A Robust Estimator of $R = P(X > Y)$ of Heavy-tailed Distributions and its Sampling Distributions

Dais George

Catholicate College, Pathanamthitta, Kerala, India

*Corresponding Author: daissaji@rediffmail.com

Copyright ©2014 Horizon Research Publishing All rights reserved.

Abstract Heavy-tailed distributions have wide applications in life-time contexts, especially in reliability and risk modeling. So we consider the estimation problem of reliability $R = P(X > Y)$ for small samples, when $X$ and $Y$ are two independent but not identically distributed random variables belonging to the family of heavy-tailed distributions, using a robust estimator, namely the harmonic moment estimator. Extensive simulation studies are carried out to study the performance of this estimator. The relative efficiency of the estimator with the well known Hill estimator is studied. We obtain the sampling distribution of the parameters of the distribution as well as that of estimator of $R$ which will help us to study the properties of the estimators. Also we find out the asymptotic confidence intervals of $R$ and its performance is studied with respect to average width and the coverage probability, through simulations.

Keywords Harmonic Moment Estimator, Heavy-tailed Distributions, Pareto Distribution, Reliability Function

1 Introduction

Heavy-tails are characteristics of many phenomena where the probability of a single huge value impacts heavily. Record-breaking insurance losses, financial log returns, file sizes stored on a server, transmission rates of files etc. are examples of random variables following heavy-tailed distributions. The distribution of a random variable $X$ is said to have heavy-tail if

$$P[X > x] = x^{-\alpha}L(x),$$

where $L$ is slowly varying. That is, for $x > 0$

$$\lim_{t \to \infty} \frac{L(tx)}{L(t)} = 1.$$

Regardless of the distribution for small values of the random variable if the asymptotic shape of the distribution is hyperbolic, it is heavy tailed. Heavy-tailed distributions have wide applications in lifetime contexts, especially in reliability and risk modeling. We consider the estimation problem of reliability $R = P(X > Y)$ for small and large samples, when $X$ and $Y$ are two independent but not identically distributed random variables belonging to the family of heavy-tailed distributions. The heavy-tailed distributions like the student’s $t$ and Pareto families have been used to model data with high kurtosis see, [4]. The simplest heavy tailed distribution is the Pareto distribution. It is hyperbolic over its entire range.

The Pareto distribution is a power law probability distribution that coincides with social, scientific, geophysical, actuarial and many other types of observable phenomena. The univariate Pareto distribution is a simple model for non-negative data with a power law probability tail, at least approximately for large values of $X$. It is a useful model in the analysis of income data, reliability studies, risk modeling and business failure data [14]. Reference [1] gave an extensive historical survey of its uses in the context of income distribution. References [6], [16] and [19] discussed the application of the Pareto distribution in various fields. Moreover Pareto Type I distribution has a position of importance in the field of life testing because this distribution can be considered as a distribution of failure time of a component as follows.

Suppose the life time of a particular component have an exponential distribution with failure rate $\lambda$ and threshold parameter or guarantee time $c$. Now in a population of components, there could be variation in $\lambda$ values because of small fluctuations in manufacturing tolerances. So the component can be regarded as one belonging to a random sub population (see, [15]). The effect of the changes in $\lambda$ is accommodated by assuming $\lambda$ as a random variable following a gamma distribution with scale parameter $k$ and shape parameter $\alpha$.

Then the failure time $X$ of a component selected at random from such a mixed population has a Pareto distribution [12] with density function

$$f(x; \alpha, c) = \int_0^\infty \lambda e^{-\lambda(x-c)} \frac{\alpha^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda} d\lambda = \alpha c^\alpha x^{-(\alpha+1)}, \quad x > c, \quad \alpha > 0.$$
Stress-strength model describes the life of a component which has a random strength (X), subject to random stress (Y). The component fails at the instant that the stress applied to it exceeds the strength and the component will function satisfactorily when \( X > Y \). Reliability of a component (R) during a given period is taken to be the probability that the strength exceeds the stress \( P(X > Y) \) during the interval.

The problem of estimating \( R \) when X and Y are independent exponential random variables is discussed in [18], [13] and [3]. In [2], the author studied the case when X and Y are independent Burr Random variables of Type XII. For more studies see, [17], [7], [9], [8] and [5].

Because of the real life applications of the Pareto distribution and because most of the real data (heavily-tailed) under consideration recently are Pareto tailed, the estimate of \( R \) and related inference is important. So in this study we are estimating the desired probability \( R = P(X > Y) \) where X and Y are two independent but not identically distributed random variables following Pareto type I distribution, with parameters \( \alpha_1 \) and \( \alpha_2 \). This paper is organized as follows. Section 2 discusses about Pareto type I distribution. Section 3 focus on the estimation of tail index using the harmonic moment estimation method. Estimation of reliability function of the Pareto type I distribution using the above mentioned tail index estimator with its performance have been successfully studied in Section 4. We obtain the sampling distribution of the harmonic moment estimator of tail index and of reliability R in Section 5. In Section 6 we estimate the asymptotic confidence interval of R and its performance is studied.

2 Pareto Type-I Distribution

The simplest heavy-tailed distribution is the Pareto distribution which is hyperbolic over its entire range and has the probability density function,

\[ P(x) = \alpha c^\alpha x^{-(\alpha + 1)}, \quad \alpha, c > 0, x \geq c \]  

(2)

and its cumulative distribution function is given by:

\[ F(x) = P[X \leq x] = 1 - (c/x)^\alpha, \quad x \geq c \]

where \( \alpha \) is the shape parameter and tail index and \( c \) represents the smallest value, that the random variable can take. As \( \alpha \) decreases, an arbitrarily large portion of the probability mass may be present in the tail of the distribution.

3 Harmonic Moment Estimator

To obtain the estimate of \( R \), we have to estimate \( \alpha_1 \) and \( \alpha_2 \). For that we are using the harmonic moment estimator.

The harmonic moment estimator is introduced in [10] for estimating the tail index of distributions, in the maximum domain of attractions of the Frechet distribution, with \( F(x) \) as in (1). Some distributions satisfying (1) are the Burr, Frechet, F, inverse Gamma, log Gamma and Pareto. For more details see, [10]. This tail index estimator is important because one uses only a small number of the largest order statistics when estimating the tail index.

For the Pareto type I distribution with probability density function (2), we obtain the estimate of \( \alpha \) as

\[ \hat{\alpha} = \frac{\frac{1}{\theta} \sum_{i=1}^{n} \left( \frac{x_i}{c} \right)^{\frac{1}{\theta}}} {\theta \left( 1 - \frac{1}{\theta} \sum_{i=1}^{n} \left( \frac{x_i}{c} \right)^{\frac{1}{\theta}} \right)}, \quad \theta \in R^+ \]  

(3)

Here we are assumed that there is an exact Pareto tail beyond some high threshold \( u \), but the form \( F(x) \) is not specified for \( x < u \).

That is

\[ \exists \ u > 0 \ \text{such that} \ x > u \ \Rightarrow \ P(X > x) = dx^{-\alpha}; \ d, \ \alpha > 0 \]

so that the harmonic moment estimator for \( \alpha \) using the upper order statistics (descending) above a random threshold \( X_{(k+1)} \) is

\[ \hat{\alpha}_{k,n}(\theta) = \frac{\frac{1}{\theta} \sum_{i=1}^{k} \left( \frac{x_{(k+1)}}{c} \right)^{\frac{1}{\theta}}} {\theta \left( 1 - \frac{1}{\theta} \sum_{i=1}^{k} \left( \frac{x_{(k+1)}}{c} \right)^{\frac{1}{\theta}} \right)}, \quad \theta \in R^+, \ k \in \{1, 2, \ldots, n-1\} \]

Moreover it is a robust estimator.

The harmonic moment estimator satisfies certain very important properties. For details see, [10].

4 Estimation of R

Let X and Y be random variables distributed as Pareto type I distribution with parameters \( \alpha_1 \) and \( \alpha_2 \) respectively.

The probability density functions of X and Y are given by

\[ f(x) = \alpha_1 c^{\alpha_1} x^{-(\alpha_1 + 1)}, \quad \alpha_1, c > 0, \ x \geq c \]  

(5)

and

\[ f(y) = \alpha_2 c^{\alpha_2} y^{-(\alpha_2 + 1)}, \quad \alpha_2, c > 0, \ y \geq c \]  

(6)

Using equations (5) and (6), the function R will be

\[ R = \int_c^\infty f(x) F_Y(x) dx \]

(7)

\[ R = \frac{\alpha_2}{\alpha_1 + \alpha_2} \]

Clearly R depends only on \( \alpha_1 \) and \( \alpha_2 \). As \( \alpha_1 \to \alpha_2 \), using L’Hospital rule, \( P(X > Y) \to 1/2 \). Note that

\[ R = \frac{1}{1+a}, \quad a = \frac{\alpha_1}{\alpha_2} \]

and

\[ \frac{dR}{da} = -1 \left( \frac{1}{(1+a)^2} \right) \leq 0. \]
that is for fixed $\alpha$, the MSE is symmetric with respect to $\alpha$. Thus the choice of the tuning parameter $\alpha$ in a particular case is important.

We have conducted a simulation study to estimate $R$ of the Pareto type I distribution using simulated random samples from the Pareto type I distribution with sample sizes $n=15, 20$ and $30$ with $\alpha=0.8$ and $\alpha=1.6$.

The following steps will be considered for obtaining the numerical results.

**Step I** Generate $1000$ random samples $X_i$, $i=1, 2, \ldots, n$ from the Pareto type I distribution with sample sizes $n=15, 20$ and $30$ with $\alpha=0.8$ and $\alpha=1.6$.

**Step II** Using the harmonic moment estimator, obtain $1000$ estimates of $\hat{\alpha}_1$.

**Step III** Similarly, generate $1000$ random samples from the Pareto distribution with parameter $\alpha_2$ and obtain $\hat{\alpha}_2$.

**Step IV** Using equation (7), we obtain the estimate of $R$ using the harmonic moment estimates of $\alpha_1$ and $\alpha_2$ respectively.

### 4.1 Performance Study

An extensive numerical investigation is carried out to study the performance of the estimator using the simulated random samples from the Pareto Type I distribution in terms of average bias and average MSE.

The results are given in Tables 1 and 2. The absolute bias of the harmonic moment estimate is very small. The MSE is symmetric with respect to $\alpha_1$ and $\alpha_2$. Though we have considered different choice for the parameter $\alpha_1$, there is no appreciable difference in the average bias and average MSE accordingly.

### Remark:

Given $X^{(k+1)} \geq u$, a choice of $\theta$ is given by

$$\theta = \sqrt{\frac{k^2 + 8k + k}{2\alpha}} < \infty$$

minimizes $M.S.E(\hat{\alpha}_{k,n}(\theta))$.

The harmonic moment estimator to the well known Hill estimator [11] was compared by Henry in [10] and it was shown that the harmonic moment estimator outperforms the Hill estimator for estimating the tail index of a distribution in small sample situations, in terms of bias, variance and mean squared error. Also the harmonic moment estimator works well in terms of efficiency for large $\theta$. That is for $\theta > (\sqrt{1/\epsilon - 1})/\alpha$, since the asymptotic relative efficiency of Hill estimator to $\hat{\alpha}_{k,n}(\theta)$ is

$$\lim_{k(n) \to \infty} \frac{AV ar(Hill_{k,n})}{AV ar(\hat{\alpha}_{k,n}(\theta))} = \frac{\theta_\alpha(\theta^2 + 2)}{(\theta + 1)^2} = 1 - (\alpha + 1)^{-2}.$$

Then $(\alpha^\theta + 1)^{-2} < \epsilon \leftrightarrow \theta > (\sqrt{1/\epsilon - 1})/\alpha$.

Moreover, the harmonic moment estimator is a robust estimator. For the Hill estimator, the asymptotic B.P is zero where as for harmonic moment estimator, the asymptotic B.P is in $(0, \frac{1}{2})$ depending on the true value of $\alpha$ and the choice of the tuning parameter $\theta$.

### 5 Sampling distribution of Estimators

In this section we obtain the sampling distribution of the parameters of the Pareto type I distribution and harmonic moment estimators of reliability $R$. We apply the Pearson’s technique for the two estimators to obtain the sampling distribution. Knowing the distribution, we can study the properties of the estimators.

Pearson’s criterion for fixing the distribution family in a particular case is

$$K = \frac{\beta_1(\beta_2 + 3)^2}{4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)}$$

where $\beta_1$ and $\beta_2$ are the measures of skewness and kurtosis respectively. According to the values of $K$, there are seven types of Pearson curves. They are:

- Type I when $K < 0$
- Type II when $K = 0$, $\beta_1 = 0$ and $\beta_2 < 3$
- Type III when $2\beta_2 = 3\beta_1 - 6 = 0$ and $K = \infty$
- Type IV when $0 < K < 1$
- Type V when $K = 1$
- Type VI when $K > 1$
- Type VII when $K = 0$, $\beta_1 = 0$ and $\beta_2 > 3$

The steps to be followed for obtaining the sampling distribution of the estimators are

**Step 1:** Generate $1000$ random samples $X_1, X_2, \ldots, X_n$ from Pareto type I distribution with sample sizes $n=15, 20, 30$ with shape parameter $\alpha_1=0.5$ and $1.6$. Using these data we obtain $1000$ estimates of the harmonic moment estimates of $\alpha_1$. Similarly we can obtain the harmonic moment estimates of Pareto type I with parameter $\alpha_2$. Then using (7) we can obtain $1000$ estimates of $R$.

### Table 1. Estimators of $R$ with $\alpha_1 = 0.8$, $\alpha_2 = 1.5$

<table>
<thead>
<tr>
<th>(n,m)</th>
<th>$R_{Har}$</th>
<th>Bias</th>
<th>M.S.E</th>
</tr>
</thead>
<tbody>
<tr>
<td>(15,15)</td>
<td>0.6528</td>
<td>-0.0006</td>
<td>0.004</td>
</tr>
<tr>
<td>(15,20)</td>
<td>0.6509</td>
<td>-0.0002</td>
<td>0.003</td>
</tr>
<tr>
<td>(15,30)</td>
<td>0.6510</td>
<td>-0.0012</td>
<td>0.0021</td>
</tr>
<tr>
<td>(20,15)</td>
<td>0.6536</td>
<td>0.0013</td>
<td>0.0048</td>
</tr>
<tr>
<td>(20,20)</td>
<td>0.6532</td>
<td>0.0015</td>
<td>0.0052</td>
</tr>
<tr>
<td>(20,30)</td>
<td>0.6527</td>
<td>0.0026</td>
<td>0.0056</td>
</tr>
<tr>
<td>(30,15)</td>
<td>0.6524</td>
<td>0.0004</td>
<td>0.0034</td>
</tr>
<tr>
<td>(30,20)</td>
<td>0.6512</td>
<td>0.0002</td>
<td>0.0028</td>
</tr>
<tr>
<td>(30,30)</td>
<td>0.6503</td>
<td>-0.0031</td>
<td>0.0042</td>
</tr>
</tbody>
</table>

### Table 2. Estimators of $R$ with $\alpha_1 = 1.6$, $\alpha_2 = 1.5$

<table>
<thead>
<tr>
<th>(n,m)</th>
<th>$R_{Har}$</th>
<th>Bias</th>
<th>M.S.E</th>
</tr>
</thead>
<tbody>
<tr>
<td>(15,15)</td>
<td>0.4840</td>
<td>0.00001</td>
<td>0.00009</td>
</tr>
<tr>
<td>(15,20)</td>
<td>0.4841</td>
<td>-0.0009</td>
<td>0.0008</td>
</tr>
<tr>
<td>(15,30)</td>
<td>0.4820</td>
<td>-0.0079</td>
<td>0.0074</td>
</tr>
<tr>
<td>(20,15)</td>
<td>0.4849</td>
<td>0.0013</td>
<td>0.0008</td>
</tr>
<tr>
<td>(20,20)</td>
<td>0.4844</td>
<td>0.0016</td>
<td>0.0043</td>
</tr>
<tr>
<td>(20,30)</td>
<td>0.4845</td>
<td>-0.0001</td>
<td>0.0062</td>
</tr>
<tr>
<td>(30,15)</td>
<td>0.4862</td>
<td>0.0024</td>
<td>0.0012</td>
</tr>
<tr>
<td>(30,20)</td>
<td>0.4858</td>
<td>0.0033</td>
<td>0.0036</td>
</tr>
<tr>
<td>(30,30)</td>
<td>0.4846</td>
<td>0.0006</td>
<td>0.0052</td>
</tr>
</tbody>
</table>
Table 3. The Sampling Distribution for Harmonic Moment Estimator of Parameter $\alpha_1$ and Reliability $R$:

<table>
<thead>
<tr>
<th>$\alpha_2$, $\alpha_1$</th>
<th>For $\alpha_1$</th>
<th>For $R_{Har}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_2$</td>
<td>$\alpha_1$</td>
<td>n</td>
</tr>
<tr>
<td>1.5</td>
<td>0.5</td>
<td>15</td>
</tr>
<tr>
<td>20</td>
<td>0.0017</td>
<td>IV</td>
</tr>
<tr>
<td>30</td>
<td>-0.0017</td>
<td>I</td>
</tr>
<tr>
<td>1.5</td>
<td>1.6</td>
<td>15</td>
</tr>
<tr>
<td>20</td>
<td>-0.0007</td>
<td>I</td>
</tr>
<tr>
<td>30</td>
<td>-0.2901</td>
<td>I</td>
</tr>
</tbody>
</table>

Step 2: Calculate Pearson’s coefficient $K$ and then determine the types of Pearson curves.

Table 2 shows the sampling distribution of the parameter $\alpha_1$ and of Reliability $R$. Since the harmonic moment estimators outperform for small samples we conduct the numerical study of harmonic moment estimators for sample sizes $n=15, 20$ and $30$. Using SAS package, we did the computational works.

From Table 3, it can be noted that the sampling distributions of the harmonic moment estimators of the shape parameter ($\alpha$) are of type I and IV when $\alpha_1 < \alpha_2$ and $\alpha_1 > \alpha_2$ where as the sampling distributions of the Harmonic moment estimates of the reliability function $R$ are of Types I and IV, when $\alpha_1 < \alpha_2$ and type I only when $\alpha_1 > \alpha_2$.

Each type of Pearson’s system has a specific probability density function. The type I Pearson curves has the density function

$$f(x) = k \left(1 + \frac{x}{\alpha_1}\right)^{L_1} \left(1 - \frac{x}{\alpha_2}\right)^{L_2}, \quad -\alpha_1 < x < \alpha_2;$$

$L_1 < -1$, where $\alpha_1$, $\alpha_2$, $L_1$ and $L_2$ are the parameters of the family of distributions and $k$ is constant. Type I is Beta distribution.

The density function of the type IV Pearson curves is,

$$f(x) = k \left(1 + \frac{x^2}{a^2}\right)^{-\lambda} \exp\left(-\nu \tan^{-1}\frac{x}{a}\right)^{L_2},$$

$$-\infty < x < \infty,$$

where $k$ is a constant and $a, \lambda$ and $\nu$ are the parameters of the distribution of type IV Pearson curve.

6 Asymptotic Confidence Interval of $R$

The point estimator of $R$ is obtained using harmonic moment estimates of $\alpha_1$ and $\alpha_2$ as $R_{Har} = R(\hat{\alpha}_1, \hat{\alpha}_2)$.

Then using delta method,

$$b_1(\alpha_1, \alpha_2) = \frac{-\alpha_2}{(\alpha_1 + \alpha_2)^2}$$ (8)

$$b_2(\alpha_1, \alpha_2) = \frac{-\alpha_1}{\alpha_2} b_1(\alpha_1, \alpha_2)$$ (9)

and

$$Var(\hat{R}) = \alpha_1 b_1^2(\alpha_1, \alpha_2) \left[\frac{(\alpha_1 \theta_1 + 1)^2}{\theta_1(\alpha_1 \theta_1 + 2)} + \frac{\alpha_1 (\alpha_2 \theta_2 + 1)^2}{\alpha_2 \theta_2(\alpha_2 \theta_2 + 2)}\right].$$

Table 4. Asymptotic Confidence Intervals of $R$ with $\alpha_1 = 0.8$, $\alpha_2 = 1.5$

<table>
<thead>
<tr>
<th>(n,m)</th>
<th>Average Confidence Length</th>
<th>Coverage Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>(15,15)</td>
<td>0.3102</td>
<td>0.9282</td>
</tr>
<tr>
<td>(15,20)</td>
<td>0.3009</td>
<td>0.9367</td>
</tr>
<tr>
<td>(15,30)</td>
<td>0.2890</td>
<td>0.9299</td>
</tr>
<tr>
<td>(20,15)</td>
<td>0.3004</td>
<td>0.9224</td>
</tr>
<tr>
<td>(20,20)</td>
<td>0.2816</td>
<td>0.9313</td>
</tr>
<tr>
<td>(20,30)</td>
<td>0.2718</td>
<td>0.9334</td>
</tr>
<tr>
<td>(30,15)</td>
<td>0.2866</td>
<td>0.9301</td>
</tr>
<tr>
<td>(30,20)</td>
<td>0.2642</td>
<td>0.9331</td>
</tr>
<tr>
<td>(30,30)</td>
<td>0.2370</td>
<td>0.9382</td>
</tr>
</tbody>
</table>

and as $m \to \infty$, $n \to \infty$, we have

$$\frac{\hat{R} - R}{\alpha_1 b_1(\alpha_1, \alpha_2) \sqrt{(\alpha_1 \theta_1 + 1)^2 \theta_1(\alpha_1 \theta_1 + 2) + \frac{\alpha_1 (\alpha_2 \theta_2 + 1)^2}{\alpha_2 \theta_2(\alpha_2 \theta_2 + 2)}}} \to N(0, 1),$$

Then the asymptotic 95% confidence interval for $R_{Har}$ is given by

$$\hat{R} \pm 1.96\alpha_1 b_1(\alpha_1, \alpha_2) \sqrt{(\alpha_1 \theta_1 + 1)^2 \theta_1(\alpha_1 \theta_1 + 2) + \frac{\alpha_1 (\alpha_2 \theta_2 + 1)^2}{\alpha_2 \theta_2(\alpha_2 \theta_2 + 2)}}$$ (10)

and the average length of the asymptotic 95% confidence interval of $R_{Har}$ is

$$\frac{1}{N} \sum_{i=1}^{n} 2(1.96)\alpha_1 b_1(\alpha_1, \alpha_2) \sqrt{(\alpha_1 \theta_1 + 1)^2 \theta_1(\alpha_1 \theta_1 + 2) + \frac{\alpha_1 (\alpha_2 \theta_2 + 1)^2}{\alpha_2 \theta_2(\alpha_2 \theta_2 + 2)}}. $$

6.1 Performance Study

Now we study the performance of the asymptotic confidence interval with respect to the average confidence length and coverage probability. The results are given in Tables 4 and 5 respectively.

From tables 4 and 5, it is clear that the average confidence length is reasonably small and the coverage probability is somewhat close to the nominal value in the case of asymptotic confidence interval. So asymptotic confidence interval can be used for finding the confidence interval of $R$.

7 Conclusion

We have estimated $R=P(X > Y)$ with respect to two independent Pareto type I distributions with shape
parameters $\alpha_1$ and $\alpha_2$. This study can be used in two dimensions. Firstly for comparing two distributions with common base distribution (in medical studies, for studying the effectiveness of drugs) and secondly as a measure of reliability or for conducting stress-strength analysis. As a failure time distribution, this measure can be used as a measure of reliability for Pareto type I distribution. In practical situations sometimes we have to deal mostly with small samples. For small samples, the harmonic moment estimator, which maintains satisfactorily high performance over a specified range of departures from the ideal model and also not too less efficient than the M.L.E and the Hill estimator. The simulation study also suggested the effectiveness of the harmonic moment estimate of $R$ for small samples. Moreover it is a robust estimator. But it is not much good for for large samples. Also we find the sampling distribution of both the tail index estimator and the estimator of $R$. This will help us to study the properties of the estimators. We estimated the asymptotic confidence interval which performs well with respect to the average length of the confidence interval and coverage probability. Further this study can be extended to the general class of distributions with Pareto tail having index $\alpha$, whose MLE is not a robust estimator.

REFERENCES


