Special Proceedings: ISBN#: 978-81-950383-0-5
23rd Annual Conference, 24-28 February 2021; pp 93-99

Benini Distribution: A Less Known Income-Size Distribution

V.S. Vaidyanathan
Department of Statistics,
Pondicherry University,
Puducherry

Received: 11 June 2021; Revised: 20 June 2021; Accepted: 24 June 2021

Abstract

Size distributions have potential applications in economics and actuarial science. Pareto and log-normal family of distributions are quite popular income-size distributions studied broadly in the literature. However, some less known size distributions are available but are not studied extensively. This paper discusses the properties and applications of one such probability distribution namely, Benini distribution.

Key words: Benini distribution; Maximum likelihood estimation; Lehmann-Scheffe theorem; Sufficient statistic; Claim amount.

AMS Subject Classifications: 60E05, 62F10

1. Introduction

Size distributions have potential applications in modeling income sizes, claim sizes and particle sizes, *etc.* These distributions usually are long- tailed and are skewed. For example, in modeling claim sizes, it is often the case that claims made by insurers are unusually high in very few cases, and therefore a long-tailed skewed distribution having support in the positive real line can be used to model the claim sizes. Pareto, log-normal and Gamma family of distributions are some well known choices in the family of size distributions to model size amounts. The cumulative distribution function (cdf) of a two-parameter Pareto distribution is given by

$$F(x) = 1 - (\frac{x}{x_0})^{-\alpha}, x \geq x_0 > 0, \alpha > 0, \tag{1}$$

where x_0 and α respectively denote the location and shape parameters. The corresponding probability density function (pdf) is

$$f(x) = \frac{\alpha x_0^{\alpha}}{x^{\alpha+1}}, x \geq x_0 > 0, \alpha > 0. \tag{2}$$

This distribution was developed by the Italian economist Vilfredo Pareto in the year 1895. Pareto observed a decreasing (downward) linear relationship between the logarithm of income

Corresponding Author: V.S. Vaidyanathan
Email: vaidya.stats@gmail.com

(x) and the logarithm of (T_x), the number of persons receiving income greater than x, $x \geq x_0$. He formulated the relationship as

$$log(T_x) = A - \alpha log(x), x \geq x_0 > 0.$$

Upon normalizing by the number of income receivers, the distribution function of X is obtained as given in (1). Various modified versions of Pareto distributions are available in the literature and are broadly categorized as Type I, Type II, ..., Type V Pareto distributions. A review of Pareto distributions along with its applications, properties and characterizations can be found in Arnold (2015). Soon after the publication by Pareto, economists, both applied and theoretical, started working on this new income-size distribution and found that it reasonably provided a good fit to model many economic phenomena. However, Rodolfo Benini, an Italian statistician and demographer, while modeling the distribution of wealth left as will (legacy), found that assuming a quadratic relationship provided a better fit for modeling legacies instead of a linear relationship as assumed in Pareto distribution. He obtained a new size distribution involving the quadratic term which came to be known as Benini distribution (Benini, 1905). The cdf and pdf of Benini distribution are respectively given by

$$F(x) = 1 - exp\{-\alpha log(\frac{x}{x_0}) - \beta(log(\frac{x}{x_0}))^2\}, x \geq x_0 > 0, \alpha, \beta \geq 0; (\alpha, \beta) \neq (0, 0) \qquad (3)$$

and

$$f(x) = exp\{-\alpha log(\frac{x}{x_0}) - \beta(log(\frac{x}{x_0}))^2\} \left[\frac{\alpha}{x} + \frac{2\beta log(\frac{x}{x_0})}{x}\right], x \geq x_0 > 0, \alpha, \beta \geq 0; (\alpha, \beta) \neq (0, 0). \qquad (4)$$

A distribution function having the form given in (3) with corresponding density function as given in (4) is known as Benini distribution and denoted as Benini(x_0, α, β). Here, x_0 denote the scale parameter and α, β denote the shape parameters of the distribution. Sphilberg (1977) has independently obtained this distribution in the context of modeling loss amounts due to fire accidents. Though Benini distribution has been established long ago, review of literature reveals that very little research has been carried in studying its properties and characterization. The present article is an attempt to fill this gap by deriving some properties and results of this distribution. In Section 2, some properties of two-parameter Benini distribution are discussed and new results are derived. Parametric Estimation is carried out in Section 3 and an illustration of the same in modeling insurance claim amount data is provided in Section 4. Conclusion of the paper is given in Section 5.

2. Two-Parameter Benini Distribution: Properties and New Resuts

Taking $\alpha = 0$ in (3) and (4), we obtain two-parameter Benini distribution having cdf and pdf respectively as

$$\begin{aligned} F(x) &= 1 - exp\{-\beta(log(\frac{x}{x_0}))^2\} \\ &= 1 - \left(\frac{x}{x_0}\right)^{-\beta(log(x) - log(x_0))}, x \geq x_0 > 0, \beta > 0 \end{aligned} \qquad (5)$$

and

$$f(x) = \left[\frac{2\beta log(\frac{x}{x_0})}{x}\right] exp\{-\beta(log(\frac{x}{x_0}))^2\}, x \geq x_0 > 0, \beta > 0. \tag{6}$$

The plot of the density and cumulative distribution functions of the two-parameter Benini distribution for different values of the shape parameter β and $x_0 = 1$ is depicted in Figure 1. From the plots, it can be observed that the density curve becomes more peaked with thin-longer tails for large values of the shape parameter.

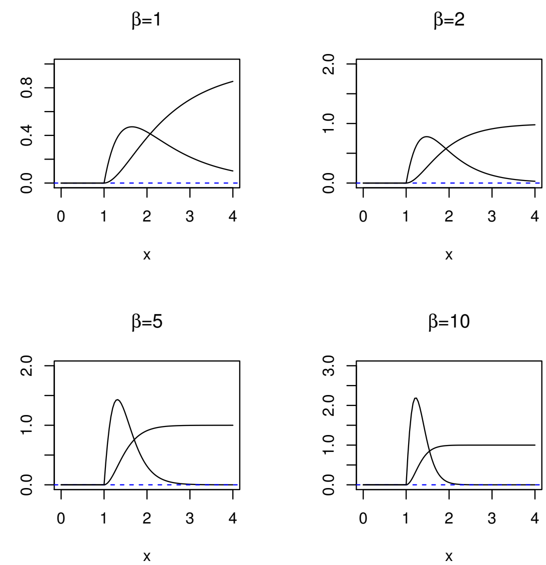

Figure 1: Plot of density and cumulative distribution functions of two-parameter Benini distribution

2.1. Properties

- The quantile function corresponding to (5) is given by

$$F^{-1}(u) = x_0 exp\left\{\sqrt{[\frac{-1}{\beta}]log(1-u)}\right\}, 0 < u < 1. \tag{7}$$

  Given a u, x_0 and β, random samples can be generated from two-parameter Benini distribution by inverse transformation method using (7).

- Equating the cdf given in (5) to 0.5, the median m of two-parameter Benini distribution is obtained as $m = x_0 exp\left\{\sqrt{[\frac{1}{\beta}]log(2)}\right\}$.

- Klieber (2013) has shown that Benini distribution suffers from the problem of moment indeterminacy *i.e.,* moments of all order exists but the distribution is not determined by the moment generating function. This property is true for log-normal distribution also.

- From (6), it can be seen that for fixed x_0, $\sum_{i=1}^{n}[log(\frac{X_i}{x_0})]^2$ is sufficient statistic for β based on n independent and identically distributed (iid) samples.

- Also, for fixed x_0, Benini distribution belong to one-parameter exponential family of distributions and therefore $\sum_{i=1}^{n}[log(\frac{X_i}{x_0})]^2$ is a complete sufficient statistic for β.

2.2. New Results

Let the random variable X has two-parameter Benini distribution with cdf and pdf as given in (5) and (6) respectively.

Result 1: Let $Y = log(\frac{X}{x_0})$. Then the distribution of Y is Rayleigh.
Proof: Consider the cdf of Y namely, $F(y)$, which is given by

$$\begin{aligned} F(y) &= P[Y \leq y] \\ &= P[log(\frac{X}{x_0}) \leq y] \\ &= P[X \leq e^y x_0] \\ &= 1 - \left(\frac{e^y x_0}{x_0}\right)^{-\beta(log(e^y x_0) - log(x_0))} \\ &= 1 - e^{-\beta y^2}, y \geq 0. \end{aligned}$$

This shows the distribution of Y is Rayleigh.

Result 2: Let $Y = aX$. Then Y is distributed as two-parameter Benini with parameters ax_0 and β.
Proof: The cdf of Y is

$$\begin{aligned} F(y) &= P[Y \leq y] = P[aX \leq y] \\ &= P[X \leq \frac{y}{a}] \\ &= 1 - \left(\frac{y}{ax_0}\right)^{-\beta(log(\frac{y}{a}) - log(x_0))} \\ &= 1 - \left(\frac{y}{ax_0}\right)^{-\beta(log(y) - log(ax_0))} \qquad , y \geq ax_0 > 0. \end{aligned}$$

Comparing with the cdf of Benini distribution given in (5), it is clear that Y is Benini with parameters ax_0 and β.

Result 3: Distribution of the first order statistic.
Proof: The cdf of the first order statistic $X_{(1)} = min\{X_1, X_2, ..., X_n\}$ based on n iid observations from two-parameter Benini distribution is given by

$$\begin{aligned} F(x_{(1)}) &= 1 - [1 - F(x)]^n \\ &= 1 - \left[1 - \left(1 - \left(\frac{x}{x_0}\right)^{-\beta(log(x) - log(x_0))}\right)\right]^n \\ &= 1 - \left(\frac{x}{x_0}\right)^{-n\beta(log(x) - log(x_0))} . \end{aligned}$$

Thus $X_{(1)}$ is distributed as Benini with parameters x_0 and $n\beta$ *i.e.,* the first order statistic is also distributed as Benini.

Result 4: Distribution of the sufficient statistic of the shape parameter β.
Proof: As mentioned in Section 2.1, for fixed x_0, $\sum_{i=1}^{n}[log(\frac{X_i}{x_0})]^2$ is sufficient for the shape parameter β. From Result 1, the distribution of $Y = log(\frac{X}{x_0})$ is found to be Rayleigh with parameter β. Since the sum of squares of n iid Rayleigh random variables has Gamma distribution, $\sum_{i=1}^{n}[log(\frac{X_i}{x_0})]^2$ has Gamma distribution with parameters n and β. In other words, the distribution of the sufficient statistic for the shape parameter β is Gamma(n, β).

3. Parametric Estimation

In this section, estimation of the parameters of two-parameter Benini distribution by maximum likelihood (ML) method is discussed. The likelihood function based on n iid samples from two-parameter Benini distribution is given by

$$L(\beta, x_0|x_1, x_2, ..., x_n) = \prod_{i=1}^{n}\left[\frac{2\beta log(\frac{x_i}{x_0})}{x_i}\right] exp\{-\beta(log(\frac{x_i}{x_0}))^2\}, x_i \geq x_0, \beta > 0.$$

Using the likelihood function, it is easy to verify that the ML estimator of x_0 is $\hat{X}_{(0)} = min(X_1, X_2, ..., X_n)$ and that of β is $\hat{\beta} = \frac{n}{\sum_{i=1}^{n}[log(\frac{X_i}{x_0})]^2}$. The distribution of $\hat{X}_{(0)}$ is Benini as established already in Result 3. To derive the distribution of $\hat{\beta}$, we proceed as follows. Note that for fixed x_0, the distribution of $\sum_{i=1}^{n}[log(\frac{X_i}{x_0})]^2$ is Gamma with parameters n and β as established in Result 4. Let $\hat{\beta}$ be denoted as $\hat{\beta} = \frac{n}{Y}$, where $Y = \sum_{i=1}^{n}[log(\frac{X_i}{x_0})]^2$, and Y follow Gamma(n, β). The distribution of $\frac{1}{Y}$ is inverse Gamma with pdf

$$f\left(\frac{1}{y}\right) = \frac{\beta^n}{\Gamma(n)} e^{-\beta(\frac{1}{y})} \left(\frac{1}{y}\right)^{n+1}, y > 0.$$

Hence, the pdf of $Z = \frac{n}{Y}$ is

$$f(z) = \frac{\beta^n}{\Gamma(n+1)} e^{-\beta(\frac{z}{n})} z^{n+1}, z > 0. \tag{8}$$

In other words, (8) is the probability density function of the ML estimator of β. Using (8), the expected value and variance of $\hat{\beta}$ is obtained as

$$E(\hat{\beta}) = \frac{n\beta}{(n-1)} \tag{9}$$

and

$$V(\hat{\beta}) = \frac{n^2\beta^2}{(n-1)} \left[\frac{1}{(n-1)(n-2)}\right]. \tag{10}$$

From (9), it is clear that the $E(\hat{\beta}) \neq \beta$ *i.e.,* the ML estimator $\hat{\beta}$ is not unbiased for β. However, as $n \longrightarrow \infty$ in (9) and (10), $E(\hat{\beta}) \longrightarrow \beta$ and $V(\hat{\beta}) \longrightarrow 0$, implying $\hat{\beta}$ is a consistent estimator of β. Using (9), an unbiased estimator of β is obtained as

$$\begin{aligned} T &= \frac{(n-1)}{n}\hat{\beta} \\ &= \frac{(n-1)}{n}\frac{n}{\sum_{i=1}^{n}[log(\frac{X_i}{x_0})]^2}. \end{aligned}$$

Remark 1: For fixed x_0, as mentioned in 2.1, Benini distribution belong to the one-parameter exponential family of densities with complete sufficient statistic $\sum_{i=1}^{n}[log(\frac{X_i}{x_0})]^2$. Since T is a function of complete sufficient statistic, by the application of Lehmann-Scheffe theorem, T is a minimum variance unbiased estimator of β. The variance of T is given by $V(T) = \frac{\beta^2}{(n-2)}$.

Remark 2: Since $\sum_{i=1}^{n}[log(\frac{X_i}{x_0})]^2$ is sufficient for β, uniformly most powerful (UMP) test on β can be derived using the sufficient statistic. Suppose we wish to test the hypothesis $H : \beta \leq \beta_0$ against the alternate $K : \beta > \beta_0$, then the UMP test will reject the hypothesis H for large value of the sufficient statistic. Since for large values of β, the distribution has thin-long tail, the above test can be used to check whether the given data is from a Benini distribution with thin-long tail. The p-value of the test can be computed using Gamma distribution.

4. Illustration

To illustrate the computation of the estimate of the parameters, a real-life data on insurance claim amount due to fire loss is considered in this section. The data set corresponds to Danish reinsurance claim data and consists of 2167 insurance claim amounts (in millions of Danish Krone) during the period from 3rd January 1980 until 31st December 1990. This data

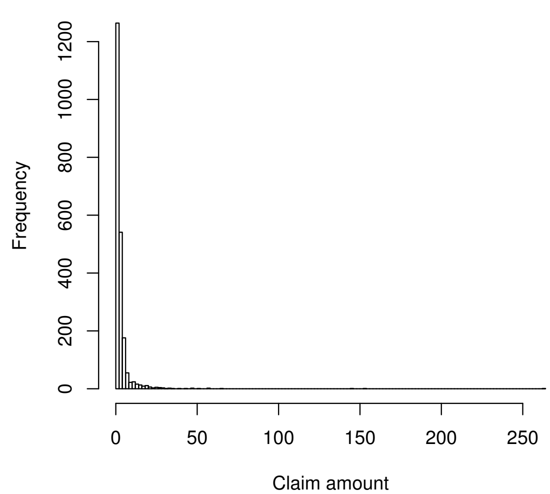

Figure 2: Histogram of claim amount

set has been used by Nadarajah and Bakar (2014) to fit composite models. The summary statistics of the data are: Minimum value is 1.000, first quartile 1.321, median 1.778, mean 3.385, third quartile 2.967 amd maximum value 263.250. The moment based skewness of claim amount is found to be 18.749 which indicates the data is positively skewed. The histogram of the claim amount is depicted in Figure 2. The histogram reveals that most of the claim amounts are small barring a few which are very large. Thus, it is reasonable to assume a positively skewed probability distribution having long-tail to model the claim amounts. Since Benini distribution is positively skewed and has long-tail, it is chosen to model the insurance claim amounts. As mentioned in the previous section, the ML estimator of the scale parameter x_0 is $\hat{X}_{(0)} = min(X_1, X_2, ..., X_n)$. Since the minimum value of claim amounts is one, $\hat{X}_{(0)} = 1$. The ML estimate of β is obtained using $\hat{\beta} = \frac{n}{\sum_{i=1}^{n}[log(\frac{X_i}{x_0})]^2}$ and is found to be 0.8828. The estimated standard error of $\hat{\beta}$ is 0.0189.

5. Conclusion

In this paper, a less known probability distribution namely, Benini distribution is reviewed and some new properties and distributional results are derived. The results obtained are used to find the distribution of the maximum likelihood estimator of the parameters of the distribution. In addition, minimum variance unbiased estimation of the shape parameter is carried out.

References

Arnold, Barry C. (2015). *Pareto Distributions*. Second Edition. CRC Press, Boca Raton, Florida.

Benini, R. (1905). I diagrammi a scala logaritmica (a proposito della graduazione per valore delle successioni ereditarie in Italia, Francia e Inghilterra). *Giornale degli Economisti*, **Series II, 16**, 222–231.

Klieber, C. (2013). On the moment indeterminacy of the Benini distribution. *Statistical Papers*, **54(4)**, 1121-1130.

Nadarajah, S. and Bakar, S. A. A. (2014). New composite models for the Danish fire insurance data. *Scandinavian Actuarial Journal*, **2**, 180-187.

Shpilberg, D. (1977). The probability distribution of fire loss amount. *The Journal of Risk and Insurance*, **44(1)**, 103-115.