Data augmentation based methods for estimating the parameters of the Feller-Pareto Distribution: Theory and applications

Indranil Ghosh[1]

[1]University of North Carolina, Wilmington, USA

Keywords and Phrases: Pareto distribution, Feller-Pareto distribution, Data augmentation, Bayesian estimation, Convergence of MCMC procedure.

Abstract

In income and wealth data modeling Pareto distribution and its several variants play an important role. Both univariate and multivariate variations of this model have been extensively used as a suitable model for various non-negative socio-economic variables, for pertinent details, see Arnold (2015). In this article, we consider the most general Feller-Pareto (FP, in short) distribution, which subsumes all four types of Pareto distributions and show that it can be represented as a mixture of a conditional generalized gamma and an unconditional gamma distribution. Using this strategy, we consider a data augmentation based method (under the envelope of Bayesian paradigm) to estimate the parameters of the FP distribution. This mixture representation allows us to easily derive conditional Jeffery's type non informative priors. For illustrative purposes, one data set is considered to exhibit the utility of the proposed method.

1 Introduction

Applications of several Pareto models and its various generalizations in modeling socio-economic phenomena are well established in literature. For an excellent survey on several Pareto models along with its stochastic properties see Arnold (2015) and the references cited therein. In the hierarchy of several Pareto models, Pareto (Type IV) is the most general which subsumes three other models. Feller (1971) came up with a different representation for the Pareto (Type IV) model, expressing it as a ratio of two independent gamma variables, a distribution alternatively known as Beta distribution of the second kind. According to Feller, if $Y_i \sim \Gamma(\delta_i, 1)$, $i = 1, 2$, are independent random variables, if for $\mu \in \mathbb{R}$, $\sigma > 0$, $\gamma > 0$, we define $W = \mu + \sigma \left(\frac{Y_2}{Y_1}\right)^{\gamma}$, then W has a Feller-Pareto (FP, henceforth, in short) distribution, and we write $W \sim FP(\mu, \sigma, \delta_1, \delta_2, \gamma)$. The corresponding density can be easily obtained as

$$g(w) = \frac{\left(\frac{w-\mu}{\sigma}\right)^{\delta_1/\gamma - 1}}{\gamma\Gamma(\delta_1)\Gamma(\delta_2)} \frac{\Gamma\left(\delta_1 + \delta_2\right)}{\left(1 + \left(\frac{w-\mu}{\sigma}\right)^{1/\gamma}\right)^{\delta_1+\delta_2}} I(w > \mu). \tag{1.1}$$

There is another construction of the FP distribution as defined in the next. Let Y_1 and Y_2 be two independent radon variables having gamma distribution with scale parameter σ and shape parameters δ_1, δ_2. Then, $X = \mu + \gamma \left(\frac{Y_2}{Y_1}\right)^{\gamma}$ has FP$(\mu, \sigma, \delta_1, \delta_2, \gamma)$. Indeed Arnold (2015) has shown that the FP distribution is a generalization of the Pareto (IV) distribution.

The FP family appears to be an unimodal distribution which subsumes a variety of continuous probability models as special cases. For example, it includes Pareto (type I), Pareto (type II), Pareto (type III), and Pareto (type IV) which are identified as special cases by appropriately selecting model parameters in Eq. (1.1) given below:

- Pareto (I)(σ, α)=FP$(\sigma, \sigma, 1, \alpha, 1)$,
- Pareto (II)(μ, σ, α)=FP$(\mu, \sigma, 1, \alpha, 1)$,
- Pareto(III)(μ, σ, γ)=FP$(\mu, \sigma, \gamma, 1, 1)$,
- Pareto(IV)$(\mu, \sigma, \gamma, \alpha)$=FP$(\mu, \sigma, \gamma, \alpha, 1)$,

Another special case of the FP distribution is the transformed beta family which includes several well-known probability models such as Burr, Generalized Pareto, and Inverse Burr among others. Noticeably, a salient feature of these distributions is that they possess relatively high probability in the upper tail. However, it is also interesting to note that there are some distributions that exhibit distinctly non-Paretian behavior in the upper tail. For instance, Log-logistic, Inverse Pareto, and Inverse Paralogistic -each is a special case of Inverse Burr have relatively "light" tails as independently observed by Brazauskas (2002). Application of such probability models covers a wide spectrum of areas ranging from actuarial science, economics, finance to health science domain and telecommunications, for distributions of variables such as sizes of insurance claims, incomes in a population of people, stock price fluctuations, duration of responses to medical treatment, and length of telephone calls. For pertinent details, see Arnold (1983, 2015); Johnson et al. (1994); Klein and Moeschberger (1997). Moreover, some of these distributions are relevant within much broader classes of probability models. For example, a generalized Pareto distribution arises in semiparametric modeling of upper observations in samples from distributions which are regularly varying or in the domain of attraction of extreme value distributions, see Embrechts et al. (1997). This motivates us to consider the study of this distribution from the estimation perspective.

Next, without loss of generality, we consider $\mu = 0$. Alternatively, we can estimate μ from a given data as $w_{1:n} = \min_{1 \leq i \leq n} w_i$, which is a consistent estimate of μ. Then, we can subtract from w and set μ equal to zero. Kalbfleisch and Prentice (1980) identified Eq. (1.1) as a generalized F density. However, the seemingly unattractive and complicated (although it does have a closed form density) expression of the density function discouraged researchers to investigate further about this model. In the literature attempts have been made to discuss structural properties of the FP distribution as well as methodologies related to methods of estimation under a frequentist approach. A not-exhaustive list of such references are mentioned as follows. Tahmasebi and Behboodian (2010) have discussed and derived the exact analytical expressions of entropy for the FP family and order statistics of FP subfamilies. Dutang et al. (2022) developed an `R package actuar` for implementing FP distribution in actuarial applications numerically. Odubote and Oluyede (2014) discussed a weighted version of a Feller-Pareto distribution and discussed several useful structural properties. Brazauskas (2022) derived the exact form of Fisher Information matrix for the FP distribution. However, none of the above cited references have discussed the estimation of the model parameters under a frequentist approach. Additionally, classical estimation appears to have serious limitations in efficiently estimating the parameters of the FP model that has a total of 5 parameters. In particular, one is faced with the following problems:

- The classical maximum likelihood method of estimation may not perform satisfactorily well, because we have 4 parameters (assuming $\mu = 0$) in total, and the associated likelihood is not a well behaved function as it involves gamma functions. Moreover, even if we estimates for the parameters, it is quite difficult to examine mathematically or otherwise, whether or not the estimated values are global or local maximum.

- From Eq.(1.1), with $\mu = 0$, the k-th order moment ($k \geq 2$) of FP distribution will be

$$E\left(W^k\right) = \frac{\sigma^k \Gamma(k\gamma + \delta_2)\Gamma(\delta_1 - k\gamma)}{\Gamma(\delta_1)\Gamma(\delta_2)}.$$

  It is quite obvious from this expression that for the k-th order of moment to exist, we must have $\delta_1 > k\gamma$. This is quite a strong assumption, and at the same time we do not know for sure whether for a given data set, this condition will be satisfied. Although, one may use the method of fractional moments (which always exists), but that too, might not yield satisfactory results as several estimation methods under the classical set-up involves selecting a starting value for the parameters under study.

As a remedy, we seek a different approach in this paper. We write the density function in Eq. (1.1) as a mixture of one conditional generalized gamma and an unconditional gamma distribution. Later on, we will show that this mixture representation helps us to derive easily conditional Jeffery's type non informative prior. This is quite fascinating in the sense that we are not assuming any external prior, but by rewriting the model in terms of two known distributions by invoking the argument of data augmentation. Needless to say, similar technique might well be considered for a multivariate FP distribution. The rest of the paper is organized as follows. In Section 2, the maximum likelihood estimation method is used to estimate the model parameters for the FP distribution given in Eq. (1.1). In section 3, we discuss mixture representation of the density given in Eq. (1.1). In Section 4, we discuss a simulation study. Section 5 deals with the application of the proposed methodology to a data set having varying complexities. Finally, some concluding remarks are presented in Section 6.

2 Estimation by the method of maximum likelihood

For a random sample of size n the associated log-likelihood function drawn from the probability model in Eq. (1.1) will be

$$\begin{aligned}\log L\left(\sigma, \gamma, \delta_1, \delta_2\right) &= \left(\frac{\delta_1}{\gamma} - 1\right) \sum_{i=1}^{n} \log\left(\frac{W_i}{\sigma}\right) + n\left(\Gamma\left(\delta_1 + \delta_2\right)\right) - n\left\{\Gamma(\delta_1) + \Gamma(\delta_2)\right\} \\ &\quad -n \log\gamma - (\delta_1 + \delta_2) \sum_{i=1}^{n} \log\left(1 + \left(\frac{W_i}{\sigma}\right)^{1/\gamma}\right). \qquad (2.1)\end{aligned}$$

The corresponding maximum likelihood equations will be by differentiating Eq. (2.1) with respect to γ, σ, δ_1 and δ_2 will be

$$\frac{\partial \log L(\sigma,\gamma,\delta_1,\delta_2)}{\partial\gamma} = -(\delta_1+\delta_2)\sum_{i=1}^{n} - \frac{\left(\frac{W_i}{\sigma}\right)^{\frac{1}{\gamma}}\log\left(\frac{W_i}{\sigma}\right)}{\gamma^2\left(\left(\frac{W_i}{\sigma}\right)^{\frac{1}{\gamma}}+1\right)} - \frac{\delta_1\sum_{i=1}^{n}\log\left(\frac{W_i}{\sigma}\right)}{\gamma^2} - \frac{n}{\gamma}. \quad (2.2)$$

$$\frac{\partial \log L(\sigma,\gamma,\delta_1,\delta_2)}{\partial\sigma} = -(\delta_1+\delta_2)\sum_{i=1}^{n} - \frac{W_i\left(\frac{W_i}{\sigma}\right)^{\frac{1}{\gamma}-1}}{\gamma\sigma^2\left(\left(\frac{W_i}{\sigma}\right)^{\frac{1}{\gamma}}+1\right)} - \frac{n\left(\frac{\delta_1}{\gamma}-1\right)}{\sigma}. \quad (2.3)$$

$$\begin{aligned}\frac{\partial \log L(\sigma,\gamma,\delta_1,\delta_2)}{\partial\delta_1} &= \frac{\sum_{i=1}^{n}\log\left(\frac{W_i}{\sigma}\right)}{\gamma} - \sum_{i=1}^{n}\log\left(\left(\frac{W_i}{\sigma}\right)^{\frac{1}{\gamma}}+1\right) + n\Gamma(\delta_1+\delta_2)\psi^{(0)}(\delta_1+\delta_2) \\ &\quad -n\Gamma(\delta_1)\psi^{(0)}(\delta_1). \quad (2.4)\end{aligned}$$

$$\begin{aligned}\frac{\partial \log L(\sigma,\gamma,\delta_1,\delta_2)}{\partial\delta_2} &= -\sum_{i=1}^{n}\log\left(\left(\frac{W_i}{\sigma}\right)^{\frac{1}{\gamma}}+1\right) + n\Gamma(\delta_1+\delta_2)\psi^{(0)}(\delta_1+\delta_2) \\ &\quad -n\Gamma(\delta_2)\psi^{(0)}(\delta_2). \quad (2.5)\end{aligned}$$

The maximum likelihood estimates of the parameters $\sigma,\gamma,\delta_1,\delta_2$ are obtained by equating Eqs. (2.2)-(2.5) to zero, and $\psi^{(0)}(z) = \frac{d}{dz}\log(\Gamma(z))$, is the polygamma function.

For the elements of the Fisher information (FIM, in short), an interested reader is referred to the paper by Brazauskas (2002) with γ_1 and γ_2 need to be replaced by δ_1 and δ_2 as per the notation utilized in this paper in the probability density function in Eq. (1.1).

The asymptotic variance-covariance matrix of the MLE $\underline{\widehat{\theta}} = \left(\widehat{\sigma},\widehat{\gamma},\widehat{\delta}_1,\widehat{\delta}_2\right)$ can be obtained from the inverse of the observed FIM as

$$\mathbf{U} = \mathbf{I}^{-1} \overset{def.}{=} \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ u_{21} & u_{22} & u_{23} & u_{24} \\ u_{31} & u_{32} & u_{33} & u_{34} \\ u_{41} & u_{42} & u_{43} & u_{44} \end{bmatrix}.$$

Lehmann and Casella (2006) have discussed the asymptotic normality of the MLE under certain conditions (see, Theorem 3.10, p.449). For the FP distribution given in Eq. (1.1), it is straightforward to see that

$$\left|\frac{\partial^4}{\partial\sigma\,\partial\gamma\,\partial\delta_1\,\partial\delta_2}[\log f_{\underline{X}}(\underline{x}|\theta)]\right| = 0.$$

In addition, all the remaining necessary and sufficient conditions of Theorem 3.10 of Lehmann and Casella (2006) are satisfied, and therefore, it can be assumed that

$$\left(\widehat{\sigma}, \widehat{\gamma}, \widehat{\delta}_1, \widehat{\delta}_2\right) \overset{asymp}{\sim} N_4\left(\sigma, \gamma, \delta_1, \delta_2, (\mathbf{U}[\underline{\theta}])^{-\mathbf{1}}_{\mathbf{jj}}\right).$$

Consequently, a $100\,(1-q)\%$ approximate confidence intervals of the parameters $\widehat{\theta}_i$, will be

$$\widehat{\theta}_i \pm Z_q \times \sqrt{u_{ii}}, \tag{2.6}$$

$i = 1, 2, 3$, where Z_q is the $100q$-th upper percentile of the standard normal distribution.

2.1 Simulation Study

In this section, we conduct a Monte Carlo simulation study to evaluate the performance of the likelihood inference for the FP distribution given in Eq. (1.1). Random samples from the FP distribution are obtained using the `actuar` package in R. In particular, we consider the sample sizes $n = 50$, 75 and 100 with the following six sets of choices of the model parameters.

(a) Choice 1: $\sigma = 2$, $\gamma = 1.5$ and $\delta_1 = 0.35$, $\delta_2 = 0.35$.

(b) Choice 2: $\sigma = 2.5$, $\gamma = 2$ and $\delta_1 = 0.65$, $\delta_2 = 0.75$.

(c) Choice 3: $\sigma = 1.75$, $\gamma = 1.2$ and $\delta_1 = 1.35$, $\delta_2 = 1.40$.

(d) Choice 4: $\sigma = 1.45$, $\gamma = 1.6$ and $\delta_1 = 1.40$, $\delta_2 = 1.52$.

(e) Choice 5: $\sigma = 3$, $\gamma = 1.8$ and $\delta_1 = 1.65$, $\delta_2 = 1.75$.

(f) Choice 6: $\sigma = 3.5$, $\gamma = 2.25$ and $\delta_1 = 0.95$, $\delta_2 = 0.85$.

For each setting, 20000 sets of random samples are generated. Regarding the MLE estimates, we have mimicked the strategy adopted in Duatang et al. (2022). For each simulated random sample, we also compute the 95% approximate confidence intervals for the parameters σ, γ, δ_1, and δ_2 based on Eq. (2.6) with the asymptotic variances obtained from inverting the observed Fisher information matrix as well as the approximated variances obtained from a parametric bootstrap method with 400 bootstrap samples, for details on the bootstrapping, see Efron and Tibshirani (1993). The estimated biases and mean squared errors (MSEs) of the MLEs of σ, γ, δ_1, and δ_2 are presented in Table 2.1. The estimated coverage probabilities and average widths of the confidence intervals are presented in Table 2.2. Since the observed information need not be positive definite which results in negative asymptotic variances, see, for example, Verbeke and Molenberghs (2007). Additionally, we also represent the percentage of cases in which the asymptotic variances are negative and the confidence intervals cannot be computed in Table 2.2. In those cases that the asymptotic variances are negative, we recommend using a parametric bootstrap method as an alternative method to approximate the variances of the MLEs.

One may observe from Table 2.1, that the estimated MSEs for the four parameters σ, γ, δ_1 and δ_2 decrease as the sample size increases. However, for the estimated biases, there is not a steady decreasing pattern with the increase of sample sizes, and on the contrary, in some cases, it appears that there is a negligible amount (by $0.01 - 0.05$) of increase. Whether this is an anomaly or not will be investigated in a separate article. We also observe that the estimated MSEs of δ_2 is larger than the MSEs of σ, γ and δ_1. Next, from Table 2.2, one may also observe the following

- that the proportions of cases in which negative variance estimates are obtained is negligibly small.
- the computed approximate confidence intervals based on bootstrap variances performs satisfactorily well. Note that these approximate confidence intervals can be used as an alternative when the asymptotic variances are negative, for pertinent details, see Ghosh and Ng (2019) and the references cited therein.

3 Mixture representation

In this section, we represent the mixture representation of the FP distribution. Suppose, $X_1 \sim \Gamma(\delta_1, 1)$ and $X_2 \sim \Gamma(\delta_2, 1)$ and they are independent. Let us define $Z_i = X_i^{\gamma}$, $i = 1, 2$, with $\gamma > 0$. Then, the distribution of Z_i will be

$$f(z_i) = \frac{1}{\gamma \Gamma(\delta_i)} \exp\left(-z_i^{1/\gamma}\right) z_i^{\delta_i/\gamma - 1} \quad z_i > 0.$$

Note that the above density function has been independently derived and explored by Stacy (1962). The joint density of Z_1 and Z_2 will be

Table 2.1: Simulated biases and MSEs of the MLEs of the parameters in the FP distribution for various choices of $\sigma, \gamma, \delta_1, \delta_2$ and n.

Parameter choice	n	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Choice 1	50	0.231	0.153	0.101	0.169	0.423	0.568	0.342	2.754
	75	0.213	0.141	0.095	0.127	0.393	0.428	0.313	2.538
	100	0.158	0.122	0.087	0.325	0.365	0.242	0.207	2.032
Choice 2	50	0.091	0.157	0.138	0.147	0.233	0.329	0.268	3.028
	75	0.081	0.133	0.115	0.138	0.201	0.311	0.224	2.987
	100	0.079	0.118	0.103	0.106	0.189	0.287	0.219	2.567
Choice 3	50	0.130	0.156	0.033	0.126	0.144	0.023	0.151	2.468
	75	0.132	0.143	0.028	0.098	0.125	0.021	0.127	2.275
	100	0.126	0.118	0.021	0.074	0.104	0.018	0.112	1.986
Choice 4	50	0.085	0.192	0.128	0.137	0.142	0.123	0.295	3.228
	75	0.091	0.154	0.119	0.126	0.132	0.114	0.233	3.107
	100	0.084	0.148	0.103	0.115	0.125	0.107	0.221	2.978
Choice 5	50	0.026	0.021	0.019	0.151	0.166	0.103	0.228	4.237
	75	0.024	0.017	0.015	0.141	0.152	0.087	0.213	3.893
	100	0.023	0.010	0.016	0.132	0.134	0.066	0.175	3.218
Choice 6	50	0.115	0.128	0.133	0.067	0.121	0.020	0.265	2.249
	75	0.102	0.098	0.114	0.054	0.113	0.021	0.218	2.177
	100	0.093	0.088	0.101	0.036	0.102	0.018	0.194	2.037

Table 2.2: Simulated coverage probabilities (CP) and average widths (AW) of the MLEs of the parameters in the FP distribution for various choices of $\sigma, \gamma, \delta_1, \delta_2$.

Parameter choice		Based on asymptotic variances from inverting **I**									Based on bootstrap variances							
		σ		γ		δ_1		δ_2			σ		γ		δ_1		δ_2	
	n	CP	AW	CP	AW	CP	AW	CP	AW	% of negative variance	CP	AW	CP	AW	CP	AW	CP	AW
Choice 1	50	0.950	1.377	0.952	0.563	0.992	2.361	0.0475	0.913	1.388	0.938	0.535	0.922	1.732	0.912	0.932	0.928	0.917
	75	0.939	1.243	0.953	0.487	0.986	1.261	0.090	0.814	1.344	0.938	0.485	0.917	1.534	0.902	0.925	0.908	0.928
	100	0.935	1.137	0.959	0.432	0.982	0.584	0.120	0.923	1.299	0.935	0.452	0.927	1.335	0.875	0.911	0.896	0.922
Choice 2	50	0.905	1.444	0.940	0.577	0.997	2.569	0.130	0.912	1.508	0.950	0.570	0.958	1.137	0.928	0.918	0.948	0.931
	75	0.882	1.292	0.943	0.498	0.993	1.189	0.170	0.940	1.422	0.945	0.508	0.959	1.032	917	.906	.917	0.927
	100	0.853	1.203	0.943	0.448	0.988	1.032	0.090	0.925	1.362	0.944	0.469	0.954	0.812	0.903	0.912	0.894	0.905
Choice 3	50	0.950	1.371	0.949	0.561	0.993	2.481	0.110	0.918	1.392	0.943	0.535	0.926	1.643	0.932	0.868	0.932	0.936
	75	0.941	1.229	0.956	0.484	0.986	1.264	0.170	0.904	1.345	0.935	0.486	0.914	1.345	0.875	0.782	0.916	0.913
	100	0.930	1.117	0.955	0.428	0.978	1.172	0.110	0.911	1.292	0.934	0.449	0.924	0.733	0.835	0.638	0.835	0.908
Choice 4	50	0.904	1.437	0.940	0.575	0.997	2.556	0.130	0.936	1.478	0.945	0.564	0.955	1.542	0.917	0.886	0.953	0.951
	75	0.882	1.289	0.943	0.496	0.993	1.218	0.090	0.935	1.406	0.944	0.504	0.953	1.046	0.901	0.863	0.927	0.938
	100	0.853	1.199	0.943	0.447	0.988	1.043	0.100	0.921	1.373	0.947	0.474	0.947	0.938	0.893	0.843	0.906	0.923
Choice 5	50	0.915	1.428	0.944	0.657	0.982	1.768	0.127	0.945	1.389	0.951	0.546	0.936	1.238	0.926	0.893	0.918	0.936
	75	0.869	1.285	0.924	0.483	0.973	1.528	0.108	0.938	1.201	0.934	0.501	0.913	1.032	0.893	0.865	0.889	0.928
	100	0.842	1.152	0.913	0.436	0.934	1.134	0.094	0.922	1.158	0.915	0.472	0.906	0.942	0.836	0.843	0.873	0.915
Choice 6	50	0.896	1.634	0.938	0.893	1.032	1.489	0.134	0.942	1.385	0.953	0.726	0.946	1.413	0.943	0.886	0.905	0.933
	75	0.874	1.518	0.943	0.496	0.975	1.017	0.090	0.964	1.406	0.932	0.518	0.942	1.025	0.938	0.834	0.868	0.928
	100	0.848	1.243	0.943	0.447	0.981	0.936	0.100	0.912	1.373	0.914	0.462	0.934	0.967	0.903	0.805	0.831	0.921

$$f(z_1, z_2) = \frac{1}{\gamma^2\Gamma(\delta_1)\Gamma(\delta_2)} \exp\left\{-(z_1^{1/\gamma} + z_2^{1/\gamma})\right\} z_1^{\delta_1/\gamma-1} z_2^{\delta_2/\gamma-1}, \quad z_i > 0, i = 1, 2.$$

Next, let us consider the following transformation. $W = \sigma\frac{Z_1}{Z_2}$, with $\sigma > 0$, and $U = Z_2$. We are interested in the distribution of W. The Jacobian of the above transformation is $|J| = \frac{u}{\sigma}$. Then, the joint distribution of W and U will be

$$f(u, w) = \frac{1}{\gamma^2\sigma\Gamma(\delta_1)\Gamma(\delta_2)} \left(\frac{uw}{\sigma}\right)^{\delta_1/\gamma-1} u^{\delta_2/\gamma-1} \left(\frac{u}{\sigma}\right) \exp\left(-\left(\left(\frac{uw}{\sigma}\right)^{1/\gamma} + u^{1/\gamma}\right)\right) I(w > 0, u > 0).$$

Hence, the marginal distribution of W will be

$$\begin{aligned} g(w) &= \frac{\left(\frac{w}{\sigma}\right)^{\delta_1/\gamma-1}}{\gamma^2\sigma\Gamma(\delta_1)\Gamma(\delta_2)} \int_0^\infty u^{\frac{\delta_1+\delta_2}{\gamma}-1} \exp\left(- u^{1/\gamma}(\left(\frac{w}{\sigma}\right)^{1/\gamma} + 1)\right) du \\ &= \frac{\left(\frac{w}{\sigma}\right)^{\delta_1/\gamma-1}}{\sigma\gamma\Gamma(\delta_1)\Gamma(\delta_2)} \frac{\Gamma(\delta_1 + \delta_2)}{\left(1 + \left(\frac{w}{\sigma}\right)^{1/\gamma}\right)^{\delta_1+\delta_2}} I(w > 0). \end{aligned} \tag{3.1}$$

Noticeably, it is the ratio of two independent Stacy random variables with the same shape parameter. Jordanova, P.K et al. (2023) discussed and studied the distribution of the ratio of two independent Stacy random variables when both the shape parameters for the numerator and denominator random variables is equal to one. However, it must be noted that Jordanova, P.K et al. (2023) did not work on a subset of this current work. Precisely, the authors in that paper work on different sets of the values of the powers in the numerator and denominator, and just have an intersection when these powers are equal to 1.

Next, observe that we can rewrite Eq. (3.1) as

$$g(w^*) = \int_0^\infty f_1(w^*|u) f_2(u) du,$$

where $W^* = \frac{W}{\sigma}$ and

- $W|U = u \sim \Gamma\left(\frac{\delta_1}{\gamma}, u\right)$, with

$$f_1(w|u) = \frac{u^{\delta_1/\gamma}}{\gamma\Gamma(\delta_1)} w^{\delta_1/\gamma-1} \exp\left\{-uw^{1/\gamma}\right\} I(w > 0).$$

- $U \sim \Gamma\left(\frac{\delta_2}{\gamma}, 1\right)$.

In this case, the associated complete data likelihood takes the following form:

$$
\begin{aligned}
& L\Bigg(\delta_1, \delta_2, \gamma, W_1, W_2, \cdots, W_n, U_1, U_2, \cdots, U_n\Bigg) \\
&= \frac{1}{(\gamma\sigma)^n}\left(\frac{1}{\Gamma(\delta_1)}\right)^n \left(\frac{1}{\Gamma(\delta_2)}\right)^n \prod_{i=1}^{n} u_i^{\delta_2/\gamma-1} \exp\left\{-\sum_{i=1}^{n} u_i^{1/\gamma}\right\} \\
&\times \prod_{i=1}^{n} (u_i w_i)^{\delta_1/\gamma} \exp\left\{-\sum_{i=1}^{n} (u_i w_i)^{1/\gamma}\right\} \\
&= \left(\frac{1}{\gamma\sigma\Gamma(\delta_1)\Gamma(\delta_2)}\right)^n \prod_{i=1}^{n} u_i^{(\delta_1+\delta_2)/\gamma-1} \prod_{i=1}^{n} \left(\frac{w_i}{\sigma}\right)^{\delta_1/\gamma-1} \exp\left\{-\sum_{i=1}^{n} u_i^{1/\gamma}\left(1+\left(\frac{w_i}{\sigma}\right)^{1/\gamma}\right)\right\}.
\end{aligned}
\tag{3.2}
$$

Next, we try to derive from Eq.(3.2), the full conditionals of $(\delta_1, \delta_2, \gamma, \sigma, U_1, U_2, \cdots, U_n)$, given $(W_1, W_2, \cdots, W_n)$ as follows:

- For $i = 1, 2, \cdots, n$,
  $U_i | u_{-i}, \underline{w}, \gamma, \delta_1, \delta_2, \sigma \propto u_i^{(\delta_1+\delta_2)/\gamma-1} \exp\left\{-u_i^{1/\gamma}\left(1+\left(\frac{w_i}{\sigma}\right)^{1/\gamma}\right)\right\}$
  $\sim$ Generalized gamma $\left((\delta_1+\delta_2)/\gamma, 1/\left(1+\left(\frac{w_i}{\sigma}\right)^{1/\gamma}\right), \gamma\right)$.

- $$
\begin{aligned}
&\gamma | \underline{u}, \underline{w}, \delta_1, \delta_2, \sigma \\
&\propto \exp\left\{\frac{1}{\gamma}\left((\delta_1+\delta_2)\sum_{i=1}^{n}\log u_i + \delta_1 \sum_{i=1}^{n} \log\frac{w_i}{\sigma}\right) - \sum_{i=1}^{n} u_i^{1/\gamma}\left(1+\left(\frac{w_i}{\sigma}\right)^{1/\gamma}\right)\right\}\left(\frac{1}{\gamma\sigma}\right)^n.
\end{aligned}
$$

- $\delta_2 | \underline{u}, \underline{w}, \delta_1, \gamma, \sigma \propto \left(\frac{1}{\sigma\Gamma(\delta_2)}\right)^n \exp\left\{\frac{\delta_2}{\gamma}\sum_{i=1}^{n}\log u_i\right\}$. Note that this function is log concave.

- $\delta_1 | \underline{u}, \underline{w}, \delta_2, \gamma, \sigma \propto \left(\frac{1}{\Gamma(\delta_1)}\right)^n \exp\left\{\frac{\delta_2}{\gamma}\sum_{i=1}^{n}\left(\log u_i + \log\frac{w_i}{\sigma}\right)\right\}$. Note that this function is also log concave.

- Finally, $\sigma | \underline{u}, \underline{w}, \delta_1, \delta_2, \gamma \propto \left(\frac{1}{\gamma\Gamma(\delta_1)\Gamma(\delta_2)}\right)^n \exp\Bigg(-\sum_{i=1}^{n} u_i^{1/\gamma}\left(1+\left(\frac{w_i}{\sigma}\right)^{1/\gamma}\right)\Bigg)$.

Therefore, all $(n+4)$ full conditionals can be sampled using Acceptance Rejection (AR) sampling using the R-package `ars`.

On the choice of hyperparameters for the priors:

In this case, the hyperparameter values for the Gamma priors are obtained via the method of matching first two theoretical moments with the sample moments. Needless to say that there are other available strategies of selecting hyperparameters, such as the procedure adopted by Giannone, D. et al. (2013) in relation to vector autoregression models, Singh, P., & Hellander, A. (2018) in the context of hyperparameter optimization, etc. However, the methodology adopted in such references might be more applicable to spatial-temporal models rather than the model that we have here. Additionally, whether such strategies will be beneficial (in the sense of computational efficiency, computation time) is a subject matter of a future study. Furthermore, it is safe to opine that we are not claiming that this set of priors with these specific choice of hyperparameters is optimum for the associated Bayesian analysis, but, in our cases, we have tried several other prior choices and found minimal changes in the final estimates. Also, the acceptance probability across found to be between $(0.36, 0.73)$, which indicates that the chain mixing was satisfactory. A full scale study regarding a wide range of prior choices needs to be done which is beyond the scope of this present paper.

3.1 Comment on the Convergence of MCMC Procedure

For the convergence diagnostics of the adopted MCMC procedure, the method of Gelman and Rubin (1992) is utilized. It involves two stages. The first step, before sampling begins, involves obtaining an over-dispersed estimate of the target distribution(s) and using these to generate the starting points for the desired number of independent chains (in our case ,we consider running a MCMC procedure with $m = 10$ parallel independent chains of length $2k$=40000 each). The second step involves using the last $k = 1000$ iterations to re-estimate the target distribution of the scalar quantity of interest (in our case σ, γ, δ_1 and δ_2). The convergence of the MCMC convergence is monitored by the following quantity given below:

$$\sqrt{\hat{R}} = \sqrt{\left\{\frac{k-1}{n} + \frac{(m+1)B}{mkW}\right\} \frac{df}{(df-2)}},$$

where B is the variance between the means from the $m = 10$ parallel chains, W

is the average of the $m = 10$ within-chain variances and df is the degrees of freedom of the approximating t density, for details, see pp. 465, Eq. (20) of Gelman & Rubin (1992). Convergence is said to have achieved once the value of $\sqrt{\hat{R}}$ is near 1 for all scalar estimands of interest for a sufficiently large k ($k \to \infty$). It is to be noted here that Gelman and Rubin (1992) had proposed the method for Gibbs sampler but here we have applied their procedure on the output of MCMC procedure which is a particular case of Gibbs sampling. For illustrative purposes, we provide the values of $\sqrt{\hat{R}}$ for all of our scalar quantities of interest under the partial dependent conjugate prior set up (provided in Tables 6.1 − 6.3 in Appendix).

4 Simulation study

For the associated Bayesian analysis, we consider the following two different sets of prior choices, namely the non-informative & improper and conjugate prior set up. At first, we consider the non-informative prior set-up and we label it as Choice 1.

Choice 1 (Non-informative prior set-up):

$$\Pi(\sigma) \propto \frac{1}{(1+\sigma)^2} I\left(\sigma > 0\right).$$

$$\Pi(\gamma) \propto \frac{1}{(1+\gamma)^2} I\left(\gamma > 0\right).$$

$$\Pi(\delta_1) \propto \frac{1}{(1+\delta_1)^2} I\left(\delta_1 > 0\right).$$

$$\Pi(\delta_2) \propto \frac{1}{(1+\delta_2)^2} I\left(\delta_2 > 0\right).$$

Therefore, the joint prior in this case will be $\Pi(\sigma) \times \Pi(\gamma) \times \Pi(\delta_1) \times \Pi(\delta_2)$.

The associated posterior summary is represented in Table 4.1.

Choice 2 (Conjugate prior set-up):

Next, under the conjugate prior set-up, we consider the same prior choices for all the parameters, i.e., $(\delta_1, \delta_2, \gamma, \sigma)$ each follows a Gamma distribution with shape=0.2, and rate=0.2 respectively. We consider using WINBUGS (a software for computing Bayesian analysis) with `R` interface to conduct the simulation study. We do not claim that the prior choice made here is optimum, but among different choices of the hyperparameters made for this study, this choice appears to be producing satisfactory results for the Bayesian analysis based on a random sample of size $n = 100$. The associated posterior summary is provided in Table and 4.2.

For both the cases, we consider the following four sets of true parameter choices.

(a) Choice 1: $\sigma = 0.07$, $\gamma = 0.25$ and $\delta_1 = 0.75$, $\delta_2 = 0.82$.

(b) Choice 2: $\sigma = 0.158$, $\gamma = 3.10$ and $\delta_1 = 0.65$, $\delta_2 = 0.75$.

(c) Choice 3: $\sigma = 4.98$, $\gamma = 6.70$ and $\delta_1 = 0.57$, $\delta_2 = 0.61$.

(d) Choice 4: $\sigma = 3.22$, $\gamma = 3.50$ and $\delta_1 = 0.52$, $\delta_2 = 0.53$.

Remark. From the posterior simulation studies as given in Tables 4.1 and 4.2, one may comment the following:

- Under the non-informative prior set-up the associated 95% HPD intervals are slightly wider.
- We cannot make a general comment as to whether prior conjugacy in this scenario will perform uniformly better or not. A full scale study involving other possible prior choices and hyperparameter values along with a flat prior of the form $\Pi(\theta) \propto 1$ will be the subject matter of a separate article.

5 Real data application

In this section, we consider a data on fault-trace lengths from Clark et al. (1999). As per the authors, mimicking their words, "It has often been observed that fault-trace lengths tend to follow a power-law or Pareto distribution, at least for sufficiently large lengths." The applicability of the FP model under the classical paradigm (using the MLE method) has already been discussed in Clark et al. (1999). In this paper, we discuss the Bayesian estimation of the FP model parameters. One prominent reason for selecting the FP distribution is that it is the most general model in the hierarchy of Pareto distribution and many other distributions have been assumed to fit this particular data set which assumes a Pareto like behavior (especially mimicking the tail-behavior). Here, we consider the estimates of the parameters using moment matching (i.e, equating the sample moments with the theoretical moments) strategy and treat them as the initial prior choice for the

Table 4.1: Posterior summary for the FP model under the non-informative prior assumption.

Parameter choices	$\widehat{\sigma}$		$\widehat{\gamma}$		$\widehat{\delta_1}$		$\widehat{\delta_2}$	
	PM	95% HPD	PM	95% HPD	PM	95% HPD	PM	95% HPD
Choice 1	0.0713	(0.03269, 1.3226)	0.2434	(0.1729, 0.8924)	0.7348	(0.5314, 0.8821)	0.8016	(0.3478, 1.0112)
Choice 2	0.1684	(0.1242, 1.0823)	3.2653	(1.4522, 3.6592)	0.6491	(0.3835, 0.8754)	0.7156	(0.3458, 0.9622)
Choice 3	5.0321	(3.1642, 6.0106)	6.2336	(2.1326, 7.8725)	0.5559	(0.4904, 0.9525)	0.5998	(0.3547, 0.9767)
Choice 4	3.2678	(1.324, 4.2134)	3.485	(1.4782, 4.4218)	0.5078	(0.4012, 0.7459)	0.4982	(0.2756, 0.9166)

Table 4.2: Posterior summary for the FP model under the conjugate prior assumption.

Parameter choices	$\widehat{\sigma}$		$\widehat{\gamma}$		$\widehat{\delta_1}$		$\widehat{\delta_2}$	
	PM	95% HPD	PM	95% HPD	PM	95% HPD	PM	95% HPD
Choice 1	0.0725	(0.05148, 0.2316)	0.2456	(0.1831, 0.7161)	0.7403	(0.5618, 0.8503)	0.8026	(0.4268, 1.0654)
Choice 2	0.1502	(0.1369, 0.8324)	3.1138	(1.6361, 3.5842)	0.6582	(0.3841, 0.9765)	0.7015	(0.3149, 0.9232)
Choice 3	4.942	(3.8634, 6.2044)	0.5427	(0.2105, 0.7684)	0.5902	(0.3824, 0.9523)	0.5938	(0.3326, 0.9813)
Choice 4	3.2535	(1.559, 4.1314)	3.4673	(1.5728, 4.4281)	0.5169	(0.3928, 0.7326)	0.5128	(0.1875, 0.9107)

Bayesian analysis. The summary of the Bayesian output is given in Table 5.1. Regarding the convergence of the adopted MCMC, we provide the $\sqrt{\hat{R}}$ values given in the Appendix. For the Bayesian analysis, we consider the following two different sets of prior choices:

- Choice 1 (Non-informative prior set-up):

$$\Pi(\sigma) \propto \frac{1}{\sigma} I\left(\sigma > 0\right).$$

$$\Pi(\gamma) \propto \frac{1}{\gamma} I\left(\gamma > 0\right).$$

$$\Pi(\delta_1) \propto \frac{1}{\left(1 + \delta_1\right)^2} I\left(\delta_1 > 0\right).$$

$$\Pi(\delta_2) \propto \frac{1}{\left(1 + \delta_2\right)^2} I\left(\delta_2 > 0\right).$$

- Choice 2 (Conjugate prior set-up):
  - $\Pi(\sigma) \sim Gamma(shape = 0.27, rate = 0.35)$.
  - $\Pi(\gamma) \sim Gamma(shape = 1.27, rate = 2.09)$.
  - $\Pi(\delta_1) \sim Gamma(shape = 1.09, rate = 2.12)$.
  - $\Pi(\delta_2) \sim Gamma(shape = 1.45, rate = 3.50)$.

From the 95% HPD intervals for the conjugate prior set-up, it appears that the Bayesian estimation under the non-informative (and improper) prior set-up is less efficient as expected. The summary of the output is given in Table 4.1.

Sensitivity analysis with respect to the hyperparameters:

An anonymous referee has expressed concern that the results of this analysis might be influenced by the choice of hyperparameters. As an initial effort to investigate hyperparameters sensitivity, we re-analyze the data set with two other sets of prior choices that are given as follows (for the conjugate prior set-up):

1. **Choice 3:**
   - $\Pi(\sigma) \sim Gamma(shape = 0.32, rate = 0.49)$.

Table 5.1: Posterior summary for the 4 parameter FP model on fault-trace lengths data.

Sampling Algorithm	$\widehat{\sigma}$		$\widehat{\gamma}$		$\widehat{\delta_1}$		$\widehat{\delta_2}$	
	P.M.	95% HPD	P.M.	95% HPD	P.M.	95% HPD	P.M.	95% HPD
Non-conjugate prior (Choice 1)	2.4163	(1.345, 3.721)	1.6892	(1.146, 2.923)	3.2754	(2.563, 4.659)	2.2347	(1.785, 3.267)
Conjugate prior (Choice 2)	2.4213	(1.525, 3.423)	1.6647	(1.268, 2.534)	3.2649	(2.785, 3.867)	2.2218	(1.867, 3.104)

- $\Pi(\gamma) \sim Gamma(shape = 2.49, rate = 3.28)$.
- $\Pi(\delta_1) \sim Gamma(shape = 1.53, rate = 2.58)$.
- $\Pi(\delta_2) \sim Gamma(shape = 1.68, rate = 1.39)$.

2. **Choice 4:**

- $\Pi(\sigma) \sim Gamma(shape = 0.68, rate = 0.89)$.
- $\Pi(\gamma) \sim Gamma(shape = 2.38, rate = 2.73)$.
- $\Pi(\delta_1) \sim Gamma(shape = 1.78, rate = 3.46)$.
- $\Pi(\delta_2) \sim Gamma(shape = 1.95, rate = 2.34)$.

The posterior summaries based on the above two different prior choices (with varying choices of the hyperparameters) are given in Tables 5.2. It appears that with these two sets of new choices of hyperparameters for the respective priors the final conclusion remains the same, i.e., the performance of the Bayesian inference under the conjugate prior set-up performs slightly better as compared to that of under the non-informative prior set-up.

6 Concluding remarks

In this paper, we have discussed estimation of the model parameters of the Feller-Pareto distribution under both frequentist and Bayesian paradigm. Under the frequentist approach, we consider the MLE method and also discussed the asymptotic normality of the estimates under certain regularity conditions. Noticeably, as mentioned in Section 1, estimation of the parameters under the classical approach is usually hindered by the fact that the likelihood function is not well behaved, and the parameter space have constraints. Some of these problems can be avoided by using computing environment such as `R` using specific packages that are designed to handle such types of complicated models in terms of their estimation. Nevertheless, effficient estimation of model parameters for the most general members of the Pareto-family, namely the Feller-Pareto distribution under the classical approach is difficult to achieve, as the existence of moments has some conditions that are hard to satisfy from a practical point of view. However, by rewriting the density function in Eq. (1.1) as a mixture of two well-known continuous probability models and by invoking the strategy of data augmentation, we have conducted a Bayesian analysis under both non-informative & improper and conjugate prior set-up. Based on our study,

Table 5.2: Posterior summary for the 4 parameter FP model on fault-trace lengths data under different conjugate prior choices (Choice 3 and Choice 4.

Sampling Algorithm	$\widehat{\sigma}$		$\widehat{\gamma}$		$\widehat{\delta_1}$		$\widehat{\delta_2}$	
	P.M.	95% HPD	P.M.	95% HPD	P.M.	95% HPD	P.M.	95% HPD
Choice 3	2.4039	(1.312, 3.589)	1.6735	(1.138, 2.891)	3.2587	(2.607, 3.587)	2.2108	(1.793, 3.117)
Choice 4	2.4168	(1.486, 3.479)	1.6803	(1.275, 2.586)	3.2604	(2.674, 3.673)	2.2346	(1.831, 3.168)

it is difficult to comment as to which of the two procedures (MLE and Bayesian) will be performing uniformly better than the other. Now, in the absence of large samples and in the presence of prior information, one can hope that the inferential aspect under the Bayesian paradigm will be better than any strategy under the classical set-up. The results of a small simulation study is encouraging. We have also discussed the utility of the proposed strategy of estimation of the Feller-Pareto distribution in the context of a real-life data set. A more comprehensive study is required to explore the usefulness of Bayesian estimation from a standard Bayesian analysis point of view.

Several possible works in the future direction can be considered, including but not limited to

- choice of other types of improper priors as opposed to the ones assumed in this paper and then conduct a Bayesian analysis.
- selecting a dependent & full conditional priors from the exponential family as opposed to subjective conjugate and proper priors assumed in this paper.
- estimation of the model parameters by other methods under then classical paradigm, such as the method of maximum product spacing distance, Cramer Von mises, method of ordinary and weighted least square estimators etc., and have a comparison study to see the relative efficiency of each of these estimation strategies in the context of such type of probability models.

Appendix

In this section, we discuss the proofs of the results mentioned in Section 3.

Log concavity of distributions corresponding to mixture representation in Section 3 :

First of all, note that a real valued function $g()$ is said to be log-concave on the interval (a, b) if the function $\log g()$ is a concave function on (a, b). Equivalently, one can say, $\frac{g^{'}()}{g()}$ is monotonically decreasing on (a, b) or $(\log^{``} g() < 0,)$ where $``$ stands for the double derivative and $'$ stands for the first order derivative respectively.

Result 1: The kernel of the conditional density of

$$\gamma|\underline{u}, \underline{w}, \delta_1, \delta_2 \propto \exp\left\{\frac{1}{\gamma}\left((\delta_1+\delta_2)\sum_{i=1}^{n}\log u_i + \delta_1\sum_{i=1}^{n}\log w_i\right) - \sum_{i=1}^{n} u_i^{1/\gamma}\left(1+w_i^{1/\gamma}\right)\right\}\left(\frac{1}{\gamma}\right)^n$$

is log concave and integrable.

Proof. We need to show that the kernel of this distribution is log concave and integrable. Let us define $A_1 = \sum_{i=1}^{n} u_i^{1/\gamma}\left(1 + w_i^{1/\gamma}\right)$. Then

$$A_1' = \frac{\partial A_1}{\partial \gamma} = \log\left(1/\gamma\right)\left[A_1 + \sum_{i=1}^{n}\left(u_i w_i\right)^{1/\gamma}\right].$$

Therefore,

$$\frac{A_1'}{A_1} = 1 + \log\left(1/\gamma\right)\left\{\frac{\sum_{i=1}^{n}\left(u_i w_i\right)^{1/\gamma}}{\sum_{i=1}^{n}\left(u_i w_i\right)^{1/\gamma} + \sum_{i=1}^{n}\left(u_i\right)^{1/\gamma}}\right\} = 1 + \log\left(1/\gamma\right)\left\{\frac{1}{1 + \frac{\sum_{i=1}^{n}(u_i)^{1/\gamma}}{\sum_{i=1}^{n}(u_i w_i)^{1/\gamma}}}\right\}. \tag{6.1}$$

Note that, from Eq. (6.1), for integer values $\gamma > 0$ and with $(u_i, w_i) > 0, \quad \forall i = 1, 2, \cdots, n$, as γ increasing, the quantity $\log(1/\gamma)$ is decreasing. Also, the term inside the second parenthesis is decreasing. Hence A_1 is log-concave.

Next, consider

$$A_2 = \frac{1}{\gamma}\left(\left(\delta_1 + \delta_2\right)\sum_{i=1}^{n}\log u_i + \delta_1\sum_{i=1}^{n}\log w_i\right)$$

.

For the sake of notational simplicity, we write

$$B = \left(\delta_1 + \delta_2\right)\sum_{i=1}^{n}\log u_i + \delta_1\sum_{i=1}^{n}\log w_i.$$

Note that B is positive (based on the argument as mentioned earlier.) Also, $A_2 = \frac{B}{\gamma}$.

Therefore,

$$\frac{\partial^2 \log A_2}{\partial \gamma^2} = \frac{1}{\gamma^2} > 0.$$

So, A_2 is log-convex. Again, since A_1 is log-concave $-A_1$ is log-convex. Now, our kernel for the conditional density

$$A_2 + (-A_1).$$

This is sum of two log-convex functions, and therefore, it is log-convex.

We must make a note of the following:

- For integer values of γ, $\log(1/\gamma)$ is a decreasing function.
- For fractional values of γ (i.e., $0 < \gamma < 1$) $\log(1/\gamma)$ is increasing.

- For any values of γ, and with any fractional positive values of (u_i, w_i) the ratio $\frac{1}{1+\frac{\sum_{i=1}^{n}(u_i)^{1/\gamma}}{\sum_{i=1}^{n}(u_i w_i)^{1/\gamma}}}$ is decreasing.
- But, for integer values of (u_i, w_i) and for any values of γ, the above ratio is increasing.

  However, since the quantity is of the form $\frac{1}{1+\text{positive quantity}}$, it will be always < 1.

Similarly, one can establish the log-concavity (and/or log-convexity) of all the other conditional distributions.

On the integrability of all the conditional density functions for AR sampling:

Here we begin with the following:

1. Since the conditional density of U_i given $U_{-i}, \gamma, \delta_1, \delta_2$ is a generalized gamma, indeed it is integrable.
2. The conditional density of γ is given by (without the normalizing constant)

$$\gamma|\underline{u}, \underline{w}, \delta_1, \delta_2 \propto \exp\left\{\frac{1}{\gamma}\left((\delta_1+\delta_2)D_1+\delta_1 D_2\right) - \sum_{i=1}^{n} u_i^{1/\gamma}\left(1+w_i^{1/\gamma}\right)\right\}\left(\frac{1}{\gamma}\right)^n,$$

where $D_1 = \sum_{i=1}^{n} \log u_i$, and $D_2 = \sum_{i=1}^{n} \log w_i$. Next, if consider the transformation $\theta = \frac{1}{\gamma}$, then the conditional density of θ given $\underline{u}, \underline{w}, \delta_1, \delta_2$ will be

$$g\left(\theta|\underline{u}, \underline{w}, \delta_1, \delta_2\right) \propto \exp\left\{\theta\left((\delta_1+\delta_2)D_1+\delta_1 D_2\right) - \sum_{i=1}^{n} u_i^{\theta}\left(1+w_i^{\theta}\right)\right\}\theta^{n-2} = M_1 M_2, \tag{6.2}$$

where $M_1 = \exp\left(\theta T_1\right)\theta^{n-2}$, $M_2 = \exp\left(\sum_{i=1}^{n} u_i^{\theta}\left(1+w_i^{\theta}\right)\right)$, and $T_1 = (\delta_1+\delta_2)D_1 + \delta_1 D_2$.

Next, note that $0 < \gamma < \infty$ will imply $0 < \theta < \infty$. Also, $(u_i, w_i) \geq 0$. So, the function M_2 will exhibit the following:

- $\lim_{\theta\to 0} M_2 = \exp(-2n) < \infty$, provided $n < \infty$.
- $\lim_{\theta\to\infty} M_2 \Longrightarrow 0$.

Table 6.1: Values of $\sqrt{\hat{R}}$ for n=100.

Chain number	$\sqrt{\hat{R}_\sigma}$	$\sqrt{\hat{R}_\gamma}$	$\sqrt{\hat{R}_{\delta_1}}$	$\sqrt{\hat{R}_{\delta_2}}$
1	0.81	1.07	1.08	0.71
2	0.92	1.09	1.10	0.86
3	1.01	1.05	0.83	0.85
4	0.96	1.02	0.77	1.01
5	1.01	0.96	0.87	1.02
6	0.97	1.01	1.06	0.96
7	0.89	0.93	0.98	0.87
8	0.91	0.95	0.84	0.85
9	0.92	0.82	0.86	0.88
10	0.96	0.83	0.89	0.95

Therefore, M_2 is a bounded function on the interval $[0, \infty)$.

Next, let us consider the integrability of the quantity M_1

$$\begin{aligned} \int_0^\infty M_1 d\theta &= \int_0^\infty \exp(\theta T_1)\, \theta^{n-2} d\theta \\ &= \int_0^\infty \sum_{j=0}^\infty \frac{(\theta T_1)^j}{j!} \theta^{n-2} d\theta \\ &= \sum_{j=0}^\infty \frac{(T_1)^j}{j!} \int_0^\infty \theta^{n-2+j} d\theta. \end{aligned}$$

This integral will be bounded and integrable, provided $n + j < 2$. This appears to be a impossible condition. Hence the proof.

Table 6.2: Values of $\sqrt{\hat{R}}$ for n=200.

Chain number	$\sqrt{\hat{R}_\sigma}$	$\sqrt{\hat{R}_\gamma}$	$\sqrt{\hat{R}_{\delta_1}}$	$\sqrt{\hat{R}_{\delta_2}}$
1	0.83	0.84	0.93	0.77
2	0.84	0.91	1.04	0.79
3	1.01	1.03	0.93	0.81
4	0.81	0.95	0.89	1.01
5	0.83	0.96	0.92	0.77
6	0.87	0.92	0.86	0.95
7	0.76	0.81	0.92	0.91
8	0.82	0.84	0.85	0.93
9	0.97	0.88	0.87	0.96
10	0.96	0.95	0.92	0.89

References

[1] Arnold, B.C. (2008). Pareto and generalized Pareto distributions. In Modeling Income Distributions and Lorenz Curves. Edited by: Chotikapanich D. Springer, New York.

[2] Arnold, B. C. (2014). Univariate and multivariate Pareto models. Journal of Statistical Distributions and Applications, 1, 1-16.

[3] Arnold, B.C. (2015). Pareto Distributions, 2nd edition. Series: Chapman & Hall/CRC Monographs on Statistics & Applied Probability.

[4] Brazauskas, V. (2002). Fisher information matrix for the Feller–Pareto distribution. Statistics & probability letters, 59(2), 159-167.

[5] Clark, R. M., Cox, S. J. D., & Laslett, G. M. (1999). Generalizations of power-law distributions applicable to sampled fault-trace lengths: model choice, parameter estimation and caveats. Geophysical Journal International, 136(2), 357-372.

[6] Dutang, C., Goulet, V., & Langevin, N. (2022). Feller-Pareto and related distributions: Numerical implementation and actuarial applications. Journal of Statistical Software, 103, 1-22.

Table 6.3: Values of $\sqrt{\hat{R}}$ for n=500.

Chain number	$\sqrt{\hat{R}_\sigma}$	$\sqrt{\hat{R}_\gamma}$	$\sqrt{\hat{R}_{\delta_1}}$	$\sqrt{\hat{R}_{\delta_2}}$
1	0.81	1.07	1.01	0.71
2	0.92	1.09	1.03	0.86
3	1.01	0.95	0.87	0.98
4	0.78	0.89	0.84	0.94
5	1.02	0.83	0.84	0.93
6	0.84	0.91	0.86	0.94
7	0.77	0.93	0.92	0.78
8	0.85	0.82	0.79	0.92
9	0.90	0.89	0.97	0.95
10	0.92	0.93	0.94	0.99

[7] Efron, B., & Tibshirani, R. J. (1994). An introduction to the bootstrap. CRC press, New York.

[8] Ghosh, I., & Ng, H. K. T. (2019). A class of skewed distributions with applications in environmental data. Communications in Statistics: Case Studies, Data Analysis and Applications, 5(4), 346-365.

[9] Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (1995). *Bayesian Data Analysis.* Chapman and Hall, New York.

[10] Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 7, 457-511.

[11] Giannone, D., Lenza, M., & Primiceri, G. E. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, 97(2), 436-451.

[12] Kalbfleisch , J.D, and Prentice, R.L.The Statistical Analysis of Failure Time Data. Wiley, New York, 1980.

[13] Lehmann, E. L., & Casella, G. (2006). Theory of point estimation. Springer Science & Business Media.

[14] Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS-a Bayesian modeling framework: concepts, structure, and extensibility. Statistics and computing, 10(4), 325-337.

[15] Odubote, O., & Oluyede, B. O. (2014). Theoretical Properties of the Weighted Feller-Pareto Distributions. Asian Journal of Mathematics and Applications, 2014, 1.

[16] Jordanova, P.K., Savov, M., Tchorbadjieff, A., & Stehlík, M. (2023) Mixed Poisson process with Stacy mixing variable, Stochastic Analysis and Applications, DOI: 10.1080/07362994.2023.2242471.

[17] Robert, C.P. and Casella, G. (1999). *Monte Carlo Statistical Methods.* Springer, New York.

[18] Rubenstein, R.Y. (1981). *Simulation and the Monte Carlo Method.* John Wiley and Sons, New York.

[19] Singh, P., & Hellander, A. (2018). Hyperparameter optimization for approximate bayesian computation. *In 2018 Winter Simulation Conference (WSC) (pp. 1718-1729). IEEE.*

[20] Stacy, E.W. (1962) A generalization of the Gamma distribution, Annals of Mathematical Statistics, vol. 33 (3), pp. 1187-1192.

[21] Tahmasebi, S., & Behboodian, J. (2010). Shannon entropy for the Feller-Pareto (FP) family and order statistics of FP subfamilies. Applied Mathematical Sciences, 4(10), 495-504.

[22] Verbeke, G., & Molenberghs, G. (2007). What can go wrong with the score test? The American Statistician, 61(4), 289-290.