frontiers | Frontiers in Applied Mathematics and Statistics

TYPE Original Research
PUBLISHED 20 March 2023
DOI 10.3389/fams.2023.1137036

OPEN ACCESS

EDITED BY
Fabrizio Maturo,
Universitas Mercatorum, Italy

REVIEWED BY
Fuxia Cheng,
Illinois State University, United States
Abdelfateh Beghriche,
Université Frères Mentouri Constantine 1, Algeria
Seghier Fatma Zohra,
University of Skikda, Algeria

*CORRESPONDENCE
Halim Zeghdoudi
✉ halimzeghdoudi77@gmail.com

SPECIALTY SECTION
This article was submitted to Statistics and Probability, a section of the journal Frontiers in Applied Mathematics and Statistics

RECEIVED 03 January 2023
ACCEPTED 23 February 2023
PUBLISHED 20 March 2023

CITATION
Benchettah MH, Zeghdoudi H and Raman V (2023) On composite length-biased exponential-Pareto distribution: Properties, simulation, and application in actuarial science . *Front. Appl. Math. Stat.* 9:1137036. doi: 10.3389/fams.2023.1137036

On composite length-biased exponential-Pareto distribution: Properties, simulation, and application in actuarial science

Moulouk Halima Benchettah[1], Halim Zeghdoudi[1]* and Vinoth Raman[2]

[1]LaPS Laboratory, Badji-Mokhtar University, Annaba, Algeria, [2]Department of Quality Measurement and Evaluation, Deanship of Quality and Academic Accreditation, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

The composite length-biased exponential-Pareto (CLBEP) distribution is a new composite distribution that is introduced in this article. This model's probability density function, moments, and quantiles, among other statistical characteristics, are determined mathematically. The parameters' maximum-likelihood estimation and stochastic ordering are discussed. A comparison study with other new composite and conventional distributions is also included. Specifically, using two actual fire insurance data sets, the goodness of fit of this new model is contrasted with the composite exponential-Pareto, composite lognormal-Pareto, and composite Rayleigh-Pareto distributions (Algerian and Danish fire insurance losses).

2010 AMS subject classifications: 62E10; 60E05.

KEYWORDS

composite distribution, Pareto distribution, length-biased exponential, maximum-likelihood estimation, quantile function

1. Introduction

Currently, digital methods are being used in the fields of biology, economics, physical sciences, statistical sciences, and other fields. In the applications of other fields as well as in daily life, the statistical sciences are essential. Probability distributions are frequently the foundation of statistical science because many problems in these fields frequently do not follow one of the fundamental probability distributions. Actuarial science and finance generally use common distributions to express their data on payments, quantity and number of claims, and premium computation. Examples of these distributions are exponential, Poisson, length-biased exponential, and Pareto.

The length-biased exponential distribution, on the other hand, offers a wide range of practical applications in several industries (reliability, actuarial science, survival analysis, and mathematical financiers). The lifetime of a phenomenon with no memory, no aging, no wear and tear, or the profits of an insurance company, or various models of surpluses and financial assets, are frequently modeled using the length-biased exponential distribution.

The modeling of unimodal insurance loss data with a long tail appeals to actuaries. Distributions that may replicate the heavy tail of insurance loss data are necessary to provide a sufficiently precise estimate of the degree of connected business risk, including gamma, Pareto, length-biased exponential, Rayleigh lognormal, and Weibull.

For example, if there are both modest and significant losses, insurance companies may experience losses. When modeling very large losses, practitioners seem to favor the Pareto distribution for size distribution. Length-biased exponential, lognormal, Rayleigh, or Weibull models are preferred when the losses are composed of smaller values with high frequencies and larger losses with low frequencies [1]. Nevertheless, no conventional size model can simultaneously account for losses that are both minor and significant. Unlike length-biased, lognormal, Rayleigh, or Weibull exponential models, which have a positive general fit but fit the tail poorly, Pareto models actually fit the tail well.

When modeling data that have heavy tails, the composite distributions appear appropriate. For instance, the one-parameter exponential-Pareto (exp-Pareto) model and the one-parameter inverse gamma-Pareto (IG-Pareto) model have both been proposed as potential models for the modeling of insurance data. When they are fitted to well-known insurance data sets, such as the Danish fire insurance data set, they still are unable to perform satisfactorily. So, the model needs to be improved. By exponentiation of the random variable linked to the probability density function (pdf) of an inverse gamma-Pareto distribution, Liu and Ananda [2] suggested an improved version of the one-parameter IG-Pareto model. Their suggested model outperformed the original model significantly across several data sets. Furthermore, there are other composite models such as the composite lognormal-Pareto (cLP) model (see Scollnik [3] and composite Rayleigh-Pareto (cRP) model (see Benatmane et al. [4]). For more details see [5–12].

As a result, we suggest, in this study, a novel composite distribution that blends length-biased and Pareto exponential distributions. This effort aims to introduce a new composite distribution. As a result, the CLBEP distribution has a single parameter. It is simple to determine mathematical qualities in an explicit form. Due to its composition (two types of distributions that can be simulated for survival analysis and actuarial purposes), this new distribution offers advantages. Many real-life data sets can be analyzed using the CLBEP model, which provides suitable fits to these data sets.

The current article is structured as follows: The composite length-biased exponential-Pareto distribution and some of its statistical characteristics are discussed in Section 2. The estimation of parameters is addressed in Section 3. A numerical example with a comparison of various classical and composite models using two real data sets is provided in Section 4.

2. Formulation of the CLBEP distribution

For many theoretical issues, the length-biased exponential and Pareto distributions might not be adequate. We created the composite length-biased exponential-Pareto (CLBEP) distribution, based on the composite transformation, to have a flexible model. Let T be an arbitrary random variable with density function

$$f(t;\theta)=\begin{cases} cf_1(t) & 0<t\le\theta \\ cf_2(t) & \theta\le t<\infty \end{cases}$$

where f_1 is a length-biased exponential density, f_2 is a two-parameter Pareto density, and c is the normalizing constant. Hence,

$$f_1(t)=\frac{t}{\lambda^2}\exp\left(-\frac{t}{\lambda}\right), 0<t\le\infty,$$

$$f_2(t)=\frac{\alpha\theta^{\alpha}}{t^{\alpha+1}}, t>\theta,$$

where λ, α, and θ are unknown non-negative parameters. To obtain a composite smooth density function, we use the continuity and differentiability conditions at the threshold point θ, i.e.,

$$\begin{cases} f_1(\theta)=f_2(\theta) \\ \frac{d}{dt}f_1(\theta)=\frac{d}{dt}f_2(\theta). \end{cases}$$

These two restrictions give

$$\begin{cases} \frac{\theta}{\lambda^2}\exp\left(-\frac{\theta}{\lambda}\right)=\frac{\alpha}{\theta} \\ \frac{(\lambda-\theta)}{\lambda^3}\exp(-\frac{\theta}{\lambda})=\frac{-\alpha(\alpha+1)}{\theta^2}. \end{cases}$$

After some calculation, we get

$$\begin{cases} 2-\beta+\left(\beta^2\exp(-\beta)\right)=0 \\ \alpha=\beta^2\exp(-\beta) \end{cases} \text{with } \beta=\frac{\theta}{\lambda}.$$

Using the numerical methods, we find

$$\begin{cases} \beta=2.5118 \\ \alpha=0.51181. \end{cases}$$

To find the normalizing constant, we use the density condition ($\int_0^\infty f(t,\theta)dt=1$), which has

$$\begin{aligned} c &= \frac{1}{e^{-\frac{\theta}{\lambda}}\left(\frac{\theta}{\lambda}+1\right)+2} \\ &= \frac{1}{e^{-\beta}\left(\beta+1\right)+2}=0.43766. \end{aligned}$$

Since $f(t;\theta)$ can be expressed as

$$f(t;\theta)=\begin{cases} \frac{2.7613t}{\theta^2}\exp\left(\frac{-2.5118t}{\theta}\right) & 0<t\le\theta \\ \frac{0.224}{t^{1.51181}}\theta^{0.51181} & \theta\le t<\infty. \end{cases} \quad (1)$$

2.1. Statistical properties of the CLBEP distribution

In this subsection, many statistical properties are presented, such as the behavior of PDF and quantile function, as well as the moments and stochastic ordering.

$$\lim_{t\to 0}f(t;\theta)=0 \text{ and } \lim_{t\to\infty}f(t;\theta)=0.$$

The following proposition states that there is one shape for the PDF of the CLBEP distribution. Furthermore, the plots of PDF for some parameter value of the proposed model are presented in Figure 1.

Proposition 1. The PDF $f(t;\theta)$ in Equation (1) of the CLBEP distribution is unimodal for $\theta > 0$.

Proof. *The first derivative of $f(t;\theta)$ is*

$$\frac{df(t;\theta)}{dt} = \begin{cases} -\frac{0.0002}{\theta^3}\exp\left(-2.5118\frac{t}{\theta}\right)(34679t - 138070\theta) & 0 < t \leq \theta \\ -\frac{0.33865}{t^{2.5118}}\theta^{0.51181} & \theta \leq t < \infty. \end{cases}$$

The CLBEP distribution is unimodal with maximum value at the point $\hat{t} = 0.398\,14\theta$, where the unique mode is $t_{mod} = 0.398\,14\theta$.

2.2. Cumulative distribution function and moments of the CLBEP distribution

The cumulative distribution function (c.d.f.) of this composite model is

$$F(t;\theta) = \begin{cases} 0.43768\left(1 - (1 + \frac{2.5118t}{\theta})\exp\left(\frac{-2.5118t}{\theta}\right)\right) & 0 < t \leq \theta \\ 1 - 0.43768\left(\frac{\theta}{t}\right)^{0.51181} & \theta \leq t < \infty \end{cases} \quad (2)$$

The *kth* moment about the origin of the CLBEP distribution can be obtained as:

$$E\left(T^k\right) = -\frac{0.43768\theta^k}{(2.5118)^k}\left(\Gamma\left(k+2, 2.5118\right) - \Gamma\left(k+2\right)\right) + 0.2240^{0.51181}\left(\frac{\mathbf{x}^{k-1.5118}}{k-1.5118}\right)\Bigg|_{\theta}^{\infty},$$

which $E\left(T^k\right) = \infty$ (infinite), for $k \geq 2$.

The mean of the CLBEP distribution is given by

$$E(T) = 0.160\,03\theta + 0.437\,67.$$

2.3. The quantile function of the CLBEP distribution

The quantile function of the CLBEP distribution is given in the following theorem.

Theorem 1. *The quantile function of the CLBEP distribution is*

$$F_T^{-1}(u) = \begin{cases} \frac{-\theta}{2.5118} - \frac{\theta}{2.5118}W_{-1}\left(\frac{u}{0.43768e} - e^{-1}\right) & \text{if } 0 < u < u_0 \\ \theta\left(\frac{0.43768}{(1-u)}\right)^{1.9539} & \text{if } u_0 < u < 1 \end{cases}$$

where $u_0 = 0.43768$.

Proof. *For $u \in]0; u_0[$, we have to solve the equation $F(t) = u$ with respect to t, $t > 0$*

$$0.43768\left(1 - (1 + \frac{2.5118t}{\theta})\exp\left(\frac{-2.5118t}{\theta}\right)\right) = u$$

$$-(1+\frac{2.5118t}{\theta})\exp\left(\frac{-2.5118t}{\theta}\right) = \frac{u}{0.43768} - 1$$

Multiplying by e^{-1} both sides, we find

$$-(1+\frac{2.5118t}{\theta})\exp\left(\frac{-2.5118t}{\theta}\right)e^{-1} = \left(\frac{u}{0.43768} - 1\right)e^{-1}$$

$$W(z)exp(W(z)) = z$$

We see that $-(1 + \frac{2.5118t}{\theta})$ is the Lambert W function of the real argument $(\frac{u}{0.43768} - 1)e^{-1}$. Then, we have

$$W\left(\frac{u}{0.43768e} - e^{-1}\right) = -1 - \frac{2.5118t}{\theta}. \quad (3)$$

Moreover, for any $\theta, t > 0$, it is immediate that $-(1 + \frac{2.5118t}{\theta}) < 0$, and it can also be checked that $(\frac{u}{0.43768e} - e^{-1}) \in \left]\frac{-1}{e}; 0\right[$ since $u \in]0; u_0[$

Therefore, by taking into account the properties of the negative branch of the Lambert W function, Equation (3) becomes

$$W_{-1}\left(\frac{u}{0.43768e} - e^{-1}\right) = -\left(1 + \frac{2.5118t}{\theta}\right).$$

Finally, $\forall\theta > 0$, $t = F_T^{-1}(u)$,

$$F_T^{-1}(u) = \frac{-\theta}{2.5118} - \frac{\theta}{2.5118}W_{-1}\left(\frac{u}{0.43768e} - e^{-1}\right), \text{where } 0 < u < u_0.$$

Now, for $u \in]u_0; 1[$, we have to solve the equation $F(t) = u$ with respect to t, $t > 0$

$$1 - 0.43768\left(\frac{\theta}{t}\right)^{0.51181} = u$$

it is easy to find

$$F_T^{-1}(u) = \theta\left(\frac{0.43768}{(1-u)}\right)^{1.9539} \quad \text{where } u_0 < u < 1.$$

2.4. Stochastic ordering

Consider two random variables Z_1 and Z_2. Then, Z_1 is said to be smaller than Z_2 in the following cases:

1. Stochastic order ($Z_1 <_S Z_2$), if $F_{Z_1}(t) < F_{Z_2}(t), \forall t$.
2. Convex order ($Z_1 \leq_{cx} X_2$), if for all convex functions Ψ and provided expectation exists, $E[\Psi(Z_1)] \leq E[\Psi(Z_2)]$.
3. Hazard rate order ($Z_1 \leq_{hr} Z_2$), if $h_{Z_1}(t) \geq h_{Z_2}(t), \forall t$.
4. Likelihood ratio order ($Z_1 <_{lr} Z_2$), if $\frac{f_{Z_1}(t)}{f_{Z_2}(t)}$ is decreasing in t.

Remark 1. Likelihood ratio order $\Rightarrow$ hazard rate order $\Rightarrow$ stochastic order, If $E[Z_1] = E[Z_2]$, then convex order $\Leftrightarrow$ stochastic order.

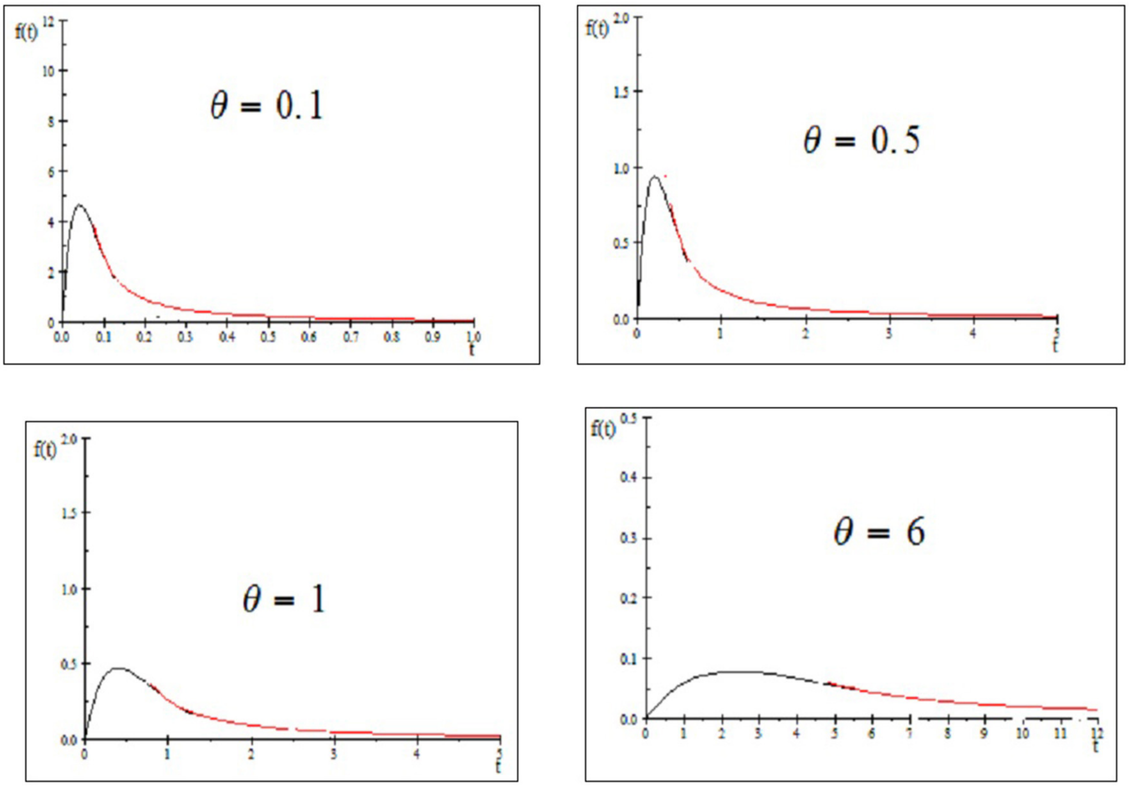

FIGURE 1
The plots of PDF for some parameter value of θ.

Theorem 2. Let $Z_i \backsim CLBEP$ distribution (θ_i); $i = 1, 2$ be two random variables. If $\theta_1 \leq \theta_2$, then $Z_1 <_{lr} Z_2, Z_1 <_{hr} Z_2; Z_1 <_S Z_2$ and $Z_1 \leq_{cx} Z_2$.

Proof.

Case I: $0 < t \leq \theta$

We have

$$\frac{f_{Z_1}(t;\theta_1)}{f_{Z_2}(t;\theta_2)} = \frac{\theta_2^2}{\theta_1^2}\exp\left(\frac{-2.511\,8t}{\theta_1} + \frac{2.511\,8t}{\theta_2}\right).$$

Using the $\ln\left(\frac{f_{Z_1}(t;\theta_1)}{f_{Z_2}(t;\theta_2)}\right)$ for simplification, we can find

$$\frac{d}{dt}\ln\left(\frac{f_{Z_1}(t;\theta_1)}{f_{Z_2}(t;\theta_2)}\right) = \frac{2.5118\theta_2^2}{\theta_1^2}\left(\frac{\theta_1-\theta_2}{\theta_1\theta_2}\right).$$

To this end, if $\theta_1 \leq \theta_2$, we have $\frac{d}{dt}\ln\left(\frac{f_{Z_1}(t;\theta_1)}{f_{Z_2}(t;\theta_2)}\right) \leq 0$. This means that $Z_1 <_{lr} Z_2$.

Case II: $\theta \leq t < \infty$

We have

$$f_{Z_1}(t;\theta_1) - f_{Z_2}(t;\theta_2) = \frac{0.261\,48}{t^{1.511\,81}}\left(\theta_1^{0.51181} - \theta_2^{0.51181}\right).$$

We can see, if $\theta_1 \leq \theta_2$, then $f_{Z_2}(t;\theta_2) \leq f_{Z_2}(t;\theta_2)$. Furthermore, according to Remark 1, the theorem is proved.

3. Generating random values from the CLBEP distribution

3.1. Parameter estimation

In this section, we will introduce two methods of estimating the unknown parameter θ.

3.1.1. An *ad hoc* procedure based on percentiles

The following *ad hoc* procedure provides a closed form for the parameter θ, estimated using percentiles. Let $t_1, t_2, ..., t_n$ be a random sample from the CLBEP model. Assume that $t_1 \leq t_2 \leq ... \leq t_n$ and $t_m \leq \theta \leq t_{m+1}$. Based on percentiles, the parameter θ can be estimated, as the *pth* percentile, where $p = F(\theta)$

$$p = 0.43768\left(1 - (1 + \frac{2.5118\theta}{\theta})\exp\left(\frac{-2.5118\theta}{\theta}\right)\right) = 0.312\,99.$$

According to Klugman et al. [1], we have a smooth empirical estimate of the *pth* percentile given by

$$\hat{\theta} = (1-h)t_m + ht_{m+1}$$

with

$$\begin{cases} m = [(n+1)p] \\ h = (n+1)p - m. \end{cases} \quad (4)$$

The Pareto distribution or the length-biased exponential distribution will be a more superior model than the composite length-biased exponential-Pareto distribution according as $\hat{\theta}$ is closer to t_1 or t_n.

3.1.2. Maximum-likelihood estimation

Assume again that $t_1 \leq t_2 \leq ... \leq t_n$ and $t_m \leq \theta \leq t_{m+1}$. Then, the likelihood function is

$$L(t_1, ..., t_n; \theta) = \prod_{i=1}^{n} f(t_i) = \prod_{i=1}^{m} f_1(t_i) \prod_{i=m+1}^{n} f_2(t_i)$$

$$= \prod_{i=1}^{m} 2.7613 \frac{t_i}{\theta^2} \exp\left(-2.51181 \frac{t_i}{\theta}\right) \prod_{i=m+1}^{n} \frac{(0.224)}{t_{i.}^{1.51181}} \theta^{0.51181}$$

$$= (2.7613)^m (0.224)^{n-m} \theta^{0.51181(n-m)-2m} e^{\frac{-2.51181}{\theta} \sum_{i=1}^{m} t_i} \frac{\prod_{i=1}^{m} t_i}{\prod_{i=m+1}^{n} t_i^{1.51181}}$$

$$= k\theta^{0.51181n - 2.51181m} e^{\left(\frac{-2.51181}{\theta} \sum_{i=1}^{m} t_i\right)}.$$

with

$$k = (2.7613)^m (0.224)^{n-m} \frac{\prod_{i=1}^{m} t_i}{\prod_{i=m+1}^{n} t_i^{1.51181}}.$$

Define $\ln L = \ln k + (0.51181n - 2.51181m) \ln \theta + \left(\frac{-2.51181}{\theta} \sum_{i=1}^{m} t_i\right)$.

Differentiating $\ln L$ with respect to θ gives

$$\frac{d \ln L}{d\theta} = \frac{0.51181n - 2.51181m}{\theta} + \frac{2.51181 \sum_{i=1}^{m} t_i}{\theta^2}.$$

Hence, the solution of the likelihood equation $\frac{d \ln L}{d\theta} = 0$ is

$$\hat{\theta} = \frac{2.51181 m \bar{t}_m}{2.251181m - 0.51181n}, \text{ if } \frac{n}{m} \neq 4.9 \text{ and } \bar{t}_m = \frac{\sum_{i=1}^{m} t_i}{m}. \quad (5)$$

Since this estimator requires the value of m, we recommend the following algorithm (see Teodorescu and Vernic [13]):

4. Numerical and application examples

In this section, the estimation procedure described in Section 3 has been explained using two data samples generated from the *CLBEP* model. The generating algorithm used is based on the inversion of the c.d.f. (Equation 2).

```
Step 1. Estimate m as above, from Equation (4).
Step 2. Evaluate θ̂ from Equation (5).
Step 3. Verify if θ̂ is in between t_m ≤ θ̂ ≤ t_{m+1}. If
so, then θ̂ is the MLE. If not, use Algorithm 2.
An alternative algorithm would be to replace Step
1 with the consideration of all possible values
for m and the achievement for each of them of the
verification of step 3:
```

Algorithm 1. Estimate θ using MLE.

```
Step 1. For each m (m = 1,2,...,n − 1), evaluate θ̂_m
from Equation (4). Check if θ̂_m is in between t_m ≤
θ̂ ≤ t_{m+1}. If yes, then θ̂_m is the MLE. If no, go to
next m.
Step 2. If there is no solution to θ, try an
alternative model.
```

Algorithm 2. Estimate θ using percentiles.

TABLE 1 Test for $\theta = 5$.

Classes	n_i	f_i	*Theoretical freq.*, p_i	$\frac{n(f_i - p_i)^2}{p_i}$ (CLBEP)
[0, 2)	16	0.1480	0.1358	0.1184
[2, 4)	22	0.2037	0.1688	0.7748
[4, 6)	13	0.1203	0.1060	0.2053
[6, 10)	14	0.1296	0.1070	0.5120
[10, 60)	27	0.25	0.2151	0.6115
[60, 7000)	16	0.1480	0.1306	0.2507
$\sum$	108	1	χ^2*distance* :	2.4730

4.1. Example

The data set given in this subsection, consisting of 108 values, was sampled from a length-biased exponential-Pareto population with parameter $\theta = 5$ (see Table 1 in the Appendix).

The estimated values of the parameter are:

- By Algorithm 1, $m = 39 : \hat{\theta}_1 = 5.053\,6$.
- By Algorithm 2, MLE Step 1: $\hat{\theta}_2 = 4.9812$.
- By Algorithm 2, MLE Step 2 : $\hat{\theta}_3 = 4.9810$.

We notice that Algorithm 2 in Step 1 gives a more accurate value. We also applied the χ^2 test to check the distribution fitting, and the results for $\hat{\theta}_3$ are given in Tables 1–4.

The χ^2 distances calculated for all the estimated values of the parameters are

$$\chi^2 distance\left(\hat{\theta}_1\right) = 2.4591$$
$$\chi^2 distance\left(\hat{\theta}_2\right) = 2.4819$$
$$\chi^2 distance\left(\hat{\theta}_3\right) = 2.4818.$$

The χ^2 test accepts the length-biased exponential Pareto model for all values of the parameter as expected, which $d^2\left(\hat{\theta}_1\right)$ is a minimum.

TABLE 2 Test for $\hat{\theta}_1 = 5.0536$.

Classes	n_i	f_i	*Theoretical freq.,* p_i	$\frac{n(f_i-p_i)^2}{p_i}(CLBEP)$
$[0, 2)$	16	0.1480	0.1338	0.1628
$[2, 4)$	22	0.2037	0.1679	0.8223
$[4, 6)$	13	0.1203	0.10648	0.1931
$[6, 10)$	14	0.1296	0.1076	0.4833
$[10, 60)$	27	0.25	0.2162	0.5681
$[60, 7000)$	16	0.1480	0.1314	0.2291
$\sum$	108	1	$\chi^2 distance:$	2.4591

TABLE 3 Test for $\hat{\theta}_2 = 4.9812$.

Classes	n_i	f_i	*Theoretical freq.,* p_i	$\frac{n(f_i-p_i)^2}{p_i}(CLBEP)$
$[0, 2)$	16	0.1480	0.1365	0.1045
$[2, 4)$	22	0.2036	0.1692	0.7587
$[4, 6)$	13	0.1202	0.1059	0.2098
$[6, 10)$	14	0.1295	0.1068	0.5226
$[10, 60)$	27	0.25	0.2146	0.6275
$[60, 7000)$	16	0.1480	0.1304	0.2585
$\sum$	108	1	$\chi^2 distance:$	2.4819

TABLE 4 Test for $\hat{\theta}_3 = 4.9810$.

Classes	n_i	f_i	*Theoretical freq.,* p_i	$\frac{n(f_i-p_i)^2}{p_i}(CLBEP)$
$[0, 2)$	16	0.1480	0.1366	0.1043
$[2, 4)$	22	0.2035	0.1692	0.7587
$[4, 6)$	13	0.1202	0.10594	0.2099
$[6, 10)$	14	0.1295	0.1068	0.5225
$[10, 60)$	27	0.25	0.2146	0.6275
$[60, 7000)$	16	0.1480	0.1304	0.2585
$\sum$	108	1	$\chi^2 distance:$	2.4817

4.2. Goodness of fit

In this subsection, we apply the composite length-biased exponential-Pareto model to two real insurance data sets.

Data set I: is 100 Algerian (SAA company) fire insurance losses (see Appendix).

We provide in Table 5 the estimated value of fitted models and the values of the $-LL, AIC, AICc$, and BIC evaluated at the maximum-likelihood estimators.

Data set II: is 2, 156 Danish fire insurance losses.

We use the same analysis, we find

Tables 5, 6 indicate that the CLBEP model outperforms classical distributions, composite Rayleigh-Pareto, composite exponential-Pareto, and composite lognormal-Pareto models in terms of $-LL$, AIC, $AICc$, and BIC for data sets I and II. In addition, in data set II, the Pareto model outperforms the conventional model since it covers a larger loss ($n = 2, 156$).

TABLE 5 Estimated values of fitted models and $-LL, AIC, AICc$, and BIC data set I.

Distributions	Parameters	-LL	AIC	AICc	BIC
Pareto	$\hat{\theta} = 0.12$, $\hat{\alpha} = 0.486$	165.80	333.62	333.68	338.83
Exponential	$\hat{\lambda} = 0.7855$	124.14	250.27	250.32	252.88
Length-biased expo	$\hat{\theta} = 0.6965$	115.97	233.94	233.98	236.55
c exponential-P	$\hat{\theta} = 1.867$	4.58	9.75	9.79	12.35
c lognormal-P	$\hat{\theta} = 1.241$, $\hat{\alpha} = 1.267$	4.28	8.92	8.98	14.13
c Rayleigh-P	$\hat{\theta} = 2.654$	4.08	8.76	8.80	11.36
CLBEP	$\hat{\theta} = 0.21$	3.75	7.84	7.92	10.38

TABLE 6 Estimated values of fitted models and $-LL, AIC, AICc$, and BIC data set II.

Distributions	Parameters	-LL	AIC	AICc	BIC
Pareto	$\hat{\theta} = 0.313$, $\hat{\alpha} = 0.546$	5.67	11.35	11.353	22.702
Exponential	$\hat{\lambda} = 0.417$	4041.1	8084.2	8084.2	8089.9
Length-biased expo	$\hat{\theta} = 0.834$	6447	12, 892	12, 892.001	12898
c Exponential-P	$\hat{\theta} = 2.477$	6.58	12.56	12.562	18.23
c Lognormal-P	$\hat{\theta} = 1.385$, $\hat{\alpha} = 1.436$	3.88	7.76	7.762	19.11
c Rayleigh-P	$\hat{\theta} = 2.848$	3.57	7.72	7.721	13.39
CLBEP	$\hat{\theta} = 2.3$	3.52	7.62	7.621	13.36

5. Conclusion

A unique distribution known as the composite length-biased exponential Pareto generated is suggested for application. Some of the mathematical features of this distribution include the quantile function, stochastic ordering, moments of the CLBEP, and maximum-likelihood estimation. In contrast to other conventional and new composite distributions, the distribution proposed in this work gives very satisfactory results. The goodness of fit of this novel model is compared to different conventional and new composite models, such as composite exponential-Pareto, composite lognormal-Pareto, and composite Rayleigh-Pareto distributions, using two real fire insurance data sets (Algerian and Danish fire insurance losses). Compared to the standard models, the composite models provided a far better fit to the data. The composite exponential-Pareto, composite lognormal-Pareto, and composite Rayleigh-Pareto distributions do not fit as well as the CLBEP model provides. We predict that researchers interested in statistical sciences and their applications, such as dependability and actuarial sciences, will be drawn to the CLBEP model. A future research may examine the Bayesian estimation of the CLBEP parameter, introducing the truncated version of the CLBEP distribution. In addition, it is interesting to use similar composite distributions to model the epidemic problem.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Acknowledgments

The authors acknowledge the Editor, FM and the referees of this journal for their constant encouragement to finalize the paper.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fams.2023.1137036/full#supplementary-material

References

1. Klugman SA, Panjer HH, Willmot G. *Loss Models: From Data to Decisions*. New York, NY: Wiley (2008).

2. Liu B, Ananda MMA. Analyzing insurance data with an exponentiated composite Inverse-Gamma Pareto Model. *Commun Stat Theory Methods*. (2022) 2022:399. doi: 10.1080/03610926.2022.2050399

3. Scollnik DPM. On composite Lognormal-Pareto models. *Scand Actuarial J*. (2007) 2007:20–33. doi: 10.1080/03461230601110447

4. Benatmane, C, Zeghdoudi H, Shanker R, Lazri, N. Composite Rayleigh-Pareto distribution: application to real fire insurance losses data set. *J Stat Manag Syst*. (2021) 24:545–57. doi: 10.1080/09720510.2020.1759253

5. Elbatal I, Aryal G. A new generalization of the exponential Pareto distribution. *J Inf Optim Sci*. (2017) 38:675–97. doi: 10.1080/02522667.2016.1220079

6. EL-Sagheer RM, Mahmoud MAW, Abdallah SHM. Statistical inferences for new Weibull-Pareto distribution under an adaptive type-ii progressive censored data. *J Stat Manag Syst*. (2018) 21:1021–57. doi: 10.1080/09720510.2018.1467628

7. Aminzadeh MS, Deng, M. Bayesian predictive modeling for Inverse Gamma-Pareto composite distribution. *Commun Stat Theory Methods*. (2019) 48:1938–54. doi: 10.1080/03610926.2018.1440595

8. Cebrian A, Denuit M, Lambert PH. Generalized Pareto fit to the society of Actuaries' large claims database. *North Am Actuarial J*. (2003) 7:18–36. doi: 10.1080/10920277.2003.10596098

9. Preda V, Ciumara R. On composite models: Weibull-Pareto lognormal-Pareto. A comparative study. *Rom J Econ Forecast*. (2006) 3:32–46.

10. Cooray K, Ananda MA. Modeling actuarial data with a composite Lognormal-Pareto model. *Scand Actuarial J*. (2005) 5:321–34. doi: 10.1080/03461230510009763

11. Teodorescu S, Vernic R. Some composite exponential-Pareto models for actuarial prediction. *Rom J Econ Forecast*. (2009) 12:82–100.

12. Teodorescu S, Vernic R. On composite pareto models. *Math Rep*. (2013)15:11–29.

13. Teodorescu S, Vernic R. A composite exponential–pareto distribution. The Annals of the "Ovidius" University of Constanta. *Math Series*. (2006) XIV:99–108.