Asian Journal of Statistical Sciences
Vol. 2, No-1, (2022), pp. 61-81

ARF INDIA
Academic Open Access Publishing
Journal homepage link:
http://arfjournals.com/ajss

On a New Generalization of Cauchy Distribution

K.Jayakumar[a] and Fasna.K[b]

[a] *Department of Statistics, University of Calicut, Kerala - 673 635, India.*
[b] *Department of Statistics, University of Calicut, Kerala -673 635, India.*

ARTICLE HISTORY
Compiled August 18, 2022

Received 05 January 2022; Accepted 20 June 2022

ABSTRACT
In this paper, we introduce a new four-parameter distribution called the new generalized Cauchy distribution (NGC). The structural properties of the new distribution are discussed. Expressions for the quantiles, mode, mean deviation, and distribution of order statistics are derived. It is shown that the distribution belongs to the class of subexponential distributions. NGC has regularly varying tails and is a member of the class of heavy-tailed distributions. It is shown that the tail weight of NGC is higher as compared to the Cauchy distribution. Parameters of NGC distribution are estimated by the percentile method, method of quantile least square, Cramer-Von Mises method, and method of maximum likelihood. Monte Carlo simulation is performed in order to investigate the performance of quantile least square estimates, Cramer-Von Mises estimates, and maximum likelihood estimates. The existence and uniqueness of maximum likelihood estimates are proved. The application of two real data sets shows the performance of the new model over other generalizations of Cauchy distribution.

KEYWORDS
Cramer-Von Mises method; Heavy tailed; Maximum likelihood estimation; Method of quantile least-square; Regular variation

1. Introduction

The first analysis of the properties of the Cauchy distribution was published by the French mathematician Poisson in 1824. The Cauchy distribution had first appeared in the works of Pierre de Fermat and was then studied by many researchers such as Issac Newton, Gottfried Leibniz, and others. Based on [1], the Cauchy distribution becomes associated with Cauchy when Cauchy responded to an article by [2] criticizing a method of interpolation proposed by Cauchy. The Cauchy distribution named after Augustin Cauchy; is a continuous probability distribution and is also known as the Lorentz distribution or Breit-Wigner distribution. It is also the distribution of the ratio of two independent normally distributed random variables with a mean zero.
The Cauchy distribution has no moments, and therefore the law of large numbers does not apply, which motivates researchers to generalize the Cauchy distribution. [3] proposed a generalization of the Cauchy distribution, [4] introduced the wrapped

CONTACT K.Jayakumar[a]. Email:jkumar19@rediffmail.com

Cauchy distribution, skew Cauchy distribution was studied in [5], another class of skew-Cauchy distribution was introduced by [6], [7] introduced a generalization of the skew-Cauchy distribution and recently [8] used the beta family studied in [9] to generate the so-called Beta Cauchy distribution.
The Cauchy distribution is used in statistics as the canonical example of pathological distribution since both its expected value and variance are undefined. In mathematics, it is closely related to the Poisson kernel, which is the fundamental solution for the Laplace equation. It is one of the distributions that are stable.
The Cauchy distribution resembles the normal distribution family of curves, while the resemblance is there it has a taller peak than normal. That means it is a heavy tail probability distribution and unlike the normal distribution, its fat tails decay much more slowly. This distribution is often used in Spectroscopy, hydrology, electric permittivity, etc.
A random variable X is said to have Cauchy distribution with parameters μ and θ if its pdf is given by

$$g(x) = \frac{1}{\pi\theta}\frac{1}{(1+(\frac{x-\mu}{\theta})^2)}; \quad -\infty \leq x \leq \infty, -\infty \leq \mu \leq \infty, \theta > 0 \tag{1}$$

and the cdf of X is given by

$$G(x) = \frac{1}{\pi}\arctan\left(\frac{x-\mu}{\theta}\right) + 0.5. \tag{2}$$

This article introduces a new four-parameter Cauchy distribution, called the New generalized Cauchy (NGC) distribution. The NGC distribution has regularly varying tails, it belongs to the class of long-tailed distributions and is a member of the dominated variation distribution. Hence it belongs to the class of subexponential distributions. Also, it can be seen that $\lim_{x\to\infty} h(x) = 0, \quad h(x)$ is the hazard rate of NGC.
As a consequence of these, we have the following:
If $X_1, X_2, ..., X_n$ are i.i.d random variables and if $S_n = X_1 + X_2 + ... + X_n$, then

$$P(S_n > x) \sim nP(X_i > x), \quad as \quad x \to \infty.$$

That is, if

$$V_n = \max_{1\leq i\leq n} X_i,$$

then

$$P(S_n > x) \sim nP(X_i > x) \sim P(V_n > x).$$

Hence, for large x, the event $S_n > x$ is due to the event $V_n > x$. That is, exceedances of the high threshold by the sum are due to the exceedances of this threshold by the largest value in the sample.
One of the classical fields of applications for subexponential distributions is insurance mathematics. Such distributions are used as realistic models for describing the sizes of real-life insurance claims which can have distributions with very heavy tails. This

interpretation suggests one way of defining a heavy-tailed distribution: the tail of the sum is essentially determined by the tail of the maximum. This intuitive approach leads to the definition of a sufficiently large class of heavy-tailed distributions.
Two classes of heavy-tailed distribution have been most successful, the distributions with regularly varying tails, and the subexponential distributions. In various fields of applied mathematics, we observe power-law behavior. To describe the deviation from pure power laws, the notion of regular variation was introduced. The regular variation appears in various fields of applied probability, so as queuing theory, extreme value theory, renewal theory, theory of summation of random variables, and point process theory.

Quantile estimation methods can be used for estimating parameters of different distributions, particularly in the case when we cannot use the method of moments, for heavy-tailed distributions. We focus on the percentile method, the quantile least squares method, and its modifications.
The paper is organized as follows. In Section 2, we introduce a new generalized Cauchy distribution, discuss the shape of the density function and distribution function of the model. We derive the quantiles, mode, and Mean deviation, Analytical shapes of the reliability functions of the model under study, pdf of order statistics, and their moments are derived in Section 3. In section 4, study the tail properties of distributions, and it is shown that the distribution has a regularly varying tail, and belongs to the class of subexponential distributions. In Section 5, quantile least square estimation methods, Cramer-Von Mises estimation method, and method of maximum likelihood estimation is explored, and we evaluate the performance of quantiles, Cramer-Von Mises, and maximum likelihood estimates using Simulation. We analyze two real data sets to illustrate the use of the proposed distribution in Section 6. In Section 7, concluding remarks are presented.

2. New Generalized Cauchy Distribution

We introduce a new generalization of the Cauchy distribution as a competitor for several generalizations of the Cauchy distribution.
[10] introduced a method of adding parameters to distribution and based on this method, several extensions of existing distributions have appeared in the literature. [11] generalized the Marshall-Olkin scheme and introduced a family of distributions generated through truncated negative binomial. Note that, both the Marshall-Olkin scheme and its generalization through truncated negative binomial arise as the distribution of random minimum or random maximum. For the applications of random minimum or random maximum in various fields, see [10]. Recently, several generalizations of Uniform distributions have appeared in the literature. [12] introduced Marshall-Olkin extended Uniform distribution with pdf

$$g(y;\alpha,\beta)=\frac{\alpha\beta}{(\alpha\beta+(1-\alpha)y)^2};\quad 0\leq y\leq 1,\alpha>0,\beta>0, \tag{3}$$

and expressed it as a mixture distribution with exponential distribution as mixing density. Using the approach of [11], [13]defined a generalized Uniform distribution with pdf

$$g(y;\alpha,\beta) = \frac{(1-\alpha)\beta\alpha^{\beta}}{(1-\alpha^{\beta})(y(1-\alpha)+\alpha)^{\beta+1}}; \quad 0 \le y \le 1, \alpha > 0, \beta > 0, \tag{4}$$

and studied its properties.
Note that, a uniform random variable Y can be transformed to a Cauchy random variable X through the transformation $X = \mu + \theta \tan[(Y-0.5)\pi]$.
When Y has the pdf (4), we obtain a new distribution which we call New Generalized Cauchy (NGC) distribution with four parameters.
The pdf of $NGC(\alpha,\beta,\mu,\theta)$ distribution thus obtained is

$$f(x) = \frac{\beta\alpha^{\beta}(1-\alpha)}{\pi(1-\alpha^{\beta})} \frac{[(0.5+\frac{1}{\pi}\arctan(\frac{x-\mu}{\theta}))(1-\alpha)+\alpha]^{-(\beta+1)}}{(1+(\frac{x-\mu}{\theta})^2)}, \tag{5}$$

where $\infty \le x \le \infty, -\infty \le \mu \le \infty, \alpha, \beta, \theta > 0$,
and the cumulative distribution function(cdf) of X is given by

$$F(x) = \frac{1-\alpha^{\beta}[(0.5+\frac{1}{\pi}\arctan(\frac{x-\mu}{\theta}))(1-\alpha)+\alpha]^{-\beta}}{1-\alpha^{\beta}}. \tag{6}$$

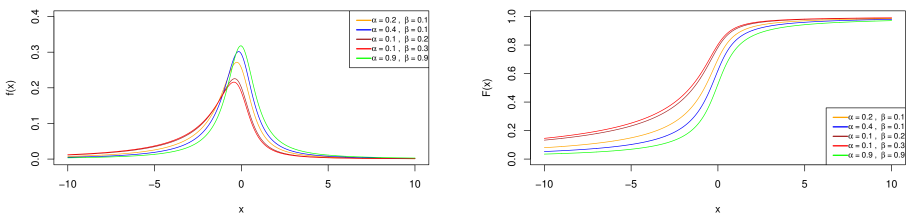
For convenience, let $\theta = 1$ and $\mu = 0$.
Then

$$f(x) = \frac{\beta\alpha^{\beta}(1-\alpha)}{\pi(1-\alpha^{\beta})} \frac{[(0.5+\frac{1}{\pi}\arctan(x))(1-\alpha)+\alpha]^{-(\beta+1)}}{(1+x^2)}, \tag{7}$$

where $-\infty \le x \le \infty, \alpha > 0, \beta > 0$.

Remark 1. When $\beta = 1$ and $\alpha \to 1$, NGC reduces to Cauchy distribution.

The pdf and cdf plots of $NGC(\alpha,\beta,\mu,\theta)$ for various values of the parameters are presented in Figure 1.



(a) Plots of the pdf of $NGC(\alpha,\beta,\mu,\theta)$ distribution

(b) Plots of the cdf of $NGC(\alpha,\beta,\mu,\theta)$ distribution

Figure 1. Pdf and cdf plot of $NGC(\alpha,\beta,\mu,\theta)$ distribution

3. Properties of the New Generalized Cauchy Distribution

Theorem 3.1. *The limit of the NGC density function as $x \to \pm\infty$ is zero.*

Proof. Trivial and hence omitted. □

3.1. Quantile function

The q^{th} quantile x_q of the NGC random variable is given by

$$x_q = \mu + \theta \tan\left[\pi\left[\frac{\alpha}{1-\alpha}[(1-q(1-\alpha^{\beta}))^{-\frac{1}{\beta}} - 1] - 0.5\right]\right]. \qquad (8)$$

In particular, Median is given by,

$$x_{0.5} = \mu + \theta \tan\left[\pi\left[\frac{\alpha}{1-\alpha}[(1-0.5(1-\alpha^{\beta}))^{-\frac{1}{\beta}} - 1] - 0.5\right]\right]. \qquad (9)$$

Remark 2. The mode of the $NGC(\alpha, \beta)$ is the solution of the equation $k(x) = 0$, where

$$k(x) = 2(\mu - x)\pi\left[\left(0.5 + \frac{1}{\pi}\arctan\left(\frac{x-\mu}{\theta}\right)(1-\alpha)\right) + \alpha\right] - \theta(1-\alpha)(1+\beta).$$

3.2. Mean Deviation

The mean deviation about the median, can be used as measure of the degree of scatter in a population. Let M be the median of the NGC distribution given by (9).
The mean deviation about the median can be calculated as

$$\delta(X) = E|X - M| = \int_{-\infty}^{\infty} |x - M| f(x)dx,$$

Hence,we obtain the following equation $\delta = \mu - 2J(M)$ where J(q) is

$$J(q) = \frac{(1-\alpha)\beta\alpha^{\beta}}{1-\alpha^{\beta}} \int_{-\infty}^{q} x \frac{[(0.5 + \frac{1}{\pi}\arctan(\frac{x-\mu}{\theta}))(1-\alpha) + \alpha]^{-(\beta+1)}}{\pi(1 + (\frac{x-\mu}{\theta})^2)} dx. \qquad (10)$$

One can easily compute this integral numerically in software such as MATLAB, Mathcad, R, and others and hence obtain the mean deviation about the median as desired.

3.3. Stochastic Ordering

Stochastic orders used in many areas of probability and statistics. Such areas include reliability theory, survival analysis, queueing theory, biology, economics, insurance, and actuarial science (see, [14]). Let X and Y be two random variables having cdf's F and G respectively, and denote by $\bar{F} = 1 - F$ and $\bar{G} = 1 - G$ their respective survival functions, with corresponding pdf's f, g. The random variable X is said to be smaller than Y in the:

(1) stochastic order (denoted as $X \leq_{st} Y$) if $\bar{F}(x) \leq \bar{G}(x)$ for all x;
(2) likelihood ratio order (denoted as $X \leq_{lr} Y$) if $\frac{f(x)}{g(x)}$ is decreasing in $x \geq 0$;
(3) hazard rate order (denoted as $X \leq_{hr} Y$) if $\frac{\bar{F}(x)}{\bar{G}(x)}$is decreasing in $x \geq 0$;

(4) reversed hazard rate order (denoted as $X \leq_{rhr} Y$) if $\frac{F(x)}{G(x)}$ is decreasing in $x \geq 0$.

The four stochastic orders defined above are related to each other, have the following implications (see,[14]):

$$X \leq_{rhr} Y \Leftarrow X \leq_{lr} Y \Rightarrow X \leq_{hr} Y \Rightarrow X \leq_{st} Y. \tag{11}$$

The NGC is ordered with respect to the strongest "likelihood ratio" ordering as shown in the following theorem. It shows the flexibility of the NGC distribution.

Theorem 3.2. *Let $X \sim NGC(\alpha_1, \beta_1)$ and $Y \sim NGC(\alpha_2, \beta_2)$. If $\beta_1 = \beta_2 = \beta$ and $\alpha_1 < \alpha_2$; then $X \leq_{lr} Y$ hence $X \leq_{rhr} Y$,$X \leq_{hr} Y$ and $X \leq_{st} Y$.*

Proof. The likelihood ratio is

$$\frac{g_X(y)}{g_Y(y)} = \frac{{\alpha_1}^{\beta}(1-\alpha_1)(1-\alpha_2^{\beta})[(0.5+\frac{1}{\pi}\arctan(x))(1-\alpha_1)+\alpha_1]^{-(\beta+1)}}{{\alpha_2}^{\beta}(1-\alpha_2)(1-\alpha_1^{\beta})[(0.5+\frac{1}{\pi}\arctan(x))(1-\alpha_2)+\alpha_2]^{-(\beta+1)}}$$

Thus,

$$\frac{d\log}{dx}\left[\frac{g_X(y)}{g_Y(y)}\right]$$

$$= \frac{(1+\beta)}{\pi(1+x^2)}\left[\frac{(1-\alpha_2)}{[(0.5+\frac{1}{\pi}\arctan(x))(1-\alpha_2)+\alpha_2]} - \frac{(1-\alpha_1)}{[(0.5+\frac{1}{\pi}\arctan(x))(1-\alpha_1)+\alpha_1]}\right]$$

$$< 0.$$

Now, if $\beta_1 = \beta_2 = \beta$ and $\alpha_1 < \alpha_2$, then $\frac{d\log}{dx}\left[\frac{g_X(y)}{g_Y(y)}\right] < 0$, which implies that $X \leq_{lr} Y$ hence $X \leq_{rhr} Y$,$X \leq_{hr} Y$ and $X \leq_{st} Y$. $\square$

3.4. Reliability Analysis

The reliability function is the characterization of an explanatory that maps a set of events, usually associated with the failure of some system onto time. It is the probability that the system will survive beyond a specified time, is defined by $R(t) = 1 - F(t)$. The Reliability function of $NGC(\alpha, \beta, \mu, \theta)$ is given by,

$$R(t) = 1 - \left[\frac{1-\alpha^{\beta}\left[(0.5+\frac{1}{\pi}\arctan(\frac{t-\mu}{\theta}))(1-\alpha)+\alpha\right]^{-\beta}}{1-\alpha^{\beta}}\right]. \tag{12}$$

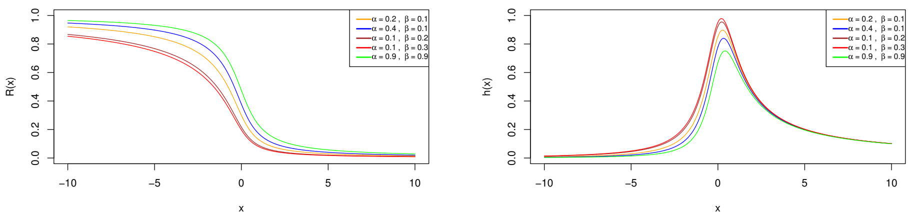
The other characteristic of interest of a random variable is the hazard rate function defined by

$$h(t) = \frac{f(t)}{1-F(t)},$$

The hazard rate function of $NGC(\alpha, \beta, \mu, \theta)$ is given by,

$$h(t) = \frac{\frac{\beta\alpha^{\beta}(1-\alpha)}{\pi(1-\alpha^{\beta})} \frac{[(0.5+\frac{1}{\pi}\arctan(\frac{t-\mu}{\theta}))(1-\alpha)+\alpha]^{-(\beta+1)}}{(1+(\frac{t-\mu}{\theta})^2)}}{1 - \left[\frac{1-\alpha^{\beta}[(0.5+\frac{1}{\pi}\arctan(\frac{t-\mu}{\theta}))(1-\alpha)+\alpha]^{-\beta}}{1-\alpha^{\beta}}\right]}. \quad (13)$$

The behaviour of reliability and hazard rate function of $NGC(\alpha, \beta, \mu, \theta)$ for various choices of the values of the parameters are presented in Figure 2. The cumulative



(a) Reliability function when $\mu = 0$ and $\theta = 1$ (b) Hazard rate function when $\mu = 0$ and $\theta = 1$

Figure 2. Reliability and hazard rate plot of $NGC(\alpha, \beta, \mu, \theta)$ distribution

hazard rate function of NGC distribution $H(t)$ is given by,

$$H(t) = -\ln R(t)$$

$$= -\ln\left[1 - \left[\frac{1-\alpha^{\beta}[(0.5+\frac{1}{\pi}\arctan(\frac{t-\mu}{\theta}))(1-\alpha)+\alpha]^{-\beta}}{1-\alpha^{\beta}}\right]\right]. \quad (14)$$

It is important to know that the units for H(t) are the cumulative probability of failure per unit of time, distance, or cycles.

Theorem 3.3. *The limit of the hazard rate function of NGC distribution as $t \to \pm\infty$ is zero.*

Proof. Trivial and hence omitted. □

3.5. Order Statistics

Let $X_1, X_2, ..., X_n$ be a random sample from $NGC(\alpha, \beta, \mu, \theta)$. Also, let $X_{(1)}, X_{(2)}, ..., X_{(n)}$, denote the corresponding order statistics. Then the pdf and cdf of k^{th} order statistics, are given by

$$f_X(x) = \frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}[1-F(x)]^{n-k} f(x)$$

$$= \frac{n!}{(k-1)!(n-k)!} \frac{1}{1-\alpha^{\beta}} \frac{\beta\alpha^{\beta}[(0.5+\frac{1}{\pi}\arctan(\frac{t-\mu}{\theta}))(1-\alpha)+\alpha]^{-(\beta+1)}(1-\alpha)}{\pi(1+(\frac{t-\mu}{\theta})^2)}$$

$$\left[\frac{1-\alpha^{\beta}[(0.5+1/\pi\arctan(x))(1-\alpha)+\alpha]^{-\beta}}{1-\alpha^{\beta}}\right]^{k-1}$$

$$\left[1-\frac{1-\alpha^{\beta}[(0.5+1/\pi\arctan(\frac{t-\mu}{\theta}))(1-\alpha)+\alpha]^{-\beta}}{1-\alpha^{\beta}}\right]^{n-k} \tag{15}$$

and

$$F_X(x)=\sum_{j=k}^{n}\binom{n}{j}[F(x)]^j[1-F(x)]^{n-j}$$

$$=\sum_{j=k}^{n}\binom{n}{j}\left[\frac{1-\alpha^{\beta}[(0.5+\frac{1}{\pi}\arctan(\frac{x-\mu}{\theta}))(1-\alpha)+\alpha]^{-\beta}}{1-\alpha^{\beta}}\right]^{j}$$

$$\left[1-\frac{1-\alpha^{\beta}[(0.5+\frac{1}{\pi}\arctan(\frac{x-\mu}{\theta}))(1-\alpha)+\alpha]^{-\beta}}{1-\alpha^{\beta}}\right]^{n-j} \tag{16}$$

respectively.
The pdf of the minimum is,

$$f_{X_{(1)}}(x)=\frac{n\beta\alpha^{\beta}[(0.5+\frac{1}{\pi}\arctan(\frac{x-\mu}{\theta}))(1-\alpha)+\alpha]^{-(\beta+1)}(1-\alpha)}{\pi(1-\alpha^{\beta})(1+(\frac{x-\mu}{\theta})^2)}$$

$$\left[1-\frac{1-\alpha^{\beta}[(0.5+\frac{1}{\pi}\arctan(\frac{x-\mu}{\theta}))(1-\alpha)+\alpha]^{-\beta}}{1-\alpha^{\beta}}\right]^{n-1} \tag{17}$$

and the pdf of the maximum is,

$$f_{X_{(n)}}(x)=\frac{n\beta\alpha^{\beta}[(0.5+\frac{1}{\pi}\arctan(\frac{x-\mu}{\theta}))(1-\alpha)+\alpha]^{-(\beta+1)}(1-\alpha)}{\pi(1-\alpha^{\beta})(1+(\frac{x-\mu}{\theta})^2)}$$

$$\left[\frac{1-\alpha^{\beta}[(0.5+\frac{1}{\pi}\arctan(\frac{x-\mu}{\theta}))(1-\alpha)+\alpha]^{-\beta}}{1-\alpha^{\beta}}\right]^{n-1}. \tag{18}$$

4. Tail Behaviour

Here we study the tail behaviour of *NGC* distribution.
The *NGC* has a heavy tail, that is it takes extreme value with high probability. This feature empirically distinguishes *NGC* from the normal and many other distributions.

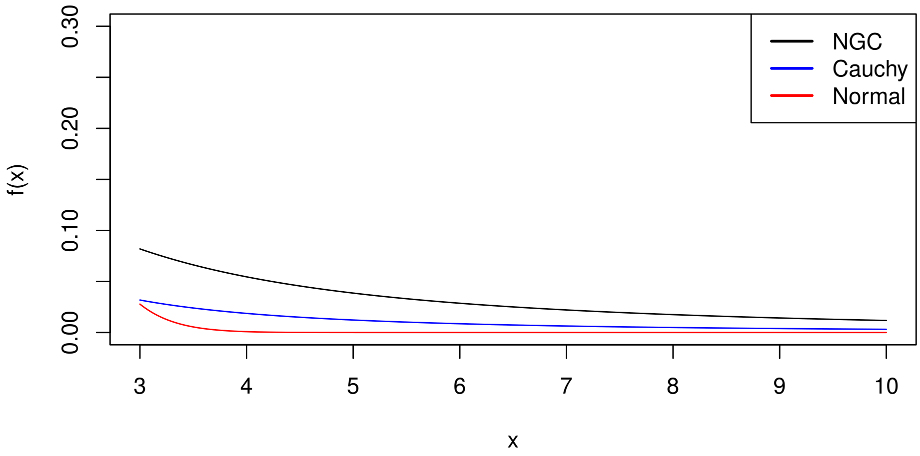


Figure 3. Comparison of tails of Cauchy, normal and NGC densities.

Figure 3 plots the tails of the density of NGC and compare them with Cauchy and normal densities.
Here we plot the tail density of NGC by using parameter values $\mu = 0, \theta = 1$ and $\alpha > 1$, $\beta \geq 1$ and compare with standard Cauchy and standard normal densities. The NGC distribution has tails thicker than Cauchy and normal.

We can easily show that $\limsup_{x\to\infty} f(x)e^{\lambda x} = \infty$ for any $\lambda > 0$, and hence the density f is heavy tailed.
Definition: An ultimately positive function f is called regularly varying at infinity with index $\gamma \epsilon R$ if for any fixed $c > 0$,

$$\lim_{x\to\infty} \frac{f(cx)}{f(x)} = c^{\gamma}.$$

The following theorem establishes that the NGC density function given in (7) is a function with regularly varying tails.

Theorem 4.1. *The density function of NGC distribution is a function with regularly varying tails.*

Proof. Using the density function (7), we have

$$\lim_{x\to\infty} \frac{f(cx)}{f(x)} = \lim_{x\to\infty} \frac{[(0.5 + \frac{1}{\pi}\arctan(cx))(1-\alpha) + \alpha]^{-(\beta+1)}(1+x^2)}{[(0.5 + \frac{1}{\pi}\arctan(x))(1-\alpha) + \alpha]^{-(\beta+1)}(1+(cx)^2)}.$$

Applying limits, the above simplifies to

$$\lim_{x\to\infty} \frac{f(cx)}{f(x)} = \frac{1}{c^2},$$

and hence we have the desired result. □

Definition: An ultimately positive function f is long-tailed and is said to belong to class L if and only if

$$\lim_{x\to\infty}\frac{f(x+y)}{f(x)}=1,\quad for\ \ all\ \ y>0.$$

Theorem 4.2. *The NGC distribution belongs to the class L.*

Proof.

$$\lim_{x\to\infty}\frac{f(x+y)}{f(x)}=\frac{[(0.5+\frac{1}{\pi}\arctan(x+y))(1-\alpha)+\alpha]^{-(\beta+1)}(1+x^2)}{[(0.5+\frac{1}{\pi}\arctan(x))(1-\alpha)+\alpha]^{-(\beta+1)}(1+(x+y)^2)}=1,$$

and hence f belongs to the class L. □

Definition: An ultimately positive function f belong to the class D of dominated variation distributions if there exists $c>0$

$$\lim_{x\to\infty}\frac{f(x)}{f(2x)}=c,\quad for\ \ all\ \ x>0.$$

Theorem 4.3. *The NGC distribution belongs to the class D dominated variation distributions .*

Proof.

$$\lim_{x\to\infty}\frac{f(x)}{f(2x)}=\lim_{x\to\infty}\frac{[(0.5+\frac{1}{\pi}\arctan(x+y))(1-\alpha)+\alpha]^{-(\beta+1)}(1+(2x)^2)}{[(0.5+\frac{1}{\pi}\arctan(2x))(1-\alpha)+\alpha]^{-(\beta+1)}(1+x^2)}.$$

Applying limits, the above simplifies to

$$\lim_{x\to\infty}\frac{f(x)}{f(2x)}=4,$$

and hence f belongs to the class of dominated variation distributions. □

5. Parameter Estimation

In this section, we use the percentile method, the quantile least squares method and its modification, the method of Cramer-von-Mises, and the method of maximum likelihood for estimation of parameters of NGC distributions.

5.1. The percentile method

In the percentile method, distribution quantiles are compared with sample quantiles. The number of required quantiles depends on the number of distribution parameters. The percentile method (PM) allows for the estimation of unknown parameters

$\theta_1, \theta_2, ..., \theta_s$ of the continuous random variable X distribution with cumulative distribution function $F(., \theta_1, \theta_2, ..., \theta_s)$ by comparing theoretical quantiles and empirical quantiles ([15], [16]). Then, the function for which we calculate the global minimum has the following form:

$$G(\theta_1, \theta_2, ..., \theta_s) = \sum_{i=1}^{n} (X_{i/n:n} - Q_{i/n})^2 \tag{19}$$

where $X_{i/n:n}$ is the sample quantile of order $p_i = \frac{i}{n}$ from the i.i.d. sample $X_1, X_2, ..., X_n$ and $Q_{i/n} = F^{-1}(\frac{i}{n}, \theta_1, \theta_2, ..., \theta_n)$.
Let $X_1, X_2, ..., X_n$ be an i.i.d. sample with cdf F. Let us denote by $X_{p_i:n}$ the sample quantile of order $p_i, i = 1, ..., s$. Estimators of the parameters $\theta_1, \theta_2, \theta_3 ..., \theta_s$ are the statistics $\hat{\theta_1}^{pm}, \hat{\theta_2}^{pm}, ..., \hat{\theta_s}^{pm}$ that are solutions of the equations:

$$X_{p_1:n} = F^{-1}(p_1, \theta_1, \theta_2, ..., \theta_s),$$

$$X_{p_2:n} = F^{-1}(p_2, \theta_1, \theta_2, ..., \theta_s), \tag{20}$$

$$.....$$

$$X_{p_s:n} = F^{-1}(p_s, \theta_1, \theta_2, ..., \theta_s).$$

where F^{-1} is the inverse of F.
When estimating parameters $\theta_1, \theta_2, \theta_3, \theta_4$ for the random variable X with cdf $F(., \theta_1, \theta_2, \theta_3, \theta_4)$, frequently the quantiles of orders $p_1, p_2, p_3, and p_4$ are chosen, such that $p_1 + p_2 + p_3 + p_4 = 1$.
For the NGC distribution, the cdf is $F(x) = \frac{1-\alpha^{\beta}[(0.5+\frac{1}{\pi}\arctan(\frac{x-\mu}{\theta}))(1-\alpha)+\alpha]^{-\beta}}{1-\alpha^{\beta}}$ and we consider the quantiles p_1, p_2, p_3 and p_4.
The estimators of μ, θ, α and β are the simultaneous solutions of the equations

$$X_{p_1:n} = \mu + \theta \tan\left[\pi\left[\frac{\alpha}{1-\alpha}[(1 - p_1(1-\alpha^{\beta}))^{-\frac{1}{\beta}} - 1] - 0.5\right]\right],$$

$$X_{p_2:n} = \mu + \theta \tan\left[\pi\left[\frac{\alpha}{1-\alpha}[(1 - p_2(1-\alpha^{\beta}))^{-\frac{1}{\beta}} - 1] - 0.5\right]\right],$$

$$X_{p_3:n} = \mu + \theta \tan\left[\pi\left[\frac{\alpha}{1-\alpha}[(1 - p_3(1-\alpha^{\beta}))^{-\frac{1}{\beta}} - 1] - 0.5\right]\right] \tag{21}$$

and

$$X_{p_4:n} = \mu + \theta \tan\left[\pi\left[\frac{\alpha}{1-\alpha}[(1-p_4(1-\alpha^\beta))^{-\frac{1}{\beta}} - 1] - 0.5\right]\right].$$

5.2. Quantile least squares method and its modification

The quantile least squares method (QLSM) estimates the unknown parameters $\theta_1, \theta_2, ..., \theta_s$ of random variable X with cdf F by minimizing the sum of squares of the differences between theoretical and empirical quantiles ([17]; [16]). Then, the function for which we calculate the global minimum has the following form:

$$G(\theta_1, \theta_2, ..., \theta_s) = \sum_{i=1}^{n}(X_{i/n:n} - Q_{i/n})^2 \tag{22}$$

where $Xi/n : n$ is the sample quantile of order $p_i = \frac{i}{n}$ from the i.i.d. sample $X_1, X_2, ..., X_n$ and $Q_{i/n} = F^{-1}(\frac{i}{n}, \theta_1, \theta_2, ..., \theta_n)$.
The estimators of parameters $\theta_1, \theta_2, ..., \theta_s$ obtained by QLSM are denoted by $\hat{\theta_1}^{qls}, \hat{\theta_2}^{qls}, ..., \hat{\theta_s}^{qls}$.
Estimation using all available quantile orders can, however, in some cases cannot be feasible. For the *NGC* distribution extreme statistics have infinite variance, which means that the mean squared errors of estimators based on them are very large. Therefore, the smallest and largest order statistics must be rejected for the estimation of the *NGC* distribution parameters. The modification of the QLSM is rejecting a fixed number of quantiles, which is known as the truncated quantile least squares method (TQLSM). In this case, the estimators of distribution parameters $\theta_1, \theta_2, ..., \theta_s$ of the random variable X with distribution function $F(., \theta_1, \theta_2, ..., \theta_s)$ are statistics $\hat{\theta_1}^{tqls}, \hat{\theta_2}^{tqls}, ..., \hat{\theta_s}^{tqls}$, for which the following expression reaches a global minimum:

$$G(\theta_1, \theta_2, ..., \theta_s) = \sum_{i \epsilon I_n}(X_{p_i:n} - Q_{p_i})^2 \tag{23}$$

where $p_i = \frac{i}{n}$ and I_n is the subset of $1, 2, ..., n$. For symmetric or close to symmetric distributions, it is recommended to skip k quantiles, where k is an even number, that is $\frac{k}{2}$ the smallest and $\frac{k}{2}$ the largest quantiles. Then, the function (23) takes the form:

$$G(\theta_1, \theta_2, ..., \theta_s) = \sum_{i=1+k/2}^{n-k/2}(X_{p_i:n} - Q_{p_i})^2 \tag{24}$$

which is to be minimized.
The application of the TQLSM for the *NGC* distribution is related to the minimization of the function:

$$G(\alpha, \beta, \mu, \theta) = \sum_{i=1+k/2}^{n-k/2}\left[X_{p_i:n} - \left[\mu + \theta \tan\left[\pi\left[\frac{\alpha}{1-\alpha}[(1-p_i(1-\alpha^\beta))^{-\frac{1}{\beta}} - 1] - 0.5\right]\right]\right]\right]^2. \tag{25}$$

Therefore the estimators of μ, θ, α and β are the simultaneous solutions of the equations $\frac{\partial G}{\partial \mu} = 0$, $\frac{\partial G}{\partial \theta} = 0$, $\frac{\partial G}{\partial \alpha} = 0$ and $\frac{\partial G}{\partial \beta} = 0$, where k is a fixed even number, $X_{p_i:n}$ is the quantile from the i.i.d. sample $X_1, X_2, ..., X_n$ and $p_i = \frac{i}{n}$ for $i = 1 + \frac{k}{2}, ..., , n - \frac{k}{2}$.

5.3. Method of Cramer-von Mises

Cramer-von-Mises type minimum distance estimators are based on minimizing the distance between the theoretical and empirical cumulative distribution functions. [18] provided empirical evidence that the bias of these estimators is smaller than the bias of other minimum distance estimators. The Cramer-von-Mises estimators $\hat{\alpha}_{CME}$,$\hat{\beta}_{CME}$,$\hat{\mu}_{CME}$, and $\hat{\theta}_{CME}$, are the values of α,β,μ, and θ, minimizing

$$C(\alpha, \beta, \mu, \theta) = \frac{1}{12n} + \sum_{i=1}^{n} \left[F(t_i \mid \alpha, \beta, \mu, \theta,) - \frac{2i-1}{2n} \right]^2 .$$

Differentiating the above equation partially, with respect to the parameters α, β ,μ and θ respectively and equating them to zero, we get the normal equations. Since the normal equations are non-linear, we can use iterative method to obtain the solution.

5.4. Maximum Likelihood Estimation(MLE)

If the parameters of the NGC distribution are not known, then the maximum likelihood estimates of the parameters can be obtained as follows: For analytical simplicity, let assume that $\mu = 0$ and $\theta = 1$.
Consider a random sample $(x_1, x_2, ..., x_n)$ of size n, from the $NGC(\alpha, \beta, \mu, \theta)$ distribution where $\mu = 0$ and $\theta = 1$. Then, the log likelihood function is given by,

$$\log L = n \log \beta - n \log(1 - \alpha^{\beta}) + n\beta \log \alpha + n \log(1 - \alpha) - n \log \pi$$

$$-(\beta+1) \sum_{i=1}^{n} \log[(0.5 + \frac{1}{\pi} \arctan(x_i))(1-\alpha) + \alpha] + \sum_{i=1}^{n} \log(1 + {x_i}^2). \quad (26)$$

The likelihood equations are,

$$\frac{\partial \log L}{\partial \alpha} = \frac{n\beta\alpha^{\beta-1}}{1-\alpha^{\beta}} + \frac{n\beta}{\alpha} - \frac{n}{1-\alpha} - (\beta+1) \sum_{i=1}^{n} \frac{(0.5 - \frac{1}{\pi} \arctan(x_i))}{[(0.5 + \frac{1}{\pi} \arctan(x_i))(1-\alpha) + \alpha]}$$

$$= 0, \quad (27)$$

and

$$\frac{\partial \log L}{\partial \beta} = \frac{n\alpha^{\beta} \log \alpha}{1-\alpha^{\beta}} + \frac{n}{\beta} + n \log \alpha - \sum_{i=1}^{n} \log[(0.5 + \frac{1}{\pi} \arctan(x_i))(1-\alpha) + \alpha]$$

$$= 0. \quad (28)$$

These equations do not have explicit solutions and they have to be obtained numerically using statistical software like *nlm* package in R.

Theorem 5.1. *Let $g_1(\alpha;\beta,x)$ denote the function on the right-hand side (RHS) of Eq. (27), where β is the true value of the parameter. Then there exists a unique solution for $g_1(\alpha;\beta,x)=0$, for $\hat{\alpha}\epsilon(0,\infty)$.*

Proof. We have

$$g_1(\alpha;\beta,x)=\frac{n\beta\alpha^{\beta-1}}{1-\alpha^{\beta}}+\frac{n\beta}{\alpha}-\frac{n}{1-\alpha}-(\beta+1)\sum_{i=1}^{n}\frac{(0.5-\frac{1}{\pi}\arctan(x_i))}{[(0.5+\frac{1}{\pi}\arctan(x_i))(1-\alpha)+\alpha]}.$$

Now

$$\lim_{\alpha\to 0}g_1(\alpha;\beta,x)=0+\infty-n-(\beta+1)\sum_{i=1}^{n}\frac{(0.5-\frac{1}{\pi}\arctan(x_i))}{(0.5+\frac{1}{\pi}\arctan(x_i))}=\infty,$$

On the other hand

$$\lim_{\alpha\to\infty}g_1(\alpha;\beta,x)=-\infty.$$

Therefore there exists at least one root, say $\hat{\alpha}\epsilon(0,\infty)$ such that $g_1(\alpha;\beta,x)=0$
To show uniqueness, the first derivative of $g_1(\alpha;\beta,x)=0$ is

$$\frac{\partial g_1(\alpha;\beta,x)}{\partial\alpha}<0,$$

Hence there exist a solution for $g_1(\alpha;\beta,x)=0$, and root, $\hat{\alpha}$ is unique. □

Theorem 5.2. *Let $g_2(\beta;\alpha,x)=0$ denote the function on the right-hand side (RHS) of Eq. (28), where α is the true value of the parameter. Then there exists a unique solution for $g_2(\beta;\alpha,x)=0$, for $\hat{\beta}\epsilon(0,\infty)$.*

Proof. We have

$$g_2(\beta;\alpha,x)=\frac{n\alpha^{\beta}\log\alpha}{1-\alpha^{\beta}}+\frac{n}{\beta}+n\log\alpha-\sum_{i=1}^{n}\log[(0.5+\frac{1}{\pi}\arctan(x_i))(1-\alpha)+\alpha].$$

Now

$$\lim_{\beta\to 0}g_2(\beta;\alpha,x)=\infty+n\log(\alpha)-\sum_{i=1}^{n}\log[(0.5+\frac{1}{\pi}\arctan(x_i))(1-\alpha)+\alpha]+0=\infty,$$

On the other hand

$$\lim_{\beta\to\infty}g_2(\beta;\alpha,x)=0+n\log(\alpha)-\sum_{i=1}^{n}\log[(0.5+\frac{1}{\pi}\arctan(x_i))(1-\alpha)+\alpha]+0<0.$$

Therefore there exists at least one root, say $\hat{\beta}\epsilon(0,\infty)$ such that $g_2(\beta;\alpha,x)=0$
To show uniqueness, the first derivative of $g_2(\beta;\alpha,x)=0$ is

$$\frac{\partial g_2(\beta;\alpha,x)}{\partial\beta}<0,$$

Hence there exists a solution for $g_2(\beta;\alpha,x)=0$, and root, $\hat{\beta}$ is unique. □

5.5. Simulation study

We conduct Monte Carlo simulation studies to compare the performance of the estimators discussed in the previous sections and the process is repeated 10000 times. We evaluate the performance of the estimators based on bias and mean squared error. Methods are compared for sample sizes $n=100$ and 500.
For each estimate we calculate the bias, mean-squared error. The statistics are obtained using the following formulae.
$Bias(\hat{\alpha})=\frac{1}{n}\sum_{i=1}^{n}(\hat{\alpha}-\alpha)$ $Bias(\hat{\beta})=\frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}-\beta)$
$Bias(\hat{\mu})=\frac{1}{n}\sum_{i=1}^{n}(\hat{\mu}-\mu)$ $Bias(\hat{\theta})=\frac{1}{n}\sum_{i=1}^{n}(\hat{\theta}-\theta)$
$MSE(\hat{\alpha})=\frac{1}{n}\sum_{i=1}^{n}(\hat{\alpha}-\alpha)^2$ $MSE(\hat{\beta})=\frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}-\beta)^2$
$MSE(\hat{\mu})=\frac{1}{n}\sum_{i=1}^{n}(\hat{\mu}-\mu)^2$ $MSE(\hat{\theta})=\frac{1}{n}\sum_{i=1}^{n}(\hat{\theta}-\theta)^2$

The bias (estimate-actual), and the mean square errors (MSE) of the parameter estimates for the percentile method, the truncated quantile least squares method, method of Cramer-von-Mises, and Maximum likelihood estimation are presented in Table 1 and 2.
From Tables 1 and 2, we note that the TQLM performs well for estimating the model parameters. Also, as the sample size increases, the biases (estimate minus actual) and the MSEs of the average estimates of truncated quantile least square estimates decrease as expected.
The following observations can be drawn from Tables 1 and 2.
1. All the estimators show the property of consistency, i.e., the MSE decreases as the sample size increases.
2. The bias of all parameters decreases with an increasing n for all the methods of estimations.
3. The bias of $\hat{\mu}$,$\hat{\theta}$, generally increases with an increasing mu, theta for any given mu,theta and n and for all methods of estimation.
4. In terms of MSE, all the methods of estimation produce smaller MSE for $\hat{\alpha}$ compared to that of other parameters.

The results of the analysis indicate that rejecting a number of the smallest and the largest quantiles significantly improved the estimators as compared to the QLSM which rejects only extreme statistics. The quantile least squares method has an advantage over the percentile method in that there is no need to determine the ranks of used order statistics. In the case of the NGC distribution, the application of minimum or maximum statistics leads to very large mean squared errors of the parameter estimators because extreme statistics have infinite variances. Rejecting extreme order statistics significantly improves the properties of the estimators. Hence, we suggest the truncated quantile least squares method.

Table 1. Simulation result for $\alpha = 0.5, \beta = 0.1, \mu = 1$ and $\theta = 2$.

n	*Est.*	*PM*	*TQLM* ($k = 40$)	*TQLM* ($k = 60$)	*CVM*	*MLE*
100	$Bias(\hat{\alpha})$	0.0003	-0.0001	-0.0004	-0.0004	0.0004
	$MSE(\hat{\alpha})$	0.0001	1.797×10^{-5}	0.0011	0.0002	0.0002
	$Bias(\hat{\beta})$	0.0003	0.002	0.0003	6.180×10^{-5}	0.0013
	$MSE(\hat{\beta})$	0.0001	0.0043	0.0001	3.819×10^{-6}	0.0026
	$Bias(\hat{\mu})$	9.140×10^{-5}	-0.0001	-0.0001	-0.0003	-0.0005
	$MSE(\hat{\mu})$	8.355×10^{-6}	2.142×10^{-5}	2.978×10^{-5}	0.0001	0.0002
	$Bias(\hat{\theta})$	-0.0001	-0.0007	-0.001	-0.0013	-0.0018
	$MSE(\hat{\theta})$	2.392×10^{-5}	0.0006	0.001	0.001	0.0019
500	$Bias(\hat{\alpha})$	0.0004	-1.626×10^{-5}	1.724×10^{-6}	-0.0004	0.0004
	$MSE(\hat{\alpha})$	0.0002	2.645×10^{-7}	2.975×10^{-9}	0.0001	0.0002
	$Bias(\hat{\beta})$	0.0002	0.005	0.0036	0.0002	0.0017
	$MSE(\hat{\beta})$	7.380×10^{-5}	0.0259	0.0129	4.504×10^{-5}	0.003
	$Bias(\hat{\mu})$	-0.001	-5.435×10^{-5}	-0.0004	-0.0003	-0.0006
	$MSE(\hat{\mu})$	0.001	2.954×10^{-6}	0.0002	0.0001	0.0003
	$Bias(\hat{\theta})$	-0.001	-0.001	-0.0007	-0.001	-0.0015
	$MSE(\hat{\theta})$	0.003	0.001	0.0005	0.0011	0.0024

Table 2. Simulation result for $\alpha = 0.9, \beta = 0.5, \mu = 2$ and $\theta = 1$.

n	*Est.*	*PM*	*TQLM* ($k = 40$)	*TQLM* ($k = 60$)	*CVM*	*MLE*
100	$Bias(\hat{\alpha})$	-6.796×10^{-5}	-0.0008	-0.0008	-0.0008	9.998×10^{-5}
	$MSE(\hat{\alpha})$	4.619×10^{-6}	0.0007	0.0007	0.0007	9.9975×10^{-6}
	$Bias(\hat{\beta})$	-4.296×10^{-5}	0.0012	3.083×10^{-5}	0.0002	-0.0003
	$MSE(\hat{\beta})$	1.846×10^{-5}	0.0015	9.505×10^{-7}	4.446×10^{-5}	9.355×10^{-5}
	$Bias(\hat{\mu})$	-0.0009	-0.0011	-0.0014	-0.0014	-0.0003
	$MSE(\hat{\mu})$	0.0008	0.0012	0.0021	0.0019	9.856×10^{-5}
	$Bias(\hat{\theta})$	0.0008	-0.0009	-0.0008	-0.0009	-0.0006
	$MSE(\hat{\theta})$	0.0007	0.0008	0.0007	0.0009	0.0003
500	$Bias(\hat{\alpha})$	9.589×10^{-5}	-0.0008	-0.0005	-0.0008	-0.0007
	$MSE(\hat{\alpha})$	9.195×10^{-5}	0.0007	0.0002	0.0008	0.0005
	$Bias(\hat{\beta})$	-0.0003	0.0017	0.0034	0.0007	-0.0004
	$MSE(\hat{\beta})$	9.577×10^{-5}	0.0032	0.1178	0.0005	0.0002
	$Bias(\hat{\mu})$	-0.0021	-0.001	-0.0012	-0.0013	-0.0014
	$MSE(\hat{\mu})$	0.0048	0.0011	0.0016	0.0017	0.0019
	$Bias(\hat{\theta})$	-0.0007	-0.0009	-0.0003	-0.0009	-0.0006
	$MSE(\hat{\theta})$	0.0005	0.0009	9.532×10^{-6}	0.0009	0.0003

6. Applications

In this section we consider two real-life data sets and compare the fit of the *NGC* distribution with the following distributions:
(a) Two parameter Cauchy distribution having pdf

$$f(x;\mu,\theta) = \frac{1}{\pi\theta}\frac{1}{(1 + (\frac{x-\mu}{\theta})^2)},$$

where $-\infty < x < \infty, -\infty < \mu < \infty, \theta > 0$.
(b)Three parameter Skew Cauchy (SC)distribution introduced by [6] with pdf

$$f(x;\mu,\theta,\lambda) = \frac{1}{\pi\theta}\frac{1}{(1 + (\frac{x-\mu}{\theta})^2)}\left[1 + \frac{\lambda(x-\mu)}{\sqrt{\theta^2 + (1+\lambda^2)(x-\mu)^2}}\right],$$

where $-\infty < x < \infty, -\infty < \mu, \lambda < \infty, \theta > 0$.
(c)Transmuted Cauchy (TC)distribution introduced by [19] with pdf

$$f(x;(\lambda,\mu,\theta)) = \frac{1}{\pi\theta}\frac{1}{(1+(\frac{x-\mu}{\theta})^2)}\left[(1+\lambda)-2\lambda\left(\frac{1}{\pi}\arctan\left(\frac{x-\mu}{\theta}\right)+0.5\right)\right],$$

where $-\infty < x < \infty, -1 \leq \lambda \leq 1, -\infty < \mu < \infty, \theta > 0$.
(d)New generalized Pareto (NGP)distribution introduced by [20] with pdf

$$f(x;(\alpha,\beta,\gamma,\theta)) = \frac{\alpha\beta^{\alpha}\theta(1-\gamma)\gamma^{\theta}}{1-\gamma^{\theta}}\frac{x^{\alpha\theta-1}}{(\gamma x^{\alpha}+(1-\gamma)\beta^{\alpha})^{\theta+1}},$$

where $x > \beta, \alpha, \beta, \gamma, \theta > 0$
The values of the log-likelihood functions($-\ln(L)$), AIC(Akaike Information Criterion), AICC(Akaike Information Criterion with correction), and BIC(Bayesian Information Criterion) are calculated for the five distributions in order to verify which distribution fits better to data. The better distribution corresponds to smaller $-\ln(L)$, AIC, AICC and BIC values. Here, $AIC = -2\ln(L) + 2k$, $AICC = -2\ln(L) + (\frac{2kn}{n-k-1})$ and $BIC = -2\ln(L) + k\ln(n)$, where L is the likelihood function evaluated at the maximum likelihood estimates, k is the number of parameters, and n is the sample size.

6.1. First data set

The real data set corresponds to the data set from [21] on breaking stress of carbon fibers (in Gba): The data set is
3.70, 2.74, 2.73, 2.50, 3.60, 3.11, 3.27, 2.87, 1.47, 3.11, 4.42, 3.22, 1.69, 3.28, 3.09, 1.87, 3.15, 4.90, 3.75, 2.43, 2.95, 2.97, 3.39, 2.67, 2.93, 3.22, 3.39, 2.81, 4.20, 3.33, 2.55, 3.31, 3.31, 2.85, 2.56, 2.35, 2.55, 2.59, 2.38, 2.81, 2.77, 2.17, 2.83, 1.92, 1.41, 3.68, 2.97, 2.76, 4.91, 3.68, 1.84, 1.59, 3.19, 1.57, 0.81, 5.56, 1.73, 1.59, 2.00, 1.71, 2.17, 1.17, 5.08, 2.48, 1.18, 3.51, 2.17, 1.69, 1.25, 4.38, 1.84, 2.48,0.85, 1.61,2.79,4.70, 2.03, 1.80, 1.57, 1.08, 2.03, 1.61,2.12, 1.89, 2.05, 3.65.
The data is approximately symmetric with skewness = 0.541 and kurtosis = 0.141
The descriptive statistics of the above data set are given in Table 3. The values in Table 4 show that the NGC distribution leads to a better fit to the other four models.

Table 3. The descriptive statistics of first data set.

Min	1st Q	Median	Mean	3rd Q	Max
0.810	1.875	2.700	2.673	3.257	5.560

Figure 4, shows the fitted density curves for the first data set.

6.2. Second data set

The second data set (http://www.ibge.gov.br/seriesestatisticas/exibedados.php?idnivel=-BR& idserie=PRECO101), is the INPC data which represents the national index of consumer prices in Brazil since 1979. The INPC index measures the cost of living of households with head employees. The data set is given below.
The data is skewed-to-the right with skewness = 1.800 and kurtosis = 4.183.

Table 4. Parameter estimates and goodness of fit for various models fitted for the first data set.

Model	parameter estimates	log L	AIC	AICC	BIC	K-S	p-value
Cauchy	$\hat{\mu} = 2.3966$ $\hat{\theta} = 0.1000$	-239.3104	484.6208	484.9134	491.9838	0.5806	0.002
SC	$\hat{\mu} = 2.5586$ $\hat{\theta} = 0.1100$ $\hat{\lambda} = 0.6136$	-136.5362	279.0724	279.365	286.4354	0.2203	0.004
TC	$\hat{\lambda} = 0.6023$ $\hat{\mu} = 2.6911$ $\hat{\theta} = 0.1100$	-136.478	278.956	279.2486	286.319	0.3177	0.005
NGP	$\hat{\alpha} = 0.7017$ $\hat{\beta} = 1.4363$ $\hat{\gamma} = 0.9996$ $\hat{\theta} = 1.1291$	-141.719	291.42	291.913	301.237	0.5060	0.002
NGC	$\hat{\mu} = 2.6474$ $\hat{\theta} = 0.6047$ $\hat{\alpha} = 0.9954$ $\hat{\beta} = 10.2729$	-134.1803	276.3606	276.8544	286.1776	0.1035	0.3147

0.69	0.44	0.13	0.03	0.17	0.37	2.47	0.62	0.57	1.39	0.39
0.97	0.42	0.12	-0.11	0.50	0.39	2.70	0.31	0.84	0.30	0.55
0.43	0.49	0.27	0.70	0.73	0.82	3.39	1.07	0.48	-0.05	0.74
0.30	0.62	0.23	0.91	0.50	0.18	1.57	0.74	0.49	0.09	0.07
0.25	0.42	0.38	0.73	0.40	0.04	0.83	1.29	0.77	0.13	0.05
0.59	0.43	0.40	0.44	0.41	-0.06	0.86	0.94	0.55	0.05	0.47
0.32	0.16	0.54	0.57	0.57	0.99	1.15	0.44	0.29	0.61	1.28
0.31	-0.02	0.58	0.86	0.39	1.38	0.61	0.79	0.16	0.74	1.29
0.26	0.11	0.15	0.44	0.83	1.37	0.09	1.11	0.43	0.94	0.65
0.26	-0.07	0.00	0.17	0.54	1.46	0.68	0.60	1.21	0.96	0.42
-0.18	-0.28	0.49	0.15	0.18	0.68	0.34	1.20	0.29	1.51	2.46
0.11	0.15	0.54	0.29	0.35	0.45	0.38	1.33	0.71	1.40	2.18
-0.31	0.72	0.85	0.10	0.11	0.81	0.02	1.28	1.46	1.17	2.10
-0.49	0.45	0.57	-0.03	0.60	0.33	0.50	0.93	1.65	1.02	2.49
1.62	1.01	1.44								

The descriptive statistics of the above data set are given in Table 5. The values in Table 6 shows that the *NGC* distribution leads to a better fit to the other four models.

Table 5. The descriptive statistics of second data set.

Min	1st Q	Median	Mean	3rd Q	Max
0.00000	0.290	0.500	0.6646	0.8600	3.3900

Figure 5, shows the fitted density curves for the second data set.

7. Concluding remarks

In this paper, we have introduced and studied a new family of distributions called the New Generalized Cauchy (*NGC*) distribution which extends the Cauchy distribution. We have studied various distributional characteristics of the model. The *NGC* distribution is heavy-tailed and belongs to the class of subexponential distributions. It has regularly varying tails and is a competitor of a number of existing generalizations of Cauchy. We have considered estimation of parameters of *NGC* distribution using the

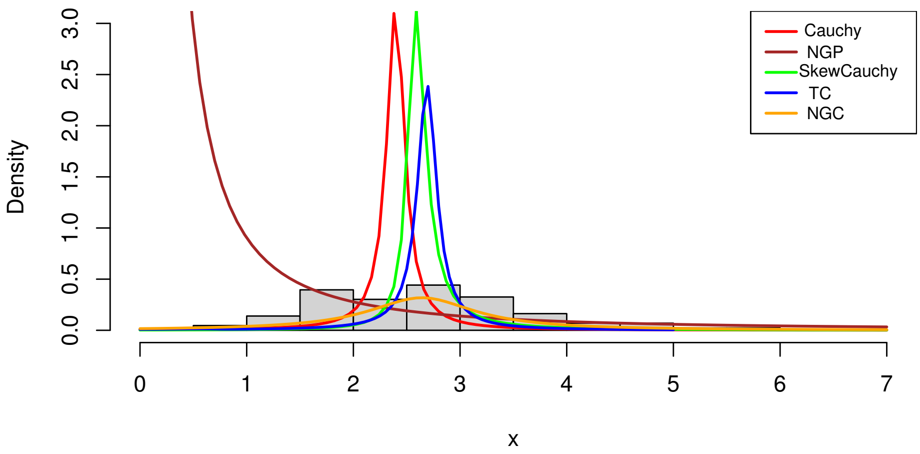


Figure 4. Fitted pdf plots of first data set

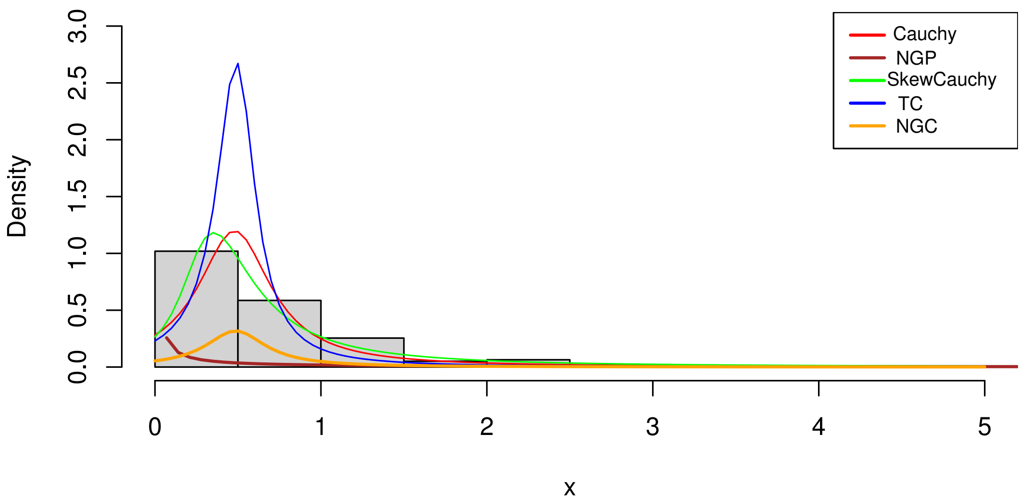


Figure 5. Fitted pdf plots of second data set.

Table 6. Parameter estimates and goodness of fit for various models fitted for the second data set.

Model	parameter estimates	log L	AIC	AICC	BIC	K-S	p-value
Cauchy	$\hat{\mu} = 0.4792$ $\hat{\theta} = 0.2656$	-139.3542	284.7083	284.8653	293.8771	0.5372	0.0303
SC	$\hat{\lambda} = 1.1888$ $\hat{\mu} = 0.2424$ $\hat{\theta} = 0.3275$	-132.7465	271.4929	271.6499	280.6617	0.4158	0.2013
TC	$\hat{\lambda} = 0.2527$ $\hat{\mu} = 0.4995$ $\hat{\theta} = 0.1400$	-132.0012	270.0025	270.1593	276.1711	0.2076	0.2628
NGP	$\hat{\alpha} = 0.0490$ $\hat{\beta} = 0.0020$ $\hat{\gamma} = 0.0867$ $\hat{\theta} = 0.0001$	-138.308	284.616	284.879	296.840	0.8579	0.3260
NGC	$\hat{\mu} = 0.4858$ $\hat{\theta} = 0.2193$ $\hat{\alpha} = 0.9993$ $\hat{\beta} = 3.7722$	-125.3175	258.635	258.8982	270.86	0.1249	0.6541

method of quantile least square, Cramer-Von Mises method, and method of maximum likelihood. In the case of NGC distribution, which is a heavy tailed distribution, rejecting a fixed number of the smallest and the largest quantiles significantly improves the properties of the parameter estimators so that the TQLSM method performs well for estimating the model parameters. The heavy-tailed properties of the model make it appropriate for modeling a number of real-life situations. The applicability of the model is demonstrated by using two real data sets. From Tables 4 and 6, we observe a better performance of our scheme than the existing distributions.

Acknowledgements

The authors wish to thank to the Editor-in-Chief and referees for their comments which improved the original version of the paper.

References

[1] Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). Continuous Univariate Distributions, vol. 1 (second edition), John Wiley and Sons, Inc., New York, .

[2] Bienayme, I. J. (1853). Remarques sur les differences qui distinguent l'interpolation de M. Cauchy de la methode des moindres carres et qui assurent la superiorite de cette method, C. R. Hebd. Seances Acad. Sci. Paris, 37, 5–13.

[3] Rider, P.R. (1957). Generalized cauchy distribution, Annals of the Institute of Statistical Mathematics, 9 , 215–223.

[4] Batschelet, E. (1981). Circular Statistics in Biology, Academic Press, San Diego.

[5] Arnold, B.C. and Beaver, R.J.(2000). The skew-Cauchy distribution, Statistics and Probability Letters, 49, 285–290.

[6] Behboodian, J., Jamalizadeh, A. and Balakrishnan, N.(2006). A new class of skew-Cauchy distributions, Statistics and Probability Letters, 76, 1488-1493.

[7] Huang, W. J., Chen, Y. H. (2007). Generalized skew-Cauchy distribution, Statistics and Probability Letters, 77, 1137–1147.

[8] Alshawarbeh, E., Famoye, F. and Lee, C. (2013). Beta-Cauchy distribution: some properties and applications, Journal of Statistical Theory and Applications, 12, 378-391.

[9] Eugene, N., Lee, C. and Famoye, F. (2002). The beta-normal distribution and its applications, Communications in Statistics: Theory and Methods, 31, 497–512.

[10] Marshall, A. W. and Olkin, I.(1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families, Biometrika, 84, 641-652.

[11] Nadarajah, K., Jayakumar, K. and Ristic, M.M. (2013). A new family of lifetime models, Journal of Statistical Computation and Simulation, 83, 1389-1404.

[12] Jose, K. K., and Krishna, E. (2011). Marshall-Olkin extended uniform distribution, ProbStat Forum, 4, 78-88.

[13] Jayakumar, K. and Sankaran, K. K. (2016). On a generalization of uniform distributions and its properties, Statistica, 76, 83-91.

[14] Shaked, M. and Shanthikumar, J.G. (2007). Stochastic Orders. Springer, New York.

[15] Wywial, J. (2004). Introduction to Statistical Inference. Publishing House of University of Economics in Katowice, Katowice.

[16] Castillo, E., Hadi, A.S., Balakrishnan, N. and Sarabia, J.M. (2004). Extreme value and related models with application in engineering and science. Wiley Interscience, A John Wiley & Sons, Inc. New Jersey.

[17] Gilchrist, W.G. (2000). Statistical modelling with quantile functions. Chapman & Hall/CRT, Boca Raton.

[18] Macdonald, P.D.M. (1971). 'Comment on "An estimation procedure for mixtures of distributions" by Choi and Bulgren', Journal of the Royal Statistical Society B, 33, 326–329.

[19] Ball, C., Rimal, B. and Chhetri, S. (2012). A New Generalized Cauchy Distribution with an Application to Annual One Day Maximum Rainfall Data, Statistics, Optimization and Information computing,9,123-136.

[20] Jayakumar, K., Bindu, K. and Hamedani. G.G.(2020). On a new generalization of Pareto distribution and its applications, Communication in statistics- simulation and computation, 49, 1264-1284.

[21] Nichols, M.D., Padgett, W.J. (2006). A bootstrap controll chart for Weibull percentiles, Quality and Reliability Engineering International, 22, 141-151.