Research Journal of Mathematical and Statistical Sciences ______________________________ISSN 2320-6047
Vol. **11(2),** 1-5, September **(2023)** *Res. J. Mathematical and Statistical Sci.*

Weighted Pareto Distribution: Statistical Properties and Estimation

Isha Gupta* **and Shivani Chowdhary**
Department of Statistics; University of Jammu, Jammu, India
ig.isha@yahoo.in

Available online at: www.isca.in, www.isca.me
Received 12th March 2021, revised 8th February 2023, accepted 4th July 2023

Abstract

In this paper, we have introduced a new class of Pareto distribution i.e. weighted Pareto distribution. Some structural properties of the distribution including behavior of probability density function, cumulative distribution function, reliability, hazard function, moments, entropy and order statistics are studied and derived. Also, by using different methods of estimation we obtain estimate of parameter of distribution.

Keywords: Pareto distribution, Weighted distribution, Reliability function, Hazard function, Order Statistics, Moments, Moment estimator, Maximum likelihood estimator, Minimum variance unbiased estimator, Shannon entropy.

Introduction

Pareto distribution is a skewed heavy tailed distribution that is sometimes used to model the distribution of incomes and is named after the economist Vilfredo Pareto. The study of weighted distributions can be used for better comprehension of standard distributions and provides techniques of extending distributions for further flexibility to fit a data set.

There are many applications in which the available data are a biased sample instead. Fisher[1] modeled a biased sample using a weight function w(x) ≥ 0, and constructed a weighted distribution with a density that is proportional to w(x)* g(x).

Applications of weighted distribution to biased samples in various areas including medicine, ecology, reliability, and branching processes can be seen in Patil and Rao[2], Patil et al.[3], Gupta and Kirmani[4], Gupta and Kundu[5], Shakhatreh[6], Dey et al.[7], Das and Kundu[8], Gupta and Gupta[9], Gupta and Gupta[10-12], Saghir et al.[13], Sen and Maiti[14,] Gupta and Gupta[15,16], Mandouh and Mohamed[17] and the references therein.

The objective of this paper is twofold: to study the properties of the weighted Pareto distribution (WPD) as well as to estimate the parameters of the model from frequencist view points. The article is organized as follows. Pareto distribution is discussed in section 1. Weighted Pareto distribution is introduced in Section 2. In Section 3 we study reliability function and hazard rate. Expressions for the moments, moment generating function and characteristics function of the WPD are presented in Section 4. In section 5, expression for order statistics and Shannon entropy are derived. Section 6 deals with different methods of estimation of parameters. Section 7 deals with estimation of Reliability function, Hazard function and Shannon's entropy. The paper ends with a brief conclusion in Section 8.

Weighted Pareto Distribution

The Probability density function of basic Pareto distribution with shape parameter θ is given by

$$g(x)=\frac{\theta}{x^{\theta+1}};x\geq 1 \qquad (1)$$

$$E\left(X^{n}\right)=\begin{cases}\dfrac{\theta}{\theta-n};0<n<\theta\\ \infty;n\geq\theta\end{cases}$$

Mean= $E(X)=\dfrac{\theta}{\theta-1};if\theta>1$

Variance = $\dfrac{\theta}{(\theta-1)^{2}};if\theta>2$

The probability density function of weighted distribution is given by

$$f(x)=\frac{w(x)\times g(x)}{E[w(x)]};$$

where $w(x)$ is a weight function which is non – negative.

Here, we take $w(x)=x^{k}, k=1,2,3......$ and $g(x)$ as the probability density function of Pareto distribution with expected value

$E_g\left(X^k\right)<\infty$.

Definition: A non- negative continuous random variable X is said to follow weighted Pareto distribution with shape parameter k and θ if its probability density function is given by

$$f(x,\theta,k)=\frac{\theta-k}{x^{(\theta-k+1)}};x\geq 1,0<k<\theta \quad (2)$$

and the corresponding cumulative density function is given by

$$F_X(x)=P(X\leq x)$$

$$=1-(x)^{-\theta+k};x\geq 1,0<k<\theta \quad (3)$$

Special Cases: i. When $k=0$, the weighted Pareto distribution reduces to Pareto distribution. ii. When $k=1$, the weighted Pareto distribution reduces to length or size biased Pareto distribution.

Reliability Analysis

In this subsection, we present the reliability function and the hazard function for the proposed weighted Pareto distribution. The reliability is defined as the probability that a system will survive beyond a specified time and is given by

$$R(x)=1-F_X(x)$$

$$=1-\left[1-(x)^{-\theta+k}\right]$$

$$=(x)^{-\theta+k} \quad (4)$$

The hazard function is also known as hazard rate or failure rate and is given by

$$h(x)=\frac{f(x,\theta,k)}{R(x)}$$

$$=\frac{\theta-k}{x} \quad (5)$$

Statistical Properties

In this section, we shall discuss properties of Weighted Pareto distribution specially moments and its associated measures, moment generating function, mode, harmonic mean and order statistics.

Moments and associated measures: Suppose X is a random variable which follows weighted Pareto distribution with shape parameter θ and k, then r^{th} order raw moment of random variable X $\mu_r';r=1,2,3......$ is given by

$$\mu_r'=E\left(X^r\right)$$

$$=\frac{\theta-k}{\theta-k-r} \quad (6)$$

Thus, Mean $E(X)=\dfrac{\theta-k}{\theta-k-1}$;

$$E\left(X^2\right)=\frac{\theta-k}{\theta-k-2};$$

$$\text{Variance}=\frac{\theta-k}{\theta-k-2}-\left(\frac{\theta-k}{\theta-k-1}\right)^2$$

Coefficient of variation is given by C.V.(X)

$$=\frac{\sqrt{\frac{\theta-k}{\theta-k-2}-\left(\frac{\theta-k}{\theta-k-1}\right)^2}}{\frac{\theta-k}{\theta-k-1}}$$

Mode: In order to find mode of weighted Pareto distribution, we take the logarithm of its probability density function as follow:

$$\log f(x,\theta,k)=\log(\theta-k)-(\theta-k+1)\log x$$

On differentiating the above equation w.r.t. 'x' and equating to zero, we get modal value

$$x=\theta-k+1 \quad (7)$$

Harmonic mean: The harmonic means of a random variable X which follow weighted Pareto distribution is given by

$$H=E\left(\frac{1}{X}\right)$$

$$=(\theta-k)\int_1^\infty \frac{1}{x^{(\theta-k+1)}}dx$$

$$=\frac{\theta-k}{\theta-k+1} \quad (8)$$

Moment generating function: In this sub section, we derived the moment generating function of weighted Pareto distribution; we begin with the well known definition of moment generating function given by

$$M_X(t)=E\left(e^{tX}\right)$$

$$=\int_1^{\infty}\sum_{r=0}^{\infty}\frac{t^r x^r}{r!}f(x,\theta,k)dx$$

$$=\sum_{r=0}^{\infty}\frac{t^r}{r!}\times\mu_r^{'}$$

$$=\sum_{r=0}^{\infty}\frac{t^r}{r!}\times\frac{\theta-k}{\theta-k-r};t\varepsilon\Re \quad (9)$$

Cumulant generating function: The cumulant generating function (CGF) of X is obtained as

$$K_X(t)=\log M_X(t)$$

$$=\log\sum_{r=0}^{\infty}\frac{t^r}{r!}\times\frac{\theta-k}{\theta-k-r};t\varepsilon\Re \quad (10)$$

Characteristic function: The characteristic function of a random variable following Pareto distribution is given by

$$\phi_X(t)=E\left(e^{itX}\right)$$

$$=\sum_{r=0}^{\infty}\frac{(it)^r}{r!}\times\frac{\theta-k}{\theta-k-r};t\varepsilon\Re \quad (11)$$

Order Statistics

Order statistics make their appearance in many statistical theory and practice. We know that if $X_{(1)},X_{(2)},X_{(3)}.....X_{(n)}$ denotes the order statistics of a random sample $X_1,X_2,X_3....X_n$ from a continuous population with c.d.f. $F_X(x)$ and p.d.f. $f_X(x)$, then the pdf of r^{th} order statistics $X_{(r)}$ is given by

$$g_{X_{(r)}}(x)=\frac{n!}{(r-1)!(n-r)!}\times g_X(x)\times\left[G_X(x)\right]^{r-1}$$

Now using the p.d.f and c.d.f of weighted Pareto distribution, we get the probability density function of rth order statistics of weighted Pareto distribution given by

$$f_{X_{(r)}}(x)=\frac{n!}{(r-1)!(n-r)!}\times f(x)\times\left[F_X(x)\right]^{r-1}$$

$$=\frac{n!}{(r-1)!(n-r)!}\times\frac{\theta-k}{x^{(\theta-k+1)}}\left[1-(x)^{-\theta+k}\right] \quad (12)$$

The probability density function of smallest order statistics $X_{(1)}$ is given by

$$f_{X_{(1)}}(x)=n\times\frac{\theta-k}{x^{(\theta-k+1)}}\times\left[x^{-\theta+k}\right]^{n-1} \quad (13)$$

The probability density function of largest order statistics is given by

$$f_{X_{(n)}}(x)=n\times\frac{\theta-k}{x^{(\theta-k+1)}}\times\left[1-x^{-\theta+k}\right]^{n-1} \quad (14)$$

Shannon's Entropy: Every probability distribution has some kind of uncertainty associated with it and entropy is used to measure this uncertainty as well as randomness of a probabilistic system. The concept of Entropy was introduced by Shannon[18] as a measure of information, which provides a quantitative measure of uncertainty. Let X be a random variable which follow weighted Pareto distribution with probability density function f(x) given by (2.1), then the Shannon's entropy is given by

$$H(x)=-E[\log f(x)]$$

$$E[\log f(x)]=\log(\theta-k)-(\theta-k+1)E[\log(x)]$$

In order to compute entropy, we need to find $E[\log(x)]$.

$$E[\log(x)]=\frac{k-\theta+1}{\theta-k}$$

Hence Shannon entropy is given by

$$H(x)=-\log(\theta-k)+(\theta-k+1)\times\frac{k-\theta+1}{\theta-k} \quad (15)$$

Estimation of Parameters

In this section, we obtain estimate of parameter θ of Weighted Pareto distribution using different methods of estimation:

Method of Moments: Moment estimator of parameter θ is obtained by equating the sample moment with the corresponding population moment. Let $x_1,x_2,x_3......$ be a sequence of random variables from weighted Pareto distribution whose probability density function is given by equation (1), then Sample moment is given by

$$m_1^{'}=\bar{x}$$

Population moment is given by

$$\mu_1' = E(X) = \frac{\theta - k}{\theta - k - 1}$$

On equating the above two equations, we get moment estimator of parameter θ and k given by

$$\hat{\theta} = k + \frac{\bar{x}}{\bar{x} - 1} \quad (16)$$

And

$$\hat{k} = \theta - \frac{\bar{x}}{\bar{x} - 1} \quad (17)$$

Maximum Likelihood Estimation: Maximum likelihood estimation has been the most widely used method of estimation of parameter of a distribution. If a random sample is drawn from Weighted Pareto distribution with pdf given by Equation (1), then the likelihood function is given by

$$L(\underline{x}, \theta) = (\theta - k)^n \times \prod_{i=1}^{n} \frac{1}{(x_i)^{\theta - k + 1}}$$

Log of likelihood function is given by

$$LogL(\underline{x}, \theta, k) = nLog(\theta - k) + \sum_{i=1}^{n} \log(x_i)^{k - \theta - 1}$$

Thus, maximum likelihood estimator of θ and k are given by

$$\hat{\theta}_{Mle} = \frac{n}{\sum_{i=1}^{n} \log x_i} + k \quad (18)$$

And

$$\hat{k}_{Mle} = \theta - \frac{n}{\sum_{i=1}^{n} \log x_i} \quad (19)$$

Method of Minimum Variance Unbiased Estimation: Let a random sample is drawn from weighted Pareto distribution then the log of likelihood function is given by

$$\log L(\underline{x}, \theta) = n \log(\theta - k) + \log \prod_{i=1}^{n} \frac{1}{(x_i)^{\theta - k + 1}}$$

$$\frac{d}{d\theta} \log L(\underline{x}, \theta) = \frac{T(\underline{x}) - \theta}{V[T(\underline{x})]}$$

where $T(\underline{x}) = \frac{n}{\sum_{i=1}^{n} \log x_i} + k$ and $V[T(\underline{x})] = \frac{\theta - k}{\sum_{i=1}^{n} \log x_i}$

Hence, Minimum variance unbiased estimator of θ is

$$\frac{n}{\sum_{i=1}^{n} \log x_i} + k \quad (20)$$

And Minimum variance unbiased estimator of k is

$$\theta - \frac{n}{\sum_{i=1}^{n} \log x_i} \quad (21)$$

Estimation of Reliability function, Hazard function and Shannon Entropy

On using the invariance property of maximum likelihood estimators, the maximum likelihood estimators of reliability function is obtained by replacing θ and k by $\hat{\theta}$ and $\hat{k}$ respectively, i.e

$$\hat{R}_{Mle} = (x)^{-\hat{\theta}_{Mle} + k}$$

$$= x^{\left(\frac{-n}{\sum_{i=1}^{n} \log x_i}\right)} \quad (22)$$

Maximum likelihood estimator of Hazard function is

$$\hat{h}_{Mle} = \frac{\hat{\theta}_{Mle} - k}{x}$$

$$= \frac{n}{x \sum_{i=1}^{n} \log x_i} \quad (23)$$

Maximum likelihood estimator of Shannon’s entropy is given by

$$\hat{H}(x)_{Mle} = -\log(\hat{\theta}_{Mle} - k) + \frac{(\hat{\theta}_{Mle} - k + 1)(k - \hat{\theta}_{Mle} + 1)}{(\hat{\theta}_{Mle} - k)}$$

$$= -\log\left[\frac{n}{\sum \log x_i}\right] + \frac{\left(\frac{n}{\sum \log x_i} + 1\right)\left[1 - \frac{n}{\sum \log x}\right]}{\frac{n}{\sum \log x_i}} \quad (24)$$

Conclusion

In this study, a new distribution called Weighted Pareto distribution is introduced. A detailed study on the statistical properties of the new distribution is presented. Estimates of parameter of Weighted Pareto Distribution are also obtained by using three different methods of estimation. Finally, maximum likelihood estimator of reliability function, hazard function and Shannon's entropy is obtained.

References

1. Fisher, R. A. (1934). The effect of methods of ascertainment upon the estimation of frequencies. *Annals of eugenics*, 6(1), 13-25.

2. Patil, G. P. and Rao, C. R. (1978). Weighted distributions and size-biased sampling with application to wildlife populations and human families. *Biometrics*, 34, 179 – 189.

3. Patil, G. P., Rao, C. R. and Ratnaparkhi, M. V. (1986). On discrete weighted distributions and their use in model choice for observed data. *Commun. Statist.-Theory Meth.*, 15(3), 907–918.

4. Gupta, R. C. and Kirmani, S. N. U. A. (1990). The role of weighted distribution in stochastic modeling. *Commun. Statist.—Theory Meth.*, 19(9), 3147–3162.

5. Gupta, R. D. and Kundu, D. (2009). A new class of Weighted Exponential distribution. *Statistics*, 43, 621-634.

6. Shakhatreh, M. K. (2012). A two parameter of Weighted Exponential distribution. *Statistics and Probability Letters*, 82, 252 - 261.

7. Dey, S., Dey, T. and Anis, M. Z. (2013). Weighted Weibull Distribution: Properties and Estimation. *Journal of Statistical Theory and Practice*.

8. Das, S. and Kundu, D. (2016). On Weighted Exponential distribution and its length biased version. *Journal of the Indian society for Probability and Statistics*, 17(1), 1 –21.

9. Gupta, I. and Gupta, R. (2016). Bayesian Inference about the Estimation and Credible Intervals for a Continuous Distribution Under Bounded Loss Function. *Global Journal for Research Journals*, 5(9), 332-340.

10. Gupta, I. and Gupta, R. (2017). Estimation of parameter and reliability function of Exponentiated Inverted Weibull Distribution using Classical and Bayesian approach. *International Journal of Recent Scientific Research*, 8(7), 18117-18819.

11. Gupta, I. and Gupta, R. (2017). Bayesian and E-Bayesian estimation of the unknown shape parameter of Exponentiated Inverted Weibull Distribution using different loss functions. *International Education and Research Journal*, 3(5), 408-416.

12. Gupta, I. and Gupta, R. (2017). Bayesian and E- Bayesian Method of Estimation of Parameter of Rayleigh Distribution- A Bayesian Approach under Linex Loss Function. *International Journal of Statistics and Systems,* 12(4), 791-796.

13. Saghir A., Hamedani G. G., Tazeem S., and Khadim, A. (2017). Weighted Diatribution: A brief review, perspective and characterizations. *International Journal of Statistics and Probability*, 6(3).

14. Sen, S., Chandra, N. and Maiti, S. S. (2017). The Weighted Gamma distribution, properties and applications. *Journal of Reliability and Statistical Studies*, 10(1), 43-58.

15. Gupta, I. and Gupta, R. (2018). Bayesian and Non Bayesian Method of Estimation of Scale Parameter of Gamma Distribution under Symmetric and Asymmetric Loss Functions. *World Scientific News: An International Scientific Journal*, 172–191.

16. Gupta, I. and Gupta, R. (2018). Sequential Analysis of unknown parameter of Exponentiated Inverted Weibull Distribution. *International Journal of Applied Mathematics and Statistics*, 57(3), 169-179.

17. Mandouh, R. M and Mohamed, M. A. G. (2020). Log Weighted Pareto Distribution and Its Statistical Properties. *Journal of Data Science*, 18(1), 161-189

18. Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379 - 423.