AUSTRIAN JOURNAL OF STATISTICS
Volume 39 (2010), Number 4, 329–340

On the Ratio of two Independent Exponentiated Pareto Variables

M. Masoom Ali[1], Manisha Pal[2] and Jungsoo Woo[3]

[1] Dept. of Mathematical Sciences, Ball State University, Indiana, USA
[2] Dept. of Statistics, University of Calcutta, India
[3] Dept. of Statistics, Yeungnam University, Gyongsan, South Korea

Abstract: In this paper we derive the distribution of the ratio of two independent exponentiated Pareto random variables, X and Y, and study its properties. We also find the UMVUE of $\Pr(X < Y)$, and the UMVUE of its variance. As some of the expressions could not be expressed in closed forms, some special functions have been used to evaluate them.

Zusammenfassung: In dieser Arbeit leiten wir die Verteilung des Quotienten zweier unabhängiger exponenzierter Pareto Zufallsvariablen X and Y her und studieren seine Eigenschaften. Wir finden auch den UMVUE von $\Pr(X < Y)$ und den UMVUE seiner Varianz. Da einige Ausdrücke nicht in geschlossener Form dargestellt werden können, wurden einige spezielle Funktionen verwendet, um diese auszuwerten.

Keywords: Exponentiated Pareto Distribution, Generalized Hypergeometric Function, UMVUE.

1 Introduction

The Pareto distribution is a power law probability distribution having cumulative distribution function (cdf)

$$F(x) = 1 - \left(\frac{\beta}{x}\right)^c , \qquad x \geq \beta , \quad c > 0 , \tag{1}$$

where β and c are, respectively, the threshold and shape parameters.

The distribution is found to coincide with many social, scientific, geophysical, actuarial, and various other types of observable phenomena. Some examples where the Pareto distribution gives good fit are the sizes of human settlements, the values of oil reserves in oil fields, the standardized price returns on individual stocks, sizes of meteorites, etc. An extension/generalization of the Pareto distribution is the exponentiated Pareto distribution with cdf

$$G(x) = F^{\alpha}(x) , \qquad \alpha > 0 . \tag{2}$$

Here α denotes the exponentiating parameter, and for $\alpha = 1$ the distribution reduces to the standard Pareto distribution (1).

R. C. Gupta, Gupta, and Gupta (1998) introduced the exponentiated exponential distribution, and its properties were studied by R. D. Gupta and Kundu (2001). Pal, Ali, and Woo (2006) studied certain aspects of the exponentiated Weibull distribution. M. M. Ali, Pal, and Woo (2007) studied several exponentiated distributions, including the exponentiated Pareto distribution, and discussed their properties. They showed that the exponentiated Pareto distribution gives a good fit to the tail-distribution of Nasdaq data.

The problem of estimating the probability that a random variable X is less than another random variable Y arises in many practical situations, like biometry, reliability study, etc. This problem has been studied by many authors for different distributions of X and Y, see, for example Pal, Ali, and Woo (2005), M. Ali, Pal, and Woo (2005), and M. M. Ali, Pal, and Woo (2009).

In this paper, we find the distribution of the ratio of two independent exponentiated Pareto random variables X and Y and study its properties. Some special functions have been used to evaluate terms that cannot be expressed in closed form. We also find the UMVUE of $\Pr(X < Y)$ and the UMVUE of its variance. Finally, we obtain the UMVUEs of $\Pr(X < Y)$ and its variance by analyzing a simulated data set and a real-life data set.

2 Distribution of the Ratio Y/X

Let X and Y be independent exponentiated Pareto random variables having cdf (2) with parameters α_1, β_1, c and α_2, β_2, c, respectively. Using formula 3.197(3) in Gradshteyn and Ryzhik (1965), the cdf of the ratio $Q = Y/X$ is obtained as

$$F_Q(z) = F\left(-\alpha_2, 1; \alpha_1 + 1; \left(\tfrac{\beta_2}{\beta_1 z}\right)^c\right), \qquad \text{if } z \geq \tfrac{\beta_2}{\beta_1}, \tag{3}$$

where $F(a, b; c; z) = \sum_{i=0}^{\infty} \frac{(a)_i (b)_i}{(c)_i} \frac{z^i}{i!}$ is the hypergeometric function and $(w)_i = w(w + 1) \cdots (w + i - 1)$ with $(w)_0 = 1$.

From (3) and formula 15.2.1 in Abramowitz and Stegun (1972), the pdf of Q is therefore given by

$$f_Q(z) = \frac{c\alpha_2}{1+\alpha_1}\left(\tfrac{\beta_2}{\beta_1}\right)^c z^{-c-1} F\left(-\alpha_2 + 1, 2; \alpha_1 + 2; \left(\tfrac{\beta_2}{\beta_1 z}\right)^c\right), \qquad \text{if } z \geq \tfrac{\beta_2}{\beta_1}. \tag{4}$$

The kth moment of Q about the origin is obtained as

$$\begin{aligned}
\mathrm{E}(Q^k) &= \frac{c\alpha_2}{1+\alpha_1}\left(\frac{\beta_2}{\beta_1}\right)^c \sum_{i=0}^{\infty} \frac{(-\alpha_2+1)_i (2)_i}{(\alpha_1+2)_i i!}\left(\frac{\beta_2}{\beta_1}\right)^{ci} \int_{\beta_2/\beta_1}^{\infty} z^{-c(i+1)-1+k} dz \\
&= \frac{c\alpha_2}{1+\alpha_1}\left(\frac{\beta_2}{\beta_1}\right)^k \sum_{i=0}^{\infty} \frac{(-\alpha_2+1)_i (2)_i}{(\alpha_1+2)_i} \frac{1}{c(i+1)-k} \frac{1}{i!} \\
&= \frac{\alpha_2}{1+\alpha_1}\left(\frac{\beta_2}{\beta_1}\right)^k \left(1-\frac{k}{c}\right)^{-1} \sum_{i=0}^{\infty} \frac{(-\alpha_2+1)_i (2)_i}{(\alpha_1+2)_i} \frac{(1-k/c)}{(i+1)-k/c} \frac{1}{i!} \\
&= \frac{\alpha_2}{1+\alpha_1}\left(\frac{\beta_2}{\beta_1}\right)^k \left(1-\frac{k}{c}\right)^{-1} {}_3F_2\left(1-\alpha_2, 2; 1-\tfrac{k}{c}; \alpha_1+2, 2-\tfrac{k}{c}; 1\right), \quad \text{if } k < c,
\end{aligned} \tag{5}$$

where

$${}_3F_2(a, b; c; p; q; 1) = \sum_{i=0}^{\infty} \frac{(a)_i (b)_i (c)_i}{(p)_i (q)_i} \frac{z^i}{i!}$$

is the generalized hypergeometric function in Gradshteyn and Ryzhik (1965).

Figure 1 shows the density curves for different combinations of (α_1, α_2) when $c = 5$ and $\beta_2/\beta_1 = 2$. Table 1 provides the asymptotic means, variances, and coefficients of skewness of the density (4) when $c = 5$, for different combinations of (α_1, α_2). (We use formula 9.14(1) in Gradshteyn and Ryzhik (1965) for the computation.) The figures indicate that the distribution of Q is skewed to the right.

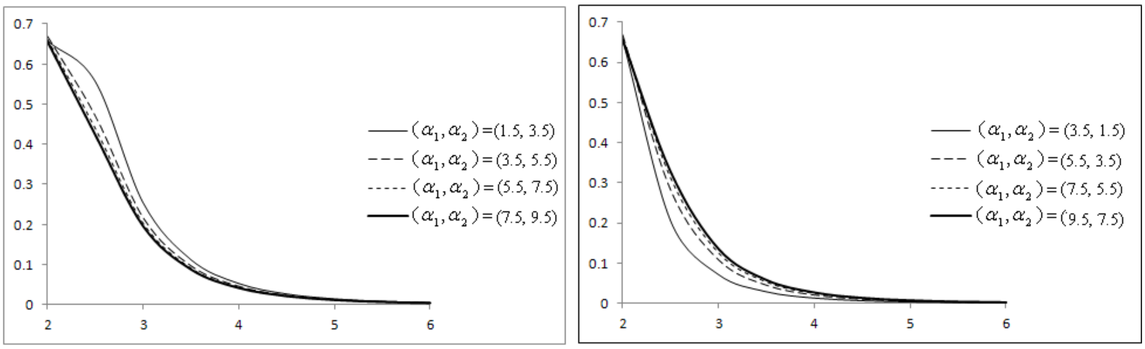

Figure 1: Density curves for $c = 5$ and $\beta_2/\beta_1 = 2$ for $\alpha_1 < \alpha_2$ (left) and $\alpha_1 > \alpha_2$ (right).

Table 1: Asymptotic mean, variance, and coefficient of skewness of the density (4) with $c = 5$ (units of mean and variance are β_2/β_1 and $(\beta_2/\beta_1)^2$, respectively).

α_1, α_2	Mean	Variance	Skewness
(1.5, 3.5)	1.20386	0.18101	2.7391
(3.5, 1.5)	0.91806	0.10180	2.6677
(3.5, 5.5)	1.14940	0.18179	2.5964
(5.5, 3.5)	0.97778	0.18092	2.5840
(5.5, 7.5)	1.12560	0.18122	2.5561
(7.5, 5.5)	1.00342	0.14379	2.5519
(7.5, 9.5)	1.11269	0.18079	2.5373
(9.5, 7.5)	1.01776	0.15115	2.5354

From Table 1, we observe that for $c = 5$:

- the distribution is skewed to the right;
- for $\alpha_1 > \alpha_2$, the distribution has smaller variance than that when $\alpha_1 < \alpha_2$.

3 Distribution of the Ratio X/(X+Y)

Consider the ratio $R = X/(X+Y)$, where X and Y are independent exponentiated Pareto random variables having the cdf (2) with parameters α_1, β_1, c and α_2, β_2, c, respectively.

From (3) we obtain the cdf of the ratio R as

$$\begin{aligned} F_R(z) &= \Pr\left(\frac{Y}{X} > \frac{1-z}{z}\right) \\ &= 1 - F\left(-\alpha_2, 1; \alpha_1 + 1; \left(\frac{\beta_2}{\beta_1}\frac{z}{1-z}\right)^c\right), \qquad \text{if } 0 < z < \tfrac{\beta_1}{\beta_1+\beta_2} \end{aligned} \tag{6}$$

and its density function is

$$\begin{aligned} f_R(z) &= \frac{c\alpha_2}{1+\alpha_1}\left(\frac{\beta_2}{\beta_1}\right)^c \frac{z^{c-1}}{(1-z)^{c+1}} \\ &\quad \times F\left(1-\alpha_2, 2; \alpha_1 + 2; \left(\frac{\beta_2}{\beta_1}\frac{z}{1-z}\right)^c\right), \qquad \text{if } 0 < z < \tfrac{\beta_1}{\beta_1+\beta_2} \end{aligned} \tag{7}$$

Using formula 7.512(5) in Gradshteyn and Ryzhik (1965), the kth moment of R about the origin is obtained as

$$\begin{aligned} \mathrm{E}(R^k) &= \frac{c\alpha_2}{1+\alpha_1}\left(\frac{\beta_2}{\beta_1}\right)^c \int_0^{\frac{\beta_1}{\beta_1+\beta_2}} z^{k+c-1}(1-z)^{-c-1} F\left(1-\alpha_2, 2; \alpha_1+2; \left(\frac{\beta_2}{\beta_1}\frac{z}{1-z}\right)^c\right) dz \\ &= \frac{c\alpha_2}{1+\alpha_1}\left(\frac{\beta_2}{\beta_1}\right)^c \int_0^{\frac{\beta_1}{\beta_2}} u^{k+c-1}(1+u)^{-k} F\left(1-\alpha_2, 2; \alpha_1+2; \left(\frac{\beta_2}{\beta_1}u\right)^c\right) du \\ &= \frac{c\alpha_2}{1+\alpha_1}\left(\frac{\beta_1}{\beta_2}\right)^k \int_0^1 v^{k+c-1}\left(1+\frac{\beta_1}{\beta_2}v\right)^{-k} F(1-\alpha_2, 2; \alpha_1+2; v^c) dv \\ &= \frac{c\alpha_2}{1+\alpha_1}\left(\frac{\beta_1}{\beta_2}\right)^k \sum_{i=0}^{\infty}\left(\frac{\beta_1}{\beta_2}\right)^i \frac{{}_{(-k)}P_i}{i!} \int_0^1 v^{k+c+i-1} F(1-\alpha_2, 2; \alpha_1+2; v^c) dv \\ &= \frac{\alpha_2}{1+\alpha_1}\left(\frac{\beta_1}{\beta_2}\right)^k \sum_{i=0}^{\infty}\left(\frac{\beta_1}{\beta_2}\right)^i \frac{{}_{(-k)}P_i}{i!} \int_0^1 v^{\frac{k+c+i}{c}-1} F(1-\alpha_2, 2; \alpha_1+2; v) dv \\ &= \frac{\alpha_2}{1+\alpha_1}\left(\frac{\beta_1}{\beta_2}\right)^k \sum_{i=0}^{\infty}\left(\frac{\beta_1}{\beta_2}\right)^i \frac{{}_{(-k)}P_i}{i!(k+c+i)} \\ &\quad \times {}_3F_2(1-\alpha_2, 2; (k+c+i)/c; \alpha_1+2, (k+2c+i)/c; 1), \qquad \text{if } \beta_2 > \beta_1, \end{aligned}$$

where ${}_{(a)}P_i = a(a-1)\cdots(a-i+1)$ and ${}_{(a)}P_0 = 0$.

Figure 2 shows the density curves and Table 2 provides the asymptotic means, variances, and coefficients of skewness for the distribution (7) for different combinations of α_1, α_2, when $c = 5$, $\beta_1 = 1$, and $\beta_2 = 2$. The figure also shows that the distribution is skewed to the left.

Table 2 indicates that

- the distribution is skewed to the left;
- its mean and variance increase slightly as α_1 increases when $\alpha_1 > \alpha_2$ but the mean slightly decreases and the variance slightly increases as α_1 increases when $\alpha_1 < \alpha_2$.

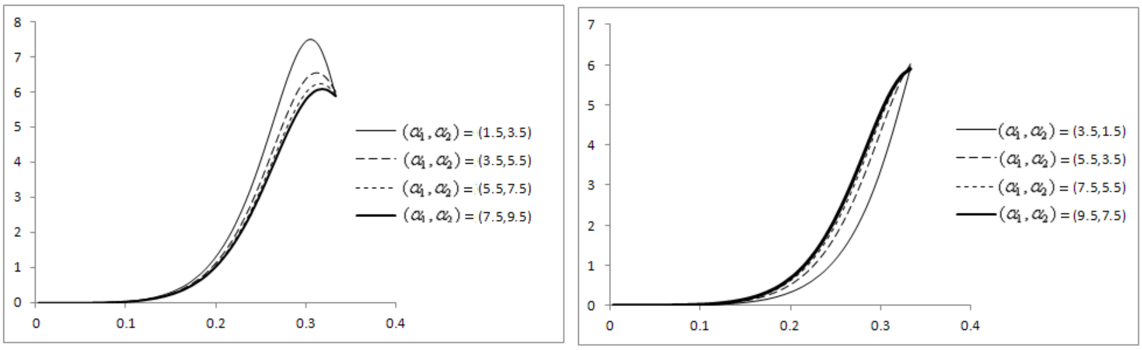

Figure 2: Density curves for $c = 5$, $\beta_1 = 1$, $\beta_2 = 2$ for $\alpha_1 < \alpha_2$ (left) and $\alpha_1 > \alpha_2$ (right).

Table 2: Asymptotic mean, variance, and coefficient of skewness for the distribution of R when $c = 5$ and $\beta_1 = 1$, $\beta_2 = 2$.

α_1, α_2	Mean	Variance	Skewness
(1.5, 3.5)	0.15237	0.01781	−0.7436
(3.5, 1.5)	0.07623	0.01824	−0.9984
(3.5, 5.5)	0.14275	0.01932	−0.4002
(5.5, 3.5)	0.10631	0.02001	−0.6023
(5.5, 7.5)	0.13727	0.01983	−0.2415
(7.5, 5.5)	0.10875	0.02064	−0.4527
(7.5, 9.5)	0.12844	0.01999	−0.1695
(9.5, 7.5)	0.11026	0.02091	−0.3931

4 Estimation of Pr(X < Y)

Now we attempt to find the uniformly minimum-variance unbiased estimator (UMVUE) of $\xi = \Pr(X < Y)$, where X and Y are independently distributed as exponentiated Pareto, having cdf's

$$G_X(x) = \left[1 - \left(\frac{\beta_1}{x}\right)^c\right]^{\alpha_1}, \qquad x \geq \beta_1\,,\ \alpha_1 > 0\,,\ c > 0$$

$$G_Y(y) = \left[1 - \left(\frac{\beta_2}{y}\right)^c\right]^{\alpha_2}, \qquad y \geq \beta_2\,,\ \alpha_2 > 0\,,\ c > 0\,.$$

From (3) we have

$$F_Q(1) = \xi = \Pr\left(\frac{Y}{X} > 1\right) = \begin{cases} 1 - F\left(-\alpha_2, 1; \alpha_1 + 1; \left(\frac{\beta_2}{\beta_1}\right)^c\right), & \text{if } \beta_2 \leq \beta_1, \\ F\left(-\alpha_1, 1; \alpha_2 + 1; \left(\frac{\beta_1}{\beta_2}\right)^c\right), & \text{if } \beta_2 > \beta_1, \end{cases}$$

where $F(a, b; c; z)$ is the hypergeometric function. For $\beta_1 = \beta_2$,

$$R = 1 - F(-\alpha_2, 1; \alpha_1 + 1; 1)\,,$$

which depends only on the exponentiating parameters. We shall assume that β_1, β_2 and c are known.

UMVUE of ξ:

Let $(X_1, \ldots, X_m)$ and $(Y_1, \ldots, Y_n)$ be independent random samples of sizes m and n from the distributions of X and Y, respectively, where $m, n > 2$.

Define $U_i = -\log(1 - (\beta_1/X_i)^c)$, $i = 1, \ldots, m$, and $V_i = -\log(1 - (\beta_2/Y_i)^c)$, $i = 1, \ldots, n$. Then $(U_1, \ldots, U_m)$ and $(V_1, \ldots, V_n)$ form independent random samples from an Exponential(α_1) and an Exponential(α_2) distribution, respectively.

An unbiased estimator of ξ is given by

$$Z = \begin{cases} 1, \text{ if } X_1 < Y_1, \\ 0, \text{ otherwise.} \end{cases}$$

Since the complete sufficient statistics for α_1 and α_2 are $U^* = \sum_{i=1}^m U_i$ and $V^* = \sum_{i=1}^n V_i$, respectively, by the Lehmann-Scheffé Theorem the UMVUE of R is

$$\begin{aligned} \hat{\xi} &= \mathrm{E}(Z|U^*, V^*) = \Pr(X_1 < Y_1|U^*, V^*) = \Pr(U_1 > V_1|U^*, V^*) \\ &= \Pr\left(\frac{U_1}{U^*} > \frac{V_1}{V^*}\frac{V^*}{U^*}\middle| U^*, V^*\right) . \end{aligned} \tag{8}$$

Since U_1/U^* and V_1/V^* are independently distributed, with $U_1/U^* \sim$ Beta$(1, m)$ and $V_1/V^* \sim$ Beta$(1, n)$, (8) gives

$$\hat{\xi} = \begin{cases} n\int_0^1 (1-v)^{n-1}(1-av)^m dv\,, & \text{if } a < 1, \\ n\int_0^{1/a} (1-v)^{n-1}(1-av)^m dv\,, & \text{if } a > 1, \end{cases}$$

where $a = V^*/U^*$.

UMVUE of ξ^2:

An unbiased estimator of ξ^2 is given by

$$Z_1 = \begin{cases} 1, \text{ if } X_1 < Y_1, X_2 < Y_2, \\ 0, \text{ otherwise.} \end{cases}$$

Hence, by the Lehmann-Scheffé Theorem the UMVUE of R^2 is

$$\widehat{\xi^2} = \Pr\left(\frac{U_1}{U^*} > \frac{V_1}{V^*}\frac{V^*}{U^*}, \frac{U_2}{U^*} > \frac{V_2}{V^*}\frac{V^*}{U^*}\middle| U^*, V^*\right) .$$

$T_1 = U_1/U^*$ and $T_2 = U_2/U^*$ are jointly distributed with pdf

$$f_{T_1,T_2}(t_1, t_2) = \frac{\Gamma(m)}{\Gamma(m-2)\Gamma^2(1)}(1 - t_1 - t_2)^{m-3}\,, \qquad t_1, t_2 > 0\,, \quad t_1 + t_2 \leq 1\,.$$

Similarly, $W_1 = V_1/V^*$ and $W_2 = V_2/V^*$ are jointly distributed with pdf

$$f_{W_1,W_2}(w_1, w_2) = \frac{\Gamma(n)}{\Gamma(n-2)\Gamma^2(1)}(1 - w_1 - w_2)^{n-3}\,, \qquad w_1, w_2 > 0\,, \quad w_1 + w_2 \leq 1\,.$$

Hence,

$$\widehat{\xi}^2 = \int\limits_{\substack{w_1>0,w_2>0\\ w_1+w_2\le 1}} \Pr(T_1 > aW_1, T_2 > aW_2) f_{W_1,W_2}(w_1,w_2)dw_1dw_2\,.$$

Now, for given $W_1 = w_1, W_2 = w_2$,

- if $a \ge 1$

$$\Pr(T_1 > aw_1, T_2 > aw_2) = \begin{cases} 0, & \text{if } w_1 + w_2 \ge 1/a, \\ \int\limits_{aw_1}^{1-aw_2} \int\limits_{aw_2}^{1-t_1} f_{T_1,T_2}(t_1,t_2)dt_2dt_1, & \text{if } 0 < w_1, w_2 < 1/a, \\ & \quad w_1 + w_2 \le 1/a. \end{cases}$$

- if $a \le 1$

$$\Pr(T_1 > aw_1, T_2 > aw_2) = \begin{cases} 0, & \text{if } w_1 + w_2 \ge 1, \\ \int\limits_{aw_1}^{1-aw_2} \int_{aw_2}^{1-t_1} f_{T_1,T_2}(t_1,t_2)dt_2dt_1 & \text{if } 0 < w_1\ ,w_2 < 1, \\ & \quad w_1 + w_2 \le 1. \end{cases}$$

Applying the transformation $T_3 = \frac{T_2}{1-T_1}$, we note that T_1 and T_3 are independently distributed with $T_1 \sim \text{Beta}(1, m-1)$ and $T_3 \sim \text{Beta}(1, m-2)$. Then, for given $W_1 = w_1$, $W_2 = w_2$,

$$\Pr(T_1 > aw_1, T_2 > aw_2) = \begin{cases} 0, & \text{if } a \ge 1,\, w_1 + w_2 \ge 1/a \\ (m-1)(aw_2)^{m-2}(1-aw_1), & \text{if } a \ge 1,\, 0 < w_1, w_2 < 1/a, \\ & \quad w_1 + w_2 \le 1/a \\ & \text{or } a \le 1,\, 0 < w_1, w_2 < 1, \\ & \quad w_1 + w_2 \le 1. \end{cases}$$

Therefore, we have the following:

- if $a \le 1$

$$\begin{aligned}
\widehat{\xi}^2 &= \int\limits_{\substack{w_1>0,w_2>0\\ w_1+w_2\le 1}} \Pr(T_1 > aW_1, T_2 > aW_2) f_{W_1,W_2}(w_1,w_2)dw_1dw_2 \\
&= \frac{\Gamma(m-2)\Gamma(n-2)}{\Gamma(m+n-3)} a^{m-2} \int_0^1 (1-w_1)^{m+n-4}(1-aw_1)dw_1 \\
&= \Gamma(m-2)\Gamma(n-2)a^{m-2} \left\{ \frac{1}{\Gamma(m+n-2)} - \frac{a}{\Gamma(m+n-1)} \right\},
\end{aligned}$$

- if $a > 1$

$$\widehat{\xi^2} = \int\limits_{\substack{0 < w_1 < 1/a, 0 < w_2 < 1/a \\ w_1 + w_2 \le 1/a}} \Pr(T_1 > aW_1, T_2 > aW_2) f_{W_1,W_2}(w_1, w_2) dw_1 dw_2$$

$$= \frac{\Gamma(m-2)\Gamma(n-2)}{\Gamma(m+n-3)} a^{m-2} \int_0^{1/a} B_{\frac{1-aw_1}{a(1-w_1)}}(m-1, n-2)(1-w_1)^{m+n-4}(1-aw_1) dw_1$$

$$= \frac{\Gamma(m-2)\Gamma(n-2)}{\Gamma(m+n-3)} a^{m-2} \int_0^{1/a} B_{\frac{1-aw_1}{a(1-w_1)}}(m-1, n-2)(1-w_1)^{m+n-4}(1-aw_1) dw_1 \, ,$$

where

$$B_x(r,k) = \int_0^x z^{r-1}(1-z)^{k-1} dz \, .$$

The UMVUE of $\mathrm{var}(\hat{\xi})$ is then given by $\widehat{\mathrm{var}}(\hat{\xi}) = \hat{\xi}^2 - \widehat{\xi^2}$.

5 Data Analysis

We now analyze two sets of data, a simulated one and a real life one, to give UMVUEs of $\xi = \Pr(X < Y)$ and its variance.

Example 1 (Simulated Data)

The following two independent samples were generated from exponentiated Pareto distributions with parameters $\alpha_1 = 2$, $\beta_1 = 1.5$, $c = 2$ and $\alpha_2 = 1.5$, $\beta_2 = 1.25$, $c = 2$, respectively. The first distribution corresponds to that of X and the second one to that of Y. The observations are as follows:

Sample 1: 1.775, 1.571, 2.484, 4.258, 1.518, 1.509, 1.609, 1.558, 4.425, 1.601, 6.511, 1.513, 7.986, 1.998, 8.316, 2.075, 1.503, 2.089, 1.508, 5.544, 1.506, 2.282, 2.093, 1.682, 1.817.

Sample 2: 2.110, 1.255, 1.559, 1.256, 1.436, 1.549, 1.402, 1.266, 1.708, 1.370, 1.689, 1.253, 1.910, 2.122, 1.455, 1.262, 1.804, 6.916, 1.352, 2.105, 2.047, 1.646, 3.648, 1.259, 1.500, 1.317, 1.311, 1.253, 1.643, 1.907.

The true value of $\xi = \Pr(X < Y)$ is 0.3145. Here, $a = 1.22898$. Hence, the UMVUE of ξ and its variance are $\widehat{\xi} = n \int_0^{1/a} (1-v)^{n-1}(1-av)^m dv = 0.4931$ and $\hat{\mathrm{var}}(\hat{\xi}) = 0.2432$, respectively.

Example 2 (Real Life Data)

Data on the major rice crop in the crop year 2001-2002 (April 1, 2001 to March 31, 2002) from two Tambols – Nongyang and Nonghan - of Amphoe Sansai in the Chiang Mai province, Thailand, have been used as samples (data source Watthanacheewakul and Suwattee, 2010). Each Tambol consists of a set of 228 and 281 farmer households, respectively. Samples of sizes 23 and 28 were drawn from each Tambol. The sampled data are shown in Table 3.

We first fitted exponentiated Pareto distributions with parameters α_1, β_1, c and α_2, β_2, c, respectively, to the data on X and Y and then checked the goodness of fit using exponentiated Pareto probability plots.

Table 3: The Major Rice Crop in Kilograms from the two Tambols for the Crop Year 2001-2002 (April 1, 2001 to March 31, 2002).

Nongyang X	Nonghan Y	Nongyang X	Nonghan Y
3440	2300	18000	3200
3200	1800	3150	1900
5400	2000	8500	2200
3800	2400	4500	3500
4300	12000	4250	6000
7000	2800	3500	2300
3700	3000	3000	2850
6000	2250	3600	3700
3250	2100	5000	1900
3500	2700	3000	
3000	1800	4000	
3400	3500	3600	
5000	2500	4270	
3600	2600	5800	

Let $(X_1, \ldots, X_m)$ and $(Y_1, \ldots, Y_n)$ denote the random samples on X and Y, and let the corresponding ordered statistics be $X_{(1)} < \cdots < X_{(m)}$ and $Y_{(1)} < \cdots < Y_{(n)})$. Then, the likelihood function for $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2, c)$ is given by

$$\begin{aligned} L(\theta) &= c^{m+n}\alpha_1^m \alpha_2^n \beta_1^{cm} \beta_2^{cn} \prod_{i=1}^{m} x_{(i)}^{-(c+1)} \prod_{i=1}^{n} y_{(i)}^{-(c+1)} \\ &\times \prod_{i=1}^{m} (1 - (\beta_1/x_{(i)})^c)^{\alpha_1 - 1} \prod_{i=1}^{n} (1 - (\beta_2/y_{(i)})^c)^{\alpha_2 - 1} I_{x_{(1)}>\beta_1, y_{(1)}>\beta_2}\,, \end{aligned}$$

where the indicator function is defined as

$$I_{x_{(1)}>\beta_1, y_{(1)}>\beta_2} = \begin{cases} 1, \text{ if } x_{(1)} > \beta_1, y_{(1)} > \beta_2, \\ 0, \text{ otherwise.} \end{cases}$$

Now, for $\alpha_1 < 1$, $\alpha_2 < 1$, as β_1 approaches $x_{(1)}$ and β_2 approaches $y_{(1)}$, $L(\theta)$ tends to ∞. This implies that the maximum likelihood estimators (MLEs) of α_1, α_2, and c do not exist. We therefore obtain modified MLEs of the unknown parameters using the procedure proposed by Raqab, Madi, and Kundu (2007).

Since the likelihood function is maximized at $(\beta_1, \beta_2) = (x_{(1)}, y_{(1)})$, the modified MLEs of β_1 and β_2 are $\hat{\beta}_1 = x_{(1)}$, $\hat{\beta}_2 = y_{(1)}$. The modified likelihood function for

$\theta^* = (\alpha_1, \alpha_2, c)$ is then given by

$$L_{mod}(\theta^*) = c^{m^*+n^*} \alpha_1^{m^*} \alpha_2^{n^*} x_{(1)}^{cm^*} y_{(1)}^{cn^*} \prod_{i=2}^{m} x_{(i)}^{-(c+1)} \prod_{i=2}^{n} y_{(i)}^{-(c+1)} \prod_{i=2}^{m} (1 - (x_{(1)}/x_{(i)})^c)^{\alpha_1 - 1} \times \prod_{i=2}^{n} (1 - (y_{(1)}/y_{(i)})^c)^{\alpha_2 - 1} ,$$

where $m^* = m - 1$, $n^* = n - 1$, and the log likelihood is given by

$$\begin{aligned} \log L(\theta) = &(m^* + n^*) \log(c) + m^* \log(\alpha_1) + n^* \log(\alpha_2) + c(m^* \log(x_{(1)}) + n^* \log(y_{(1)})) \\ &-(c+1) \left(\sum_{i=2}^{m} \log(x_i) + \sum_{i=2}^{n} \log(y_i) \right) + (\alpha_1 - 1) \sum_{i=2}^{m} \log(1 - (x_{(1)}/x_{(i)})^c) \\ &+(\alpha_2 - 1) \sum_{i=2}^{n} \log(1 - (y_{(1)}/y_{(i)})^c) . \end{aligned}$$

The modified MLEs of α_1 and α_2 are then obtained as

$$\hat{\alpha}_1 = -\frac{m^*}{\sum_{i=2}^{m} \log(1 - (x_{(1)}/x_{(i)})^c)} , \qquad \hat{\alpha}_2 = -\frac{n^*}{\sum_{i=2}^{n} \log(1 - (y_{(1)}/y_{(i)})^c)}$$

and the modified MLE $\hat{c}$ of c is a solution of the non-linear equation

$$c = g(c) ,$$

where

$$\begin{aligned} g(c) = (m^* + n^*) &\left[(\hat{\alpha}_1 - 1) \left\{ \sum_{i=2}^{m} \frac{\left(\frac{x_{(1)}}{x_{(i)}}\right)^c \log\left(\frac{x_{(1)}}{x_{(i)}}\right)}{1 - \left(\frac{x_{(1)}}{x_{(i)}}\right)^c} \right\} \right. \\ &\left. +(\hat{\alpha}_2 - 1) \left\{ \sum_{i=2}^{n} \frac{\left(\frac{y_{(1)}}{y_{(i)}}\right)^c \log\left(\frac{y_{(1)}}{y_{(i)}}\right)}{1 - \left(\frac{y_{(1)}}{y_{(i)}}\right)^c} \right\} + \sum_{i=2}^{m} \log\left(\frac{x_{(i)}}{x_{(1)}}\right) + \sum_{i=2}^{n} \log\left(\frac{y_{(i)}}{y_{(1)}}\right) \right]^{-1} . \end{aligned}$$

For the given data set, the modified MLEs of the unknown parameters came out as $\hat{\beta}_1 = 1800$, $\hat{\beta}_2 = 3000$, $\hat{c} = 3.1399$, $\hat{\alpha}_1 = 1.8014$, $\hat{\alpha}_2 = 1.5482$. The goodness of fit of the distributions has been checked using probability plots (see Figure 3). These plots show that the exponentiated Pareto distribution fits fairly well to the two data sets.

Assuming the true values of the parameters β_1, β_2, and c to be known and given by their modified MLEs, i.e., $\beta_1 = 1800$, $\beta_2 = 3000$, $c = 3.1399$, we obtain the UMVUE of $\xi = \Pr(X < Y)$ as $\widehat{\xi} = 0.5714$, and $\widehat{\mathrm{var}}(\hat{\xi}) = 0.3265$.

6 Conclusion

The paper studies the distributions of the ratios Y/X and $X/(X+Y)$, when X and Y are distributed independently as exponentiated Pareto. These distributions find importance in

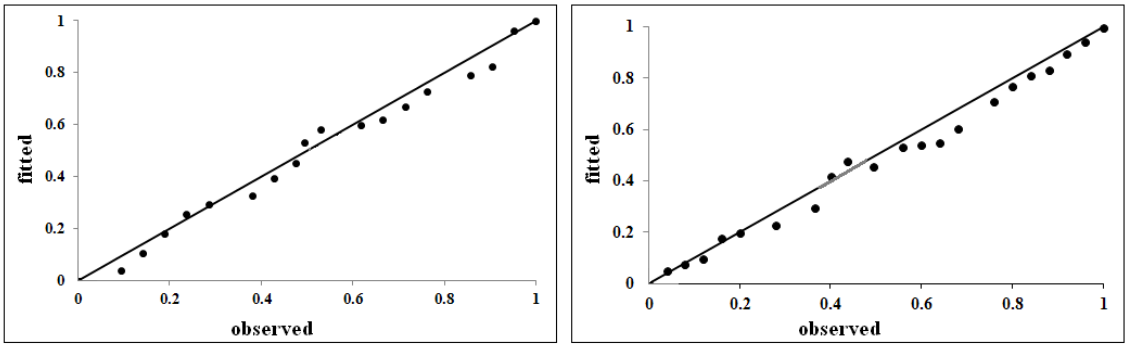

Figure 3: Probability plots for the distributions fitted to data set 1 (left) and 2 (right).

many situations, like studying the proportion of human settlement in a place, proportion of oil reserves in an oil field, etc. They are also useful in computing the reliability function $\Pr(X < Y)$, when X and Y denote the stress and strength variables, respectively. The paper, further, finds the UMVUE of $\Pr(X < Y)$, and also UMVUE of its variance. A simulated data set and a real data set are analyzed to estimate $\Pr(X < Y)$ and its variance.

Acknowledgement

The authors thank the Editor and the anonymous referee for the fruitful suggestions, which immensely helped to improve the presentation.

References

Abramowitz, M., and Stegun, I. A. (1972). *Handbook of Mathematical Functions*. New York: Dover Publications, Inc.

Ali, M., Pal, M., and Woo, J. (2005). Inference on P(Y<X) in generalized uniform distributions. *Calcutta Statistical Association Bulletin*, *57*, 35-48.

Ali, M. M., Pal, M., and Woo, J. (2007). Some exponentiated distributions. *The Korean Communications in Statistics*, *14*, 93-109.

Ali, M. M., Pal, M., and Woo, J. (2009). Estimation of Pr(Y<X) when X and Y belong to different distribution families. *Journal of Probability and Statistical Science*, *8*, 19-33.

Gradshteyn, I. S., and Ryzhik, I. M. (1965). *Tables of Integral, Series, and Products*. New York: Academic Press.

Gupta, R. C., Gupta, P. L., and Gupta, R. D. (1998). Modelling failure time data by Lehman alternatives. *Communications in Statistics, Theory and Methods*, *27*, 887-904.

Gupta, R. D., and Kundu, D. (2001). Exponentiated exponential family: An alternative to Gamma and Weibull distributions. *Biometrical Journal*, *43*, 117-130.

Pal, M., Ali, M. M., and Woo, J. (2005). Estimation and testing of P(Y>X) in two-parameter exponential distributions. *Statistics*, *39*, 415-428.

Pal, M., Ali, M. M., and Woo, J. (2006). Exponentiated Weibull distribution. *Statistica*, *66*, 139-148.

Raqab, M. Z., Madi, M. T., and Kundu, D. (2007). Estimation of P(Y<X) for a 3-parameter generalized exponential distribution. *Communications in Statististics, Theory and Methods*, *37*, 2854-2864.

Watthanacheewakul, L., and Suwattee, P. (2010, March 17-19, 2010). Comparisons of several Pareto population means. In *Proceedings of the International MultiConference of Engineers and Computer Scientists 2010* (Vol. III). Hong Kong.

Authors' addresses:

M. Masoom Ali
Department of Mathematical Sciences
Ball State University
Muncie, IN 47306
USA

Manisha Pal
Department of Statistics
Calcutta University
35 Ballugumge Circular Road
Kolkata - 700019
India
E-mail: manishapal2@gmail.com

Jungsoo Woo
Department of Statistics
Yeungnam University
Gyongsan 712-749
South Korea