Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/isl/120203

On the Negative Binomial-Discrete Erlang-Truncated Exponential Mixture

A. R. El-Alosey[1,*] *and H. Eledum*[2]

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

Received: 17 Mar. 2022, Revised: 08 Aug. 2022, Accepted: 10 Aug. 2022
Published online: 1 Feb. 2023

Abstract: This paper introduces a new two parameters discrete distribution by mixing the negative binomial with the discrete Erlang-truncated exponential distribution; the proposed distribution is called the negative binomial-discrete Erlang-truncated exponential (NBDETE) distribution. The quantile function, moments, moment generating function, Renyi entropy, order statistics, stress-strength reliability, and simulating the random sample are some of the statistical features of the NBDETE that are investigated. The method of maximum likelihood procedure is adopted for estimating the model parameters. A Monte Carlo simulation is performed to assess the performance of the accuracy of point estimates for NBDETE. Finally, an application of the new distribution is illustrated in a real data set. The proposed NBDETE distribution is the right-skewed over-dispersed, and it can accommodate a constant hazard rate function.

Keywords: Mixture of distributions, hazard function, maximum likelihood estimation, quantile functions, Stress-strength parameter.

1 Introduction

The Negative Binomial distribution (NB) is used to model the count data and is employed as a functional form that overcomes the over-dispersion restriction of the Poisson distribution. In other words, the NB distribution relaxes the equality of mean and variance property of the Poisson distribution. This distribution is produced by mixing the Poisson and gamma distributions. The NB distribution has become increasingly popular as a more flexible alternative to counting data with over-dispersion. Many studies have been made on the mixtures of the negative binomial distribution. Gomez-Deniz et al. [1] introduced the univariate and multivariate versions of the negative binomial-inverse Gaussian distributions. The negative binomial-Beta Exponential distribution is studied by Pudprommarat and Bodhisuwan [2]. Shanker et al. [3] proposed the quasi-negative binomial- discrete Akash distribution. The negative binomial logistic distribution is derived by Ravikumar et al. [4].

The discrete Erlang-Truncated Exponential (DETE) distribution is used to model the count data and has a constant hazard rate function [5]. This distribution is defined by discretizing the continuous Erlang-Truncated Exponential (ETE) distribution. Recently, a few researchers mixed the continuous Erlang-Truncated Exponential distribution with some other distributions. For example, Kongrod et al. [6] proposed the negative binomial-Erlang distribution by mixing ETE with NB distribution, while Shrahili et al. [7] mixed the ETE with beta distribution to produce the beta Erlang truncated exponential distribution (BETE). However, to the best of our knowledge of the literature, no one has mixed the DETE with any other discrete or continuous distribution to get a new distribution.

In this paper, we introduce a new mixture of negative binomial distribution by mixing the negative binomial and the zero-truncated discrete Erlang-truncated exponential distribution, which is called the negative binomial-discrete Erlang-truncated exponential distribution (NBDETE). We investigate some statistical properties of NBDETE and estimate its parameters using the maximum likelihood method.

The remainder of this paper is structured as follows: The new proposed distribution is presented in section 2. Some statistical properties such as quantile function, moment generating function, Renyi entropy, order statistics, stress-strength

* Corresponding author e-mail: alaa.abdelmaksoud@science.tanta.edu.eg; a_r_elalosey@yahoo.com

parameter, and simulating the random sample are demonstrated in section 3. Section 4 provides the maximum likelihood method to estimate NBDETE parameters. The Monte Carlo simulation is performed in section 5. Section 6 demonstrates an application of the proposed distribution NBDETE. Finally, section 7 provides some concluding remarks.

1.1 The Negative Binomial (NB) and Discrete Erlang-Truncated (DETE) Distributions

The negative binomial distribution is a probability distribution that is used with discrete random variables. This distribution models the number of successes in a sequence of independent and identically distributed Bernoulli trials before a specified number of failures occurs. Moreover, NB generalized the Geometric distribution.

Suppose X is the BN random variable with parameters $n > 0$ and $0 \le p \le 1$, then the pmf for X is given by

$$f_X(x;n,p) = \binom{n+x-1}{n-1} p^n (1-p)^x \quad, x = 0,1,2 \tag{1}$$

where n is the number of successes and p is the probability of success.

The pmf of the random variable N having the discrete Erlang-Truncated distribution with parameters $0 \le p \le 1$ and $\beta > 0$ is defined as [5]

$$f_N(n;p,\ \beta) = (1-p^{\beta})\ p^{\beta(n-1)}\ ; \qquad n = 1,2,3,\ldots \tag{2}$$

2 Negative binomial- Discrete Erlang-truncated Exponential distribution (NBDETE)

This section discusses the probability mass and cumulative distribution functions of the proposed Negative Binomial-Discrete Erlang-truncated Exponential (NBDETE) distribution.

2.1 Probability mass and cumulative distribution functions for NBDETE distribution

Assuming that the parameter n of the Negative binomial distribution in Eq.(1) follows a zero-truncated discrete Erlang-Truncated Exponential (DETE) distribution of Eq.(2), then the pmf of the Negative binomial mixture of the DETE distribution (NBDETE) can be obtained as

$$\begin{aligned} f_X(x;\ p,\ \beta) &= \sum_{n=1}^{\infty} f_X(x;n,p)\, f_N(n;\ p,\ \beta) \\ &= \frac{p(1-p^{\beta})\ (1-p)^x}{\left(1-p^{\beta+1}\right)^{x+1}};\ \ x = 0,1,2,\ldots \end{aligned} \tag{3}$$

with the corresponding cdf as

$$F_X(a;\ p,\ \beta) = 1 - \left(\frac{1-p}{1-p^{\beta+1}}\right)^{a+1};\quad a = 0,1,2,\ldots \tag{4}$$

where, $0 \le p \le 1$ and $\beta > 0$ are shape parameters.

Fig.1 and Fig.2 respectively demonstrate pmf and cdf of NBDETE distribution for combination values of p and β. From Fig. 1, it is observed that the proposed distribution is right-skewed, and its pmf is a decreasing function.

2.2 Survival and Hazard functions

The survival function of X is

$$s_X(x;\ p,\ \beta\ ) = 1 - F_X(x-1;\ p,\ \beta) = \left(\frac{1-p}{1-p^{\beta+1}}\right)^x$$

The hazard function is as follows

$$h_X(x;\ p,\ \beta\ ) = \frac{f(x;\ p,\ \beta)}{S_X(x;\ p,\ \beta\ )} = \frac{p(1-p^{\beta})}{1-p^{\beta+1}} \tag{5}$$

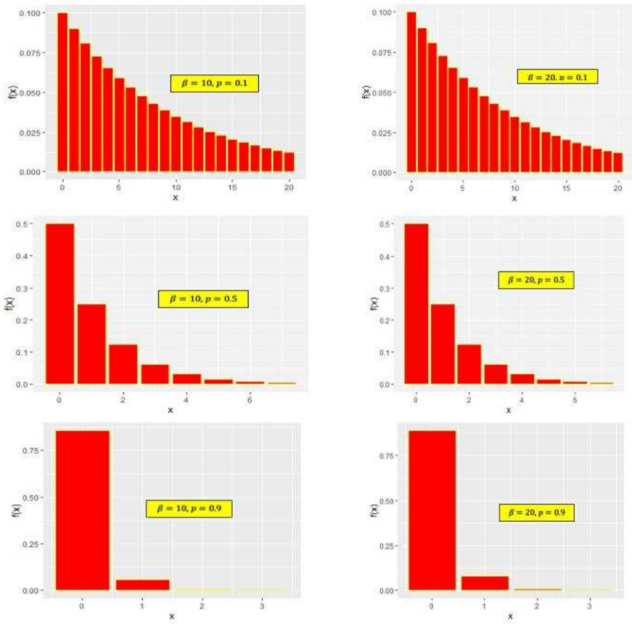

Fig. 1: The pmfs of NBDETE for some values of p and β.

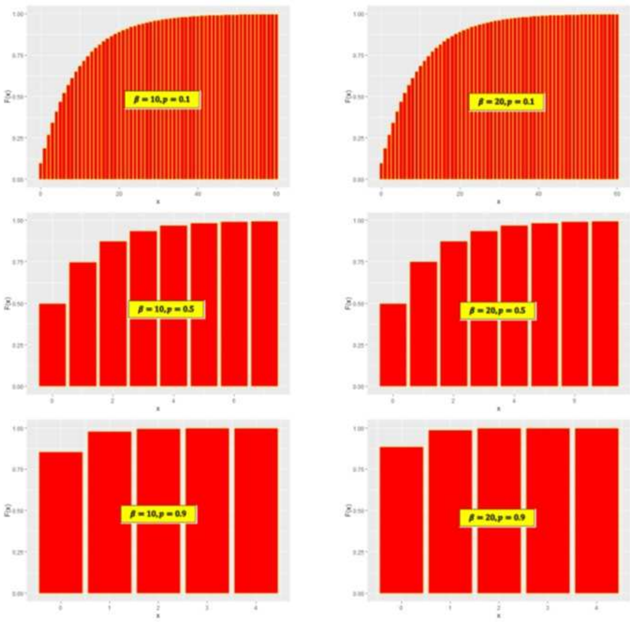

Fig. 2: The cdfs of NBDETE for some values of p and β.

Fig. 3 as well as Table 1 display the behavior of the hazard rate function of NBDETE for various values of β and p. Based on Fig. 3 and Table 1, it is observed that the hazard rate function of the proposed distribution is constant, and increases with p and β.

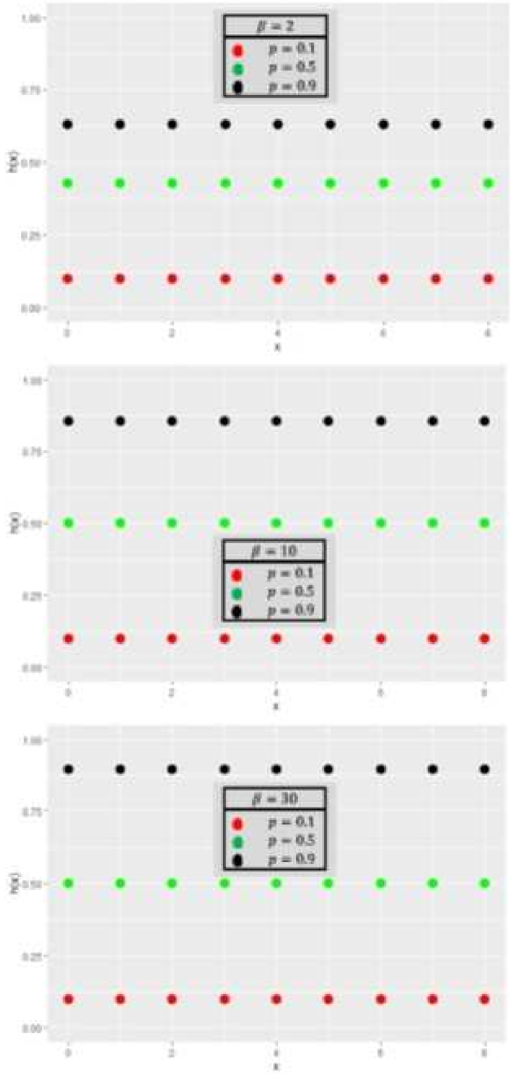

Fig. 3: Hazard function of NBDETE for different values of p and β.

Table 1: Hazard function of NBDETE for different values of p and β

	p=0.1	p=0.5	p=0.9
β=2	0.0990991	0.4285710	0.6309960
β=8	0.1000000	0.4997557	0.8542676
β=20	0.1000000	0.5000000	0.8960335

3 Distributional properties

Some statistical properties of the NBDETE distribution including quantile function, moments, moment generating function, Rényi entropy, order statistics, stress-strength reliability, and simulating the random sample, were obtained in this section.

3.1 Quantile function

The quantile of order $0 < r < 1$, can be calculated by inverting the cdf in Eq (4) as follows

$$F_X(Q;p,\ \beta) = 1 - \left(\frac{1-p}{1-p^{\beta+1}}\right)^{Q+1}$$

Then $F_X^{-1}(r) = \min\{x \in R : F_X(x) \geq r\}$

$$1 - \left(\frac{1-p}{1-p^{\beta+1}}\right)^{Q+1} = r$$

Thus, The r^{th} quantile is

$$Q(r;\alpha,\ \beta,\ p) = \frac{\log_2(1-r)}{\log_2(1-p) - \log_2\left(1-p^{\beta+1}\right)} - 1 \tag{6}$$

The median of NBDETE can be obtained by substituting by $r = \frac{1}{2}$ in Eq.(6) as follows:

$$Q_{0.5} = Q(r;\alpha,\beta,p) = \frac{-1}{\log_2(1-p) - \log_2(1-p^{\beta+1})} - 1$$

3.2 The moment generating function

In this subsection, the moment generating function of a random variable X having the NBDETE distribution with parameters $(p,\ \beta)$ is derived as

$$\begin{aligned} M_X(t) &= E\left(e^{tx}\right) \\ &= \frac{p(1-p^{\beta}-p^{\beta+1}-p^{2\beta+1})}{(1-p^{\beta+1})(1-e^{t}(1-p)-p^{\beta+1})} \end{aligned} \tag{7}$$

Using Eq.(7), we can compute the first moment (mean) of the NBDETE distribution as follows

$$\mu_1 = E(X) = \left.\frac{dM_X(t)}{dt}\right|_{t=0} = \frac{(1-p)}{p(1-p^{\beta})} \tag{8}$$

The second moment about the origin is

$$\mu_2 = E\left(X^2\right) = \frac{(1-p)}{p(1-p^{\beta})} + \frac{2(1-p)^2}{p^2(1-p^{\beta})^2}$$

As a result, the NBDETE distribution's variance is given by

$$\begin{aligned} \sigma^2 &= V(X) = \mu_2 - \mu_1^2 \\ &= \frac{(1-p)}{p(1-p^{\beta})} + \frac{(1-p)^2}{p^2(1-p^{\beta})^2} \end{aligned} \tag{9}$$

It is obvious from Eqs. (8) and (9) that

$$\sigma^2 = \mu + \mu^2 > \mu$$

This demonstrates that the NBDETE distribution is always over-dispersed, allowing to be utilized with data that is over-dispersed (the variance greater than the mean).

The mean and variance of the NBDETE distribution for various combinations of p and β are shown in Table 2. It is obvious that when p and β grow up, both the mean and variance dropdown. As a result, the two parameters of the suggested distribution can be utilized to fit most count data sets.

Table 2: Mean and Variance of NBDETE distribution for different values of p and β

p	0.01		0.1		0.1	
β	Mean	Variance	Mean	Variance	Mean	Variance
2	9.09091	91.7355	1.33333	3.11111	0.584795	0.926781
5	9.00009	90.0017	1.03226	2.09781	0.271327	0.344945
8	9.00000	90	1.00392	2.01178	0.195092	0.233152
20	9.00000	90	1	2	0.126489	0.142489

The third moment about the origin is

$$\begin{aligned}\mu_3 &= E\left(X^3\right)\\ &= \frac{(1-p)}{p\left(1-p^{\beta}\right)} + \frac{6(1-p)^2}{p^2\left(1-p^{\beta}\right)^2}\\ &\quad + \frac{6(1-p)^3}{p^3\left(1-p^{\beta}\right)^3}\end{aligned}$$

The fourth moment about the origin is

$$\begin{aligned}\mu_4 &= E\left(X^4\right)\\ &= \frac{(1-p)}{p\left(1-p^{\beta}\right)} + \frac{14(1-p)^2}{p^2\left(1-p^{\beta}\right)^2}\\ &\quad + \frac{36(1-p)^3}{p^3\left(1-p^{\beta}\right)^3} + \frac{24(1-p)^4}{p^4\left(1-p^{\beta}\right)^4}\end{aligned}$$

The NBDETE distribution has the coefficient of variation $(C.V)$, coefficient of Skewness $(\sqrt{\beta_1})$, the coefficient of Kurtosis (β_2), and the index of dispersion (γ) as

$$C.V = \frac{\sigma}{\mu_1} = \sqrt{\frac{1-p^{\beta+1}}{1-p}}$$

$$\begin{aligned}\sqrt{\beta_1} &= \frac{\mu_3 - 3\mu\sigma^2 - \mu^3}{(\sigma^2)^{\frac{3}{2}}}\\ &= \left[\frac{(1-p)}{p\left(1-p^{\beta}\right)} + \frac{3(1-p)^2}{p^2\left(1-p^{\beta}\right)^2} + \frac{2(1-p)^3}{p^3\left(1-p^{\beta}\right)^3}\right]\\ &\quad \div \left[\frac{(1-p)}{p\left(1-p^{\beta}\right)} + \frac{(1-p)^2}{p^2\left(1-p^{\beta}\right)^2}\right]^{\frac{3}{2}}\end{aligned}$$

$$\beta_2 = \frac{\mu_4 - 4\mu\mu_3 + 6\mu_2\mu^2 - 3\mu^4}{(\sigma^2)^2}$$
$$= \left[1 + \frac{10(1-p)}{p\left(1-p^\beta\right)} + \frac{18(1-p)^2}{p^2\left(1-p^\beta\right)^2} + \frac{4(1-p)^3}{p^3\left(1-p^\beta\right)^3}\right]$$
$$\div \left[\frac{(1-p)}{p\left(1-p^\beta\right)} + \frac{2(1-p)^2}{p^2\left(1-p^\beta\right)^2} + \frac{(1-p)^3}{p^3\left(1-p^\beta\right)^3}\right]$$
$$\gamma = \frac{\sigma^2}{\mu_1} = 1 + \frac{(1-p)}{p\left(1-p^\beta\right)}$$

Fig. 4 shows the nature and behavior of the NBDETE distribution's the coefficient of variation, coefficient of Skewness, coefficient of Kurtosis , and the index of dispersion for different values of parameters p and β.

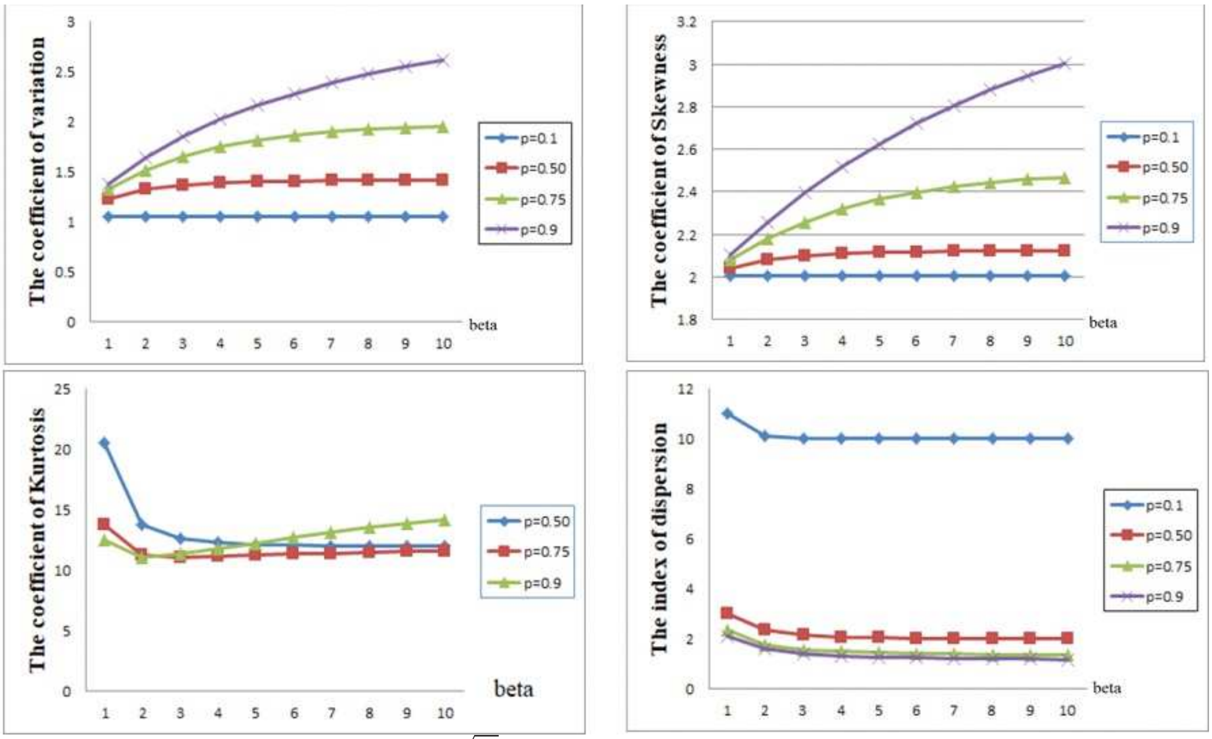

Fig. 4: $C.V$, $\sqrt{\beta_1}$, β_2 and γ of the NBDETE distribution.

Based on Fig. 4 it is clear that the coefficient of variation, coefficient of skewness and the index of dispersion decrease when p increases.

3.3 Entropy

Statistical entropy is a probabilistic measure of uncertainty or ignorance about the outcome of a random experiment, as well as a measure of that uncertainty's decrease. The Rényi entropy, which has been developed and utilized in a range of disciplines and contexts, is one of many entropy and information indices. This measure is defined as

$$I_R(\gamma) = \frac{1}{1-\gamma} \log\left[\sum_{x=0}^{\infty} f^\gamma(x)\right]$$

where $\gamma > 0$ and $\gamma \neq 0$
The Rényi entropy for a random variable X with a pmf of the NBDETE distribution in Eq.(3) is

$$I_R(\gamma) = \frac{1}{1-\gamma} \log \left[\frac{p^{\gamma} 1 - p^{\beta})^{\gamma}}{\left(1 - p^{\beta+1}\right)^{\gamma} - (1-p)^{\gamma}} \right]$$

3.4 Order statistics

The purpose of this subsection is to develop some general equations concerning order statistics for the NBDETE distribution. More precisely, let $f_k(x; p,\ \beta)$ and $F_k(x; p,\ \beta)$ be the pmf and cdf of the k^{th} order statistic of a random sample $X_1, X_2, \ldots, X_n$; of size n, respectively, derived from NBDETE $(p,\ \beta)$. The k^{th} order statistic's pmf is

$$\begin{aligned} &f_k(x; p\ \beta) \\ &= \frac{n!}{(k-1)!\,(n-k)!} [F(x; p,\ \beta)]^{k-1} [1 - F(x; p,\ \beta)]^{n-k} f(x; p,\ \beta) \\ &= \frac{n!}{(k-1)!\,(n-k)!} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} \frac{p(1-p^{\beta})}{1-p^{\beta+1}} \\ &\quad \times \left(\frac{1-p}{1-p^{\beta+1}} \right)^{x(n-k+j+1)+n+j-k} \end{aligned}$$

The k^{th} order statistic's cdf is

$$\begin{aligned} F_k(x; p,\ \beta) &= \sum_{i=k}^{n} \binom{n}{i} [F(x; \lambda,\ \beta)]^i [1 - F(x; \lambda,\ \beta)]^{n-i} \\ &= \sum_{i=k}^{n} \sum_{j=0}^{n} (-1)^j \binom{n}{i} \binom{n}{j} \left(\frac{1-p}{1-p^{\beta+1}} \right)^{(x+1)(n-i+j)} \end{aligned}$$

Consider the minimum order statistics as $X_{(1)} = min(X_1, X_2, \ldots, X_n)$, the maximum order statistics as $X_{(n)} = max(X_1, X_2, \ldots, X_n)$, and the medium order statistics as $X_{(m+1)}$ with$m = \frac{n}{2}$. Therefore, the pmf of the minimum, maximum and median respectively are,

$$f_1(x; p, \beta) = \frac{np(1-p^{\beta})}{1-p^{\beta+1}} \left(\frac{1-p}{1-p^{\beta+1}} \right)^{n(x+1)-1}$$

$$\begin{aligned} f_n(x; p,\ \beta) &= \frac{np(1-p^{\beta})}{1-p^{\beta+1}} \left(\frac{1-p}{1-p^{\beta+1}} \right)^{x} \\ &\times \left[1 - \left(\frac{1-p}{1-p^{\beta+1}} \right)^{(x+1)} \right]^{n-1} \end{aligned}$$

$$\begin{aligned} f_{m+1}(x; p,\ \beta) &= \frac{n!}{(m)!\,(n-m+1)!} \frac{p(1-p^{\beta})}{1-p^{\beta+1}} \\ &\times \left(\frac{1-p}{1-p^{\beta+1}} \right)^{(n-m)(x+1)-1} \\ &\times \left[1 - \left(\frac{1-p}{1-p^{\beta+1}} \right)^{(x+1)} \right]^{m} \end{aligned}$$

3.5 Stress-strength parameter

In reliability analysis, the quantity $R = P(X > Y)$ is called the stress-strength model. This parameter has many applications in engineering, biostatistics, quality control, military, medicine, and psychology [8,9]. The stress-strength model can be used in the medical field when conducting a case-control study assuming that Y and X represent the outcomes of a treatment and a control group, respectively, after that the ineffectiveness of the treatment is measured by R. In reliability analysis the stress-strength model can be applied using Y as the strength of the component and X is stress, while R and $(1-R)$ refer to the probabilities of system performance and system failure, respectively. The discrete version of the stress-strength parameter is specified as

$$R = P(X > Y) = \sum_{x=0}^{\infty} f_X(x) F_Y(x)$$

where the pmf and cdf of the independent discrete random variables X and Y, respectively, are denoted by f_X and F_Y, respectively. Now, let $X \sim NBDETE(p_1, \beta_1)$ and $Y \sim NBDETE(p_2, \beta_2)$. Using Eqs. (3), and (4), we get

$$\begin{aligned} R &= \sum_{x=0}^{\infty} \left[\frac{p_1(1-{p_1}^{\beta_1})}{\left(1-{p_1}^{\beta_1+1}\right)} \left(\frac{1-p_1}{1-{p_1}^{\beta_1+1}} \right)^x \right] \\ &\quad \times \left[1 - \left(\frac{1-p_2}{1-{p_2}^{\beta_2+1}} \right)^{x+1} \right] \\ &= 1 - \frac{p_1(1-{p_1}^{\beta_1})(1-p_2)}{\left(1-{p_1}^{\beta_1+1}\right)\left(1-{p_2}^{\beta_2+1}\right) - (1-p_1)(1-p_2)} \end{aligned} \tag{10}$$

3.6 Simulating the Random Sample

Random numbers from the NBDETE distribution can be obtained by equating the cdf of the distribution in Eq.(4) with a uniform random number and inverting the expression; that is the random number from NBDETE is gotten by solving the following equation for x

$$1 - \left(\frac{1-p}{1-p^{\beta+1}} \right)^{x+1} = u$$

The random sample from NBDETE can be further explained as

$$x = \frac{\log(1-u)}{\log(1-p) - \log(1-p^{\beta+1})} - 1 \tag{11}$$

where u is an arbitrary continuous uniform point within $(0,1)$.

4 Maximum likelihood estimation

This section aims to derive the maximum likelihood estimation (MLE) for parameters of the proposed distribution NBDETE.
Let $X_1, X_2, \ldots, X_n$ be a random sample of size n having the NBDETE distribution. The log-likelihood of the NBDETE distribution is

$$\begin{aligned} \ell &= n \log \left[\frac{p(1-p^{\beta})}{1-p^{\beta+1}} \right] \\ &\quad - \sum_{i=1}^{n} x_i \log \left[\left(\frac{1-p}{1-p^{\beta+1}} \right) \right] \end{aligned} \tag{12}$$

differentiating Eq.(12) partially with respect to the shape parameters p and β of NBDETE distribution we get the likelihood equations as

$$\begin{aligned}\frac{\partial \ell}{\partial p} =& \frac{(1+\beta)(\sum_{i=1}^{n} x_i + n)p^{\beta}}{1-p^{\beta+1}} \\ &+ \frac{n(1-p^{\beta}-\beta p^{\beta})}{p(1-p^{\beta})} - \frac{\sum_{i=1}^{n} x_i}{1-p} \\ =& 0\end{aligned} \tag{13}$$

and

$$\begin{aligned}\frac{\partial \ell}{\partial \beta} =& \frac{p^{\beta+1}(\sum_{i=1}^{n} x_i + n)\log p}{1-p^{\beta+1}} - \frac{np^{\beta}\log p}{1-p^{\beta}} \\ =& 0\end{aligned} \tag{14}$$

The solutions of Eq.s (13), and (14) provide the MLEs of p and β, which can be obtained by numerical methods. Since the MLE of the vector of unknown parameters $\theta = (p,\ \beta\ )^T$ cannot be derived in closed forms, it is, therefore, hard to derive the exact distribution of the MLEs.
The second partial derivatives are given below

$$\begin{aligned}\frac{\partial^2 \ell}{\partial p^2} =& \frac{n\left[1-\left(2+\beta+\beta^2\right)p^{\beta}+(1+\beta)p^{2\beta}\right]}{p^2\left(1-p^{\beta}\right)^2} \\ &+ \frac{(\sum_{i=1}^{n} x_i + n)p^{\beta}\left(\beta(\beta+1)p+(\beta^2+1\right)p^{\beta}}{\left(1-p^{\beta+1}\right)^2} \\ &- \frac{\sum_{i=1}^{n} x_i}{(1-p)^2}\end{aligned}$$

$$\frac{\partial^2 \ell}{\partial \beta^2} = (\log p)^2 p^{\beta}\left[\frac{(\sum_{i=1}^{n} x_i + n)p}{\left(1-p^{\beta+1}\right)^2} - \frac{n}{(1-p^{\beta})^2}\right]$$

$$\begin{aligned}\frac{\partial^2 \ell}{\partial p \partial \beta} =& \frac{(\sum_{i=1}^{n} x_i + n)p^{\beta}\left\{1-p^{\beta+1}+(\beta+1)\log p\right\}}{\left(1-p^{\beta+1}\right)^2} \\ &- \frac{np^{\beta}\left[1-p^{\beta}+\beta\log p\right]}{(1-p^{\beta})^2}\end{aligned}$$

The asymptotic distribution of the MLE $\hat{\theta}$ was introduced by Lawless [10] as

$$\left(\hat{\theta}-\theta\right) \rightarrow N\left(0,\ I^{-1}(\theta)\right)$$

where $I^{-1}(\theta)$ is the inverse Fisher's information matrix of unknown parameters $\theta = (p,\ \beta)^T$ as follows

$$I_{Y(p,\beta)}(\theta) = \begin{bmatrix} -E\left(\frac{\partial^2 \ell}{\partial p^2}\right) & -E\left(\frac{\partial^2 \ell}{\partial p \partial \beta}\right) \\ -E\left(\frac{\partial^2 \ell}{\partial p \partial \beta}\right) & -E\left(\frac{\partial^2 \ell}{\partial \beta^2}\right) \end{bmatrix}$$

On the other hand, Fisher's information matrix can be computed by using the approximation

$$I_Y\left(\hat{\theta}\right) = \begin{bmatrix} -\left.\frac{\partial^2 \ell}{\partial p^2}\right|_{(\hat{p},\hat{\beta})} & -\left.\frac{\partial^2 \ell}{\partial p \partial \beta}\right|_{(\hat{p},\hat{\beta})} \\ -\left.\frac{\partial^2 \ell}{\partial p \partial \beta}\right|_{(\hat{p},\hat{\beta})} & -\left.\frac{\partial^2 \ell}{\partial \beta^2}\right|_{(\hat{p},\hat{\beta})} \end{bmatrix}$$

where $\hat{p}$ and $\hat{\beta}$ are the MLEs of p and β respectively.

5 Simulation

In this section, the Monte Carlo simulation is performed to assess the performance of the accuracy of point estimates for $NBDETE(p, \beta)$ distribution using MLE with 10000 replications. The simulation studies were conducted with sample sizes of n = 30, 100, 300, and 500 and we considered some different values for the parameters p and β. The evaluation of point estimation was performed based on the empirical mean and the mean squared error (MSE). The formula of Eq.(11) is used to generate the sample data from $NBDETE(p, \beta)$ while the R *mledist* function of the *fitdistrplus* package is utilized to perform the MLE calculations. The empirical results are given in Table 3. All results in Table 3 indicate that the estimates are close to the true values of the parameters for these sample sizes. Moreover, as the sample size increases, the MSEs decrease as expected, which means that the maximum likelihood method can be used effectively for estimating the parameters of the NBDETE distribution.

Table 3: Simulation results (Empirical means and the MSEs)

	p=0.1, β=20		p=0.1, β=25		p=0.1, β=30	
n	$\hat{p}$	$\hat{\beta}$	$\hat{p}$	$\hat{\beta}$	$\hat{p}$	$\hat{\beta}$
30	0.1026 (0.000348)	21.1677 (1.570335)	0.1031 (0.000343)	26.4457 (2.410703)	0.1038 (0.000359)	29.5005 (0.589064)
100	0.1013 (0.000092)	21.1047 (1.358507)	0.101 (0.000104)	26.3298 (1.915543)	0.1009 (0.000091)	29.482 (0.535464)
300	0.1004 (0.000031)	21.0474 (1.208421)	0.1002 (0.000033)	26.2692 (1.73825)	0.1003 (0.000031)	29.4581 (0.531226)
500	0.1004 (0.000017)	21.0285 (1.139734)	0.1003 (0.000018)	26.2725 (1.728282)	0.1003 (0.000017)	29.4331 (0.511755)
	p=0.2, β=20		p=0.2, β=25		p=0.2, β=30	
n	$\hat{p}$	$\hat{\beta}$	$\hat{p}$	$\hat{\beta}$	$\hat{p}$	$\hat{\beta}$
30	0.2063 (0.001135)	23.1912 (16.88861)	0.2065 (0.001176)	26.6218 (3.095766)	0.2055 (0.001207)	29.7337 (0.650176)
100	0.2022 (0.00034)	23.1866 (10.87487)	0.2017 (0.000331)	26.4341 (2.245908)	0.2016 (0.000332)	29.7645 (0.547142)
300	0.2007 (0.000111)	23.0591 (9.735587)	0.2008 (0.000108)	26.2904 (1.8049)	0.2003 (0.000107)	29.9211 (0.452541)
500	0.2004 (0.000064)	22.8224 (8.345134)	0.2007 (0.000059)	26.2847 (1.772303)	0.2 (0.000069)	29.7852 (0.403019)

Note:The MSE is between parentheses

6 Application

This section demonstrates an application of the proposed distribution NBDETE. Therefore, to assess the goodness of fit for NBDETE and to examine its performance compared with some related distributions, NBDETE is fitted to right-skewed over-dispersed real lifetime count data set. This data set was reported by Consul and Jain [11] explaining the accidents to 647 women working on Shells for 5 weeks and it is extremely right-skewed (2.1212) and over-dispersed since the sample variance (0.69190) is greater than the respective sample mean (0.46522). The data set was recently used by Nekoukhou et al. [12], Gomez-Deniz [13], and El-Alosey & Eledum [14] in the application of generalization of the geometric distribution, discrete generalized exponential distribution of a second type (DGE2), and discrete Extended Erlang-Truncated Exponential, respectively. To compare the proposed NBDETE distribution with its related distributions involving discrete Erlang-Truncated Exponential (DETE) [5], Generalized Rayleigh distribution (DGR) [15], discrete Burr (DBD) [16], and DGE_2, we used -log-likelihood (-log(L)), and the χ^2 (chi-square) statistic as criteria for comparison. Note that the R functions *mledist* and *chisq.test* are used to perform the calculations of the MLE and chi-square, respectively.

Table 4 demonstrates the observed values and the corresponding expected values computed using each distribution, while Table 5 explains the MLE estimators, -LogL, and the χ^2 statistics with the corresponding p-values.

Table 4: The observed and expected values of the data set

Count	Observed	NBDETE	DETE	DGR	DBD	DGE_2
0	447	441.7	441.7	448.12	447.44	446.92
1	132	140.3	140.3	122.6	142.01	133.63
2	42	44.62	44.62	53	35.97	44.43
3	21	14.24	14.24	18.15	12.65	15.02
4	3	4.6	4.6	4.41	5.77	5.16
>5	2	1.54	1.54	0.72	3.16	1.84
Total	647	647	647	647	647	647

Table 5: Parameters estimates, -Log L, χ^2 statistic and p-value for the selected distributions of the data set

Model	MLEs		-Log L	statistic	p-value
NBDETE	$\hat{\beta}$=54.467	$\hat{p}$=0.6830	592.479	4.612	0.465
DETE	$\hat{\beta}$=0.417	$\hat{p}$=0.0639	592.479	4.613	0.4649
DGR	$\hat{\alpha}$=0.2196	$\hat{p}$=0.8123	592.544	6.172	0.2899
DBD	$\hat{\alpha}$=1.642	$\hat{\theta}$=0.1841	597.955	8.979	0.1099
DGE_2	$\hat{\alpha}$=0.898	$\hat{p}$=0.3379	592.183	3.448	0.6312

From the results in Table 5, it is observed that the values of -logL for the proposed NBDETE besides DETE is (592.479) which is the most minimum among the other related distributions (the smaller the better). On the other hand, this value together with the values of χ^2 statistic and their corresponding p-values explains that the proposed NBDETE and DETE distribution are the most appropriate model to fit this data set. Moreover, all the studied distributions are appropriate for fitting this data set.

7 Conclusions

This paper developed a new two parameters mixture of negative binomial distribution called the negative binomial-discrete Erlang-truncated exponential (NBDETE) distribution, which is created by combining the negative binomial with the zero-truncated discrete Erlang-truncated exponential distribution. We look at some of NBDETE's statistical properties and use the maximum likelihood method to estimate its parameters. The pmf of NBDETE is a right-skewed decreasing function while the hazard function is a constant. The proposed NBDETE is fitted to a right-skewed over-dispersed real lifetime count data set, while the Monte Carlo simulation is performed to assess the performance of the accuracy of point estimates for the distribution using MLE. The application together with the Monte Carlo simulation revealed that; the proposed distribution is the most convenient model to fit a right-skewed over-dispersed real lifetime count data set., further, the maximum likelihood method can be used effectively for estimating the parameters of the NBDETE distribution. We recommend the proposed distribution for modeling the over-dispersed right-skewed real-life count data sets adequacy.

Availability of data and materials

Readers who are interested in contacting the authors can do so.

Conflicts of Interest

There are no conflicts of interest declared by the authors for the publication of this paper..

Financial funding and support

There is no funding for this project to report.

References

[1] E. Gomez-Deniz, J. Sarabia, and E. Calderin-Ojeda, Univariate and multivariate versions of the negative binomial-inverse Gaussian distributions with applications, *Insurance: Mathematics and Economics*, **42(1)**, 39-49 (2008), https://doi.org/10.1016/j.insmatheco.2006.12.001.

[2] C. Pudprommarat and W. Bodhisuwan, A new mixed negative binomial distribution, *Journal of Applied Sciences*, **12**, 1853-1858 (2012), doi: 10.3923/jas.2012.1853.1858.

[3] R. Shanker, K. K. Shuklam and R. Shanker, A quasi Negative binomial- Discrete Akash distribution and its applications to ecology, *International Journal of Statistics and Applied Mathematics*, **3(3)**, 111-119 (2018).

[4] K. Ravikumar, M. Thomas and C. B. Prasanth, Negative Binomial Logistic Distribution, *Turkish Journal of Computer and Mathematics Education*, **12(11)**, 5963-5976 (2021).

[5] A. R. El-Alosey, Discrete Erlang-Truncated Exponential Distribution, *International Journal of Statistics and Applied Mathematics*, **6(1)**, 230-236 (2021).

[6] S. Kongrod, W. Bodhisuwan and P. Payakkapong, The negative binomial-Erlang distribution with applications, *International Journal of Pure and Applied Mathematics*, **92(3)**, 389-401 (2014), doi: http://dx.doi.org/10.12732/ijpam.v92i3.7.

[7] M. Shrahili, I. Elbatal, I. Muhammad and M. Muhammad, Properties and applications of beta Erlang-truncated exponential distribution, *Journal of Mathematics and Computer Science*, **22**, 16-37 (2021).

[8] S. Kotz, M. Lumelskii and M. Pensky, *The Stress-Strength Model and its Generalizations: Theory and Applications*, Work Scientific, New-York (2003).

[9] L. Ventura and W. Racugno, Recent advances on Bayesian inference for $P(X<Y)$, *Bayesian analysis*, **6(2)**,1–75 (2011).

[10] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley and Sons, New York, 2nd edition (2002).

[11] P. C. Consul and G. C. Jain, A Generalization of the Poisson Distribution, *Technometrics*, **15**, 791-799 (1973),DOI: 10.1080/00401706.1973.10489112.

[12] V. Nekoukhou, M. H. Alamatsaz and H. Bidram, Discrete generalized exponential distribution of a second type, *Statistics. A Journal of Theoretical and Applied Statistics*, **47(4)**, 876-887(2013), doi: 10.1080/02331888.2011.633707.

[13] E. Gomez-Deniz, Another generalization of the geometric distribution. *TEST*, **19(2)**, 399-415(2010), doi: 10.1007/s11749-009-0169-3.

[14] A. R. El-Alosey and H. Eledum, Discrete Extended Erlang-Truncated Exponential distribution and its Applications, *Applied Mathematics & Information Sciences*, **16(1)**, 127-138 (2022), doi: http://dx.doi.org/10.18576/amis/160113.

[15] M. H. Alamatsaz, S. T. Dey and S. S. Harandi, Discrete generalized Rayleigh distribution, *Pakistan journal of statistics*, **32(1)**, 1-20 (2016).

[16] H. Krishna and P. S. Pundir, Discrete Burr and discrete Pareto distributions, *Statistical Methodology*, **6(1)**, 177-188 (2009), doi:10.1016/j.stamet.2008.07.001.