JOURNAL OF STATISTICS SCIENCES
Volume 1, Number 1, June 2009, pp. 55-62


# ON GENERALIZATION OF THE PARETO DISTRIBUTION

*A. K. Olapade*

***ABSTRACT:*** *In this paper, we present a probability density function that generalizes the Pareto distribution. The cumulative distribution function and the moments of the generalized distribution are obtained while the distribution of some order statistics of the distribution are established. Included in this paper are the estimation of the parameters of the generalized Pareto distribution and some theorems that relate it to other statistical distributions are established.*

***2000 Mathematics Subject Classifications:*** *Primary 62E15; Secondary 62E10.*

***Keywords:*** *Generalized pareto distribution, Gamma distribution, Exponential distribution, Homogeneous differential equation.*

## 1. INTRODUCTION

A continuously distributed random variable $X$ is said to follow a Pareto distribution if

$$f(x) = \frac{1}{(1+x)^2},\ x > 0 \tag{1.1}$$

with cumulative distribution function

$$F(x) = 1 - \frac{1}{1+x},\ x > 0. \tag{1.2}$$

It is well known in general that generalized models being more flexible than ordinary models, are usually prefered in analysing most data sets. This has prompted several authors to embark upon investigating the properties and applications of generalized models. Patil and Taillie (1994) presented the probability density function of a generalized Pareto distribution as

$$f(x; \beta, \theta) = \frac{\beta\theta}{(1+\theta x)^{\beta+1}},\ x > 0, \theta > 0, \beta > 0, \tag{1.3}$$

where $\beta$ and $\theta$ are the shape and scale parameters respectively.

In this paper, we further generalize the Pareto distribution with the probability density function

$$f(x;\lambda,\beta,\theta)=\frac{\lambda^{\beta}\theta\beta}{(\lambda+\theta x)^{\beta+1}},\ x>0,\lambda>0,\theta>0,\beta>0. \tag{1.4}$$

Equation (1.4) is what we refer to as a generalized Pareto distribution where $\lambda$ and $\beta$ are the shape parameters and $\theta$ is a scale parameter.

If the location parameter $\mu$ is introduced, we have

$$f(x;\lambda,\beta,\theta,\mu)=\frac{\lambda^{\beta}\theta\beta}{(\lambda+\theta(x-\mu))^{\beta+1}},\ x>0,\lambda>0,\theta>0,\beta>0,\mu>0. \tag{1.5}$$

Hence we can name the distribution (1.5) as a four parameter generalized Pareto distribution.

### 1.1. Cumulative Distribution Function of the Generalized Pareto Distribution

For a random variable $X$ that has the probability density function shown in the equation (1.4), the cumulative distribution function is given as

$$F(x;\lambda,\beta,\theta)=\int_0^x\frac{\lambda^{\beta}\theta\beta}{(\lambda+\theta t)^{\beta+1}}dt$$

$$=-\frac{\lambda^{\beta}}{(\lambda+\theta t)^{\beta}}\Big|_0^x=1-\left(\frac{\lambda}{\lambda+\theta x}\right)^{\beta},\ 0<x<\infty. \tag{1.6}$$

The probability that a generalized Pareto random variable $X$ lies in the interval $(\alpha_1,\alpha_2)$ is given as

$$pr(\alpha_1<X<\alpha_2)=\left(\frac{\lambda}{\lambda+\theta\alpha_1}\right)^{\beta}-\left(\frac{\lambda}{\lambda+\theta\alpha_2}\right)^{\beta},\ \text{for } \alpha_1<\alpha_2. \tag{1.7}$$

## 2. MOMENTS OF THE GENERALIZED PARETO DISTRIBUTION

The moments of generalized Pareto distribution are shown below.

The first moment:

$$E[X]=\int_{-\infty}^{\infty}xf(x)dx \tag{2.1}$$

$$=\lambda^{\beta}\theta\beta\int_0^{\infty}\frac{x}{(\lambda+\theta x)^{\beta+1}}dx=\frac{\theta\beta}{\lambda}\int_0^{\infty}\frac{x}{(1+\frac{\theta x}{\lambda})^{\beta+1}}dx. \tag{2.2}$$

Let $\frac{\theta x}{\lambda} = z$, $x = \frac{\lambda}{\theta} z$ and $dx = \frac{\lambda}{\theta} dz$, then

$$E[X] = \frac{\lambda\beta}{\theta} \int_0^{\infty} \frac{z}{(1+z)^{\beta+1}} dz = \frac{\lambda\beta}{\theta} B(2, \beta - 1), \tag{2.3}$$

where $B(.,.)$ is a complete beta function. Therefore, the mean of $X$ is obtained as

$$\mu = \frac{\lambda\beta}{\theta} \frac{\Gamma 2\Gamma(\beta-1)}{\Gamma(\beta+1)} = \frac{\lambda}{\theta(\beta-1)}, \quad \beta \neq 1.$$

The second moment:

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx \tag{2.4}$$

$$= \lambda^{\beta}\theta\beta \int_0^{\infty} \frac{x^2}{(\lambda+\theta x)^{\beta+1}} dx = \frac{\theta\beta}{\lambda} \int_0^{\infty} \frac{x^2}{(1+\frac{\theta x}{\lambda})^{\beta+1}} dx. \tag{2.5}$$

Using the substitutions in first moment

$$E[X^2] = \frac{\lambda^2\beta}{\theta^2} \int_0^{\infty} \frac{z^2}{(1+z)^{\beta+1}} dz = \frac{\lambda^2\beta}{\theta^2} B(3, \beta - 2) \tag{2.6}$$

$$= \frac{\lambda^2\beta}{\theta^2} \frac{\Gamma 3\Gamma(\beta-2)}{\Gamma(\beta+1)} = \frac{2\lambda^2}{\theta^2(\beta-1)(\beta-2)}, = \beta \neq 1, 2. \tag{2.7}$$

Variance of the generalized Pareto distribution ($\sigma^2$) is now obtained as

$$\sigma^2 = E[X^2] - (E[X])^2$$

$$= \frac{2\lambda^2}{\theta^2(\beta-1)(\beta-2)} - \frac{\lambda^2}{\theta^2(\beta-1)^2}$$

$$= \frac{\lambda^2}{\theta^2(\beta-1)} \left[ \frac{2}{\beta-2} - \frac{1}{\beta-1} \right]$$

$$= \frac{\lambda^2\beta}{\theta^2(\beta-1)^2(\beta-2)}, \beta > 2. \tag{2.8}$$

The $n^{th}$ moment: We obtained the $n^{th}$ moment of the generalized Pareto distribution as follows

$$E[X^n] = x^n \int_{-\infty}^{\infty} x^n f(x)dx \tag{2.9}$$

$$= \lambda^{\beta}\theta\beta \int_0^{\infty} \frac{x^n}{(\lambda+\theta x)^{\beta+1}} dx = \frac{\theta\beta}{\lambda} \int_0^{\infty} \frac{x^n}{(1+\frac{\theta x}{\lambda})^{\beta+1}} dx. \tag{2.10}$$

Using the substitutions in first moment

$$E[X^n] = \frac{\lambda^n\beta}{\theta^n} \int_0^{\infty} \frac{z^n}{(1+z)^{\beta+1}} dz = \frac{\lambda^n\beta}{\theta^n} B(n+1, \beta-n) \tag{2.11}$$

$$= \frac{\lambda^n\beta}{\theta^n} \frac{\Gamma(n+1)\Gamma(\beta-n)}{\Gamma(\beta+1)} = \frac{\lambda^n}{\theta^n} \frac{\Gamma(n+1)\Gamma(\beta-n)}{\Gamma\beta}, \beta > n. \tag{2.12}$$

## 3. ORDER STATISTICS OF THE GENERALIZED PARETO DISTRIBUTION

Let $X_1, X_2, ..., X_n$ be $n$ independently continuous random variables from the generalized Pareto distribution and let $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$ be the corresponding order statistics. Let $F_{X_{(r:n)}}(x)$, $(r = 1, 2, ..., n)$ be the cumulative distribution function of the $r^{th}$ order statistics $X_{(r:n)}$ and $f_{X_{(r:n)}}(x)$ denotes its probability density function. David (1970) gives the probability density function of $X_{(r:n)}$ as

$$f_{X_{(r:n)}}(x) = \frac{1}{B(r, n-r+1)} P^{r-1}(x)[1-P(x)]^{n-r}p(x). \tag{3.1}$$

For the generalized Pareto distribution with probability density function and cumulative distribution function given in the equations (1.4) and (1.6) respectively, by substituting $f(x)$ for $p(x)$ and $F(x)$ for $P(x)$ in the equation (3.1), we have

$$f_{X(r:n)}(x) = \frac{1}{B(r, n-r+1)}\left[1-\left(\frac{\lambda}{\lambda+\theta x}\right)^{\beta}\right]^{r-1}\left[\left(\frac{\lambda}{\lambda+\theta x}\right)^{\beta}\right]^{n-r}\frac{\lambda^{\beta}\theta\beta}{(\lambda+\theta x)^{\beta+1}}$$

$$= \frac{\theta\beta}{B(r, n-r+1)}\left[1-\left(\frac{\lambda}{\lambda+\theta x}\right)^{\beta}\right]^{r-1}\left[\left(\frac{\lambda}{\lambda+\theta x}\right)^{\beta}\right]^{n-r+1}\frac{1}{(\lambda+\theta x)}. \tag{3.2}$$

### 3.1. The Probability Density Function of the Minimum and Maximum Observations of Generalized Pareto Distribution

The minimum observation is denoted as $X_{1:n}$ and its probability density function is obtained by making $r = 1$ in the equation (3.2) to have

$$f_{X_{(1)}}(x) = \frac{\lambda^{n\beta}\theta n\beta}{(\lambda+\theta x)^{n\beta+1}},\ x > 0, \lambda > 0, \theta > 0, \beta > 0; \tag{3.3}$$

which is another generalized Pareto with parameter $(\lambda, \theta, n\beta)$.

The maximum observation is denoted by $X_{n:n}$ and its probability density function is obtained by making $r = n$ in the equation (3.2) to have

$$f_{X(n)}(x) = \frac{n\lambda^{\beta}\theta\beta}{(\lambda+\theta x)^{n\beta+1}}\left[1-\left(\frac{\lambda}{\lambda+\theta x}\right)^{\beta}\right]^{n-1},\ x > 0, \lambda > 0, \theta > 0, \beta > 0. \tag{3.4}$$

These densities of the order statistics are very useful as we shall see in the later section.

**Theorem 3.1:** The probability density function of the sample median in a random sample of size $2n - 1$ from the generalized Pareto distribution is proportional to the product of the probability density function of the minimum observation and $(n - 1)^{th}$ power of the cumulative distribution function of the generalized Pareto distribution.

**Proof:** If $X_1, X_2, \ldots, X_{2n-1}$ are from the generalized Pareto distribution such that $X_1 \leq X_2 \leq \ldots \leq X_{2n-1}$, then

$$f_{X_{(n)}}(x) = \frac{\lambda^{\beta}\theta\beta}{B(n,n)(\lambda+\theta x)^{\beta+1}}\left[1-\left(\frac{\lambda}{\lambda+\theta x}\right)^{\beta}\right]^{n-1}\left[\left(\frac{\lambda}{\lambda+\theta x}\right)^{\beta}\right]^{n-1} \tag{3.5}$$

$$= \frac{\lambda^{\beta}\theta\beta}{B(n,n)(\lambda+\theta x)^{\beta+1}}\left[1-\left(\frac{\lambda}{\lambda+\theta x}\right)^{\beta}\right]^{n-1}$$

$$= \frac{\Gamma(2n)}{n(\Gamma n)^2}\frac{\lambda^{n\beta}\theta n\beta}{(\lambda+\theta x)^{n\beta+1}}[F(x)]^{n-1}$$

$$= \frac{\Gamma(2n)}{n(\Gamma n)^2}f_{X_{(1)}}(x)[F(x)]^{n-1}, \tag{3.6}$$

where $\frac{\Gamma(2n)}{n(\Gamma n)^2}$ is the proportionality constant. This completes the prove.

## 4. ESTIMATION OF THE PARAMETERS OF THE GENERALIZED PARETO DISTRIBUTION

Suppose a random variable $X$ that has the generalized Pareto distribution have mean $\mu$ and variance $\sigma^2$. Then by the method of moment, equating the theoretical mean and the theoretical variance to the experimental mean and experimental variance respectively, we obtain

$$\mu = \frac{\lambda}{\theta(\beta - 1)} \tag{4.1}$$

and

$$\sigma^2 = \frac{\lambda^2 \beta}{\theta^2 (\beta - 1)^2 (\beta - 2)}. \tag{4.2}$$

By substituting for $\mu$ in the equation (4.2) and solve for $\beta$, we have

$$\hat{\beta} = \frac{\sigma^2 + 2\mu^2}{\sigma^2}. \tag{4.3}$$

Also, from the equation (4.1),

$$\hat{\lambda} = \mu\theta(\hat{\beta} - 1). \tag{4.4}$$

Since $\theta$ is a function of $\sigma$ as a measure of dispersion, we obtained the estimates of $\lambda$, $\beta$ and $\theta$ of the generalized Pareto distribution having obtained the estimates of $\mu$ and $\sigma$ by method of moment.

## 5. SOME RELATIONSHIPS BETWEEN THE GENERALIZED PARETO AND OTHER STATISTICAL DISTRIBUTIONS

In this section, we shall state and prove some theorems that relate the generalized Pareto distribution to some other statistical distributions.

**Theorem 5.1:** Suppose $Y$ is a continuously distributed random variable with probability density function $f_Y(y)$, then the random variable $X = e^{\frac{Y}{\beta}} - \frac{\lambda}{\theta}$ has a generalized Pareto distribution with parameters ($\beta$, $\theta$, $\lambda$) if $Y$ is exponentially distributed.

**Proof:** The probability density function of an exponential random variable $Y$ is

$$f_Y(y) = e^{-y},\ y > 0.$$

By omitting all constant, the density of $X$ can be written as

$$f_X(x) \ \alpha \ (\lambda + \theta x)^{-(\beta+1)}. \tag{5.1}$$

Since any density function proportional to the right hand side of the equation (5.1) is that of a generalized Pareto random variable, the proof is complete.

**Theorem 5.2:** Suppose $X_1$ and $X_2$ are independently distributed continuous random variables. If $X_1$ has an exponential distribution with probability density function

$$f(x_1) = e^{-x_1}, x_1 > 0$$

and $X_2$ has a gamma distribution with probability density function

$$f(x_2) = \frac{1}{\Gamma(\beta)} x_2^{\beta-1} e^{-x_2},\ x_2 > 0,\ \beta > 0, \tag{5.2}$$

then the random variable $Y = \dfrac{\lambda X_1}{\theta X_2}$ has a generalized Pareto distribution with parameters $\lambda$, $\theta$, and $\beta$.

**Proof:** The joint density function of $X_1$ and $X_2$ is

$$f(x_1, x_2) = \frac{1}{\Gamma(\beta)} x_2^{\beta-1} e^{-(x_1+x_2)}. \tag{5.3}$$

Let $y_1 = \dfrac{\lambda x_1}{\theta x_2}$ and $y_2 = x_2$, so that we obtain the density of the random variable $Y_1$ as

$$f(y_1) = \frac{\theta}{\lambda \Gamma(\beta)} \int_0^\infty y_2^{\beta} e^{-y_2\left(\frac{\theta}{\lambda} y_1 + 1\right)} dy_2$$

$$\frac{\lambda^\beta \theta \beta}{(\lambda + \theta y_1)^{\beta+1}},\ y_1 > 0,\ \lambda > 0,\ \theta > 0,\ \beta > 0. \tag{5.4}$$

Since this is the probability density function of a generalized Pareto random variable $Y_1$, with parameters $\lambda$, $\theta$, and $\beta$, the proof is complete.

**Theorem 5.3:** The random variable $X$ is generalized Pareto with probability density function $f$ given in the equation (1.4) if and only if $f$ satisfies the homogeneous differential equation

$$(\lambda + \theta x) f' + (\beta + 1)\theta f = 0 \tag{5.5}$$

where prime denotes differentiation and $f$ denotes $f(x)$.

**Proof:** Suppose $X$ is a generalized Pareto random variable with probability density function given in the equation (1.4), it is easily shown that the $f$ above satisfies the equation (5.5).

Conversely, if we assume that the function in the equation (1.4) satisfies the equation (5.5). Separating the variables in the equation (5.5) and integrating, we have

$$\ln f = -(\beta + 1)\ln(\lambda + \theta x) + \ln k, \tag{5.6}$$

where $k$ is a constant of integration. Obviously from equation (5.6)

$$f = \frac{k}{(\lambda + \theta x)^{\beta+1}}. \tag{5.7}$$

The normalizing constant $k = \lambda^{\beta}\theta\beta$ and this completes the prove.

## REFERENCES

[1] H. A. David, Order Statistics. John Wiley, New York (1970).

[2] G. P. Patil and C. Taillie. Statistical Evaluation of the Attainment of Interim Cleanup Standards at Hazardous Waste Sites. Environmental and Ecological Statistics 1, (1994), 333-351.

**A. K. Olapade**
Department of Mathematics
Obafemi Awolowo University
Ile-Ife, Nigeria
*E-mail: akolapad@oauife.edu.ng*