AIMS Mathematics, 8(5): 10266–10282.
DOI: 10.3934/math.2023520
Received: 21 October 2022
Revised: 15 February 2023
Accepted: 20 February 2023
Published: 28 February 2023

http://www.aimspress.com/journal/Math

Research article

Discrete Erlang-2 distribution and its application to leukemia and COVID-19

Mohamed Ahmed Mosilhy[1,2,*]

[1] Department of Mathematics, Faculty of Science, Cairo University, Egypt
[2] Department of Statistics, Faculty of Science, University of Tabuk, KSA

* **Correspondence:** Emails: mosilhy@sci.cu.edu.eg, moelsayed@ut.edu.sa; Tel: +201143611101, +966506164520.

Abstract: Via the survival discretization method, this research revealed a novel discrete one-parameter distribution known as the discrete Erlang-2 distribution (DE2). The new distribution has numerous surprising improvements over many conventional discrete distributions, particularly when analyzing excessively dispersed count data. Moments and moments-generating functions, a few descriptive measures (central tendency and dispersion), monotonicity of the probability mass function, and the hazard rate function are just a few of the statistical aspects of the postulated distribution that have been developed. The single parameter of the DE2 distribution was estimated via the maximum likelihood technique. Real-world datasets, leukemia and COVID-19, were applied to analyze the effectiveness of the recommended distribution.

Keywords: leukemia and COVID-19 data; discretization; reliability analysis; Erlang-2 distribution; maximum likelihood estimator; Fano factor
Mathematics Subject Classification: 37M15, 30G25, 44A55

1. Introduction

Many researchers in diverse fields, notably in the field of medical statistics, periodically run into the issue where continuous random variables are not always evaluated on a continuous scale but are usually recorded using discrete random variables. In actual testing experiments like radio-sensitivity and growth kinetics, estimating the lifetime of patients or equipment on a weekly or daily discrete scale is frequently impossible. Survival analysis typically expresses outcomes in terms of days, such as the possibility of tumor treatment following fractionated radiotherapy, the duration of vacation time taken by lung cancer patients while receiving treatment or the interval between remission and relapse.

The geometric and negative binomial distributions are discrete analogues for the exponential and

gamma distributions inside this situation. These discrete distributions contain monotonic hazard rate functions, which makes them inappropriate in a number of circumstances.

The research exposes a large number of discrete lifetime distributions; see [1, 2]. Roy [3] analyzed the notion of discrete concentration, considering it a straightforward method that, given a continuous model, can generate a discrete life distribution model; the discrete normal distribution was developed using this idea. Utilizing the same methodology as for discretizing continuous probability distributions. The discrete Burr and Pareto distributions were derived by Krishna and Pundir in [4]. The discrete Gamma distribution was first defined by Chakraborty and Chakravarty [5]. The second type of generalized exponential discrete was inserted by Alamatsaz et al. [6], and the exponentiated discrete from the Weibull distribution was pressed in [7]. Alamatsaz et al. [8] introduced a two-parameter discrete generalized Rayleigh distribution, Hussain et al. [9] presented a two-parameter discrete Lindley distribution. The discretization of weighted Exponential distribution was emerged in [10] and Jayakumar and Babu [11] devolped a discrete additive Weibull geometric distribution.

Very recently, El-Morshedy et al. [12] introduced the discrete Bilal distribution, El-Morshedy [13] presented the discrete Gompertz distribution, Eldeeb et al. [14] helmed the discrete analog of inverted Topp-Leone distribution, El-Alosey [15] established the discrete Erlang-truncated Exponential distribution distribution, Ahsan-ul-Haq et al. [16] proposed discrete type-II half-logistics exponential distribution and Eledum and El-Alosey, in [17], presented the discrete extended Erlang-truncated exponential distribution.

In this work, we create a distinctive one-parameter discrete distribution of the DE2 distribution by adopting the continuous survival technique in Section 2. We realize some of the basic statistical characteristics of the DE2 distribution in Section 3. Section 4 evaluating the estimation parameter using the maximum likelihood method. Moreover, we validated the suggested distribution using the datasets for leukemia and COVID-19 in Section 5. Finally, we provide some concluding remarks in Section 6.

- Erlang-2 distrebution (E2 distribution)

Erlang created the Erlang distribution to analyze the number of concurrent calls that might be received by the switching station operators. This research on telephone traffic engineering has been broadened to take into account queueing systems' waiting times more generally. The subject of stochastic processes also makes use of the distribution.

The probability density function (pdf) of the Erlang distribution is

$$f_E(x;k,\lambda) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} \quad \text{for } x, \lambda \geq 0. \tag{1.1}$$

The parameter λ is known as the rate parameter, while the value k is referred as the shape parameter. The Erlang distribution's cumulative distribution function (cdf) is

$$F_E(x;k,\lambda) = 1 - \sum_{j=0}^{k-1} \left(\frac{(x\lambda)^j e^{-\lambda x}}{j!} \right). \tag{1.2}$$

The Erlang-2 distribution (E2 distribution) is defined by setting $k = 2$ in Eq (1.1). Moreover, its probability density function (pdf) and cumulative distribution function (cdf) are respectively given by:

$$f_{E2}(x;\lambda) = \lambda^2 x e^{-\lambda x}, \quad \text{for } x, \lambda \geq 0, \tag{1.3}$$

$$F_{E2}(x;\lambda) = 1 - \Big(1 + \lambda x\Big)\ e^{-\lambda x}. \tag{1.4}$$

The survival function of E2 distribution is:

$$S_{E2}(x) = \Big(1 + \lambda x\Big)\ e^{-\lambda x}. \tag{1.5}$$

The hazard rate function (failure rate function) of E2 distribution is:

$$h_{E2}(x) = \frac{\lambda^2\ x}{1 + \lambda x}. \tag{1.6}$$

The reverse hazard rate function of E2 distribution is:

$$h_{r(E2)}(x) = \frac{\lambda^2 x}{e^{\lambda x} - (1 + \lambda x)},\ \ \lambda \neq 0. \tag{1.7}$$

The Mills ratio function of E2 distribution is:

$$Mills\ ratio_{E2}(x) = \frac{e^{\lambda x} - (1 + \lambda x)}{\lambda^2 x},\ \ \lambda \neq 0. \tag{1.8}$$

2. Discrete Erlang-2 distrebution (DE2 distribution)

The surviving discretization method can be used to create a discrete duplicate of any continuous random variable, see [3, 18]. A discrete random variable Y can be defined as equal to $[X]$, which is the floor function of X and is the greatest integer less than or equal to X, given a continuous random variable X with a survival function $S_X(x)$. The probability mass function (pmf) of Y is then provided by

$$f(y) = P_r(Y = y) = S_X(y) - S_X(y + 1).$$

This idea is used to suggest a one-parameter discrete probability distribution by discretizing the parameterized version of the given in Eq (1.3). First re-parameterization of Erlang-2 distribution in Eq (1.3) is done by taking $p = \lambda$, which lead us to the formula of pmf for discrete Erlang-2 distribution (DE2 distribution), as follows:

$$f_{DE2}(y;p) = P_r(Y = y) = e^{-py}\Big((1 - e^{-p})\ (py + 1) - pe^{-p}\Big), \tag{2.1}$$

where p is the rate parameter, $\ p \geq 0\ \ and\ \ y = 0, 1, 2, \ldots$.

Figure 1 shows that the sensitivity of the rate parameter p make a modification in the shape of pmf, at different values of p which has small difference between them, and also it has been noticed that the distribution's mode shifts to the left as the rate parameter rises.

The corresponding cdf of DE2 distribution is

$$F_{DE2}(y;p) = P_r(Y \leq y) = \begin{cases} 1 - e^{-p(y+1)}(1 + p\ (y + 1)), & y \geq 0, \\ 0, & elsewhere. \end{cases} \tag{2.2}$$

Figure 2 shows that the sensitivity of the rate parameter p, will change the shape of cdf, at different values which has small difference between them. On the other hand, it can be observed that the cdf approaches one quickly as p increases, increasing the likelihood of high numbers.

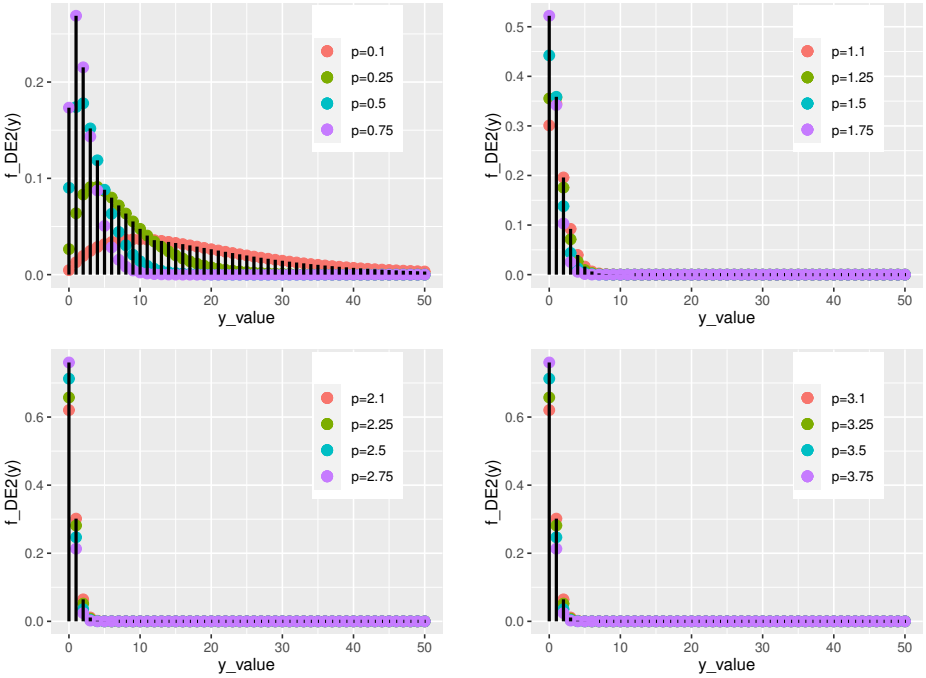

Figure 1. The pmf of DE2 for different values of p.

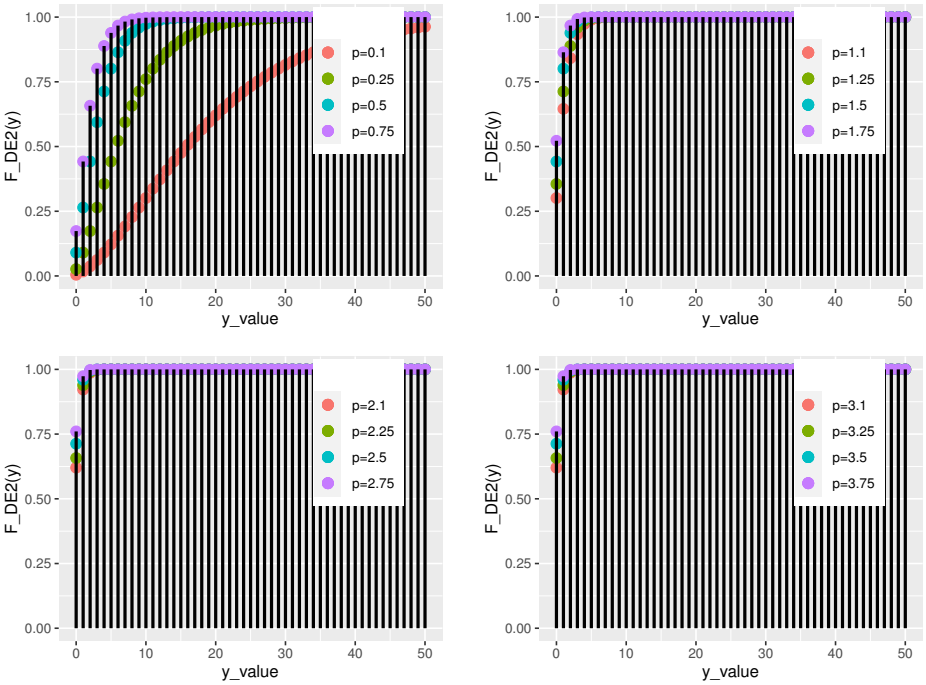

Figure 2. The cdf of DE2 for different values of p.

3. Statistical properties of DE2 distribution

Some of the statistical characteristics of the mentioned distribution, DE2, were inferred in this section.

3.1. Survival, hazard rate, reverse hazard rate and Mills ratio functions of (DE2)

The definitions of the survival and Mills ratio functions, as well as the discrete hazard rate (also known as the discrete failure rate) and reverse hazard rate functions, are covered in this subsection, see [19, 20].

3.1.1. Survival function

The classical definition of the survival function for the DE2 distribution is as follows:

$$\begin{aligned} S_{DE2}(y;p) &= p_r(Y \geq y) = p_r(Y = y) + p_r(Y > y) \\ &= p_r(Y = y) + [1 - p_r(Y \leq y)] \\ &= f_{DE2}(y;p) + [1 - F_{DE2}(y;p)] \\ &= e^{-py}\Big((1 - e^{-p})\,(py + 1) - pe^{-p}\Big) + \Big[1 - \Big(1 - e^{-p(y+1)}(1 + p\,(y + 1))\Big)\Big] \\ &= e^{-py}\,(py + 1), \end{aligned} \tag{3.1}$$

where $p \geq 0$ and $y \in \mathbb{N}$.

From Figure 3, it's interesting to observe that when p rises, the survival function of DE2 falls off at various values of the rate parameter p.

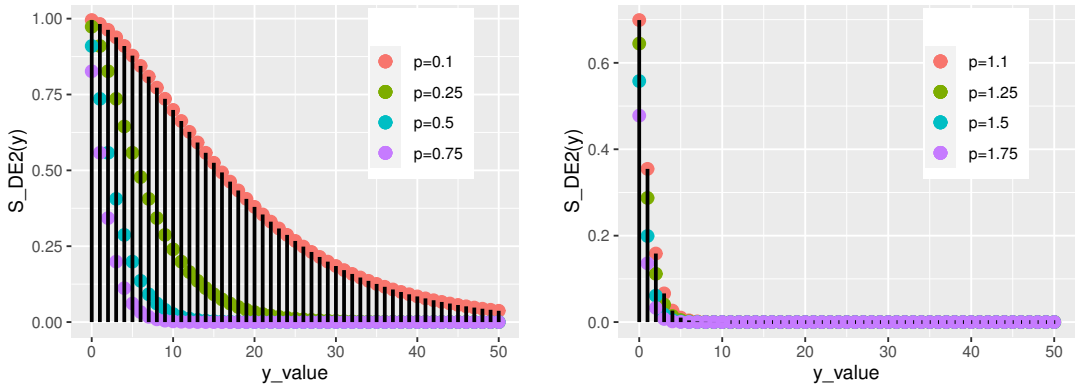

Figure 3. The survival function of DE2 for different values of p.

3.1.2. Hazard rate function

The "hazard rate" (or "failure rate") function of the DE2 distribution is often defined as follows:

$$h_{DE2}(y;p) = \frac{f_{DE2}(y;p)}{S_{DE2}(y;p)} = 1 - \frac{e^{-p}(1 + p(y + 1))}{1 + py}, \tag{3.2}$$

where $p \geq 0$ and $y \in \mathbb{N}$.

From Figure 4, at different values of the rate parameter p, the hazard rate function of DE2 increases as p increases.

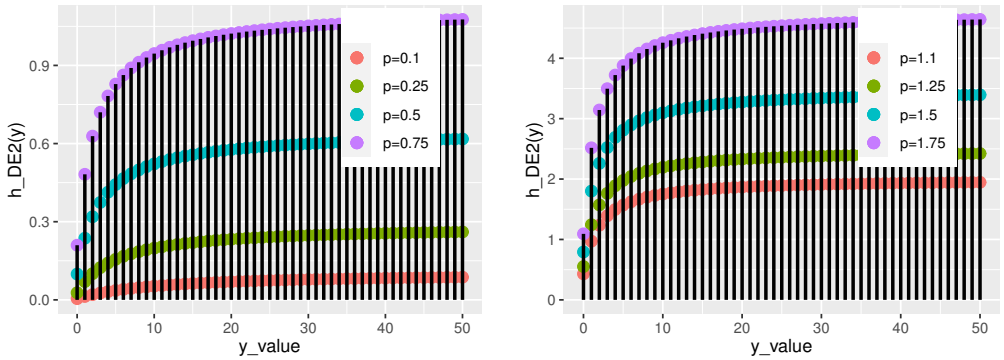

Figure 4. The hazard function of DE2 for different values of p.

3.1.3. The reverse hazard rate function

It is customary to define the "reverse hazard rate" function of the DE2 distribution as follows:

$$h_{r\ DE2}(y;p) = \frac{f_{DE2}(y;p)}{F_{DE2}(y;p)} = \frac{py(e^p-1)+e^p-2}{e^{p(y+1)}+p(y+1)-1}, \quad p \neq 0, \tag{3.3}$$

where $p > 0$ and $y \in \mathbb{N}$.

From Figure 5, we can see that at various values of the ratee parameter p, the reverse hazard function of DE2 reduces as p grows.

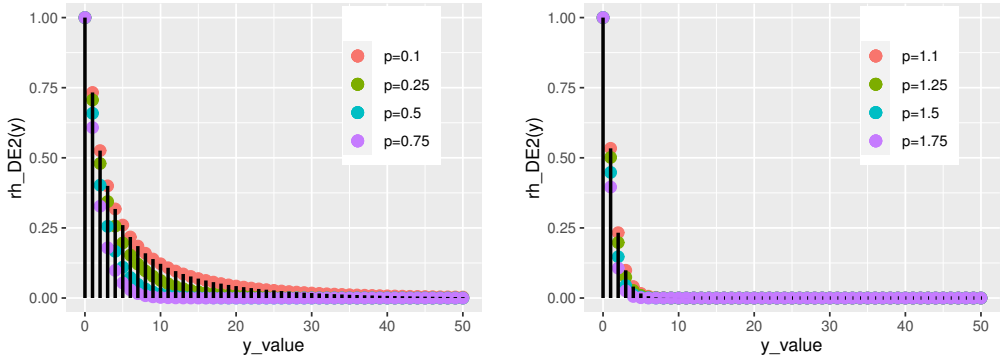

Figure 5. The reverse hazard function of DE2 for different values of p.

3.1.4. The Mills ratio function

Additionally, the Mills ratio function for the DE2 distribution is properly defined:

$$\text{Mills ratio}_{DE2}(y;p) = \frac{1}{h_{r\ DE2}(y;p)} = \frac{e^{p(y+1)}+p(y+1)-1}{py(e^p-1)+e^p-2}, \tag{3.4}$$

where $p > 0$ and $y \in \mathbb{N}$.

Figure 6 demonstrates the Mills ratio function of DE2 for various rate parameter values.

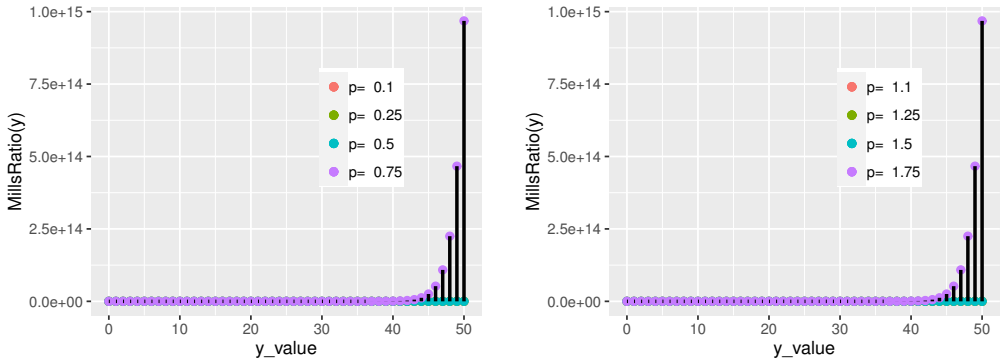

Figure 6. The Mills ratio function of DE2 for different values of p.

3.2. Moments and moment generating function of DE2 distribution

For a random variable Y with both the DE2 distribution and parameter p, the moments and moments generating function are generated in this subsection.

Theorem 3.1. *Let Y be a random variable having the DE2 distribution, then the moment generating function (mgf) of Y is*

$$M_Y(t) = \frac{e^{-(p-t)}(e^{-p} + p - 1) - e^{-p}(p + 1) + 1}{(1 - e^{-(p-t)})^2}, \tag{3.5}$$

where $p > 0$ and $p \neq t$.

Proof. We know that

$$\begin{aligned} M_Y(t) &= E(e^{ty}) = \sum_{ally} e^{ty} f_{DE2}(y; p) \\ &= \sum_{y=0}^{\infty} e^{ty}\left[e^{-py}\left((1 - e^{-p})\,(py + 1) - pe^{-p}\right)\right] \\ &= \sum_{y=0}^{\infty} e^{y(t-p)}\left((1 - e^{-p})\,(py + 1) - pe^{-p}\right) \\ &= (1 - e^{-p}) \sum_{y=0}^{\infty} e^{y(t-p)}\,(py + 1) - pe^{-p} \sum_{y=0}^{\infty} e^{y(t-p)}. \end{aligned}$$

Using "Wolfram Mathematica" we obtain

$$\begin{aligned} M_Y(t) &= (1 - e^{-p})\left[\frac{1 + e^{t-p}(p - 1)}{(1 - e^{t-p})^2}\right] - pe^{-p}\left[\frac{1}{1 - e^{t-p}}\right] \\ &= \frac{e^{t-p}(e^{-p} + p - 1) - e^{-p}(p + 1) + 1}{(1 - e^{t-p})^2} \end{aligned}$$

$$= \frac{e^{-(p-t)}(e^{-p}+p-1)-e^{-p}(p+1)+1}{(1-e^{-(p-t)})^2},$$

where $p > 0$ and $p \neq t$. So the theorem is proved.

As a result, using Eq (3.5), the DE2 distribution's first moment, or the mean, can be calculated as follows:

$$\mu_1 = E(Y) = \frac{dM_Y(t)}{dt}\bigg|_{t=0} = \frac{e^{-p}(p+1)-e^{-2p}}{(1-e^{-p})^2}, \quad p \neq 0. \tag{3.6}$$

Also, the second moment of the DE2 distribution is

$$\mu_2 = E(Y^2) = \frac{d^2M_Y(t)}{dt^2}\bigg|_{t=0} = \frac{3pe^{-2p}+e^{-p}(p+1)-e^{-3p}}{(1-e^{-p})^3}, \quad p \neq 0. \tag{3.7}$$

And so the variance of the DE2 distribution using Eqs (3.6) and (3.7) is

$$\begin{aligned} \mathrm{Var}_{DE2}[Y] &= E(Y^2) - [E(Y)]^2 \\ &= \frac{e^{-3p}(1-p)-e^{-2p}(p^2+2)+e^{-p}(p+1)}{(1-e^{-p})^4}, \quad p \neq 0. \end{aligned} \tag{3.8}$$

Consequently, the standard deviation of the DE2 distribution is

$$\begin{aligned} SD_{DE2} &= \sqrt{\frac{e^{-3p}(1-p)+e^{-p}(p+1)-e^{-2p}(p^2+2)}{(1-e^{-p})^4}} \\ &= \frac{\sqrt{e^{-3p}(1-p)+e^{-p}(p+1)-e^{-2p}}}{(1-e^{-p})^2}, \quad p \neq 0. \end{aligned} \tag{3.9}$$

The following relation yields the r^{th} moment of the DE2 distribution:

$$\mu_r = E(Y^r) = \frac{d^rM_Y(t)}{dt^r}\bigg|_{t=0}.$$

For example, the third moment of the DE2 distribution is

$$\begin{aligned} \mu_3 = E(Y^3) &= \frac{d^3M_Y(t)}{dt^3}\bigg|_{t=0} \\ &= \frac{-e^{-4p}+e^{-p}(p+1)+e^{-3p}(7p-3)+e^{-2p}(10p+3)}{(1-e^{-p})^4}, \quad p \neq 0. \end{aligned}$$

and so on.

3.3. Monotonicity of the DE2 distribution

3.3.1. Monotonicity of pmf in the DE2 distribution

In order to determine the critical point for the monotonicity of pmf for the DE2 distribution, we first set the first derivative of pmf to zero, as shown below:

$$f'_{DE2}(y;p) = p^2e^{-py}\Big(y(e^{-p}-1)+e^{-p}\Big), \quad p > 0 \text{ and } y \in \mathbb{N}. \tag{3.10}$$

Put $f'_{DE2}(y;p) = 0$ to obtain the pmf critical point as follows:

$$f'_{DE2}(y_0;p) = 0 \Rightarrow y_0 = \frac{e^{-p}}{1-e^{-p}}, \quad p \neq 0. \tag{3.11}$$

Because $p > 0$, $f_{DE2}(y;p)$ is monotonic dicreasing when $y > y_0$ and monotonic increasing when $y < y_0$.

3.3.2. Monotonicity of hazard rate function in the DE2 distribution

The hazard rate function's monotonic behavior, as demonstrated in [21], must be investigated. To do this, apply the same test as before:

$$h'_{DE2}(y;p) = \frac{e^p p^2}{(1+p(y+1))^2}.$$

We notice that, the hazard rate function is monotonic increasing for all value of Y and p. Since $h'_{DE2}(y;p) > 0$, $\forall p \geq 0$ and $\forall y \in \mathbb{N}$. Figure 4 shows the increasing of the hazard rate function at different rates of the rate parameter p.

3.4. Mode, skewness, kurtosis and Fano factor of DE2 distribution

The Fano factor, mode, skewness, kurtosis, and other critical measures (central tendency and dispersion) that are essential for discrete distribution in a wide range of fields, particularly medical statistics, are inferred in this subsection.

3.4.1. The mode of DE2 distribution

In the first, we take out the critical point of pmf for DE2 distribution by obtaining the first derivative of pmf, which we introduce in Eq (3.10) and set it to zero using Eq (3.11), yielding the point $y_0 = \frac{1}{e^p-1}$. Calculate the second derivative of $f_{DE2}(y;p)$ to determine whether this point is the local maximum or local minimum:

$$f''_{DE2}(y_0;p) = p^2 e^{-p\left(1+\frac{1}{e^p-1}\right)}(1-e^p), \quad p > 0 \text{ and } y = 0,1,2,\ldots.$$

As $p > 0$, so $(1-e^p) < 0$, $f''_{DE2}(y_0;p) < 0$, and y_0 is the local maximum point of pmf, this means that

$$Mode_{DE2} = y_0 = \frac{1}{e^p-1}, \quad p > 0. \tag{3.12}$$

3.4.2. The skewness of DE2 distribution (SK_{DE2})

In order to quantify the lack of symmetry in a probability distribution, there are several skewness measures that have been proposed in the literature. Karl Pearson's measure is the most widely applied of these and is denoted by the following formula:

$$SK_{DE2} = \frac{Mean - Mode}{\text{Standard Deviation}}.$$

Using Eqs (3.6), (3.9) and (3.12) we get the coefficient of skewness of DE2 distribution is

$$SK_{DE2} = \sqrt{\frac{p^2\ e^{-p}}{\left((1-p)e^{-2p} - e^{-p}(p^2+2) + p + 1\right)}},\quad p \neq 0. \tag{3.13}$$

We notice that, since $p > 0$ the coefficient of skewness is positive for any value of p, so the DE2 distribution is right skewness.

3.4.3. The kurtosis of DE2 distribution (CK_{DE2})

Using the classical definition of the coefficient of kurtosis, which is

$$CK_{DE2} = E\left(\frac{y - Mean}{SD}\right)^4 = \sum_{y=0}^{\infty}\left(\frac{y - E(y)}{SD}\right)^4 f_{DE2}(y;p).$$

Using Eqs (2.1), (3.6) and (3.9) and "Wolfram Mathematica" we obtained

$$\begin{aligned} CK_{DE2} = &\frac{\left[e^{-6p}(3 - 14p - 4p^2) - e^{-5p}(21 - 31p + 14p^2 + 6p^3)\right]}{\left[e^{-3p}(p-1) - e^{-p}(p+1) + e^{-2p}(p^2+2)\right]^2} \\ &+ \frac{\left[e^{-4p}(34 - 36p^2 - 3p^4) - e^{-3p}(21 + 31p + 14p^2 - 6p^3)\right]}{\left[e^{-3p}(p-1) - e^{-p}(p+1) + e^{-2p}(p^2+2)\right]^2} \\ &+ \frac{\left[(1-p)e^{-7p} + e^{-2p}(3 + 14p - 4p^2) + e^{-p}(p+1)\right]}{\left[e^{-3p}(p-1) - e^{-p}(p+1) + e^{-2p}(p^2+2)\right]^2},\quad p \neq 0. \end{aligned} \tag{3.14}$$

3.4.4. Fano factor of DE2 distribution (FF_{DE2})

The Fano factor (dispersion index), like the coefficient of variation, is a statistical measure of the dispersion of a probability distribution of a Fano noise. It bears the name Ugo Fano, an Italian American physicist (see [22]). The variance to mean ratio is known as the "Fano factor" (FF_{DE2}), and it is as follows:

$$FF_{DE2} = \frac{Variance}{Mean}.$$

It indicates whether a certain model is suitable for over- or under-dispersed datasets and is widely used in ecology as a standard measure for measuring clustering (over dispersion) or repulsion (under dispersion). The distribution is over- or under-dispersed if $FF_{DE2} > 1$ or $FF_{DE2} < 1$. The (FF_{DE2}) of the DE2 distribution, using Eqs (3.6) and (3.8), is given by

$$FF_{DE2} = \frac{1 + (1-p)e^{-2p} - e^{-p}(p^2+2) + p}{(1-e^{-p})^2\,(1 - e^{-p} + p)},\quad p \neq 0. \tag{3.15}$$

Table 1 provides descriptive statistics of the DE2 distribution at various parameter levels. It is evident that when the form parameter's value increases, the skewness and kurtosis decrease. The proposed distribution is successful for overscattered (underscattered) data when $FF_{DE2} > 1$ (< 1). The second observation is that the variance is larger than the mean.

Table 1. Mean, variance, skewness, kurtosis and Fano facto of DE2 for various values of the parameter p at $n = 50$.

p	Mean	Variance	SK_{DE2}	CK_{DE2}	FF_{DE2}
$p = 0.10$	17.8527	129.8252	0.7117	2.8184	7.2720
$p = 0.25$	7.4982	31.9934	1.3918	5.8138	4.2668
$p = 0.50$	3.5003	8.0786	1.3950	5.9404	2.3079
$p = 0.75$	2.1678	3.6284	1.3784	5.8745	1.6738
$p = 1.10$	1.3217	1.7148	1.3653	5.7682	1.2974
$p = 1.25$	1.1050	1.3364	1.3678	5.7273	1.2093
$p = 1.50$	0.8418	0.9353	1.3878	5.6811	1.1111
$p = 1.75$	0.6558	0.6890	1.4313	5.6807	1.0506
$p = 2.10$	0.4735	0.4746	1.5368	5.8006	1.0024
$p = 2.25$	0.4141	0.4100	1.5987	5.9094	0.9901
$p = 2.50$	0.3330	0.3252	1.7247	6.1874	0.9767
$p = 2.75$	0.2689	0.2607	1.8798	6.6111	0.9696

4. Maximum likelihood fstimation

The purpose of this part is to discover the parameters of the recommended DE2 distribution while using a maximum likelihood estimate (MLE).

If we construct a random sample of size n with the DE2 distribution, $Y_1, Y_2, \ldots, Y_n$, the likelihood function is given by

$$L_{DE2}(y_i, p) = \prod_{i=1}^{n} f_{DE2}(y_i; p) = \prod_{i=1}^{n} e^{-py_i}\Big((1 - e^{-p})\,(py_i + 1) - pe^{-p}\Big).$$

The log-likelihood function is given by

$$\begin{aligned} l_{DE2}(y_i, p) &= \ln\Big[L_{DE2}(y_i, p)\Big] \\ &= \sum_{i=1}^{n} \ln\Big[e^{-py_i}((1 - e^{-p})\,(py_i + 1) - pe^{-p})\Big] \\ &= \sum_{i=1}^{n}\Big[- py_i + \ln\Big((1 - e^{-p})\,(py_i + 1) - pe^{-p}\Big)\Big]. \end{aligned} \tag{4.1}$$

Differentiating Eq (4.1) with respect to the parameter p, we have the following equation

$$\frac{\partial l_{DE2}(y_i, p)}{\partial p} = -\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \frac{(1 + e^{-p}(p - 1))y_i + pe^{-p}}{p(1 - e^{-p})y_i - e^{-p}(1 + p) + 1}. \tag{4.2}$$

Now, putting

$$\frac{\partial l_{DE2}(y_i, p)}{\partial p} = 0,$$

$$\sum_{i=1}^{n} \frac{(1 + e^{-p}(p-1))y_i + pe^{-p}}{p(1-e^{-p})y_i - e^{-p}(1+p) + 1} = nE(y), \tag{4.3}$$

where $E(y)$ is the mean of y.

The maximum likelihood estimator (MLE) of the parameter $\hat{p}$ is obtained from the solution of Eq (4.3). Finding the exact form of the MLE of the parameter $\hat{p}$ is not achievable owing to the non-linear functions involved in the equation. It must therefore be solved quantitatively. The *R* statistical programming is used to do this.

5. Applications

The proposed distribution DE2 is fitted to three right-skewed and overly dispersed real lifetime counting datasets in this section. The first two datasets coming from a clinical trial were discussed in relation to individuals with acute leukemia's ability to maintain remission [23, 24]. Furthermore, the third dataset is fitting the COVID-19 death cases in some specific consecutive days over a given time period.

To test the goodness of fit of the proposed distribution, compare it with some related distributions. Such as the negative binomial distribution (NBD) [27], discrete Erlang-truncated Exponential distribution (DETE) [15], discrete Burr distribution (DBD) [4], discrete type-II half-logistics exponential distribution (DHLE_2) [16], and Poisson distribution [28].

In order to determine which distribution fits these three datasets accurately, some measures of goodness of fit are produced. These measurements include the values of the log-likelihood function -logL, the Kolmogorov-Smirnov (K-S) statistic along with its *P*-value, and the Akaike information criterion (AIC). The best distribution is the one that has the smallest -logL, AIC, and K-S with a *P*-value greater than 0.05.

Table 2 computes summary results for the three datasets in terms of central tendency and dispersion measures (minimum, maximum, mean, median (med.), mode, skewness (SK_{DE2}), kurtosis (CK_{DE2}), and Fano factor (FF_{DE2})). Because $(SK_{DE2}) > 0$, all datasets are clearly right-skewed, and $FF_{DE2} > 1$, they are overly dispersed.

Table 2. Summary statistics for datasets 1–3.

	n	Min	Max	Mean	Med.	Mode	SK_{DE2}	CK_{DE2}	FF_{DE2}
dataset 1	21	1	30	9.5	8	4.5	1.5	4.8	5.6
dataset 2	30	1	35	12.5	9.5	6.03	0.7	2.9	5.7
dataset 3	32	0	59	7.4	6	3.5	4.3	22.3	13.5

5.1. Application of DE2 distribution to the Leukomia dataset

The results of a clinical trial testing 6-mercaptopurine (6-MP) [25], to a placebo in patients with acute leukemia were reported by two patient groups. The first group, insert the remission times (in

weeks) of 21 patients who were given the drug (6-MP), as introduced in Table 3. The second group presents the remission times of 30 patients given the placebo, as shown in Table 4. The data can be found in [24, 26]. We show the results of measures of goodness of fit that are applied to two datasets in Tables 5 and 6.

Table 3. A group of 21 patients with leukemia (6-MP).

Dataset 1										
1	1	2	3	4	5	8	8	8	8	8
8	8	11	11	11	11	13	13	27	30	

Table 4. A group of 30 patients with leukemia (Placebo).

Dataset 2									
1	1	2	4	4	6	6	6	6	7
8	9	9	9	9	10	12	13	14	15
19	20	20	20	20	20	20	21	30	35

Table 5. Parameters estimates (P-E), -logL, AIC, K-S and *P*-value of DE2, NBD, DETE, DBD, $DHLE_2$ and Poisson distributions for dataset 1.

Dist.	Parameters estimates			-logL	AIC	K-S	*P*-value
DE2	–	–	$\hat{p}$=0.201	66.649	135.298	0.253	0.137
NBD	$\hat{r}$ = 2.418	–	$\hat{p}$ = 0.203	66.729	137.458	0.532	0.004
DETE	$\hat{\alpha}$ = 4.465	–	$\hat{p}$ = 0.978	69.295	142.59	0.309	0.036
DBD	$\hat{\alpha}$ = 208.66	–	$\hat{p}$ = 0.998	74.501	153.002	0.337	0.017
$DHLE_2$	$\hat{\alpha}$ = 394.95	–	$\hat{p}$ = 0.024	96.286	206.572	0.902	$2.8e^{-15}$
Poisson	$\hat{\lambda}$ = 9.476	–	–	90.504	183.009	0.5003	0.018

Table 6. Parameters estimates, -logL, AIC, K-S and *P*-value for DE2, NBD, DETE, DBD, $DHLE_2$ and Poisson distributions of dataset 2.

Dist.	Parameters estimates			-logL	AIC	K-S	*P*-value
DE2	–	–	$\hat{p}$ = 0.154	103.06	208.12	0.144	0.559
NBD	$\hat{r}$ = 2.455	–	$\hat{p}$ = 0.164	103.02	210.04	0.515	0.0001
DETE	$\hat{\alpha}$ = 4.313	–	$\hat{p}$ = 0.982	107.018	218.036	0.249	0.0482
DBD	$\hat{\alpha}$ = 199.6	–	$\hat{p}$ = 0.998	119.98	243.95	0.362	0.0008
$DHLE_2$	$\hat{\alpha}$ = 5212.3	–	$\hat{p}$ = 0.007	150.766	304.532	0.933	0.0000
Poisson	$\hat{\lambda}$ = 12.534	–	–	145.29	292.58	0.500	0.0002

By comparing the goodness of fit statistics in Tables 5 and 6 among the six distributions, it is clear

that the proposed distribution is the only one that fits the two datasets due to the lowest -logL, AIC, and K-S in addition to *P*-values greater than 0.05.

5.2. Application of DE2 distribution to the COVID-19 dataset

DE2 distribution is fitted to the more well-known fields of COVID-19 survival times in Australia [29]. This dataset belongs to Australia of 32 days, that is recorded from 3 September 2020 to 4 October 2020. This data is made up of cases that are added on a daily basis, which are shown in Table 7.

Table 7. The number of daily deaths in Australia due to the COVID-19.

Dataset 3											
6	15	59	11	5	9	8	11	7	9	6	7
6	0	8	8	5	7	5	2	3	5	2	8
1	2	3	7	4	2	2	3				

Table 8 displays the -logL, AIC, and Kolmogorov-Smirnov (K-S) statistics, as well as the *P*-value for each model.

We notice that, in Table 8, the best distribution is DE2 since the smallest of -logL, AIC, and K-S have a *P*-value greater than 0.05.

Table 8. Parameters estimates, -logL, AIC, K-S and *P*-value for DE2, NBD, DETE, DBD, $DHLE_2$ and Poisson distributions of dataset 3.

Dist.	Parameters estimates			-logL	AIC	K-S	*P*-value
DE2	–	–	$\hat{p} = 0.254$	96.984	195.967	0.154	0.431
NBD	$\hat{r} = 1.606$	–	$\hat{p} = 0.179$	96.699	197.398	0.518	0.001
DETE	$\hat{\alpha} = 4.541$	–	$\hat{p} = 0.972$	98.017	200.033	0.255	0.031
DBD	$\hat{\alpha} = 4.258$	–	$\hat{p} = 0.873$	107.148	218.295	0.272	0.018
$DHLE_2$	$\hat{\alpha} = 2.561$	–	$\hat{p} = 0.568$	140.282	284.564	0.684	$1.99e^{-13}$
Poisson	$\hat{\lambda} = 7.375$	–	–	164.917	331.833	0.500	0.002

6. Conclusions

This article uses a survival discretizing method to derive a new one-parameter discrete Erlang-2 distribution (DE2) from a one-parameter continuous Erlang-2 distribution (E2 distribution). We explore certain statistical characteristics of the postulated distribution, including measures of central tendency, measures of dispersion, monotonicity, moments, and the moment generating function. The maximum likelihood method is used to estimate model parameters. Furthermore, the proposed distribution is unimodal (gets this mode) and has a useful effect on the overly dispersed dataset, as measured by the Fano factor as the rate parameter is increased. Furthermore, we used DE2 on two datasets that discussed the results of a clinical trial reporting on the ability to maintain remission in acute leukemia patients and compared them to a few discrete distributions that are connected.

Additionally, we applied DE2 to fit the COVID-19 death cases dataset. Finally, the sitting work seeks to draw broader applications in engineering, medicine, and other areas of research (In the future, we plan to present a paper in this area).

Acknowledgements

We appreciate the handling editor and the two anonymous reviewers for their insightful criticism and recommendations, which have significantly improved the presentation of the manuscript.

Conflict of interest

The author declares having no competing interest.

References

1. C. Bracquemond, O. Gaudoin, A survey on discrete lifetime distribution, *Int. J. Reliab. Qual. Saf. Eng.*, **10** (2003), 69–98. http://doi.org/10.1142/S0218539303001007
2. S. Chakraborty, Generating discrete analogues of continuous probability distributions-a survey of methods and constructions, *J. Stat. Distrib. Appl.*, **2** (2015), 6. http://doi.org/10.1186/s40488-015-0028-6
3. D. Roy, The discrete normal distribution, *Commun. Stat. Theory Meth.*, **32** (2003), 1871–1883. http://doi.org/10.1081/STA-120023256
4. H. Krishna, P. S. Pundir, Discrete Burr and discrete Pareto distributions, *Stat. Meth.*, **6** (2009), 177–188. http://doi.org/10.1016/j.stamet.2008.07.001
5. S. Chakraborty, D. Chakravarty, Discrete Gamma distribution: properties and parameter estimations, *Commun. Stat. Theory Meth.*, **41** (2012), 3301–3324. http://doi.org/10.1080/03610926.2011.563014
6. V. Nekoukhou, M. H. Alamatsaz, H. Bidram, Discrete generalized exponential distribution of a second type, *J. Theor. Appl. Stat.*, **47** (2013), 876–887. http://doi.org/10.1080/02331888.2011.633707
7. V. Nekoukhou, H. Bidram, The exponentiated discrete Weibull distribution, *Sort Stat. Oper. Res. Trans.*, **39** (2015), 127–146.
8. M. H. Alamatsaz, S. Dey, T. Dey, S. S. Harandi, Discrete generalized Rayleigh distribution, *Pak. J. Stat.*, **32** (2016), 1–20.
9. T. Hussain, M. Aslam, M. Ahmad, A two parameter discrete lindley distribution, *Rev. Colomb. Estad.*, **39** (2016), 45–61. http://doi.org/10.15446/rce.v39n1.55138
10. P. Khongthip, M. Patummasut, W. Bodhisuwan, The discrete weighted exponential distribution and its applications, *Songklanakarin J. Sci. Technol.*, **40** (2018), 1105–1114. http://doi.org/10.14456/sjst-psu.2018.137

11. K. Jayakumar, M. G. Babu, Discrete additive Weibull geometric distribution, *J. Stat. Theory Appl.*, **18** (2019), 33–45. http://doi.org/10.2991/jsta.d.190306.005

12. E. Altun, M. El-Morshedy, M. S. Eliwa, A study on discrete Bilal distribution with properties and applications on integer-valued autoregressive process, *Revstat Stat. J.*, **18** (2020), 70–99.

13. M. S. Eliwa, Z. A. Alhussain, M. El-Morshedy, Discrete Gompertz-G family of distributions for over- and under-dispersed data with properties, estimation, and applications, *Mathematics*, **8** (2020), 358. http://doi.org/10.3390/math8030358

14. A. S. Eldeeb, M. Ahsan-ul-Haq, A. Baba, A discrete analog of inverted Topp-Leone distribution: properties, estimation and applications, *Int. J. Anal. Appl.*, **19** (2021), 695–708. http://doi.org/10.28924/2291-8639-19-2021-695

15. A. R. El-Alosey, Discrete Erlang-truncated exponential distribution, *Int. J. Stat. Appl. Math.*, **6** (2021), 230–236. http://doi.org/10.22271/maths.2021.v6.i1c.653

16. M. Ahsan-ul-Haq, A. Babar, S. S. Hashmi, A. S. Alghamd, A. Z. Afify, The discrete type-II half-logistic exponential distribution with applications to COVID-19 data, *Pak. J. Stat. Oper. Res.*, **17** (2021), 921–932. http://doi.org/10.18187/pjsor.v17i4.3772

17. A. R. El-Alosey, H. Eledum, Discrete extended Erlang-truncated exponential distribution and its applications, *Appl. Math. Inf. Sci.*, **16** (2021), 127–138. http://doi.org/10.18576/amis/160113

18. D. Roy, Reliability measures in the discrete bivariate set-up and related characterization results for a bivariate geometric distribution, *J. Multivar. Anal.*, **46** (1993), 362–373. http://doi.org/10.1006/jmva.1993.1065

19. M. Xie, O. Gaudoin, C. Bracquemond, Redefining failure rate function for discrete distribution, *Int. J. Reliab. Qual. Saf. Eng.*, **9** (2002), 75–85. http://doi.org/10.1142/S0218539302000822

20. M. Bebbington, C. D. Lai, M. Wellington, R. Zitikis, The discrete additive Weibull distribution: a bathtub-shaped hazard for discontinuous failure data, *Reliab. Eng. Syst. Saf.*, **106** (2012), 37–44. http://doi.org/10.1016/j.ress.2012.06.009

21. P. L. Gupta, R. C. Gupta, R. C. Tripathi, On the monotonic properties of discrete failure rates, *J. Stat. Plann. Infer.*, **65** (1997), 55–68. http://doi.org/10.1016/S0378-3758(97)00064-5

22. U. Fano, Ionization yield of radiations. II. The fluctuations of the number of ions, *Phys. Rev.*, **72** (1974), 26–29. http://doi.org/10.1103/PhysRev.72.26

23. E. A. Gehan, Estimating survival functions from the life table, *J. Chronic Dis.*, **21** (1969), 629–644. http://doi.org/10.1016/0021-9681(69)90035-6

24. J. F. Lawless, *Statistical models and methods for lifetime data*, John Wiley Sons Inc., 2002. http://doi.org/10.1002/9781118033005

25. E. Freireich, E. Gehan, E. Frei, L. Schroeder, I. Wolman, R. Anbari, et al., The effect of 6-mercaptopurine on the duration of steroid induced remission in acute leukemia, *Blood*, **21** (1963), 699–716. http://doi.org/10.1182/blood.V21.6.699.699

26. J. P. Klein, M. L. Moeschberger, *Survival analysis*, Springer Verlag, 1997.

27. K. Aramidis, An EM algorithm for estimating negative binomial parameters, *Aust. New Z. J. Stat.*, **41** (1999), 213–221. http://doi.org/10.1111/1467-842X.00075

28. P. C. Consul, G. C. Jain, A generalization of the Poisson distribution, *Technometrics*, **15** (1973), 791–799. http://doi.org/10.2307/1267389

29. E. M. Almetwally, G. M. Ibrahim, Discrete Alpha power inverse Lomax distribution with application of COVID-19 data, *Int. J. Appl. Math.*, **9** (2020), 49–60.