STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **13**, May 2025, pp 1789–1799.
Published online in International Academic Press (www.IAPress.org)

A Bayesian Method for Estimation of the Entropy in the Presence of Outliers Based on the Contaminated Pareto Model

Mahsa Salajegheh, Mehdi Jabbari Nooghabi*, Kheirolah Okhli

Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran

Abstract Shannon entropy and Fisher information are pivotal in the information theory area. The presence of outliers in data and using an inappropriate model may cause misleading inferential results in the amount of information. Our aim of this paper is to compute the amount of Shannon entropy and Fisher information that exists in the Pareto distribution in the presence of multiple outliers. Unlike the existing methods in the literature, we present a good method for the estimation of Shannon entropy and Fisher information to cope with the allowing for the possibility of outliers. In this regard, we focus on the Bayesian approach proposed by [32] based on the contaminated Pareto distribution. We implement the Gibbs sampler which is a simple and rational method for computing Bayesian estimation of Shannon entropy and Fisher information under different loss functions. Some simulation studies are conducted to investigate the performance of the proposed methodology under various sample sizes and the number of outliers. In the end, two examples of real insurance claim data are studied to illustrate the superiority of the proposed model in analyzing datasets and computing the amount of Shannon entropy and Fisher information.

Keywords Shannon entropy, Fisher information, Outliers, Insurance claim data, Contaminated Pareto model, Bayesian analysis

AMS 2010 subject classifications 94A17, 62G32

DOI: 10.19139/soic-2310-5070-1997

1. Introduction

Term entropy, is one of the most notable terms in statistics and information theory, which originally goes back to the works of [8] and [6] in thermodynamics. The idea of information-theoretic entropy was redefined by [33] and later by [37] in Cybernetics. The concept of entropy in information theory is the uncertainty involved in predicting the value of a random variable. Several kinds of entropy have been introduced hitherto and Shannon entropy is the most famous kind of entropy. After [33], a large number of papers, books, and monographs have been published on its extensions and applications over the past years, which can be mentioned to [30, 35, 9, 2, 3, 7, 20, 38, 1, 26, 18, 34, 31].

The Shannon entropy of a random variable X assuming its values in D_x with probability density function $f(x;\theta)$ is defined by

$$\begin{aligned} En(X) &= E\Big(-\ln\big(f(x;\theta)\big)\Big) \\ &= -\int_{D_x} f(x;\theta)\ln\big(f(x;\theta)\big)dx. \end{aligned} \tag{1}$$

*Correspondence to: Mehdi Jabbari Nooghabi (Email: jabbarinm@um.ac.ir). Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

Fisher information is one of the other forms of information measure which is considered in this paper. The Fisher information was presented for the first time by [14] which gives a way of measuring the amount of information that a random variable carries about an unknown parameter. Fisher information in a random variable X is

$$\begin{aligned} I_X(\theta) &= E\left(\left[\frac{\partial}{\partial\theta}\ln\big(f(x;\theta)\big)\right]^2\right) \\ &= -E\left(\frac{\partial^2}{\partial\theta^2}\ln\big(f(x;\theta)\big)\right) \end{aligned} \tag{2}$$

In some statistical data analyses, there are some conditions where some of the points in a sample data are too small or too large compared to the rest of the observations which are called outliers ([4]). An important cause of arising outliers is the contaminations that theoretically do not come from the same distribution that the remaining observations come from ([28]). One of the most considered methodologies to cope with the allowing for the possibility of outliers is to use an appropriate model. Interested readers can refer to [11], [32], [19], [29], and [23] for estimating several distributions in the presence of outliers. In this paper, we will obtain the amount of information such as Shannon entropy and Fisher information when outliers exist in a sample.

The Pareto distribution was expressed by Pareto in 1897 as a model for the analysis of income. Later, [5] showed The Pareto distribution has an important role in modeling insurance claims data, especially in automobile insurance problems. [12, 13] used the famous Dixit model for the analysis of data in the presence of outliers by implementing the Pareto distribution. The amount of Shannon entropy in the insurance field is important because know that from the point of insurance managers, the best insurance policy for an insurance company is one that lasts for a long period of time and has more uncertainty in terms of claims. [22] characterized the insurance demand in terms of the entropy of the underlying probability distribution for losses. [24] obtained the amount of information such as Shannon entropy, Tsallis entropy, Fisher information, and Kullback-Leibler distance that exists in the Pareto distribution in the presence of a small number of outliers. They used the marginal distribution of the famous Dixit model (not the joint distribution of the Dixit model) for obtained Tsallis entropy, Fisher information, and Kullback-Leibler distance that may cause misleading inferential conclusions. Also, [17] obtained the maximum likelihood and Bayesian estimators of entropy under different loss functions in the presence of outliers by the marginal distribution of the Dixit model. In this paper, we present an appropriate, simple, and without restriction to the number of outliers and sample size method for obtaining the amount of information such as Shannon entropy and Fisher information when outliers exist in a Pareto sample. Also unlike [24] and [17], this paper will use the joint distribution to compute the amount of information in the Pareto sample in the presence of outliers.

[32] assumed a Pareto sample in the presence of upper outliers and presented a Bayesian approach based on the contaminated Pareto (CP) model using the Gibbs sampler. Following [32], [29] considered the contaminated exponential distribution and extended a Bayesian methodology using the Gibbs sampler for positive-valued insurance data when outliers exist in the sample. They discussed that the Bayesian analysis can be simplified by using Markov chain Monte Carlo such as the Gibbs sampler. Our aim in this paper, compute the amount of information such as Shannon entropy and Fisher information in the presence of outliers with a focus on the proposed Bayesian approach by [32]. In this regard, we obtained the amount of Shannon entropy and Fisher information for the CP model with a focus on the Bayesian approach under different loss functions when the outliers exist in the sample. Numerical studies demonstrate that the proposed method can be utilized with a small, moderate, or large sample size and various numbers of outliers. Furthermore, we obtain a preferable loss function in our study.

The rest of the paper is outlined as follows. Section 2 presents the formulation of the CP model and obtains Shannon entropy and Fisher information on all parameters of the proposed CP model. The results and analysis of the simulation studies with different numbers of outliers and sample sizes (small, moderate, and large) are exhibited in Section 3. In Section 4, we present two examples of real insurance claim datasets to illustrate the outperformance of the proposed methodology. Finally, a brief conclusion of the article is discussed in Section 5.

2. Theory information

According to [32], we start with a definition and notation of the CP model with a focus on the Bayesian methodology in this section. Suppose $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ be a random sample from $CP(\alpha, \theta, \beta, \epsilon)$. The probability density function of X_i is

$$f_{CP}(x_i; \alpha, \theta, \beta, \epsilon) = \epsilon f_P(x_i; \alpha, \theta\beta) + (1-\epsilon) f_P(x_i; \alpha, \theta), \qquad i = 1, 2, \ldots, n\,, \tag{3}$$

where $f_P(\alpha, \beta)$ denotes the probability density function of the Pareto distribution with the shape α and scale β parameters. Scollnik (2015) [32] re-expressed the CP model (3) to the following form

$$f_{CP}(x_i; \alpha, \theta, \beta, \delta_i) = \frac{\alpha\theta^{\alpha}\beta^{\alpha\delta_i}}{x_i^{\alpha+1}} \boldsymbol{I}\big(x_i - \theta\beta^{\delta_i}\big), \qquad i = 1, 2, \ldots, n\,, \tag{4}$$

where α, θ and β are shape, scale and contaminated parameters, respectively. Also, $\boldsymbol{I}$ is the indicator function and $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_n)^\top$ are independent Bernoulli random variables with an identical probability of success given by ϵ. In other words, we have

$$X_i|\boldsymbol{\delta} \sim Pa\big(\alpha, \theta\beta^{\delta_i}\big), \qquad i = 1, 2, \ldots, n\,, \tag{5}$$

where $Pa(.,.)$ denotes the Pareto distribution. Thus, the conditional likelihood is given by

$$L(\boldsymbol{x}|\alpha, \theta, \beta, \boldsymbol{\delta}, \epsilon) = \frac{\alpha^n \theta^{n\alpha} \beta^{\alpha \sum_{i=1}^n \delta_i}}{\prod_{i=1}^n x_i^{\alpha+1}} \prod_{i=1}^n \boldsymbol{I}\big(x_i - \theta\beta^{\delta_i}\big). \tag{6}$$

Furthermore, Scollnik [32] presented the full conditional posterior density of $\epsilon, \boldsymbol{\delta}, \alpha, \theta$ and β are given by

$$(\epsilon|\boldsymbol{x}, \alpha, \theta, \beta, \boldsymbol{\delta}) \sim Beta\Big(b_1 + \sum_{i=1}^n \delta_i, n + b_2 - \sum_{i=1}^n \delta_i\Big), \tag{7}$$

$$(\delta_i|\boldsymbol{x}, \alpha, \theta, \beta, \epsilon) \sim Ber\Big(\frac{\epsilon\beta^{\alpha}}{1-\epsilon+\epsilon\beta^{\alpha}}\Big), \quad \beta\theta < x_i, i = 1, 2, \ldots, n, \tag{8}$$

$$(\alpha|\boldsymbol{x}, \theta, \beta, \boldsymbol{\delta}) \sim Gamma\Big(n + a_1, a_2 + \sum_{i=1}^n \ln x_i - n\ln(\theta) - \sum_{i=1}^n \delta_i \ln(\beta)\Big), \ \alpha > 2, \tag{9}$$

$$(\theta|\boldsymbol{x}, \alpha, \beta, \boldsymbol{\delta}) \sim Gamma\Big(t_1 + n\alpha, t_2)\Big), \quad \theta < min_i(\frac{x_i}{\beta^{\delta_i}}), \ i = 1, 2, \ldots, n, \tag{10}$$

and

$$(\beta|\boldsymbol{x}, \alpha, \theta, \boldsymbol{\delta}) \sim Gamma\Big(\alpha\sum_{i=1}^n \delta_i, \lambda\Big), \quad 1 < \beta^* < \beta < \frac{min_{i\ni\delta_i=1}(x_i)}{\theta}, \tag{11}$$

respectively. Keep in mind that $Ber(.,.)$ and $Gamma(.,.)$ indicate the Bernoulli and gamma distribution, respectively, and $a_1, a_2, b_1, b_2, t_1, t_2$ and λ are the hyperparameter values (see Scollnik (2015) [32] for details).

Now in the next subsections, we obtain the amount of information such as Shannon entropy and Fisher information based on the proposed model and by implementing Gibbs sampler.

2.1. Shannon entropy

Shannon entropy of the CP model is obtained in the following theorem based on the Bayesian methodology. This quantity is derived for both forms, marginal and joint posterior distributions of the contaminated Pareto in the presence of outliers.

Theorem 1
Let random variable X_i follow a the CP model (4), so the Shannon entropy of X_i is

$$En(X_i) = \ln\Big(\frac{\theta\beta^{\delta_i}}{\alpha} e^{\frac{\alpha+1}{\alpha}}\Big). \tag{12}$$

Proof
See A. □

Theorem 2
Suppose $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ be a random sample from $CP(\alpha, \theta\beta^{\delta_i})$, $i = 1, 2, \ldots, n$, then the Shannon entropy for the joint distribution of the contaminated Pareto in the presence of outliers is

$$En(\boldsymbol{X}) = \ln\Big(\frac{\theta^n \beta^{\sum_{i=1}^n \delta_i}}{\alpha^n} e^{n+\frac{n}{\alpha}}\Big). \tag{13}$$

Proof
Based on Shannon entropy Eq. (1) and the form of the conditional likelihood (6), we have

$$\begin{aligned} En(\boldsymbol{X}) &= E\Big(-\ln\Big(f(\boldsymbol{x}, \alpha, \theta, \beta, \boldsymbol{\delta}, \epsilon)\Big)\Big) \\ &= E\Big(-\ln\Big(\prod_{i=1}^n f(x_i, \alpha, \theta, \beta, \delta_i, \epsilon)\Big)\Big) \\ &= \sum_{i=1}^n En(X_i). \end{aligned}$$

Using Theorem 1, therefore we have

$$\begin{aligned} En(\boldsymbol{X}) &= \sum_{i=1}^n \ln\Big(\frac{\theta\beta^{\delta_i}}{\alpha} e^{\frac{\alpha+1}{\alpha}}\Big) \\ &= \ln\Big(\frac{\theta^n \beta^{\sum_{i=1}^n \delta_i}}{\alpha^n} e^{n+\frac{n}{\alpha}}\Big). \end{aligned}$$

□

The next subsection allocated to computing the Fisher information for the CP distribution.

2.2. Fisher information

The following theorem describe Fisher information of all parameters of the proposed CP model.

Theorem 3
Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a random sample from the proposed CP model (4), so the amount of Fisher information there exist in $\boldsymbol{X}$ about parameters α, θ and β are

$$I_{\boldsymbol{X}}(\alpha) = \frac{n}{\alpha^2}, \tag{14}$$

$$I_{\boldsymbol{X}}(\theta) = \frac{n\alpha}{\theta^2}, \tag{15}$$

and

$$I_{\boldsymbol{X}}(\beta) = \frac{\alpha\sum_{i=1}^n \delta_i}{\beta^2}, \tag{16}$$

respectively.

Proof
See B. □

In order to compute Bayesian estimates of the Shannon entropy and Fisher information, samples are generated from the full conditional posterior distribution by implementing the Gibbs sampler ([15, 27, 16]). Refer to the C for details on the Gibbs algorithm.

3. Simulation study

In this section, we investigate the performance of the proposed CP model for computing Shannon entropy and Fisher information via simulation for small, moderate, and large samples and various numbers of outliers. In this regard, we take $n = 15, 20, 25, 50, 100, 1000$ aiming to have the small, moderate, and large samples and various numbers of outliers ranging from 1 to 5, i.e. $\sum_{i=1}^{n} \delta_i = 1, 2, 3, 4, 5$. In addition, we take the presumed parameters $\alpha = 3, \theta = 50,000, \beta = 7$ and $N = 10,000$ with a burn-in period of 2000 during each replication of 1000 trials. Also, the assumed hyperparameter values are $a_1 = 10, t_1 = 3, \lambda = 1, \beta^* = 1.1, b_1 = 0.18$ and $b_2 = 3.5$ whereas a_2 and t_2 are calculated the conditional prior mean is close to the maximum likelihood estimate. In order to monitor the convergence of the MCMC simulations, we employ the scale reduction factor estimate, recommended by Gelman et al. (2013). The scale factors for all MCMC chains for the sequences of ϵ, α, θ and β are within $1.0000 - 1.00001$, indicating their convergence.

In order to simulate outliers, we consider $\delta_i = 1$ in the CP model (5) and generate outliers of size $\sum_{i=1}^{n} \delta_i = k$ from $Pa\big(\alpha, \theta\beta^{\delta_i}\big)$. Also to simulate main (without outlier) samples, we consider $\delta_i = 0$ in the CP model (5) and generate samples of size $n - k$ from $Pa\big(\alpha, \theta\beta^{\delta_i}\big)$.

3.1. Estimation of Shannon entropy under different loss functions in the presence multiple of outliers

This subsection presents the Bayes estimation of Shannon entropy under different loss functions, to investigate the performance of the proposed methodology in the presence of multiple outliers. In addition, we compare the Bayes estimation obtained under different loss functions. In this regard, consider squared error loss function (SELF), precautionary loss function (PLF), and DeGroot loss function (DLF) presented by [21], [25], and [10], respectively.

Table 1. The Bayes estimation of Shannon entropy under presumed parameters $\alpha = 3, \theta = 50,000$ and $\beta = 7$.

n	$\sum_{i=1}^{n} \delta_i$	Exact value	SELF	PLF	DLF	n	$\sum_{i=1}^{n} \delta_i$	Exact value	SELF	PLF	DLF
15	1	167.8	220.2	202.9	186.1	50	1	554.7	661.8	650.2	639.2
15	2	169.7	290.0	233.9	197.6	50	2	556.6	701.0	678.6	654.9
15	3	171.7	484.5	291.6	211.5	50	3	558.6	776.9	721.0	656.7
15	4	173.6	1287.7	428.6	229.7	50	4	560.5	923.1	792.4	667.9
15	5	175.5	4611.1	738.1	248.3	50	5	562.5	1174.4	888.4	680.4
20	1	223.0	283.8	267.6	251.8	100	1	1107.4	1303.4	1293.6	1284.9
20	2	225.0	352.9	298.6	252.4	100	2	1109.3	1345.1	1330.5	1315.8
20	3	226.9	514.5	354.4	266.4	100	3	1111.3	1384.9	1363.7	1340.1
20	4	228.9	1004.2	463.4	286.6	100	4	1113.2	1446.5	1411.3	1370.8
20	5	230.8	2817.2	717.0	319.5	100	5	1115.2	1516.2	1461.3	1391.8
25	1	278.3	344.9	329.6	314.8	1000	1	11056.4	12838.1	12829.2	12821.2
25	2	280.3	407.8	362.9	324.6	1000	2	11058.4	12870.5	12861.3	12853.0
25	3	282.2	542.7	408.5	319.2	1000	3	11060.3	12912.8	12903.2	12894.5
25	4	284.1	874.8	500.3	326.5	1000	4	11062.3	12954.7	12944.7	12935.6
25	5	286.1	2134.6	737.4	387.3	1000	5	11064.2	12993.3	12983.0	12973.6

Tables 1 and 2 report the Bayes estimation of Shannon entropy under SELF, PLF and DLF. From the reported results in Table 1, it can be seen that the exact value of Shannon entropy is an increasing function with respect to the sample size and the number of outliers. The Bayes estimation of Shannon entropy under SELF, PLF and DLF is an increasing function with respect to the sample size. As expected, the Bayes estimation of Shannon entropy under different loss functions increases by increasing the number of outliers, indicating that the CP model and its Bayesian method are efficient when outliers are in the sample. Also, the Bayes estimation of Shannon entropy is large for the outlier sample, indicating that it is interesting for an insurance company. From the reported results in Table 2, it can be seen that the MSEs of Shannon entropy under the DLF are smaller than the SELF and PLF for various sample sizes and the number of outliers.

Table 2. MSEs of Shannon entropy under presumed parameters $\alpha = 3, \theta = 50,000$ and $\beta = 7$.

n	$\sum_{i=1}^{n} \delta_i$	SELF	PLF	DLF	n	$\sum_{i=1}^{n} \delta_i$	SELF	PLF	DLF
15	1	11273	1350	861	50	1	24034	11036	9647
15	2	156970	4192	1517	50	2	94958	18095	13930
15	3	2008908	14636	2486	50	3	227842	32212	21646
15	4	29256111	57686	3913	50	4	2549096	83261	32693
15	5	15026676590	2428404	6507	50	5	6446234	139050	43479
20	1	11981	2091	1514	100	1	51229	39195	38031
20	2	123340	6018	2778	100	2	74966	53696	51889
20	3	1507755	17844	4361	100	3	113949	70010	65676
20	4	215722476	188473	6871	100	4	179194	92609	85683
20	5	2095917704	1065438	10538	100	5	350339	126516	109197
25	1	26773	3923	2430	1000	1	3265066	3199229	3198396
25	2	117094	7775	4073	1000	2	3374657	3307342	3306622
25	3	723067	18507	6548	1000	3	3522397	3452092	3451432
25	4	10613291	64606	9827	1000	4	3678705	3603766	3603017
25	5	234026620	358367	15294	1000	5	3814124	3735283	3734566

3.2. Estimation of Fisher information under different loss functions in the presence multiple of outliers

This subsection presents the Bayes estimation of Fisher information under different loss functions, to investigate the performance of the proposed methodology in the presence of multiple outliers. In this regard, we compute the Bayes estimation of Fisher information of all parameters of the proposed CP model under SELF, PLF and DLF.

Table 3 reports the Bayes estimation of Fisher information of all parameters under different loss functions. From the reported results in Table 3, it can be seen that the Bayes estimation of Fisher information of α, θ and β parameters under DLF is closer than the exact value of Fisher information. It is clear that the exact value of Fisher information of α and θ parameters is an increasing function with respect to the sample size. While, the exact value of Fisher information of $beta$ parameter increases when the number of outliers increased. As expected, the results of simulation show that the Bayes estimation of Fisher information of α and θ parameter under SELF, PLF and DLF is an increasing function with respect to the sample size n. Furthermore by increasing the number of outliers, the Bayes estimation of Fisher information of α parameter under SELF, PLF and DLF increases, meanwhile the Bayes estimation of Fisher information of θ parameter under SELF, PLF and DLF decreased.

Table 3. Fisher information under presumed parameters $\alpha = 3, \theta = 50,000$ and $\beta = 7$.

n	$\sum_{i=1}^{n} \delta_i$	Exact value			SELF			PLF			DLF		
		$I_{\boldsymbol{X}}(\alpha)$	$I_{\boldsymbol{X}}(\theta)$	$I_{\boldsymbol{X}}(\beta)$	$I_{\boldsymbol{X}}(\alpha)$	$I_{\boldsymbol{X}}(\theta)$	$I_{\boldsymbol{X}}(\beta)$	$I_{\boldsymbol{X}}(\alpha)$	$I_{\boldsymbol{X}}(\theta)$	$I_{\boldsymbol{X}}(\beta)$	$I_{\boldsymbol{X}}(\alpha)$	$I_{\boldsymbol{X}}(\theta)$	$I_{\boldsymbol{X}}(\beta)$
15	1	1.7	1.80E-08	6.1224E-02	69	3.172E-09	1.130E-02	29	4.689E-09	4.211E-02	8	8.264E-09	1.935E-01
15	2	1.7	1.80E-08	1.2245E-01	517	1.343E-09	7.067E-03	118	2.600E-09	3.216E-02	19	5.917E-09	1.782E-01
15	3	1.7	1.80E-08	1.8367E-01	4738	4.909E-10	2.478E-03	527	1.330E-09	1.566E-02	43	4.219E-09	1.211E-01
15	4	1.7	1.80E-08	2.4490E-01	73459	1.323E-10	5.660E-04	3030	5.831E-10	5.860E-03	96	2.937E-09	7.431E-02
15	5	1.7	1.80E-08	3.0612E-01	1262575	3.247E-11	1.286E-04	17322	2.471E-10	2.248E-03	180	2.161E-09	4.932E-02
20	1	2.2	2.40E-08	6.1224E-02	66	4.848E-09	1.121E-02	34	6.533E-09	4.147E-02	14	9.771E-09	1.761E-01
20	2	2.2	2.40E-08	1.2245E-01	395	2.275E-09	6.634E-03	108	4.082E-09	3.240E-02	15	1.028E-08	2.340E-01
20	3	2.2	2.40E-08	1.8367E-01	2700	9.744E-10	2.772E-03	403	2.305E-09	1.799E-02	30	7.667E-09	1.754E-01
20	4	2.2	2.40E-08	2.4490E-01	24543	3.475E-10	8.744E-04	1710	1.187E-09	8.104E-03	69	5.329E-09	1.079E-01
20	5	2.2	2.40E-08	3.0612E-01	309666	1.015E-10	2.360E-04	8916	5.360E-10	3.226E-03	185	3.338E-09	5.851E-02
25	1	2.8	3.00E-08	6.1224E-02	61	6.859E-09	1.245E-02	36	8.759E-09	4.446E-02	18	1.200E-08	1.739E-01
25	2	2.8	3.00E-08	1.2245E-01	283	3.588E-09	8.373E-03	103	5.641E-09	3.767E-02	27	1.041E-08	2.092E-01
25	3	2.8	3.00E-08	1.8367E-01	1584	1.708E-09	3.268E-03	287	3.713E-09	2.170E-02	22	1.227E-08	2.345E-01
25	4	2.8	3.00E-08	2.4490E-01	10349	7.280E-10	1.260E-03	1038	2.090E-09	1.105E-02	31	1.106E-08	1.959E-01
25	5	2.8	3.00E-08	3.0612E-01	120885	2.251E-10	3.752E-04	5699	9.332E-10	4.532E-03	180	4.729E-09	7.549E-02
50	1	5.6	6.00E-08	6.1224E-02	75	1.657E-08	1.568E-02	60	1.839E-08	4.579E-02	46	2.074E-08	1.319E-01
50	2	5.6	6.00E-08	1.2245E-01	143	1.233E-08	1.455E-02	100	1.457E-08	4.830E-02	63	1.809E-08	1.720E-01
50	3	5.6	6.00E-08	1.8367E-01	381	8.148E-09	7.615E-03	192	1.112E-08	3.467E-02	65	1.848E-08	2.055E-01
50	4	5.6	6.00E-08	2.4490E-01	1309	4.869E-09	3.485E-03	449	7.871E-09	2.159E-02	81	1.758E-08	2.023E-01
50	5	5.6	6.00E-08	3.0612E-01	4569	2.834E-09	1.967E-03	1028	5.555E-09	1.531E-02	101	1.647E-08	2.008E-01
100	1	11.1	1.20E-07	6.1224E-02	123	3.592E-08	2.034E-02	112	3.765E-08	4.671E-02	101	3.953E-08	1.005E-01
100	2	11.1	1.20E-07	1.2245E-01	179	2.985E-08	2.019E-02	157	3.180E-08	4.801E-02	136	3.414E-08	1.099E-01
100	3	11.1	1.20E-07	1.8367E-01	249	2.556E-08	1.735E-02	209	2.775E-08	4.490E-02	169	3.077E-08	1.170E-01
100	4	11.1	1.20E-07	2.4490E-01	390	2.073E-08	1.327E-02	304	2.332E-08	4.021E-02	218	2.728E-08	1.302E-01
100	5	11.1	1.20E-07	3.0612E-01	607	1.704E-08	9.432E-03	431	1.995E-08	3.340E-02	258	2.542E-08	1.362E-01
1000	1	111.1	1.20E-06	6.1224E-02	1018	3.961E-07	1.990E-02	1009	3.979E-07	4.020E-02	1000	3.997E-07	7.374E-02
1000	2	111.1	1.20E-06	1.2245E-01	1051	3.898E-07	2.814E-02	1042	3.916E-07	5.089E-02	1032	3.934E-07	8.457E-02
1000	3	111.1	1.20E-06	1.8367E-01	1096	3.818E-07	3.328E-02	1086	3.836E-07	5.781E-02	1075	3.854E-07	9.144E-02
1000	4	111.1	1.20E-06	2.4490E-01	1141	3.741E-07	3.353E-02	1130	3.760E-07	5.672E-02	1119	3.778E-07	8.884E-02
1000	5	111.1	1.20E-06	3.0612E-01	1185	3.672E-07	3.575E-02	1173	3.691E-07	5.883E-02	1161	3.709E-07	8.993E-02

4. Real insurance claim data analysis

In this section, we assume two examples of real insurance claim datasets for illustrative purposes. In here, we take $N = 100,000$ with a burn-in period of $20,000$. As it occurs in practice, we assume the values of all parameters of the CP model are unknown.

4.1. Example 1: medical insurance data

In the first example, we select a medical insurance example which is discussed initially by [13]. This dataset is related to an insurance company in Iran that offers medical insurance as one of its services. later, [32] fit the CP outlier model to this example and obtained Bayesian estimation for all parameters. [32] showed that the proposed CP model is a better fit for the medical insurance data.

In order to use the proposed Bayesian approach, we assume the hyperparameter values $a_1 = 0.001, t_1 = 1000, \lambda = 1, \beta^* = 1.1, b_1 = 0.18$ and $b_2 = 3.5$ and calculated the hyperparameter values a_2 and t_2 based on the same methodology which is discussed in Section 3. Table 4 reports the Bayes estimation of Shannon entropy and Fisher information of all parameters under different loss functions. The insurance companies are interested in that there is more uncertainty in terms of claims, particularly in the context of the outlier claim data.

Table 4. Estimation results for the medical insurance data.

	SELF	PLF	DLF
En(X)	291.729	290.623	289.476
$I_{\boldsymbol{X}}(\alpha)$	2.450	2.292	2.145
$I_{\boldsymbol{X}}(\theta)$	7.908E-09	8.176E-09	8.452E-09
$I_{\boldsymbol{X}}(\beta)$	0.207	0.765	2.824

4.2. Example 2: motor insurance data

In the second example, we select a motor insurance example which is discussed initially by Dixit and Jabbari Nooghabi [12]. This dataset is related to an insurance company in Iran that offers motor insurance as one of its services. later, Scollnik [32] fit the CP outlier model to this example and obtained Bayesian estimation for all parameters. Scollnik [32] showed that the proposed CP model is a better fit for the motor insurance data.

In order to use the proposed Bayesian approach, we assume the hyperparameter values $a_1 = 10, t_1 = 1000, \lambda = 1, \beta^* = 1.1, b_1 = 2.17$ and $b_2 = 19.57$ and calculated the hyperparameter values a_2 and t_2 based on the same methodology which is discussed in Section 3. Table 5 reports the Bayes estimation of Shannon entropy and Fisher information of all parameters under different loss functions.

Table 5. Estimation results for the motor insurance data.

	SELF	PLF	DLF
En(X)	357.174	342.060	328.035
$I_{\boldsymbol{X}}(\alpha)$	138.853	92.056	61.031
$I_{\boldsymbol{X}}(\theta)$	1.901E-11	2.333E-11	2.863E-11
$I_{\boldsymbol{X}}(\beta)$	512.230	814.095	1293.852

5. Conclusion

[24] and [17] used the marginal distribution of the famous Dixit model (not the joint distribution of the Dixit model) for obtained the amount of information in the Pareto sample with the presence of outliers that may cause misleading inferential conclusions. In this paper, an appropriate and new method without restriction to the number of outliers and sample size for computing the amount of information such as Shannon entropy and Fisher information has been proposed when outliers exist in a Pareto sample. We focused on the Bayesian approach proposed by [32] based on the contaminated Pareto distribution. We implemented the Gibbs sampler which is a simple and rational method for computing Bayesian estimation of Shannon entropy and Fisher information under different loss functions. Some simulation studies are conducted to showed the performance of the proposed methodology under various sample sizes and the number of outliers. The results of simulation exhibited that that the CP model and its Bayesian method are efficient when outliers are in the sample. Also, the DLF is a preferable loss function for the Bayesian approach based on the CP model. In the end, two examples of real insurance claim data are studied to illustrate the superiority of the CP model in analyzing datasets and computing the amount of Shannon entropy and Fisher information. It may be point out that the proposed approach can be extended for other distributions and obtained the other forms of information measure sach as Renyi entropy, Tsallis entropy and Kullback-Leibler divergence.

Appendix

A.

Proof of Theorem 1. Based on Shannon entropy Eq. (1) and the form of CP model (4), we have

$$\begin{aligned} En(X_i) &= -\int_{\theta\beta^{\delta_i}}^{\infty} \frac{\alpha\theta^{\alpha}\beta^{\alpha\delta_i}}{x_i^{\alpha+1}} \ln\left(\frac{\alpha\theta^{\alpha}\beta^{\alpha\delta_i}}{x_i^{\alpha+1}}\right) dx_i \\ &= -\ln\left(\alpha\theta^{\alpha}\beta^{\alpha\delta_i}\right) + (\alpha+1)\alpha\theta^{\alpha}\beta^{\alpha\delta_i} \int_{\theta\beta^{\delta_i}}^{\infty} \frac{\ln(x_i)}{x_i^{\alpha+1}} dx_i. \end{aligned}$$

Using the method of integral part by part, we have

$$\begin{aligned} En(X_i) &= -\ln\left(\alpha\theta^{\alpha}\beta^{\alpha\delta_i}\right) + (\alpha+1)\Big(\ln\left(\theta\beta^{\delta_i}\right) + \frac{1}{\alpha}\Big) \\ &= \ln\left(\frac{\theta\beta^{\delta_i}}{\alpha}\right) + \frac{\alpha+1}{\alpha} \\ &= \ln\left(\frac{\theta\beta^{\delta_i}}{\alpha} e^{\frac{\alpha+1}{\alpha}}\right). \end{aligned}$$

B.

Proof of Theorem 3. Based on the conditional likelihood (6), we have

$$\ln\Big(L(\boldsymbol{x}|\alpha,\theta,\beta,\boldsymbol{\delta},\epsilon)\Big) \propto n\ln(\alpha) + n\alpha\ln(\theta) + \alpha\sum_{i=1}^{n}\delta_i\ln(\beta) - (\alpha+1)\sum_{i=1}^{n}\ln(x_i).$$

Therefore

$$\frac{\partial^2}{\partial\alpha^2}\ln\Big(L(\boldsymbol{x}|\alpha,\theta,\beta,\boldsymbol{\delta},\epsilon)\Big) = -\frac{n}{\alpha^2},$$
$$\frac{\partial^2}{\partial\theta^2}\ln\Big(L(\boldsymbol{x}|\alpha,\theta,\beta,\boldsymbol{\delta},\epsilon)\Big) = -\frac{n\alpha}{\theta^2},$$
$$\frac{\partial^2}{\partial\beta^2}\ln\Big(L(\boldsymbol{x}|\alpha,\theta,\beta,\boldsymbol{\delta},\epsilon)\Big) = -\frac{\alpha\sum_{i=1}^n\delta_i}{\beta^2}.$$

Using Eq. (2), we have

$$I_{\boldsymbol{X}}(\alpha) = \frac{n}{\alpha^2},$$
$$I_{\boldsymbol{X}}(\theta) = \frac{n\alpha}{\theta^2},$$
$$I_{\boldsymbol{X}}(\beta) = \frac{\alpha\sum_{i=1}^n\delta_i}{\beta^2}.$$

C.

The number of iterations is initially set to r=0, and then, based on the appropriate initial parameter values, the following steps are repeated (N times).

(i) Generate samples for ϵ_{r+1}, α_{r+1}, θ_{r+1}, β_{r+1}, and $\delta_{r+1,i}$, $i=1,2,\ldots,n$ from Eq. (7) to (11), respectively.
(ii) Compute the value of Shannon entropy from Eq. (13).
(iii) Compute the value of Fisher information there exist in $\boldsymbol{X}$ about parameters α,θ and β from Eq. (14) to (16), respectively.

ACKNOWLEDGMENTS

The author is thankful to the editors and the referees for their valuable comments.

REFERENCES

1. A. A. H. Ahmadini, A. S. Hassan, A. N. Zaky, and S. S. Alshqaq, *Bayesian inference of dynamic cumulative residual entropy from Pareto II distribution with application to COVID-19*, AIMS Math, vol. 6, no. 3, pp. 2196–2216, 2020.
2. M. Asadi, N. Ebrahimi, G. G. Hamedani, and E. S. Soofi, *Information measures for Pareto distributions and order statistics*, Advances in distribution theory, order statistics, and inference, pp. 207–223, 2006.
3. S. Baratpour, J. Ahmadi, and N. R. Arghami, *Entropy properties of record statistics*, Statistical Papers, vol. 48, pp. 197–213, 2007.
4. V. Barnett, and T. Lewis, *Outlier in statistical data*, 3rd ed., Chichester: John Wiley and Sons, New York, 1994.
5. G. A. Benktander *note on the most dangerous and skewest class of distribution*, ASTIN Bulletin: The Journal of the IAA, vol. 2, no. 3, pp. 387–390, 1963.
6. L. Boltzmann, *Neitere Studien uber das Warmegleichgewicht unter Gasmolekulen*, K. Akad. Wiss. (Wein) Sitzb, vol. 66, 275, 1872.
7. Y. Cho, H. Sun, and K. Lee, *An estimation of the entropy for a Rayleigh distribution based on doubly-generalized Type-II hybrid censored samples*, Entropy, vol. 16, no. 7, pp. 3655–3669, 2014.
8. R. Clausius, *Abhaudlungen uber die mechanische Warmetheorie Friedrich*, Vieweg, Braunschweig, 1864.
9. T. Cover, and J. Thomas, *Elements of Information Theory*, Wiley, New York. 2006.
10. M. H. DeGroot, *Optimal statistical decisions*, John Wiley and Sons, New Jersey, 2005.
11. U. J. Dixit, *Characterization of the gamma distribution in the presence of k outliers*, In Bulletin Bombay Mathematical Colloquium, vol. 4, pp. 54–59, 1987.
12. U. J. Dixit, and M. J. Nooghabi, *Efficient estimation in the Pareto distribution with the presence of outliers*, Statistical Methodology, vol. 8, no. 4, pp. 340–355, 2011.
13. U. J. Dixit, and M. Jabbari Nooghabi, *Efficient estimation of the parameters of the Pareto distribution in the presence of outliers*, Communications for Statistical Applications and Methods, vol. 18, no. 6, pp. 817–835, 2011.
14. R. A. Fisher, *Theory of statistical estimation*, In Mathematical proceedings of the Cambridge philosophical society, vol. 22, no. 5, pp. 700–725, 1925.

15. A. Gelfand, and A. Smith, *Sampling based approaches to calculating marginal densities*, Technical Report, 415, 1989.
16. A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*, Chapman and Hall/CRC, 2013.
17. A. S. Hassan, E. A. Elsherpieny, and R. E. Mohamed, *Classical and Bayesian estimation of entropy for Pareto distribution in presence of outliers with application*, Sankhya A, vol. 85, no. 1, pp. 707–740, 2023.
18. A. S. Hassan, and A. N. Zaky, *Entropy Bayesian estimation for Lomax distribution based on record*, Thailand Statistician, vol. 19, no.1, pp. 95–114, 2021.
19. Karimi, H.; Nasiri, P. *Estimation parameter of R= P (Y¡ X) for length-biased weighted Lomax distributions in the presence of outliers*, Mathematical and Computational Applications, vol. 23, no. 1, 9, 2018.
20. V. Kumar, and N. Singh, *Characterization of Lifetime Distribution Based on Generalized Interval Entropy*, Statistics, Optimization & Information Computing, vol. 6, no. 4, pp. 547–559., 2018.
21. A. M. Legendre, *New Methods for the Determination of the Orbits of Comets*, Courcier, Paris, 1805.
22. H. Nakata, S. Yasuyuki, and T. Mari, *Entropy characterisation of insurance demand: Theory and evidence*, Research Inst. of Economy, Trade and Industry, 2010.
23. M. J. Nooghabi, and M. Naderi, *Stress–strength reliability inference for the Pareto distribution with outliers*, Journal of Computational and Applied Mathematics, vol. 404, 113911, 2022.
24. M. J. Nooghabi, and E. K. Nooghabi, *On entropy of a Pareto distribution in the presence of outliers*, Communications in Statistics-Theory and Methods, 2016, vol 45, no. 17, pp. 5234–5250, 2016.
25. J. G. Norstrom, *The use of precautionary loss functions in risk analysis*, IEEE Transactions on reliability, vol 45, no. 3, pp. 400–403, 1996.
26. H. A. Noughabi, H. A. Noughabi, and J. Jarrahiferiz, *Informational energy and entropy applied to testing exponentiality*, Statistics, Optimization & Information Computing, 2020, vol 8, no. 1, pp. 220–228, 2020.
27. I. Ntzoufras, *Bayesian modeling using WinBUGS*, John Wiley and Sons, vol. 698, 2011.
28. K. Okhli, and M. J. Nooghabi, *On the three-component mixture of exponential distributions: A Bayesian framework to model data with multiple lower and upper outliers*, Mathematics and computers in simulation, vol 208, pp. 480–500, 2023.
29. K. Okhli, and M. J. Nooghabi, *On the contaminated exponential distribution: A theoretical Bayesian approach for modeling positive-valued insurance claim data with outliers*, Applied Mathematics and Computation, vol 392, 125712, 2021.
30. A. Renyi, *On Measures of Entropy and Information*, Univ. California Press, Berkeley, Calif, vol. 4, pp. 547–561, 1961.
31. M. Salajegheh, M. J. Nooghabi, and K. Okhli, *A Bayesian approach for modeling heavy tailed insurance claim data based on the contaminated lognormal distribution*, METRON, vol. 2025, https://doi.org/10.1007/s40300-025-00288-9
32. D. P. Scollnik, *A Pareto scale-inflated outlier model and its Bayesian analysis*, Scandinavian Actuarial Journal, vol. 2015, no. 3, pp. 201–220, 2015.
33. C. E. Shannon, *A mathematical theory of communication*, The Bell system technical journal, vol. 27, no. 3, pp. 379–423, 1984.
34. E. Sohrabi, and M. Jabbari Nooghabi, *Estimation of the parameters of Weibull distribution and net premium against outliers*, Iranian Journal of Insurance Research, vol. 13, no. 1 pp. 43-–60, 2024.
35. C. Tsallis, *Possible generalization of Boltzmann-Gibbs statistics*, Journal of statistical physics, vol. 52, pp. 479–487, 1988.
36. I. Verdinelli, and L. Wasserman, *Bayesian analysis of outlier problems using the Gibbs sampler*, Statistics and Computing, vol. 1, pp. 105–117, 1991.
37. N. Weiner, *Cybernetics*, Wiley, New York, 1949.
38. V. Zardasht, *Results on relative mean residual life and relative cumulative residual entropy*, Statistics, Optimization & Information Computing, vol. 7, no. 1, pp. 150–159, 2019.