AN INFERENTIAL STUDY OF DISCRETE BURR-HATKE EXPONENTIAL DISTRIBUTION UNDER COMPLETE AND CENSORED DATA

ARVIND PANDEY

RAVINDRA PRATAP SINGH*

ABHISHEK TYAGI

•

Department of Statistics, Central University of Rajasthan, Rajasthan-305817, India
Department of Statistics, Chaudhary Charan Singh University, Meerut-250004, India
arvindmzu@gmail.com
stats.rpsingh@gmail.com*
abhishektyagi033@gmail.com

Abstract

In this article, a new one-parameter discrete distribution called discrete Burr-Hatke exponential distribution is introduced and its mathematical characteristics are thoroughly investigated. The proposed distribution is capable of modelling over-dispersed, positively skewed, decreasing failure rate, and randomly right-censored data. We have also introduced many statistical properties including moments, skewness, kurtosis, mean residual life and mean past lifetime, index of dispersion, coefficient of variation, stress strength parameter, quantile function, and order statistics. Method of maximum likelihood is used to estimate unknown model's parameter under complete and censored data. In addition, a technique for generating randomly right-censored data from the proposed model is provided. To evaluate the behaviour of the estimator with complete and censored data, two simulation studies are presented. Two complete and two censored datasets from various disciplines are studied to demonstrate the significance of the suggested distribution in comparison to the existing discrete probability distributions.

Keywords: Burr-Hatke exponential distribution, Method of maximum likelihood, Discrete distribution, Random censoring, Simulation study

1. Introduction

Many continuous lifetime models have been proposed and investigated in reliability theory. However, measuring the life of a component on a continuous scale is frequently impossible or inconvenient. For example, in reliability engineering, the lifetime of an on/off switching device, in survival analysis, the survival times for those suffering from diseases such as lung cancer or the period from remission to relapse may be recorded as the number of days/weeks etc. Furthermore, the count phenomenon arises in a wide range of practical scenarios, including the number of earthquakes that occur in a calendar year, the number of absences, the number of accidents, the number of species kinds in ecology, the number of insurance claims, the number of deaths/daily cases due to the COVID-19 pandemic observed over a specified duration and so on. In all of these circumstances, it is more appropriate to measure these characteristics on a discrete scale rather than a continuous analogue.
Although there are several conventional discrete distributions such as the Binomial, Poisson, Geometric etc and recently developed discrete models to analyse above discussed characteristics. The research for new discrete distributions that are appropriate under various scenarios is still

underway. One prominent area of study in this field is the development of discrete distributions by discretizing suitable continuous probability distributions. Discretization of continuous distribution can be accomplished by a variety of methods. Out of which one of the most widely used methods is [1]. In this approach, he proposed discrete normal distribution using the survival function of its continuous counterpart. Chakraborty in [2] named this technique the survival discretization method. One of the most important advantages of this method is that the produced discrete distribution has the same functional form of the survival function as its continuous version. As a result of this feature, many of the reliability characteristics of the distribution remain unchanged. According to this methodology, for a given continuous random variable (RV) $'Y'$ with survival function (SF) $S_Y(y) = P(Y \geq y)$, the random variable $X = [Y] =$ largest integer less than or equal to Y will have the probability mass function (PMF),

$$\begin{aligned} \mathrm{P}(X = x) &= P(x \leq Y \leq x + 1) \\ &= P\left(Y \geq x\right) - P\left(Y \geq x + 1\right) \\ &= S_Y(x) - S_Y(x + 1); \ x = 0, 1, 2, ... \end{aligned} \tag{1}$$

Many scholars have discretized various well-known continuous distributions using this approach. For instance, [3] investigated the discrete Rayleigh distribution, [4] researched the discrete Maxwell distribution. In addition, [5] investigated the discrete Burr and discrete Pareto distribution. Discrete inverse Weibull distribution developed by [6] . Discrete-continuous Burr III distribution defined by [7]. For more studies on discrete distribution, one can refer to [8], [9], [10], [11] and the references cited therein. Recently, [12] developed a discrete analogue of the odd Weibull-G family of distributions: properties, classical and Bayesian estimation with applications to count data of the number of new coronavirus cases.

In many circumstances, data collection is restricted by constraints such as time or money, making it hard to obtain the entire dataset. This form of incomplete data is referred to as censored data. Various censoring mechanisms are available in the literature to examine these datasets. One of the greatly applicable censorship is random censoring. This scheme consists of studies in which subjects can be censored at any time during the experiment period. Random censoring can be seen in clinical trials or medical studies where patients do not finish the course of treatment and leave before the endpoint. Randomly censored lifetime data are common in many applications such as medical science, biology, reliability studies, and so on, and must be properly analysed to make correct inferences and appropriate research conclusions. Random censoring has been widely investigated in the literature for continuous models see [13]. The censoring technique has also been studied merely under discrete models, namely [14] and [15]. Recently, [16] developed discrete inverted Nadarajah-Haghighi distribution and estimated its parameters under complete and random right-censored censored data.

The majority of existing discrete models were developed to assess count data and, in most cases, they do not accurately analyse the censored data. These situations motivate us to develop a more appropriate discrete distribution that is not capable only of analysing count data but also well enough for modelling censored data. Therefore, in this article, we have proposed a discrete analogue of the Burr-Hatke exponential model by using approach (1) and named it as discrete Burr-Hatke exponential (DBHE) distribution. Hence the ultimate objectives of developing the DBHE model is as follows, a) To construct a discrete model capable of modelling both complete and censored data, b) To design a discrete model with more flexibility and fewer parameters so that the form of diverse distributional properties can be easily handled, c) Numerous practical studies, such as newly developed engineering systems and infant mortality, have shown decreasing failure rate; consequently, we wish to construct a discrete model with a decreasing failure rate function, d) To develop a model that can fit positively skewed, leptokurtic and over-dispersed real data, e) To produce a discrete model that can provide consistently better fits than other well-known discrete models in the existing statistical literature.

The rest of the article is structured as follows: Section 2 introduces the DBHE distribution. Some significant distributional and survival features are investigated in Section 3. In Section 4, we use the maximum likelihood estimation approach to estimate the parameter of the DBHE distribution with complete data and also present numerical illustrations based on empirical and real-world datasets. Section 5 discusses the maximum likelihood estimator (MLE) for the model's parameter under randomly right-censored data and it also includes the technique for generating censored observations from the proposed model. The numerical examples using randomly right-censored empirical and real data have also been presented in section 5. Section 6 concludes with some final observations.

2. The DBHE distribution

The Burr–Hatke exponential (BHE) distribution was proposed by [17].The probability density function (PDF) and SF of the BHE distribution are given as

$$f\left(y,\theta\right)=\frac{\theta\left(2+\theta y\right)}{\left(1+\theta y\right)^{2}}\exp\left(-\theta y\right);y\geq 0,\theta>0, \tag{2}$$

$$S\left(y,\theta\right)=P\left(Y>y\right)=\frac{\exp\left(-\theta y\right)}{\left(1+\theta y\right)};y\geq 0,\theta>0, \tag{3}$$

respectively. The BHE distribution is rightly skewed with decreasing hazard rate function (HRF). This model is very useful to analyse reliability/medical data which have the pattern of decreasing hazard rate. Since it has been generalized by exponential baseline distribution so it may be regarded as an alternative to the several one-parameter exponential families of distributions. Now, using a methodology (1) the PMF of the DBHE model can be obtained as

$$P_X\left(x,\theta\right)=\left(\frac{1}{\left(1+\theta x\right)}-\frac{\exp(-\theta)}{\left(1+\theta+\theta x\right)}\right)\exp\left(-\theta x\right),x=0,1,2...;\theta>0. \tag{4}$$

The CDF corresponding to Equation (4) is given by,

$$F_X(x,\theta)=1-\frac{\exp\left(-\theta\left(x+1\right)\right)}{\left(1+\theta+\theta x\right)},x=0,1,2,...;\theta>0. \tag{5}$$

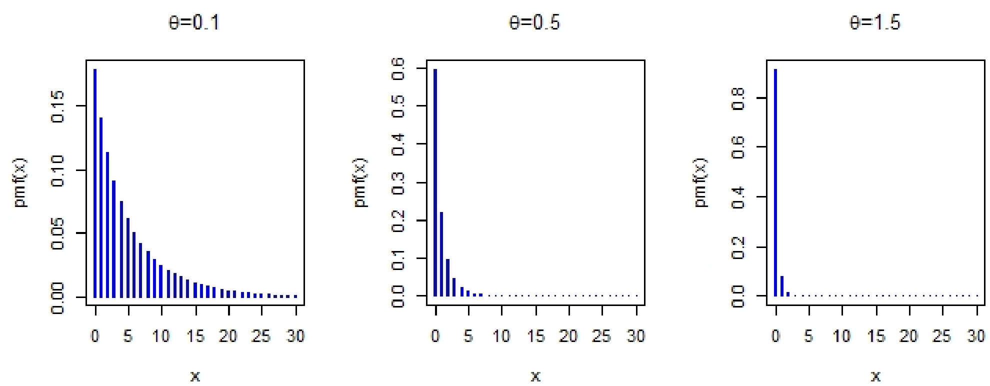

Figure 1: *The PMF plots of the DBHE model for different values of θ.*

Figure 1 shows the PMF plots for different values of the model parameter. From Figure 1, we can conclude that the PMF of the DBHE distribution is unimodal and right-skewed. Also, the behaviour of the PMF at endpoints are as follows:

- $\lim_{x\to 0} P_X(x,\theta)=1-\frac{\exp(-\theta)}{(1+\theta)}$,
- $\lim_{x\to\infty} P_X\left(x,\theta\right)=\lim_{\theta\to 0} P_X\left(x,\theta\right)=\lim_{\theta\to\infty} P_X\left(x,\theta\right)=0.$

3. Distributional Properties

3.1. Recurrence Relation for Probabilities

To obtain the probability mass on various values of X, we can use the following recursive relation

$$P_X\left(x+1,\theta\right)=\left(\frac{1}{(1+\theta+\theta x)}-\frac{\exp\left(-\theta\right)}{(1+2\theta+\theta x)}\right)\left(\frac{1}{(1+\theta x)}-\frac{\exp(-\theta)}{(1+\theta+\theta x)}\right)^{-1}\exp\left(-\theta\right)P_X\left(x,\theta\right).$$

It is observable that $\{P_X\left(x+1\right)\}^2<P_X\left(x\right)P_X\left(x+1\right)$ for all x. As a result, the DBHE distribution is log-convex. Due to this convexity, the proposed distribution has a non-increasing failure rate [18].

3.2. Moments, Skewness and Kurtosis

Moments of a probability distribution are an important tool for measuring its different properties such as mean, variance, skewness, kurtosis, etc. If $F\left(x\right)$ is the CDF of a discrete random variable, then the r^{th} raw moments of this random variable can be obtained by using the following formula:

$$E(X^r)=\sum_{x=0}^{\infty}\left\{\left((x+1)^r-x^r\right)\left(1-F\left(x\right)\right)\right\}.$$

Using the above expression, the r^{th} raw moment denoted by $\mu_r^{'}$ of the DBHE distribution can be written as

$$\mu_r^{'}=E(X^r)=\exp\left(-\theta\right)\sum_{x=0}^{\infty}\frac{\left((x+1)^r-x^r\right)}{(1+\theta+\theta x)}\exp\left(-\theta x\right).\tag{6}$$

Using the ratio test, we can easily observe that, the expression in Equation (6) is convergent. It implies the existence of the r^{th} moment of the proposed distribution.
Now, using Equation (6), the first four-row moments of the DBHE distribution are

$$\mu_1^{'}=E(X)=\exp\left(-\theta\right)\sum_{x=0}^{\infty}\frac{\exp\left(-\theta x\right)}{(1+\theta+\theta x)},\tag{7}$$

$$\mu_2^{'}=E(X^2)=\exp\left(-\theta\right)\sum_{x=0}^{\infty}\frac{(2x+1)}{(1+\theta+\theta x)}\exp\left(-\theta x\right),\tag{8}$$

$$\mu_3^{'}=E(X^3)=\exp\left(-\theta\right)\sum_{x=0}^{\infty}\frac{\left(3x^2+3x+1\right)}{(1+\theta+\theta x)}\exp\left(-\theta x\right),\tag{9}$$

$$\mu_4^{'}=E(X^4)=\exp\left(-\theta\right)\sum_{x=0}^{\infty}\frac{\left(4x^3+6x^2+4x+1\right)}{(1+\theta+\theta x)}\exp\left(-\theta x\right).\tag{10}$$

The variance of the DBHE distribution is given by ,

$$\begin{aligned}Var(X)&=E\left(X^2\right)-E(X)^2\\&=\left(\sum_{x=0}^{\infty}\frac{(2x+1)\exp\left(-\theta x\right)}{(1+\theta+\theta x)}\right)-\left(\exp\left(-\theta\right)\sum_{x=0}^{\infty}\frac{\exp\left(-\theta x\right)}{(1+\theta+\theta x)}\right)^2.\end{aligned}$$

Using above raw moments in (7)-(10), we can easily find the skewness and kurtosis from the following relations

$$K=\frac{E(X^4)-4E(X^2)E\left(X\right)+6E(X^2)(E\left(X\right))^2-3(E\left(X\right))^4}{\left(Var(X)\right)^2}.$$

Table 1 presents some numerical results of the mean, variance, skewness and kurtosis for the DBHE distribution for different values of θ.

Table 1: *Mean, Variance, Skewness and kurtosis for different values of θ.*

Measure ↓ $\theta\rightarrow$	0.1	0.2	0.3	0.5	0.7	0.9	1	1.5	2
Mean	4.6575	1.8326	0.9674	0.3692	0.1701	0.0863	0.0629	0.0149	0.0041
Variance	52.5614	13.4619	5.7640	1.7577	0.7144	0.3336	0.2356	0.0505	0.0130
Skewness	4.8141	5.4455	6.9073	11.9474	20.6287	34.9582	45.2208	156.6311	517.1785
Kurtosis	10.7294	10.8777	12.1065	17.1201	26.0157	40.5689	50.8788	160.0551	504.2940

From Table 1, it is clear that:

1. As the parameter's value increases, the values of mean and variance of the DBHE distribution decrease, whereas the values of skewness and kurtosis increase.
2. The proposed model is appropriate for modelling positively skewed and leptokurtic data.

3.3. Index of Dispersion and Coefficient of Variation

The index of dispersion (IOD) is a measure used to determine the possibility of over-dispersion (under-dispersion) of the model under study. An IOD greater than one indicates over-dispersion, whereas an IOD lower than one indicates under-dispersion. Equi-dispersion is indicated when the IOD is equal to one. The expression for IOD of the DBHE distribution is

$$IOD\left(X\right)=\frac{Var(X)}{E\left(X\right)}=\frac{\left(\sum\limits_{x=0}^{\infty}\frac{(2x+1)\exp(-\theta x)}{(1+\theta+\theta x)}\right)-\left(\exp\left(-\theta\right)\sum\limits_{x=0}^{\infty}\frac{\exp(-\theta x)}{(1+\theta+\theta x)}\right)^{2}}{\exp\left(-\theta\right)\sum\limits_{x=0}^{\infty}\frac{\exp(-\theta x)}{(1+\theta+\theta x)}}. \tag{11}$$

Furthermore, the coefficient of variation (COV) is a measure of data variability. The COV measure is commonly used to compare the variability of independent samples. The larger the coefficient of variation (COV), the more erratic the data. If X follows DBHE model, the COV of DBHE may be represented as

$$COV\left(X\right)=\frac{\left(Var(X)\right)^{1/2}}{E\left(X\right)}=\frac{\left(\left(\exp\left(-\theta\right)\sum\limits_{x=0}^{\infty}\frac{(2x+1)\exp(-\theta x)}{(1+\theta+\theta x)}\right)-\left(\exp\left(-\theta\right)\sum\limits_{x=0}^{\infty}\frac{\exp(-\theta x)}{(1+\theta+\theta x)}\right)^{2}\right)^{1/2}}{\exp\left(-\theta\right)\sum\limits_{x=0}^{\infty}\frac{\exp(-\theta x)}{(1+\theta+\theta x)}}. \tag{12}$$

The numerical values of IOD and COV are shown in Table 2 for a variety of model parameter values.

Table 2: *Index of dispersion and coefficient of variation of DBHE for different values of θ.*

Measure ↓ $\theta\rightarrow$	0.1	0.2	0.3	0.5	0.7	0.9	1	1.5	2
IOD	11.2853	7.3457	5.9583	4.7613	4.1991	3.8674	3.7481	3.3862	3.2143
COV	1.5566	2.0021	2.4818	3.5913	4.9680	6.6959	7.7212	15.0746	28.1402

From Table 2, it is observable that, when the parameter's value increases, the IOD decreases and the COV increases. Since, IOD>1 indicating that the proposed model is appropriate for modelling over-dispersed data.

3.4. Quantile Function

The point x_q is known as the q^{th} quantile of a discrete random variable X if it satisfies $P\left(X\leq x_q\right)\geq q$ and $P\left(X>x_q\right)>1-q$ that is $F\left(x_q-1\right)<q\leq F\left(x_q\right)$ (See, [19]).

Using this result, the q^{th} quantile of DBHE distribution can be obtained by

$$x_q = \left\lceil \frac{1}{\theta} \left\{ \log \left(\frac{1}{(1+\theta x_q)} - \frac{\exp(-\theta)}{(1+\theta+\theta x_q)} \right) - \log q \right\} \right\rceil, \tag{13}$$

where $\lceil . \rceil$ is the ceiling function that returns the smallest integer greater than or equal to its argument.
A random number (integer) can be easily sampled from the proposed distribution by using Equation (13) when q be a uniform random number drawn from a Uniform distribution on the unit interval, i.e. U(0,1). In particular, if we put $q = 0.5$, we will get the value of the median of the proposed distribution.

3.5. Order Statistics

Order statistics have several applications in reliability engineering and life testing. Let $X_1, X_2, ..., X_n$ be a random sample from DBHE distribution. Also, let $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$, denote the corresponding order statistics. Then, the CDF of r^{th} order statistic, say, $Z = X_{(r)}$, is given by

$$\begin{aligned} F_r(z,\theta) &= \sum_{i=r}^{n} \binom{n}{j} F^i(z) \left[1 - F(z,\theta)\right]^{n-i} \\ &= \sum_{i=1}^{r} \sum_{k=0}^{n-i} (-1)^k \binom{n}{i} \binom{n-i}{k} \left\{ 1 - \frac{\exp(-\theta(z+1))}{(1+\theta+\theta z)} \right\}^{(i+k)}. \end{aligned} \tag{14}$$

The corresponding PMF of r^{th} order statistic is

$$\begin{aligned} f_r(z) &= F_r(z) - F_r(z-1) \\ &= \sum_{i=1}^{r} \sum_{k=0}^{n-i} (-1)^k \binom{n}{i} \binom{n-i}{k} \left[\left\{ 1 - \frac{\exp(-\theta(z+1))}{(1+\theta+\theta z)} \right\}^{(i+k)} - \left\{ 1 - \frac{\exp(-\theta z)}{(1+\theta z)} \right\}^{(i+k)} \right]. \end{aligned} \tag{15}$$

Particularly, by putting $r = 1$ and $r = n$ in Equation (15), we can obtain the PMF of minimum$\left(X_{(1)}, X_{(2)}, ..., X_{(n)}\right)$ and the PMF of maximum$\left(X_{(1)}, X_{(2)}, ..., X_{(n)}\right)$, respectively.

3.6. Survival Characteristics

The Survival function of the proposed distribution is

$$S(x,\theta) = P(X > x) = \frac{\exp(-\theta x)}{(1+\theta x)} \; ; x = 0, 1, 2, ...$$

The hazard rate is a reliability characteristic that describes the system's failure behaviour over time. The discrete HRF for the DBHE distribution is given by

$$h(x,\theta) = P(X = x | X \geq x) = \frac{P(X = x)}{S(x-1,\theta)} = \frac{(1+\theta+\theta x - \exp(-\theta)(1+\theta x))}{(1+\theta+\theta x)}; x = 0, 1, 2, ..., \tag{16}$$

provided that $S(x-1,\theta) > 0$.

Figure 2 shows the HRF plots of the DBHE distribution for different values of θ. It is noted that the shape of the HRF is decreasing.
The reverse hazard rate function of the DBHE distribution is given by

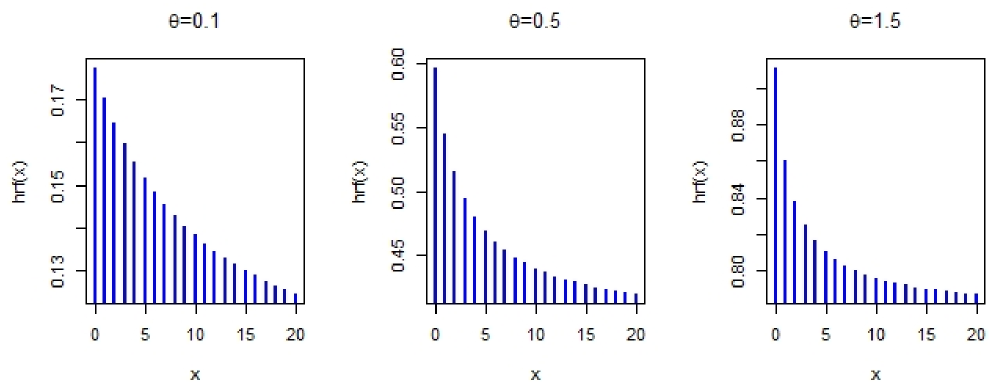

Figure 2: *The HRF plots of the DBHE model for different values of θ.*

$$h^{*}\left(x,\theta\right)=P\left(X=x|X\leq x\right)=\frac{P\left(X=x\right)}{F\left(x,\theta\right)}=\frac{\left(\frac{1}{(1+\theta x)}-\frac{\exp(-\theta)}{(1+\theta+\theta x)}\right)}{\left(1-\frac{\exp(-\theta(x+1))}{(1+\theta+\theta x)}\right)}\exp\left(-\theta x\right). \tag{17}$$

The second rate of failure of the proposed model is given by

$$h^{**}\left(x,\theta\right)=\log\left\{\frac{S\left(x-1\right)}{S\left(x\right)}\right\}=\theta+\log\left(1+\theta+\theta x\right)-\log\left(1+\theta x\right). \tag{18}$$

3.7. Mean Residual and Mean Past Lifetime

The mean residual life (MRL) function, which represents the ageing mechanism, is broadly used in a wide variety of fields, including reliability engineering, survival analysis, biomedical research, and among others. In the literature, it is widely established that the MRL function uniquely characterises the distribution function F since it comprises all of the model's data. In discrete setup, the MRL, represented by the symbol $m(i)$, may be defined as follows:

$$m(i)=E(Y-i|Y\geq i)=\frac{1}{S(i)}\sum_{j=i+1}^{\infty}S(j);\ i=0,1,2,...,$$

where $S\left(.\right)$ is SF. If X has DBHE distribution with parameter θ, then the MRL function of X is

$$m(i)=\frac{(1+\theta i)}{\exp(-\theta i)}\sum_{j=i+1}^{\infty}\frac{\exp(-\theta j)}{(1+\theta j)}.$$

A function is known as the mean past life (MPL) function or expected inactivity time function (EITF) denoted by $m^{*}(i)$, is used to estimate the amount of time since the failure of X if the system has failed at some point before 'i'. In a discrete setting, the MPL function can be defined as

$$m^{*}(i)=E(i-X|X<i)=\frac{1}{F(i-1)}\sum_{k=1}^{i}F(k-1);\ i=1,2,....$$

By replacing the CDF (5) in the expression of $m^{*}(i)$, we can easily obtain the MPL for the proposed model.

3.8. Stress–Strength Parameter

Stress–strength analysis has been extensively used in reliability modelling. Suppose the random variable X and Y denotes the strength and stress of a system (both X and Y are in the positive domain), respectively, then the stress strength reliability $R = P[X > Y]$ can be defined as

$$R = P[X > Y] = \sum_{x=0}^{\infty} P_X(x) F_Y(x),$$

where $P_X(x)$ and $F_Y(x)$ respectively, denote the PMF and CDF of the independent discrete random variables X and Y. Let $X \sim DBHE(\theta_1)$ and $Y \sim DBHE(\theta_2)$, then R of the DBHE is,

$$R = \sum_{x=0}^{\infty} \left\{ \left(1 - \frac{\exp(-\theta_1(x+1))}{(1+\theta_1+\theta_1 x)}\right) \left(\frac{1}{(1+\theta_2 x)} - \frac{\exp(-\theta_2)}{(1+\theta_2+\theta_2 x)}\right) \exp(-\theta_2 x) \right\}. \quad (19)$$

Since, it is difficult to obtain the expression of R in explicit form therefore we perform a numerical analysis of R for different values of θ_1 and θ_2. The numerical outputs of R are presented in Table 3.

Table 3: *The numerical values of R for different combinations of θ_1 and θ_2.*

$\theta_1 \downarrow \theta_2 \rightarrow$	0.05	0.1	0.25	0.5	1	2	5
0.05	0.51381	0.35905	0.18823	0.09638	0.03710	0.00830	0.00020
0.1	0.66849	0.51372	0.30089	0.16324	0.06501	0.01478	0.00036
0.25	0.81902	0.69564	0.46859	0.27722	0.11696	0.02740	0.00067
0.5	0.87830	0.77949	0.56485	0.35267	0.15502	0.03718	0.00092
1	0.90091	0.81440	0.61111	0.39304	0.17724	0.04320	0.00107
2	0.90559	0.82202	0.62219	0.40351	0.18342	0.04496	0.00112
5	0.90593	0.82258	0.62304	0.40435	0.18394	0.04511	0.00112

From this table, we observe that for any fixed value of θ_1, R decreases as θ_2 increases, whereas for a fixed value of θ_2, as θ_1 increases, the value of R also increases.

4. Analysis of complete data under DBHE distribution

In this section, we estimate the unknown parameter of the DBHE distribution using the MLE method. An algorithm for generating random data is presented. We also present numerical examples based on empirical and real-world datasets to demonstrate the utility of the proposed approach for evaluating complete data.

4.1. Maximum Likelihood Estimation with Complete Data

Suppose $\underline{x} = (x_1, x_2, ..., x_n)$ be a random sample from DBHE distribution then the log-likelihood function can be written as

$$\log L(\underline{x}; \theta) = -\theta \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \log\left(\frac{1}{(1+\theta x_i)} - \frac{\exp(-\theta)}{(1+\theta+\theta x_i)}\right). \quad (20)$$

By differentiating Equation (22) with respect to the parameter θ, we get the non-linear likelihood equation as follows

$$\sum_{i=1}^{n} \left[\left(\frac{\exp(-\theta)}{(1+\theta+\theta x_i)}\right) \left(\frac{1+x_i}{(1+\theta+\theta x_i)} + 1\right) - \frac{x_i}{(1+\theta x_i)^2} \right] \left[\frac{1}{(1+\theta x_i)} - \frac{\exp(-\theta)}{(1+\theta+\theta x_i)} \right]^{-1} - \sum_{i=1}^{n} x_i = 0. \quad (21)$$

The solution of Equation (21) gives the MLE of θ. However, there is no explicit form for the solution of Equation (21). Therefore, Equation (21) has to be solved by using iterative methods such as Newton-Raphson, Nelder-Mead etc.

4.2. Numerical Illustration Using Simulated Data

In this subsection, we perform a Monte Carlo simulation study to show how well the MLE can estimate the unknown parameter of the DBHE distribution. Therefore, we conduct a simulation study with replication number 1,000. The true parameter values are used as $\theta = 0.05$, $\theta = 0.25$, and $\theta = 0.5$. There is no stated reason for using these parameter values. It may be used in several different ways. Random samples from the DBHE distribution are generated with $n = 15, 20, 25, ..., 100$ sample sizes using Equation (13). The simulation results are interpreted based on the mean square errors (MSEs) and absolute biases (ABs) where

$$MSE = \frac{1}{1000}\sum_{j=1}^{1000}\left(\hat{\theta}^j - \theta\right)^2 \text{and } AB = \frac{1}{1000}\sum_{j=1}^{1000}\left|\hat{\theta}^j - \theta\right|,$$

here, $\hat{\theta}$ is an estimate of θ.
The simulation results are graphically summarized and displayed in Figure 3.

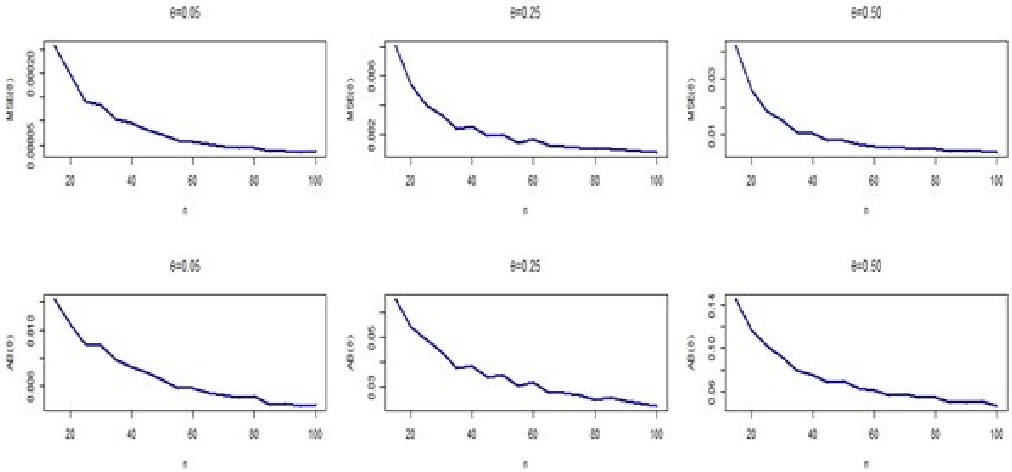

Figure 3: *Plots for MSEs and ABs for different values of θ for complete data.*

Figure 3 illustrates that the MSEs of the MLEs tend to zero as n approaches infinity. This demonstrates the consistency of the estimator. Furthermore, when n increases, the ABs is also declined to zero.

4.3. Real Data Analysis

In this section, we illustrate the utility of the DBHE distribution by examining two real-world datasets. Several criteria are used to compare fitted models, including the -logL, the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the Hannan Quinn information criterion (HQIC), and the Chi-square (χ^2) statistic with its associated P-value. The descriptive summaries of the datasets are shown in Table 4. From this table, we can see that the IOD for all datasets is greater than 1, indicating that the considered datasets can only be modelled by discrete distributions with overdispersion phenomena. The comparing models to DBHE distribution are listed in Table 5.

Table 4: *Descriptive Statistics of the Datasets.*

Data	n	Mean	Variance	Skewness	Kurtosis	IOD	COV
Dataset I	100	0.67	1.1526	2.4697	4.532	1.7203	1.6024
Dataset II	400	0.5475	1.1256	9.7478	15.6829	2.0558	1.9378

Table 5: *The competitive models of the DBHE distribution.*

Distribution	Abbreviation	Parameter(s)	Author(s)
Geometric	Geo	θ	-
Discrete Lindley	DLi	λ	[20]
Discrete Lindley-Two Parameter	DLi-II	p, β	[21]
Discrete Pareto	DPa	β	[5]
Discrete linear failure rate	DLFR	λ_1,λ_2	[22]
Discrete inverse Weibull	DIW	α,β	[6]
Discrete log-logistic	DLogL	δ,λ	[23]
Discrete Nielsen	DN	p, θ	[24]
Negative Binomial	NB	μ,Θ	-
Zero-Inflated Negative Binomial	ZINB	μ, Θ, ω	-
Poisson- Lindley	PL	θ	[25]
Generalized Poisson-Lindley	GPL	θ, α	[26]

Dataset I: The first dataset, consists of the recordings of the total number of carious teeth among the four deciduous molars in a sample of 100 children 10 and 11 years old [5]. The expected frequency of the fitted models along with their MLE, standard error (SE), -logL, and goodness of fit measures are presented in Table 6. Since, the values of -logL, χ^2 test statistic, AIC, BIC, CAIC, and HQIC of DBHE distribution are smallest among those of other considered models, hence this new distribution appears to be a very suitable model for this dataset. Similarly, the higher P-value corresponding to χ^2 statistic for DBHE distribution show its dominance on other candidate models in terms of model fitting.

Table 6: *The MLE (SEs) and goodness of fit statistics for different models under dataset I.*

X	Observed Frequency	DBHE	Geo	DLi	DLi-II	DPa	DLFR	DIW	DLogL
0	64	62.80	59.88	57.13	59.88	69.04	59.9	63.3	62.73
1	17	21.37	24.02	26.88	24.02	15.37	24.01	22.48	22.42
2	10	8.60	9.64	10.45	9.64	6.01	9.63	6.44	7.01
3	6	3.78	3.87	3.71	3.87	3.01	3.86	2.76	2.98
>=4	3	3.45	2.59	1.83	2.59	6.57	2.6	5.02	4.86
Total	100	100	100	100	100	100	100	100	100
MLE (SE)		0.55043 (0.064)	0.59879, (0.038)	0.274 (0.029)	0.401 (0.269), 0.478 (0.529)	0.184 (0.032)	0.401 (0.056), 1.0 (0.044)	0.633 (0.049), 1.576 (0.251)	0.745 (0.101), 1.768 (0.267)
$-\log$ L		112.328	112.474	113.68	112.475	116.83	112.470	116.275	115.470
χ^2		1.575	3.347	6.638	3.347	3.225	3.340	3.503	2.783
D.F.		2	2	2	1	2	1	1	1
P-value		0.455	0.188	0.036	0.067	0.199	0.068	0.061	0.095
AIC		226.656	226.947	229.36	228.950	235.66	228.940	236.550	234.940
BIC		229.261	229.552	232.96	234.160	238.27	234.150	241.760	240.150
CAIC		226.697	226.988	229.39	229.073	235.70	229.063	236.673	235.063
HQIC		227.710	228.001	230.41	231.058	236.72	231.048	238.658	237.048

Dataset II: The second dataset represents the number of chromatid aberrations in 24 hours [28]. The expected frequency of the fitted models along with their MLE, SE, -logL, and goodness of fit measures are presented in Table 7. On comparison of the values of -logL, χ^2 test statistic, P-value, AIC, BIC, CAIC, and HQIC, we again found that the DBHE distribution is the best model than the other five models understudy for this dataset.

Table 7: *The MLE (SEs) and Goodness of fit statistics for different models under dataset II.*

X	Observed Frequency	DBHE	DN	NB	ZINB	PL	GPL
0	268	269.36	270.14	270.18	270.18	257.02	269.24
1	87	80.48	79.40	78.55	78.55	93.39	78.70
2	26	29.28	29.21	29.84	29.84	32.76	30.86
3	9	11.76	11.88	12.22	12.22	11.21	12.55
4	4	5.01	5.11	5.19	5.19	3.77	5.13
5	2	2.22	2.28	2.25	2.25	1.25	2.09
6	1	0.90	1.05	0.99	0.99	0.41	0.85
7	3	0.47	0.93	0.78	0.78	0.13	0.35
Total	400	400	400	400	400	400	400
MLE (SEs)		0.63026 (0.037)	0.5301 (0.0601), 1.1089 (0.2179)	0.5475 (011539), 0.6200 (0.1270)	0.5475 (0.1701), 0.6200 (0.3383), 0.00008 (0.2989)	2.379 (0.169)	1.576 (0.259), 0.473 (0.159)
− log L		399.342	399.410	399.860	399.860	399.857	400.553
χ_2		1.781	1.924	2.416	2.416	6.283	2.940
D.F.		3	2	2	1	3	2
P-value		0.619	0.382	0.299	0.120	0.098	0.229
AIC		800.683	802.820	803.720	805.720	801.714	805.106
BIC		804.675	810.803	811.703	817.694	805.706	813.089
CAIC		800.693	802.850	803.750	805.781	801.724	805.136
HQIC		802.264	805.981	806.881	810.462	803.295	808.267

5. Analysis of randomly censored data under DBHE distribution

In this section, we derive the MLE of the unknown parameter of the DBHE distribution for random rightly-censored data. For the DBHE model, an algorithm for generating random right-censored data is presented. We also present numerical examples based on empirical and real-world datasets to show the usefulness of the proposed approach for evaluating random censored data.

5.1. Maximum Likelihood Estimation with Randomly Censored Data

Due to the availability of right-censored observations, the contribution of the i^{th} individual for the likelihood function based on a random sample (x_i, d_i) of size n is given by

$$L_i = [f(x_i)]^{d_i} [S(x_i)]^{1-d_i},$$

where d_i is a censoring indicator variable, that is, $d_i = 1$ for an observed lifetime and $d_i = 0$ for a censored lifetime $(i = 1, 2, 3, ..., n)$. Assuming the DBHE model, the likelihood function for θ is given by

$$L(\theta|\underline{x}, \underline{d}) = \prod_{i=1}^{n} \left\{ \left(\frac{1}{(1+\theta x_i)} - \frac{\exp(-\theta)}{(1+\theta+\theta x_i)} \right) \exp(-\theta x_i) \right\}^{d_i} \left\{ \frac{\exp(-\theta x_i)}{(1+\theta x_i)} \right\}^{1-d_i}, \tag{22}$$

where $\underline{d} = (d_1, d_2, ..., d_n)$. The corresponding log-likelihood function is

$$\log L(\theta|\underline{x}, \underline{d}) = \sum_{i=1}^{n} d_i \log \left\{ \frac{1}{(1+\theta x_i)} - \frac{\exp(-\theta)}{(1+\theta+\theta x_i)} \right\} + \sum_{i=1}^{n} (d_i - 1) \log(1+\theta x_i) - \theta \sum_{i=1}^{n} x_i. \tag{23}$$

Taking the first derivative of Equation (23) w.r.t. θ and setting this derivative equal to zero, we can obtain the likelihood equation for the parameter θ. Although, it is hard to find a closed-form expression of MLE for the parameter θ using this likelihood equation, therefore, we can use an appropriate numerical methodology such as the Newton-Raphson iteration method to obtain the MLE of θ.

5.2. Algorithm to Simulate Random Right-Censored Data

We present a simple approach in this part for generating random right-censored data from the suggested model. The algorithm is as follows:

Step 1: Fix the values of the parameter θ.

Step 2: Draw n random pseudo from $Uniform(0,1)$ i.e. $u_i \sim U(0,1); i = 1,2,...,n$.

Step 3: Obtain $x_i^{/} = F^{-1}(u_i;\theta); i = 1,2,...,n$, where $F^{-1}(\bullet)$ is defined in Equation (13).

Step 4: Draw n random pseudo from $c_i \sim U(0, \max(x_i^{/})); i = 1,2,...,n$. This is the distribution that controls the censorship mechanism.

Step 5: If $x_i^{/} \leq c_i$, then $x_i = [x_i^{/}]$ and $d_i = 1, i = 1,2,...,n$, else, $x_i = [c_i]$ and $d_i = 0, i = 1,2,...,n$. Hence, pairs of values $(x_1, d_1), (x_2, d_2), ..., (x_n, d_n)$ are obtained as the random right-censored data.

5.3. Numerical Illustration Using Simulated Random Right-Censored Data

This subsection portrays a simulation study to evaluate the performance of the MLE using randomly right-censored data. The whole study is based on randomly chosen samples from the DBHE distribution of sizes 20, 25, ...,100. The values of θ are set to 0.05, 0.25, and 0.50. The procedure described above is used to generate the requisite random right-censored data. All simulation findings are based on 1000 replications for different settings of parameter values and sample sizes. Based on these 1000 values, we estimated the MSE and AB of the parameter estimate, and the resultant graphs are given in Figure 4.

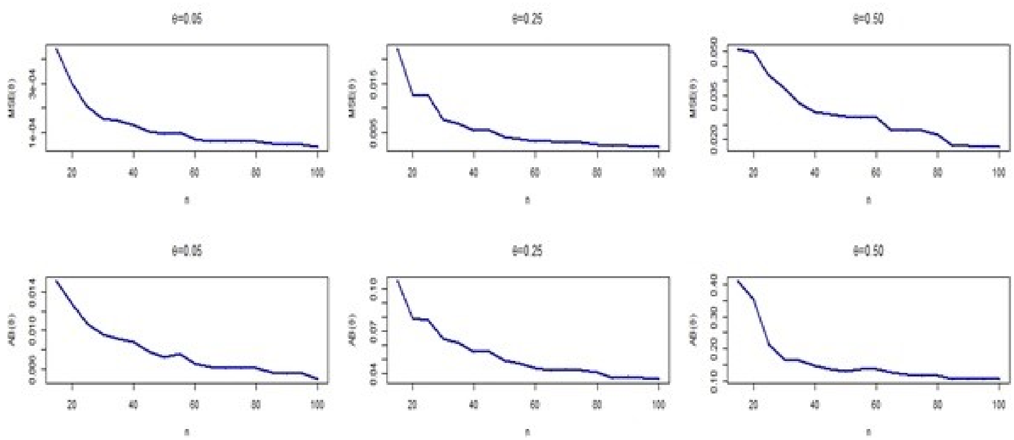

Figure 4: *Plots for MSEs and ABs for different values of θ under censored data.*

As seen in Figure 4, the MSEs of the MLE approach θ as n approaches infinity. This illustrates the estimator's consistency. Additionally, when n increases, the ABs is also tending to zero.

5.4. Application to Real Data Analysis

Here, we examine two real datasets to illustrate the applicability of the DBHE model to randomly censored data. The following datasets and their fitting are described as follows:
Dataset III: This dataset is obtained from [29]. The data below are remission times, in weeks, for a group of 30 patients with leukaemia who received similar treatment.
1, 1, 2, 4, 4, 6, 6, 6, 7, 8, 9, 9, 10, 12, 13, 14, 18, 19, 24, 26, 29, 31*, 42, 45*, 50*, 57, 60, 71*, 85*, 91.
The observations with asterisks indicate censored times. The MLE (SE) of the θ for the given dataset is 0.0201 (0.0008). Now, we have been used Kolmogorov-Smirnov (K-S) test to check whether the given data follows DBHE distribution or not. The calculated value of the K-S test is 0.13333 and P-value is equal to 0.9525. These values announce that the DBHE distribution can be used to model this data.

Dataset IV: Here, we analyze another real dataset obtained from [29]. The data below show survival times (in months) of patients with Hodgkin's disease who were treated with nitrogen mustards.
1.05, 2.92, 3.61, 4.20, 4.49, 6.72, 7.31, 9.08, 9.11, 14.49*, 16.85, 18.82*, 26.59*, 30.26*, 41.34*.
The asterisks observations represent censored times. For the provided dataset, the MLE (SE) of the θ is 0.0311 (0.0027). We have also performed the K-S test to see whether the data distribution fits the DBHE distribution or not, and it is found that the K-S test has a value of 0.2 and a P-value of 0.9383. So, it can be seen that the DBHE distribution fits the data very well.

6. Conclusions

In this paper, we have proposed discrete Burr-Hatke exponential distribution. It is observed that with one parameter, this model has great flexibility in terms of fitting as it is capable of modelling right-skewed, decreasing failure rate, and over-dispersed counts datasets. Some of its fundamental properties have been discussed in detail. The unknown parameter of the DBHE distribution with complete and censored data has been estimated by using the maximum likelihood approach. We have provided an algorithm to generate randomly right-censored data. Additionally, the performance of the estimator under complete and censored data have been examined through an extensive simulation study. Finally, the flexibility of the DBHE distribution has been empirically proven by using four real-life applications consisting of two complete and two censored datasets. Hence, we can conclude that the proposed model will serve a wide spectrum of applications in various domains such as medical, reliability, survival analysis, etc.

Acknowledgement

The authors would like to express their gratitude to the editor for their care of the paper and giving us the the opportunity to make the paper better.

References

[1] Roy, D. (2003). The discrete normal distribution. *Commun. Statist. Theor. Meth.* 32(10):1871–1883.
[2] Chakraborty, S. (2015). Generating discrete analogues of continuous probability distributions-A survey of methods and constructions. *Journal of Statistical Distributions and Applications,* 2(1), 6.
[3] Roy, D. (2004). Discrete Rayleigh distribution. *IEEE Trans. Reliab.* 53:255–260.
[4] Krishna, H., & Pundir, P. S. (2007). Discrete maxwell distribution. *InterStat,* 3.
[5] Krishna, H., & Pundir, P. S. (2009). Discrete Burr and discrete Pareto distributions. *Statistical Methodology,* 6(2), 177-188.

[6] Jazi, M. A., Lai, C. D., & Alamatsaz, M. H. (2010). A discrete inverse Weibull distribution and estimation of its parameters. *Statistical Methodology,* 7(2), 121-132.

[7] Al-Huniti, A. A., & AL-Dayian, G. R. (2012). Discrete Burr type III distribution. *American Journal of Mathematics and Statistics,* 2(5), 145-152.

[8] Alamatsaz, M. H., Dey, S., Dey, T., & Harandi, S. S. (2016). Discrete generalized Rayleigh distribution. *Pakistan journal of statistics,* 32(1).

[9] Jayakumar, K., & Babu, M. G. (2018). Discrete Weibull geometric distribution and its properties. *Communications in Statistics-Theory and Methods,* 47(7), 1767-1783.

[10] Tyagi, A., Choudhary, N., and Singh, B. (2019). Discrete additive Perks-Weibull distribution: Properties and applications. *Life Cycle Reliability and Safety Engineering,* 8(3), 183-199.

[11] Tyagi, A., Choudhary, N., & Singh, B. (2020). A new discrete distribution: Theory and applications to discrete failure lifetime and count data. J. Appl. Probab. Statist, 15, 117-143.

[12] El-Morshedy, M., Eliwa, M. S., & Tyagi, A. (2021). A discrete analogue of odd Weibull-G family of distributions: properties, classical and Bayesian estimation with applications to count data. *Journal of Applied Statistics,* 1-25.

[13] Garg, R., Dube, M., & Krishna, H. (2020). Estimation of parameters and reliability characteristics in Lindley distribution using randomly censored data. *Statistics, Optimization & Information Computing,* 8(1), 80-97.

[14] Krishna, H., & Goel, N. (2017). Maximum likelihood and Bayes estimation in randomly censored geometric distribution. *Journal of Probability and Statistics,* 2017.

[15] Achcar, J. A., Martinez, E. Z., de Freitas, B. C. L., & de Oliveira Peres, M. V. (2021). Classical and Bayesian inference approaches for the exponentiated discrete Weibull model with censored data and a cure fraction. *Pakistan Journal of Statistics and Operation Research,* 467-481.

[16] Singh, B., Singh, R. P., Nayal, A. S., & Tyagi, A. (2022). Discrete Inverted Nadarajah-Haghighi Distribution: Properties and Classical Estimation with Application to Complete and Censored data. *Statistics, Optimization & Information Computing,* 10(4), 1293-1313.

[17] Yadav, A. S., Altun, E., & Yousof, H. M. (2019). Burr–Hatke Exponential Distribution: A Decreasing Failure Rate Model Statistical Inference and Applications. *Annals of Data Science,* 1-20.

[18] Gupta, P. L., Gupta, R. C., and Tripathi, R. C. (1997). On the monotonic properties of discrete failure rates. *Journal of Statistical Planning and Inference,* 65(2), 255-268.

[19] Rohatgi, V. K., & Saleh, A. M. E. (2015). An introduction to probability and statistics. *John Wiley & Sons.*

[20] Gomez-Deniz, E., & Calderin-Ojeda, E. (2011). The discrete Lindley distribution: properties and applications. *Journal of Statistical Computation and Simulation,* 81(11), 1405-1416.

[21] Hussain, T., Aslam, M., & Ahmad, M. (2016). A two parameter discrete Lindley distribution. *Revista Colombiana de Estad?stica,* 39(1), 45-61.

[22] Kumar, C., Tripathi, Y. M., & Rastogi, M. K. (2017). On a discrete analogue of linear failure rate distribution. *American journal of mathematical and management sciences,* 36(3), 229-246.

[23] Para, B. A., & Jan, T. R. (2016). Discrete version of log-logistic distribution and its applications in genetics. *Int. J. Mod. Math. Sci,* 14(4), 407-422.

[24] Castellares, F., Lemonte, A. J., & Santos, M. A. (2020). On the Nielsen distribution. *Brazilian Journal of Probability and Statistics, 34(1),* 90-111.

[25] Sankaran, M. (1970). 275. note: The discrete Poisson-Lindley distribution. *Biometrics,* 145-149.

[26] Mahmoudi, E., & Zakerzadeh, H. (2010). Generalized Poisson–Lindley distribution. *Communications in Statistics—Theory and Methods, 39(10),* 1785-1798.

[27] Phyo, I. (1973). Use of a Chain Binomial in the Epidemiology of Caries. *Journal of dental research,* 52(4), 750-752.

[28] Castellares, F., Lemonte, A. J., & Santos, M. A. (2020). On the Nielsen distribution. *Brazilian Journal of Probability and Statistics,* 34(1), 90-111.

[29] Lawless, J. F. (2011). Statistical models and methods for lifetime data (Vol. 362). *John Wiley & Sons.*