A New Family of Distributions in the Class of the Alpha Power Transformation with Applications to Income

Abtract

In this paper, the alpha-power transformation is used to propose a new class of probability distribution which is particularly useful for analysis of lifetime data. A special case is considered in details. Various mathematical properties of the proposed model including quantile function, moments and moment generating function, residual life, mean residual life and order statistics are derived. The maximum likelihood method is used to estimate the unknown parameters. A real data set is analyzed using the new distribution, and a simulation study is carried out to asses the performance of the new family.

Keywords: alpha power transformation; new alpha power transformed; new extended alpha power transformation; Dagum distribution; moment generating function; maximum likelihood estimation; Gini coefficient.

1 Introduction

Statistical distributions are very useful in describing and predicting real world phenomena. Numerous classical distributions have been extensively used over the past few decades for modeling data in several areas. Examples of the use of statistical distributions are to be found in several areas such as engineering, actuarial science, environmental, biological studies, bio-medical analysis, demography, economics, finance and insurance. However, due to the increase in the complexity of the industrial processes and due to the expansion of relations in the business world, where there is a clear need for extended forms of these distributions.

Furthermore, in many practical fields, classical distributions do not provide a proper fit to real data and is often an approximation rather than exactness. So, in such situations we need either completely new distributions modified or generalized forms of the existing distributions.

Recent developments on several methods for generating new families of continuous univariate distributions that extend well-known distributions and at the same time provide added flexibility in modelling data. Many other methods have been developed for the purpose of generating family of lifetime distributions. Such methods occurred as a result of the most influential work by [(1)] for proposing statistical distributions via system of differential equation approach. Another method, based on differential equation, was proposed by [(2)]. The method of combining two symmetric distributions to form a skewed distribution was first proposed by [(3)], who introduced the skew normal distribution by introducing an extra parameter to the normal distribution to bring more flexibility to the normal distribution.

Since 1990, methodologies of generating new statistical distributions aimed at introducing additional parameters to the existing distributions or combining existing distributions. Often introducing an extra parameter brings more flexibility to a class of distribution functions, and it can be very useful for data analysis purposes. This method has been widely used to generate the so-called exponentiated family; this name was first used by [(4)]. They applied this idea to introduce an extra parameter to a two-parameter Weibull distribution with one scale (λ) and one shape (β) parameter.

[(5)] introduced another method for adding an extra parameter from lifetime distribution by using the survival function of the distribution $S(x)$.

$$G(x) = 1 - \frac{\alpha S(x)}{1-(1-\alpha)S(x)},\ \alpha > 0 \tag{1.1}$$

They considered two special cases; namely, when X follows exponential or Weibull distribution using Eq.(1.1) and derived several properties of this proposed model.

[(6)] introduced a pioneering idea by constructing what is known as the beta generated distributions. They proposed using the beta distribution as the generator to develop the beta generated distributions. The cdf of a beta-generated distribution is defined as

$$G(x) = \int_0^{F(x)} b(t)\ dt,\ \alpha, \beta > 0 \tag{1.2}$$

where $b(t)$ is the pdf. of the beta random variable and $F(x)$ is the cdf of any random variable X. The pdf corresponding to the beta-generated distribution is given by

$$g(x) = \frac{1}{B(\alpha,\beta)}\ f(x)\ F^{\alpha-1}\ (1-F(x))^{\beta-1}, \tag{1.3}$$

where $f(x)$, $F(X)$ are the pdf and cdf of any random variable, and $B(\alpha, \beta)$ denotes the beta function. [(7)] extended the beta-generated family of distributions to the Kumaraswamy generated family (KW-G) by replacing the beta distribution in beta-generated with the Kumaraswamy distribution.

[(8)] introduced a general method for generating families of continuous distributions called the transformed-transformer (T-X) family, which allows the use of any continuous p.d.f. as the generator instead of beta or Kumaraswamy. The cdf of the (T-X) distribution is $G(x) = \int_{d1}^{W(F(x))} r(t)dt$, where r(t) is the p.d.f. of the random variable $T \in [d1, d2]$ for $-\infty \leq d1 < d2 \leq \infty$. The function W(F(x)) is monotonic and absolutely continuous. [(9)] considered the function W(F(x)) to be the quantile function of a random variable Y and defined the T-R$\{$Y$\}$ family.

[(10)] proposed a powerful method called the alpha power transformation (APT) for adding an extra parameter to a family of distributions to obtain a new family. The parameter α provides more flexibility to the new family. Let F(x) and f(x) be the cdf and pdf of a random variable X, then the α-power transformation of $F(x)$ for $x \in \mathbb{R}$, is defined as follows:

$$G_{APT}(x) = \begin{cases} \frac{\alpha^{F(x)}-1}{\alpha-1} & \text{if} \quad \alpha > 0, \alpha \neq 1, x \in \mathbb{R} \\ F(x) & \text{if} \quad \alpha = 1. \end{cases} \tag{1.4}$$

And the corresponding pdf as follows:

$$g_{APT}(x) = \begin{cases} \frac{1}{\alpha-1}\log\alpha f(x)\alpha^{F(x)} & \text{if} \quad \alpha > 0, \alpha \neq 1, \\ f(x) & \text{if} \quad \alpha = 1. \end{cases} \tag{1.5}$$

where α is a shape parameter. Using (1.4), they introduced the alpha power exponential (APE) distribution and studied the main properties as well as estimation of the parameters of the proposed distribution.

Many authors studied the APT method to re-extend some univariate distributions in particular cases; for example, [(11)] introduced the alpha power Weibull distribution. [(12)] derived a closed form expression for moment properties of the alpha power generalized exponential distribution. [(13)] studied the general mathematical properties of the APT family and considered the alpha power exponentiated Weibull distribution. [(14)] discussed the estimation of the parameters of the APE distribution using nine methods of estimation. [(15)] introduced the exponentiated generalized alpha power family of distributions to extend several other distributions. They used the new family and develop a new distribution, called the exponentiated generalized alpha power exponential distribution.

Another technique of APT is used to obtain a new class of lifetime distributions by [(16)]; which is named new alpha power transformed (NAPT) family. The cdf of the proposed family is defined by the following expression

$$G_{NAPT}(x;\Theta) = \begin{cases} \frac{F(x;\varepsilon)\alpha^{F(x;\varepsilon)}}{\alpha} & \text{if} \quad \alpha,\varepsilon > 0, \alpha \neq 1 \\ F(x;\varepsilon) & \text{if} \quad \alpha = 1. \end{cases} \tag{1.6}$$

where $\Theta = (\alpha, \varepsilon)$. The probability density function (pdf) corresponding to Eq.(1.6) is given by

$$g_{NAPT}(x;\Theta) = \frac{1}{\alpha}\Big[f(x;\Theta)\alpha^{F(x;\Theta)}\Big]\Big[1 + \log(\alpha)F(x;\Theta)\Big], \quad \text{if} \quad \alpha > 0, \alpha \neq 1. \tag{1.7}$$

For practical utility, they considered a new distribution with three-parameter special model named as alpha power transformed Weibull (NAPTW) distribution to evaluate the efficiency of the proposed class.

[(17)] proposed a new family of APT to construct a new class of lifetime distributions, called the Zubair-G family which has the cdf as follows

$$G(x;\alpha,\varepsilon) = \frac{e^{\alpha F(x;\varepsilon)^2} - 1}{e^{\alpha} - 1}, \quad \text{if} \quad \alpha,\varepsilon > 0, x \in \mathbb{R}. \tag{1.8}$$

The probability density function (pdf) of the Zubair-G family is given by

$$g(x;\alpha,\varepsilon) = \frac{1}{e^{\alpha} - 1}\Big[2\alpha f(x;\varepsilon)F(x;\varepsilon)e^{\alpha F(x;\varepsilon)^2}\Big] \tag{1.9}$$

Recently, [(18)] introduced a new method to add an additional parameter to the existing distributions.

His effort led to a new family of lifetime distributions, called the new extended alpha power transformation (NEAPT) family of distributions, which has the cdf as follows:

$$G_{NEAPT}(x;\alpha,\xi) = \frac{\alpha^{F(x;\in) - e^{F(x;\in)}}}{\alpha - e^{\alpha}}; \quad \text{if} \quad \alpha,\xi > 0, \alpha \neq e, x \in R. \tag{1.10}$$

The probability density function (pdf) is given by

$$g_{NEAPT}(x;\alpha,\xi) = \frac{1}{\alpha - e^{\alpha}}\Big[f(x;\xi)(\alpha^{F(x;\xi)}log\alpha - \alpha e^{\alpha F(x;\xi)})\Big]; \quad x \in R. \tag{1.11}$$

A special sub-case was considered in details, two parameters Weibull distribution based on NEAPT by [(18)].

The aim of this research is to present an extra parameter to a family of distribution functions to bring more flexibility to the given family. We call this new family of distribution the extended α-power transformation $(ExAPT)$ family. The proposed family in the class of α-power transformation method is flexible and could be used to analyze a wide class of data. The rest of the paper is organized as follows. In Section 2, we introduce the $(ExAPT)$ family, and discuss some general properties of this family of distributions. A special sub-model of the proposed family along with the graphical sketching of its pdf and hazard is discussed in Section 3. In Section 4, some mathematical properties are obtained. Maximum likelihood estimates of the model parameters are obtained in Section 5. Simulation study is conducted in Section 6. In Section 7 a real data set is applied. Finally, concluding remarks are provided in Section 8.

2 Proposed Family of Distributions

Here we introduce an extension of the alpha power transformation (APT) to create a new class of distributions. Let $F(x)$ be the cdf of a random variable $X \in [d1, d2]$, for $-\infty \leq d1 < d2 \leq \infty$, and let $W(\cdot)$ be a differentiable non-decreasing function satisfying $W(x) \to d1 \; as \; x \to 0 \; and \; W(x) \to d2 \; as \; x \to 1$. The cdf of the proposed family is defined by the following expression

$$G_{ExAPT}(x;\theta) = \begin{cases} \frac{\alpha^{kW(F(x;\xi))}-1}{\alpha-1} & \text{for} \quad \alpha, \xi > 0, \alpha \neq 1, \\ kW(F(x;\xi)) & \text{for} \quad \alpha = 1. \end{cases} \quad x \in \mathbb{R} \tag{2.1}$$

where $\theta = (\alpha, \xi)^T$ and $k = [W(1)]^{-1}$. The pdf of the new family of distributions is given by

$$g_{ExAPT}(x;\theta) = \begin{cases} \frac{1}{\alpha-1} log(\alpha) k f(x;\xi) \big(\frac{\partial W(F(x;\xi))}{\partial x}\big) \alpha^{kW(F(x;\xi))} & \text{for} \quad \alpha > 0, \alpha \neq 1, \\ k f(x;\xi) \big(\frac{\partial W(F(x;\xi))}{\partial x}\big) & \text{for} \quad \alpha = 1. \end{cases} \tag{2.2}$$

The corresponding survival function (sf), hazard rate function (hrf), reversed hazard rate function (rhrf), and cumulative hazard rate function (chrf), are respectively, given by

$$\begin{aligned} S(x;\theta) =& 1 - G_{ExAPT}(x;\theta) \\ =& \frac{\alpha}{\alpha-1} \Big[1 - \alpha^{kW(F(x;\xi))-1} \Big], \qquad \alpha > 0, \alpha \neq 1, x \in \mathbb{R}. \end{aligned} \tag{2.3}$$

$$\begin{aligned} h(x;\theta) =& \frac{g_{ExAPT}(x;\theta)}{1 - G_{ExAPT}(x;\theta)} \\ =& \frac{log(\alpha) k f\big(x;\xi\big) \big(\frac{\partial W(F(x;\xi))}{\partial x}\big) \alpha^{kW(F(x;\xi))}}{\alpha - \alpha^{kW(F(x;\xi))}}, \quad \alpha > 0, \alpha \neq 1, x \in \mathbb{R}. \end{aligned} \tag{2.4}$$

$$\begin{aligned} r(x;\theta) =& \frac{g_{ExAPT}(x;\theta)}{G_{ExAPT}(x;\theta)} \\ =& \frac{log(\alpha) k f(x;\xi) \big(\frac{\partial W(F(x;\xi))}{\partial x}\big) \alpha^{kW(F(x;\xi)}}{\alpha^{kW(F(x;\xi))} - 1}, \quad \alpha > 0, \alpha \neq 1, x \in \mathbb{R}. \end{aligned} \tag{2.5}$$

$$\begin{aligned} H(x;\theta) =& - log(1 - G_{ExAPT}(x;\theta)) \\ =& - log \Big[\frac{\alpha - \alpha^{kW(F(x;\xi))}}{\alpha - 1} \Big], \qquad \alpha > 0, \alpha \neq 1, x \in \mathbb{R}. \end{aligned} \tag{2.6}$$

The main reasons behind using the $(ExAPT)$ family in practice reside in its flexibility regarding adding additional parameters to modify the existing distributions. It is used for the purpose of presenting the extended version of the basic distribution that contains a closed form for cdf, sf and hrf, and, finally, providing better fits than other modified models.

3 Some Statistical Properties

This subsection considers some basic mathematical properties of the proposed family.

3.1 Quantile Function

The quantile function of $(ExAPT)$ random variable X is given by

$$x = F^{-1}\Big[W^{-1}\Big(\frac{log(u(\alpha-1)+1)}{klog\alpha}\Big)\Big] \tag{3.1}$$

Using the previous equation, we can generate random sample from $(ExAPT)$ family by using U as uniform random number.

3.2 Moments and Moment Generating Function

Moments are very important and helps to describe the important properties of the distribution; For example, central tendency, dispersion, skewness and kurtosis. The r^{th} moment of the $(ExAPT)$ family of distributions are derived as follows

$$\mu_r^{'} = \int_{-\infty}^{\infty} x^r \frac{1}{\alpha-1} log(\alpha)kf(x;\xi)\Big(\frac{\partial W(F(x;\xi))}{\partial x}\Big)\alpha^{kW(F(x;\xi))}\, dx. \tag{3.2}$$

Using the series representation in the form $\alpha^{\nu} = \sum_{i=0}^{\infty} \frac{(log\alpha)^i}{i!}\nu^i$, so the expression (3.2) can be re-write as

$$\mu_r^{'} = \frac{1}{\alpha-1}\sum_{i=0}^{\infty}\frac{(log(\alpha))^{i+1}}{i!}k^{i+1}\int_{-\infty}^{\infty} x^r f(x;\xi)\Big(\frac{\partial W(F(x;\xi))}{\partial x}\Big)[W(F(x;\xi)]^i\, dx, \tag{3.3}$$

then,

$$\mu_r^{'} = \frac{1}{\alpha-1}\sum_{i=0}^{\infty}\frac{(log(\alpha))^{i+1}}{i!}\, k^{i+1}\eta_{r,i}, \tag{3.4}$$

where,

$$\eta_{r,i} = \int_{-\infty}^{\infty} x^r f(x;\xi)\Big(\frac{\partial W(F(x;\xi))}{\partial x}\Big)[W(F(x;\xi)]^i\, dx. \tag{3.5}$$

Furthermore, a general expression for moment generating function (mgf) of the $(ExAPT)$ random variable X is

$$\mu_x(t) = \frac{1}{\alpha-1}\sum_{i,r=0}^{\infty}\frac{(log(\alpha))^{i+1}}{i!r!}k^{i+1}t^r\eta_{r,i} \tag{3.6}$$

3.3 Mean Residual and Reverse Residual Life

The mean residual life is the expected additional lifetime that a component has survived until time t. More specifically, if the random variable x represents the life of a component, then the mean residual life is given by $m(t) = E(X - t/X > t)$, see [(19)]. A second measure of interest in reversed time is the reversed mean residual life. Suppose a device has failed before attaining age t. Then, the random variable $X_t = t - X/(X < t)$ is the time passed since the device has failed, conditioned on the fact that its lifetime is less than t, and this is referred to as the reversed residual life or inactivity time of X, see [(20)].

The mean residual life and reversed residual life associated with a lifetime random variable are of interested in numerous areas of applied sciences such as survival analysis, biometry, actuarial studies and risk management. The mean residual lifetime of $(ExAPT)$ random variable X denoted by $R_{(t)}$ is derived as

$$R_{(t)}(x) = \frac{S(x+t)}{S(t)},$$

$$R_{(t)}(x) = \frac{\alpha - \alpha^{kW(F(x+t;\xi))}}{\alpha - \alpha^{kW(F(x;\xi))}} \tag{3.7}$$

Additionally, the reverse residual lifetime of the $(ExAPT)$ random variable denoted by $\bar{R}_{(t)}$ is

$$\bar{R}_{(t)} = \frac{S(x-t)}{S(t)},$$

$$\bar{R}_{(t)} = \frac{\alpha - \alpha^{kW(F(x-t;\xi))}}{\alpha - \alpha^{kW(F(x;\xi))}} \tag{3.8}$$

3.4 Order Statistics

Let $X_1, X_2, ..., X_n$ be a random sample of size n taken independently from the $(ExAPT)$ distribution with parameters α and ξ. Let $X_{1:n}, X_{2:n}, ..., X_{n:n}$ be the corresponding order statistics. Then, from [(21)], the density of $X_{r:n}$ for $(r = 1, 2, ..., n)$ is given by

$$g_{r:k}(x;\theta) = \frac{g(x;\theta)}{B(r, n-r+1)} \sum_{i=0}^{n-r} \binom{n-r}{i} (-1)^i \, [G(x;\theta)]^{i+r-1}, \tag{3.9}$$

using (2.1) and (2.2) in (3.9), we get the density of the r^{th} order statistic for $(ExAPT)$ family.

4 Model Description

The Dagum distribution was introduced by [(22)] for modeling personal income data as an alternative to the Pareto and log-normal models. This distribution has been extensively used in various fields such as, income and wealth data, meterological data, reliability and survival analysis. The Dagum distribution (also called the inverse Burr) is the reciprocal transformation of the Burr XII. [(27)] said that Dagum referred to his model as the generalized logistic-Burr distribution. Dagum distribution has three Types: Type I with three parameters, Type II and Type III with four parameters, for detial see [(26)]. The current study will focus on Type I only.

The continuous random variable X is to have a three parameters Dagum distribution, abbreviated as $X \sim Dag(\beta, \lambda, \delta)$, if its density probability function (pdf) is given as

$$f(x;\beta,\lambda,\delta) = \beta\,\lambda\,\delta\,x^{-\delta-1}\,[1+\lambda\,x^{-\delta}]^{-\beta-1}, x > 0. \tag{4.1}$$

The cdf of Equation (5.5) is given by

$$F(x;\beta,\lambda,\delta) = [1+\lambda\,x^{-\delta}]^{-\beta}, x > 0, \beta, \lambda, \delta > 0, \tag{4.2}$$

where β and δ are shape parameter and they are both positive, while $\lambda > 0$ is a scale parameter. A special case of the Dagum distribution when $\beta = 1$ is the log-logistic distribution, see [(23)].

Another family of distribution that have been used to model income data is the generalized gamma (GG) distribution introduced by [(28)]. It provides a flexible family with a variety of shapes and hazard functions for modeling duration time. It includes the exponential, Weibull, gamma and Rayleigh distributions, among others as special sub-models. It is suitable for modeling data with different types of hazard rate function: increasing, decreasing, bathtub and unimodal. The GG distribution has been used in many applications such as engineering, hydrology and survival analysis. The probability density function of the generalized gamma distribution GG(a, b ,c) is given by

$$f(x;a,b,c) = \frac{a}{c\,\Gamma b}\left[\frac{x}{c}\right]^{(ab-1)} e^{-(\frac{x}{c})^{a}}, x \geqslant 0, a, b, c > 0, \tag{4.3}$$

where $\Gamma(.)$ is the gamma function, b and a are shape parameters, and c is a scale parameter.

An important and useful model of the extended α-power family is where we use the GG (a,1,1) distribution for the $F(\cdot)$ function and use the cumulative distribution function of $Dag(1,\lambda,\delta)$ distribution for the $W(\cdot)$ function with $d1 = 0$ and $d2 = \infty$. In this case, the cdf of the Extended α-Power-Dagum-Generalized gamma *(ExAPT-DG)* model is given by

$$G(x;\theta) = \begin{cases} \frac{\alpha^{\frac{(1+\lambda)}{1+\lambda(1-e^{-x^{a}})^{-\delta}}}-1}{\alpha-1}, & \text{for} \quad \alpha,\lambda,\delta,a > 0, \alpha \neq 1, \\ \frac{(1+\lambda)}{1+\lambda(1-e^{-x^{a}})^{-\delta}}, & \text{for} \quad \alpha = 1. \end{cases} \quad x \in \mathbb{R} \tag{4.4}$$

The probability density function (pdf) corresponding to (4.4) is given by

$$g(x;\theta) = \begin{cases} \frac{log(\alpha) a\lambda\delta x^{a-1}\, e^{-x^{a}} (1+\lambda)\alpha^{\frac{(1+\lambda)}{1+\lambda(1-e^{-x^{a}})^{-\delta}}}}{(\alpha-1)(1-e^{-x^{a}})^{(\delta+1)}[1+\lambda(1-e^{-x^{a}})^{-\delta}]^{2}}, & \text{for} \quad \alpha > 0, \alpha \neq 1, \\ \frac{a\lambda\delta x^{a-1}\, e^{-x^{a}}(1+\lambda)}{(1-e^{-x^{a}})^{(\delta+1)}[1+\lambda(1-e^{-x^{a}})^{-\delta}]^{2}}, & \text{for} \quad \alpha = 1. \end{cases} \tag{4.5}$$

The (sf), (hrf), (rhrf), and (chrf) of *(ExAPT-DG)* distribution, are respectively, given by

$$S(x;\theta) = \frac{\alpha - \alpha^{\frac{(1+\lambda)}{1+\lambda(1-e^{-x^{a}})^{-\delta}}}}{\alpha - 1}, \tag{4.6}$$

$$h(x;\theta) = \frac{log(\alpha) a\lambda\delta\, x^{a-1}\, e^{-x^{a}}(1+\lambda)\alpha^{\frac{(1+\lambda)}{1+\lambda(1-e^{-x^{a}})^{-\delta}}}}{(1-e^{-x^{a}})^{(\delta+1)}[1+\lambda(1-e^{-x^{a}})^{-\delta}]^{2}[\alpha - \alpha^{\frac{(1+\lambda)}{1+\lambda(1-e^{-x^{a}})^{-\delta}}}]}, \tag{4.7}$$

$$r(x;\theta) = \frac{log(\alpha) a\lambda\delta x^{a-1}\, e^{-x^{a}}(1+\lambda)\alpha^{\frac{(1+\lambda)}{1+\lambda(1-e^{-x^{a}})^{-\delta}}}}{\alpha^{\frac{(1+\lambda)}{1+\lambda(1-e^{-x^{a}})^{-\delta}}} - 1}, \tag{4.8}$$

$$H(x;\theta) = -\,log\left[\frac{\alpha - \alpha^{\frac{(1+\lambda)}{1+\lambda(1-e^{-x^{a}})^{-\delta}}}}{\alpha - 1}\right]. \tag{4.9}$$

The plots of of the density and hrf's of the *(ExAPT-DG)* distribution are given in Figures 1 and 2, for different values of parameters.

5 Estimation

Let $X_1, X_2, ..., X_n$ be a simple random sample of size n from the *ExAPT-DG* distribution, then from the pdf in Eq.(4.5) the likelihood function will be in the form

$$\begin{aligned} L(\alpha,\lambda,\mathbf{a},\delta;\underline{x}) = \left(\frac{1}{\alpha-1}\right)^{n} (log(\alpha))^{n}\; a^{n}\lambda^{n}\; \delta^{n}(1+\lambda)^{n} e^{-\sum_{i=1}^{n} x_i^{a}} \prod_{i=1}^{n} x_i^{a-1} \\ \prod_{i=1}^{n}[1+\lambda(1-e^{-x_i^{a}})^{-\delta}]^{2} \prod_{i=1}^{n}\left[1-e^{-x_i^{a}}\right]^{(\delta+1)} \prod_{i=1}^{n}\left[\alpha^{\frac{(1+\lambda)}{1+\lambda(1-e^{(-x_i)^{a}})^{-\delta}}}\right]. \end{aligned} \tag{5.1}$$

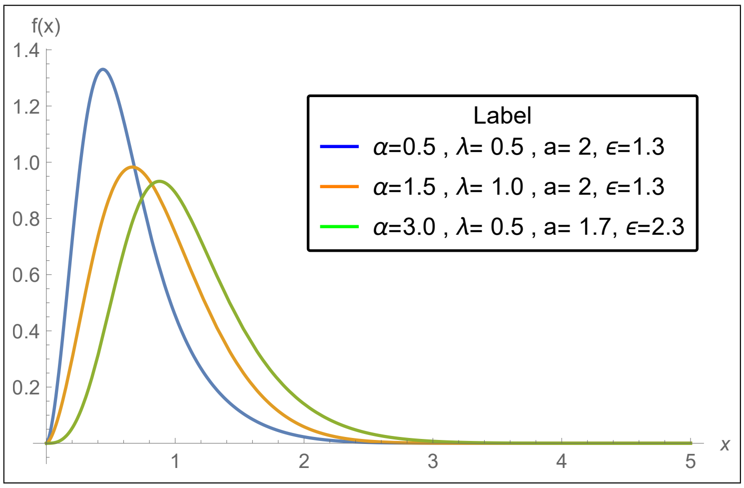

Figure 1: Density functions of *(ExAPT-DG)* distribution, for different values of parameters

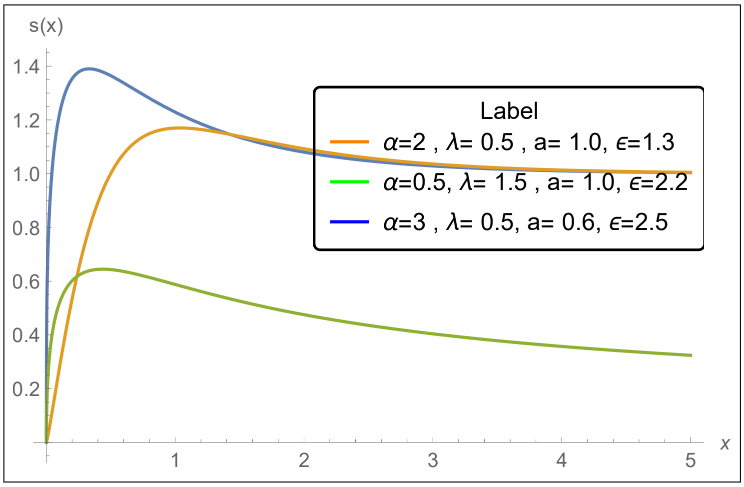

Figure 2: Hazard rate functions of *(ExAPT-DG)* distribution,for different values of parameters

The log likelihood function is

$$
\begin{aligned}
l = & \; nlog(log\alpha) - nlog(\alpha - 1) + nlog(a) + nlog(\lambda) + nlog(\delta) + nlog(1+\lambda) - \\
& \sum_{i=1}^{n} x_i^a + (a-1)\sum_{i=1}^{n} log(x_i) - 2\sum_{i=1}^{n} log(1 + \lambda(1 - e^{-x_i^a})^{-\delta}) \\
& +(\delta+1)\sum_{i=1}^{n} log(1 - e^{-x_i^a}) + log\alpha \sum_{i=1}^{n}\Big[\frac{(1+\lambda)}{1+\lambda(1-e^{(-x_i)^a})^{-\delta}}\Big].
\end{aligned} \tag{5.2}
$$

The maximum likelihood estimators of α, λ, a and δ can be obtained by differentiating l with respect to α, λ, a and δ and equating the results to zero.

$$\frac{\partial l}{\partial \alpha} = \frac{n}{\hat{\alpha} log(\hat{\alpha})} - \frac{n}{\hat{\alpha}-1} + \frac{1}{\hat{\alpha}} \sum_{i=1}^{n} \frac{(1+\hat{\lambda})}{1+\hat{\lambda}(1-e^{(-x_i)^{\hat{a}}})^{-\hat{\delta}}} = 0, \tag{5.3}$$

$$\begin{aligned} \frac{\partial l}{\partial \lambda} &= \frac{n}{\hat{\lambda}} + \frac{n}{1+\hat{\lambda}} - 2\sum_{i=1}^{n} \Big[\frac{(1-e^{-x_i^{\hat{a}}})^{-\hat{\delta}}}{(1+\hat{\lambda}(1-e^{-x_i^{\hat{a}}})^{-\hat{\delta}}}\Big] + \\ log\hat{\alpha} &\sum_{i=1}^{n} \Big[\frac{1}{1+\hat{\lambda}(1-e^{-x_i^{\hat{a}}})^{-\hat{\delta}}} - \frac{(1-e^{-x_i^{\hat{a}}})^{-\hat{\delta}}(1+\hat{\lambda})}{(1+\hat{\lambda}(1-e^{-x_i^{\hat{a}}})^{-\hat{\delta}})^2}\Big] = 0, \end{aligned} \tag{5.4}$$

$$\begin{aligned} \frac{\partial l}{\partial a} = \frac{n}{\hat{a}} &+ \sum_{i=1}^{n} log(x_i) x_i^{\hat{a}} + (\hat{\delta}+1) \sum_{i=1}^{n} \frac{e^{-x_i^{\hat{a}}}\ x_i^{\hat{a}}\ log(x_i)}{(1-e^{-x_i^{\hat{a}}})} \\ &-2\ \hat{\lambda}\ \hat{\delta} \sum_{i=1}^{n} \frac{e^{-x_i^{\hat{a}}}\ x_i^{\hat{a}}\ log(x_i)\ (1-e^{-x_i^{\hat{a}}})^{-(\hat{\delta}+1)}}{1+\hat{\lambda}(1-e^{-x_i^{\hat{a}}})^{-\hat{\delta}}} + log(\hat{\alpha}) \\ &\sum_{i=1}^{n} \frac{(1+\hat{\lambda})\hat{\lambda}\hat{\delta} e^{-x_i^{\hat{a}}}\ x_i^{\hat{a}}\ log(x_i)(1-e^{-x_i^{\hat{a}}})^{-(\hat{\delta}+1)}}{(1+\hat{\lambda}(1-e^{-x_i^{\hat{a}}})^{-\hat{\delta}})^2} = 0, \end{aligned} \tag{5.5}$$

and,

$$\begin{aligned} \frac{\partial l}{\partial \delta} = \frac{n}{\hat{\delta}} &- \sum_{i=1}^{n} log(1-e^{-x_i^{\hat{a}}}) + 2\sum_{i=1}^{n} \frac{\hat{\lambda}\ (1-e^{-x_i^{\hat{a}}})^{-\hat{\delta}}\ log(1-e^{-x_i^{\hat{a}}})}{1+\hat{\lambda}(1-e^{-x_i^{\hat{a}}})^{-\hat{\delta}}} \\ &+ log(\hat{\alpha}) \sum_{i=1}^{n} \frac{\hat{\lambda}(1+\hat{\lambda})(1-e^{-x_i^{\hat{a}}})\ log(1-e^{-x_i^{\hat{a}}})}{(1+\hat{\lambda}(1-e^{-x_i^{\hat{a}}})^{-\hat{\delta}})^2} = 0. \end{aligned} \tag{5.6}$$

The system of the non-linear equations cannot be solved explicitly, so the maximum likelihood estimators of α, λ, a and δ can be obtained by using a numerical technique.

6 Simulation study

A simulation study is carried out to asses the performance of the maximum likelihood estimation for *(ExAPT-DG)* by sample size n. The performance of the estimates for the parameters has been studied in term of their ; the biases and the mean square errors (MSEs) using the *Mathematica* 11 package. The numerical steps are listed as follows:

1. N=1000 random samples of sizes; $n = 30, 50\ and\ 100$ from the *(ExAPT-DG)* distribution are generated;
2. Selected initial guess values for the parameters are used;
3. The maximum likelihood estimates (mles) of *(ExAPT-DG)* model are evaluated for each sample size.
4. Calculate the biases and mean square error (MSE) of these estimators using Mathematica 11.

Some results of the simulation study are shown in the following Tables .

Table 1: The parameter estimation from *(ExAPT-DG)* distribution using maximum likelihood method.

n	par	Set 1 $(\alpha = 1.6; \lambda = 2.7; a = 0.8; \delta = 0.6)$			Set 2 $(\alpha = 2.5; \lambda = 1.7; a = 1.9; \delta = 1.9)$		
		MLE	Bias	MSE	MLE	Bias	MSE
30	α	1.2446	0.1262	0.1287	2.3891	0.0122	0.0338
	λ	1.2001	2.2494	2.2509	1.0783	0.3864	0.3901
	a	0.7333	0.0044	0.0052	1.5949	0.0931	0.0994
	δ	0.3754	0.0504	0.0504	1.4927	0.1658	0.1680
50	α	1.2727	0.1071	0.1112	2.4427	0.0032	0.0432
	λ	2.4343	0.0705	0.0782	1.1548	0.2972	0.3039
	a	0.7868	0.00017	0.0013	1.8067	0.0086	0.0218
	δ	0.4174	0.0333	0.0334	1.7753	0.0155	0.0224
100	α	1.5040	0.0092	0.0174	2.4604	0.0015	0.0325
	λ	2.4808	0.0480	0.0594	1.2570	0.1961	0.2019
	a	0.7973	0.00007	0.0012	1.8560	0.0019	0.0114
	δ	0.5050	0.0090	0.0092	1.8257	0.0055	0.0119

Table 2: The parameter estimation from *(ExAPT-DG)* distribution using maximum likelihood method.

n	par	Set 1 $(\alpha = 1.5; \lambda = 0.5; a = 1.3; \delta = 0.5)$			Set 2 $(\alpha = 0.8; \lambda = 0.7; a = 1.3; \delta = 2.7)$		
		MLE	Bias	MSE	MLE	Bias	MSE
30	α	1.2206	0.0780	0.0788	0.4969	0.0918	0.0921
	λ	0.3128	0.0350	0.0351	0.5223	0.0315	0.0316
	a	0.8786	0.1775	0.1777	0.7752	0.2753	0.2759
	δ	0.4473	0.0027	0.0028	2.1802	0.2701	0.2703
50	α	1.3458	0.0237	0.0258	0.7591	0.0016	0.0030
	λ	0.3988	0.0102	0.0105	0.5418	0.0250	0.0253
	a	1.2114	0.0078	0.0085	1.0601	0.0570	0.0585
	δ	0.4181	0.0066	0.0067	2.2325	0.2184	0.2196
100	α	1.6635	0.0267	0.0350	0.7746	0.0006	0.0020
	λ	0.4854	0.0002	0.0006	0.5730	0.0161	0.0164
	a	1.4148	0.0132	0.0166	1.2372	0.0039	0.0061
	δ	0.4149	0.0052	0.0056	2.3162	0.1472	0.1494

Table 3: The parameter estimation from *(ExAPT-DG)* distribution using maximum likelihood method.

n	par	Set 1 $(\alpha = 0.8; \lambda = 1.5; a = 2.7; \delta = 1.3)$			Set 2 $(\alpha = 2; \lambda = 1.5; a = 2.7; \delta = 1.3)$		
		MLE	Bias	MSE	MLE	Bias	MSE
30	α	0.4764	0.1040	0.1047	1.4280	0.3271	0.3340
	λ	0.7835	0.5133	0.5134	1.0378	0.2135	0.2160
	a	1.2248	0.1761	0.2214	2.3650	0.1122	0.1269
	δ	1.0309	0.0723	0.0725	1.0221	0.0772	0.0778
50	α	0.6862	0.0129	0.0134	1.5537	0.1991	0.2102
	λ	1.1929	0.0942	0.0949	1.3368	0.0266	0.0315
	a	1.7206	0.9592	0.9598	2.6433	0.0032	0.0290
	δ	1.1099	0.0361	0.0362	1.1245	0.0307	0.0326
100	α	0.7224	0.0060	0.0076	1.8225	0.0315	0.0567
	λ	1.3228	0.0313	0.0331	1.4417	0.0033	0.0124
	a	2.3864	0.0982	0.1029	2.6701	0.0001	0.0254
	δ	1.2345	0.0042	0.0051	1.2447	0.0030	0.0064

Table 4: The parameter estimation from *(ExAPT-DG)* distribution using maximum likelihood method.

n	par	Set 1 $(\alpha = 0.5; \lambda = 1.5; a = 2.7; \delta = 1.3)$			Set 2 $(\alpha = 3; \lambda = 0.5; a = 0.7; \delta = 0.9)$		
		MLE	Bias	MSE	MLE	Bias	MSE
30	α	0.4094	0.0082	0.0123	2.6912	0.0954	0.0991
	λ	1.2787	0.0489	0.0573	0.4527	0.0022	0.0023
	a	2.6836	0.00026	0.0411	0.5751	0.0155	0.0157
	δ	1.2906	0.00008	0.0058	0.5782	0.1034	0.1035
50	α	0.4110	0.0079	0.0104	2.7720	0.0519	0.0602
	λ	1.2778	0.0439	0.0547	0.5125	0.0002	0.0005
	a	2.6732	0.0007	0.0241	0.6771	0.0005	0.0011
	δ	1.2847	0.0002	0.0035	0.6838	0.0467	0.0470
100	α	0.4472	0.0027	0.0047	2.9560	0.0019	0.0126
	λ	1.4523	0.0022	0.0063	0.4632	0.0013	0.0016
	a	2.6472	0.0027	0.0142	0.7307	0.0009	0.0018
	δ	1.2907	0.0024	0.0031	0.7537	0.0213	0.0218

It is noticed, from Tables 1-4, it can easily be detected, that the ML averages are very close to the true values of the parameters. Also, bias and MSE are decrease when the sample size increases. This is indicative of the fact that the estimates are consistent and approaches the population parameter values as the sample size increases.

7 Application

A sample is taken from the Egyptian household income, expenditure and consumption survey (HIECS) for a period of time in $2015\backslash 2016$. The original sample size is 11988 households and the sample taken was of 1200 households. The data represents the income of household in units of LE 10000. An attempt was made to fit the *(ExAPT-DG)* to the sample at hand. The computational software *Mathematica* 11 was used to solve the likelihood normal equations from Eq.(5.3) to Eq.(5.6). The descriptive statistics for the distribution of income are shown in the following table.

Table 5: Descriptive statistics for the distribution of income in period $2015\backslash2016$

Obs.	Min	Max	Mean	Var.	Skewness	Kurtosis
1200	0.522	21.5435	4.0269	4.9840	2.2687	12.3323

Solving the likelihood normal equations yield the following results $\alpha = 8.8931$, $\lambda = 0.4463$, $a = 0.7752$, $\delta = 10.7622$. The cdf using these parameters and the empirical distribution of income household data can show how the data fits of the concerning distribution and Figure 3 shows how the two curves of the empirical and the cumulative distributions are almost identical.

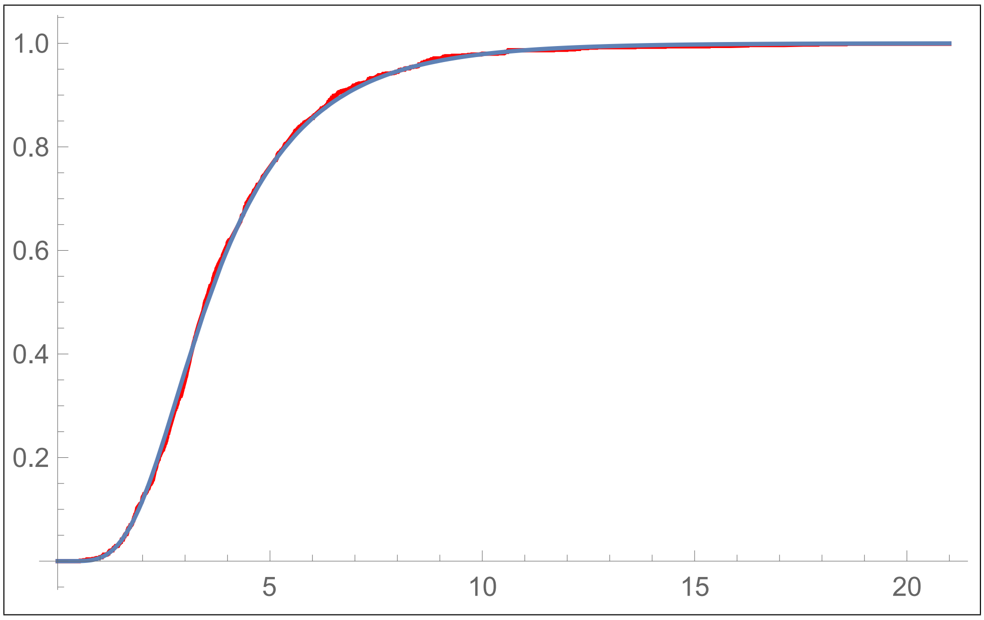

Figure 3: Empirical and Cumulative Distribution

The data is tested if it fits the distribution or not where the null hypothesis and its alternative will be:

H_0: The dataset follow the ExAPT-DG distribution.

against

H_1: The dataset do not follow the ExAPT-DG distribution.

After using different tests such as Kolmogorov Smirnov, Anderson Darling and Cramer Von Mises. The next table shows results of the diffrent kinds of tests and test statistics for each test and also the p values:

Table 6: Results Using Different Goodness of Fit Tests

Test Name	Test statistics	pValue
Kolmogorov Smirnov	0.02	0.66
Anderson Darling	0.5	0.66
Cramer Von Mises	0.1	0.60

The results of goodness of fit test, in the previous table, for the real data of income using different tests show that the null hypothesis the datasets have the same distribution can not be rejected at the 5 percent of significance level.

7.1 Income Inequality Measures

The Gini index is the Gini coefficient expressed as a percentage, and is equal to the Gini coefficient multiplied by 100. The Gini coefficient is equal to half of the relative mean difference (a measure of dispersion). The Gini coefficient is a measure of income inequality which is an indication of social welfare. It is defined as a ratio with values between 0 and 1. 0 corresponds to perfect income equality (i.e. everyone has the same income) and one corresponds to perfect income inequality (i.e. 1 person has all the income, while everyone else has zero income), see [(30)].

The Gini index can be expressed mathematically for a cumulative distribution function as shown in the following formula

$$G = 1 - \frac{1}{\mu}\int_0^\infty (1 - F(x))^2 dx, \tag{7.1}$$

where μ is the mean of the distribution, see [(29)].

Applying the previous Eq.(7.1) using the cumulative distribution of *(ExAPT-DG)* as in Eq.(4.4) and the parameters $\alpha = 8.8931$, $\lambda = 0.4463$, $a = 0.7752$ and $\delta = 10.7622$ of the real income data of HIECS $2015\backslash2016$ survey, then the following equation

$$G = 1 - \frac{1}{\mu_1}\int_0^\infty \Big(1 - \frac{\alpha^{\frac{(1+\lambda)}{1+\lambda(1-e^{-x^a})^{-\delta}}} - 1}{\alpha - 1}\Big)^2 dx, \tag{7.2}$$

where, $\mu_1 = \int_0^\infty \frac{log(\alpha) a\lambda\delta x^a \; e^{-x^a}(1+\lambda)\alpha^{\frac{(1+\lambda)}{1+\lambda(1-e^{-x^a})^{-\delta}}}}{(\alpha-1)(1-e^{-x^a})^{(\delta+1)}[1+\lambda(1-e^{-x^a})^{-\delta}]^2}$. According to previous equation and these parameters, we get that Gini index equals 0.275.

8 CONCLUSIONS

In this article, a new family of distributions, called the $(ExAPT)$ family of distributions is presented. General expressions for some of mathematical properties of the new family are derived. A special sub-model of the new family, called *(ExAPT-DG)* distribution is considered. The estimation of the model parameters through maximum likelihood method is discussed. A numerical evaluation is carried out to examine the performance of mles for *(ExAPT-DG)* model. The new model was fitted to a real data set from the income of Egyptian household.

References

[1] Pearson, K. X. Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Philosophical Transactions Of The Royal Society Of London.(A.).*, 343-414 (1895).

[2] Burr, I. Cumulative frequency functions. *The Annals Of Mathematical Statistics*. **13**, 215-232 (1942).

[3] Azzalini, A. A class of distributions which includes the normal ones.. *Scandinavian Journal Of Statistics*. **12** pp. 171-178 (1985).

[4] Mudholkar, G. & Srivastava, D. Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Transactions On Reliability*. **42**, 299-302 (1993).

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

[6] Eugene, N., Lee, C. & Famoye, F. Beta-normal distribution and its applications.. *Communications In Statistics-Theory And Methods*. **31**, 497-512 (2002).

[7] Jones, M. Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology*. **6**, 70-81 (2009).

[8] Alzaatreh, A., Lee, C. & Famoye, F. A new method for generating families of continuous distributions.. *Metron*. **71**, 63-79 (2013).

[9] Aljarrah, M., Lee, C. & Famoye, F. On generating TX family of distributions using quantile functions.. *Journal Of Statistical Distributions And Applications*. **1**, 1-17 (2014).

[10] Mahdavi, A. & Kundu, D. A new method for generating distributions with an application to exponential distribution. *Communications In Statistics-Theory And Methods*. **46**, 6543-6557 (2017)

[11] Nassar, M., Alzaatreh, A., Mead, M. & Abo-Kasem, O. Alpha power Weibull distribution: Properties and applications. *Communications In Statistics-Theory And Methods*. **46**, 10236-10252 (2017).

[12] Nadarajah, S. & Okorie, I. On the moments of the alpha power transformed generalized exponential distribution. *Ozone: Science Engineering*. **40**, 330-335 (2018).

[13] Mead, M., Cordeiro, G., Afify, A. & Al Mofleh, H. The alpha power transformation family: properties and applications. *Pakistan Journal Of Statistics And Operation Research*. pp. 525-545 (2019).

[14] Nassar, M., Afify, A. & Shakhatreh, M. Estimation methods of alpha power exponential distribution with applications to engineering and medical data. *Pakistan Journal Of Statistics And Operation Research*. pp. 149-166 (2020).

[15] ElSherpieny, E. & Almetwally, E. The Exponentiated Generalized Alpha Power Family of Distribution: Properties and Applications. *Pakistan Journal Of Statistics And Operation Research*. pp. 349-367 (2022).

[16] Elbatal, I., Ahmad, Z., Elgarhy, M. & Almarashi, A. A new alpha power transformed family of distributions: properties and applications to the Weibull model. *Journal Of Nonlinear Science And Applications*. **12**, 1-20 (2018).

[17] Ahmad, Z. The Zubair-G family of distributions: Properties and applications. *Annals Of Data Science*. **7**, 195-208 (2018).

[18] Ahmad, Z., Elgarhy, M. & Abbas, N. A new extended alpha power transformed family of distributions: properties and applications. *Journal Of Statistical Modelling: Theory And Applications*. **1**, 13-27 (2018).

[19] Gupta, R. & Bradley, D. Representing the mean residual life in terms of the failure rate. *Mathematical And Computer Modelling*. **37**, 1271-1280 (2003).

[20] Nair, U., Sankaran, P. & Balakrishnan, N. Reliability modelling and analysis in discrete time. (Academic Press,2018).

[21] David, H. Order statistics, (2nd ed).. (New York: J. Wiley,1981).

[22] Dagum, C. A New Model for Personal Income Distribution: Specification and Estimation. *Economie Appliquée*. **30**, 413-437 (1977).

[23] Domma, F., Giordano, S. & Zenga, M. Maximum likelihood estimation in Dagum distribution with censored samples. *Journal Of Applied Statistics*. **38**, 2971-2985 (2011).

[24] Domma, F. & Condino, F. The beta-dagum distribution: definition and properties. *Communications In Statistics-Theory And Methods*. **42**, 4070-4090 (2013).

[25] Kleiber, C. & Kotz, S. Statistical size distributions in economics and actuarial sciences. (John Wiley Sons,2003).

[26] Kleiber, C. A guide to the Dagum distributions. *Modeling Income Distributions And Lorenz Curves*. pp. 97-117 (2008).

[27] Oluyede, B. & Rajasooriya, S. The Mc-Dagum distribution and its statistical properties with applications. *Asian Journal Of Mathematics And Applications*. **2013** (2013).

[28] Stacy, E. A generalization of the gamma distribution. *The Annals Of Mathematical Statistics*. pp. 1187-1192 (1962).

[29] Dedduwakumara, D. & Prendergast, L. Interval estimators for inequality measures using grouped data. *Research School On Statistics And Data Science*. pp. 238-252 (2019).

[30] Dixon, P., Weiner, J., MitchellOlds, T. & Woodley, R. Erratum to Bootstrapping the Gini Coefficient of Inequality. *Ecology*. **69**, 1307 (1988).