M-Estimations of Shape and Scale Parameters by Order Statistics in Least Informative Distributions on q-deformed logarithm

Mehmet Niyazi ÇANKAYA[1*]

ABSTRACT: The maximum $\log_q$ likelihood estimation (MLqE) method is used to estimate robustly parameters recently. In robust estimation method, the least informative distribution (LID) proposed by Huber is a convex combination of two probability density functions f_0 and f_1. In this study, the recently proposed least informative distributions (LIDs) in MLqE are used to estimate parameters. This paper also studies on the objective functions proposed by maximum $\log_q$-likelihood principle (MLqE) originally derived by $\log_q$-likelihood. The role and capability of order statistics in LIDs in MLqE are examined while getting the estimates of shape and scale parameters. The distance measure for evaluation of fitting performance is given to choose a value for the parameter q in $\log_q$ when the objective functions derived from MLqE are used. The simulation and real data application are given. Thus, we compare the fitting performance of objective functions constructed by MLE on log, MLqE on $\log_q$ and LIDs with order statistics in MLqE. We observed that order statistic chosen for density f_1 in LID in MLqE has a new objective function to fit the data sets. In the simulation, we make two contaminations into artificial data sets. The first contamination is inliers derived by order statistics and the second one is outliers. Thus, we observe that the new objective function can give satisfactory results.

Keywords: Distribution, efficiency, estimation, order statistics, robustness.

[1] Mehmet Niyazi ÇANKAYA(**Orcid ID: 0000-0002-2933-857X**), Uşak Üniversitesi, Uygulamalı Bilimler Yüksekokulu, Uluslararası Ticaret Bölümü, 1. Eylül Kampüsü, Uşak, Türkiye

*Sorumlu Yazar / Corresponding Author: Mehmet Niyazi ÇANKAYA, mehmet.cankaya@usak.edu.tr

Mehmet Niyazi ÇANKAYA	10(3): 1984-1996, 2020

M-Estimations of Shape and Scale Parameters by Order Statistics in Least Informative Distributions on q-deformed logarithm

INTRODUCTION

M-estimation is a very general method based on likelihood estimation method to estimate the model parameters of underlying distribution. Maximum likelihood estimation (MLE) method can have logarithm (log) function to proceed the analytical tractability while getting the estimators of parameters. In MLE, log function can be replaced by $\log_q$ so we have a generalized form of log. The parameter q is responsible to have different forms of log. The robust estimation based on the objective function is originally proposed by (Huber, 1964). The robust estimation is a generalization of maximum log-likelihood estimation method. In the robust method, the objective functions which can give better fitting than probability density function $f(x; \boldsymbol{\theta})$ are used for modeling. The generalized likelihood estimation method works mainly a mapping form of a p.d. function $f(x; \boldsymbol{\theta})$, i.e., $\rho(x; \boldsymbol{\theta}) = \Lambda\big(f(x; \boldsymbol{\theta})\big)$. If Λ is chosen as log function, then log-likelihood principle (LP) works. One can choose $\log_q$ function for Λ. In this case, LP still works but the different forms of p.d. function $f(x; \boldsymbol{\theta})$ is obtained due to the properties of the function Λ. According to the chosen function Λ, ρ can be a member of escort distribution (Jizba, 2004). The origin of objective functions is mainly based on the escort distributions which do not have the normalizing constant that makes a function defined on the interval [0,1]. LIDs are a member of the escort distributions as a rich family. Different examples of objective functions can be read from (Huber, 1981; Shevlyakov et al., 2008; Ni and Huo, 2009; Hampel et al. 2011; Andrews and Hampel, 2015) and references therein.

The objective function which will be used to model the data set should adjust the efficiency. In this direction, we will propose objective functions which use the role of order statistics (OS) and the parameter q. OS is responsible to fit data set by means of many p.d. functions. In our case, the objective function has a p.d. function from OS but it also has the properties of underlying distribution. Our case is mainly based on LID proposed by (Huber, 1964). The generalized form of LIDs for arbitrary p.d. functions has recently been studied by (Çankaya and Korbel, 2018). Note that M-functions from (Huber, 1981; Ni and Huo, 2009; Hampel et al. 2011; Andrews and Hampel, 2015) are mainly strict functions and they are not flexible to get the different forms of functions. They have strict tuning constant which can only conduct some part of function instead of scanning the function completely. For example, Huber M-objective function is a strict function and it is normal in the middle and Laplace at the tails on the real line (Huber, 1964). In our study, we can have a neighborhood of a function by means of parameter q. Further, LIDs based on $\log_q$ are capable to fit data set efficiently as well because LIDs in $\log_q$ have two p.d. functions f_0 and f_1 which work with the parameter q which is responsible to have different forms of functions. Note that using parameter q is advantage to manage efficiency and robustness (Çankaya and Korbel, 2018).

The main aim of this study is to use the objective functions based on $\log_q$. In this perspective, we apply the order statistics (OS) to LIDs in MLqE to test the role and capability of OS for the estimations of parameters. Thus, we also test the performance of OS at LID in MLqE. Further, OS can give an advantange to manage efficiency and so the data sets can be fitted well. We can also manage robustness via parameter q in $\log_q$. Since LIDs in MLqE have the underlying distribution f_0 and the contaminated distribution f_1, we consider to use order statistics for f_1 in order to manage the robustness and efficiency at the same time. The estimators are produced by use of the objective functions from M-functions. Since the estimators are produced by use of likelihood type estimation method, they can be called as M-estimators. The numerical values of M-estimators are replaced by parameters of the underlying distribution F_0 which is cumulative distribution function (c.d. function). Thus, the probability value (p-

value) of Kolmogorov-Smirnov (KS) test statistic for the underlying distribution F_0 is computed (Huber-Carol et al., 2012). We can observe the role of p-values of KS in order to adjust the tuning constant q. This can be an alternative approach if the integrals in information geometry (Çankaya and Korbel, 2018; Bozdogan, 1987, Csaki, 1981; Çankaya and Korbel, 2017) are computational extensive and cannot be convergence to finite value, which is an another approach given by this paper when compared with OS used for f_1 in LID.

The remainder of this study is as follows: Materials and methods give the tools used to estimate the parameters in a p.d. function $f(x_i; \boldsymbol{\theta})$. We introduce the well-known OS and the popular Weibull distribution. We express the goodness of fit test how to test the fitting competence. The optimization as important stage at the computational evaluation of objective functions is carried out by genetic algorithm. Simulation is provided to show how the procedure OS in LID works especially. The real data application is provided. Thus, we observe the cooperative cross-cheking between the p-values of KS test statistic and illustrative figures for the evaluation of fitting performance. The last section is divided for the conclusion.

MATERIALS AND METHODS

Maximum log and $\log_q$ likelihood estimation methods

The well-known maximum likelihood estimation (MLE) method is given by

$$L(\boldsymbol{\theta}, x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}), \tag{1}$$

where x_i, $i = 1, 2, \ldots, n$ are observations and n is a sample size drawn from a hypothetical parametric model f. In the MLE method, the main working principle is based on the p.d. function $f(x_i; \boldsymbol{\theta})$, because the parameters $\boldsymbol{\theta}$ are estimated via using the likelihood function L.

In order to get analytically simple expressions, we can take the log function of both sides of the given expression in equation (1). Thus, we have the following expression given by

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

The maximum $\log_q$ likelihood estimation method (Ferrari and Yang, 2010; Giuzio et al., 2016) is a generalization of the log-likelihood estimation method in the M-functions (Huber, 1964; Hampel et al. 2011). The maximum $\log_q$ likelihood estimation (MLqE) method is given by

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

where $q \in \mathbb{R}\backslash\{1\}$. $\log_q(f) = \frac{f^{1-q}-1}{1-q}$ is derived from Tsallis q-entropy (Tsallis, 1988; Elze, 2004). In MLqE method, let us introduce a part again as the following form:

$$\rho_{\log_q}(x; \boldsymbol{\theta}) = \log_q(f_0(x; \boldsymbol{\theta})), \tag{4}$$

which is an objective function based on $\log_q$. If $q \to 1$, $\rho_{\log_q}$ drops to $\rho_{\log}(x; \boldsymbol{\theta}) = \log(f_0(x; \boldsymbol{\theta}))$ is a part from MLE method. $\log_q$ is a strictly monotonic function (Huber, 1964; Çankaya and Korbel, 2018).

Since M-estimation method in M-function is a generalization of ML estimation method, we can consider to use M-functions to estimate the parameters $\boldsymbol{\theta}$ of function f simultaneously as well. M-functions in (Hampel et al. 2011) are strict/rigid functions and they do not have a rich family when compared by MLqE method, because MLqE can be applied for any p.d. function $f(x_i; \boldsymbol{\theta})$. Further, it is noted that $f(x_i; \boldsymbol{\theta})$ can be an escort distribution or p.d. function. The working principle of modeling is based on the objective functions. These are mainly escort distributions and it is not essential to have a normalizing constant in an escort distribution (Jizba, 2004; Jizba and Korbel, 2016; Çankaya and Korbel; 2017).

Order statistics

Let X be a random variable whose distribution function (d.f.) and probability density function (p.d. function) are $F(x; \boldsymbol{\theta})$ and $f(x; \boldsymbol{\theta})$ respectively. Let $X_1, X_2, \dots, X_n$ have independent and identical d.f. $F(x; \boldsymbol{\theta})$ and p.d. function $f(x; \boldsymbol{\theta})$. $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ denote the order statistics (OS) of these random variables. For $1 \leq r \leq n$, the p.d. function of $X_{r:n}$ is given by

$$f_{r:n}(x; \boldsymbol{\theta}) = \frac{n!}{(r-1)!(n-r)!} [F(x; \boldsymbol{\theta})]^{r-1} [1 - F(x; \boldsymbol{\theta})]^{n-r} f(x; \boldsymbol{\theta}), n, r \in \mathbb{Z}^+, \quad (5)$$

factorials are replaced by $n! = \Gamma(n + 1), (r - 1)! = \Gamma(r)$ and $(n - r)! = \Gamma(n - r + 1), n, r > 0$ (Arnold, 1992; Prudnikov, 1986).

Least informative distributions

The least informative distribution (LID) is given by

$$f_\varepsilon(x; \boldsymbol{\theta}) = (1 - \varepsilon) f_0(x; \boldsymbol{\theta}) + \varepsilon f_1(x; \boldsymbol{\theta}). \tag{6}$$

f_ε indicates the contaminated distribution. The constant ε is the contamination rate. f_0 is the underlying distribution and f_1 is the contamination into underlying distribution f_0 and so f_1 is a contamination distribution (Huber, 1964; Huber, 1981).

Least informative distributions based on $\log_q$

The following objective function is derived by using LIDs and MLqE method

$$\rho_{\log_q}\big(f_0(x; \boldsymbol{\theta}), f_1(x; \boldsymbol{\theta})\big) = f_0(x; \boldsymbol{\theta})^{-q}(f_1(x; \boldsymbol{\theta}) - f_0(x; \boldsymbol{\theta})), \tag{7}$$

where f_0 and f_1 are p.d. function. If f_1 is zero, then equation (7) is $-f_0^{1-q}$. The main part in $\log_q$ can be obtained. Thus, we can see the role of f_1. Since f_1 exists in the LIDs in MLqE, the objective function can have a high fitting capability on a data set. It is possible to apply different distributions into LIDs in MLqE (Godambe, 1960). Note that LIDs in MLqE have a big framework when one wants to use it for modeling (Gelfand and Fomin, 1963; Suyari, 2006; Çankaya and Korbel, 2018; Huber, 1964; Huber, 1981). Note that when OS form in equation (5) for a p.d. function $f(x; \boldsymbol{\theta})$ is chosen for $f_1(x; \boldsymbol{\theta})$ at LID in MLqE, we can have a new objective functions to fit data set efficiently. In the working principle of

Mehmet Niyazi ÇANKAYA   **10(3): 1984-1996, 2020**

M-Estimations of Shape and Scale Parameters by Order Statistics in Least Informative Distributions on q-deformed logarithm

LIDs including any f_0 and f_1 at the same time, it is not essential to have a strict/rigid function which can be used to fit data set.

Weibull distribution

Weibull distribution is flexible and has main parameters which are shape and scale in a parametric model. The c.d. function of the Weibull distribution is given by

$$F(x; a, b) = 1 - \exp\left\{-\left(\frac{x}{b}\right)^a\right\}, x > 0; a, b > 0. \tag{8}$$

The p.d. function $f(x) = \frac{d}{dx}F(x)$ is given by

$$f(x, a, b) = \frac{a}{b}\left(\frac{x}{b}\right)^{a-1}\exp\left\{-\left(\frac{x}{b}\right)^a\right\}, x > 0; a, b > 0, \tag{9}$$

a and b are the shape and scale parameters respectively, it is represented by $W(a, b)$ (Rinne, 2008). Since Weibull distribution is a member of exponential family and also Weibull has a kernel $\exp\left(-\left(\frac{x}{b}\right)^a\right)$, finiteness of integral is satisfied (Malik, 1992; Prudnikov et al., 1986; Çankaya, 2018). Thus, we can use $W(a, b)$ for modeling. We will use $W(a, b)$ in our numerical experiment.

M-estimation method based on the objective functions from M-functions

In the M-estimation method, the objective function ρ is minimized according to parameters $\boldsymbol{\theta}$. M-estimators $\widehat{\boldsymbol{\theta}}$ from M-functions ρ are produced by

$$\widehat{\boldsymbol{\theta}} := \sum_{i=1}^{n} \rho(x_i; \boldsymbol{\theta}) \atop \arg\min \boldsymbol{\theta} \tag{10}$$

$x = (x_1, x_2, \ldots, x_n)$ is a vector of observations (Huber, 1964; Huber, 1981).

Goodness of fit test: Distance measures for evaluation of fitting performance

Since the main nature of distribution structure of a real data set is not known exactly, the different objective functions which can be alternative or accommodate to each other should be used for analyzing a real data set. The objective functions will be used to estimate the parameters of p.d. function $f(x; \boldsymbol{\theta})$. Let us remind that LIDs in $\log_q$ have underlying and contamination distributions. Researcher should keep the characteristic of underlying p.d. distribution f_U or c.d. distribution F_U, because we prefer to get robust and efficient M-estimates from M-functions which can be insensitive to contamination into underlying distribution. Since the main characteristic of a real data set can be expressed by means of underlying distribution, we can use the cumulative distribution (c.d.) functions for any statistical distributions to detect the performance of robustness and efficiency. As it is expected, we can use this approach to evaluate the fitting performance if M-estimates from M-functions are plugged into to c.d. function $F_U(x; \widehat{\boldsymbol{\theta}}_M)$. $\widehat{\boldsymbol{\theta}}_M$ are M-estimators produced by objective or M-functions. $\widehat{\boldsymbol{\theta}}_M := \widehat{\boldsymbol{\theta}}_{M_0}$ are values called as M-estimates from M-estimators $\widehat{\boldsymbol{\theta}}_M$. Note that the information criterions (ICs) based on log and $\log_q$ such as Akaike and Bayesian ICs should not be used, because log and $\log_q$ are based on the different sense and they are not comparable to determine the fitting performance of the objective functions. For this reason, we prefer to use distance measure evaluated from underlying distributions. Since we use the c.d. functions, c.d. functions for Weibull, normal, etc. are defined on the closed interval $[0,1]$, that is, they are at a same base for comparison among M-estimates plugged into $F_U(x; \widehat{\boldsymbol{\theta}}_{M_0})$. $\widehat{\boldsymbol{\theta}}_{M_0}$

Mehmet Niyazi ÇANKAYA	10(3): 1984-1996, 2020

M-Estimations of Shape and Scale Parameters by Order Statistics in Least Informative Distributions on q-deformed logarithm

represent the estimates which are numerical values obtained after the optimization of objective functions in equations (2), (4) and (7) according to parameters $\boldsymbol{\theta}$ is performed.

$$KS = \max_{1 \le i \le n} \left| F_U(x_{(i)}; \widehat{\boldsymbol{\theta}}_M) - \frac{i}{n+1} \right| \tag{11}$$

is Kolmogorov-Smirnov (KS) test statistic. KS test statistic gets the maximum value of differences evaluated by absolute distance between $F_U(x_{(i)}; \widehat{\boldsymbol{\theta}}_M)$ and $\frac{i}{n+1}$. The biggest value of p-value of KS test statistic indicates that the best fitting is accomplished by underlying distribution function $F_U(x_{(i)}; \widehat{\boldsymbol{\theta}}_{M_0})$. $x_{(i)}$ represents the ordered form from smallest to highest values for variable x. i indicates a number for the order in the sample size n, i.e., $i = 1, 2, \ldots, n$ (Huber-Carol et al. 2012; Huber, 1964; Huber, 1981). The different values for q are tried until the highest p-value of KS test statistic is obtained. f_0 is assumed to be a hypothetical model which can be real nature of observations in a data set. Note that using f_0 is an idea proposed by Huber's variance covariance matrix (Huber, 1964; Huber, 1981). This is why we prefer to use KS statistic as an alternative way to Huber's variance covariance matrix in information geometry. Thus, computational error for evaluating the integral of Fisher information (FI) in information geometry (Çankaya and Korbel, 2018) can be skipped. We choose OS which has an extensive analytical expression in expected value of FI. Since main function is $W(a, b)$, we can use c.d. function $F_U(x_{(i)}; \widehat{\boldsymbol{\theta}}_{M_0})$ of $W(a, b)$ to get p-value of KS test statistic. Using KS is simple and effective way to determine value of parameter q.

Numerical experiment: Simulation and real data applications

Knowing the design of the artificial random numbers from a p.d. function $f(x; \boldsymbol{\theta})$ is an important indicator to observe the fitting performance of the proposed objective functions. OS into LIDs in MLqE is used. Artificial data sets and real data set which has non-known behavior are used. We apply objective functions derived by equations (2), (4) and (7) to estimate te shape and scale parameters. Since we use order statistic for p.d. function f_1 in the LID in MLqE, we add some ordered artificial data set from the same artificial data set. Inliers and outliers are derived by items in order:

1. **xxx**=randWei(n,au,bu); 2. **xx**=sort(**xxx**);for j=1:nc xc(j)=**xx**(ceil(n*r)); xc(j)=2*xc(j); end 4. **x**=[**xx**(1:nu) **xc**];

'nc' and 'xc' represent the number of sample size and artificial data set for contamination, respectively. **x** is a vector of artificial numbers. **x** has contamination from some ordered data set for **xxx**. r is a rate of order. r is taken as 1 which shows the biggest last value of **xx** to generate outliers.

'randWei' is a written function in MATLAB2013a. $F(x) = y$, y is uniformly distributed random variable defined on interval [0,1]. The artificial numbers for variable x of F are generated by $x = F^{-1}(u)$ (Huber-Carol et al., 2012). Thus, we have artificial numbers for distribution F with 'au' and 'bu' which show values of parameters for underlying distribution f_0. We use $W(a, b)$ with 'au' and 'bu'. Since Weibull is flexible and the popular distribution for modeling a data set and is finite for $x \to \infty$ (Malik, 1992), it can be used for fitting. We can also observe the performance of Weibull when it is applied to LID in MLqE and MLqE.

Optimization of M-functions

MLE, MLqE and LID in MLqE have their corresponding M-functions with parameters $\boldsymbol{\theta}$. Optimization is a process to find the global point of function $G(x; \boldsymbol{\theta})$ minimized according to parameters $\boldsymbol{\theta}$. Since we have data from function G, that is, x is a variable which can have different values which are regarded as observations, we have a sampling form of function G. Thus, $\boldsymbol{\theta}$ will be estimated by

canditate estimators $\widehat{\boldsymbol{\theta}}$ from $\sum_{i=1}^{n} \rho(x_i; \boldsymbol{\theta})$ (Huber, 1964; Huber, 1981). Since we use $\log_q$, optimization is an important stage to convergence global point of function G. Thus, we will have estimators which can obtained by optimization at the global point of function G (Örkcü et al., 2015). For optimization of function G, we use genetic algorithm (GA) which is produced by evaluationary framework. Even if a function is non-smooth, GA can find the global point in function G. We use GA to avoid convergence to local points in G. Further, GA has nice property which is that search region for values of parameters is used instead of using initial point while performing optimization according to parameters $\boldsymbol{\theta}$. HGA at a module in MATLAB 2013a is used. Optimization is performed by 'ga' function. 'CrossoverFcn' and 'crossoversinglepoint' are selected to have less numerical computation error. The search space for values of parameter of p.d. function is necessary for input in 'ga'. The search space of $\rho(x; \boldsymbol{\theta}), \boldsymbol{\theta} = (a, b)$, optimized according to the parameters a and b is the closed range $[0, 10^{10}]$ for both parameters.

RESULTS AND DISCUSSION

The results of simulation for the estimations of shape and scale parameters

Tables 1-2 and Table 3 give inliers and outliers case for values of parameters, respectively. In Tables 1-3, $\hat{\theta}$, $\widehat{Var}(\widehat{\boldsymbol{\theta}})$ and $\widehat{MSE}(\widehat{\boldsymbol{\theta}})$ represent estimates, simulation variance and mean squared error of M-estimators, respectively. $Norm(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = d_1^2 + d_2^2 + \cdots + d_p^2$. $\boldsymbol{D} = \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}, \boldsymbol{D} = (d_1, d_2, \ldots, d_p)$ (Huber-Carol et al., 2012). We have different sample sizes $n_1 = 30, n_2 = 50$ and $n_3 = 100$. The number of replication for each sample size is 10^4. The contamination rate ε is 0.1. Inliers and outliers were explained by numerical experiment section. The objective functions $\rho_{\log}(x; \boldsymbol{\theta})$, $\rho_{\log_q}(x; \boldsymbol{\theta})$ and $\rho_{\log_q}(f_0(x; \boldsymbol{\theta}), f_1(x; \boldsymbol{\theta}))$ are used to estimate a and b. To determine the values of tuning constant q for robustness, our tryings are performed until the smallest values of $\widehat{MSE}(\widehat{\boldsymbol{\theta}})$ are obtained. As it is logically expected, there is only one objective function which fits data well for the given values of q in $\log_q$ and r in OS. For the simulation, the known value of r is already used for p.d. function of OS.

For inliers in Tables 1-2, we observed that the estimates of $\hat{a}$ and $\hat{b}$ from LIDOSq are better than that of MLqE. When sample sizes are increased, $\widehat{MSE}(\widehat{\boldsymbol{\theta}})$ values get smaller values. In outlier case, there is a small situation which cannot be satisfied at a small computational error. It is also noted that when the sample size gets larger, size of contamination is increased, so this can be reason of small increasing in the sample size from $n = 50$ to $n = 100$ at $a = 2$ and $b = 10$ in Table 3. For the outlier cases from Table 3, we observed that the estimates of $\hat{a}$ from LIDOSq and the estimates of $\hat{b}$ from MLqE have small $\widehat{MSE}$ values, which shows us that the shape parameter a can be predicted well if the objective function ρ from LIDOSq is used. In the same comment, MLqE should be used to get small $\widehat{MSE}$ values for the estimates of scale parameter b.

Table 1. Inliers case: $\hat{a}$ and $\hat{b}$ from ρ based on MLE, MLqE and LIDOSq when $a = 3$ and $b = 5$

r	n	$\boldsymbol{\theta}$	MLE		MLqE		LIDOSq	
					$q = 1.67$		$q = 1.24$	
			$a = 3$	$b = 5$	$a = 3$	$b = 5$	$a = 3$	$b = 5$
		$\hat{\boldsymbol{\theta}}$	3.6194	4.5253	3.0575	4.5599	2.9910	4.6486
0.25	30	$\widehat{Var}(\hat{\boldsymbol{\theta}})$	0.3881	0.1017	0.2588	0.0981	0.1890	0.0974
		$\widehat{MSE}(\hat{\boldsymbol{\theta}})$	0.7718	0.3270	0.2621	0.2918	0.1891	0.2209
		$Norm(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$	1.0988		0.5539		0.4100	
					$q = 1.67$		$q = 1.03$	
		$\hat{\boldsymbol{\theta}}$	3.5495	4.5169	2.9669	4.5406	3.0070	4.6154
0.25	50	$\widehat{Var}(\hat{\boldsymbol{\theta}})$	0.2156	0.0598	0.1504	0.0577	0.0981	0.0552
		$\widehat{MSE}(\hat{\boldsymbol{\theta}})$	0.5175	0.2932	0.1515	0.2687	0.0981	0.2031
		$Norm(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$	0.8107		0.4202		0.3012	
					$q = 1.49$		$q = 0.92$	
		$\hat{\boldsymbol{\theta}}$	3.5118	4.5129	3.0295	4.5208	3.0099	4.5893
0.25	100	$\widehat{Var}(\hat{\boldsymbol{\theta}})$	0.1042	0.0302	0.0772	0.0288	0.0423	0.0273
		$\widehat{MSE}(\hat{\boldsymbol{\theta}})$	0.3661	0.2675	0.0781	0.2584	0.0424	0.1960
		$Norm(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$	0.6336		0.3365		0.2384	
					$q = 1.57$		$q = 1.42$	
		$\hat{\boldsymbol{\theta}}$	3.8580	4.6213	3.1896	4.6013	3.1440	4.7269
0.5	30	$\widehat{Var}(\hat{\boldsymbol{\theta}})$	0.4267	0.1037	0.2832	0.0973	0.2306	0.0955
		$\widehat{MSE}(\hat{\boldsymbol{\theta}})$	1.1629	0.2471	0.3191	0.2563	0.2513	0.1702
		$Norm(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$	1.4100		0.5754		0.4215	
					$q = 1.66$		$q = 1.44$	
		$\hat{\boldsymbol{\theta}}$	3.7960	4.6211	3.0502	4.5918	3.0215	4.6999
0.5	50	$\widehat{Var}(\hat{\boldsymbol{\theta}})$	0.2466	0.0615	0.1645	0.0581	0.0993	0.0557
		$\widehat{MSE}(\hat{\boldsymbol{\theta}})$	0.8802	0.2051	0.1670	0.2248	0.0997	0.1458
		$Norm(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$	1.0853		0.3918		0.2455	
					$q = 1.61$		$q = 1.3$	
		$\hat{\boldsymbol{\theta}}$	3.7616	4.6231	3.0175	4.5806	3.0360	4.6727
0.5	100	$\widehat{Var}(\hat{\boldsymbol{\theta}})$	0.1138	0.0310	0.0839	0.0298	0.0466	0.0279
		$\widehat{MSE}(\hat{\boldsymbol{\theta}})$	0.6939	0.1730	0.0842	0.2057	0.0479	0.1350
		$Norm(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$	0.8669		0.2899		0.1829	
					$q = 1.89$		$q = 2.06$	
		$\hat{\boldsymbol{\theta}}$	3.8353	4.7620	2.9925	4.6951	2.9976	4.8135
0.75	30	$\widehat{Var}(\hat{\boldsymbol{\theta}})$	0.4546	0.1065	0.2655	0.1016	0.2604	0.1022
		$\widehat{MSE}(\hat{\boldsymbol{\theta}})$	1.1523	0.1631	0.2656	0.1946	0.2604	0.1370
		$Norm(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$	1.3154		0.4602		0.3974	
					$q = 1.95$		$q = 1.75$	
		$\hat{\boldsymbol{\theta}}$	3.7763	4.7554	2.8721	4.6742	3.0365	4.7805
0.75	50	$\widehat{Var}(\hat{\boldsymbol{\theta}})$	0.2500	0.0625	0.1500	0.0598	0.1472	0.0585
		$\widehat{MSE}(\hat{\boldsymbol{\theta}})$	0.8527	0.1224	0.1664	0.1660	0.1486	0.1067
		$Norm(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$	0.9751		0.3324		0.2553	
					$q = 1.85$		$q = 1.64$	
		$\hat{\boldsymbol{\theta}}$	3.7386	4.7547	2.8482	4.6616	3.0544	4.7496
0.75	100	$\widehat{Var}(\hat{\boldsymbol{\theta}})$	0.1200	0.0320	0.0859	0.0311	0.1019	0.0297
		$\widehat{MSE}(\hat{\boldsymbol{\theta}})$	0.6656	0.0922	0.1089	0.1456	0.1049	0.0924
		$Norm(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$	0.7578		0.2545		0.1973	

Mehmet Niyazi ÇANKAYA   **10(3): 1984-1996, 2020**

M-Estimations of Shape and Scale Parameters by Order Statistics in Least Informative Distributions on q-deformed logarithm

Table 2. Inliers case: $\hat{a}$ and $\hat{b}$ from ρ based on MLE, MLqE and LIDOSq when $a = 2$ and $b = 10$

r	n	θ	MLE		MLqE		LIDOSq	
					$q = 1.98$		$q = 1.55$	
			$a = 2$	$b = 10$	$a = 2$	$b = 10$	$a = 2$	$b = 10$
		$\hat{\theta}$	2.4039	8.6137	2.0680	9.2462	2.0443	9.5880
0.25	30	$\widehat{Var}(\hat{\theta})$	0.1723	0.8165	0.0936	0.8506	0.0671	0.9298
		$\widehat{MSE}(\hat{\theta})$	0.3354	2.7382	0.0982	1.4189	0.0690	1.0995
		$Norm(\hat{\theta}, \theta)$	3.0736		1.5171		1.1685	
					$q = 1.97$		$q = 1.54$	
		$\hat{\theta}$	2.3671	8.5931	2.0186	9.1898	2.0014	9.5249
0.25	50	$\widehat{Var}(\hat{\theta})$	0.0950	0.4840	0.0533	0.4902	0.0358	0.5463
		$\widehat{MSE}(\hat{\theta})$	0.2297	2.4634	0.0537	1.1465	0.0358	0.7720
		$Norm(\hat{\theta}, \theta)$	2.6931		1.2002		0.8078	
					$q = 1.91$		$q = 1.49$	
		$\hat{\theta}$	2.3458	8.5938	1.9967	9.1334	1.9966	9.4596
0.25	100	$\widehat{Var}(\hat{\theta})$	0.0469	0.2538	0.0281	0.2537	0.0313	0.3039
		$\widehat{MSE}(\hat{\theta})$	0.1665	2.2313	0.0281	1.0046	0.0313	0.5959
		$Norm(\hat{\theta}, \theta)$	2.3978		1.0327		0.6272	
					$q = 1.97$		$q = 1.44$	
		$\hat{\theta}$	2.5596	8.8893	2.1113	9.3266	2.1336	9.6510
0.5	30	$\widehat{Var}(\hat{\theta})$	0.1935	0.8768	0.1005	0.8654	0.0793	0.9094
		$\widehat{MSE}(\hat{\theta})$	0.5067	2.1105	0.1129	1.3190	0.0971	1.0312
		$Norm(\hat{\theta}, \theta)$	2.6172		1.4319		1.1283	
					$q = 2.05$		$q = 1.62$	
		$\hat{\theta}$	2.5352	8.8880	2.0446	9.3175	2.0504	9.6616
0.5	50	$\widehat{Var}(\hat{\theta})$	0.1103	0.5280	0.0569	0.5207	0.0352	0.5508
		$\widehat{MSE}(\hat{\theta})$	0.3967	1.7645	0.0589	0.9865	0.0377	0.6653
		$Norm(\hat{\theta}, \theta)$	2.1612		1.0454		0.7030	
					$q = 2.03$		$q = 1.67$	
		$\hat{\theta}$	2.5056	8.8885	2.0017	9.2777	2.0145	9.6375
0.5	100	$\widehat{Var}(\hat{\theta})$	0.0545	0.2553	0.0300	0.2535	0.0108	0.2759
		$\widehat{MSE}(\hat{\theta})$	0.3102	1.4907	0.0300	0.7751	0.0110	0.4073
		$Norm(\hat{\theta}, \theta)$	1.8009		0.8051		0.4183	
					$q = 1.79$		$q = 1.56$	
		$\hat{\theta}$	2.5591	9.3062	2.1744	9.5089	2.1414	9.8888
0.75	30	$\widehat{Var}(\hat{\theta})$	0.1929	0.9065	0.1076	0.8931	0.0930	0.9578
		$\widehat{MSE}(\hat{\theta})$	0.5056	1.3878	0.1380	1.1343	0.1130	0.9702
		$Norm(\hat{\theta}, \theta)$	1.8934		1.2723		1.0832	
					$q = 2.07$		$q = 1.65$	
		$\hat{\theta}$	2.5178	9.2975	2.0487	9.5670	2.0729	9.8761
0.75	50	$\widehat{Var}(\hat{\theta})$	0.1114	0.5432	0.0579	0.5450	0.0482	0.5722
		$\widehat{MSE}(\hat{\theta})$	0.3795	1.0368	0.0603	0.7325	0.0535	0.5876
		$Norm(\hat{\theta}, \theta)$	1.4163		0.7928		0.6411	
					$q = 2.07$		$q = 1.65$	
		$\hat{\theta}$	2.4962	9.2794	2.0026	9.5153	2.0457	9.8142
0.75	100	$\widehat{Var}(\hat{\theta})$	0.0540	0.2688	0.0309	0.2683	0.0207	0.2801
		$\widehat{MSE}(\hat{\theta})$	0.3002	0.7881	0.0309	0.5032	0.0228	0.3146
		$Norm(\hat{\theta}, \theta)$	1.0883		0.5341		0.3374	

Table 3. Outliers case: $\hat{a}$ and $\hat{b}$ from ρ based on MLE, MLqE and LIDOSq when $a = 3$ and $b = 5$

n	$\boldsymbol{\theta}$	MLE		MLqE $q = 0.77$		LIDOSq $q = 0.0275$	
		$a = 3$	$b = 5$	$a = 3$	$b = 5$	$a = 3$	$b = 5$
30	$\widehat{\boldsymbol{\theta}}$	1.6325	5.9845	4.0598	4.6354	3.3172	7.0909
	$\widehat{Var}(\boldsymbol{\theta})$	0.0191	0.1524	0.6648	0.1112	0.5830	2.4881
	$\widehat{MSE}(\boldsymbol{\theta})$	1.8891	1.1216	1.7880	0.2441	0.6836	6.8598
	$Norm(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta})$	3.0107		2.0321		7.5434	
				$q = 0.82$		$q = 0.0275$	
50	$\widehat{\boldsymbol{\theta}}$	1.5829	6.0438	3.7733	4.6668	3.0104	6.7880
	$\widehat{Var}(\boldsymbol{\theta})$	0.0128	0.0983	0.5787	0.0789	0.6290	3.1528
	$\widehat{MSE}(\boldsymbol{\theta})$	2.0210	1.1877	1.1766	0.1899	0.6292	6.3498
	$Norm(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta})$	3.2087		1.3665		6.9790	
				$q = 0.82$		$q = 0.02$	
100	$\widehat{\boldsymbol{\theta}}$	1.5286	6.1231	3.8261	4.6377	2.6398	5.7445
	$\widehat{Var}(\boldsymbol{\theta})$	0.0075	0.0550	0.1884	0.0350	0.3734	1.6099
	$\widehat{MSE}(\boldsymbol{\theta})$	2.1727	1.3163	0.8709	0.1663	0.5031	2.1643
	$Norm(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta})$	3.4890		1.0372		2.6674	
				$q = 0.83$		$q = 0.025$	
		$a = 2$	$b = 10$	$a = 2$	$b = 10$	$a = 2$	$b = 10$
30	$\widehat{\boldsymbol{\theta}}$	1.2807	11.7543	1.9953	9.5624	2.0507	12.5667
	$\widehat{Var}(\boldsymbol{\theta})$	0.0192	1.4457	0.5919	1.2707	0.2401	8.5996
	$\widehat{MSE}(\boldsymbol{\theta})$	0.5366	4.5232	0.5919	1.4623	0.2427	15.1877
	$Norm(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta})$	5.0598		2.0542		15.4304	
				$q = 0.84$		$q = 0.02$	
50	$\widehat{\boldsymbol{\theta}}$	1.2301	11.8787	1.8505	9.6980	1.9440	11.8211
	$\widehat{Var}(\boldsymbol{\theta})$	0.0126	1.0803	0.4572	0.9621	0.2063	7.3526
	$\widehat{MSE}(\boldsymbol{\theta})$	0.6053	4.6098	0.4796	1.0533	0.2094	10.6689
	$Norm(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta})$	5.2151		1.5329		10.8783	
				$q = 0.835$		$q = 0.01$	
100	$\widehat{\boldsymbol{\theta}}$	1.1772	12.0162	2.0038	9.4158	2.0029	10.7047
	$\widehat{Var}(\boldsymbol{\theta})$	0.0081	0.8980	0.4024	0.6709	0.2099	5.7911
	$\widehat{MSE}(\boldsymbol{\theta})$	0.6851	4.9631	0.4024	1.0122	0.2099	6.2877
	$Norm(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta})$	5.6482		1.4146		6.4976	

Real data application for the estimations of shape and scale parameters of Weibull distribution

A real data set (Proschan, 1963) represents the time for the consecutive failures of the air conditioning system of a fleet of Boeing 720 jet airplanes. The hours of flying time between failures are analyzed for the real data application. The data are 23, 261, 87, 7, 120, 14, 62, 47, 225, 71, 246, 21, 42, 20, 5, 12, 120, 11, 3, 14, 71, 11, 14, 11, 16, 90, 1, 16, 52, 95. The data set is fitted by the objective functions in Table 4.

Mehmet Niyazi ÇANKAYA	10(3): 1984-1996, 2020

M-Estimations of Shape and Scale Parameters by Order Statistics in Least Informative Distributions on q-deformed logarithm

Table 4. Estimates of parameters a and b by using different objective functions based on log, $\log_q$ and LID in $\log_q$

M-functions: Objective functions	$\hat{a}$	$\hat{b}$	p-values of KS
$\rho_{\log}(f_0 = W(a,b))$	0.8536	54.6180	0.4370
$\rho_{\log_{q=0.9135}}(f_0 = W(a,b))$	0.8507	47.6563	0.7612
$\rho_{\log_{q=0.03,r=1.95}}(f_0 = W(a,b), f_1 = OSW(a,b))$	0.8590	47.1482	**0.7701**

$OSW(a,b)$ represents the order statistics (OS) of $W(a,b)$. Table 4 shows M-estimates from M-estimators produced by objective function ρ. As much as the p value of KS test statistic is near to one, the best fitting with estimates $\hat{a}$ and $\hat{b}$ of $W(a,b)$ can be accomplished by means of $F_U(x_{(i)}; \hat{\boldsymbol{\theta}}_{M_0})$, as given by the section for goodness of fit test. The best fit has been performed by $\rho_{\log_{q=0.03,r=1.95}}$ which is LID with OS in MLqE. We prefer to use the non-integer order for OS to manage the efficiency of function ρ, i.e., if r = 1.95, then the highest value for p-values of KS has been obtained. When the overall results in Table 4 are observed, it is seen that using OS for f_1 in LID in MLqE gave the best fitting. Thus, the efficiency via LID with OS for f_1 in MLqE is managed well. Further, using OS can give an advantage to model the data set which can have inlier(s) due to the nature of OS.
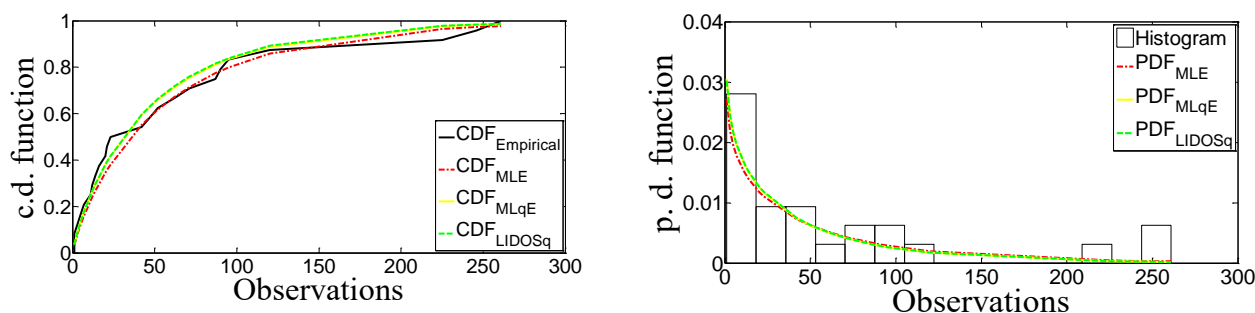

Figure 1. The fitted c.d. and p.d. functions superimposed onto emprical CDF (c.d. function) and PDF (p.d. function)

Figure 1 is given for illustrative purpose. Thus, it is also possible to observe the accordance between the fitted densities which are c.d. and p.d. functions and p-values of KS test statistic. Since we use three objective functions which are $\rho_{\log}$, $\rho_{\log_{q=0.9135}}$ and $\rho_{\log_{q=0.03,r=1.95}}$ in Table 4, we have three M-estimates produced by CDF_{MLE} for $\rho_{\log}$, CDF_{MLqE} for $\rho_{\log_q}$ and CDF_{LIDOSq} for $\rho_{\log_{q,r}}$.

General comments for simulation real data application

It is noted that if f_1 in LID is chosen as $1 - f_1$, then OS into LID in MLqE can work properly and gave the satisfactory results for simulation and real data application. As a result, when the shape of objective function and the structure of data set are accommodated to each other well, MSE values for simulation are small. Due to the page number restriction, we did not give some results of simulation. In our non-given tryings, there were situations in which MLqE and LIDOSq are more better than the estimations of shape parameter a especially.

CONCLUSION

The present paper has studied on the applicability of order statistics in LIDs in MLqE. The p-value of KS test statistic has been proposed to use the determine the value of parameter q in $\log_q$. The OS into

LID in MLqE can work properly and gave the satisfactory results if f_1 in LID is chosen as $1 - f_1$. Otherwise, they did not give small MSE values in the simulation. In addition, in the application of the real data sets, the p-values of KS are enough high when $1 - f_1$ is taken. It has been observed that if we choose $1 - f_1$ for f_1, we can have satisfactory results for capability of modeling. The shape and scale parameters as main parameters in modeling procedure have been considered to fit the data sets. We have made a comparison among the objective functions which take its inheritance rooted from likelihood estimation method based on $\log$ and $\log_q$. Since q manages the overall shape of function, we consider only these type functions. The different generalized logarithms and their performance at the different estimation methods will be investigated for future studies. The further properties for asymptotic behavior, asymptotic expansions and robustness for the used objective functions at here will be studied extensively as a separate study in the future.

REFERENCES

Andrews DF, Hampel FR, 2015. Robust estimates of location: Survey and advances. Princeton University Press.

Arnold BC, Balakrishnan N, Nagaraja HN, 1992. A first course in order statistics (Vol. 54). Siam.

Bozdogan H, 1987. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika 52(3):345-370.

Csaki F, 1981. Second international symposium on information theory. Académiai Kiadó, Budapest.

Çankaya MN, Korbel J, 2017. On statistical properties of Jizba–Arimitsu hybrid entropy. Physica A: Statistical Mechanics and its Applications 475: 1-10.

Çankaya MN, Korbel J, 2018. Least informative distributions in maximum q-log-likelihood estimation. Physica A: Statistical Mechanics and its Applications 509: 140-150.

Çankaya MN, 2018. Asymmetric bimodal exponential power distribution on the real line. Entropy 20(1): 1-23.

Elze HT, 2004. Introduction: Quantum Theory and Beneath? In Decoherence and Entropy in Complex Systems. Springer. Berlin, Heidelberg, 119-124.

Ferrari D, Yang Y, 2010. Maximum Lq-likelihood estimation. The Annals of Statistics 38(2): 753-783.

Giuzio M, Ferrari D, Paterlini S, 2016. Sparse and robust normal and t-portfolios by penalized Lq-likelihood minimization. European Journal of Operational Research 250(1): 251-261.

Gelfand I, Fomin S, 1963. Calculus of Variations. Prentice-Hall Inc. Englewood Cliffs. NJ.

Godambe VP, 1960. An optimum property of regular maximum likelihood estimation. The Annals of Mathematical Statistics 31(4):1208-1211.

Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA, 2011. Robust statistics: the approach based on influence functions. John Wiley & Sons. Vol. 196.

Huber-Carol, C, Balakrishnan, N, Nikulin M, Mesbah M. (Eds.), 2012. Goodness-of-fit tests and model validity. Springer Science & Business Media.

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

Jizba P, 2004. Information theory and generalized statistics. In Decoherence and Entropy in Complex Systems (pp. 362-376). Springer, Berlin, Heidelberg.

Jizba P, Korbel J, 2016. On q-non-extensive statistics with non-Tsallisian entropy. Physica A: Statistical Mechanics and its Applications 444: 808-827.

Malik SC, Arora S, 1992. Mathematical analysis. New Age International.

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

Mehmet Niyazi ÇANKAYA   **10(3): 1984-1996, 2020**

M-Estimations of Shape and Scale Parameters by Order Statistics in Least Informative Distributions on q-deformed logarithm

Örkcü HH, Özsoy VS, Aksoy E, Dogan MI, 2015. Estimating the parameters of 3-p Weibull distribution using particle swarm optimization: A comprehensive experimental comparison. Applied Mathematics and Computation 268: 201-226.

Proschan F, 1963. Theoretical explanation of observed decreasing failure rate. Technometrics 5(3), 375-383.

Prudnikov AP, Brychkov, IA, Marichev OI, 1986. Integrals and series: special functions (Vol. 2). CRC Press.

Rinne, H, 2008. The Weibull distribution: A handbook. CRC press.

Shevlyakov G, Morgenthaler S, Shurygin A, 2008. Redescending M-estimators. Journal of Statistical Planning and Inference 138(10): 2906-2917.

Suyari H, 2006. Mathematical structures derived from the q-multinomial coefficient in Tsallis statistics. Physica A: Statistical Mechanics and its Applications 368(1): 63-82.

Tsallis C, 1988. Possible generalization of Boltzmann-Gibbs statistics. Journal of statistical physics 52(1-2): 479-487.