100$ into $f_{\text{DOIPEx}}(y;\alpha,\beta,\lambda)$ as (3.1), it is observed that $\sum_{y=1}^{100} f_{\text{DOIPEx}}(y;\alpha,\beta,\lambda)=1$ for every set of parameters (specific parameter values), thus substituting parameter α, β, and λ and $y=1,2,\ldots,100$ into the formulas for $E(Y)$ and $V(Y)$ to obtain the results for mean, variance, and ID. The results show the evidence of overdispersion (Case 1, 2, 3, 6), underdispersion (Case 4) and equidispersion (Case 5). In addition, the ID seems to increase with the increase of α for fixed β and λ, as well as β for fixed α and λ, or λ for fixed α and β.

Table 1 Values of the mean, variance, and ID of DOIPEx for varying values of α,β and λ

Case	α	β	λ	$E(Y)$	$V(Y)$	ID
1	1	0.5	0.7	0.5978	1.2725	2.1286
2	1	0.7	0.3	2.3074	9.0377	3.9168
3	3	1	0.5	3.1658	5.5332	1.7478
4	3	5	0.7	4.0672	3.7841	0.9304
5	5	5	0.5	6.9797	7.1032	1.0177
6	5	10	0.3	14.0461	19.2519	1.3706

3.6 Estimation of Parameters

In this section, the parameter estimation of DOIPEx(α,β,λ) will be done based on maximum likelihood estimation (MLE) procedure. Consider $(y_1,y_2,\ldots,y_n)$ is a random sample of size n from the DOIPEx distribution in (3.1). The likelihood function of DOIPEx(α,β,λ) is given by

$$L(\alpha,\beta,\lambda)=\prod_{i=1}^{n}\left\{\begin{matrix}\left(\frac{[1-exp(-\lambda y_i-\lambda)]^\alpha}{[1-(1-\beta)\,exp(-\lambda y_i-\lambda)]^\alpha}\right)\\ -\left(\frac{[1-exp(-\lambda y_i)]^\alpha}{[1-(1-\beta)\,exp(-\lambda y_i)]^\alpha}\right)\end{matrix}\right\},$$

then the corresponding log-likelihood function can be calculated as

$$log\,L(\alpha,\beta,\lambda)=\sum_{i=1}^{n} log\left\{\begin{matrix}\left(\frac{[1-exp(-\lambda y_i-\lambda)]^\alpha}{[1-(1-\beta)\,exp(-\lambda y_i-\lambda)]^\alpha}\right)\\ -\left(\frac{[1-exp(-\lambda y_i)]^\alpha}{[1-(1-\beta)\,exp(-\lambda y_i)]^\alpha}\right)\end{matrix}\right\}.$$

Hence, the likelihood equations are

$$\frac{\partial}{\partial\alpha}log\,L(\alpha,\beta,\lambda)=\frac{\partial}{\partial\alpha}\sum_{i=1}^{n} log\left\{\begin{matrix}\left(\frac{[1-exp(-\lambda y_i-\lambda)]^\alpha}{[1-(1-\beta)\,exp(-\lambda y_i-\lambda)]^\alpha}\right)\\ -\left(\frac{[1-exp(-\lambda y_i)]^\alpha}{[1-(1-\beta)\,exp(-\lambda y_i)]^\alpha}\right)\end{matrix}\right\}$$
$$=\sum_{i=1}^{n}\left\{\frac{1}{\frac{A^\alpha}{B^\alpha}-\frac{C^\alpha}{D^\alpha}}\left[\frac{A^\alpha}{B^\alpha}log\left(\frac{A}{B}\right)-\frac{C^\alpha}{D^\alpha}log\left(\frac{C}{D}\right)\right]\right\},$$

$$\frac{\partial}{\partial\beta}log\,L(\alpha,\beta,\lambda)=\frac{\partial}{\partial\beta}\sum_{i=1}^{n} log\left\{\begin{matrix}\left(\frac{[1-exp(-\lambda y_i-\lambda)]^\alpha}{[1-(1-\beta)\,exp(-\lambda y_i-\lambda)]^\alpha}\right)\\ -\left(\frac{[1-exp(-\lambda y_i)]^\alpha}{[1-(1-\beta)\,exp(-\lambda y_i)]^\alpha}\right)\end{matrix}\right\}$$
$$=\sum_{i=1}^{n}\left\{\frac{-\alpha}{\frac{A^\alpha}{B^\alpha}-\frac{C^\alpha}{D^\alpha}}\left[\frac{A^\alpha}{B^{\alpha+1}}exp(-\lambda y_i-\lambda)-\frac{C^\alpha}{D^{\alpha+1}}exp(-\lambda y_i)\right]\right\},$$

$$\frac{\partial}{\partial\lambda}log\,L(\alpha,\beta,\lambda)=\frac{\partial}{\partial\lambda}\sum_{i=1}^{n} log\left\{\begin{matrix}\left(\frac{[1-exp(-\lambda y_i-\lambda)]^\alpha}{[1-(1-\beta)\,exp(-\lambda y_i-\lambda)]^\alpha}\right)\\ -\left(\frac{[1-exp(-\lambda y_i)]^\alpha}{[1-(1-\beta)\,exp(-\lambda y_i)]^\alpha}\right)\end{matrix}\right\}$$
$$=\sum_{i=1}^{n}\left\{\frac{1}{\frac{A^\alpha}{B^\alpha}-\frac{C^\alpha}{D^\alpha}}\left[\begin{matrix}\left(\begin{matrix}\alpha(1-\beta)(-y_i-1)\frac{A^\alpha}{B^{\alpha+1}}exp(-\lambda y_i-\lambda)\\ +2(y_i+1)\frac{A^{\alpha-1}}{B^\alpha}exp(-\lambda y_i-\lambda)\end{matrix}\right)\\ -\left((-2y_i)(1-\beta)\frac{C^\alpha}{D^{\alpha+1}}exp(-\lambda y_i)+(2y_i)\frac{C^{\alpha-1}}{D^\alpha}exp(-\lambda y_i)\right)\end{matrix}\right]\right\},$$

Where $A=1-exp(-\lambda y_i-\lambda)$, $B=1-(1-\beta)\,exp(-\lambda y_i-\lambda)$, $C=1-exp(-\lambda y_i)$ and $D=1-(1-\beta)\,exp(-\lambda y_i)$.

With the above equating to zero, these equations cannot be solved analytically, and they have to be obtained numerically by using the Newton-Raphson procedure via the nonlinear minimization (nlm) function in the statistical package of R language (13).

3.7 Applications

In this section, we provide the applications of the DOIPEx distribution to real-life count data sets with different features, namely: underdispersion, equidispersion, and overdispersion. A comparison of the proposed DOIPEx distribution with its sub-models (i.e., DEx, DEEx, and DMOEx distributions) and the Poisson distribution has been performed by using the Chi-square test, Kolmogorov-Smirnov (K-S) test along with the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The better distribution is the distribution that corresponds to smaller values of AIC, BIC, and a larger *p*-value of Chi-square and K-S from the goodness-of-fit test.

3.7.1 Application 1: Underdispersed Data

The data set obtained from (14) provides information about the word lengths in a Slovak poem. The mean, variance, and dispersion index of the 117 counts are 2.8889, 0.9100, and 0.3150, respectively, indicating underdispersed data. In evaluating DOIPEx with other distributions, Table 2 shows that if we use AIC or BIC as the tool for comparison, the DMOEx gives a better fit than the others. At the same time, the DOIPEx performance is quite competitive. However, if we use the Chi-square or K-S goodness-of-fit test to compare fitting distributions, then the DOIPEx distribution best fits the data. Figure 5 shows a comparison between observed and expected values of fitted distributions.

Table 2 Observed and expected frequencies for the number of syllables of words in a Slovak poem.

No. of runs	Observed	Fitting distributions				
		Poisson	DEx	DEEx	DMOEx	DOIPEx
0	0	6.5	30.1	0.0 } 46.1	1.0 } 7.6	**0.5** } **6.3**
1	7	18.8	22.3	4.5	6.6	**5.8**
2	33	27.2	16.6	41.6	30.7	33.5
3	49	26.2	12.3	43.0	51.8	50.3
4	22	18.9	9.2	18.9	22.0 } 26.9	21.1
5	6	10.9	6.8	9.0	4.9	5.9
Estimated parameters		$\hat{\lambda}$ =2.889	$\hat{\lambda}$ = 0.297	$\hat{\alpha}$ = 35.791 $\hat{\lambda}$ = 1.221	$\hat{\beta}$ = 659.691 $\hat{\lambda}$ = 1.925	$\hat{\alpha}$ = 1.532 $\hat{\boldsymbol{\beta}}$ =165.558 $\hat{\lambda}$ = 1.697
-Log likelihood		188.738	259.372	163.654	161.020	160.568
AIC		379.476	520.744	331.308	326.040	327.136
BIC		382.238	523.506	336.832	331.564	335.423
Chi-squares		37.696	184.206	3.176	0.4179	0.164
Degree of freedom		4	4	1	1	1
p-value		<0.0001	<0.0001	0.0747	0.5180	0.6856
K-S		0.1708	0.4321	0.0525	0.0233	0.0133
p-value		0.0022	<0.0001	0.9032	1.0000	1.0000

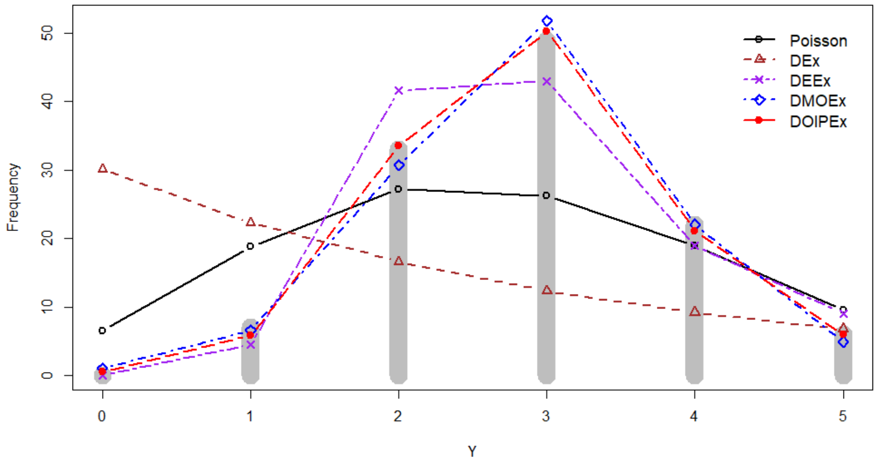

Figure 5 Empirical and fitted distributions plot for Application 1

3.7.2 Application 2: Equidispersed Data

The data set in Table 3 is the number of accidents per day which resulted in a claim to the insurance company (15). The mean and variance of the 365 counts are 2.0137 and 2.0190, and the index of dispersion is 1.0026, which gives evidence of equidispersion. Examination of fitting distributions in Table 3 indicates that if we consider AIC or BIC, the Poisson distribution gives a better fit compared to the others, while the performance of DMOEx and DOIPEx is quite competitive. However, if we consider the Chi-square or K-S goodness-of-fit test to compare fitting distributions, then the DOIPEx distribution provides the best fit to the data. Figure 6 shows a comparison between observed and expected values of fitted distributions.

Table 3 Observed and expected frequencies of the number of accidents per day which resulted in a claim to the insurance company.

No. of claims/day	Observed	Fitting distributions				
		Poisson	DEx	DEEx	DMOEx	DOIPEx
0	47	48.7	121.1	40.8	51.5	46.6
1	97	98.1	80.9	117.3	92.5	99.5
2	109	98.8	54.1	97.5	103.1	103.3
3	62	66.3	36.1	56.2	68.4	65.2
4	25	33.4	24.1	28.3	31.3	30.5
5	16	13.4	16.1	13.4	11.9	12.4
6	9	6.2	10.8	11.5	6.3	7.5
Estimated parameters		$\hat{\lambda}$ =2.014	$\hat{\lambda}$ = 0.403	$\hat{\alpha}$ = 3.617 $\hat{\lambda}$ = 0.789	$\hat{\beta}$ =11.953 $\hat{\lambda}$ = 1.087	$\hat{\alpha}$ = 1.346 $\hat{\beta}$ = 6.186 $\hat{\lambda}$ = 0.997
-Log likelihood		622.426	699.005	626.400	623.282	622.519
AIC		1246.852	1400.010	1256.800	1250.564	1251.038
BIC		1250.752	1403.910	1264.600	1258.364	1262.738
Chi-squares		5.234	123.130	7.838	5.394	2.875
Degree of freedom		5	5	4	4	3
p-value		0.3880	<0.0001	0.0977	0.2492	0.4113
K-S		0.0202	0.2031	0.0387	0.0186	0.0140
p-value		0.9984	<0.0001	0.6465	0.9996	1.0000

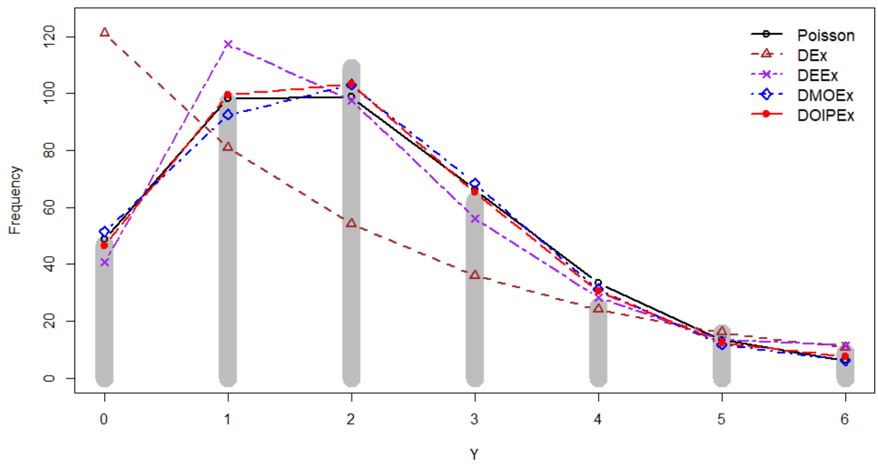

Figure 6 Empirical and fitted distributions plot for Application 2.

3.7.3 Application 3: Overdispersed Data

The last data set taken from Flynn (16), represent the 4406 counts of number of hospital stays of United States' residents aged 66 and over. In Table 4, the data contain evidence of overdispersion with a mean = 0.2960, variance = 0.5571 and index of dispersion = 1.8821. We discuss the appropriateness of the distributions through the goodness-of-fit; Chi-square and K-S test as well as the two information-based criteria; AIC and BIC. The results show that the DOIPEx is competing well against the others and gives a satisfactory fit (see Figure 7).

Table 4 Observed and expected frequencies of the number of hospital stays of United States' residents aged 66 and over.

No. of hospital stays	Observed	Fitting distributions				
		Poisson	DEx	DEEx	DMOEx	DOIPEx
0	3541	3277.1	3399.8	3543.1	3544.3	3544.7
1	599	970.0	776.4	575.6	580.8	588.4
2	176	143.6	177.3	186.8	178.1	170.4
3	48	14.2	40.5	64.8	63.7	61.8
4	20	1.1	9.3	23.0	24.0	24.2
5	12	0.1 } 15.3	2.1	8.2	9.2	9.8
6	5	0.0	0.5 } 12.0	2.9 } 12.7	3.6	**4.0**
7	1	0.0	0.1	1.0	1.4 } 5.9	**1.6** } **6.7**
8	4	0.0	0.0	0.6	0.9	**1.1**
Estimated parameters		$\hat{\lambda}$ =0.296	$\hat{\lambda}$ = 1.477	$\hat{\alpha}$ = 0.492 $\hat{\lambda}$ = 1.028	$\hat{\beta}$ = 0.382 $\hat{\lambda}$ = 0.944	$\hat{\alpha}$ = 9.126 $\boldsymbol{\hat{\beta} = 0.035}$ $\hat{\lambda}$ = 0.888
-Log likelihood		3304.509	3067.988	3012.809	3010.337	3009.007
AIC		6611.018	6137.976	6029.618	6024.674	6024.014
BIC		6617.409	6144.367	6042.399	6037.455	6043.186
Chi-squares		535.930	122.980	13.056	8.931	6.263
Degree of freedom		2	3	3	4	3
p-value		<0.0001	<0.0001	0.0045	0.0629	0.0995
K-S		0.0599	0.0320	0.0048	0.0034	0.0028
p-value		<0.0001	0.0002	1.0000	1.0000	1.0000

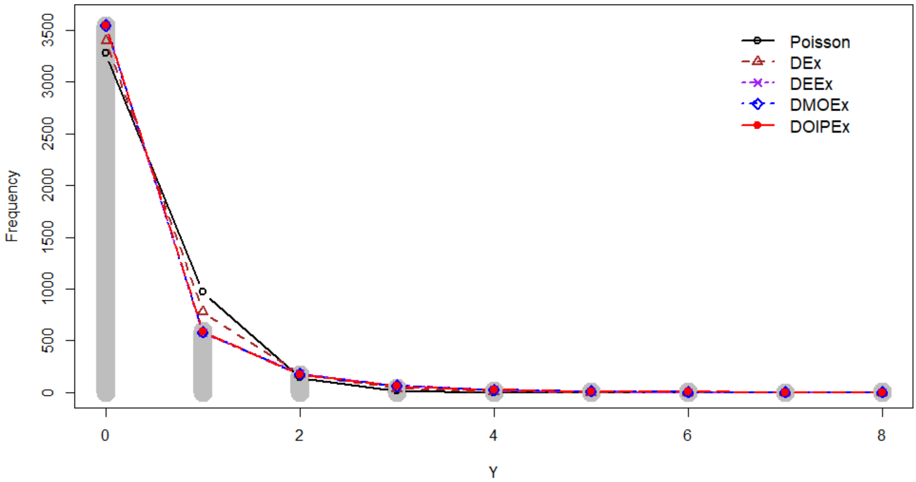

Figure 7 Empirical and fitted distributions plot for Application 3.

4. Conclusions

This paper presents the DOIPEx distribution, which is obtained by discretizing the continuous lifetime model of the OIPEx distribution. The DMOEx, DEEx, and DEx distributions are all special cases of the DOIPEx distribution. The quantile function, hazard and reversed hazard functions, and order statistics of the proposed distribution have been discussed. Additionally, the parameter estimation is carried out via the maximum likelihood method. Finally, three real data sets with different features, namely: underdispersion, equidispersion, and overdispersion have been considered for the goodness of fit test. The results show that the DOIPEx distribution provides the best fit compared to the Poisson and the three special sub-models, and we expect this distribution can serve as an alternative approach to modeling count data.

Acknowledgements

The authors are thankful to reviewers for their valuable suggestions to improve our paper.

Declaration of conflicting interests

The authors declared that they have no conflicts of interest in the research, authorship, and this article's publication.

References

1. Al-Babtain AA, Ahmed AHN, Afify AZ. A new discrete analog of the continuous Lindley distribution, with reliability applications. Entropy. 2020;22(6):603.
2. El-Morshedy M, Eliwa MS, Nagy H. A new two-parameter exponentiated discrete Lindley distribution: properties, estimation and applications. J. Appl. Stat. 2020;47(2):354-75.
3. Sangpoom S, Bodhisuwan W. The discrete asymmetric Laplace distribution. J. Stat. Theory. Pract. 2016;10(1):73-86.
4. Alamatsaz MH, Dey S, Dey T, Harandi SS. Discrete generalized rayleigh distribution. Pak. J. Stat. 2016;32(1):1-20.
5. Para BA, Jan TR. On discrete three parameter burr Type XII and discrete lomax distributions and their applications to model count data from medical science. Biom. Biostat. Int. J. 2016;4(2):70-82.
6. Jayakumar K, Babu MG. Discrete additive Weibull geometric distribution. J. Stat. Theory Appl. 2019;18(1):33-45.
7. Prieto F, Gómez-Déniz E, Sarabia JM. Modelling road accident blackspots data with the discrete generalized Pareto distribution. Accident Anal. Prev. 2014;71:38-49.
8. Nekoukhou V, Bidram H. The exponentiated discrete Weibull distribution. Sort-stat. Oper. Res. T. 2015;39(1):127-46.

9. Aldahlan MA, Afify AZ. A new three-parameter exponential distribution with applications in reliability and engineering. J. Nonlinear Sci. Appl. 2020;13(5):258-69.
10. Roy D. The discrete normal distribution. Commun. Stat.-Theor. M. 2003;32(10): 1871-83.
11. Bracquemond C, Gaudoin O. A survey on discrete lifetime distributions. Int. J. Reliab. Qual. Saf. Eng. 2003;10(1):69-98.
12. Sato H, Ikota M, Sugimoto A, Masuda H. A new defect distribution meteorology with a consistent discrete exponential formula and its applications. IEEE T. Semiconduct. M. 1998;12(4):409-18.
13. R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria. 2023.
14. Wimmer G, Köhler R, Grotjahn R, Altmann G. Toward a theory of word length distributions. J. Quant. Linguist. 1994;1(1):98-106.
15. Douglas JB. Analysis with standard contagious distributions. International Co-operative Publishing House, Fairland MD. 1980.
16. Flynn M, Francis LA. More flexible GLMs zero-inflated models and hybrid models. Casualty Actuarial Society E-Forum, Winter. 2009;148-224.

