A New Method for Generalizing Burr and Related Distributions

Tanujit Charkaborty\textsuperscript{a*}, Suchismita Das\textsuperscript{b}, Swarup Chattopadhyay\textsuperscript{c}

\textsuperscript{a}Statistical Quality Control and Operations Research Unit, Indian Statistical Institute, 203, B.T. Road, Kolkata-700108, India. (Mail: tanujit.r@isical.ac.in)
\textsuperscript{b}Department of Data Science, S P Jain School of Global Management, Mumbai-400070, India. (Mail: suchismita.das@spjain.org)
\textsuperscript{c}Machine Intelligence Unit, Indian Statistical Institute, 203, B.T. Road, Kolkata-700108, India. (Mail: swarup.r@isical.ac.in)

Abstract

A new method has been proposed to generalize Burr-XII distribution, also called Burr distribution, by adding an extra parameter to an existing Burr distribution for more flexibility. In this method, the exponent of the Burr distribution is modeled using a nonlinear function of the data and one additional parameter. The models of this newly introduced generalized Burr family can significantly increase the flexibility of the former Burr distribution with respect to the density and hazard rate shapes. Families expanded using the method proposed here is heavy-tailed and belongs to the maximum domain of attractions of the Frechet distribution. The method is further applied to yield three-parameter classical Pareto and generalized exponentiated distributions which shows the broader application of the proposed idea of generalization. A relevant model of the new generalized Burr family has been considered in detail, with particular emphasis on the hazard functions, stochastic orders, estimation procedures, and testing methods are derived. Finally, as empirical evidence, the new distribution is applied to the analysis of large-scale heavy-tailed network data and compared with other commonly used distributions available for fitting degree distributions of networks. Experimental results suggest that the proposed Burr distribution with nonlinear exponent better fits the large-scale heavy-tailed networks better than the popularly used Marshall-Olkin generalization of Burr and exponentiated Burr distributions.

Key Words: Burr distribution; exponentiated distributions; stochastic ordering; reliability properties; Maximum likelihood.

\textsuperscript{*}Corresponding author.
E-mail addresses: tanujit.r@isical.ac.in

© 2020 by the author(s). Distributed under a Creative Commons CC BY license.
1 Introduction

The development of statistical distributions is one of the oldest research topics in the field of statistics. There has been a renewed interest in developing more flexible statistical distributions in recent decades. Since the seminal work by Karl Pearson in 1895 [35], several general methods have been developed for generating family of distributions. Pearson presented a systematic approach for generating statistical distributions to model non-symmetric type of data using differential equation. The Pearson system of continuous distributions is a system for which every probability density function (PDF) $f(x)$ satisfies the differential equation of the following form:

$$\frac{1}{f(x)} \frac{df(x)}{dx} = \frac{a + x}{b_0 + b_1 x + b_2 x^2},$$

(1)

where $a, b_0, b_1,$ and $b_2$ are the shape parameters. Different types of distributions correspond to different forms of solution to Eqn. (1). The form of solution of Eqn. (1) depends on the root of the equation

$$b_0 + b_1 x + b_2 x^2 = 0.$$

Pearson presented four types of distributions [14] and characterizations of Pearson type distributions are also available in the literature [4]. Irving W. Burr proposed another important development in this category. In Burr’s method [9], system of distributions satisfy the following differential equation:

$$dF = F(1 - F)g(x)dx,$$

(2)

where $0 \leq F \leq 1$ and $g(x)$ is a non-negative function over $x$. Twelve different solutions to the Eqn. (2) in the form of cumulative distribution functions (CDF) were given and named as Burr Types I-XII distributions. The Burr type-XII distribution, being a member of the Burr system has gained more attention in the last decade due to its potential use in practical situations. The Burr Type-XII, simply called Burr distribution, is highly useful for fitting heavy-tailed data sets from the field of reliability, economics, hydrology, actuarial science, and network science among many others [23]. Burr distribution also emerges as a suitable model to describe stationary states of complex and non-equilibrium systems [37, 36]. The main advantage of Burr distribution from the extreme value statistics’ point of view is that it has algebraic tails which are effective for modeling failures that occur with lesser frequency than with corresponding models based on exponential tails [39]. The CDF and PDF of the Burr distribution are defined as follows.
Definition 1 A random variable $X$ follows Burr distribution with parameters $c$, $\lambda$, $\alpha$, if the CDF is of the following form:

$$F(x; c, \lambda, \alpha) = 1 - \left[1 + \left(\frac{x}{\lambda}\right)^c\right]^{-\alpha}, \quad x > 0, \quad c, \alpha, \lambda > 0,$$

where $\alpha$ and $c$ are shape parameters and $\lambda$ is a scale parameter. The density function of the Burr distribution is given by

$$f(x; c, \lambda, \alpha) = c\alpha\lambda^{-c}x^{c-1}\left[1 + \left(\frac{x}{\lambda}\right)^c\right]^{-\alpha-1}, \quad x > 0, \quad c, \alpha, \lambda > 0.$$ (4)

The corresponding survival function is given by

$$S(x; c, \lambda, \alpha) = \left[1 + \left(\frac{x}{\lambda}\right)^c\right]^{-\alpha}, \quad x > 0.$$ (5)

The hazard rate function is given by

$$h(x; c, \lambda, \alpha) = \frac{f(x; c, \lambda, \alpha)}{S(x; c, \lambda, \alpha)} = c\alpha\lambda^{-c}x^{c-1}\left[1 + \left(\frac{x}{\lambda}\right)^c\right]^{-1}, \quad x > 0.$$ (6)

It is interesting to note that the CDF in (3) is regularly varying at infinity, viz. they satisfy for some $\gamma > 0$, called the tail-index, and for all $t > 0$,

$$\lim_{x \to \infty} \frac{1 - F(tx)}{1 - F(x)} = t^{-\gamma}.$$

This result on the shape parameter of the Burr distribution suggests that it is well-suited for modeling extreme values as a heavy-tailed distribution. The bivariate and multivariate extensions of Burr distribution are available in the past literature [13, 40]. Consequently, numerous modifications or generalizations of the Burr distribution, using different compounding and weighting techniques, have been suggested. See, for instance, the beta Burr distribution introduced by [34], the McDonald Burr distribution introduced by [17], order-statistics based generalized Burr distribution [5], the Marshall-Olkin Burr (MO Burr) distribution developed by [22], modified Burr distribution [21] the exponentiated Burr distribution introduced by [25] and the log-Weibull Burr distribution developed by [1]. Though all these modifications and generalizations resulted well in specific data analysis but all of these assumed constant exponents for the Burr distribution which causes potential failure when working with large-scale heavy-tailed data sets from various applied domains [23]. In this work, the primary hypothesis is that the exponent $\alpha$ of the Burr distribution is not a constant and varies according to a nonlinear function $g$ which depends on the data.
This article aims to introduce an extra parameter to Burr, Pareto, and exponentiated distributions to bring more flexibility to the given families when dealing with real-world large-scale heavy-tailed data sets. We introduce a new method of generalization, namely the shape-parameter transformation (SPT) method, where the non-negative shape parameter is assumed to be expressible as a nonlinear function of the empirical data and adds an additional parameter to the base distribution. The proposed SPT method is straightforward to use; hence it can effectively be used for data analysis purposes. The proposed SPT method is first applied to a Burr distribution to yield generalized Burr (GBurr) family of distributions. The newly introduced GBurr family has some excellent statistical properties and belongs to the maximum domain of attractions of the Frechet distribution. The method presented in this paper adds an additional parameter in the distribution and can be a competitor to the popularly used Marshall-Olkin (MO) [28] and Lehmann alternatives-based [26] method of generalizing continuous distributions. The SPT method is further applied to the Pareto and the exponentiated family of distributions which shows the broader applicability of the proposed SPT method. Further, a specific nonlinear variant of the GBurr family, we call it NBurr distribution, is discussed in detail. Complementary theoretical aspects are studied, such as shapes, asymptotes, quantiles, stochastic ordering, reliability parameter, and inferential statistics. The application of the proposed NBurr distribution is shown using large-scale heavy-tailed network data sets from various disciplines.

The rest of the paper is organized as follows. We discuss the SPT method and its application to the Burr, Pareto, and exponentiated family of distributions in Section 2 along with some structural properties. In Section 3, we study a simple model from the GBurr family of distributions and discuss its stochastic and inferential characteristics. The empirical application of this new NBurr distribution to various large scale heavy-tailed network data sets is presented in Section 4. Finally, we conclude this paper in Section 5.

2 Shape Parameter Transformation (SPT) Method

In this section, we introduce the SPT method to generalize Burr, Pareto, and exponentiated distributions which model the shape parameter of these families as a nonlinear function of the data. It also adds an additional parameter to the families as mentioned above of distributions for flexible modeling of real-life data sets. The SPT method is first applied to yield generalized Burr distributions which are regularly varying distributions at
infinity and are heavy-tailed. Furthermore, this new SPT method is applied to Pareto distribution and exponentiated family of distributions to generalize these families.

2.1 The generalized Burr (GBurr) family

Definition 2 A continuous random variable \( X \) follows generalized Burr (Burr) family of distributions if and only if it has the following CDF:

\[
F_{\text{GBurr}}(x; c, \lambda, \alpha, \beta) = 1 - \left[ 1 + \left( \frac{x}{\lambda} \right)^c \right]^{-\alpha} g\left( \frac{x}{\lambda}; \alpha, \beta \right); \quad x \geq 0, \quad \alpha, \lambda, c > 0, \tag{7}
\]

and \( F(x) = 0 \) if \( x < 0 \), where the real valued continuous, positive function \( g : (0, \infty) \to \mathbb{R}^+ \) is differentiable on \((0, \infty)\). The shape parameter \( \alpha \) of the Burr distribution is replaced with \( g(z) = g(x/\lambda; \alpha, \beta) = g(x/\lambda) \), say, where \( \beta \) is an additional shape parameter and \( g(z) \) satisfies the following conditions:

1. The function \( g(z) \) is strictly positive and have finite limit at infinity, viz.

\[
\lim_{z \to \infty} g(z) = \alpha (> 0). \tag{8}
\]

2. \( \lim_{z \to 0^+} (1 + z^c)^{g(z)} = 1 \) and \( \lim_{z \to \infty} (1 + z^c)^{g(z)} = \infty. \)

3. \( \frac{g'(z)}{g(z)} \geq -\frac{cz^{c-1}}{(1 + z^c) \log(1 + z)} \), \( z > 0 \), where \( g'(z) = \frac{d}{dz} |g(z)|. \)

It is very easy to verify that:

1. \( F_{\text{GBurr}}(x) \) is non-decreasing since,

\[
c \left( \frac{x}{\lambda} \right)^{c-1} g\left( \frac{x}{\lambda} \right) + g'\left( \frac{x}{\lambda} \right) \left( 1 + \left( \frac{x}{\lambda} \right)^c \right) \log \left( 1 + \frac{x}{\lambda} \right) > 0, \quad x > 0. \]

2. \( \lim_{x \to -\infty} F_{\text{GBurr}}(x) = F_{\text{GBurr}}(-\infty) = 0, \lim_{x \to \infty} F_{\text{GBurr}}(x) = 1. \)

3. \( F_{\text{GBurr}}(x) \) is right continuous.

Thus, (7) is a standard CDF and it can also be expressed as follows:

\[
F_{\text{GBurr}}(x) = 1 - \exp \left[ -g\left( \frac{x}{\lambda} \right) \log \left( 1 + \left( \frac{x}{\lambda} \right)^c \right) \right], \quad x > 0. \tag{9}
\]
The corresponding survival function is given by

\[ S_{\text{GBurr}}(x) = \begin{cases} 
1 + \left(\frac{x}{\lambda}\right)^c \exp\left(\frac{c g(x/\lambda)}{\log(1 + x c \lambda)}\right), & x > 0 \\
1, & x \leq 0.
\end{cases} \]  

(10)

The probability density function is

\[ f_{\text{GBurr}}(x) = \begin{cases} 
\frac{c}{\lambda} \left[ \frac{e^{\frac{c}{\lambda}}}{1 + \left(\frac{x}{\lambda}\right)^c} g\left(\frac{x}{\lambda}\right) + \frac{1}{\lambda} g'\left(\frac{x}{\lambda}\right) \log \left(1 + \left(\frac{x}{\lambda}\right)^c\right) \right] \left[ 1 + \left(\frac{x}{\lambda}\right)^c \right]^{-\frac{c g(x/\lambda)}{\log(1 + x c \lambda)}}, & x > 0 \\
0, & x \leq 0.
\end{cases} \]  

(11)

The hazard rate function is given by

\[ h_{\text{GBurr}}(x) = \begin{cases} 
\frac{c}{\lambda} \left(\frac{x}{\lambda}\right)^{c-1} g\left(\frac{x}{\lambda}\right) + \frac{1}{\lambda} g'\left(\frac{x}{\lambda}\right) \log \left(1 + \left(\frac{x}{\lambda}\right)^c\right), & x > 0 \\
0, & x \leq 0.
\end{cases} \]  

(12)

The CDF (7) of the GBurr family of distributions is a function with regularly varying tails and belongs to the maximum domain of attraction (MDA) of the Frechet distribution with index \( \alpha > 0 \), viz. \( F \in MDA(\Phi_\alpha) \), with \( \Phi_\alpha = \exp\{-x^{-\alpha}\}, x > 0, \alpha > 0 \) [15].

**Theorem 1** GBurr family of distributions are: (a) heavy-tailed; (b) right tail-equivalent to a Pareto distribution; and (c) belongs to the MDA of the Frechet distribution.

**Proof:** (a) For the GBurr family of distributions, we note that

\[ \lim_{x \to \infty} \exp\{kx\} (1 - F(x)) = \lim_{x \to \infty} \exp\left[kx - g\left(\frac{x}{\lambda}\right) \log \left(1 + \frac{x^c}{\lambda^c}\right)\right] = \infty, \]

where \( k, \lambda, c > 0 \).

(b) To show the tail-equivalent property, we show

\[ \lim_{x \to \infty} \frac{1 - F(x)}{1 - G(x)} = \lim_{x \to \infty} \frac{\exp\left[-g\left(\frac{x}{\lambda}\right) \log \left(1 + \frac{x^c}{\lambda^c}\right)\right]}{\exp\left[-\alpha \log \left(1 + \frac{x^c}{\lambda^c}\right)\right]} = 1, \forall \lambda, c > 0, \]

where \( G(x) \) is the CDF of the Pareto Type-II distribution.
(c) It should be noted that any function \( g \) as defined in Definition 2 satisfying \( \lim_{z \to \infty} g(z) = \alpha > 0 \), is slowly varying at infinity:

\[
\lim_{z \to \infty} \frac{g(tz)}{g(z)} = 1, \forall \, t > 0.
\]

Now, \( \lim_{x \to \infty} \frac{1 - F(tx)}{1 - F(x)} = \lim_{x \to \infty} \frac{\left(1 + \frac{tx^c}{\lambda^c}\right)^{-g\left(\frac{x}{\lambda}\right)} - g\left(\frac{z}{\lambda}\right)}{\left(1 + \frac{x^c}{\lambda^c}\right)^{-g\left(\frac{z}{\lambda}\right)}} = t^{-\alpha}, \forall \, t > 0 \) and \( \lambda, c > 0 \).

It can be seen that Burr distribution belongs to this GBurr family and corresponds to the simplest choice \( g(z) = \alpha \). We give some examples of nonlinear \( g(z) \) satisfying the condition given in Definition 2 and present some distributions belonging to the GBurr family in Table 1. In order to select a model, one can choose the nonlinear exponent \( g(z) \) that meets the empirical characteristics of the given data sets. Obviously, this example list of GBurr models can further be expanded to more models by introducing some other forms of \( g(z) \) satisfying \( \lim_{z \to \infty} g(z) = \alpha > 0 \) or by increasing the number of parameters in the function \( g \).

<table>
<thead>
<tr>
<th>( g(z) ), ( \forall , z &gt; 0 )</th>
<th>( f(z) ), ( \forall , x &gt; 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha - \frac{\beta}{x + 1} )</td>
<td>( \beta \leq \alpha )</td>
</tr>
<tr>
<td>( \alpha \left[ \frac{x}{\lambda^c} \right]^\beta )</td>
<td>( \beta &gt; -1 )</td>
</tr>
<tr>
<td>( \alpha \left[ \frac{\log(1 + x)}{1 + \log(1 + x)} \right]^{\beta} )</td>
<td>( \beta &gt; -1 )</td>
</tr>
</tbody>
</table>

We can also obtain some popular size distributions as the particular cases of the GBurr family:

- By taking the constant \( g \) function, viz., when \( g(z) = 1 \), then GBurr distributions reduces to the Fisk distribution [16].
- For \( \alpha \to \infty \) and \( \beta = 0 \), GBurr family of distributions converges to the Weibull distribution [24].
- By taking \( c = 1 \) and \( \beta = 0 \), GBurr distribution reduces to the Lomax distribution [27].
2.2 Generalized Classical Pareto Distribution

In this section, we apply the SPT method to the classical Pareto distribution to yield generalized classical Pareto (GCP) distributions as defined below.

**Definition 3** A continuous random variable \(X\) follows GCP family of distributions if and only if it has the following CDF:

\[
F_{GCP}(x) = 1 - \left( \frac{x}{\gamma} \right)^{-m\left( \frac{x}{\gamma} \right)}; \quad x \geq \gamma
\]

(13)

and \(F(x) = 0\) if \(x \leq \gamma\), \(\gamma > 0\) is a scale parameter, \(m : (1, \infty) \to \mathbb{R}^+\) is a real, continuous, positive function which is differentiable on \((1, \infty)\), and the function \(m\) satisfies the following conditions:

1. \(m\) is strictly positive and have finite limit at infinity, viz.

\[
\lim_{z \to \infty} m(z) = \alpha > 0.
\]

(14)

2. \(\lim_{z \to 1^+} z^m(z) = 1\) and \(\lim_{z \to \infty} z^m(z) = \infty\).

3. \(\frac{m'(z)}{m(z)} > -\frac{1}{z \log(z)}, \quad z > 1\).

It is noted that the condition (3) is equivalent to

\[
\frac{d}{dz} \log(g(z)) \geq -\frac{d}{dz} \log[\log(z)], \quad z > 1.
\]

It is very easy to verify that \(F(x)\) in (13) is a standard distribution function and any continuous random variable \(X\) satisfying the above-mentioned conditions are called as GCP distributions. An alternative expression for \(F(x)\) can also be given as follows:

\[
F_{GCP}(x) = 1 - \exp\left[-m\left( \frac{x}{\gamma} \right) \log\left( \frac{x}{\gamma} \right)\right], \quad x \geq \gamma.
\]

The survival function is given by

\[
S_{GCP}(x) = \begin{cases} \left( \frac{x}{\gamma} \right)^{-m\left( \frac{x}{\gamma} \right)} & \text{if } x > \gamma \\ 1 & \text{if } x \leq \gamma. \end{cases}
\]

(15)
The PDF for the family of GCP distributions is given by
\[
f_{GCP}(x) = \begin{cases} 
m \left( \frac{x}{\gamma} \right) + m' \left( \frac{x}{\gamma} \right) \left( \frac{x}{\gamma} \right) \log \left( \frac{x}{\gamma} \right) \frac{S_{GCP}(x)}{x} & \text{if } x > \gamma \\
0 & \text{if } x \leq \gamma. \end{cases} \tag{16}
\]

The hazard function is given by
\[
h_{GCP}(x) = \begin{cases} 
g \left( \frac{x}{\gamma} \right) + m' \left( \frac{x}{\gamma} \right) \left( \frac{x}{\gamma} \right) \log \left( \frac{x}{\gamma} \right) \frac{x}{\gamma} & \text{if } x > \gamma \\
0 & \text{if } x \leq \gamma. \end{cases} \tag{17}
\]

Table 2 shows some examples of GCP distributions satisfying condition (14) (i.e., \( \lim_{x \to \infty} m(z) = \alpha > 0 \)). It is noted that the simplest choice \( m(z) = \alpha \) leads the classical Pareto distribution. The rest of the models are completely new. In order to select a model, one can choose the nonlinear exponent \( m(z) \) that meets the empirical characteristics of the given data sets. Obviously, this example list of GCP models can further be expanded to more models by introducing some other forms of \( m(z) \) satisfying \( \lim_{x \to \infty} m(z) = \alpha > 0 \) or by increasing the number of parameters in the function \( g \).

### Table 2

Some examples of GCP family of distributions with \( \lim_{x \to \infty} m(z) = \alpha > 0 \).

<table>
<thead>
<tr>
<th>( m(z) ), ( \forall x &gt; 1 )</th>
<th>( \beta )</th>
<th>( f(x), \forall x &gt; \gamma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha - \frac{\beta}{z} )</td>
<td>( \beta \leq \alpha )</td>
<td>( \left( \frac{x}{\gamma} \right)^{-\alpha + \beta / x} \left[ \frac{\alpha}{x} + \frac{\gamma \beta}{x} \left( \frac{\log(x/\gamma) - 1}{x} \right) \right] )</td>
</tr>
<tr>
<td>( \alpha \left( \frac{x - 1}{z} \right) )</td>
<td>( \beta &gt; -1 )</td>
<td>( \frac{\alpha}{z} \left( \frac{x}{\gamma} \right)^{-\alpha} \left[ \frac{(x/\gamma - 1)^{\beta}}{(x/\gamma - 1)} \right] \left[ \frac{(x/\gamma - 1 + \beta \log(x/\gamma))}{(x/\gamma)} \right] )</td>
</tr>
<tr>
<td>( \alpha \left[ \frac{\log(x)}{1 + \log(x)} \right] )</td>
<td>( \beta &gt; -1 )</td>
<td>( \frac{\alpha}{z} \left( \frac{x}{\gamma} \right)^{-\alpha} \left[ \frac{\log(x/\gamma)}{\log(x/\gamma) + 1} \right]^\beta \left[ \frac{\log(x/\gamma) + 1 + \beta \log(x/\gamma)}{\log(x/\gamma) + 1} \right] )</td>
</tr>
</tbody>
</table>

#### 2.3 Generalized Exponentiated Distributions

The idea of exponentiated family of distributions is based on Lehmann alternatives [26] that can add a non-negative shape parameter to any continuous probability distributions:
\[
F^*(x) = [F(x)]^\alpha, \text{ where } \alpha > 0 \text{ (real)}, \tag{18}
\]

where \( F(x) \) is a standard CDF. This new family \( F^* \) is called ‘exponentiated family’ where one raises the CDF of an existing distribution to a power.
of an additional parameter. Lehmann motivated it in the following way: 
\[ (F(x))^\alpha \] is the distribution of the maximum of \( \alpha \) independent and identically distributed variables with distribution \( F \) when \( \alpha \) is an integer and when \( \alpha \) is a rational number, \( (F(x))^\alpha \) is one-parameter family of nonparametric class of alternatives. In the theory of generalized probability distributions, Lehmann alternatives has been widely used to generate so-called exponentiated Weibull family of distribution by [29]. A systematic treatment of exponentiated weibull, exponentiated gamma, and exponentiated pareto distribution are available in [18]. Gupta and Kundu studied exponentiated exponential distribution [20] and recently Gupta et al. [19] proposed Power Normal distributions using the same idea of Lehmann’s alternatives for standard normal distribution. Nadarajah and Kotz [30] studied a list of exponentiated X family of distributions, including exponentiated gumbel, frechet, gamma, etc distributions. We use our SPT method to generalize any ‘exponentiated family’ of distributions.

**Definition 4** Any continuous random variable \( X \) with distribution function \( F(x) \) follows generalized exponentiated distribution if and only if it has the following CDF 
\[ G(x) = (F(x))^m(x); -\infty < x < \infty, \]
where \( m : (-\infty, \infty) \rightarrow (0, \infty) \) is a real, continuous function which is differentiable on \((0, \infty)\). The function \( m \) satisfies the following conditions:

1. \( m \) is strictly positive and have finite limit at infinity, viz. 
\[ \lim_{x \to \infty} m(x) = \alpha (> 0). \]

2. \( \lim_{x \to -\infty} m(x) = \beta (> \alpha). \)

3. \( m(x) < 0. \)

It is important to note that when \( m(x) = \alpha \), then \( G(x) \) corresponds to the exponentiated family of distributions [18, 30]. We give some examples of \( m(x) \) to be used for the distributions with \( x \in \mathbb{R} \).

**Example 1** Let \( m(x) \) be a function of the following form satisfying \( \lim_{x \to -\infty} \xi(x) = 2 \) and \( \lim_{x \to \infty} \xi(x) = 0 \):
\[ m(x) = \frac{(\beta - \alpha)}{2} \xi(x) + \alpha; \beta > \alpha > 0. \]
Some choices of $\xi(x)$ are as follows:

\begin{align*}
(a) \quad & \xi(x) = \begin{cases} 
e^{-x} & \text{if } x \geq 0 \\ 2 - e^x & \text{if } x < 0 \end{cases} \\
(b) \quad & \xi(x) = \begin{cases} \frac{1}{x^{2k+1}+1} & \text{if } x \geq 0 \\ 2 + \frac{1}{x^{2k+1}-1} & \text{if } x < 0 \end{cases} \\
(c) \quad & \xi(x) = \begin{cases} \frac{1}{\log(1+x)+1} & \text{if } x \geq 0 \\ 2 - \frac{1}{\log(1-x)+1} & \text{if } x < 0 \end{cases}
\end{align*}

where $k$ is any natural numbers.

**Example 2** Let $m(x)$ be a function of the following form satisfying $\lim_{x \to -\infty} \sigma(x) = 0$ and $\lim_{x \to \infty} \sigma(x) = 1$:

$$m(x) = \alpha \sigma(x) + \beta (1 - \sigma(x)); \quad \beta > \alpha > 0.$$  

Some choices of $\sigma(x)$ are as follows:

\begin{align*}
(a) \quad & \sigma(x) = \frac{1}{1 + e^{-x}}; \quad x \in \mathbb{R} \\
(b) \quad & \sigma(x) = \frac{1}{2} + \frac{1}{\pi} \arctan(x); \quad x \in \mathbb{R} \\
(c) \quad & \sigma(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp \left(-\frac{y^2}{2}\right) dy; \quad x \in \mathbb{R}
\end{align*}

It is interesting to see that

$$G(x) = F(x)^{\alpha \Phi(x) + \beta \Phi(-x)}; \quad x \in \mathbb{R},$$

is a standard CDF, where $\Phi(x)$ is a standard normal CDF. For all the above examples when $\alpha = \beta$, we get $G(x) = [F(x)]^\alpha$.

3 **NBurr Distribution: Definition and Properties**

In this section, we study a relevant model of the GBurr family of distributions with the choice of $g(z) = \alpha - \frac{\beta}{1+z}$ in Table 1. The CDF is given by

$$F_{NBurr}(x) = 1 - \left[1 + \left(\frac{x}{\lambda}\right)^c\right]^{-\left(\frac{\alpha}{1+c}\right)}, \quad x > 0,$$  

(19)
where $\alpha, \lambda, c > 0$, $\beta > -1$ and $\alpha > \beta$. We call this simple generalized form of Burr distribution as NBurr distribution (Burr distribution with nonlinear exponent). The NBurr distribution includes the Burr distribution when $\beta = 0$. The CDF in (19) can alternatively be written in the following form:

$$F_{NBurr}(x) = 1 - \exp \left[ - \left( \frac{\beta}{1 + \frac{x}{\lambda}} \right) \log \left( 1 + \left( \frac{x}{\lambda} \right)^c \right) \right], \ x > 0. \quad (20)$$

The corresponding survival function is given by,

$$S_{NBurr}(x) = \left[ 1 + \left( \frac{x}{\lambda} \right)^c \right]^{-\frac{\beta}{1 + \frac{x}{\lambda}}}, \ x > 0. \quad (21)$$

**Figure 1:** Plots of PDFs of the NBurr distribution.

The probability density function for $x > 0$ is given by

$$f_{NBurr}(x) = \left[ \frac{\beta \lambda}{(x + \lambda)^2} \log \left( 1 + \left( \frac{x}{\lambda} \right)^c \right) \right. + \left. \left( \alpha - \frac{\beta \lambda}{x + \lambda} \right) \frac{cx^{c-1}}{x^c + \lambda^c} \right] S_{NBurr}(x), \ x > 0, \quad (22)$$
and the hazard rate function is given by

\[ h_{NBurr}(x) = \frac{\beta \lambda}{(x + \lambda)^2} \log \left( 1 + \left( \frac{x}{\lambda} \right)^c \right) + \left( \alpha - \frac{\beta \lambda}{x + \lambda} \right) \frac{cx^{c-1}}{x^c + \lambda^c}, \quad x > 0. \]  

The proposed NBurr distribution satisfies the extreme value properties as given in Theorem 1. Some graphics of the NBurr model derived from this newly introduced GBurr family of distributions are illustrated in Figure 1. Remark that NBurr distribution with parameters \( \alpha, \beta, \lambda, c \), as a distribution of the MDA of the Frechet distribution, satisfies the von Mises condition:

\[ \lim_{x \to \infty} \frac{x f_{NBurr}(x)}{S_{NBurr}(x)} = \alpha > 0. \]

Next, we study the reliability properties and inferential properties of the new NBurr distributions in the next subsections.

### 3.1 Reliability properties of NBurr distribution

In this section, we study some reliability properties of the newly introduced NBurr distribution including monotonicity of hazard rates, stochastic orderings, entropy, etc. The following theorem shows that the hazard rate function of the NBurr distribution is increasing and decreasing under certain conditions on \( c \) and \( \beta \).

**Theorem 2** The hazard rate function of the NBurr distribution satisfies the following properties:

1. If \( 0 \leq c \leq 1 \) and \( \beta > 0 \) then for all \( x > 0 \), \( h_{NBurr}(x) \) is decreasing in \( x \);

2. If \( c > 1 \) and \( \beta > 0 \), then for all \( x > 0 \),
   - (a) \( h_{NBurr}(x) \) is increasing in \( x \), when \( c - 1 > \left( \frac{x}{\lambda} \right)^c \);
   - (b) \( h_{NBurr}(x) \) is decreasing in \( x \), when \( c - 1 < \left( \frac{x}{\lambda} \right)^c \);
   - (c) \( h_{NBurr}(x) \) is maximum at \( x = \lambda \left( c - 1 \right)^{1/c} \).
Proof: Differentiating (23) with respect to $x$, we have

$$h'_{NBurr}(x) = -\frac{2\beta \lambda}{(x+\lambda)^2} \log \left( \frac{x^c}{\lambda^c} \right) + \frac{2c\beta \lambda x^{c-1}}{(x+\lambda)^2 (x^c + \lambda^c)}$$

$$+ \left( \alpha - \frac{\beta \lambda}{x+\lambda} \right) \frac{cx^{c-2}}{x^c + \lambda^c} \left[ c - 1 - \frac{x^c}{\lambda^c} \right]$$

$$= \frac{2\beta \lambda}{(x+\lambda)^2} \left[ cx^{c-1} - \frac{x + \lambda}{x^c + \lambda^c} \log \left( 1 + \frac{x^c}{\lambda^c} \right) \right]$$

$$+ \left( \alpha - \frac{\beta \lambda}{x+\lambda} \right) \frac{cx^{c-2}}{(x^c + \lambda^c)^2} \left[ c - 1 - \frac{x^c}{\lambda^c} \right].$$

Now, $\alpha \geq \frac{\beta \lambda}{x+\lambda} \frac{c^c x^{c-2}}{(x^c + \lambda^c)^2} \geq 0$ and if $\beta > 0$, $\frac{2\beta \lambda}{(x+\lambda)^2 (x^c + \lambda^c)} > 0$. Thus, $h'_{NBurr}(x) \geq (\leq) 0$ if

(i) $A(x) = cx^{c-1} - \frac{x + \lambda}{x^c + \lambda^c} \log \left( 1 + \frac{x^c}{\lambda^c} \right) \geq (\leq) 0$;

(ii) $c - 1 - \frac{x^c}{\lambda^c} \geq (\leq) 0$.

Now,

$$A'(x) = \frac{c\lambda^c x^{c-2}(x+\lambda)}{(x^c + \lambda^c)^2} \left[ c - 1 - \frac{x^c}{\lambda^c} \right].$$

We can see that $A'(x) \leq 0$ if $0 \leq c \leq 1$. Therefore, $A(x)$ is decreasing in $x$ and again $A(0) = 0$. Thus, we have $A(x) \leq 0$.

Similarly we can prove that if $c > 1$, then $A(x) \geq 0$ when $c - 1 > \left( \frac{x}{\lambda} \right)^c$ and $A(x) \leq 0$ when $c - 1 < \left( \frac{x}{\lambda} \right)^c$. This implies that for $0 \leq c \leq 1$, $h_{NBurr}(x)$ is decreasing in $x$; for $c > 1$, $h_{NBurr}(x)$ is increasing in $x$, when $c - 1 > \left( \frac{x}{\lambda} \right)^c$ and $h_{NBurr}(x)$ is decreasing in $x$, when $c - 1 < \left( \frac{x}{\lambda} \right)^c$. Thus, the hazard function $h_{NBurr}(x)$ attain its maximum at $x = \lambda [c - 1]^{1/c}$.

The following example shows that the NBurr distribution does not preserves the likelihood ratio ordering. It is useful to recall that a random variable $X$ is said to be larger than another random variable $Y$ in likelihood ratio ordering (written as $X \geq_{LR} Y$) if $f_X(x)/f_Y(x)$ is an increasing function for $x > 0$.

**Example 3** Let $X$ and $Y$ be two random variables following NBurr distributions with parameters $\alpha_1, \beta_1, \lambda_1, e_1$ and $\alpha_2, \beta_2, \lambda_2, e_2$ respectively. Then,
for all \( x > 0 \), the ratio of the corresponding density functions of \( X \) and \( Y \) is given by

\[
\frac{f_{X|NBurr}(x)}{f_{Y|NBurr}(x)} = \frac{\frac{\beta_1 \lambda_1}{(x + \lambda_1)^2} \log \left(1 + \left(\frac{x}{\lambda_1}\right)^c_1\right) + \left(\alpha_1 - \frac{\beta_1 \lambda_1}{1 + \lambda_1}\right) \frac{c_1 x^{c_1-1}}{x^{c_1+1}} \left[1 + \left(\frac{x}{\lambda_1}\right)^c_1\right] - \left(\alpha_1 - \frac{\beta_1}{1 + \lambda_1}\right)}{\frac{\beta_2 \lambda_2}{(x + \lambda_2)^2} \log \left(1 + \left(\frac{x}{\lambda_2}\right)^c_2\right) + \left(\alpha_2 - \frac{\beta_2 \lambda_2}{1 + \lambda_2}\right) \frac{c_2 x^{c_2-1}}{x^{c_2+1}} \left[1 + \left(\frac{x}{\lambda_2}\right)^c_2\right] - \left(\alpha_2 - \frac{\beta_2}{1 + \lambda_2}\right)}.
\]

Now, for \( \lambda_1 = \lambda_2 = 1 \),

Case I: when \( c_1 = c_2 = 1 \), (24) reduces to

\[
\frac{f_{X|NBurr}(x)}{f_{Y|NBurr}(x)} = \frac{\frac{\beta_1}{(1+x)^2} \log (1 + x) + \left(\alpha_1 - \frac{\beta_1}{1+x}\right) \frac{1}{1+x} \left(1 + x\right) - \left(\alpha_1 - \frac{\beta_1}{1+x}\right)}{\frac{\beta_2}{(1+x)^2} \log (1 + x) + \left(\alpha_2 - \frac{\beta_2}{1+x}\right) \frac{1}{1+x} \left(1 + x\right) - \left(\alpha_2 - \frac{\beta_2}{1+x}\right)} = P_1(x), \text{say.}
\]

We can see that for \( \alpha_1 = 1.5, \alpha_2 = 2, \beta_1 = 0.5 \) and \( \beta_2 = 1.5, P_1(0.1) = 1.366, P_1(2) = 0.889 \) and \( P_1(6) = 1.426 \), which implies that \( \frac{f_{X|NBurr}(x)}{f_{Y|NBurr}(x)} \) is not monotone. Thus \( X \not\geq_{LR} Y \).

Case II: when \( c_1 \neq c_2 \), (24) reduces to

\[
\frac{f_{X|NBurr}(x)}{f_{Y|NBurr}(x)} = \frac{\frac{\beta_1}{(1+x)^2} \log (1 + x^{c_1}) + \left(\alpha_1 - \frac{\beta_1}{1+x}\right) \frac{c_1 x^{c_1-1}}{1+x^{c_1+1}} \left[1 + x^{c_1}\right] - \left(\alpha_1 - \frac{\beta_1}{1+x}\right)}{\frac{\beta_2}{(1+x)^2} \log (1 + x^{c_2}) + \left(\alpha_2 - \frac{\beta_2}{1+x}\right) \frac{c_2 x^{c_2-1}}{1+x^{c_2+1}} \left[1 + x^{c_2}\right] - \left(\alpha_2 - \frac{\beta_2}{1+x}\right)} = P_2(x), \text{say.}
\]

Again, we can see that for \( c_1 = 2, c_2 = 4, \alpha_1 = 1.5, \alpha_2 = 2, \beta_1 = 0.5 \) and \( \beta_2 = 1.5, P_2(0.5) = 1.5753, P_2(1) = 0.4843 \) and \( P_2(2) = 2.8757 \), which implies that \( \frac{f_{X|NBurr}(x)}{f_{Y|NBurr}(x)} \) is not monotone. Thus \( X \not\geq_{LR} Y \).

Hence, it can be concluded that the proposed NBurr distribution does not preserves the likelihood ratio ordering.

In the next theorem, we show that the NBurr distribution preserves the usual stochastic ordering. It is useful to remind that a random variable \( X \) is said to be larger than another random variable \( Y \) in usual stochastic ordering (written as \( X \geq_{ST} Y \)) if \( S_X(x) \geq S_Y(x) \), for all \( x > 0 \).
Theorem 3 Let $X$ and $Y$ be two random variables following NBurr distribution with parameters $\alpha_1, \beta_1, \lambda_1, c_1$ and $\alpha_2, \beta_2, \lambda_2, c_2$ respectively. Then, $X \geq_{ST} (\leq_{ST}) Y$ provided (i) $\lambda_1 \geq (\leq) \lambda_2$, (ii) $c_1 \leq (\geq) c_2$, (iii) $\alpha_1 \leq (\geq) \alpha_2$ and (iv) $\beta_1 \leq (\geq) \beta_2$, for $\beta > 0$.

Proof: $X \geq_{ST} (\leq_{ST}) Y$ if and only if for all $x > 0$, $S_{X_{NBurr}}(x) \geq (\leq) S_{Y_{NBurr}}(x)$, which is equivalent to

$$\frac{\left[1 + \left(\frac{x}{\lambda_1}\right)^{c_2}\right]^{\alpha_2} - \frac{\beta_2 \lambda_2}{x + \lambda_2}}{\left[1 + \left(\frac{x}{\lambda_1}\right)^{c_2}\right]^{\alpha_1} - \frac{\beta_1 \lambda_1}{x + \lambda_1}} \geq (\leq) 1.$$

This holds if (i) $\lambda_1 \geq (\leq) \lambda_2$, (ii) $c_1 \leq (\geq) c_2$, (iii) $\alpha_1 \leq (\geq) \alpha_2$ and (iv) $\beta_1 \leq (\geq) \beta_2$, for $\beta > 0$.

The following theorem gives the condition under which the NBurr distribution preserves the hazard rate ordering. It is well known that a random variable $X$ is said to be larger than another random variable $Y$ in hazard rate ordering (written as $X \geq_{HR} Y$) if, for all $x > 0$, $h_X(x) \leq h_Y(x)$.

Theorem 4 Let $X$ and $Y$ be two random variables following NBurr distribution with parameters $\alpha_1, \beta_1, \lambda_1, c_1$ and $\alpha_2, \beta_2, \lambda_2, c_2$ respectively. Then, $X \geq_{HR} (\leq_{HR}) Y$ provided (i) $\lambda_1 \geq (\leq) \lambda_2$, (ii) $c_1 \leq (\geq) c_2$, (iii) $\alpha_1 \leq (\geq) \alpha_2$ and (iv) $\beta_1 \leq (\geq) \beta_2$, for $\beta > 0$.

Proof: $X \geq_{HR} (\leq_{HR}) Y$ if and only if for all $x > 0$, $h_{X_{NBurr}}(x) \leq (\geq) h_{Y_{NBurr}}(x)$, which is equivalent to

$$\frac{\beta_1}{\lambda_1 \left(1 + \frac{x}{\lambda_1}\right)^2 \log \left(1 + \left(\frac{x}{\lambda_1}\right)^{c_1}\right) + \left(\alpha_1 - \frac{\beta_1}{1 + \frac{x}{\lambda_1}}\right) \frac{c_1 \left(\frac{x}{\lambda_1}\right)^{c_1 - 1}}{\lambda_1 \left(1 + \left(\frac{x}{\lambda_1}\right)^{c_1}\right)}} \leq (\geq)$$

$$\frac{\beta_2}{\lambda_2 \left(1 + \frac{x}{\lambda_2}\right)^2 \log \left(1 + \left(\frac{x}{\lambda_2}\right)^{c_2}\right) + \left(\alpha_2 - \frac{\beta_2}{1 + \frac{x}{\lambda_2}}\right) \frac{c_2 \left(\frac{x}{\lambda_2}\right)^{c_2 - 1}}{\lambda_2 \left(1 + \left(\frac{x}{\lambda_2}\right)^{c_2}\right)}.$$  

This holds if (i) $\lambda_1 \geq (\leq) \lambda_2$, (ii) $c_1 \leq (\geq) c_2$, (iii) $\alpha_1 \leq (\geq) \alpha_2$ and (iv) $\beta_1 \leq (\geq) \beta_2$, for $\beta > 0$.

Remark 1 It is interesting to note that the hazard rate function of the NBurr distribution can be both increasing and decreasing under certain conditions as given in Theorem 2. Also, NBurr distribution preserves stochastic and hazard rate orderings as shown in Theorem 3 and 4, respectively.
The Shannon entropy is a very important and well known concept in information theory as well as engineering sciences. Let $X$ be a random variable that follows NBurr distributions with parameters $\alpha, \beta, \lambda, c$. Then the Shannon’s entropy for the NBurr distribution is defined as

$$H(X) = -E[\log f_{NBurr}(x)] = -\int_{0}^{\infty} f_{NBurr}(x) \log f_{NBurr}(x) dx$$

$$= -\int_{0}^{\infty} \left( \frac{\beta \lambda}{(x+\lambda)^2} \log \left( 1 + \frac{x^c}{\lambda^c} \right) + \left( \alpha - \frac{\beta \lambda}{x+\lambda} \right) \frac{cx^{c-1}}{x^c + \lambda^c} \right) \left[ 1 + \left( \frac{x}{\lambda} \right)^c \right]^{-\left( \frac{\alpha}{1+\frac{\beta}{\lambda}} \right)} dx.$$

When we want to study the system that survived up to an age $t$, then Shannon’s entropy function is not useful in measuring the uncertainty about the residual lifetime of the system. Ebrahimi [3] has introduced residual entropy and defined as

$$H(X; t) = 1 - \int_{t}^{\infty} f_X(x) \log \left( \frac{f_X(x)}{S_X(x)} \right) dx,$$

Then the residual entropy for NBurr distributions with parameters $\alpha, \beta, \lambda, c$ is given by

$$H(X; t) = 1 - \frac{1}{S_{NBurr}(t)} \int_{t}^{\infty} f_{NBurr}(x) \log (h_{NBurr}(x)) dx$$

$$= 1 - \left[ 1 + \left( \frac{t}{\lambda} \right)^c \right]^{-\left( \frac{\alpha}{1+\frac{\beta}{\lambda}} \right)}$$

$$\int_{t}^{\infty} \left[ \frac{\beta \lambda}{(x+\lambda)^2} \log \left( 1 + \frac{x^c}{\lambda^c} \right) + \left( \alpha - \frac{\beta \lambda}{x+\lambda} \right) \frac{cx^{c-1}}{x^c + \lambda^c} \right] \left[ 1 + \left( \frac{x}{\lambda} \right)^c \right]^{-\left( \frac{\alpha}{1+\frac{\beta}{\lambda}} \right)} dx.$$

### 3.2 Parameter Estimation

Let $x_1, x_2, ..., x_n$ be a sample of size $n$ from $NBurr(\alpha, \beta, \lambda, c)$ distribution. We give procedure for parameter estimation the including the log-likelihood functions and corresponding normal equations. The log-likelihood function for the vector of parameters $\Theta = (\alpha, \beta, \lambda, c)^T$ corresponding to NBurr dis-
tributions is given by

\[ \ell \equiv \ell(x; \alpha, \beta, \lambda, c) = \sum_{i=1}^{n} \log \left[ \frac{\beta \lambda}{(\lambda + x_i)^2} \log \left( 1 + \frac{x_i^c}{\lambda^c} \right) + \left( \alpha - \frac{\beta \lambda}{\lambda + x_i} \right) c x_i^{c-1} \right] - \alpha \sum_{i=1}^{n} \log \left( 1 + \frac{x_i^c}{\lambda^c} \right) + \beta \lambda \sum_{i=1}^{n} \frac{1}{\lambda + x_i} \log \left( 1 + \frac{x_i^c}{\lambda^c} \right), \quad (25) \]

where \( n \) is the sample size, and the maximum likelihood estimates of the unknown parameter vector \((\alpha, \beta, \lambda, c)\) are those that maximize the log-likelihood function \(\ell\) in (25). The normal equations can be obtained by taking the partial derivatives of (25) w.r.t. \(\alpha, \beta, \lambda, c\) and equating them to zero:

\[ \frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^{n} \frac{c \omega_i^2 (\omega_i - 1)^{c-1}}{\beta (1 + (\omega_i - 1)^c) \log (1 + (\omega_i - 1)^c) + c \omega_i (\alpha \omega_i - \beta) (\omega_i - 1)^c} - \sum_{i=1}^{n} \log (1 + (\omega_i - 1)^c) = 0 \quad (26) \]

\[ \frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{n} \frac{(1 + (\omega_i - 1)^c) \log (1 + (\omega_i - 1)^c) - c \omega_i (\omega_i - 1)^c}{\beta (1 + (\omega_i - 1)^c) \log (1 + (\omega_i - 1)^c) + c \omega_i (\alpha \omega_i - \beta) (\omega_i - 1)^c} + \sum_{i=1}^{n} \frac{1}{\omega_i} \log (1 + (\omega_i - 1)^c) = 0 \quad (27) \]

\[ \frac{\partial \ell}{\partial \lambda} = \sum_{i=1}^{n} \frac{(\beta x_i - \beta \lambda) (\lambda^c + x_i^c)^2 \log \left( 1 + \frac{x_i^c}{\lambda^c} \right) - 2c \beta x_i (\lambda + x_i) (\lambda^c + x_i^c) - c^2 \lambda^{c-1} x_i^{c-1} (\lambda + x_i)^2 (\alpha (\lambda + x_i) - \beta \lambda)}{\beta \lambda (\lambda + x_i) (\lambda^c + x_i^c)^2 \log \left( 1 + \frac{x_i^c}{\lambda^c} \right) + c x_i^{c-1} (\alpha (\lambda + x_i) - \beta \lambda) (\lambda + x_i)^2 (\lambda^c + x_i^c)} + \frac{c \alpha}{\lambda} \sum_{i=1}^{n} \frac{x_i^c}{\lambda^c + x_i^c} + \beta \sum_{i=1}^{n} \frac{x_i \log \left( 1 + \frac{x_i^c}{\lambda^c} \right)}{\lambda + x_i} - c \beta \sum_{i=1}^{n} \frac{1}{\lambda + x_i (\lambda^c + x_i^c)} \quad (28) \]

\[ \frac{\partial \ell}{\partial c} = \sum_{i=1}^{n} \frac{\beta (\omega_i - 1)^c (1 + (\omega_i - 1)^c) \log (\omega_i - 1) + \omega_i (\omega_i - 1)^{c-1} (\alpha \omega_i - \beta) (1 + c + (\omega_i - 1)^c)}{\beta (1 + (\omega_i - 1)^c)^2 \log (1 + (\omega_i - 1)^c) + c \omega_i (\omega_i - 1)^{c-1} (\alpha \omega_i - \beta) (1 + (\omega_i - 1)^c)} - \sum_{i=1}^{n} \frac{(\alpha \omega_i - \beta) (\omega_i - 1)^c \log (\omega_i - 1)}{\omega_i (1 + (\omega_i - 1)^c)} = 0 \quad (29) \]

where \( \omega_i = 1 + \frac{x_i}{\lambda} \).
The MLEs of the four parameters for the NBurr distribution with $\alpha, \beta, \lambda$ and $c$ are obtained by setting the above partial derivatives to zero and solving them simultaneously. The closed-form solutions are not available for the equations (26), (27), (28) and (29). So, an iterative algorithm should be applied to solve these equations numerically. For practical implementation of the model, we fit the NBurr models in the whole range of the data sets with quasi-Newton BFGS numerical algorithm with initial values to be chosen as $(\hat{\alpha}_0, \hat{\beta}_0, \hat{\lambda}_0, \hat{c}_0) = (1, 1, 1, 1)$ to find the MLE estimates of the parameters.

We also present the asymptotic distributions for the NBurr distribution. The Fisher information matrix $(I)$ can be obtained by taking the expected values of the second-order and mixed partial derivatives of $\ell(x; \alpha, \beta, \lambda, c)$ w.r.t. $\alpha, \beta, \lambda$ and $c$. Since the analytical expression is hard to compute, it can be approximated by numerically investigating the $I = (I_{ij})$ matrix. The asymptotic $I$ matrix can be given as follows:

$$ I = \begin{bmatrix} -\frac{\partial^2 \ell}{\partial \alpha^2} & -\frac{\partial^2 \ell}{\partial \alpha \partial \beta} & -\frac{\partial^2 \ell}{\partial \alpha \partial \lambda} & -\frac{\partial^2 \ell}{\partial \alpha \partial c} \\ -\frac{\partial^2 \ell}{\partial \beta \partial \alpha} & -\frac{\partial^2 \ell}{\partial \beta^2} & -\frac{\partial^2 \ell}{\partial \beta \partial \lambda} & -\frac{\partial^2 \ell}{\partial \beta \partial c} \\ -\frac{\partial^2 \ell}{\partial \lambda \partial \alpha} & -\frac{\partial^2 \ell}{\partial \lambda \partial \beta} & -\frac{\partial^2 \ell}{\partial \lambda^2} & -\frac{\partial^2 \ell}{\partial \lambda \partial c} \\ -\frac{\partial^2 \ell}{\partial c \partial \alpha} & -\frac{\partial^2 \ell}{\partial c \partial \beta} & -\frac{\partial^2 \ell}{\partial c \partial \lambda} & -\frac{\partial^2 \ell}{\partial c^2} \end{bmatrix} $$

The second order partial derivatives of $\ell(x; \alpha, \beta, \lambda, c)$ w.r.t. $\alpha, \beta, \lambda$ and $c$ can be calculated but the calculations are very tedious. Hence, we omit the calculation part. The variance-covariance matrix is approximated by $M = (M_{ij})$ where $M_{ij} = I_{ij}^{-1}$. The asymptotic distribution of MLEs for $\alpha$, $\beta$, $\lambda$, and $c$ can be written as

$$ \left[ (\hat{\alpha} - \alpha), (\hat{\beta} - \beta), (\hat{\lambda} - \lambda), (\hat{c} - c) \right] \sim N_4 \left( 0, I^{-1}(\hat{\theta}) \right). $$

Then the approximate 100(1 – $k$)% confidence intervals for $\alpha$, $\beta$, $\lambda$, and $c$ are given by $\hat{\alpha} \pm Z_{k/2} \sqrt{Var(\hat{\alpha})}$, $\hat{\beta} \pm Z_{k/2} \sqrt{Var(\hat{\beta})}$, $\hat{\lambda} \pm Z_{k/2} \sqrt{Var(\hat{\lambda})}$, and $\hat{c} \pm Z_{k/2} \sqrt{Var(\hat{c})}$; where $\hat{\Theta} = (\hat{\alpha}, \hat{\beta}, \hat{\lambda}, \hat{c})$ and $Z_{k/2}$ is the upper 100 $k$-th percentile of the standard normal distribution.

### 3.3 Goodness of fit

The goodness-of-fit test refers to the measure of how well the data sets of large-scale real-world networks correspond to the fitted (hypothesized)
NBurr distribution. We have used the Chi-square statistic test and its corresponding $p$ value to determine the goodness of fit for the NBurr distribution. We obtain the $p$ values using the bootstrap resampling computational technique as follows:

- We determine the best fit of the proposed NBurr distribution to the data by estimating parameters through the MLE method. Then we calculate the Chi-square statistic for the goodness-of-fit of the best-fitted NBurr model for the data.
- We generate 50000 synthetic data sets from the NBurr distribution and calculate the Chi-square statistic for each of the synthetic data sets.
- Finally, we obtain the $p$ value for the synthetic data sets as the fraction of NBurr synthetic data sets with a Chi-square value greater than the empirical one. Higher $p$ values signifies that the proposed model is ‘most’ suitable for the data set.

In addition, we compute standard statistical measures such as Kullback-Leibler divergence (KLD), root mean squared error (RMSE), and mean absolute error (MAE) for quantifying the goodness-of-fit of the proposed NBurr distribution compared to the other commonly used heavy-tailed distributions.

4 Real-world Applications

We show the application of the NBurr distribution in the analysis of large-scale complex network data sets from various disciplines. A large-scale real-world complex network is one in which the number of nodes is of the order of thousands or millions, for example, world wide web, twitter, Facebook, wiki, amazon networks, etc. The node degree distribution of these large scale network is considered as an important structural characteristic of large scale real-world networks [2, 6, 7]. The degree distributions of these real-world networks, for example, collaboration networks, communication networks, social networks, biological networks, etc., are assumed to follow a heavy-tailed power-law distribution [31, 6]. Previous researchers reported that a baseline power law, Exponential, Pareto, lognormal, and Burr models are insufficient to fit the empirical data properly in its whole range unless some of the lower degree nodes are left out while fitting the model [12, 11,
41, 38, 10]. In recent work, Brodo et al. [8] pointed out that the recent data concentration on all these networks data shows that they no longer follow the power-law distribution.

4.1 Data

We consider large-scale network data sets from different disciplines, namely social networks, collaboration networks, citation networks, web graphs, product co-purchasing networks, temporal networks, communication networks, and ground-truth networks. We study several individual data sets from each discipline to showcase the general applicability of the proposed NBurr distribution. These data sets are publicly available at [http://snap.stanford.edu/data/index.html](http://snap.stanford.edu/data/index.html). These are standard network data sets with heavy-tail behaviors and used for modeling in the statistical analysis of networks [32, 31, 33]. An overview of these publicly available network data sets along with statistical measures (mean ($\mu$), standard deviation ($s$), etc.) are presented in Table 3. Another interesting property of these network data sets is their coefficient of variation ($s/\mu$) exceeding unity.

<table>
<thead>
<tr>
<th>Data sets</th>
<th>No. of nodes</th>
<th>No. of edges</th>
<th>Estimated parameters</th>
<th>Bootstrap chi-square value ($p$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ego-Twitter(In)</td>
<td>81,306</td>
<td>1,768,149</td>
<td>$s = 57.965, \mu = 21.747, \hat{\alpha} = 2.6654$</td>
<td>$\hat{\beta} = 3.3955, \hat{\lambda} = 0.6918, \hat{c} = 43.693, p = 0.7324$</td>
</tr>
<tr>
<td>ego-Gplus(In)</td>
<td>107,614</td>
<td>13,673,453</td>
<td>$s = 1404.8, \mu = 283.42, \hat{\alpha} = 4.9568$</td>
<td>$\hat{\beta} = 1.3737, \hat{\lambda} = 0.4503, \hat{c} = 24.924, p = 0.5837$</td>
</tr>
<tr>
<td>soc-Delicious(In)</td>
<td>536,108</td>
<td>1,365,961</td>
<td>$s = 39.826, \mu = 10.673, \hat{\alpha} = 3.7312$</td>
<td>$\hat{\beta} = 7.6500, \hat{\lambda} = 3.1845, \hat{c} = 26.999, p = 0.2197$</td>
</tr>
<tr>
<td>Athlete's-Facebook</td>
<td>13,866</td>
<td>86,859</td>
<td>$s = 17.975, \mu = 12.438, \hat{\alpha} = 1.4453$</td>
<td>$\hat{\beta} = 2.8158, \hat{\lambda} = -0.4558, \hat{c} = 24.053, p = 1.1022$</td>
</tr>
<tr>
<td>LiveJournal(In)</td>
<td>4,847,571</td>
<td>68,993,773</td>
<td>$s = 50.765, \mu = 7.1444, \hat{\alpha} = 6.0634$</td>
<td>$\hat{\beta} = 16.595, \hat{\lambda} = 5.6714, \hat{c} = 61.718, p = 0.1137$</td>
</tr>
<tr>
<td>Web</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Google(In)</td>
<td>597,414</td>
<td>3,109,539</td>
<td>$s = 69.125, \mu = 15.368, \hat{\alpha} = 2.9262$</td>
<td>$\hat{\beta} = 7.7807, \hat{\lambda} = 3.7317, \hat{c} = 79.102, p = 0.5004$</td>
</tr>
<tr>
<td>BerkStan(In)</td>
<td>183,830</td>
<td>2,722,334</td>
<td>$s = 15.313, \mu = 10.424, \hat{\alpha} = 3.0195$</td>
<td>$\hat{\beta} = 0.5426, \hat{\lambda} = 4.1159, \hat{c} = 6.3332, p = 0.3129$</td>
</tr>
<tr>
<td>Wikipedia2009(In)</td>
<td>1,864,433</td>
<td>4,507,315</td>
<td>$s = 12.316, \mu = 24.364, \hat{\alpha} = 1.4453$</td>
<td>$\hat{\beta} = 2.8158, \hat{\lambda} = -0.4558, \hat{c} = 24.053, p = 1.1022$</td>
</tr>
<tr>
<td>Product</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>co-purchasing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>amazon0505(In)</td>
<td>403,394</td>
<td>2,983,388</td>
<td>$s = 15.313, \mu = 8.3989, \hat{\alpha} = 1.8873$</td>
<td>$\hat{\beta} = 6.2857, \hat{\lambda} = 1.1626, \hat{c} = 2.2058, p = 0.9800$</td>
</tr>
<tr>
<td>networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>amazon0112(In)</td>
<td>400,727</td>
<td>3,200,444</td>
<td>$s = 13.175, \mu = 7.9865, \hat{\alpha} = 1.8873$</td>
<td>$\hat{\beta} = 6.2857, \hat{\lambda} = 1.1626, \hat{c} = 2.2058, p = 0.9800$</td>
</tr>
<tr>
<td>Temporal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sx-mathoverflow(In)</td>
<td>24,818</td>
<td>123,830</td>
<td>$s = 15.313, \mu = 8.3989, \hat{\alpha} = 1.8873$</td>
<td>$\hat{\beta} = 6.2857, \hat{\lambda} = 1.1626, \hat{c} = 2.2058, p = 0.9800$</td>
</tr>
<tr>
<td>Networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sx-supersuser(In)</td>
<td>194,085</td>
<td>1,512,334</td>
<td>$s = 15.313, \mu = 8.3989, \hat{\alpha} = 1.8873$</td>
<td>$\hat{\beta} = 6.2857, \hat{\lambda} = 1.1626, \hat{c} = 2.2058, p = 0.9800$</td>
</tr>
<tr>
<td>sx-asakubuntu(In)</td>
<td>159,316</td>
<td>1,443,339</td>
<td>$s = 15.313, \mu = 8.3989, \hat{\alpha} = 1.8873$</td>
<td>$\hat{\beta} = 6.2857, \hat{\lambda} = 1.1626, \hat{c} = 2.2058, p = 0.9800$</td>
</tr>
<tr>
<td>Communication</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email-Enron</td>
<td>36,692</td>
<td>183,831</td>
<td>$s = 15.313, \mu = 8.3989, \hat{\alpha} = 1.8873$</td>
<td>$\hat{\beta} = 6.2857, \hat{\lambda} = 1.1626, \hat{c} = 2.2058, p = 0.9800$</td>
</tr>
<tr>
<td>Wiki-Talk(In)</td>
<td>1,354,000</td>
<td>5,021,410</td>
<td>$s = 15.313, \mu = 8.3989, \hat{\alpha} = 1.8873$</td>
<td>$\hat{\beta} = 6.2857, \hat{\lambda} = 1.1626, \hat{c} = 2.2058, p = 0.9800$</td>
</tr>
<tr>
<td>Rec-Libimseti(In)</td>
<td>240,079</td>
<td>7,839,255</td>
<td>$s = 15.313, \mu = 8.3989, \hat{\alpha} = 1.8873$</td>
<td>$\hat{\beta} = 6.2857, \hat{\lambda} = 1.1626, \hat{c} = 2.2058, p = 0.9800$</td>
</tr>
<tr>
<td>Ground-Truth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>com-Orkut</td>
<td>3,072,441</td>
<td>117,185,083</td>
<td>$s = 15.313, \mu = 8.3989, \hat{\alpha} = 1.8873$</td>
<td>$\hat{\beta} = 6.2857, \hat{\lambda} = 1.1626, \hat{c} = 2.2058, p = 0.9800$</td>
</tr>
<tr>
<td>com-Youtube</td>
<td>1,134,080</td>
<td>2,983,388</td>
<td>$s = 15.313, \mu = 8.3989, \hat{\alpha} = 1.8873$</td>
<td>$\hat{\beta} = 6.2857, \hat{\lambda} = 1.1626, \hat{c} = 2.2058, p = 0.9800$</td>
</tr>
</tbody>
</table>

4.2 Experimental Results

In this section, we compare the NBurr distribution with the other seven established models, namely Power-law, Pareto, Log-normal, Power-law with
cutoff, Burr, exponentiated Burr [25], MO Burr [22] distributions. To estimate the parameters \((\alpha, \beta, \lambda, c)\) of the NBurr distribution numerically, we have used ‘optim’ function along with the quasi-Newton L-BFGS-B algorithm in R statistical software by taking the initial parameters value \((\alpha, \beta, \lambda, c) = (1, 1, 1, 1)\). The estimated values of the parameters for all the network data sets satisfied the following conditions: \(\alpha > 0, \beta > -1 \lambda > 0\) and \(c > 0\) as depicted in Table 3, which characterize the proposed NBurr distribution. Empirically it is observed that in the case of social networks, the estimated value of the parameter \(\lambda\) attains the higher values as compared to the estimated value of \(\alpha\). From Table 3 it is also clear that the proposed NBurr distribution produces higher \(p\) values (i.e. closure to 1) through bootstrapping chi-square test which suggests that the null hypothesis i.e., the data are drawn from NBurr distribution cannot be ruled out at the 0.05 level of significance. This recommends that the observed degree distribution is plausibly drawn from the NBurr distribution. Thus it can be concluded from Table 3 that the proposed NBurr distribution is effective in modeling the entire degree distribution of real-world complex networks. Also, we used some other statistical measures, viz. KLD, RMSE, and MAE to compare the performance of the proposed NBurr distribution with the competitive heavy-tailed distributions as shown in Tables 4 and 5.

**Table 4**

Table of different statistical measures of different competitive models over real-world networks

<table>
<thead>
<tr>
<th>Data sets</th>
<th>Burr</th>
<th>NBurr</th>
<th>Exponentiated Burr</th>
<th>MO Burr</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>KLD</td>
<td>MAE</td>
<td>RMSE</td>
</tr>
<tr>
<td>Social</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ego-Twitter(In)</td>
<td>2.4786</td>
<td>0.00097</td>
<td>1.1725</td>
<td>28.727</td>
</tr>
<tr>
<td>ego-Gplus(In)</td>
<td>2.7411</td>
<td>0.05621</td>
<td>0.1994</td>
<td>4.0728</td>
</tr>
<tr>
<td>soc-Delicious(In)</td>
<td>2.5914</td>
<td>0.04999</td>
<td>1.7952</td>
<td>20.291</td>
</tr>
<tr>
<td>LiveJournal(In)</td>
<td>384.81</td>
<td>0.01686</td>
<td>13.924</td>
<td>464.51</td>
</tr>
<tr>
<td>Athletes-Facebook</td>
<td>2.9584</td>
<td>0.08886</td>
<td>1.3876</td>
<td>20.288</td>
</tr>
<tr>
<td>Citation</td>
<td>2.9780</td>
<td>0.01334</td>
<td>0.4718</td>
<td>11.985</td>
</tr>
<tr>
<td>Networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cit-RepTh(In)</td>
<td>2.5476</td>
<td>0.01327</td>
<td>0.4361</td>
<td>2.9780</td>
</tr>
<tr>
<td>cit-Patents(In)</td>
<td>626.74</td>
<td>0.00203</td>
<td>6.4674</td>
<td>173.55</td>
</tr>
<tr>
<td>cit-Citeeseer(In)</td>
<td>43.719</td>
<td>0.00206</td>
<td>3.6469</td>
<td>33.484</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Web</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Google(In)</td>
<td>189.96</td>
<td>0.06699</td>
<td>8.2424</td>
<td>177.95</td>
</tr>
<tr>
<td>BerkStan(In)</td>
<td>32.321</td>
<td>0.02671</td>
<td>0.5353</td>
<td>32.388</td>
</tr>
<tr>
<td>Wikipedia2009(In)</td>
<td>176.29</td>
<td>0.03115</td>
<td>9.4514</td>
<td>52.148</td>
</tr>
<tr>
<td>Product</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>amazon060(In)</td>
<td>92.912</td>
<td>0.03115</td>
<td>7.6488</td>
<td>71.210</td>
</tr>
<tr>
<td>co-purchasing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>amazon0505(In)</td>
<td>118.17</td>
<td>0.03751</td>
<td>9.1303</td>
<td>78.727</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mx-mathoverflow(In)</td>
<td>8.8476</td>
<td>0.01442</td>
<td>1.5729</td>
<td>6.104</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email-Enron</td>
<td>76.787</td>
<td>0.0526</td>
<td>5.2356</td>
<td>74.139</td>
</tr>
<tr>
<td>Wiki-Talk(In)</td>
<td>517.42</td>
<td>0.03434</td>
<td>21.725</td>
<td>22.852</td>
</tr>
<tr>
<td>Rec-Liberati0n(In)</td>
<td>41.548</td>
<td>0.03261</td>
<td>0.9310</td>
<td>48.044</td>
</tr>
<tr>
<td>Ground-truth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wiki-Topsites</td>
<td>9.6105</td>
<td>0.01886</td>
<td>0.9193</td>
<td>6.1751</td>
</tr>
<tr>
<td>com-Orkut</td>
<td>194.24</td>
<td>0.07211</td>
<td>6.2222</td>
<td>95.580</td>
</tr>
<tr>
<td>com-Youtube</td>
<td>37.006</td>
<td>0.00112</td>
<td>0.5858</td>
<td>38.661</td>
</tr>
</tbody>
</table>

RMSE and MAE are two different variants, carrying information about
the differences between actual and predicted degree frequencies corresponding to a network. The higher similarity between actual and predicted distributions is achieved by generating smaller values of RMSE and MAE. From Tables 4 and 5, it is clear that the proposed NBurr distribution provides smaller RMSE and MAE values than other competitive distributions in almost all the networks, except a few, where Burr, MO Burr, and power-law cutoff distribution perform better than the proposed one. Power-law and Pareto distributions provide lower performance in minimizing the RMSE and MAE values compared to the other competing distributions over all the real-world networks, as clearly seen from Table 5. The performances of Burr and MO Burr distributions are alike in terms of RMSE and MAE measures. KLD is a quantity that measures the dissimilarity between two probability distributions. Thus the smaller value of KLD represents the higher similarity between the actual and the predicted distribution. From Tables 4 and 5, it is clear that the proposed NBurr distribution generates smaller KLD values compared to other competitive distributions in almost all the networks. This indicates that the observed degree distribution satisfactorily matches the proposed NBurr distribution and always superior to the state-of-the-art models in almost all the networks. Thus overall, by considering RMSE, MAE, and KLD values, the performance of the proposed NBurr distribution over real-world networks is found to be better than the other competing distributions which suggest that the observed distribution plausibly comes from the proposed NBurr distribution.

![Figure 2](image-url): Degree distribution of ego-Twitter and LiveJournal networks in log-log scale
Table 5
Table of different statistical measures of different competitive models over real-world networks

<table>
<thead>
<tr>
<th>Data sets</th>
<th>Power-law</th>
<th>Pareto Type-I</th>
<th>Log-normal</th>
<th>Power-law Cutoff</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>KLD</td>
<td>MAE</td>
<td>RMSE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>KLD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MAE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RMSE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>KLD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MAE</td>
</tr>
<tr>
<td>Social Networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ego-Twitter(In)</td>
<td>294.35</td>
<td>0.1841</td>
<td>12.847</td>
<td>564.25</td>
</tr>
<tr>
<td>ego-Cplus(In)</td>
<td>53.064</td>
<td>0.2299</td>
<td>0.9221</td>
<td>133.863</td>
</tr>
<tr>
<td>soc-Delicious(In)</td>
<td>349.66</td>
<td>0.2021</td>
<td>14.867</td>
<td>471.02</td>
</tr>
<tr>
<td>LiveJournal(In)</td>
<td>5025.2</td>
<td>0.1614</td>
<td>127.98</td>
<td>8109.9</td>
</tr>
<tr>
<td>Athletes-Facebook</td>
<td>100.16</td>
<td>0.2049</td>
<td>13.387</td>
<td>204.91</td>
</tr>
<tr>
<td>Citation Networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cit-RepFl(In)</td>
<td>5.533</td>
<td>0.1741</td>
<td>4.0821</td>
<td>12.295</td>
</tr>
<tr>
<td>cit-Patents(In)</td>
<td>27.58</td>
<td>0.2266</td>
<td>20.495</td>
<td>34.868</td>
</tr>
<tr>
<td>soc-Delicious(In)</td>
<td>349.33</td>
<td>0.2036</td>
<td>14.852</td>
<td>471.02</td>
</tr>
<tr>
<td>LiveJournal(In)</td>
<td>5025.2</td>
<td>0.1614</td>
<td>127.98</td>
<td>8109.9</td>
</tr>
<tr>
<td>Athletes-Facebook</td>
<td>100.16</td>
<td>0.2049</td>
<td>13.387</td>
<td>204.91</td>
</tr>
<tr>
<td>Web Graphs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Google(In)</td>
<td>3698.1</td>
<td>0.124</td>
<td>45.023</td>
<td>5154.1</td>
</tr>
<tr>
<td>BerkStan(In)</td>
<td>615.03</td>
<td>0.1803</td>
<td>4.0722</td>
<td>615.01</td>
</tr>
<tr>
<td>Wikipedia2009(In)</td>
<td>4371.9</td>
<td>0.1352</td>
<td>164.58</td>
<td>4371.9</td>
</tr>
<tr>
<td>Product co-purchasing networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>amazon0001(In)</td>
<td>149.54</td>
<td>0.2708</td>
<td>79.281</td>
<td>253.96</td>
</tr>
<tr>
<td>amazon0005(In)</td>
<td>1572.9</td>
<td>0.2403</td>
<td>73.603</td>
<td>2494.5</td>
</tr>
<tr>
<td>amazon012(In)</td>
<td>1564.4</td>
<td>0.2425</td>
<td>71.875</td>
<td>2462.9</td>
</tr>
<tr>
<td>Temporal Networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sx-mathoverflow(In)</td>
<td>213.91</td>
<td>0.2131</td>
<td>13.600</td>
<td>213.92</td>
</tr>
<tr>
<td>sx-superuser(In)</td>
<td>900.04</td>
<td>0.1808</td>
<td>33.837</td>
<td>900.33</td>
</tr>
<tr>
<td>sx-askubuntu(In)</td>
<td>949.66</td>
<td>0.2091</td>
<td>39.419</td>
<td>949.73</td>
</tr>
<tr>
<td>Communication Networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email-Enron</td>
<td>246.51</td>
<td>0.1779</td>
<td>14.846</td>
<td>245.25</td>
</tr>
<tr>
<td>Wiki-Talk(In)</td>
<td>77.081</td>
<td>0.2198</td>
<td>2.146</td>
<td>77.081</td>
</tr>
<tr>
<td>Rec-Limbuset(In)</td>
<td>1380.5</td>
<td>0.1342</td>
<td>15.690</td>
<td>1380.5</td>
</tr>
<tr>
<td>Ground-Truth Networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wiki-Topcats</td>
<td>566.21</td>
<td>0.1347</td>
<td>2.6145</td>
<td>580.44</td>
</tr>
<tr>
<td>com-Orkut</td>
<td>2443.6</td>
<td>0.5498</td>
<td>80.712</td>
<td>4299.3</td>
</tr>
<tr>
<td>com-Youtube</td>
<td>1380.5</td>
<td>0.1342</td>
<td>15.690</td>
<td>1380.5</td>
</tr>
</tbody>
</table>

Figure 3: Degree distribution of cit-Citeseer and BerkStan networks in log-log scale
The scatter plot of the fitted results can be used to verify the effectiveness of the proposed NBurr distribution. To do this, the log-log plots of the original frequency distribution, the estimated frequency by NBurr distribution and the frequency estimated by Burr, Exponentiated Burr, MO Burr, Power-law, Pareto Type-I, Log-normal and Power-law Cutoff distributions are drawn for all the networks under consideration. Eight such plots are given in Figures 2-5. These are the ego-Twitter network, LiveJournal network, cit-Citeseer network, Web-BerkStan network, Amazon0601 network, sx-superuser network, Email-Enron network, and com-Wiki-Topcats network. It is visually clear from Figures 2-5 that the proposed NBurr curve...
always passes through the middle of the scatter plot of the observed distribution. This signifies that the proposed NBurr distribution provides a better fit compared to the other competitive distributions in almost all of the networks. Thus, it is visually clear, through observing the plotted results, that the entire node degree distribution can be better represented by the NBurr distribution compared to other heavy-tailed distributions. Thus the proposed NBurr distribution, a modification of the Burr distribution with nonlinear exponent in the shape parameter, can be used for effective and efficient modeling of the entire degree distribution of real-world networks without ignoring the lower degree nodes. Finally, we can conclude that these heavy-tailed network data sets when modeled in the whole range using the proposed NBurr distribution shows significant improvements in comparison to state-of-the-art models.

5 Conclusion

In this paper, we present a new method called shape parameter transformation (SPT) to extend Burr and related families. The SPT method’s idea is to use a nonlinear exponent (depends on data and an additional parameter) instead of using a constant shape parameter in the Burr, Pareto, and exponentiated family of distributions. The method was first applied to the Burr distribution and a generalized Burr (GBurr) model is introduced. This newly introduced GBurr models belong to the maximum domain of attraction of the Frechet distribution and are right-tail equivalent to the Pareto distribution. Further, the SPT method is applied to the classical Pareto and exponentiated distributions. We also studied a relevant model, namely NBurr distribution, from generalized Burr family and derived various statistical properties. The practical usefulness of the proposed NBurr distribution was shown using multiple heavy-tailed network data sets from different disciplines. It is interesting to note that the NBurr distribution is a new competitor for popularly used MO Burr and exponentiated Burr distributions. An immediate extension of this work is to apply these newly introduced probability models for modeling survival and lifetime data sets. Another possible extension of this work would be to look for implementing the proposed SPT method for other size distributions available in the statistics paradigm.
Acknowledgements

The authors are thankful to Professor Gopal K Basak of Indian Statistical Institute, Kolkata for constructive comments and insightful suggestions in this paper.

References


