Heavy Tailed Pareto Distribution: Properties and Applications

K. Jayakumar[1], A. P. Kuttykrishnan[2], and Bindu Krishnan[*3]

[1]Department of Statistics, University of Calicut, Kerala, India.
[2]Sarva Shiksha Abhiyan, Thiruvananthapuram, Kerala, India.
[3]Department of Statistics, Jain University, Kochi, Kerala, India.

Abstract

In this article, we introduce a class of distributions that have heavy tails as compared to Pareto distribution of third kind, which we termed as Heavy Tailed Pareto (HP) distribution. Various structural properties of the new distribution are derived. It is shown that HP distribution is in the domain of attraction of minimum of Weibull distribution. A representation of HP distribution in terms of Weibull random variable is obtained. Two characterizations of HP distribution are obtained. The method of maximum likelihood is used for estimation of model parameters and simulation results are presented to assess the performance of new model. Marshall-Olkin Heavy Tailed Pareto (MOHP) distribution is also introduced and some of its properties are studied. It is shown that MOHP distribution is geometric extreme stable. An autoregressive time series model with the new model as marginal distribution is developed and its properties are studied.

Keywords *autoregressive model; geometric extreme stability; Marshall–Olkin distribution*

1 Introduction

Data with heavy tails have been studied by various researchers in different areas such as economics, finance, reliability, telecommunications, high speed network traffic, hydrology, insurance, linguistics, physics, biology, etc. The emergence of the Internet and World Wide Web gave a new incentive to the study of heavy-tailed distributions, in literature, due to their omnipresence in internet packet and flow data, the topological sector of the Web, size of computer files, etc. The tail dependent nature in the modelling and analysis of heavy-tailed phenomena sheds a new light in the literature. The shape of tail of heavy-tailed distributions resembles the hyperbolic shape of the Pareto distribution as a first approximation, which is characterized by the so-called tail index. The tail parameter plays a major role in modelling as it governs the thickness of tail of the distribution.

The Pareto distribution, also known as distribution of income, is a very popular heavy tailed distribution that has wider applications in different contexts. The distribution arises as tractable model in actuarial science, economics, finance, life testing, reliability, survival analysis and engineering. The family of Pareto distributions is very prominent in the literature for its competence in modelling heavy-tailed distributions. Arnold (2015) discussed various properties and applications of Pareto distribution and its extensions.

The survival function of classical Pareto distribution, denoted by $P_I(\alpha, \beta)$ and the Pareto distribution of third kind, denoted by $P_{III}(\lambda)$, are respectively given by

$$\bar{G}(x; \alpha, \beta) = \left(\frac{\beta}{x}\right)^{\alpha}; \quad x \geq \beta, \alpha, \beta > 0, \tag{1}$$

---

*Corresponding author. Email: bindusruthy@gmail.com.

and

$$\bar{G}(x;\lambda) = \frac{1}{1+x^\lambda}; \quad x > 0, \lambda > 0. \tag{2}$$

The Pareto distribution was first proposed in literature as a model to explain the allocation of income among individuals. Later, various forms of the Pareto distribution have been formulated for modelling and analysis of data from engineering, environment, geology, hydrology, actuarial science, telecommunications, reliability, risk modelling, etc. These diverse applications of Pareto distribution lead researchers to develop different generalizations of Pareto distribution.

Even though various forms of Pareto distribution such as generalized Pareto distribution by Pickands III (1975), Marshall-Olkin Pareto distribution by Thomas and Jose (2003), Pareto positive-stable by Sarabia and Prieto (2009), Kumaraswamy Pareto distribution by Bourguignon et al. (2013), transmuted Pareto (TP) distribution by Merovci and Puka (2014), Pareto ArcTan distribution by Gómez-Déniz and Calderín-Ojeda (2015), New Generalized Pareto distribution by Jayakumar et al. (2020) and so on, appeared in the literature for modelling heavy tailed data, there may arise lack of distributions for exact data fitting in many cases.

Despite the importance of Pareto distribution in data modelling, a comprehensive study on distributions heavier than Pareto distribution in the context of heavy-tailed distributions has not been considered so far. Motivated by this fact, we introduce here, a new distribution whose tail is heavier than that of Pareto distribution. The introduced distribution is a new heavy-tailed distribution that compete with Pareto distribution of third kind. Our study reveals that the new distribution is a good model to model data sets where extreme observations are significant and to model lifetime data sets.

With regard to the application for modelling time series data with Pareto marginals, Yeh et al. (1988) defined Pareto processes and discussed the applications of their model in income analysis. Pillai (1991) and Pillai et al. (1995) generalized this process in terms of semi-Pareto random variables. Balakrishna and Jayakumar (1997) introduced a bivariate minification process of the first order with semi-Pareto distribution. As the moment generating function of Pareto distribution is not in closed form, we developed here, minification time series models with our new distributions as marginal.

The paper is organized as follows. In Section 2, we introduce the Heavy Tailed Pareto distribution and discuss the shapes of density function and distribution function. Reliability properties and structural properties of the new distribution are also investigated. In Section 3, the maximum likelihood estimation of the model parameters are discussed and simulation is carried out to assess the performance of the proposed maximum likelihood estimation method. In Section 4, Marshall-Olkin Heavy Tailed Pareto (MOHP) distribution is introduced and some properties are studied. Some representations and characterizations are obtained in Section 5. A first order autoregressive minification process with MOHP distribution as marginal is developed and studied in Section 6. We conclude the article in Section 7 with some remarks on the main results and their significance.

2 Heavy Tailed Pareto Distribution

In this section, we introduce a new distribution with survival function

$$\bar{F}(x;\lambda) = \frac{1}{1+\log(1+x^\lambda)}; \quad x \geq 0, \lambda > 0, \tag{3}$$
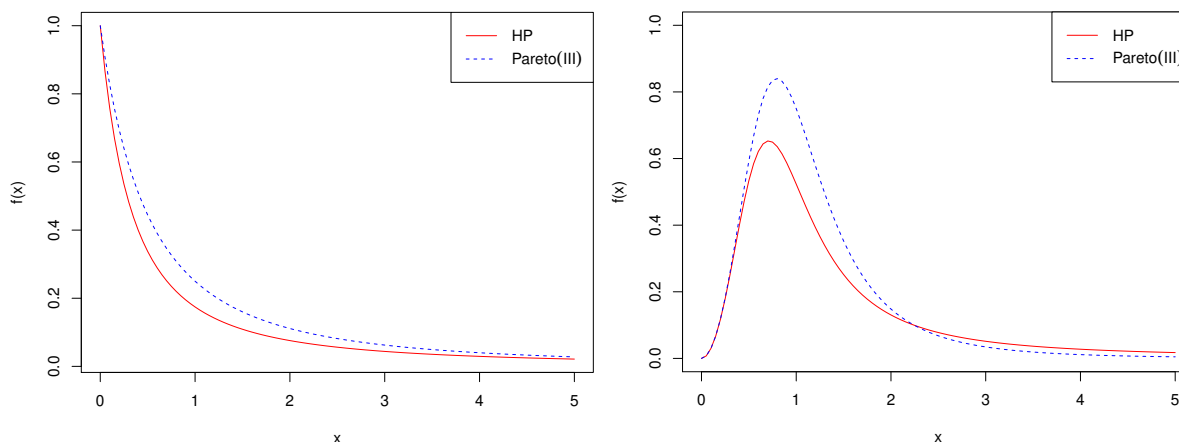
where log is the natural logarithm.

Figure 1: Pdf of HP and Pareto(III) for $\lambda = 1$ (left) and $\lambda = 3$ (right).

It is obvious that $\lim_{x \to \infty} \frac{e^{-\theta x}}{\bar{F}(x;\lambda)} = 0$, for any $\theta > 0$ and so the function (3) represents the survival function of a heavy tailed distribution. Also, it can be easily verified that (3) has heavier tail than the Pareto distribution of third kind. Hence we refer to the distribution with survival function (3) as the Heavy Tailed Pareto distribution with parameter λ, denoted by $HP(\lambda)$.

The cumulative distribution function (cdf) of $HP(\lambda)$ is given by

$$F(x;\lambda) = 1 - \frac{1}{1 + \log(1 + x^\lambda)}; \quad x \geq 0, \lambda > 0. \tag{4}$$

The probability density function (pdf) of $HP(\lambda)$ is

$$f(x;\lambda) = \frac{\lambda x^{\lambda-1}}{(1 + x^\lambda)(1 + \log(1 + x^\lambda))^2}; \quad x > 0, \lambda > 0. \tag{5}$$

From equations (2) and (3), we can see that $\bar{F}(x;\lambda) \geq \bar{G}(x;\lambda)$ and hence if $X \sim HP(\lambda)$ and $Y \sim P_{III}(\lambda)$, then X is stochastically larger than Y.

Jayakumar and Ajitha (2003) showed that $\phi(\lambda) = \frac{1}{1+\log(1+\lambda^\alpha)}$ is the Laplace transform of the probability distribution having the cumulative distribution function

$$H(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k} \int_0^\infty \frac{\Gamma(k+t)x^{\alpha(k+t)}e^{-t}}{\Gamma(t)\Gamma(1 + \alpha(k+t))}dt.$$

Hence,

$$F(x) = 1 - \phi(x) = 1 - \frac{1}{1 + \log(1 + x^\alpha)}, \quad x > 0, \ \alpha > 0,$$

is a distribution with complete monotone derivative (c.m.d.), since $\phi(x)$ is completely monotone with $\phi(0) = 1$ and $\phi(\infty) = 0$. Therefore, by Pillai and Sandhya (1990), $F(x)$ is a mixture of exponentials. Also $F(x)$ is geometrically infinitely divisible (g.i.d.) and hence infinitely divisible (i.d.). Thus the HP distribution we have introduced is a distribution with c.m.d. and hence is a mixture of exponentials and is g.i.d. For the applications of g.i.d. distribution and distribution with c.m.d., see Klebanov et al. (1985)and Aly and Bouzar (2000).

Figure 1 gives a comparison between pdf of HP distribution and pdf of Pareto distribution of third kind for different values of λ. From the graphs of Figure 1, it is very clear that the distribution with pdf (5) is heavy tailed than Pareto distribution.
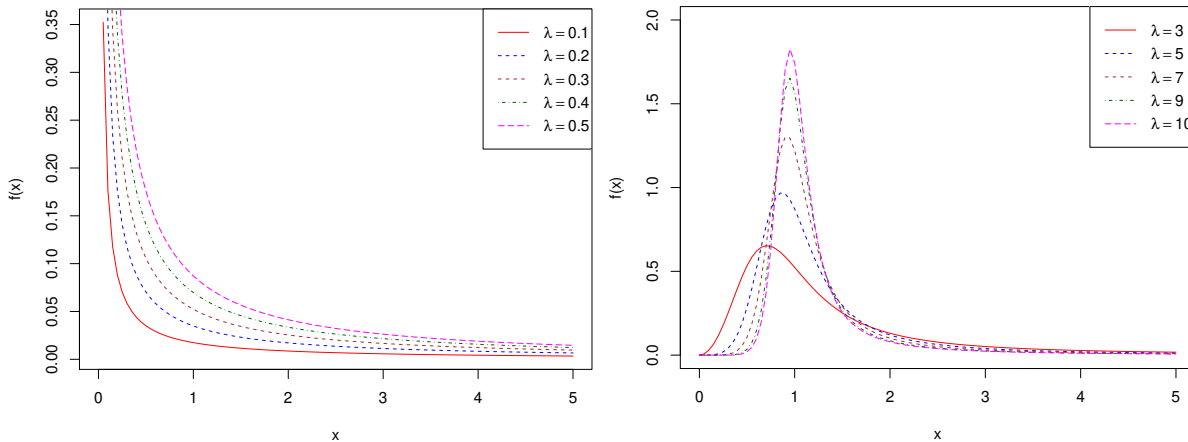
Figure 2: Pdf of HP(λ) for various values of λ.

2.1 Shapes of pdf and cdf

Graphs of pdf and cdf for various values of parameter λ are presented in Figure 2 and Figure 3 respectively.

2.2 Reliability Properties

2.2.1 Hazard Rate Function

The hazard rate function of HP(λ) is

$$h(x;\lambda) = \frac{\lambda x^{\lambda-1}}{(1+x^\lambda)(1+\log(1+x^\lambda))}. \tag{6}$$

Graph of hazard rate function for various values of λ is presented in Figure 4.

From Figure 4, it is very clear that the hazard rate function of HP is decreasing when $\lambda \leq 1$ and hence HP distribution belongs to the class of DFR distributions, if $\lambda \leq 1$.

When $\lambda > 1$, the hazard rate function behaves differently in two regions:

1. For $x \leq x_0$, the hazard rate function is increasing.

2. For $x > x_0$, the hazard rate function is decreasing, where $x_0 = \left[\frac{(\lambda-1)y}{\lambda-(\lambda-1)y}\right]^{\frac{1}{\lambda}}$ and $y = \frac{1+\log(1+x^\lambda)}{2+\log(1+x^\lambda)}$.

Now,we have, for $X \sim P_{III}(\lambda)$,

$$h_{P_{III}}(x;\lambda) = \frac{\lambda x^{\lambda-1}}{(1+x^\lambda)}. \tag{7}$$

From the hazard rate functions given by (6) and (7), we obtain $h_{HP}(x;\lambda) \leq h_{P_{III}}(x;\lambda)$ and it is shown in Figure 5. Hence , at the same age, the unit whose life distribution is HP distribution is less likely to instantaneously perish than the one whose life distribution is Pareto of third kind.
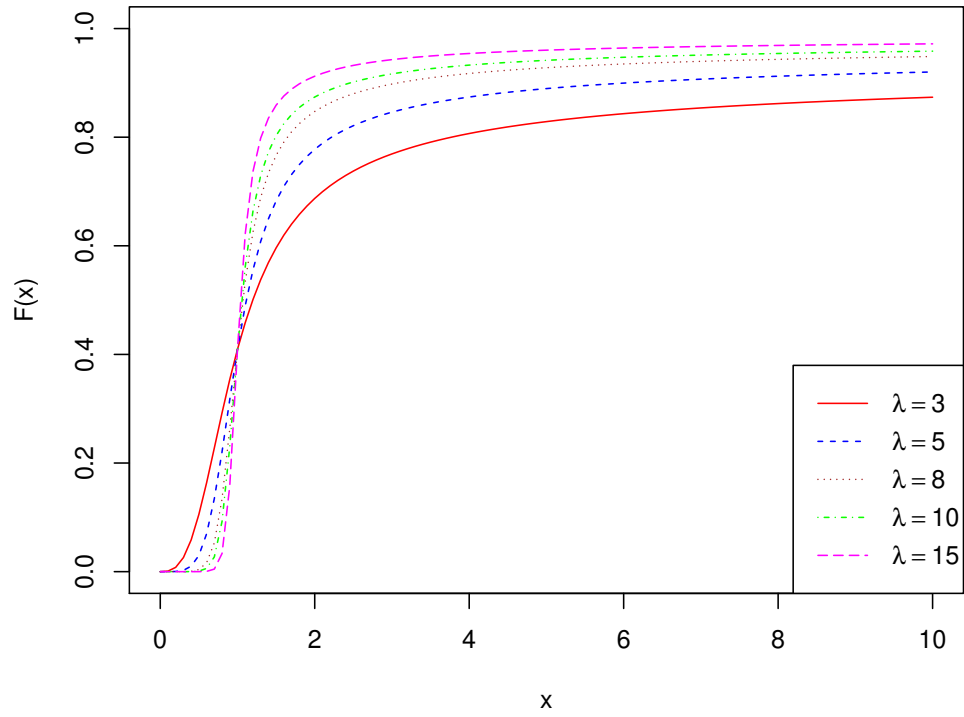
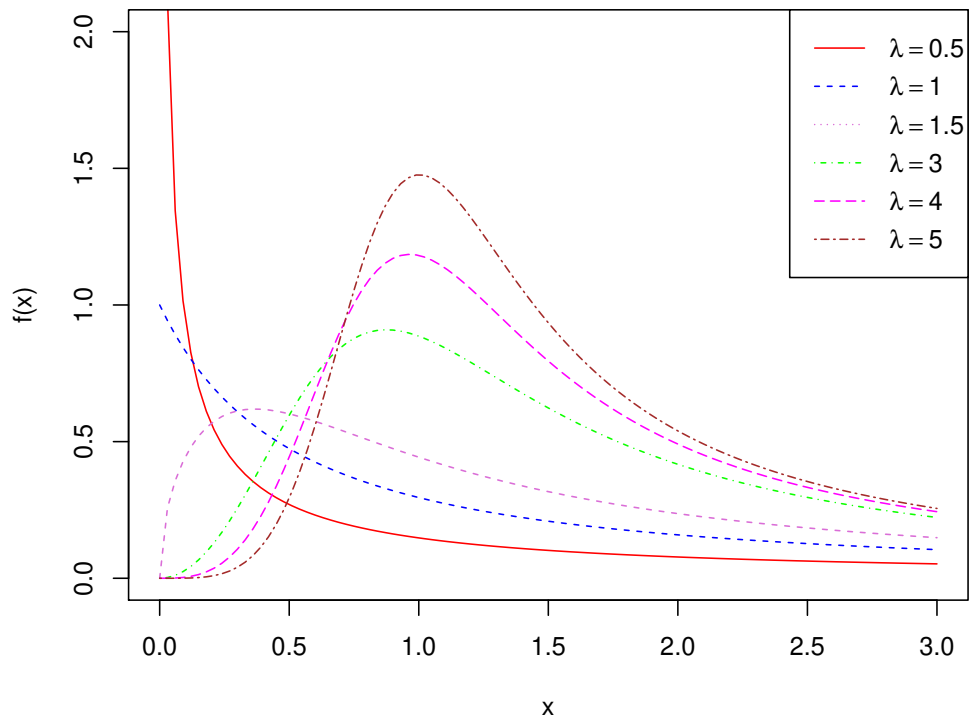Figure 3: Cdf of HP(λ)for various values of λ.

Figure 4: Hazard rate function of HP(λ) for various values of λ.
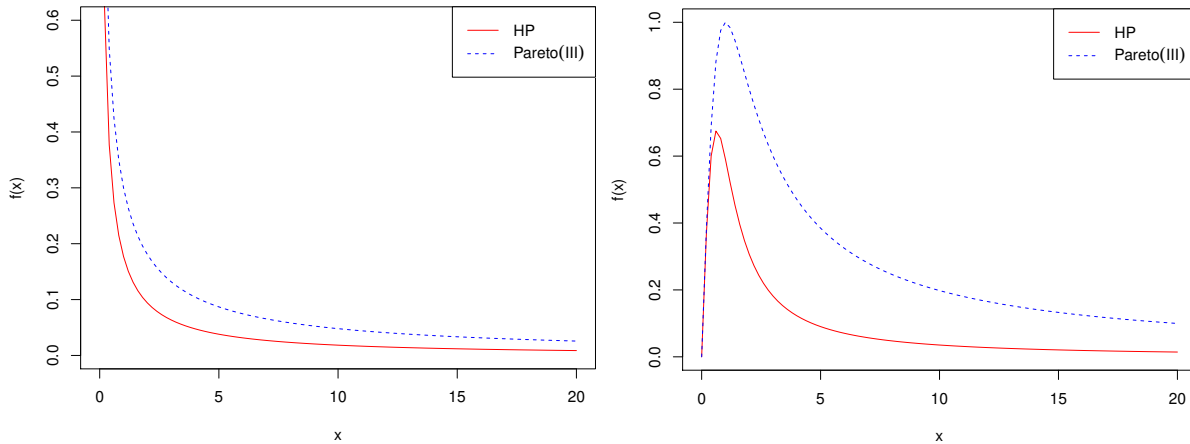
Figure 5: Hazard rate function of HP and Pareto(III) for $\lambda = 0.6$ (left) and $\lambda = 2$ (right).

2.2.2 Reverse Hazard Rate Function

Here we consider the reverse hazard rate which is useful in constructing the information matrix and in estimating the survival function for censored data.

The reverse hazard rate function of HP(λ) is given by

$$r(x; \lambda) = \frac{\lambda x^{\lambda-1}}{[1 + x^\lambda][\log(1 + x^\lambda)][1 + \log(1 + x^\lambda)]}.$$

2.2.3 Mean Residual Life Function

An important ageing property in reliability theory is the mean residual life (MRL) function which is defined simply as the expected value of the remaining lifetime beyond an age x.

The MRL function of $X \sim HP(\lambda)$ is given by

$$m(x) = \frac{\int_x^\infty \bar{F}(u)du}{\bar{F}(x)} = (1 + \log(1 + x^\lambda)) \int_x^\infty (1 + \log(1 + u^\lambda))du.$$

An estimate of MRL function can be obtained using the ordered data. Let $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ be the ordered HP data, then the MRL function is easily estimated at $x = X_{(n-k)}$, for some $k = 1, 2, ..., n - 1$ by the empirical average excess of the k data points higher than $X_{(n-k)}$ and is given by

$$\hat{m(x)} = \frac{1}{k} \sum_{j=1}^k X_{(n-j+1)} - X_{(n-k)}.$$

2.2.4 Log-Odds Rate

Wang et al. (2003) discussed a model to time to failure based on the log-odds rate and some characterizations of distributions of failure time by log-odds rate are obtained. By modelling the failure process in terms of the log-odds rate, the practioner can use the model to assess the distribution of time to failure.

Table 1: Mean, Variance, Skewness and Kurtosis of HP (λ) for different values of λ.

λ	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
Mean	0.2455	0.2662	0.2997	0.3268	0.3481	0.3649	0.3785	0.3896	0.3989	0.4068
Variance	0.3007	0.1746	0.1600	0.1594	0.1625	0.1666	0.1709	0.1750	0.1789	0.1826
Skewness	2.8293	1.6909	1.1668	0.8584	0.6580	0.5199	0.4209	0.3474	0.2916	0.2481
Kurtosis	10.7752	4.9276	3.1649	2.3799	1.9616	1.1722	1.5514	1.4414	1.3625	1.3038

They defined log-odds function as

$$LO(x) = \log \frac{F(x)}{1 - F(x)} = \log(\log(1 + x^\lambda)).$$

Hence a distribution is HP distribution if and only if $LO(x) = \log(\log(1 + x^\lambda))$. The log-odds rate (LOR) is defined as

$$LOR(t) = \frac{h(t)}{\bar{F}(t)} = \frac{\lambda t^{\lambda-1}}{(1 + t^\lambda) \log(1 + t^\lambda)},$$

where $h(.)$ is the hazard rate function.

2.3   Structural Properties

The statistical properties of the HP distribution including moments, random number generation, median, quantile function, entropies and distribution of order statistics are presented in this section.

2.3.1   Moments

The rth moment of $X \sim HP(\lambda)$ is given by

$$\mu_r^{'} = E[X^r] = \int_0^\infty x^r \frac{\lambda x^{\lambda-1}}{(1 + x^\lambda)(1 + \log(1 + x^\lambda))^2} dx$$

$$= \int_0^\infty \frac{t^{\frac{r}{\lambda}}}{[1 + t][1 + \log(1 + t)]} dt,$$

by substituting $x^\lambda = t$.

Again, by letting $\log(1 + t) = u$, the above integral becomes

$$\mu_r^{'} = \int_0^\infty \frac{(e^u - 1)^{\frac{r}{\lambda}}}{(1 + u)^2} du \qquad (8)$$

As the above integral is not easy to evaluate, we present in Table 1, the behavior of mean, variance, skewness and kurtosis for selected values of parameter λ, using R software, when the upper limit of the integral is finite. From Table 1, we can see that when λ increases, then the mean also increases but the skewness and kurtosis decreases. The value of variance appears to decrease first and then increases when $\lambda > 2$.

2.3.2   Random Number Generation

A random sample X from HP(λ) distribution can be simulated as

$$X = [e^{\frac{Y}{1-Y}} - 1]^{\frac{1}{\lambda}}; \quad Y \sim U(0,1).$$

2.3.3   Quantile Function

The qth quantile of the HP distribution is given by

$$X_q = [e^{\frac{q}{1-q}} - 1]^{\frac{1}{\lambda}}.$$

In particular, the median is

$$\text{M} = (e-1)^{\frac{1}{\lambda}}.$$

2.3.4   Rényi Entropy and Shannon Entropy

Entropy is the measure of variation or the uncertainty of a random variable X for the probability density function from the lifetime distribution.

The Rényi entropy of a random variable X with pdf $f(.)$ is defined as

$$I_R(\eta) = \frac{1}{1-\eta} \log \int_0^\infty f^\eta(x) dx; \eta > 0, \eta \neq 1.$$

The Rényi Entropy of HP(λ) is

$$
\begin{aligned}
I_R(\eta) &= \frac{1}{1-\eta} \log \int_0^\infty \frac{\lambda^\eta x^{\eta(\lambda-1)}}{(1+x^\lambda)^\eta (1+\log(1+x^\lambda))^{2\eta}} dx \\
&= \frac{1}{1-\eta} \log \int_0^\infty \frac{\lambda^{\eta-1} t^{(1-\frac{1}{\lambda})(\eta-1)}}{(1+t)^\eta (1+\log(1+t))^{2\eta}} dt,
\end{aligned}
$$

by letting $x^\lambda = t$.

The Shannon entropy of $X \sim HP(\lambda)$ is given by

$$
\begin{aligned}
E[-\log f(X)] &= E\left[-\log\left(\frac{\lambda x^{\lambda-1}}{(1+x^\lambda)(1+\log(1+x^\lambda))^2}\right)\right] \\
&= -\log\lambda - (\lambda-1)E[\log X] + E[\log(1+X^\lambda)] + 2E[\log(1+\log(1+X^\lambda))].
\end{aligned}
$$

2.3.5   Order Statistics

Let $X_1, X_2, ..., X_n$ be a random sample of size n from HP(λ) and let $X_{(1)}, X_{(2)}, ..., X_{(n)}$ denote the corresponding order statistics. Then, the pdf of ith order statistic is

$$
\begin{aligned}
f_{X_{(i)}}(x;\lambda) &= \frac{n!}{(i-1)!(n-i)!} f(x;\lambda) F^{i-1}(x;\lambda) \bar{F}^{n-i}(x;\lambda) \\
&= \frac{n!}{(i-1)!(n-i)!} \frac{\lambda x^{\lambda-1}}{(1+x^\lambda)(1+\log(1+x^\lambda))^2} \\
&\quad \left[1 - \frac{1}{1+\log(1+x^\lambda)}\right]^{i-1} \left[\frac{1}{1+\log(1+x^\lambda)}\right]^{n-i} \\
&= \frac{n!}{(i-1)!(n-i)!} \frac{\lambda x^{\lambda-1}[\log(1+x^\lambda)]^{i-1}}{[1+\log(1+x^\lambda)]^{n+1}}.
\end{aligned}
$$

Hence the pdf of the largest order statistic, $X_{(n)}$, is

$$f_{X_{(n)}}(x; \lambda) = \frac{n\lambda x^{\lambda-1}[\log(1+x^\lambda)]^{n-1}}{[1+\log(1+x^\lambda)]^{n+1}},$$

and the pdf of the smallest order statistic, $X_{(1)}$, is

$$f_{X_{(1)}}(x; \lambda) = \frac{n\lambda x^{\lambda-1}}{[1+\log(1+x^\lambda)]^{n+1}}.$$

3   Estimation of Parameters

In this section, we consider the method of maximum likelihood for the estimation of parameter of HP distribution.

Let $x_1, x_2, ..., x_n$ be a random sample of size n from this distribution with unknown parameter λ. Then, the likelihood function is

$$\begin{aligned} \ell(x; \lambda) &= \prod_{i=1}^{n} f(x; \lambda) \\ &= \frac{\lambda^n \left[\prod_{i=1}^{n} x_i\right]^{\lambda-1}}{\prod_{i=1}^{n}(1+x_i^\lambda)\left[\prod_{i=1}^{n}(1+\log(1+x_i^\lambda))\right]^2}, \end{aligned} \tag{9}$$

so that the log-likelihood function becomes

$$\log \ell = n\log\lambda + (\lambda - 1)\sum_{i=1}^{n}\log x_i - \sum_{i=1}^{n}\log(1+x_i^\lambda) - 2\sum_{i=1}^{n}\log(1+\log(1+x_i^\lambda)).$$

The first order derivative of log-likelihood function with respect to the parameter λ is given by

$$\frac{d(\log\ell)}{d\lambda} = \frac{n}{\lambda} + \sum_{i=1}^{n}\log x_i - \frac{x_i^\lambda \log x_i}{1+x_i^\lambda} - 2\sum_{i=1}^{n}\frac{x_i^\lambda \log x_i}{(1+x_i^\lambda)(1+\log(1+x_i^\lambda))}. \tag{10}$$

The maximum likelihood estimator of the parameter λ is obtained by solving the equation, $\frac{d\log\ell}{d\lambda} = 0$. The above equation (10) cannot be solved analytically and hence statistical software can be used to solve them numerically by means of iterative techniques such as the Newton-Raphson algorithm.

Let $\hat{\lambda}$ be the estimate of parameter. The asymptotic confidence interval for the parameter λ is obtained by considering the Fisher information matrix.

The Fisher information matrix corresponding to HP(λ) distribution is given by

$$I(\hat{\lambda}) = -E\left(\frac{d^2(\log\ell)}{d\lambda^2}\right)$$

where

$$\frac{d^2(\log\ell)}{d\lambda^2} = -\frac{n}{\lambda^2} - \sum_{i=1}^{n} x_i^\lambda \left(\frac{\log x_i}{1+x_i^\lambda}\right)^2 - 2\sum_{i=1}^{n}\left(\frac{x_i^\lambda(1-x_i^\lambda+\log(1+x_i^\lambda))(\log x_i)^2}{[(1+x_i^\lambda)(1+\log(1+x_i^\lambda))]^2}\right). \tag{11}$$

Table 2: MLEs and MSEs (in parentheses) for various parameter values.

$\lambda \rightarrow$	0.2	0.5	0.8	1.0	1.2	1.5	1.8	2.0
$n \downarrow$								
50	0.2113	0.5294	0.8186	1.0205	1.2457	1.5457	1.8541	2.0976
	(0.0004)	(0.0039)	(0.0158)	(0.0189)	(0.0338)	(0.0506)	(0.0752)	(0.1417)
100	0.2102	0.5289	0.8087	0.9948	1.2269	1.5254	1.8426	2.0261
	(0.0003)	(0.0027)	(0.0060)	(0.0098)	(0.0189)	(0.0237)	(0.0340)	(0.0365)
200	0.2051	0.5175	0.7966	1.0114	1.2125	1.5351	1.7936	2.0275
	(0.0001)	(0.0010)	(0.0028)	(0.0051)	(0.0088)	(0.0140)	(0.0170)	(0.0193)
500	0.2035	0.5105	0.7987	1.0089	1.2092	1.5089	1.8120	2.0205
	(0.0000)	(0.0003)	(0.0013)	(0.0024)	(0.0032)	(0.0042)	(0.0075)	(0.0084)
1000	0.2029	0.5075	0.8027	1.0009	1.2039	1.4985	1.8176	2.0142
	(0.0000)	(0.0002)	(0.0007)	(0.0007)	(0.0013)	(0.0017)	(0.0043)	(0.0053)

Using asymptotic normality results, the asymptotic confidence interval can be obtained as $\sqrt{n}(\hat{\lambda} - \lambda) \longrightarrow N(0, I^{-1}(\lambda))$. We can easily show that the HP family satisfies all the regularity conditions for $\lambda > 0$ and hence the above equation (11) holds, where $I(\lambda)$ can be estimated by $I^{-1}(\hat{\lambda})$. Also, the 95% confidence interval for the parameter λ is given by

$$\hat{\lambda} \pm 1.96 \left(\frac{I^{-1}(\hat{\lambda})}{\sqrt{n}} \right).$$

3.1 Simulation

We conduct Monte Carlo simulation study to assess the performance of maximum likelihood estimation procedure for estimating the HP parameter. Samples of sizes n = 50, 100, 200, 500 and 1000 are generated from the HP model, for different values of parameter λ. We repeated the simulation 1000 times and calculated the MLEs and mean squared errors (MSEs) of the parameter estimate. The empirical results are listed in Table 2. We note that the MSE of the parameter λ decrease as the sample size increases. The mean estimates of the parameter tend to be closer to the true parameter values.

4 Marshall-Olkin Heavy Tailed Pareto Distribution

Marshall and Olkin (1997) introduced a general method to incorporate a shape parameter in the base random variable X and have shown the flexibility of transformed model by the introduction of the shape parameter.

If $\bar{G}(x)$ is the survival function of a continuous random variable X, then

$$\bar{F}(x) = \frac{\alpha \bar{G}(x)}{1 - (1 - \alpha)\bar{G}(x)}$$

is a proper survival function. The introduction of new parameter α will make the distribution with survival function $\bar{F}(x)$ more flexible than the base distribution having survival function $\bar{G}(x)$, especially in data modelling.
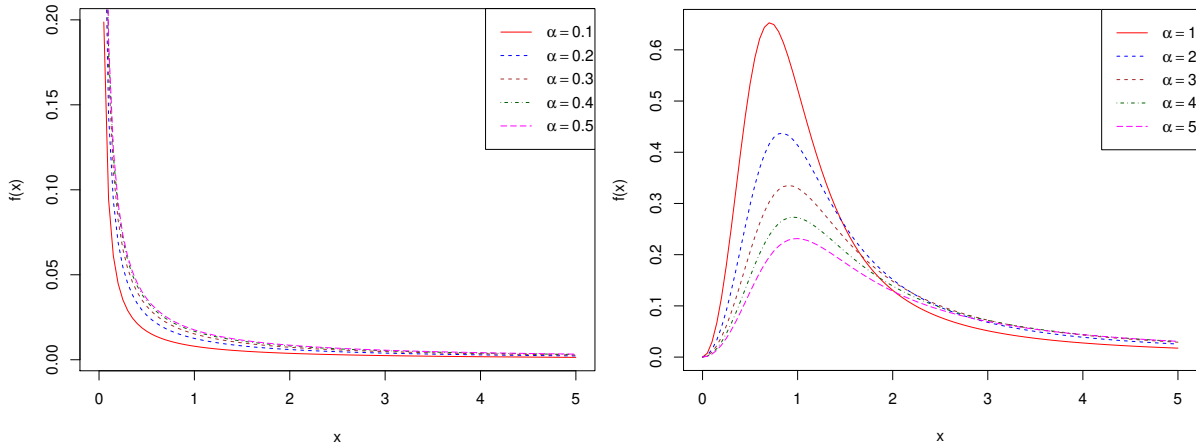
Figure 6: Pdf of MOHP(α, λ) for $\lambda = 0.1$ (left) and for $\lambda = 3$ (right).

Sankaran and Jayakumar (2008) discussed the physical interpretation of Marshall-Olkin family of distributions using proportional odds model. The Marshall and Olkin (1997) extended class of distributions offer a wide range of behaviour than the basic distributions from which they are derived.

If X is a random variable defined on $[0, \infty)$ with survival function (3), then, using the survival function defined by Marshall and Olkin (1997), we obtain a new family of distributions with survival function

$$\bar{F}(x; \alpha, \lambda) = \frac{1}{1 + \frac{1}{\alpha} \log(1 + x^\lambda)}; x \geq 0, \alpha > 0, \lambda > 0. \tag{12}$$

We refer this distribution as Marshall-Olkin Heavy Tailed Pareto distribution, denoted by MOHP(α, λ). The corresponding probability density function is defined as

$$f(x; \alpha, \lambda) = \frac{\lambda x^{\lambda-1}}{\alpha(1 + x^\lambda)(1 + \frac{1}{\alpha} \log(1 + x^\lambda))^2}; x > 0, \alpha > 0, \lambda > 0. \tag{13}$$

Note that the MOHP distribution is a class of distributions that contains HP distribution and is a viable competitor of the heavy-tailed Pareto distribution. Plots of pdf of MOHP distribution for different values of parameters α and λ are presented in Figure 6.

4.1 Properties

The hazard rate function of MOHP(α, λ) distribution is given by

$$h(x; \alpha, \lambda) = \frac{\lambda x^{\lambda-1}}{\alpha(1 + x^\lambda)(1 + \frac{1}{\alpha} \log(1 + x^\lambda))} \tag{14}$$

From the hazard rate of HP distribution and MOHP distribution, it follows that

$$h_{MOHP}(x; \alpha, \lambda) = \frac{1}{1 - (1 - \alpha)\bar{F}_{HP}(x; \lambda)} h_{HP}(x; \lambda)$$

and hence

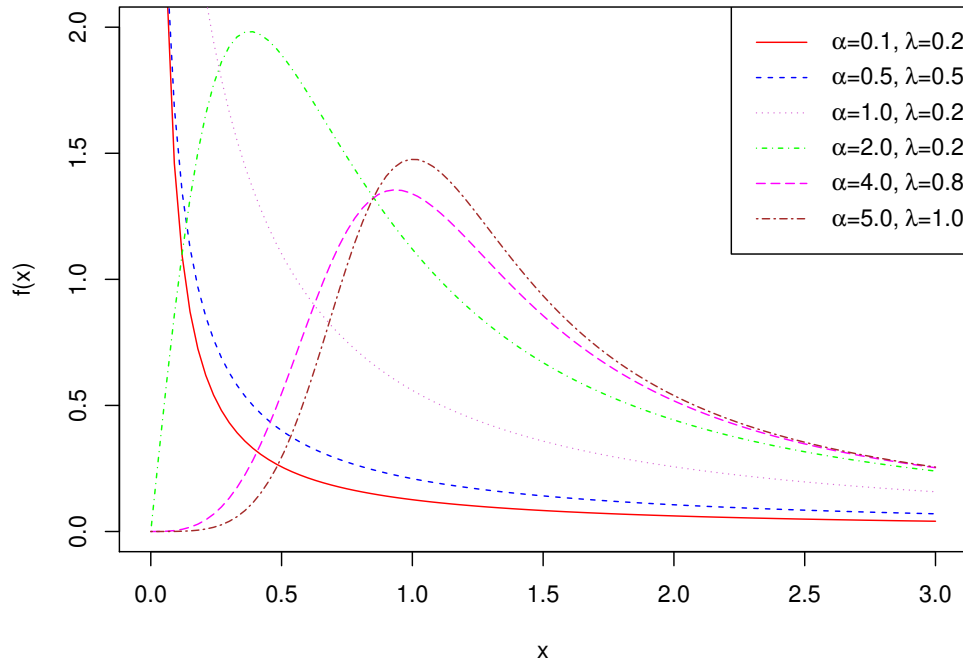$$\lim_{x \to -\infty} h_{MOHP}(x) = \lim_{x \to -\infty} \frac{h_{HP}(x)}{\alpha}$$

Figure 7: Hazard rate function of MOHP(α, λ) for various values of parameters.

and

$$\lim_{x \to \infty} h_{MOHP}(x) = \lim_{x \to \infty} h_{HP}(x).$$

Also, it can be easily verified that $\frac{h_{HP}(x)}{\alpha} \leq h_{MOHP}(x) \leq h_{HP}(x)$ for $0 < x < \infty, \alpha \geq 1$ and $h_{HP}(x) \leq h_{MOHP}(x) \leq \frac{h_{HP}(x)}{\alpha}$ for $0 < x < \infty, 0 < \alpha \leq 1$.

Plot of hazard rate function of MOHP(α, λ) for various values of α and λ is presented in Figure 7.

The quantile function of X, where $X \sim MOHP(\alpha, \lambda)$ is obtained by inverting the cdf of $MOHP(\alpha, \lambda)$ to obtain $x_u = F^{-1}(y)$ as

$$X = [e^{\frac{\alpha Y}{1-Y}} - 1]^{\frac{1}{\lambda}}.$$

Simulating the MOHP random variable is straightforward.

In particular, the median is given by

$$M = [e^{\alpha} - 1]^{\frac{1}{\lambda}}.$$

The maximum likelihood estimates of the parameters α and λ of MOHP distribution are obtained by solving the equations,

$$\frac{n}{\lambda} + \sum_{i=1}^{n} \log x_i - \frac{x_i^{\lambda} \log x_i}{1 + x_i^{\lambda}} - 2\sum_{i=1}^{n} \frac{x_i^{\lambda} \log x_i}{\alpha(1 + x_i^{\lambda})(1 + \frac{1}{\alpha}\log(1 + x_i^{\lambda}))} = 0$$

and

$$\frac{-n}{\alpha} + 2\sum_{i=1}^{n} \frac{\log(1 + x_i^{\lambda})}{\alpha^2(1 + \frac{1}{\alpha}\log(1 + x_i^{\lambda}))} = 0.$$

5 Transformations and Characterizations

The relation between HP distribution and Pareto(III) distribution is established in the following theorem.

Theorem 1. *If X follows Heavy Tailed Pareto distribution with parameter λ , then $Y = (\log(1 + X^\lambda))^{\frac{1}{\beta}}$ follows Pareto distribution of third kind with parameter β.*

Proof. If $X \sim HP(\lambda)$ and $Y = (\log(1 + X^\lambda))^{\frac{1}{\beta}}$, then by using (5) , the pdf of Y is given by

$$
\begin{aligned}
f(y; \beta) &= f(x; \lambda)|\frac{dx}{dy}| \\
&= \frac{\lambda x^{\lambda-1}}{(1 + x^\lambda)(1 + \log(1 + x^\lambda))^2} \frac{\beta(1 + x^\lambda)}{\lambda x^{\lambda-1} \log(1 + x^\lambda))^{\frac{1}{\beta}-1}} \\
&= \frac{\beta y^{\beta-1}}{(1 + y^\beta)^2}.
\end{aligned}
$$

$\square$

If we define $W_n = \min(X_1, X_2, ..., X_n)$ where X_i's are independent and identically distributed (i.i.d) random variables with distribution function $F(x)$, then the distribution function of W_n is given by

$$H_n(x) = 1 - (\bar{F}(x))^n. \tag{15}$$

The distribution $F(x)$ is said to belong to the domain of attraction for minimum of a given cumulative distribution function $H(x)$, if there exist sequences $\{a_n\}$ and $\{b_n\}$ where $a_n, b_n > 0$ are constants depending on n such that the limiting distribution

$$\lim_{n \to \infty} H_n(a_n + b_n x) = H(x), \text{ for every } x. \tag{16}$$

The next theorem proves the result related to HP and Weibull distributions with regard to the domain of attraction for minimum.

Theorem 2. *The Heavy tailed Pareto distribution is in the domain of attraction of minimum of the Weibull distribution.*

Proof. If $a_n = 0, b_n = n^{-\frac{1}{\lambda}} > 0$, then from equations (15) and (16), we have,

$$
\begin{aligned}
\lim_{n \to \infty} H_n(a_n + b_n x) &= \lim_{n \to \infty} [1 - (\bar{F}(n^{-\frac{1}{\lambda}} x))^n] \\
&= \lim_{n \to \infty} \left\{ 1 - \left[\frac{1}{1 + \log(1 + \frac{x^\lambda}{n})} \right]^n \right\} \\
&= 1 - e^{-x^\lambda},
\end{aligned}
$$

which is the cdf of Weibull distribution. $\square$

Now we establish the relationship between Weibull and Heavy Tailed Pareto distribution in the following theorem.

Theorem 3. *If X and Y are independent random variables such that X has pdf*

$$g(x) = e^{-x} \int_0^\infty \frac{1}{\Gamma(u)} x^{u-1} e^{-u} du$$

and Y follows Weibull distribution with survival function $e^{-y^\lambda}; y > 0, \lambda > 0$, then $X^{-\frac{1}{\lambda}} Y$ follows Heavy Tailed Pareto distribution.

Proof. If $U = X^{-\frac{1}{\lambda}} Y$, then its survival function is given by

$$
\begin{aligned}
\bar{F}(x) &= P[U > x] \\
&= P[X^{-\frac{1}{\lambda}} Y > x] \\
&= \int_0^\infty P[X^{-\frac{1}{\lambda}} Y > x / X = z] g(z) dz \\
&= \int_0^\infty e^{-z x^\lambda} g(z) dz.
\end{aligned}
$$

Since the Laplace transform of X with pdf $g(x)$ is $L(t) = \frac{1}{1+\log(1+t)}$, see Pillai (1991), we have,

$$\bar{F}(x) = \frac{1}{1 + \log(1 + x^\lambda)}.$$

$\square$

The following theorem gives a characterization property of HP distribution which is similar to the characterization property satisfied by Pareto (III) distribution.

Theorem 4. *A random variable $X \sim HP(\lambda)$ if and only if it satisfies the functional equation*

$$\frac{\bar{F}(u(xy))}{\bar{F}(u(x))\bar{F}(u(y))} = \frac{F(u(xy))}{F(u(x))F(u(y))}$$

where $u(z) = (e^z - 1)^{\frac{1}{\lambda}}$.

Proof. The proof is straightforward and hence is omitted. $\square$

Now we give some results related to MOHP distribution.

Theorem 5. *Marshall-Olkin Heavy tailed Pareto distribution is geometric extreme stable.*

Proof. Let $X_1, X_2, ...$ be a sequence of independent and identically distributed MOHP(α, λ) random variables with survival function (12) and let N be a geometric random variable, independent of X_i's with probability mass function (pmf)

$$P[N = n] = pq^{n-1}; 0 < p < 1, p + q = 1, n = 1, 2, ... \tag{17}$$

Define

$$U_N = \min(X_1, X_2, ..., X_N),$$
$$V_N = \max(X_1, X_2, ..., X_N).$$

Then the survival function of U_N is given by

$$\bar{F}(x) = P[U_N > x] = \frac{p\bar{F}_{MOHP}(x)}{1 - (1-p)\bar{F}_{MOHP}(x)} = \frac{1}{1 + \frac{1}{\alpha p}\log(1 + x^\lambda)}.$$

That is, $U_N \sim MOHP(\alpha p, \lambda)$. Hence MOHP distribution is geometric minimum stable.

Similarly, we can establish that MOHP distribution is geometric maximum stable. So we conclude that the MOHP distribution is geometric extreme stable.   □

The next theorem provides a characterization result of the MOHP distribution.

Theorem 6. *Let $\{X_i, i \geq 1\}$ be a sequence of i.i.d random variables with common survival function $\bar{H}$ and N be a geometric random variable with pmf* (17), *which is independent of X_i's. Let $U_N = \min(X_1, X_2, ..., X_N)$. Then $\{U_N\}$ is distributed as MOHP (p, λ) if and only if $\{X_i\}$ is distributed as HP (λ).*

Proof. Let $\bar{F}(x)$ be the survival function of U_N. Then

$$\bar{F}(x) = P[U_N > x] = \frac{p\bar{H}(x)}{1 - (1-p)\bar{H}(x)}.$$

Let $\bar{H}(x) = \frac{1}{1+\log(1+x^\lambda)}$. Then,

$$\bar{F}(x) = \frac{1}{1 + \frac{1}{p}\log(1 + x^\lambda)},$$

which is the survival function of MOHP (p, λ).

This proves the sufficiency part of the theorem and converse of the theorem is straightforward.   □

6 Autoregressive Minification Process

In this section, we develop a first order autoregressive (AR(1)) minification process with MOHP distribution as marginal distribution.

Consider an AR(1) minification process with structure

$$X_n = \begin{cases} \varepsilon_n & \text{w.p.} & \delta \\ \min(X_{n-1}, \varepsilon_n) & \text{w.p.} & 1 - \delta \end{cases} \quad 0 < \delta < 1; n \geq 1 \tag{18}$$

where $\{\varepsilon_n\}$ is a sequence of i.i.d random variables.

Theorem 7. *Consider an AR(1) model defined by* (18), *where $\{\varepsilon_n\}$ is a sequence of i.i.d random variables independent of X_n. Then $\{X_n\}$ is stationary Markovian AR(1) process with MOHP marginal if and only if $\{\varepsilon_n\}$ is a sequence of HP random variables and $X_0 \overset{d}{=} MOHP(\delta, \lambda)$.*

Proof. The model (18) can be written in terms of survival function as

$$P(X_n > x) = \delta P(\varepsilon_n > x) + (1-\delta)P(X_{n-1} > x)P(\varepsilon_n > x).$$

That is,

$$\bar{F}_{X_n}(x) = \delta\bar{F}_{\varepsilon_n}(x) + (1-\delta)\bar{F}_{X_{n-1}}(x)\bar{F}_{\varepsilon_n}(x). \tag{19}$$

Assuming stationarity,

$$\bar{F}_X(x) = \frac{\delta\bar{F}_\varepsilon(x)}{1-(1-\delta)\bar{F}_\varepsilon(x)}.$$

If

$$\bar{F}_\varepsilon(x) = \frac{1}{1+\log(1+x^\lambda)},$$

then

$$\bar{F}_X(x) = \frac{1}{1+\frac{1}{\delta}\log(1+x^\lambda)}$$

which is the survival function of MOHP(δ,λ) random variable.

Conversely, if $X_n \overset{d}{=} \text{MOHP}(\delta,\lambda)$, then,

$$\bar{F}_{\varepsilon_n}(x) = \frac{1}{1+\log(1+x^\lambda)}.$$

To prove the stationarity, we use inductive argument.

Let $X_0 \overset{d}{=} \text{MOHP}(\delta,\lambda)$ and $\varepsilon_n \overset{d}{=} \text{HP}(\lambda)$, then using (19), we have,

$$\bar{F}_{X_1}(x) = \frac{1}{1+\frac{1}{\delta}\log(1+x^\lambda)}.$$

That is, $X_1 \overset{d}{=} \text{MOHP}(\delta,\lambda)$.

If $X_{n-1} \overset{d}{=} \text{MOHP}(\delta,\lambda)$ and $\varepsilon_n \overset{d}{=} \text{HP}(\lambda)$, then we get,

$$\bar{F}_{X_n}(x) = \frac{1}{1+\frac{1}{\delta}\log(1+x^\lambda)}.$$

That is, $X_n \overset{d}{=} \text{MOHP}(\delta,\lambda)$.

Hence $\{X_n\}$ is a stationary AR(1) process with MOHP marginal if $X_0 \overset{d}{=} \text{MOHP}(\delta,\lambda)$ and ε_n's are i.i.d HP(λ) random variables. $\square$

6.1 Properties

In order to study the behaviour of the process, we simulate the sample paths of MOHP AR(1) for various values of λ and δ. The sample path of the process for

  i $\lambda = 0.5$, $\delta = 0.3$;
  ii $\lambda = 1.5$, $\delta = 0.3$;
 iii $\lambda = 3.0$, $\delta = 0.3$;
 iv $\lambda = 0.5$, $\delta = 0.8$;
  v $\lambda = 1.5$, $\delta = 0.8$;
 vi $\lambda = 3.0$, $\delta = 0.8$.

are given in Figures 8.

The joint survival function of (X_n, X_{n+1}) is given by

$$\begin{aligned}
\bar{F}_{X_n,X_{n+1}}(x,y) &= P[X_n > x, X_{n+1} > y]\\
&= [\delta\bar{F}_{X_n}(x) + (1-\delta)\bar{F}_{X_n}(\max(x,y))]\bar{F}_{\varepsilon_n}(y)
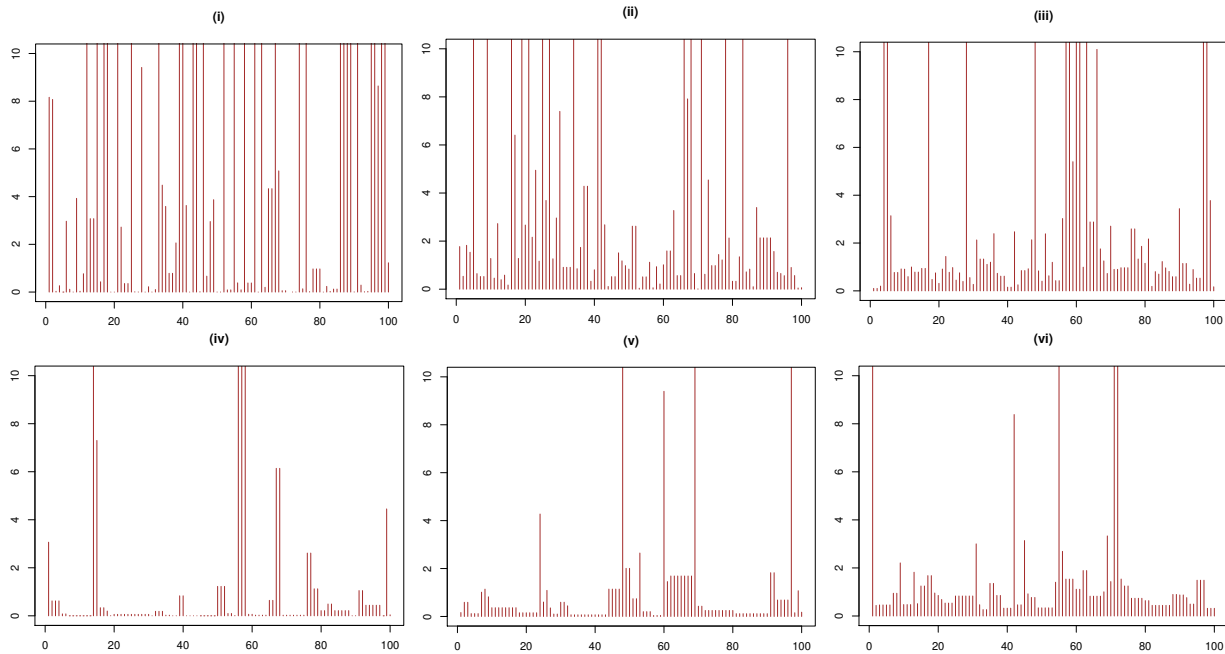\end{aligned}$$

Figure 8: Sample paths of MOHP Process.

On simplification, we get,

$$\bar{F}_{X_n, X_{n+1}}(x_n, x_{n+1}) = \frac{1}{[1 + \frac{1}{\delta} \log(1 + x_n^\lambda)][1 + \log(1 + x_{n+1}^\lambda)]},$$

if $0 < x_{n+1} < x_n < \infty$, and

$$\bar{F}_{X_n, X_{n+1}}(x_n, x_{n+1}) = \frac{\log(1 + x_{n+1}^\lambda) + \frac{1}{\delta} \log(1 + x_n^\lambda) - \log(1 + x_n^\lambda) + 1}{[1 + \frac{1}{\delta} \log(1 + x_n^\lambda)][1 + \frac{1}{\delta} \log(1 + x_{n+1}^\lambda)][1 + \log(1 + x_{n+1}^\lambda)]},$$

if $0 < x_n < x_{n+1} < \infty$, which is not symmetric in (x_n, x_{n+1}). Hence, the process is not time-reversible.

7   Conclusion

We have presented a new heavy tailed distribution called Heavy Tailed Pareto distribution that is suitable for applications in various areas such as finance, economics, reliability, survival analysis, actuarial science etc. We have derived explicit expressions for the ordinary moments, median, quantile function, Rényi entropy and Shannon entropy, and the distribution of order statistics. The model parameters are estimated by method of maximum likelihood and conducted Monte Carlo simulations to study the performance of the maximum likelihood estimates of the parameters. The estimates obtained from this method are close to the true parameter values. The Marshall-Olkin Heavy Tailed Pareto distribution was also introduced and studied some of its properties. The role of new distributions in the study of autoregressive minification process was also established and studied some of its properties.

References

Aly EEA, Bouzar N (2000). On geometric infinite divisibility and stability. *Annals of the Institute of Statistical Mathematics*, 52(4): 790–799.

Arnold B (2015). *Pareto Distributions*. CRC Press, New York, 2nd edition.

Balakrishna N, Jayakumar K (1997). Bivariate semi Pareto distributions and processes. *Statistical Papers*, 38(2): 149–165.

Bourguignon M, Silva RB, Zea LM, Cordeiro GM (2013). The Kumaraswamy Pareto distribution. *Journal of Statistical Theory and Applications*, 12(2): 129–144.

Gómez-Déniz E, Calderín-Ojeda E (2015). Modelling insurance data with the Pareto ArcTan distribution. *ASTIN Bulletin: The Journal of the IAA*, 45(3): 639–660.

Jayakumar K, Ajitha B (2003). On the geometric Mittag-Leffler distributions. *Calcutta Statistical Association Bulletin*, 54(3-4): 195–208.

Jayakumar K, Krishnan B, Hamedani G (2020). On a new generalization of Pareto distribution and its applications. *Communications in Statistics-Simulation and Computation*, 49(5): 1264–1284.

Klebanov LB, Maniya G, Melamed I (1985). A problem of Zolotarev and analogs of infinitely divisible and stable distributions in a scheme for summing a random number of random variables. *Theory of Probability & Its Applications*, 29(4): 791–794.

Marshall AW, Olkin I (1997). A new method for adding a parameter to a family of distributions with application to the exponential and weibull families. *Biometrika*, 84(3): 641–652.

Merovci F, Puka L (2014). Transmuted Pareto distribution. In: *ProbStat Forum*, volume 7, 1–11.

Pickands III J (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1): 119–131.

Pillai R (1991). Semi-Pareto processes. *Journal of Applied Probability*, 28(2): 461–465.

Pillai R, Jose K, Jayakumar K (1995). Autoregressive minification processes and the class of distributions of universal geometric minima. *Journal of the Indian Statistical Association*, 33: 53–61.

Pillai R, Sandhya E (1990). Distributions with complete monotone derivative and geometric infinite divisibility. *Advances in Applied Probability*, 22(3): 751–754.

Sankaran P, Jayakumar K (2008). On proportional odds models. *Statistical Papers*, 49(4): 779–789.

Sarabia JM, Prieto F (2009). The Pareto-positive stable distribution: A new descriptive model for city size data. *Physica A: Statistical Mechanics and its Applications*, 388(19): 4179–4191.

Thomas A, Jose K (2003). Marshall-Olkin Pareto processes. *Far East Journal of Theoretical Statistics*, 9(2): 117–132.

Wang Y, Hossain AM, Zimmer WJ (2003). Monotone log-odds rate distributions in reliability analysis. *Communications in Statistics - Theory and Methods*, 32(11): 2227–2244.

Yeh HC, Arnold BC, Robertson CA (1988). Pareto processes. *Journal of Applied Probability*, 25(2): 291–301.