A Double Transformed- Flipped-Retransformed Quantile Estimator for Skewed Distributions[1]

Dr. Mohammad Ibrahim Soliman Gaafar

Department of Statistics
Faculty of Business – Alexandria University
Mohammad.solayman@alexu.edu.eg
aswanybakkar@yahoo.com

ABSTRACT

In this paper, a hybrid three-step approach is introduced to bring the data to approximate normality. This approach uses two different transformations jointly with flipping and, in some cases, with Winsorization. The first step is to achieve approximate symmetry by transforming the data using the generalized modulus family of transformations. If the quantile to be estimated is in the longer tail, the resulting transformed sample is then Winsorized. The second step is to achieve exact sample symmetry by flipping the lower (upper) half of the transformed sample when estimating quantiles smaller (larger) than the median. The third step is to approximately Gaussianize the resulting sample using the sinh-arcsinh transformation. Estimating the quantile of the new data and then double back transforming, the new proposed nonparametric quantile estimator can be obtained. Through a simulation study, the new proposed quantile estimator is evaluated and compared with some competitor existing estimators. Simulation results show stable empirical performance and unrestricted outperformance of the proposed estimator compared to all other competitor estimators under investigation.

Keywords: *Quantiles, Robust Estimators, Modulus Family of Power Transformation, Generalized Modulus Family of Power Transformation, Sinh-arcSinh family of Transformations, Winsorized Sample, Flipped Sample.*

[1] Received in 15/1/2024, accepted in 8/2/2024.

1. INTRODUCTION

Applications of quantiles arise in diverse areas such as: Hydrology, Finance, Business, Insurance, Biology, Metrology, Economics, Reliability, Quality control, Medicine, Engineering, Demography, etc. For example, quantiles are used to determine warranty periods of goods in reliability analysis, the process capability index (PCI) in quality control. In medicine, quantiles are used to determine reference intervals within which some percentage of the values of a particular analyte in a healthy population would fall (Horn, Pesce, & Copeland, 1998). In environmental sciences, they are used to determine safety margins over which pollution of chemicals would be dangerous. In traffic engineering, they are used to set speed limits of cars and typical walking speeds. In business and finance, they are used as values at risk (VaR). Extreme quantiles of wave heights do really help design offshore platforms, breakwaters, or dikes. Quantiles are frequently used in descriptive as well as inferential statistical analysis. They help identify and capture key aspects of a distribution such as: central tendency, spread, skewness, and tail-heaviness. They help in the box-whiskers, the Q-Q, and the violin plots. The most well-known quantile is the median which is the 0.50^{th} quantile or the 50^{th} percentile.

Let $X_1, X_2, \cdots, X_n$ be a random sample of size (n) drawn from a population with distribution function $F(\cdot)$, and let $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ be the corresponding order statistics. The p^{th} population quantile, $Q(p)$, is defined as $Q(p) = inf\{x: F(x) \geq p\}$. If X is a normally distributed random variable with a mean of μ and a variance of σ^2, the p^{th} normal quantile can be given as $Q^N(p) = \mu + \sigma Z_p$, where Z_p is the p^{th} standard normal quantile. Replacing μ and σ by corresponding different estimators leads to corresponding different normal quantile estimators. The most well-known estimators of normal quantiles are:

(1) The biased normal p^{th} quantile defined as

$$\widehat{Q}_1^N(p) = \bar{X} + S.Z_p, \qquad (1.1)$$

where S is the unbiased sample standard deviation.

(2) The uniformly minimum variance unbiased estimator (UMVUE) introduced by Chakraborti and Li (2007). It is defined as

$$\widehat{Q}_2^N(p) = \bar{X} + c_n S.Z_p \qquad (1.2)$$

where $c_n = \left[\sqrt{0.5(n-1)}\,\Gamma\big(0.5(n-1)\big)/\Gamma(0.5n)\right]$.

(3) The t-based estimator used by Wang and Horn (2016). It is defined as

$$\widehat{\mathrm{Q}}_3^N(p) = \bar{X} + t_{(n-1,p)} S\sqrt{(n+1)/n}, \qquad (1.3)$$

where $t_{(n-1,p)}$ is the p^th quantile of the Student's T distribution with $(n-1)$ degrees of freedom.

Under symmetric models, and using robust location and scale estimators, Horn (1988) introduced the following robust quantile estimator:

$$\hat{Q}_H(p) = T_{b_i}(x, c_1) + t_{(n-1,p)}\sqrt{s_T^2(x, c_1) + s_{b_i}^2(x, c_2)}, \qquad (1.4)$$

*w*here $T_{b_i}(x, c_1)$ is the bi-weight location estimator with tuning constant (c_1), $s_T^2(x, c_1)$ is the bi-weight estimator of the variability of $T_{b_i}(x, c_1)$, and $s_{b_i}^2(x, c_2)$ is the square of the bi-weight estimator of dispersion with tuning constant (c_2). Due to the Equivariance-under-Monotonic-Increasing-Transformation property that quantiles have, transformed retransformed quantile estimators are applicable. Transformations can be applied to nonnormally distributed variables to bring them to approximate normality. Using Gaussian-based (e.g., $\widehat{\mathrm{Q}}_1^N(p)$, $\widehat{\mathrm{Q}}_2^N(p)$, $\widehat{\mathrm{Q}}_3^N(p)$, or $\hat{Q}_H(p)$) or symmetric-based estimators, the quantile of the transformed variable is calculated and back transformed to get an estimator of the original variable's p^th quantile. Accordingly, most quantile estimators can be divided into two broad categories; Order-Statistics-Based estimators and Gaussian-Based estimators. Estimators in the first category may be based on only one, two, or all order statistics. Huang and Brill (1999) introduced a quantile estimator based on only one order statistics. Hyndman and Fan (1996) show 9 quantile estimators each based on two order statistics. Harrell and Davis (1982), Kaigh and Lachenbruch (1982), Sfakianakis and Verginis (2008), Navruz and Özdemir (2020), and Hutson (2022) introduced some quantile estimators based on all order statistics. Estimators based on one, two, or three sample order statistics experience substantial lack of efficiency caused by the variability of individual order statistics. The efficiency of order-statistics-based estimators can be improved by forming weighted averages of all sample order statistics (see, Sheather & Marron, 1990). Estimators in the second category rely mainly on replacing location and scale estimates in the Gaussian quantile form.

Although the statistical literature is rich in various estimators of population quantiles, there is no one specific estimator that do best for all sample sizes (n), all quantile levels (p), or all different distributional shapes. For this reason, there will always be a continuous need to new quantile estimators that can do best with different patterns of the data.

This paper is organized as follows. Section 2 introduces two families of transformations on which the proposed quantile estimator relies; the generalized modulus family of power transformations introduced by Halawa (1989) and the sinh-arcsinh family of transformations introduced by Jones and Pewsey (2009). For each transformation, this section gives the form of the transformation, its inverse function, and all functions required to get the maximum likelihood estimate of each transformation's parameter(s). Section 3 is devoted to six competitor existing quantile estimators. The proposed quantile estimator is described in section 4. The design of the simulation study used to evaluate and compare the empirical performance of the proposed quantile estimator is described and the main results are reported in section 5. Section 6 gives concise conclusions of the study.

2. TWO FAMILIES OF TRANSFORMATIONS

Under location-scale models, the main goal of a transformation $T(X, \lambda)$ is to achieve some degree of normality, or at least symmetry, required for the validity of applying a certain statistical technique. This section presents the generalized modulus family of power transformations and the Sinh-arcSinh family of transformations which are valid to handle both positive and negative data.

2.1 The Generalized Modulus Family of Power Transformations

Halawa (1989) introduced the Two-domain family of power transformations which is a generalization of the Modulus family of power transformations introduced by John and Draper (1980). This transformation is defined as:

$$T_{Hal}(X;\lambda_1,\lambda_2) = \begin{cases} \left(1-(1-X)^{\lambda_1}\right)/\lambda_1 & if \quad X \le 0 \;\; and\; \lambda_1 \ne 0 \\ -ln(1-X) & if \quad X \le 0 \;\; and\; \lambda_1 = 0 \\ \left((1+X)^{\lambda_2}-1\right)/\lambda_2 & if \quad X > 0 \;\; and\; \lambda_2 \ne 0 \\ ln(1+X) & if \quad X > 0 \;\; and\; \lambda_2 = 0 \end{cases} \qquad (2.1)$$

The inverse transformation of $T_{Hal}(X;\lambda_1,\lambda_2)$ *is given as:*

$$X = T_{Hal}^{-1}(T_{Hal}) = \begin{cases} \left(1-(1-\lambda_1 T_{Hal})^{(1/\lambda_1)}\right) & if \quad T_{Hal}(X;\lambda_1,\lambda_2) \le 0 \;\; and\; \lambda_1 \ne 0 \\ 1-e^{-T_{Hal}} & if \quad T_{Hal}(X;\lambda_1,\lambda_2) \le 0 \;\; and\; \lambda_1 = 0 \\ \left((1+\lambda_2 T_{Hal})^{(1/\lambda_2)}-1\right) & if \quad T_{Hal}(X;\lambda_1,\lambda_2) > 0 \;\; and\; \lambda_2 \ne 0 \\ e^{T_{Hal}}-1 & if \quad T_{Hal}(X;\lambda_1,\lambda_2) > 0 \;\; and\; \lambda_2 = 0 \end{cases} \qquad (2.2)$$

According to the results of Van Zwet (1964), it can be shown that the transformation $T_{Hal}(X;\lambda_1,\lambda_2)$ can be used with $(0<\lambda_1<1$ and $\lambda_2>1)$ to normalize negatively skewed distributions and with $(\lambda_1 > 1$ and $0<\lambda_2<1)$ to normalize positively skewed distributions.

Assume that there exist unknown values of the transformation parameters (λ_1 and λ_2) such that the model

$$T_{Hal}\left(X-\widehat{M};\lambda_1,\lambda_2\right) = \mu_{Hal} + \sigma_{Hal}\varepsilon \qquad (2.3)$$

holds with ε having a standard normal distribution, where $(\widehat{M})$ refers to the sample median.

The Jacobian of the transformation $T_{Hal}\left(X-\widehat{M};\lambda_1,\lambda_2\right)$ is given as:

$$J_{Hal}(X;\lambda_1,\lambda_2) = \left(1+\widehat{M}-X\right)^{\lambda_1-1}\mathrm{I}\left(X\le\widehat{M}\right) + \left(1-\widehat{M}+X\right)^{\lambda_2-1}\mathrm{I}\left(X>\widehat{M}\right)$$

and hence, under model (2.3), the log-likelihood function used to estimate the parameters vector $\theta_{Hal} = (\mu_{Hal},\sigma_{Hal},\lambda_1,\lambda_2)^t$ is given as:

$$l_{Hal}(\theta_{Hal};x_1,\dots,x_n) = \frac{-n}{2}In(2\pi) - nIn(\sigma_{Hal}) - \frac{1}{2\sigma_{Hal}^2}\sum_{i=1}^{n}\left[T_{Hal}\left(x_i-\widehat{M};\lambda_1,\lambda_2\right)-\mu_{Hal}\right]^2$$

$$+\sum_{i=1}^{n}\left[\left((\lambda_1-1)ln\left(1+\widehat{M}-x_i\right)\right)\mathrm{I}\left(x_i\le\widehat{M}\right)+\left((\lambda_2-1)ln\left(1-\widehat{M}+x_i\right)\right)\mathrm{I}\left(x_i>\widehat{M}\right)\right] \qquad (2.4)$$

The partial derivatives of (2.4) with respect to μ_{Hal}, σ_{Hal}, λ_1, and λ_2 are respectively,

$$\frac{\partial l_{Hal}}{\partial \mu_{Hal}} = \frac{1}{\sigma_{Hal}^2}\sum_{i=1}^{n}\left[T_{Hal}\left(x_i-\widehat{M};\lambda_1,\lambda_2\right)-\mu_{Hal}\right], \qquad (2.5)$$

$$\frac{\partial l_{Hal}}{\partial \sigma_{Hal}} = \frac{-n}{\sigma_{Hal}} + \frac{1}{\sigma_{Hal}^3}\sum_{i=1}^{n}\left[T_{Hal}\left(x_i-\widehat{M};\lambda_1,\lambda_2\right)-\mu_{Hal}\right]^2, \qquad (2.6)$$

$$\frac{\partial l_{Hal}}{\partial \lambda_1} = \frac{-1}{\sigma_{Hal}^2}\sum_{i=1}^{n}\left[T_{Hal}\left(x_i-\widehat{M};\lambda_1,\lambda_2\right)-\mu_{Hal}\right]\times\frac{\partial T_{Hal}(x_i-\widehat{M};\lambda_1,\lambda_2)}{\partial\lambda_1} + \sum_{i=1}^{n} ln\left(1+\widehat{M}-x_i\right), \qquad (2.7)$$

$$\frac{\partial l_{Hal}}{\partial \lambda_2} = \frac{-1}{\sigma_{Hal}^2}\sum_{i=1}^{n}\left[T_{Hal}\left(x_i-\widehat{M};\lambda_1,\lambda_2\right)-\mu_{Hal}\right]\times\frac{\partial T_{Hal}(x_i-\widehat{M};\lambda_1,\lambda_2)}{\partial\lambda_2} + \sum_{i=1}^{n} ln\left(1-\widehat{M}+x_i\right) \qquad (2.8)$$

where,

$$\frac{\partial T_{Hal}}{\partial \lambda_1} = \begin{cases} -\left[T_{Hal}(x_i - \hat{M}; \lambda_1, \lambda_2) + (1 + \hat{M} - x_i)^{\lambda_1} ln(1 + \hat{M} - x_i)\right]/\lambda_1 & if \quad x_i \le \hat{M} \ \ and\, \lambda_1 \neq 0 \\ -0.5\left[ln(1 + \hat{M} - x_i)\right]^2 & if \quad x_i \le \hat{M} \ \ and\, \lambda_1 = 0 \\ 0 & if \quad x_i > \hat{M},\ \forall\, \lambda_2 \end{cases}, \quad (2.9)$$

$$\frac{\partial T_{Hal}}{\partial \lambda_2} = \begin{cases} 0 & if \quad x_i \le \hat{M},\ \forall\, \lambda_1 \\ -\left[T_{Hal}(x_i - \hat{M}; \lambda_1, \lambda_2) - (1 - \hat{M} + x_i)^{\lambda_2} ln(1 - \hat{M} + x_i)\right]/\lambda_2 & if \quad x_i > \hat{M} \ \ and\, \lambda_1 \neq 0, \\ 0.5\left[ln(1 - \hat{M} + x_i)\right]^2 & if \quad x_i > \hat{M} \ \ and\, \lambda_1 = 0 \end{cases} \quad (2.10)$$

For fixed (λ_1 and λ_2), setting (2.5) and (2.6) equal to zero, we get the estimates,

$$\hat{\mu}_{Hal} = \frac{1}{n}\sum_{i=1}^{n} T_{Hal}(x_i - \hat{M}; \lambda_1, \lambda_2), \quad (2.11)$$

And $\hat{\sigma}^2_{Hal} = \frac{1}{n}\sum_{i=1}^{n}\left[T_{Hal}(x_i - \hat{M}; \lambda_1, \lambda_2) - \hat{\mu}_{Hal}\right]^2$ (2.12)

Setting (2.7) and (2.8) equal to zero and substituting the estimators from (2.11) and (2.12) for μ_{Hal} and σ_{Hal} respectively, estimates $(\hat{\lambda}_1\ and\ \hat{\lambda}_2)$ of (λ_1 and λ_2) can be obtained using an algorithm for solving nonlinear systems of equations. The normal likelihood vector of estimators $(\hat{\theta}_{Hal})$ is determined by iteratively updating (2.11) and (2.12) by $(\hat{\lambda}_1\ and\ \hat{\lambda}_2)$ and updating (2.7) and (2.8) by $\hat{\mu}_{Hal}$ and $\hat{\sigma}_{Hal}$ until a certain convergence criterion is reached.

2.2 The Sinh-arcSinh Family of Transformations

Jones and Pewsey (2009) introduced the following two-parameter family of transformations:

$$T_{JP}(X; \alpha, \beta) = \sinh[\alpha + \beta \sinh^{-1}(X)], -\infty < \alpha < \infty \quad and \quad \beta > 0 \quad (2.13)$$

where α *serves* as a skewness parameter *and* β serves as a tail-weight parameter.

In this paper, a special case of Jones-Pewsey transformation is used. Based on the results of Van Zwet (1964) and after some algebra, it is easy to show that setting (α) equal to zero produces a transformation, $T_{JP}(X; \beta)$, that can be used to normalize symmetric distributions.

The inverse transformation of $T_{JP}(X; \beta)$ *is given as:*

$$X = T_{JP}^{-1}(T_{JP}(X; \beta), \beta) = \sinh\left[\sinh^{-1}\left(T_{JP}(X; \beta)\right)/\beta\right],\ -\infty < T_{JP}(X; \beta) < \infty \quad and \quad \beta > 0 \quad (2.14)$$

According to the results of Van Zwet (1964), it can be shown that the transformation $T_{JP}(X; \lambda_\beta)$ can be used with $(0 < \beta < 1)$ to normalize heavy-tailed symmetric distributions and with $(\beta > 1)$ to normalize light-tailed symmetric distributions.

Assume that there exists an unknown value of the transformation parameter (β) such that the model

$$T_{JP}\left(X-\widehat{M};\beta\right)=\mu_{JP}+\sigma_{JP}\varepsilon \qquad (2.15)$$

holds with ε having a standard normal distribution, where ($\widehat{M}$) refers to the sample median.

The Jacobian of the transformation $T_{JP}(X;\beta)$ is given as:

$$J_{JP}(X;\beta)=\beta\cosh\left(\beta\sinh^{-1}\left(X-\widehat{M}\right)\right)/\sqrt{1+\left(X-\widehat{M}\right)^{2}}$$

and hence, under model (2.15), the log-likelihood function used to estimate the parameters vector $\theta_{JP}=\left(\mu_{JP},\sigma_{JP},\beta\right)^{t}$ is given as:

$$l_{JP}\left(\theta_{JP};x_1,\ldots,x_n\right)=\frac{-n}{2}In(2\pi)-nIn\left(\sigma_{JP}\right)-\frac{1}{2\sigma_{JP}^2}\sum_{i=1}^{n}\left[T_{JP}\left(x_i-\widehat{M};\beta\right)-\mu_{JP}\right]^2$$

$$+\sum_{i=1}^{n}\ln\left[\beta\cosh\left(\beta\sinh^{-1}\left(x_i-\widehat{M}\right)\right)/\sqrt{1+\left(x_i-\widehat{M}\right)^{2}}\right] \qquad (2.16)$$

The partial derivatives of (2.4) with respect to μ_{JP}, σ_{PJ}, and β are respectively,

$$\frac{\partial l_{JP}}{\partial\mu_{JP}}=\frac{1}{\sigma_{JP}^2}\sum_{i=1}^{n}\left[T_{JP}\left(x_i-\widehat{M};\beta\right)-\mu_{JP}\right], \qquad (2.17)$$

$$\frac{\partial l_{JP}}{\partial\sigma_{JP}}=\frac{-n}{\sigma_{JP}}+\frac{1}{\sigma_{JP}^3}\sum_{i=1}^{n}\left[T_{JP}\left(X-\widehat{M};\beta\right)-\mu_{JP}\right]^2, \qquad (2.18)$$

$$\frac{\partial l_{JP}}{\partial\beta}=\frac{-1}{\sigma_{JP}^2}\sum_{i=1}^{n}\left[T_{JP}\left(x_i-\widehat{M};\beta\right)-\mu_{JP}\right]\times\frac{\partial T_{JP}(x_i-M;\beta)}{\partial\beta}$$

$$+\sum_{i=1}^{n}\left[(1/\beta)+\sinh^{-1}\left(x_i-\widehat{M}\right)\tanh\left(\beta\sinh^{-1}\left(x_i-\widehat{M}\right)\right)\right], \qquad (2.19)$$

where,

$$\frac{\partial\, T_{JP}(X-M;\beta)}{\partial\beta}=\sinh^{-1}\left(x_i-\widehat{M}\right)\cosh\left(\beta\sinh^{-1}\left(x_i-\widehat{M}\right)\right) \qquad (2.20)$$

For fixed (β), setting (2.17) and (2.18) equal to zero, we get the estimates,

$$\hat{\mu}_{JP}=\frac{1}{n}\sum_{i=1}^{n}T_{JP}\left(x_i-\widehat{M};\beta\right), \qquad (2.21)$$

and $\hat{\sigma}_{JP}^2=\frac{1}{n}\sum_{i=1}^{n}\left[T_{JP}\left(x_i-\widehat{M};\beta\right)-\hat{\mu}_{JP}\right]^2 \qquad (2.22)$

Using the same iterative algorithm described for the generalized modulus family of power transformations, the normal likelihood vector of estimators $\left(\hat{\theta}_{JP}=\left(\hat{\mu}_{JP},\hat{\sigma}_{JP},\hat{\beta}\right)^{t}\right)$ can be obtained.

3. COMPETITOR QUANTILE ESTIMATORS

In this section, a review of six of the best existing quantile estimators are considered.

3.1 Cheng Estimator

Based on Bernstein polynomials, Cheng (1995) introduced a nonparametric quantile estimator depending on a linear combination of order statistics. It is defined as:

$$\hat{Q}_{Ch}(p) = \sum_{i=1}^{n} \left[\binom{n-1}{i-1} p^{i-1}(1-p)^{n-i} \right] X_{(i)}, \quad 0 < p < 1 \qquad (3.1)$$

3.2 Modified Harrell-Davis Estimator

Based on a modified form of the empirical distribution function, Huang (2001) introduced an improved version of the quantile estimator introduced by Harrell and Davis (1982). This is defined as:

$$\hat{Q}_{modHD}(p) = \sum_{i=1}^{n} w_i X_{(i)}, \qquad (3.2)$$

where

$$w_i = I_{\Delta(i)}\big((n+1)p, (n+1)(1-p)\big) - I_{\Delta(i-1)}\big((n+1)p, (n+1)(1-p)\big),$$

$$\Delta(0) \equiv 0,$$

$$\Delta(i) = \sum_{j=1}^{i} \Delta^*(j)\,;\, i = 1,2,\cdots,n \text{ and}$$

$$\Delta^*(j) = \begin{cases} 0.5\left[1 - \frac{n-2}{\sqrt{n(n-1)}}\right], & i = 1, n. \\ 1/\sqrt{n(n-1)} & i = 2,3,\cdots,n-1. \end{cases}$$

The regularized incomplete beta function $\boldsymbol{I_v(a,b)}$ is defined as:

$$\boldsymbol{I_v(a,b)} = \left[\int_0^v \boldsymbol{t^{a-1}(1-t)^{b-1}.\,dt} \right] \Big/ \boldsymbol{\beta eta(a,b)}$$

According to Huang (2001), this quantile estimator is more efficient, especially for the tails of the distributions and small sample sizes.

3.3 Sfakianakis-Verginis Estimator

Sfakianakis and Verginis (2008) introduced three nonparametric quantile estimators in the forms of binomially weighted sum of all sample order statistics. In this paper, as recommended by previous studies, only the following two estimators is considered:

$$\hat{Q}_{SV2}(p) = \sum_{i=1}^{n} B(i-1;n,p)X_{(i)} + \left(2X_{(n)} - X_{(n-1)}\right)B(n;n,p), \quad (3.3)$$

$$\hat{Q}_{SV3}(p) = \sum_{i=1}^{n} B(i;n,p)X_{(i)} + \left(2X_{(1)} - X_{(2)}\right)B(0;n,p), \quad (3.4)$$

where $B(i;n,p) = \binom{n}{i} p^i (1-p)^{n-i}$

3.4 Navruz-Özdemir Estimator

Going on the same lines of Sfakianakis and Verginis (2008), Navruz and Özdemir (2022) proposed the following quantile estimator:

$$\hat{Q}_{N\ddot{o}}(p) = [(3p-1)X_{(1)} + (2-3p)X_{(2)} - (1-p)X_{(3)}]B(0;n,p) \sum_{i=1}^{n} [(1-p)B(i-1;n,p) + pB(i;n,p)]X_{(i)}$$

$$+\left[-pX_{(n-2)} + (3p-1)X_{(n-1)} + (2-3p)X_{(n)}\right]B(n;n,p) \quad (3.5)$$

3.5 Hutson-hybrid-GS Estimator

Based on a newly defined generalized expectile function, termed the sigmoidal quantile function, Hutson (2022) proposed a new quantile estimator defined as:

$$\hat{Q}_H(p) = arg\min_t \left\{(2p-1)\sum_{i=1}^{n}(x_i - t) + \frac{S}{\sqrt{n}}\sum_{i=1}^{n}\left[ln\left(1 + e^{-\sqrt{n}(x_i-t)/S}\right) + ln\left(1 + e^{\sqrt{n}(x_i-t)/S}\right)\right]\right\} \quad (3.6)$$

where (S) is the unbiased sample standard deviation.

He also introduced a hybrid quantile estimator, $\hat{Q}_{hybH}(p)$, which combines the optimal properties of the classical kernel estimator, $\hat{Q}_K(p)$, with his new estimator $\hat{Q}_H(p)$. This hybrid quantile estimator is defined as:

$$\hat{Q}_{hybH}(p) = \begin{cases} \hat{Q}_H(p) - \hat{Q}_H\left(\frac{1}{n+1}\right) + \hat{Q}_K\left(\frac{1}{n+1}\right) & if\ 0 < p \le \frac{1}{n+1} \\ \hat{Q}_K(p) & if\ \frac{1}{n+1} < p < \frac{n}{n+1} \\ \hat{Q}_H(p) - \hat{Q}_H\left(\frac{n}{n+1}\right) + \hat{Q}_K\left(\frac{n}{n+1}\right) & if\ \frac{n}{n+1} \le p < 1 \end{cases}, \quad (3.7)$$

where,

$$\hat{Q}_K(p) = \sum_{i=1}^{n}\left[\delta_i \Big/ \sum_{i=1}^{n}\delta_i\right]X_{(i)}, \quad (3.8)$$

and $\delta_i = \Phi\left(\frac{i-0.5-np}{p(1-p)}\right) - \Phi\left(\frac{i-np-1.5}{p(1-p)}\right)$

where $\Phi(\cdot)$ is the standard normal distribution function.

Hutson (2022) recommended using the proposed hybrid quantile estimator

$\hat{Q}_{hybH}(p)$, that's why it is included in the simulation study.

4. THE PROPOSED QUANTILE ESTIMATOR

In this section, the new proposed quantile estimator, $\hat{Q}_{Sol}(p)$, is described. Whether the distribution is positively or negatively skewed slightly affects how to get the estimator. For positively skewed distributions, the proposed quantile estimator, $\hat{Q}_{Sol}(p)$, can be obtained through the following detailed seven steps in a random sample of size (n):

[1] Apply the transformation $T_{Hal}\left(x_i - \widehat{M}; \lambda_1, \lambda_2\right)$ to the original sample observations and use equations (2.7) up to (2.12) to get the estimate$\left(\hat{\lambda}_1 \; and \; \hat{\lambda}_2\right)$ of the transformation parameters (λ_1 and λ_2).

[2] Find the transformed observations, Y_1, defined as $Y_1 = T_{Hal}\left(x_i - \widehat{M}; \hat{\lambda}_1, \hat{\lambda}_2\right)$.

[3] This third step depends on the quantile level to be estimated,

a) To estimate quantiles below the median, the lower half of Y_1 is flipped and the new pseudo sample is denoted as $\tilde{Y}_1$.

b) To estimate quantiles above the median,

i. Calculate $U.F. = Q_3(Y_1) + 1.5[Q_3(Y_1) - Q_1(Y_1)]$.

ii. Determine the number (m) of values of Y_1 that exceed $U.F$.

iii. Winsorize Y_1 by a proportion of $([m/n] \times 100)\%$ and denote the Winsorized sample by $\acute{Y}_1$.

iv. Flip the upper half of the sample $\acute{Y}_1$ and denote the new Winsorized pseudo sample by $\tilde{Y}_1$.

[4] Apply the transformation $T_{JP}\left(\tilde{Y}_1; \beta\right)$ to the new pseudo sample $\tilde{Y}_1$and use equations (2.19) up to (2.22) to get the estimate $(\hat{\beta})$ of the transformation parameter (β).

[5] Find the new double transformed sample, Y_2, defined as $Y_2 = T_{JP}\left(\tilde{Y}_1; \hat{\beta}\right)$.

[6] Calculate the quantile estimator $\hat{Q}_{SV3}(p)$ of Y_2.

[7] Calculate the new proposed quantile estimator of the original data $\hat{Q}_{Sol}(p)$

through the following equation:

$$\hat{Q}_{Sol}(p) = \hat{M} + T_{Hal}^{-1}\left(T_{JP}^{-1}\left(\hat{Q}_{SV3}(p); \hat{\beta}\right); \hat{\lambda}_1, \hat{\lambda}_2\right)$$

For negatively skewed distributions, the sample observations are first multiplied by (-1) then the above first 5 steps are executed to the new sample. In the sixth step, Calculate the quantile estimator $\hat{Q}_{SV3}(1-p)$ for Y_2. The new proposed quantile estimator of the original negatively skewed data $\hat{Q}_{Sol}(p)$ can be calculated through the following equation:

$$\hat{Q}_{Sol}(p) = -1 \times \left[\hat{M} + T_{Hal}^{-1}\left(T_{JP}^{-1}\left(\hat{Q}_{SV3}(1-p); \hat{\beta}\right); \hat{\lambda}_1, \hat{\lambda}_2\right)\right]$$

It worth mentioning that, if $X_1, X_2, \cdots, X_n$ is a random sample of size (n) with sample median $(\hat{M}_X)$ and $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ is the corresponding order statistics then,

a) The corresponding Winsorized sample $(\ddot{X})$ by a proportion of $([m/n] \times 100)\%$ can be expressed as:

$$\ddot{X}_{(i)} = \begin{cases} X_{(m)} & for\ i = 1,2,\cdots,m \\ X_{(i)} & for\ i = m+1, m+2, \cdots, n-m \\ X_{(n-m)} & for\ i = n-m+1, n-m+2, \cdots, n \end{cases}$$

b) The corresponding lower half flipped pseudo sample (X^*) can be expressed as:

$$X_{(i)}^* = \begin{cases} X_{(i)} & if\ i = 1,2,\cdots,\lceil 0.5(n+1)-1\rceil \\ 2\hat{M}_X - X_{(n-i+1)} & if\ i = \lceil 0.5(n+1)-1\rceil+1, \lceil 0.5(n+1)-1\rceil+2, \cdots, 2\lceil 0.5(n+1)-1\rceil \end{cases}$$

Where $\lceil\cdot\rceil$ refers to the ceiling function.

c) The corresponding upper half flipped pseudo sample (X^{**}) can be expressed as:

$$X_{(i)}^{**} = \begin{cases} 2\hat{M}_X - X_{(n-i+1)} & if\ i = 1,2,\cdots,\lceil 0.5(n+1)-1\rceil \\ X_{(i)} & if\ i = \lceil 0.5(n+1)-1\rceil+1, \lceil 0.5(n+1)-1\rceil+2, \cdots, 2\lceil 0.5(n+1)-1\rceil \end{cases}$$

5. SIMULATION STUDY

The simulation study was executed using R version 4.2.2. For each experimental situation described below, 10000 pseudo random samples, of sizes n=20, n=50, and n=100, were generated with initial seed 9831815. These samples are then used to estimate the population quantiles at seven different levels, $\boldsymbol{p = 0.025, 0.10, 0.25, 0.50, 0.75, 0.90, 0.975}$, using the first six quantile estimators described in Sections 3 and the proposed one described in section 4. The empirical performance of estimators is evaluated and compared based on the mean squared error, MSE, criterion. Four degrees of positive skewness (1.14, 2, 4, and 6.18) are covered through the standard Gumbel distribution for maximum, the standard exponential distribution, the generalized lambda distribution (GLD), and the standard lognormal distribution respectively. Two degrees of negative skewness are covered by multiplying generated data from positively skewed standard exponential and generalized lambda distributions by (-1).

For all quantile estimators and for each distribution at each and every combination of the considered different levels of "$\boldsymbol{n}$ and $\boldsymbol{p}$", the MSEs are given in tables 1 up to 6. The main results deducted from these six tables can be summarized as follows:

[1] Increasing the sample size substantially reduce the MSEs of all estimators.

[2] As the degree of skewness increases, the empirical performance of all estimators gets worse.

[3] As the value of ($\boldsymbol{p}$) gets closer to the longer tail, the MSEs of all estimators increase.

[4] As the value of ($\boldsymbol{p}$) gets closer to (0.5), the empirical performance of all estimators improves in terms of decreasing MSEs.

[5] The proposed estimator (Sol) shows the best performance in terms of least MSEs.

6. CONCLUSIONS

In this paper, a new nonparametric estimator is proposed for estimating the quantiles of skewed distributions. There are no sample restrictions from which the proposed estimator suffers. Simulation results show that the proposed estimator outperforms all other competitor estimators in most cases covered. It is recommended to estimate the population quantiles using the proposed new estimator.

Table 1: Mean squared errors (MSEs) of quantile estimators under the standard Gumbel Distribution $(\boldsymbol{Coefficient\ of\ skewness}\ (\gamma) = 1.14)$

n	p	Some Point Estimators of Population Quantiles						
		$\hat{Q}_{Ch}(p)$	$\hat{Q}_{modHD}(p)$	$\hat{Q}_{SV2}(p)$	$\hat{Q}_{SV3}(p)$	$\hat{Q}_{N\ddot{O}}(p)$	$\hat{Q}_{hybH}(p)$	$\hat{Q}_{Sol}(p)$
20	0.025	0.131076278	0.119669694	0.133145812	0.18798712	0.19312387	0.115945757	0.091853
	0.10	0.069920475	0.066040751	0.073067583	0.090049455	0.07309826	0.098762891	0.055371
	0.25	0.065457488	0.064646453	0.070367299	0.066876916	0.064586224	0.086813774	0.038895
	0.50	0.084976786	0.085307106	0.101170783	0.080472077	0.08449134	0.115459883	0.08001
	0.75	0.162032153	0.168859803	0.253191252	0.153968392	0.162152348	0.241677166	0.111387
	0.90	0.37924546	0.426921463	1.094095941	0.366742794	0.388300857	0.727241571	0.298513
	0.975	1.272205378	1.359078781	3.188389954	1.271033955	1.63947473	2.470968688	1.138279
50	0.025	0.044160567	0.04207324	0.044573454	0.077406812	0.054400735	0.053999999	0.040052
	0.10	0.027934362	0.027606416	0.028432842	0.030679156	0.027785909	0.036449209	0.021367
	0.25	0.027766372	0.027608925	0.028547708	0.027848985	0.027673057	0.03274115	0.019731
	0.50	0.036030474	0.036086982	0.038209186	0.035523482	0.03595334	0.042532303	0.032164
	0.75	0.068932904	0.070025299	0.079239104	0.067906365	0.068809455	0.082878364	0.0489527
	0.90	0.162674893	0.172057706	0.260284297	0.159880464	0.162543806	0.235899581	0.138476
	0.975	0.564144985	0.670550646	1.849245313	0.560228387	0.633029001	2.288752782	0.461035
100	0.025	0.021764108	0.021831344	0.021845269	0.034301073	0.02273737	0.026978736	0.018734
	0.10	0.015092051	0.015045138	0.015195163	0.015873997	0.01506317	0.017926506	0.011737
	0.25	0.013851008	0.013791223	0.014072744	0.013797919	0.013835515	0.01581306	0.013364
	0.50	0.018584624	0.018594858	0.019172582	0.018415577	0.01856995	0.020954264	0.016389
	0.75	0.035468738	0.035797892	0.038281634	0.035100711	0.035443741	0.041014162	0.0286942
	0.90	0.086992034	0.089554484	0.106824853	0.086211606	0.086927926	0.111203141	0.059728
	0.975	0.317571343	0.356817794	0.809077934	0.315186225	0.326647578	0.537228925	0.268753

Table 2: Mean squared errors (MSEs) of quantile estimators under the $\boldsymbol{Standard\ Exponential\ Distribution}$ $(\gamma = 2)$

n	p	Some Point Estimators of Population Quantiles						
		$\hat{Q}_{Ch}(p)$	$\hat{Q}_{modHD}(p)$	$\hat{Q}_{SV2}(p)$	$\hat{Q}_{SV3}(p)$	$\hat{Q}_{N\ddot{O}}(p)$	$\hat{Q}_{hybH}(p)$	$\hat{Q}_{Sol}(p)$
20	0.025	0.005483213	0.004330452	0.005659744	0.003291362	0.008306096	0.012775439	0.003188
	0.10	0.00841006	0.006894134	0.009262435	0.003576136	0.008561577	0.010920474	0.003509
	0.25	0.017450793	0.0162527	0.020449983	0.012071639	0.017377573	0.023631691	0.012008
	0.50	0.045393776	0.045347258	0.058708524	0.038456815	0.045472899	0.061320225	0.0349871
	0.75	0.119715556	0.126078063	0.203839631	0.110432812	0.120578509	0.184129675	0.106783
	0.90	0.322016126	0.369520719	1.005797511	0.308579987	0.328903177	0.643046123	0.30785
	0.975	1.229266145	1.316205624	3.073131405	1.227608837	1.580121078	2.294429093	1.011897
50	0.025	0.001069427	0.000795661	0.001098545	0.000406393	0.001209527	0.001050412	0.000406
	0.10	0.002527491	0.00227242	0.00267047	0.001664036	0.002518541	0.003186208	0.001634
	0.25	0.006389889	0.006187972	0.0068792	0.005461006	0.00637632	0.007618461	0.005425
	0.50	0.018488846	0.018501264	0.020528936	0.017317455	0.018470448	0.022208925	0.0157836
	0.75	0.052917814	0.05426903	0.064922455	0.05092437	0.052914706	0.065896593	0.04973
	0.90	0.148908355	0.15876751	0.24901736	0.145493598	0.148999161	0.216371029	0.14131
	0.975	0.550229782	0.664013268	1.869735234	0.545686854	0.627984348	2.317885184	0.489573
100	0.025	0.00037498	0.000306519	0.000382585	0.000170981	0.000379021	0.000442909	0.000169
	0.10	0.001078256	0.001014871	0.00113472	0.000868557	0.001076705	0.001303934	0.000861
	0.25	0.00306315	0.003013194	0.003182878	0.002836774	0.00306055	0.003508514	0.002824
	0.50	0.009543169	0.009548282	0.009987995	0.009298642	0.00953576	0.010833105	0.008888
	0.75	0.026764597	0.027100705	0.029475909	0.026295508	0.026749936	0.031343802	0.026036
	0.90	0.077456699	0.080294843	0.09862053	0.076432113	0.077435708	0.099367025	0.074616
	0.975	0.305669896	0.346467159	0.811641373	0.30307188	0.313105093	0.52042612	0.301432

Table 3: Mean squared errors (MSEs) of quantile estimators under the generalized Lambda Distribution $gld\,(-0.74575, -0.28726, -0.00989, -0.18315)$ $(\gamma = 4)$

n	p	Some Point Estimators of Population Quantiles						
		$\hat{Q}_{Ch}(p)$	$\hat{Q}_{modHD}(p)$	$\hat{Q}_{SV2}(p)$	$\hat{Q}_{SV3}(p)$	$\hat{Q}_{N\ddot{o}}(p)$	$\hat{Q}_{hybH}(p)$	$\hat{Q}_{Sol}(p)$
20	0.025	0.006903462	0.006064792	0.007048171	0.008579364	0.010049187	0.012432419	0.0058367
	0.10	0.006232235	0.005374456	0.006743224	0.004736087	0.006398992	0.008455451	0.0043981
	0.25	0.010631079	0.010007081	0.012305963	0.007944115	0.010573607	0.013925285	0.0056297
	0.50	0.027637236	0.027521199	0.036858398	0.022751777	0.027812111	0.03748721	0.019378
	0.75	0.099483064	0.105052509	0.198800767	0.087363964	0.102359818	0.148815177	0.07943
	0.90	0.437324769	0.54364707	2.04292304	0.398329939	0.401709746	0.857030485	0.324197
	0.975	2.321428612	2.979134992	9.167676188	2.270012304	1.966944752	6.11539974	1.931687
50	0.025	0.001992838	0.001871296	0.002015211	0.003378707	0.002371662	0.002293387	0.0015893
	0.10	0.002026972	0.001905784	0.002107195	0.00173788	0.002017805	0.002613007	0.001733
	0.25	0.003947563	0.003846346	0.004205144	0.003503873	0.003938786	0.0046666	0.003501
	0.50	0.010559963	0.010564486	0.011775437	0.009859305	0.010553145	0.012547035	0.010531
	0.75	0.03762452	0.038673875	0.047139398	0.035986391	0.037668293	0.04571103	0.034822
	0.90	0.157305971	0.171218719	0.322524196	0.150974949	0.158163862	0.229813212	0.134942
	0.975	1.10470255	1.603463557	5.604979704	1.079011371	0.958552892	6.815239489	0.948731
100	0.025	0.000910612	0.00090347	0.000914786	0.001483672	0.000943079	0.001096531	0.0011387
	0.10	0.000997685	0.000968117	0.00101683	0.000928073	0.000995902	0.001185144	0.001154
	0.25	0.001880567	0.001853198	0.00194709	0.001762293	0.001879004	0.002146204	0.001754
	0.50	0.005293081	0.005294298	0.005605367	0.005103892	0.005291093	0.005971231	0.004908
	0.75	0.018944892	0.019246173	0.021326273	0.018478739	0.01894464	0.021988862	0.01815
	0.90	0.078647717	0.082432404	0.10655312 5	0.077111421	0.078734586	0.101338506	0.073154
	0.975	0.604691377	0.735008558	2.386434984	0.59394327	0.580347137	1.084399557	0.497147

Table 4: Mean squared errors (MSEs) of quantile estimators under the standard lognormal distribution $(\gamma = 6.1849)$

n	p	Some Point Estimators of Population Quantiles						
		$\hat{Q}_{Ch}(p)$	$\hat{Q}_{modHD}(p)$	$\hat{Q}_{SV2}(p)$	$\hat{Q}_{SV3}(p)$	$\hat{Q}_{N\ddot{o}}(p)$	$\hat{Q}_{hybH}(p)$	$\hat{Q}_{Sol}(p)$
20	0.025	0.013033444	0.010728509	0.013389309	0.010391166	0.019299553	0.065448017	0.010087
	0.10	0.013707526	0.011553229	0.014959355	0.008133177	0.014013809	0.018602413	0.008333
	0.25	0.026832989	0.025018063	0.03147076	0.018819036	0.026754685	0.035320206	0.018811
	0.50	0.08359285	0.083021021	0.113020074	0.067638732	0.084457704	0.109014327	0.0657153
	0.75	0.393713661	0.41424409	0.890526831	0.334398635	0.413266105	0.565446827	0.260348
	0.90	2.347201902	3.018818292	12.82359659	2.086524921	2.007376045	4.453003712	1.30392
	0.975	12.94677829	16.95173081	53.53708763	12.62848366	10.4759081	34.78408765	10.42699
50	0.025	0.003272257	0.002658896	0.003339155	0.002657257	0.003905588	0.003630972	0.00267
	0.10	0.004369178	0.004057988	0.004562661	0.00346654	0.004347661	0.00564383	0.003465
	0.25	0.009270179	0.009007692	0.009919645	0.008137871	0.009253647	0.010943287	0.008128
	0.50	0.029552337	0.029538485	0.033357225	0.027279404	0.029567396	0.034937308	0.0271893
	0.75	0.135496924	0.139987634	0.177465754	0.127533294	0.135937776	0.166865715	0.116527
	0.90	0.78689145	0.857860083	1.803515614	0.747701445	0.793998636	1.09980017	0.622967
	0.975	8.170560872	12.77636754	48.46009286	7.9283118	5.981139399	58.26988896	5.380889
100	0.025	0.001310256	0.001164856	0.001328633	0.001084171	0.001357984	0.001654268	0.001089
	0.10	0.002152608	0.002073636	0.002201572	0.001922631	0.002148416	0.002602173	0.001922
	0.25	0.004707164	0.004633102	0.00488358	0.004372402	0.004703242	0.005367741	0.004362
	0.50	0.015117081	0.015119631	0.016109727	0.014464015	0.015112503	0.017129964	0.0136289
	0.75	0.069762477	0.070954387	0.079436032	0.067737549	0.06979104	0.080298213	0.065014
	0.90	0.360222264	0.380995352	0.519600315	0.350726996	0.361219838	0.459231583	0.320206
	0.975	3.88369622	4.856937247	18.5406505	3.797294991	3.576155401	6.974717239	2.631537

Table 5: Mean squared errors (MSEs) of quantile estimators under the Negative of Standard Exponential Distribution ($\gamma = -2$)

n	p	Some Point Estimators of Population Quantiles						
		$\hat{Q}_{Ch}(p)$	$\hat{Q}_{modHD}(p)$	$\hat{Q}_{SV2}(p)$	$\hat{Q}_{SV3}(p)$	$\hat{Q}_{Nö}(p)$	$\hat{Q}_{hybH}(p)$	$\hat{Q}_{Sol}(p)$
20	0.025	1.22003671	1.315420693	1.217390864	3.126938123	1.574317155	1.233638654	1.128695
	0.10	0.361993703	0.418973695	0.345035982	1.082836237	0.372230858	0.412290315	0.343301
	0.25	0.122541092	0.128934249	0.113243197	0.206132119	0.123368427	0.135854353	0.108956
	0.50	0.045393776	0.045347258	0.038456815	0.058708524	0.045472899	0.045841068	0.0367981
	0.75	0.017055962	0.015881232	0.011875872	0.019983647	0.016986139	0.015333365	0.01179862
	0.90	0.008641122	0.007031412	0.003568499	0.00953094	0.008709516	0.005257721	0.003488
	0.975	0.005528367	0.004395576	0.003235399	0.005701271	0.008293599	0.020980686	0.003149
50	0.025	0.552204908	0.67005995	0.547192658	1.842068435	0.623853608	0.66257141	0.519387
	0.10	0.146760378	0.156940968	0.143158408	0.248536034	0.146925128	0.165411315	0.138287
	0.25	0.0509184	0.052127457	0.04930088	0.061591422	0.050877423	0.055310034	0.048404
	0.50	0.018488846	0.018501264	0.017317455	0.020528936	0.018470448	0.019631906	0.01663987
	0.75	0.00646016	0.006257237	0.00552671	0.006951982	0.006447062	0.006423968	0.0054544
	0.90	0.002481133	0.002228944	0.001627486	0.002623434	0.002472634	0.002150303	0.001599
	0.975	0.001049053	0.000777379	0.00040117	0.001077881	0.001176897	0.00042122	0.000401
100	0.025	0.291807471	0.32886076	0.289471221	0.80042543	0.29843765	0.322443704	0.277313
	0.10	0.076416496	0.078855815	0.075633467	0.095205059	0.076361702	0.088912418	0.074326
	0.25	0.027193363	0.02753359	0.026713838	0.029941879	0.027179578	0.029497967	0.026431
	0.50	0.009543169	0.009548282	0.009298642	0.009987995	0.00953576	0.01031494	0.008831
	0.75	0.003107706	0.003054229	0.002863764	0.003233776	0.00310518	0.003258818	0.002844
	0.90	0.0011295	0.001066761	0.000919076	0.001164498	0.001127756	0.001123071	0.000911
	0.975	0.000370551	0.000304127	0.000172415	0.000378008	0.000374347	0.00022779	0.000171

Table 6. Mean squared errors (MSEs) of quantile estimators under the Negative of the Generalized Lambda Distribution $gld\ (-0.74575, -0.28726, -0.00989, -0.18315)\ \ (\gamma = -4)$

n	p	Some Point Estimators of Population Quantiles						
		$\hat{Q}_{Ch}(p)$	$\hat{Q}_{modHD}(p)$	$\hat{Q}_{SV2}(p)$	$\hat{Q}_{SV3}(p)$	$\hat{Q}_{Nö}(p)$	$\hat{Q}_{hybH}(p)$	$\hat{Q}_{Sol}(p)$
20	0.025	2.084337437	2.590524917	2.046236123	7.702649005	2.03904022	1.638300613	1.63339
	0.10	0.439849642	0.546280325	0.401362168	1.993758452	0.41133885	0.375578756	0.324815
	0.25	0.099420809	0.104704687	0.087334331	0.19793707	0.102342315	0.097470802	0.079027
	0.50	0.027637236	0.027521199	0.022751777	0.036858398	0.027812111	0.026567632	0.022535
	0.75	0.010518908	0.00985659	0.007662347	0.012252805	0.010462789	0.009683419	0.0073819
	0.90	0.006227876	0.005394646	0.004736864	0.006724111	0.006371941	0.005540263	0.0044312
	0.975	0.007060027	0.006251069	0.008612936	0.007200916	0.010022661	0.02632013	0.0059638
50	0.025	1.18027534	1.760767403	1.150219714	6.358695419	0.999166025	1.086665725	0.975804
	0.10	0.15436201	0.167968855	0.147702255	0.32435317	0.155408108	0.157959252	0.131043
	0.25	0.03845441	0.039611021	0.036549535	0.049044646	0.038531046	0.039407534	0.035134
	0.50	0.010559963	0.010564486	0.009859305	0.011775437	0.010553145	0.011064689	0.010531
	0.75	0.003823121	0.003724383	0.003398747	0.004074025	0.003814831	0.003908865	0.0031498
	0.90	0.002035712	0.001919053	0.001780498	0.00211239	0.002026127	0.002124893	0.001777
	0.975	0.001970164	0.001877439	0.003562547	0.001990455	0.002355424	0.004423622	0.00183549
100	0.025	0.607397033	0.743490619	0.596074125	2.490108088	0.583700182	0.532341546	0.487656
	0.10	0.07991179	0.08360951	0.078376124	0.107921784	0.080005724	0.086518371	0.074462
	0.25	0.018711953	0.019010179	0.018235094	0.021126367	0.018714683	0.019700491	0.01791
	0.50	0.005293081	0.005294298	0.005103892	0.005605367	0.005291093	0.005606111	0.005101
	0.75	0.001894537	0.001870647	0.001794438	0.001954595	0.001892719	0.0020099	0.001792
	0.90	0.000970054	0.000939038	0.000892347	0.000990215	0.000968229	0.001057647	0.000892
	0.975	0.000910261	0.000907952	0.001481495	0.000914292	0.000950389	0.001205273	0.000888

REFERENCES

Chakraborti, S. & Li, J. (2007). Confidence interval estimation of a normal percentile. The American Statistician, 61(4), 331-336.

Cheng, C. (1995). The Bernstein polynomial estimator of a smooth quantile function. Statistics and Probability Letters, 24(4), 321–330.

Halawa, A.M. (1989). *Testing for location after transformation to normality* [Unpublished thesis]. Oregon State University.

Harrell, F. E., & Davis, C. E. (1982). A new distribution-free quantile estimator. Biometrika, 69(3), 635–640.

Horn, P. S. (1988). A bi-weight prediction interval for random samples. Journal of the American Statistical Association, 83(401), 249-256.

Horn, P. S., Pesce, A. J., & Copeland, B.E. (1998). A robust approach to reference interval estimation and evaluation. Clinical Chemistry,44(3), 622-631.

Huang, M. L. (2001). On a distribution-free quantile estimator. Computational Statistics and Data Analysis, 37, 477-486.

Huang, M. L. & Brill, P. (1999). A level crossing quantile estimation method. Statistics and Probability Letters, 45(2), 111-119.

Hutson, A. D. (2022). The generalized sigmoidal quantile function. Communications in Statistics - Simulation and Computation, 53(2), 799-813.

Hyndman, R. J., & Fan, Y. (1996). Sample quantiles in statistical packages. American Statistician, 50(4), 361–365.

John, J.A. & Draper, N.R. (1980). An alternative family of transformations. Applied Statistics, 29(2), 190 – 197.

Jones, M. C. & Pewsey, A. (2009). Sinh-arcsinh distributions. Biometrika, 96(4), 761-780.

Kaigh, W. D. & Lachenbruch, P. A. (1982). A generalized quantile estimator. Communications in Statistics-Theory and Methods, 11(19), 2217-2238.

Lehman, E.L. (1983). *Theory of point estimation.* John Willey and Sons Inc., New York.

Navruz, G. & Özdemir, A. F. (2020). A new quantile estimator with weights based on a subsampling approach. British Journal of Mathematical and Statistical Psychology, 73(3),506-521.

Sfakianakis, M. E. & Verginis, D. G. (2008). A new family of nonparametric quantile estimators. Communications in Statistics – Simulation and Computation, 37(2), 337–345.

Sheather, S. J. & Marron, J.S. (1990). Kernel quantile estimators. Journal of the American Statistical Association, 85(410), 410-416.

Van Zwet, W.R. (1964). Convex Transformations of Random Variables. Mathematical Centre Tracts 7. Mathematisch Centrum, Amsterdam. MR0176511.

Wang, H. & Horn, P. S. (2016). A low-end quantile estimator from a right-skewed distribution. Communications in Statistics- Theory and Methods, 45(10), 2810-2833.

Wilcox, R. R. (2017). *Introduction to robust estimation and hypothesis testing (4th ed.).* Burlington, MA: Elsevier.

إستخدام تحويلتين مع العينة المعكوسة لتقدير مئينات التوزيعات الملتوية

د. محمد إبراهيم سليمان جعفر

ملخص البحث باللغة العربية

يتم من خلال هذا البحث تقديم مدخل هجين من ثلاث مراحل لتقريب توزيع البيانات من نمط التوزيع المعتدل (أو الطبيعي). حيث يعتمد هذا المدخل على تحويلتين مختلفتين مع استخدام فكرة العينة المعكوسة. تبدأ المرحلة الأولي بتقريب البيانات من التماثل من خلال استخدام تحويلة المقياس المعممة التي قدمها حلاوة عام 1986. وإذا كان المئين "quantile" المطلوب تقديره يقع في ذيل التوزيع (أي في الطرف الأيمن للتوزيعات موجبة الالتواء أو في الطرف الأيسر للتوزيعات سالبة الالتواء) فإن العينة المحوَّلة الناتجة يتم عمل Winsorization لها. وتهتم المرحلة الثانية بتحقيق تماثل تام في العينة من خلال إيجاد الصورة المعكوسة flipped للنصف السفلي (العلوي) للعينة المحوَّلة عند تقدير المئينات التي تقل (تزيد) عن الوسيط. وفي المرحلة الثالثة يتم استخدام تحويلة Sinh-arcSinh لتقريب نمط البيانات الناتجة من نمط التوزيع المعتدل. يتم بعد ذلك تقدير مئين البيانات المحوَّلة الجديدة التي حصلنا عيها في المرحلة الثالثة ثم استخدام الدالتين العكسيتين للتحويلتين المستخدمتين في المراحل الثلاثة السابقة وذلك للوصول إلى مقدر المئين المتين المقترح في هذا البحث. وقد تم تقييم أداء مقدر المئين المقترح ومقارنته مع بعض المقدرات المنافسة من خلال دراسة محاكاة شملت مجموعة من العوامل بمستويات مختلفة. وقد أظهرت نتائج دراسة المحاكاة استقرار الأداء التجريبي للمقدر المقترح وعدم تأثره بصغر حجم العينة وتفوقه الملحوظ على جميع المقدرات محل المقارنة.

الكلمـات الدالـة: المئينات Quantiles، المقدرات المتينة Robust، عائلة تحويلات المقياس، عائلة تحويلات المقياس المعممة، عائلة تحويلات Sinh-arcSinh، العينة المعكوسة Flipped، العينة Winsorized.

Suggested Citation according to the APA Style

Gaafar, M. I. S. (2024). A Double Transformed- Flipped- Retransformed Quantile Estimator for Skewed Distributions. *Journal of Alexandria University for Administrative Sciences*, *61*(2), 87-104.