Turkish Journal of Science & Technology
15(2), 71-78, 2020

On the Robust Estimations of Location and Scale Parameters for Least Informative Distributions

Mehmet Niyazi ÇANKAYA[1*]
[1] Department of International Trading and Finance, Applied Science School, Uşak University, Uşak, Turkey
[1]Department of Statistics, Faculty of Arts and Science, Uşak University, Uşak, Turkey
[*1] mehmet.cankaya@usak.edu.tr

(Geliş/Received: 12/02/2020; Kabul/Accepted: 09/05/2020)

Abstract: M-estimation as generalization of maximum likelihood estimation (MLE) method is well-known approach to get the robust estimations of location and scale parameters in objective function ρ especially. Maximum $\log_{\mathrm{q}}$ likelihood estimation (MLqE) method uses different objective function called as $\rho_{\log_{\mathrm{q}}}$. These objective functions are called as M-functions which can be used to fit data set. The least informative distribution (LID) is convex combination of two probability density functions f_0 and f_1. In this study, the location and scale parameters in any objective functions $\rho_{\log}$, $\rho_{\log_{\mathrm{q}}}$ and $\psi_{\log_{\mathrm{q}}}(f_0, f_1)$ which are from MLE, MLqE and LIDs in MLqE are estimated robustly and simultaneously. The probability density functions which are f_0 and f_1 underlying and contamination distributions respectively are chosen from exponential power (EP) distributions, since EP has shape parameter α to fit data efficiently. In order to estimate the location μ and scale σ parameters, Huber M-estimation, MLE of generalized t (Gt) distribution are also used. Finally, we test the fitting performance of objective functions by using a real data set. The numerical results showed that $\psi_{\log_{\mathrm{q}}}(f_0, f_1)$ is more resistance values of estimates for μ and σ when compared with other ρ functions.
Key words: Least informative distributions, maximum $\log_{\mathrm{q}}$ likelihood estimation method, robustness.

En Az Bilgilendirici Dağılımlar için Konum ve Ölçek Parametrelerinin Sağlam Tahminleri Üzerine

Öz: En çok olabilirlik tahmin (MLE) yönteminin genelleştirilmesi olarak M-tahmini, özellikle amaç fonksiyonda ρ konum ve ölçek parametrelerinin sağlam tahminlerini elde etmek için iyi bilinen bir yaklaşımdır. En çok $\log_{\mathrm{q}}$ olabilirlik tahmini (MLqE) yönteminde $\rho_{\log_{\mathrm{q}}}$ adı verilen farklı amaç fonksiyonu kullanır. Bu amaç fonksiyonlarına, veri kümesine uyması için kullanılabilen M-fonksiyonları denir. En az bilgilendirici dağılımlar, f_0 and f_1 olasılık yoğunluk fonksiyonlarının konveks kombinasyonudur. Bu çalışmada, MLqE'de LID, MLqE ve MLE yöntemleri ile herhangi bir amaç fonksiyonda $\rho_{\log}$, $\rho_{\log_{\mathrm{q}}}$ ve $\psi_{\log_{\mathrm{q}}}(f_0, f_1)$ konum ve ölçek parametreleri dayanıklı ve aynı anda tahmin edilmektedir. Dağılımın büyük bir çoğunluğu f_0 ve kontaminasyon dağılımı f_1 olmak üzere üstel kuvvet (EP) dağılımından seçilmektedir; çünkü EP, verilere verimli bir şekilde uyacak şekil parametresi α'ya sahiptir. Konum μ ve ölçek σ parametrelerini tahmin etmek için Huber M-tahmini, genelleştirilmiş t (Gt) dağılımının MLE'si de kullanılmıştır. Böylelikle, gerçek bir veri seti kullanarak amaç fonksiyonların uyum performansını test etmekteyiz. Sayısal sonuçlar aracılığı ile $\psi_{\log_{\mathrm{q}}}(f_0, f_1)$ amaç fonksiyonu diğer ρ fonksiyonları karşılaştırıldığında, μ ve σ için tahminlerin daha fazla direnç değerine sahip olduğu gösterilmiştir.

Anahtar kelimeler: En az bilgilendirici dağılımlar, maksimum $\log_{\mathrm{q}}$ olabilirlik tahmin yöntemi, dayanıklılık.

1. Introduction

In a dataset, it is possible to observe non-identical behavior. In other words, the working principle of a real world cannot be modeled by a hypothetical parametric model. Even if we assume that a dataset is a member of a parametric model with its true parameter values, a contamination into underlying distribution can be observed. In this case, robust estimation of true values of parameters of underlying distribution is a crucial role in the estimation theory. For this purpose, M-estimation method and the recently maximum $\log_{\mathrm{q}}$ likelihood estimation method are commonly used by [1-9].

The deformed logarithms are member of fractional polynomial functions. However, as it is clearly known, the estimation is a process which performs the numerical integration. The numerical integration which is based on Riemann integration rule works on the difference principle [10]. These differences can be represented by functions. If we can accomplish to construct the neighborhood of a function, then we will be capable to make

[*] Corresponding author: mehmet.cankaya@usak.edu.tr. ORCID Number of authors: [1] 0000-0002-2933-857X

different forms of non-identicality in a dataset, which shows us that the functions in the M-estimation principle is not enough reach to model not only versatile forms of non-identicality but also the different values of parametric model which represents the underlying distribution. Thus, we can make a bridge to touch the non-identical case efficiently. M-estimation method is commonly used for the estimation of location and scale parameters, because the functions in this method can be appropriate for modeling datasets which accommodate with the structure of these functions. They are generally used for the estimations of location and scale parameters. The efficiency in the M-estimation method for the estimations of the different values of parameters of underlying distribution will be a deficiency, because the structure of function for the different values of parameters cannot be modeled efficiently. Then, the functions used to estimate some parameters cannot be successful for different values of parameters in the underlying distribution in the M-estimation method that is generalization of maximum likelihood estimation (MLE) [1-9].

M-estimation is used to produce robust estimators for parameters of a probability density (p.d.) function $f_0(x;\boldsymbol{\theta})$. M-estimators are defined through an objective functions minimizing $\rho(x;\boldsymbol{\theta}) = \sum_{i=1}^{n} \Lambda[f_0(x_i;\boldsymbol{\theta})]$ over $\boldsymbol{\theta}$ [1-4]. Here, Λ is a concave function that is capable of making one to one mapping from $f_0(x;\theta)$ to Λ. M-estimators are derived by fixed functions, such as Huber, Tukey, etc. LIDs are used to produce Huber, Tukey, etc [3,11]. MLE as a special case of M-estimation is a method for estimations of parameters in a p.d. function. It is based on logarithm and does not work properly to estimate parameters in a p.d. function efficiently and robustly when data set including outlier(s) are non identically distributed, therefore we will use function $\log_q$ that mimics MLE method [5,11,12]. In our proposition, the benefit of LIDs and a p.d. function in Λ is that one can propose the objective functions from arbitrary p.d. functions to get more precise estimators for parameters in p.d. functions. The more precision can be accomplished by the parameter q and also LID in MLqE method. In the M-estimation principle, there are different functions used to fit datasets. Computational stage in the M-estimation method should be adopted according to the differentiability property of functions such as generalized divergences and deformed logarithms. For this aim, M-estimations are optimized according to parameters by using the Genetic Algorithm (GA). Thus, the local points in the optimization for the objective functions based on M-estimation can be discarded as it is pointed out that GA convergences to global points [13,14].

The aim of this study is to obtain robust estimators of location μ and scale σ parameters for LID by using MLqE and to compare with the fitting performances of objective functions for the robust estimators obtained by MLE for Generalized t (Gt) distribution and Huber M-estimation [3,11]. While performing the robust estimation procedure, we aim to get the efficient estimators as well by using another well-known robust estimation method called as MLqE. We make a comparison among the objective functions or M-functions used in the robust M-estimation method if we add the outliers into real datasets. The main motivation in LID is based on the convex combination of two functions which are f_0 and f_1 representing the underlying and contamination distributions, respectively. We use EP distribution to apply the LID in MLqE, because EP distribution has a shape parameter α which can model the peakedness of function. So, the efficiency of estimators can be performed not only the used shape parameter with fixed values but also f_1 which will be responsible to model the outliers as well [11,15-16].

The organization of this paper is as follow: Section 2 gives the preliminaries about the maximum likelihood estimation method and its generalized forms. The LIDs in MLqE are introduced. Different objective functions which will be used to fit data are given. Section 3 introduces the information criteria (IC). Section 4 is devoted to numerical illustration. Conclusions are given in the last section.

2. Preliminaries

2.1. Maximum likelihood estimation method

As it is-well-known, MLE is asymptotically unbiased with minimum variance of estimators and references therein [2,3]. Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random variables, i.e., $X_i \sim f(x;\boldsymbol{\theta}), i = 1,2,\ldots,n$ shows that $X_1, X_2, \cdots, X_n$ have identical distributions. n represents the number of sample size as a sampling version of a p.d. function f as a population. In this case, the maximum likelihood estimators (MLE) of the parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$

$$L(\boldsymbol{\theta}; x) = \prod_{i=1}^{n} f(x_i;\boldsymbol{\theta}) \tag{1}$$

is obtained by optimizing the likelihood function according to parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$. $x = (x_1, x_2, \ldots, x_n)$ is a vector of observations and the likelihood function is a function of the parameters $\theta_1, \theta_2, \ldots, \theta_p$. When the log function of both sides of the given expression in equation (1) is taken to get a tractable expression, the following expression is obtained as follow:

$$\log\big(L(\boldsymbol{\theta}; x)\big) = \sum_{i=1}^{n} \log\big(f\,(x_i; \boldsymbol{\theta})\big). \tag{2}$$

2.2. Maximum $\log_q$ likelihood estimation method

In the case of contamination, the robust and efficient estimations of the parameters of the f_0 distribution are performed by using the maximum likelihood $\log_q$ estimation (MLqE) method, which is a generalization of the likelihood estimation method and is also included in the M-functions [2,3], and

$$\log_q(L(\boldsymbol{\theta}; x)) \; = \sum_{i=1}^{n} \log_q(f\,(x_i; \boldsymbol{\theta})) \tag{3}$$

is defined. $\log_q(f) = \frac{f^{1-q}-1}{1-q}$, $0 < q < 1$. The $\log_q$ function is called the q-deformed logarithm function. q is a tuning constant used to adjust robustness and efficiency [5,11].

2.3. Least informative distributions based on $\log_q$

Let the random variables $X_1, X_2, \cdots, X_n$ be independent and non-identically distributed. In this case, there is a contaminated distribution and

$$f_\varepsilon(x_i; \boldsymbol{\theta}) = (1-\varepsilon) f_0(x_i; \boldsymbol{\theta}) + \varepsilon f_1(x_i; \boldsymbol{\theta}) \tag{4}$$

is expressed as f_ε which indicates the contaminated distribution and is called as the least informative distribution (LID). The constant ε is the contamination rate. f_0 is the underlying distribution and the estimation values of the parameters of f_0 are tried to be obtained under f_1 which is the contamination into underlying distribution f_0.

The deformed entropies and deformed logarithms derived from these entropies have found many applications in physics and many corresponding fields [22-32]. Let us rewrite the definition of LID given by

$$f_\varepsilon = (1-\varepsilon) f_0 + \varepsilon f_1. \tag{5}$$

LIDs based on the maximum q-log-likelihood was proposed to find the optimal parameters $\boldsymbol{\theta}$ in the function f_ε by [11]. The variational calculus is used to get a case in which the objective function ρ sets out a minimal change with respect to the parameter ε when there exists a contamination into the underlying distribution f_0 by a small amount of outlier distribution f_1, i.e., we set ε in f_ε as a small value close to zero. In order to remove the role of f_1, the function $\rho(f_\varepsilon)$ is derived with respect to ε and set ε to zero. Thus, we can find the estimators of parameters μ and σ by using the following equation

$$\psi_{\log_q}(f_0(x; \mu, \sigma), f_1(x; \mu, \sigma)) = \sum_{i=1}^{n} f_0\,(x_i; \mu, \sigma)^{-q} [f_1(x_i; \mu, \sigma) - f_0(x_i; \mu, \sigma)]. \tag{6}$$

$\psi_{\log_q}(f_0(x; \mu, \sigma), f_1(\mathrm{x}; \mu, \sigma))$ is called as a new objective function to fit data sets. The LIDs in MLqE are used for the estimations of parameters. The distribution f_1 in LIDs can have a role in fitting the data set as a flexible way to drive the efficiency in the estimation [3,11,12].

2.4. M-estimation method

The main idea of M-estimation method is based on the minimizing an objective function ρ with respect to parameters which will be estimated. The M-estimation method is introduced by the following equation:

$$(\hat{\mu},\hat{\sigma}) := \underset{\mu,\sigma}{\arg\min} \sum_{i=1}^{n} \rho(x;\mu,\sigma)) \tag{7}$$

$x = (x_1, x_2, \dots, x_n)$ is a vector of observations. μ and σ are location and scale parameters respectively. After minimizing the mathematical expression in equation (7), the M-estimators of parameters μ and σ are obtained. The analytical tractability of the function ρ is necessary if we get the M-estimators which will be functions of random variables such as arithmetic mean which is well-known location estimator from MLE of location parameter μ in normal distribution [3].

As it is well-known, these two parameters can be added to any arbitrary function. In our case, we consider to add these parameters into objective function ρ. The following subsection is divided for introducing how we give the mathematical expression for the objective functions which will be used to fit the data.

2.5. Objective functions as M-functions and their M-estimations

The objective functions are very popular in robustness literature to fit the data sets. They were proposed by Huber and his coworkers. The objective function ρ notated by Huber and its derived form with respected to parameters is called as the function ψ. ψ is based on the estimating equations originally proposed by Godambe [9]. Throughout this paper, we consider to use the function ψ for representing LIDs in MLqE method. Thus, we have three objective functions in this paper. These are notated by $\rho_{\log}(f_0)$, $\rho_{\log_q}(f_0)$ and $\psi_{\log_q}(f_0, f_1)$.

Let us introduce their analytical expression given below:

$$\rho_{\log}(f_0) = \sum_{i=1}^{n} \log(f_0(x;\mu,\sigma)) \tag{8}$$

is from MLE taken by log function.

$$\rho_{\log_q}(f_0) = \sum_{i=1}^{n} \log_q(f_0(x;\mu,\sigma)) \tag{9}$$

is from MLqE taken by $\log_q$ function.

The following objective function is derived by using LIDs and MLqE method, as introduced by subsection 2.3 [11].

$$\psi_{\log_q}(f_0, f_1) = \sum_{i=1}^{n} f_0(x;\mu,\sigma)^{-q}[f_1(x;\mu,\sigma) - f_0(x;\mu,\sigma)], \tag{10}$$

where f_0 and f_1 are chosen from exponential power (EP) distribution. EP is defined by

$$f_{\mathrm{EP}}(x;\ \mu,\sigma,\alpha) = \frac{\alpha}{2\sigma\Gamma\left(\frac{1}{\alpha}\right)} \exp\{-\left(\frac{|x-\mu|}{\sigma}\right)^{\alpha}\}, \tag{11}$$

where α is shape parameter to arrange the peakedness of function and references therein [15,16].

The Gt distribution is given by

$$f_{\mathrm{Gt}}(x;\ \mu,\sigma,\alpha,\nu) = \frac{\alpha}{2B\left(\frac{1}{\alpha},\nu\right)\nu^{1/\alpha}\,\sigma}(1 + \left(\frac{|x-\mu|}{\nu^{\frac{1}{\alpha}}\sigma}\right)^{\alpha})^{-(\nu+1/\alpha)}, \tag{12}$$

where α and ν are parameters for the shape of peakedness and the tail behaviour of function respectively. If the parameter ν goes to small values, then Gt distribution becomes heavy-tailed distribution. If we use a heavy-tailed distribution to fit data set, then MLEs of the parameters μ and σ will be robust. The estimation of these parameters are not performed due to fact that we need to get the robust estimations of the parameters μ and σ [17] and references therein.

3. Information Criterions for Objective Functions

Information criterion (IC) is a tool to test the fitting performance of functions. Different tools are proposed by [18,19]. After proposing objective functions, we will have another problem to test the fitting performance of $\rho_{\log}$ from MLE, $\rho_{\log_q}$ from MqLE, ψ from Huber M-Estimation and $\psi_{\log_q}(f_0, f_1)$ from LID. For this aim, robust information criterion (RIC) formulae are used to determine the value of tuning parameter q. We can consider the equations (11) and (12) including the function log from objective function in equation (2). Since we use equation (2), the lack of fit part of IC with the penalty term c_k is given by

$$IC(f_0, c_k) = -2\rho_{\log}(f_0) + c_k. \tag{13}$$

IC in equation (13) have two parts that are the lack of fit and the penalty term. The difference between AIC and BIC is due to the penalty term c_k. The correct evaluation of AIC depends on the penalty term $c_k = 2k$, which is a deficiency of AIC [18]. As an alternative to AIC, BIC was proposed when $c_k = \log(\mathrm{n})k$. We propose robust version of ICs by replacing the objective function $\rho_{\log}$ with another objective function $\psi_{\log_q}(f_0, f_1)$. Thus, we have robust versions of ICs which can be reconsidered as the following form for the objective function $\psi_{\log_q}(f_0, f_1)$ from LID

$$RIC_q(f_0, f_1, c_k) = -2\,\psi_{\log_q}(f_0, f_1) + c_k, \tag{14}$$

We make a comparison among LID in MLqE, MLqE, Huber M-estimation and MLE of location and scale parameters of Gt distribution. Note that the value of q as a tuning constant must be taken to be fixed in order to get robust estimators. For example, for a given value of q, the fitting performances of $\psi_{\log_q}(f_0, f_1)$ and $\rho_{\log_q}(f_0)$ are tested until the smallest values of IC is obtained. For the Huber M-estimation and MLE method, the lack of fit part $-2\log\mathrm{L}$ or $-2\log f$ for one sample case, i.e., $\mathrm{n} = 1$ are used, because they are based on the known logarithm function, i.e., log [11,22-32].

Since the LID has shape parameter α of EP distribution, it is logical to expect that the estimates are better than Huber's ρ function. Note that Huber's M-function is based on $\alpha = 2$ for $|y| \le k$ and $\alpha = 1$ for $|y| > k$. In other words, Huber M-estimation is normal distribution in middle, i.e. $|y| \le k$ and Laplace distribution in tail, i.e. $|y| > k$. Since Huber's M-function has the fixed values of shape parameters, there is no flexibility of Huber's M-function. For the determining of values of shape parameter α in EP distribution and tail parameter ν in Gt distribution, IC is used for the objective function $\rho_{\log}$, i.e., $\rho_{\log}(f_{EP})$ and $\rho_{\log}(f_{Gt})$ respectively. Choosing the best values of these parameters α and ν is processed while we are trying their different values of α and ν until the smallest values of IC are obtained. These parameters are also considered as tuning parameters for the sake of conducting the robust estimation procedure. For this reason, they will not be estimated [3,17].

4. Computation and Real Data Application

Optimizing $\psi_{\log_q}(f_0, f_1)$ in equation (10) according to parameters in p.d. functions f_0 and f_1 produces M-estimators $\hat{\mu}$ and $\hat{\sigma}$ from LID

$$(\hat{\mu}_{\psi}, \hat{\sigma}_{\psi}) := \underset{\mu \text{ and } \sigma}{\arg\max}\ \psi_{\log_q}(f_0, f_1) \tag{15}$$

If only f_0 is chosen for $\rho_{\log_q}$, then M-estimators $\hat{\mu}$ and $\hat{\sigma}$ will be obtained from a p.d. function f. Since $\psi_{\log_q}(f_0, f_1)$ and $\rho_{\log_q}$ are nonlinear functions according to the parameters in the p.d. function f, an optimization method as a maximization or minimization is essential to get the estimates of these parameters [2,3,11].

A data set is NCI60 cancer cell line panel. A protein data coded as BR:T-47D from Lysate Array at a website https://discover.nci.nih.gov/cellminer/ was analysed. The parameters μ and σ are estimated by using objective functions to see tendency (location parameter μ) and spread (scale parameter σ) of protein in cancer cell. The maximum and minimum values which are added as two outliers are 12.5160 at positive and -12.5160 at negative sides of the real line, respectively. Thus, the symmetry of data has been kept. The search region in HGA at a module in MATLAB 2013a for the parameters μ and σ are taken as $[-50, 50]$ and $[0, 50]$, respectively. After running GA module in MATLAB 2013a, the estimates of parameters μ and σ are obtained as given in Table 1.

Table 1 has the M-functions from LID in MLqE, i.e., $\psi_{\log_{q=0.008}}(f_0 = EP(\alpha_0 = 2.15), f_1 = EP(\alpha_1 = 1.4))$ from EP distributions with fixed values of shape parameters $\alpha_0 = 2.15$ and $\alpha_0 = 1.4$, the estimates from MLqE for the parameters μ and σ in EP distribution with fixed value of shape parameter $\alpha = 2.51$ and Huber M-function with tuning parameter $k = 1.02$. When the estimates of two parameters are compared, it is observed that the estimate of $\hat{\mu}$ from LID in MLqE as a bold does not change if two outliers are added. The estimate of $\hat{\sigma}$ has a small changing from LID in MLqE. However, Huber M-estimation and MLEs of parameters μ and σ in Gt distribution cannot have a resistance to outliers and the values of estimates of two parameters have changed. As a result, LID in MLqE has robust estimates when it is compared with other objective functions in Table 1.

Table 1. Estimates of parameters μ and σ by using different objective functions without and with two outliers for protein data in cancer cell

M-functions: Objective functions based on $\log_q$	$\hat{\mu}$	$\hat{\sigma}$	$RAIC_q$	$RBIC_q$
$\psi_{\log_{q=0.008}}(f_0 = EP(\alpha_0 = 2.15), f_1 = EP(\alpha_1 = 1.4))$	**1.5752**	1.5585	13.8654	20.0406
Two Outliers	**1.5752**	1.5578	13.8650	20.0647
$\rho_{\log_{q=0.5}}(f_0 = EP(\alpha = 2.51))$	1.3616	**1.3237**	398.7907	404.9659
Two Outliers	1.3617	**1.3237**	406.7909	412.9906
M-functions: Objective functions based on log	$\hat{\mu}$	$\hat{\sigma}$	AIC	BIC
$\psi(f_0 = EP(\alpha = 2), f_1 = EP(\alpha = 1), k = 1.02)$	1.3532	1.3851	657.3162	663.4914
Two Outliers	1.3552	1.4933	695.6350	701.8348
$\rho_{\log}(f_0 = Gt(\alpha = 2.36, \nu = 1.95))$	1.2039	2.0388	677.1910	683.3662
Two Outliers	1.2027	2.0947	706.8218	713.0215

Table 1 also shows the information criteria such as Akaike and Bayesian. $RAIC_q$ and $RBIC_q$ are abbrevations for the robust Akaike and robust Bayesian criteria in which the lack of fit part is based on q in $\log_q$[11]. They are based on q in $\log_q$. The values of $RAIC_q$ and $RBIC_q$ in two cases which are without outliers and two included outliers can be very near to each other when the objective function is $\psi_{\log_{q=0.008}}(f_0 = EP(\alpha_0 = 2.15), f_1 = EP(\alpha_1 = 1.4))$. However, when the objective function is $\rho_{\log_{q=0.5}}(f_0 = EP(\alpha = 2.51))$, the values of $RAIC_q$ and $RBIC_q$ has been changed if there exist two outliers. As a result, the LID in MLqE can have robust information criteria as well due to the fact that the estimates obtained by objective functions based on $\log_q$ could not change more when we compare the values of AIC and BIC in Table 1.

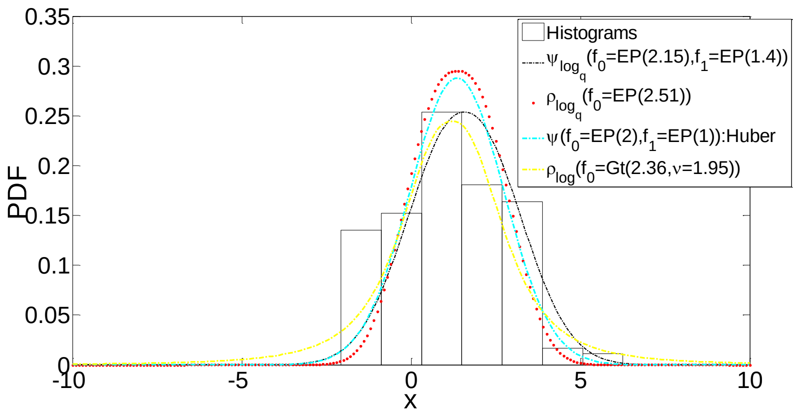

Figure 1. Robust M-estimators and MLE of parameters in f_0 and LIDs if outliers do not exist.

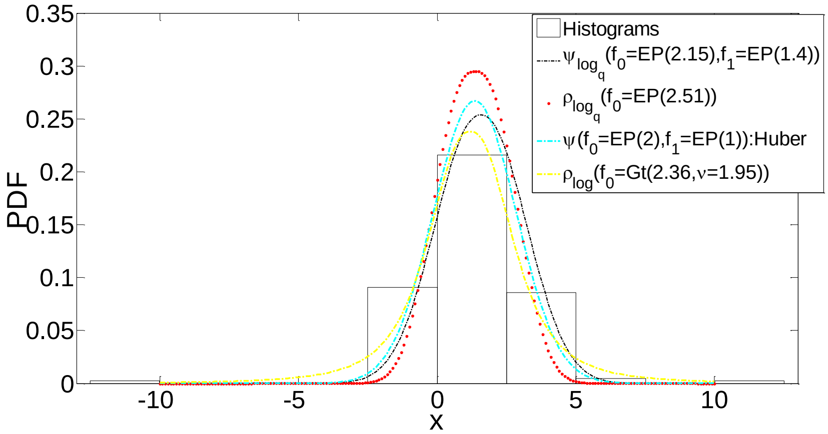

Figure 2. Robust M-estimators and MLE of parameters in f_0 and LIDs if there are two outliers.

Figures 1 and 2 are given for illustrative purpose. Four objective functions depicted by Figures 1 and 2 are used to estimate the parameters μ and σ. The p.d. functions abbreviated as PDF superimposed onto histograms are given by EP distributions with their corresponding fixed shape parameter α. Since the main aim was to plot the underlying distribution, we use the parameter α_0 from p.d. function f_0. There is also p.d. function of Gt distribution which has been used for plotting in Figures 1 and 2. After plotting the p.d. functions of EP with the estimates of M-estimators for the parameters μ and σ and the fixed values of shape and tail parameters of corresponding distribution, we can observe that $f_0 = EP(x; \hat{\mu} = 1.5752, \hat{\sigma} = 1.5585, \alpha_0 = 2.15)$ has a good competance on fitting the data set well when it is compared by the others that are $f_0 = EP(x; \hat{\mu} = 1.3616, \hat{\sigma} = 1.3237, \alpha_0 = 2.51)$, $f_0 = EP(x; \hat{\mu} = 1.3532, \hat{\sigma} = 1.3851, \alpha_0 = 2)$ and $f_0 = Gt(x; \hat{\mu} = 1.2039, \hat{\sigma} = 2.0388, \alpha_0 = 2.36, \nu = 1.95)$.

5. Conclusion

LIDs are obtained by using convex combinations of two p.d. functions for the estimations of location and scale parameters. Thus, the robust estimations of them have been performed by using the M-functions originally based on the M-estimation method. LIDs in MLqE, MLqE, M-estimation of Huber for the robust estimations of the parameters μ and σ and also MLEs the parameters μ and σ of Gt distribution have been compared if there exist two outliers into data set. So the resistance of these M-functions while conducting the robustness procedure has been observed. The role of MLqE method for LIDs have been examined when the parameters μ and σ are estimated simultaneously. The results showed the LID in MLqE can have robust to outliers not only for the estimates of two parameters but also information criteria. However, $\rho_{\log_q}$ with only one p.d. function f, Huber M-estimation and MLEs of the parameters μ and σ of Gt distribution with fixed values shape parameter α and the tail parameter ν determined by using the information criteria gave the estimates that cannot be resistance to outliers even if they are the M-functions used to get robust M-estimates of parameters. Since LIDs have two distributions f_0 and f_1 at the same time, the efficiency was observed for case in which two outliers are added into a real data set. The robustness properties of the objective function $\psi_{\log_q}(f_0, f_1)$ for all parameters in a p.d. function f will be a study in the future.

References

[1] Ni XS, Huo X. Another look at Huber's estimator: A new minimax estimator in regression with stochastically bounded noise. Journal of statistical planning and inference. 2009; 139(2):503-515.

[2] Huber PJ. Robust estimation of a location parameter. In Breakthroughs in statistics. Springer. New York, NY. 1992: 492-518.

[3] Huber PJ, Ronchetti EM. Robust statistics. John Wiley & Sons. New York. 1981.

[4] Shevlyakov G, Morgenthaler S, Shurygin A. Redescending M-estimators. Journal of Statistical Planning and Inference. 2008; 138(10): 2906-2917.
[5] Ferrari D, Yang Y. Maximum Lq-likelihood estimation. The Annals of Statistics. 2010; 38(2): 753-783.
[6] Giuzio M, Ferrari D, Paterlini S. Sparse and robust normal and t-portfolios by penalized Lq-likelihood minimization. European Journal of Operational Research. 2016; 250(1): 251-261.
[7] Andrews DF, Hampel FR. Robust estimates of location: Survey and advances. Princeton University Press. 2015
[8] Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA. Robust statistics: the approach based on influence functions. John Wiley & Sons. 2011. Vol. 196.
[9] Godambe VP. An optimum property of regular maximum likelihood estimation. The Annals of Mathematical Statistics,. 1960; 31(4):1208-1211.
[10] Malik SC, Arora S. Mathematical analysis. New Age International. 1992.
[11] Çankaya MN, Korbel J. Least informative distributions in maximum q-log-likelihood estimation. Physica A: Statistical Mechanics and its Applications. 2018; 509: 140-150.
[12] Gelfand I, Fomin S. Calculus of Variations. Prentice-Hall Inc. Englewood Cliffs. NJ.1963.
[13] Örkcü HH, Özsoy VS, Aksoy E, Dogan MI. Estimating the parameters of 3-p Weibull distribution using particle swarm optimization: A comprehensive experimental comparison. Applied Mathematics and Computation. 2015; 268: 201-226.
[14] Yalçınkaya A, Şenoğlu B, Yolcu U. Maximum likelihood estimation for the parameters of skew normal distribution using genetic algorithm. Swarm and Evolutionary Computation. 2018; 38:127-138.
[15] Çankaya MN, Bulut YM, Doğru FZ, Arslan O. A bimodal extension of the generalized gamma distribution. Revista Colombiana de Estadistica. 2015; 38(2): 353-370.
[16] Çankaya MN. Asymmetric bimodal exponential power distribution on the real line. Entropy. 2018; 20(1): 1-23.
[17] Arslan O, Genç Aİ. The skew generalized t distribution as the scale mixture of a skew exponential power distribution and its applications in robust estimation. Statistics. 2009; 43(5): 481-498.
[18] Bozdogan H. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika. 1987; 52(3):345-370.
[19] Ronchetti E. Robustness aspects of model choice. Statistica Sinica. 1997: 327-338.
[20] Çankaya MN, Yalçınkaya A, Altındağ Ö, Arslan O. On the robustness of an epsilon skew extension for Burr III distribution on the real line. Computational Statistics. 2019; 34(3): 1247-1273.
[21] Çankaya MN, Arslan O. On the robustness properties for maximum likelihood estimators of parameters in exponential power and generalized T distributions. Communications in Statistics-Theory and Methods. 2020; 49(3): 607-630.
[22] Tsallis C. Possible generalization of Boltzmann-Gibbs statistics. Journal of statistical physics. 1988; 52(1-2): 479-487.
[23] Machado JT. Fractional order generalized information. Entropy. 2014; 16(4): 2350-2361.
[24] Suyari H. Mathematical structures derived from the q-multinomial coefficient in Tsallis statistics. Physica A: Statistical Mechanics and its Applications. 2006; 368(1): 63-82.
[25] Hadjiagapiou IA. The random field Ising model with an asymmetric and anisotropic bimodal probability distribution. Physica A: Statistical Mechanics and its Applications. 2011; 390(20): 3204-3215.
[26] Hadjiagapiou IA. (2012). The random field Ising model with an asymmetric and anisotropic trimodal probability distribution. Physica A: Statistical Mechanics and its Applications, 391(13), 3541-3555.
[27] Çankaya MN, Korbel J. On statistical properties of Jizba–Arimitsu hybrid entropy. Physica A: Statistical Mechanics and its Applications. 2017; 475: 1-10.
[28] Korbel J. Rescaling the nonadditivity parameter in Tsallis thermostatistics. Physics Letters A. 2017; 381(32): 2588-2592.
[29] Jizba P, Korbel J, Zatloukal V. Tsallis thermostatics as a statistical physics of random chains. Physical Review E. 2017; 95(2): 022103.
[30] Elze HT. Introduction: Quantum Theory and Beneath?. In Decoherence and Entropy in Complex Systems. Springer. Berlin, Heidelberg. 2004: 119-124.
[31] Jizba P, Korbel J. On q-non-extensive statistics with non-Tsallisian entropy. Physica A: Statistical Mechanics and its Applications. 2016; 444: 808-827.
[32] Csaki, F. Second international symposium on information theory. Académiai Kiadó, Budapest, 1981.