A New Method for Generating Distributions with an Application to Exponential Distribution

Abbas Mahdavi[†] & Debasis Kundu[‡]

Abstract

A new method has been proposed to introduce an extra parameter to a family of distributions for more flexibility. A special case has been considered in details namely; one parameter exponential distribution. Various properties of the proposed distribution, including explicit expressions for the moments, quantiles, mode, moment generating function, mean residual lifetime, stochastic orders, order statistics and expression of the entropies are derived. The maximum likelihood estimators of unknown parameters cannot be obtained in explicit forms, and they have to be obtained by solving non-linear equations only. Further we consider an extension of the two-parameter exponential distribution also, mainly for data analysis purposes. Two data sets have been analyzed to show how the proposed models work in practice.

Key Words and Phrases: exponential distribution; hazard rate function; maximum-likelihood estimation; survival function; Fisher information matrix.

AMS Subject Classification: Primary: 62E15, Secondary: 62H10.

† Department of Statistics, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran; E-mail: a.mahdavi@vru.ac.ir. Contacting author.

‡ Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India.

1 Introduction

Adding an extra parameter to an existing family of distribution functions, is very common in the statistical distribution theory. Often introducing an extra parameter brings more flexibility to a class of distribution functions, and it can be very useful for data analysis purposes. For example, Azzalini (1985) introduced the skew normal distribution by introducing an extra parameter to the normal distribution to bring more flexibility to the normal distribution. Azzalini (1985)'s skew normal distribution takes the following form for $\lambda \in \mathbb{R}$;

$$f(x;\lambda) = 2\phi(x)\Phi(\lambda x), \quad x \in \mathbb{R}. \tag{1}$$

Here $\phi(x)$ and $\Phi(x)$ are the probability density function (PDF) and cumulative distribution function (CDF) of a standard normal distribution, and λ is the skewness parameter. It can be easily seen that $f(x;\lambda)$ is a proper PDF for any $\lambda \in \mathbb{R}$. Moreover, when $\lambda = 0$, $f(x;\lambda)$ becomes a standard normal PDF, and depending on the values of λ, it can be both positively or negatively skewed. Hence, introducing λ brings more flexibility to the standard normal distribution function. Although Azzalini (1985) introduced this method mainly for normal distribution, but it can be easily used for other symmetric distributions also.

Mudholkar and Srivastava (1993) proposed a method to introduce an extra parameter to a two-parameter Weibull distribution. If for $\alpha > 0$ and $\lambda > 0$, $(1 - e^{-\lambda x^{\alpha}})$ denotes the CDF of a two-parameter Weibull distribution, then the proposed model has the following CDF for $\beta > 0$;

$$F(x;\alpha,\lambda,\beta) = (1 - e^{-\lambda x^{\alpha}})^{\beta}, \quad \text{if} \quad x > 0, \tag{2}$$

and 0 otherwise. The two-parameter Weibull distribution has one shape (α) and one scale (λ) parameters. Mudholkar and Srivastava (1993)'s proposed exponentiated Weibull model has two shape parameters and one scale parameter. Due to presence of an extra shape parameter, the proposed exponentiated Weibull model is more flexible than the two-parameter Weibull model. In this case also, when $\beta = 1$, the proposed exponentiated Weibull distribution coincides with the two-parameter Weibull distribution. Although, Mudholkar and Srivastava (1993) proposed the exponentiated Weibull distribution, later several other exponentiated distributions have been introduced by several authors, see for example Gupta et al. (1998).

Marshall and Olkin (1997) proposed another method to introduce an additional parameter to any distribution function as follows. If $g(x)$ and $G(x)$ are the PDF and CDF, respectively of a random variable X, then the new proposed family of distribution functions has the following PDF for any $\theta \in (0, \infty)$.

$$f(x; \theta) = \frac{\theta g(x)}{(1 - (1 - \theta)(1 - G(x)))^2}; \quad x \in \mathbb{R} \tag{3}$$

Marshall and Olkin (1997) considered two special cases namely when X follows exponential or Weibull distribution and derived several properties of this proposed model. They had also provided some interesting physical interpretations also. It is clear that when $\theta = 1$, $f(x; \theta)$ matches with $g(x)$, and for different values of θ, $f(x)$ can be more flexible than $g(x)$.

Eugene et al. (2002) proposed the beta generated method that uses the beta distribution with parameters α and β as the generator to develop the beta generated distributions. The CDF of a beta-generated random variable X is defined as

$$G(x) = \int_0^{F(x)} b(t)dt, \tag{4}$$

where $b(t)$ is the PDF of a beta random variable and $F(x)$ is the CDF of any random variable X. Alzaatreh et al. (2013) introduced a new method for generating families of continuous distributions called T-X family by replacing the beta PDF with a PDF, $r(t)$, of a continuous random variable and applying a function $W(F(x))$ that satisfies some specific conditions. Recently Aljarrah et al. (2014) used quantile functions to generate T-X family of distributions. For more survey about methods to generating distributions see Lee et al. (2013) and Jones (2014).

The aim of this paper is to introduce an extra parameter to a family of distributions functions to bring more flexibility to the given family. We call this new method as α-power transformation (APT) method. The proposed APT method is very easy to use, hence it can be used quite effectively for data analysis purposes. First we discuss some general propoerties of this class of distribution functions. Then, we apply the APT method to a one-parameter exponential distribution and generated a two-parameter (scale and shape) α-power exponential (APE) distribution. It is observed that the two-parameter APE distribution has several desirable properties. It behaves very much like two parameter Weibull, gamma or generalized exponential (GE) distributions. The PDF and the hazard functions of APE distribution can take similar shapes as the Weibull, gamma or GE distributions. Hence, it can be used as an alternative to the well known Weibull, gamma or GE distributions. The CDF of APE distribution can be expressed in explicit form, hence it can be used quite conveniently for analyzing censored or truncated data also. Further we introduce three-parameter APE distribution mainly for data analysis purposes, and discuss the maximum likelihood estimation procedure of the unknown parameters. One real data set has been analyzed for illustrative purposes.

The rest of the paper is organized as follows. In Section 2, we introduce the APT method, and discuss some general properties of this family of distributions. In Section 3, we consider the APE distribution and discuss its different properties. Maximum likelihood estimators of the unknown parameters and some inferential issues are discussed in Section 4. The analysis of one real data set has been presented in Section 5. We introduce the three-parameter APE model and discuss the maximum likelihood estimation procedure of the unknown parameters in Section 6. Finally in Section 7 we conclude the paper.

2 APT Method

Let $F(x)$ be the CDF of a continuous random variable X, then the α-power transformation of $F(x)$ for $x \in \mathbb{R}$, is defined as follows:

$$F_{APT}(x) = \begin{cases} \frac{\alpha^{F(x)}-1}{\alpha-1} & \text{if} \quad \alpha > 0, \alpha \neq 1 \\ F(x) & \text{if} \qquad \alpha = 1. \end{cases} \tag{5}$$

It is immediate that $F_{APT}(x)$ is a proper CDF. If $F(x)$ is an absolute continuous distribution function with the probability density function (PDF) $f(x)$, then $F_{APT}(x)$ is also an absolute continuous distribution function with the PDF:

$$f_{APT}(x) = \begin{cases} \frac{\log \alpha}{\alpha-1} f(x)\alpha^{F(x)} & \text{if} \quad \alpha > 0, \alpha \neq 1 \\ f(x) & \text{if} \qquad \alpha = 1. \end{cases} \tag{6}$$

It is clear that for $\alpha \neq 1$, $f_{APT}(x)$ is a weighted version of $f(x)$, where the weight function

$$w(x) = \alpha^{F(x)}, \tag{7}$$

and $f_{APT}(x)$ can be written as

$$f_{APT}(x) = \frac{f(x)w(x;\alpha)}{c}. \tag{8}$$

Here the normalizing constant $c = E(w(X))$.

Weighted distributions play an important role in the statistical distribution theory, see for example Patil and Rao (1978) or Patil (2002). In this case the weight function $w(x)$ can be increasing or decreasing depending on whether $\alpha > 1$ or $\alpha < 1$. For a non-negative random variable X, $f_{APT}(x)$ has the following interpretations. Suppose a realization x of X enters the investigator's record with probability proportional to $w(x;\alpha)$, so that

$$\frac{Prob(\text{Recording}|X = y)}{Prob(\text{Recording}|X = x)} = \frac{w(y;\alpha)}{w(x;\alpha)}.$$

Here the parameter α represents recording mechanism. If Y represents the random variable of the investigator's records, then Y has the PDF $f_{APT}(x)$.

The survival reliability function $S_{APT}(x)$ and the hazard rate function (HRF) $h_{APT}(x)$ for APT distribution are in the following forms

$$S_{APT}(x) = \begin{cases} \frac{\alpha}{\alpha-1}(1-\alpha^{F(x)-1}) & \text{if} \quad \alpha \neq 1 \\ \\ 1-F(x) & \text{if} \quad \alpha = 1, \end{cases} \tag{9}$$

$$h_{APT}(x) = \begin{cases} f(x)\frac{\alpha^{F(x)-1}}{1-\alpha^{F(x)-1}}\log\alpha & \text{if} \quad \alpha \neq 1 \\ \\ \frac{f(x)}{S(x)} & \text{if} \quad \alpha = 1. \end{cases} \tag{10}$$

The p-th quantile y_p of $F_{APT}(x)$, for $\alpha \neq 1$, can be obtained as

$$y_p = F^{-1}\left\{\frac{\log(1+(\alpha-1)p)}{\log\alpha}\right\}. \tag{11}$$

If x_p denotes the p-th quantile for $F(x)$, then from (11) it follows that for $\alpha \neq 1$,

$$y_p \leq x_p \quad \text{if} \quad \frac{\log(1+(\alpha-1)p)}{\log\alpha} \leq p. \tag{12}$$

Using (12) it is possible to determine for what values of α, $F_{APT}(x)$ will be heavier tail than $F(x)$. For example, using Taylor series expansion of $\log(1+x)$, it easily follows that for α close to 1,

$$y_p < x_p \quad \text{if} \;\; \alpha < 1 \quad \text{and} \quad y_p > x_p \quad \text{if} \quad \alpha > 1. \tag{13}$$

Therefore, for α close to 1, if $\alpha < 1$, then $F(x)$ has a heavier tail than $F_{APT}(x)$, and for $\alpha > 1$, it is the other way.

If $F^{-1}(\cdot)$ exists in explicit form, then for $\alpha \neq 1$, a random sample from $F_{APT}(x)$ can be easily obtained as

$$Y = F^{-1}\left\{\frac{\log(1+(\alpha-1)U)}{\log\alpha}\right\},$$

where U is a uniform $(0,1)$. On the other hand if $F^{-1}(\cdot)$ does not exist in explicit form, but it is possible to generate a random sample from $F(x)$, then from the relation for $\alpha \neq 1$,

$$f_{APT}(x) \leq \frac{\alpha\log\alpha}{\alpha-1}f(x),$$

and using acceptance-rejection principle, a random sample from $F_{APT}(x)$ can be generated. If the hazard function and survival function of X are denoted by $h(x)$ and $S(x)$, respectively, then from (10), for $\alpha \neq 1$, the hazard function $F_{APT}(x)$ can be written as

$$h_{APT}(x) = h(x) \times \frac{\log \alpha \; S(x)}{\alpha^{S(x)} - 1}. \tag{14}$$

Therefore,

$$\lim_{x \to -\infty} h_{APT}(x) = \frac{\log \alpha}{\alpha - 1} \lim_{x \to -\infty} h(x) \quad \text{and} \quad \lim_{x \to \infty} h_{APT}(x) = \lim_{x \to \infty} h(x).$$

From (14) it follows that $h_{APT}(x)/h(x)$ is an increasing function of x for all $\alpha > 1$, and it is a decreasing function of x, for all $0 < \alpha < 1$.

We have the following results for a general distribution function $F(x)$.

RESULT 1: If $f(x)$ is a decreasing function, and $\alpha \leq 1$, then $f_{APT}(x)$ is a decreasing function.

PROOF: The result easily follows by taking $\log f_{APT}(x)$, and by using the fact that sum of two decreasing functions is a decreasing function.

RESULT 2: If $f(x)$ is a decreasing function, and $f(x)$ is log-convex, then for $\alpha \leq 1$, the hazard function $h_{APT}(x)$ is a decreasing function.

PROOF: Since

$$\frac{d^2 \log f_{APT}(x)}{dx^2} = \frac{d^2 \log f(x)}{dx^2} + \log \alpha \frac{df(x)}{dx},$$

and both the terms on the right hand side are positive, it implies that $f_{APT}(x)$ is log-convex. Hence the result follows from Bartow and Prochan (1975). ∎

Using Results 1 and 2, it easily follows that if $F(x)$ is a gamma, Weibull or generalized exponential distribution with shape parameter less than one, then for $\alpha \leq 1$, $f_{APT}(x)$ and $h_{APT}(x)$ are both decreasing functions.

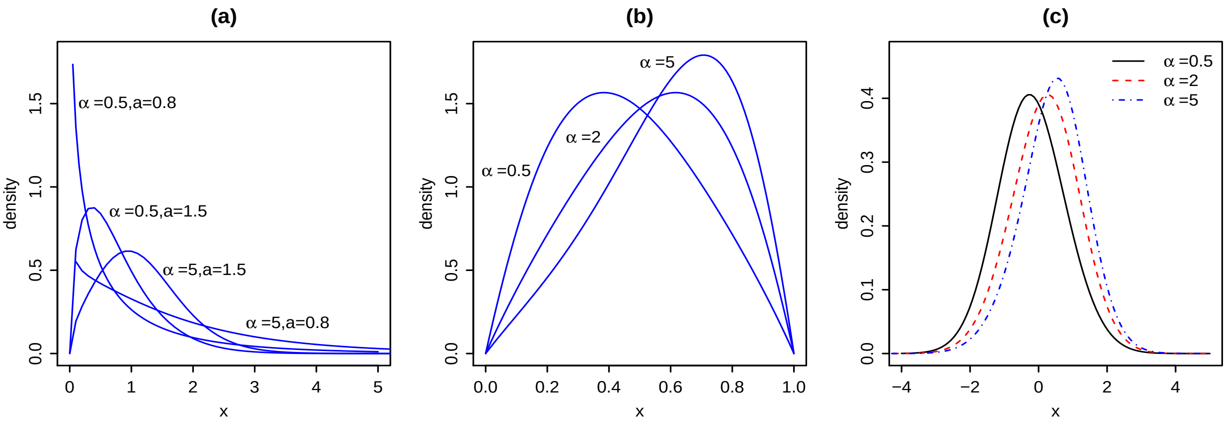

Figure 1: (a) the PDF of Weibull APT family. (b) the PDF of Beta APT family. (c) the PDF of Normal APT family.

In Figure 1, we plot the PDFs of some APT families such as when $f(x)$ is the PDFs of Weibull distribution with density function $f(x) = (a/b)(x/b)^{(a-1)}exp(-(x/b)^a)$, Beta distribution and Normal distribution. We consider in Weibull case scale parameter $b = 1$, in Beta case $a = b = 2$ and in the last case its standard form.

3 APE Distribution and its Properties

In this section we apply the APT method to a specific class of distribution functions, namely to an exponential distribution, and call this new distribution as the two-parameter APE distribution.

Definition 1: The random variable Y is said to have a two-parameter APE distribution denoted by $APE(\alpha, \lambda)$, with the shape and scale parameters as $\alpha > 0$ and $\lambda > 0$, respectively, if the PDF of Y for $y > 0$, is

$$f(y;\alpha,\lambda) = \begin{cases} \frac{\log\alpha}{\alpha-1}\lambda e^{-\lambda y}\alpha^{1-e^{-\lambda y}} & \text{if} \quad \alpha \neq 1 \\ \lambda e^{-\lambda y} & \text{if} \quad \alpha = 1, \end{cases} \tag{15}$$

and 0 otherwise.

The CDF of Y for $y > 0$, becomes

$$F(y;\alpha,\lambda) = \begin{cases} \frac{\alpha^{(1-e^{-\lambda y})}-1}{\alpha-1} & \text{if} \quad \alpha \neq 1 \\ 1 - e^{-\lambda y} & \text{if} \quad \alpha = 1, \end{cases} \tag{16}$$

and 0 otherwise. Also, the survival reliability function, $S(y)$, and the hazard rate function, $h(y)$, for APE distribution for $y > 0$, are in the following forms

$$S(y;\alpha,\lambda) = \begin{cases} \frac{\alpha}{\alpha-1}(1 - \alpha^{-e^{-\lambda y}}) & \text{if} \quad \alpha \neq 1 \\ e^{-\lambda y} & \text{if} \quad \alpha = 1, \end{cases} \tag{17}$$

$$h(y;\alpha,\lambda) = \begin{cases} \frac{\lambda e^{-\lambda y}\alpha^{-e^{-\lambda y}}\log\alpha}{1-\alpha^{-e^{-\lambda y}}} & \text{if} \quad \alpha \neq 1 \\ \lambda & \text{if} \quad \alpha = 1. \end{cases} \tag{18}$$

It can be easily seen that if $\alpha \leq e$, then $f(y;\alpha,\lambda)$ is a decreasing function of $y > 0$, and for $\alpha > e$, $f(y;\alpha,\lambda)$ is a unimodal function with mode at $(\log(\log\alpha))/\lambda$. The following result provides the shape of the HRF of an APE distribution.

Result 3: If $\alpha \leq 1$, then $h(y;\alpha,\lambda)$ is a decreasing function of $y > 0$, and for $\alpha > 1$, $h(y;\alpha,\lambda)$ is an increasing function of $y > 0$.

Proof: By taking the second derivative of $\log f(x;\alpha,\lambda)$, it easily follows that the PDF of APE(α,λ) is log-convex if $\alpha < 1$ and log-concave if $\alpha > 1$, hence the result follows from Bartow and Prochan (1975). ∎

It follows that for $\alpha < 1$, $h(y;\alpha,\lambda)$ decreases from $\lambda\log\alpha/(\alpha-1)$ to λ, and for $\alpha > 1$, it increases from $\lambda\log\alpha/(\alpha-1)$ to λ. The following Table 1 provides the comparison of the hazard function of the APE distribution with the corresponding hazard functions of Weibull, gamma and GE distributions. In all these cases the shape and scale parameters are assumed to be α and λ, respectively. It is clear from Table 1 that although the hazard function of the APE distribution is a decreasing or an increasing function depending on the shape parameter similarly as the gamma, Weibull or GE distributions, the ranges are quite different.

Table 1: Behavior of the hazard functions of the four distributions.

Parameter	Gamma	Weibull	GE	APE
$\alpha = 1$	λ	λ	λ	λ
$\alpha > 1$	Increasing from 0 to λ	Increasing from 0 to ∞	Increasing from 0 to λ	Increasing from λ to $\lambda \log \alpha/(\alpha - 1)$
$\alpha < 1$	Decreasing from ∞ to λ	Decreasing from ∞ to 0	Decreasing from ∞ to λ	Decreasing from λ to $\lambda \log \alpha/(\alpha - 1)$

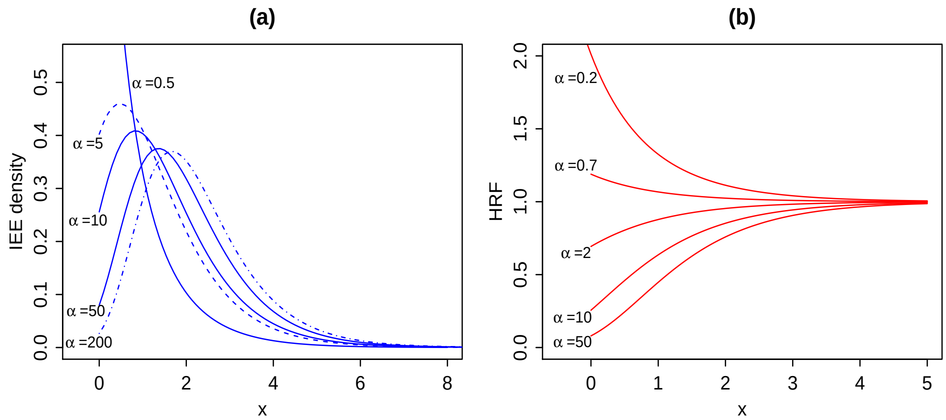

Figure 2: (a) and (b) the PDF and HRF of APE distribution with various shape parameter and fixed scale parameter $\lambda = 1$.

Figure 2 provides the PDFs and HRFs of a APE(α, λ) distribution for different values of α when the scale parameter $\lambda = 1$.

APE distribution has the following mixture representation for $\alpha > 1$. $\log(\alpha)/(\alpha - 1)$ is a decreasing function from 1- to 0, as α varies from 1+ to ∞. If $Y \sim$ APE(α, λ), then it can be represented as follows:

$$Y = \begin{cases} Y_1 & \text{with probability} \quad \frac{\log(\alpha)}{\alpha-1} \\ Y_2 & \text{with probability} \quad 1 - \frac{\log(\alpha)}{\alpha-1}, \end{cases} \tag{19}$$

where Y_1 and Y_2 have the following PDFs

$$\begin{aligned} f_{Y_1}(y) &= \lambda e^{-\lambda y}; \quad y > 0 \\ f_{Y_2}(y) &= \frac{\log(\alpha)}{\alpha - 1 - \log(\alpha)} \lambda e^{-\lambda y} \left(\alpha^{1-e^{-\lambda y}} - 1 \right); \quad y > 0, \end{aligned}$$

respectively. It is clear from the representation (19) that as α approaches 1, Y behaves like an exponential distribution, and α increases, it behaves like Y_2. Further, the random variable Y_2 has a a hidden truncation model interpretation similarly as the Azzalini's skew-normal model, see Arnold and Beaver (2000).

The generation from a APE distribution is very simple. By inverting (16), we obtain

$$Y = \frac{-1}{\lambda} \log \left\{ \frac{\log \frac{\alpha}{(\alpha-1)U+1}}{\log \alpha} \right\}. \tag{20}$$

Therefore if U follows uniform (0,1), then $Y \sim \text{APE}(\alpha, \lambda)$. The p-th quantile function of $\text{APE}(\alpha, \lambda)$ is given by

$$y_p = \frac{-1}{\lambda} \log \left\{ \frac{\log \frac{\alpha}{(\alpha-1)p+1}}{\log \alpha} \right\}, \tag{21}$$

hence the median can be obtained as

$$y_{1/2} = \frac{-1}{\lambda} \log \left\{ \frac{\log \frac{2\alpha}{\alpha+1}}{\log \alpha} \right\}. \tag{22}$$

Using the series representations

$$\alpha^{-u} = \sum_{k=0}^{\infty} \frac{(-\log \alpha)^k u^k}{k!}, \tag{23}$$

and

$$\int_0^{\infty} u^n (\log u)^m du = u^{n+1} \sum_{k=0}^{m} (-1)^k \frac{m!(\log u)^{m-k}}{(m-k)!(n+1)^{k+1}}, \tag{24}$$

the moment-generating function (MGF) of $\text{APE}(\alpha, \lambda)$ can be obtained as

$$M_Y(t) = \frac{\lambda\alpha}{1-\alpha} \sum_{k=0}^{\infty} \frac{(-\log \alpha)^{k+1}}{(k\lambda - t + \lambda)k!}, \quad t < \lambda. \tag{25}$$

Hence, the n-th moment of Y becomes

$$E(Y^n) = \frac{\alpha n!}{\lambda^n(1-\alpha)} \sum_{k=1}^{\infty} \frac{(-\log \alpha)^k}{k^n k!}. \tag{26}$$

Suppose that Y is a random variable representing a lifetime of an object. The expected additional lifetime given that a component has survived until time t is called mean residual life (MRL). The MRL function, $\mu(t)$, of a random variable Y is given by

$$\mu(t) = \frac{1}{P(Y > t)} \int_t^{\infty} P(Y > x)dx, \;\; t \geq 0. \tag{27}$$

Using (23) the MRL of $X \sim \text{APE}(\alpha, \lambda)$ for $\alpha \neq 1$, is given by

$$\mu(t) = \frac{-1}{\lambda(1 - \alpha^{-e^{-\lambda t}})} \sum_{k=1}^{\infty} \frac{(-\log\alpha)^k}{k!k} e^{-k\lambda t}, \;\; t \geq 0. \tag{28}$$

It follows that $\mu(t), t \geq 0$ is an increasing function in t for $0 < \alpha < 1$ and it is a decreasing function in t for $\alpha > 1$.

Ordering of distributions, particularly among lifetime distributions play an important role in the statistical literature. Johnson et al. (1995) have a major section on ordering of different lifetime distributions. It is already known that both gamma and GE family have likelihood ratio ordering ($<_{LR}$), which implies that both of them have ordering in hazard rate ($<_{HAZ}$) and also in distribution ($<_{ST}$). It may be recalled that if a family has a likelihood ratio ordering, it has the monotone likelihood ratio property. This implies there exists a uniformly most powerful test for any one sided hypothesis when the other parameters are known. It can be easily verified that if $Y_1 \sim \text{APE}(\alpha_1, \lambda)$ and $Y_2 \sim \text{APE}(\alpha_2, \lambda)$, then for $\alpha_1 < \alpha_2$,

$$Y_1 \leq_{LR} Y_2 \Rightarrow Y_1 \leq_{HAZ} Y_2 \Rightarrow Y_1 \leq_{ST} Y_2.$$

And if $Y_1 \sim \text{APE}(\alpha, \lambda_1, \mu)$ and $Y_2 \sim \text{APE}(\alpha, \lambda_2, \mu)$, then for $\lambda_1 < \lambda_2$,

$$Y_2 \leq_{LR} Y_1 \Rightarrow Y_2 \leq_{HAZ} Y_1 \Rightarrow Y_2 \leq_{ST} Y_1,$$

if $0 < \alpha < e$.

The entropy of a random variable measures the variation of the uncertainty. A large value of entropy indicates the greater uncertainty in the data. Some popular entropy measures are Rényi entropy (Rrnyi, 1961) or Shannon entropy (Shannon, 1951). If X is an absolute continuous random variable with PDF $f(x)$, then the Rényi entropy of X, $RE_X(\rho)$ for $\rho > 0$ and $\rho \neq 1$, is defined as

$$RE_X(\rho) = \frac{1}{1-\rho} \log \left\{ \int_{-\infty}^{\infty} f(x)^\rho dx \right\}, \tag{29}$$

and Shannon entropy is defined as $SE_X = E[-\log f(X)]$. If $Y \sim \text{APE}(\alpha, \lambda, 0)$ then the Rényi entropy of Y for $\rho > 0$ and $\rho \neq 1$, can be obtained as

$$RE_Y(\rho) = \left(\frac{\rho}{1-\rho}\right) \log \left(\frac{\alpha \log \alpha}{\alpha - 1}\right) - \log \lambda + \frac{1}{1-\rho} \log \left\{ \sum_{k=0}^{\infty} \frac{\rho^k (-\log \alpha)^k}{k!(\rho + k)} \right\}. \tag{30}$$

Further, the Shannon entropy of Y becomes

$$SE_Y = \log \left(\frac{\alpha - 1}{\lambda \alpha \log \alpha}\right) + \frac{\alpha}{1-\alpha} \sum_{k=1}^{\infty} \left\{ \frac{(-\log \alpha)^k}{kk!} - \frac{(-\log \alpha)^{k+1}}{(k+1)(k-1)!} \right\}.$$

Finally we provide an order statistics result. Let $Y_1, Y_2, ..., Y_n$ be a random sample from a APE(α, λ), and let $Y_{k:n}$ denote the k-th order statistic. The PDF of $Y_{k:n}$ for $y > 0$ is given by

$$f_{Y_{k:n}}(y) = \frac{n! \log \alpha}{(\alpha - 1)^n (k-1)!(n-k)!} \lambda e^{-\lambda y} (\alpha^{1-e^{-\lambda y}} - 1)^{k-1} (1 - \alpha^{-e^{-\lambda y}})^{n-k} \alpha^{n-k+1-e^{-\lambda y}}.$$

The q-th moment of $Y_{k:n}$ can be expressed as

$$E(Y_{k:n}^q) = \frac{n! q! (-1)^k \alpha^{n-k+1}}{\lambda^q (\alpha - 1)^n} \sum_{l=0}^{k-1} \sum_{m=0}^{n-k} \sum_{r=0}^{\infty} \frac{(-1)^{m-l} \alpha^l (l+m+1)^r (-\log \alpha)^{r+1}}{l! m! r! (k-1-l)!(n-k-m)!(r+1)^{q+1}}. \tag{31}$$

Now we discuss about the stress-strength parameter. Suppose $Y_1 \sim \text{APE}(\alpha_1, \lambda)$ and $Y_2 \sim \text{APE}(\alpha_2, \lambda)$, and Y_1 and Y_2 are independently distributed, then

Case 1: $\alpha_1 \neq 1$ and $\alpha_2 \neq 1$:

$$\begin{aligned} P(Y_1 < Y_2) &= \frac{\log \alpha_2}{(\alpha_1 - 1)(\alpha_2 - 1)} \int_0^{\infty} \lambda e^{-\lambda y} \alpha_2^{1-e^{-\lambda y}} (\alpha_1^{(1-e^{-\lambda y})} - 1) dy \\ &= \frac{\log \alpha_2}{(\alpha_1 - 1)(\alpha_2 - 1)} \int_0^1 \alpha_2^z (\alpha_1^z - 1) dz \quad \text{(by using } 1 - e^{-\lambda y} = z) \\ &= \frac{\log \alpha_2}{(\alpha_1 - 1)(\alpha_2 - 1)} \left[\frac{\alpha_1 \alpha_2 - 1}{\log(\alpha_1 \alpha_2)} - \frac{\alpha_2 - 1}{\log \alpha_2} \right] \end{aligned}$$

Case 2: $\alpha_1 = 1$, $\alpha_2 \neq 1$:

$$\begin{aligned} P(Y_1 < Y_2) &= \frac{\log \alpha_2}{(\alpha_2 - 1)} \int_0^{\infty} \lambda e^{-\lambda y} \alpha_2^{1-e^{-\lambda y}} (1 - e^{-\lambda y}) dy \\ &= \frac{\log \alpha_2}{(\alpha_2 - 1)} \int_0^1 z \alpha_2^z dz \qquad \text{(by using } 1 - e^{-\lambda y} = z) \\ &= \frac{\log \alpha_2}{(\alpha_2 - 1)} \left[\frac{1 + \alpha_2}{\log \alpha_2} - \frac{\alpha_2}{(\log \alpha_2)^2} \right] \end{aligned}$$

Case 3: $\alpha_1 \neq 1$, $\alpha_2 = 1$:

$$\begin{aligned} P(Y_1 < Y_2) &= \frac{1}{(\alpha_1 - 1)} \int_0^{\infty} \lambda e^{-\lambda y} (\alpha_1^{1-e^{-\lambda y}} - 1) dy \\ &= \frac{1}{(\alpha_1 - 1)} \left[\int_0^1 \alpha_1^z dz - 1 \right] \qquad \text{(by using } 1 - e^{-\lambda y} = z) \\ &= \frac{1}{(\alpha_1 - 1)} \left[\frac{\alpha_1 - 1}{\log \alpha_1} - 1 \right]. \end{aligned}$$

Case 4: $\alpha_1 = \alpha_2 = 1$

$$P(Y_1 < Y_2) = \int_0^{\infty} \lambda_2 e^{-\lambda_2 y} (1 - e^{-\lambda_1 y}) dy = 1 - \frac{\lambda_2}{\lambda_1 + \lambda_2} = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

4 Statistical Inference

4.1 Maximum Likelihood Estimators

Let $y_1, y_2, \ldots, y_n$ be a random sample from APE(α, λ), then the log-likelihood function becomes

$$l(\alpha, \lambda) = n \log \alpha + n \log \left(\frac{\log \alpha}{\alpha - 1} \right) + n \log \lambda - \lambda \sum_{i=1}^{n} y_i - (\log \alpha) \sum_{i=1}^{n} e^{-\lambda y_i}. \tag{32}$$

The two normal equations become

$$\frac{\partial l(\alpha, \lambda)}{\partial \alpha} = \frac{n}{\alpha} + \frac{n(\alpha - 1 - \alpha \log \alpha)}{\alpha(\alpha - 1) \log \alpha} - \frac{1}{\alpha} \sum_{i=1}^{n} e^{-\lambda y_i} = 0, \tag{33}$$

$$\frac{\partial l(\alpha, \lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} y_i + (\log \alpha) \sum_{i=1}^{n} y_i e^{-\lambda y_i} = 0. \tag{34}$$

The MLEs of α and λ can be obtained by solving (33) and (34) simultaneously, and they will be denoted by $\widehat{\alpha}$ and $\widehat{\lambda}$. Standard algorithm like Newton-Raphson method may be used to solve these non-linear equations. Alternatively, any two dimensional optimization method may be used to maximize the log-likelihood function $l(\alpha, \lambda)$ directly. Since it is a simple two-dimensional optimization problem, getting some initial guesses is also not difficult. We proceed as follows. For a fixed λ obtain $\widehat{\alpha}(\lambda)$ from (34) as

$$\widehat{\alpha}(\lambda) = \exp\left\{ \frac{\frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{\lambda}}{\frac{1}{n}\sum_{i=1}^{n} y_i e^{-\lambda y_i}} \right\},$$

and then obtain $\widehat{\lambda}$ from (33) by solving

$$1 + \frac{(\widehat{\alpha}(\lambda) - 1 - \widehat{\alpha}(\lambda)\log\widehat{\alpha}(\lambda))}{\widehat{\alpha}(\lambda)(\widehat{\alpha}(\lambda) - 1)\log\widehat{\alpha}(\lambda)} - \frac{1}{n}\sum_{i=1}^{n} e^{-\lambda y_i} = 0. \tag{35}$$

Equation (35) can be solved by simple bisection method. Once $\widehat{\lambda}$ is obtained $\widehat{\alpha}$ can be obtained as $\widehat{\alpha} = \widehat{\alpha}(\widehat{\lambda})$.

Two-parameter APE distribution is a regular family, hence we have the following asymptotic result: As $n \to \infty$, $\sqrt{n}(\widehat{\alpha} - \alpha, \widehat{\lambda} - \lambda)$ converges to a bivariate normal distribution with mean vector $\mathbf{0}$ and the variance co-variance matrix $\boldsymbol{I}_2^{-1}$, where $\boldsymbol{I}_2 = ((I_{ij}))$ is the Fisher information matrix. The elements of $\boldsymbol{I}_2 = ((I_{ij}))$ are given by

$$\begin{aligned}
I_{11} &= \frac{n}{\alpha^2} - \frac{n}{(\alpha-1)^2} + \frac{n(\log\alpha + 1)}{(\alpha\log\alpha)^2} - \frac{n}{\alpha(1-\alpha)}\sum_{k=0}^{\infty}\frac{(-\log\alpha)^{k+1}}{(k+2)k!}, \\
I_{22} &= \frac{n}{\lambda^2} + \frac{2n\alpha}{\lambda^2(\alpha-1)}\sum_{k=0}^{\infty}\frac{(-\log\alpha)^{k+2}}{(k+2)^3 k!}, \\
I_{12} &= I_{21} = \frac{n}{\lambda(\alpha-1)}\sum_{k=0}^{\infty}\frac{(-\log\alpha)^{k+1}}{(k+2)^2 k!}.
\end{aligned} \tag{36}$$

Using the above result, asymptotic confidence intervals of the unknown parameters can be easily obtained.

4.2 Testing of Hypothesis

In this section we consider different testing of hypothesis problems which are of interest. It is assumed that we have a random sample $\{y_1, \ldots, y_n\}$ from APE(α, λ). First let us consider the following testing problem on α, when the scale parameter λ is assumed to be known. Without loss of generality it is assumed that $\lambda = 1$.

$$\text{Test 1:} \quad H_0 : \alpha = \alpha_0 \quad \text{vs.} \quad \alpha > \alpha_0. \tag{37}$$

Note that when λ is known, then $T_1 = \sum_{i=1}^{n} e^{-y_i}$ is a complete sufficient statistic. The uniformly most powerful (UMP) test will be of the following form: Reject H_0 if $T_1 < c$. To determine c we need to know the distribution of T_1 under H_0. The exact distribution of T_1 is difficult to obtain, hence we can proceed in two different ways. Obtain the critical value by generating samples from y_i, alternatively use the large sample approximation of T_1. Similarly, the uniformly most powerful unbiased test of the following testing problem

$$\text{Test 2:} \quad H_0 : \alpha = \alpha_0 \quad \text{vs.} \quad \alpha \neq \alpha_0, \tag{38}$$

can be easily performed.

5 Application: coal-mining dataset

In this section, we analyze one data set to show the performance of the proposed model. The uncensored data set corresponding to intervals in days between 109 successive coal-mining disasters in Great Britain, for the period 1875-1951, published by Maguire et al. (1952). The sorted data are given as follows: 1 4 4 7 11 13 15 15 17 18 19 19 20 20 22 23 28 29 31 32 36 37 47 48 49 50 54 54 55 59 59 61 61 66 72 72 75 78 78 81 93 96 99 108 113 114 120 120 120 123 124 129 131 137 145 151 156 171 176 182 188 189 195 203 208 215 217 217 217 224 228 233 255 271 275 275 275 286 291 312 312 312 315 326 326 329 330 336 338 345 348 354 361 364 369 378 390 457 467 498 517 566 644 745 871 1312 1357 1613 1630.

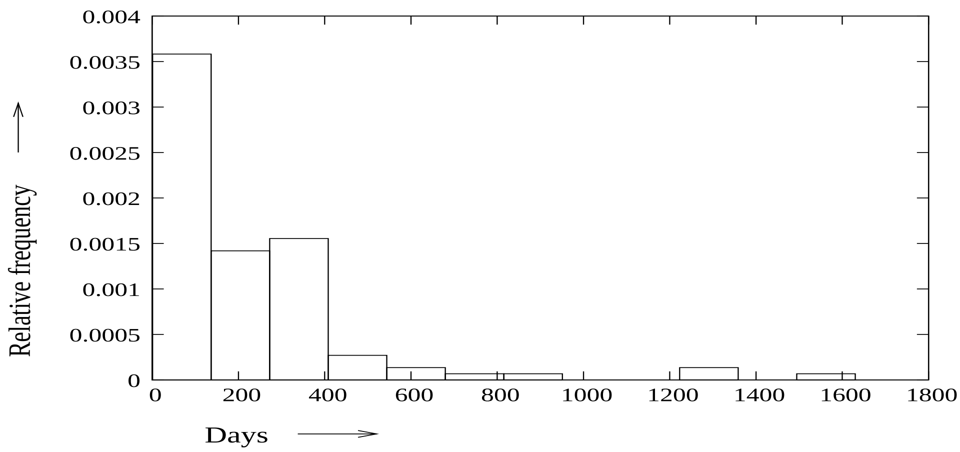

Figure 3: Relative histogram of the coal mine data set.

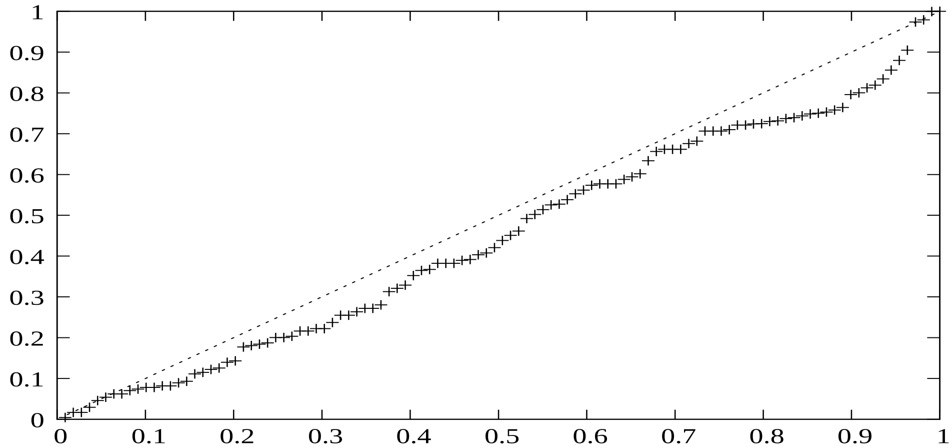

Figure 4: Scaled TTT plot of the coal mine data set.

The minimum, maximum, first quartile, median, third quartile of the above data set are: 1, 1630, 54, 145, 312, respectively. It clearly indicates that the data are skewed. The histogram and the scaled TTT plots are provided in Figure 3 and Figure 4, respectively, also indicate that the data are right skewed and the empirical hazard function is a decreasing function of time.

Now we would like to fit the proposed APE(α, λ) to the above data set. To get an idea about the MLEs λ, we plot the function provided on the left hand side of the equation (33), and it gives an idea about the initial guess value of the MLE of λ. We finally obtain the MLEs of λ and α as 0.0030 and 0.2807 with the associated log-likelihood value as -701.2132. The corresponding 95% confidence intervals based on Fisher information matrix become (0.2280, 0.3334) and (0.0028, 0.0032), respectively.

For comparison purposes, we have fitted four other two-parameter models to the same data set. In all these four distributions, the hazard function can be a decreasing function depending on the shape parameter. In all these cases it is assumed that $\alpha > 0$ and $\lambda > 0$. They respective PDFs are as follows:

- gamma distribution with PDF

$$f_{GA}(x;\alpha,\lambda) = \frac{\lambda^{\alpha}}{\Gamma\alpha} x^{\alpha-1} e^{-\lambda x} \quad x > 0.$$

- Weibull distribution with PDF

$$f_{WE}(x;\alpha,\lambda)) = \alpha\lambda(\lambda x)^{\alpha-1} e^{-(\lambda x)^{\alpha}}, \quad x > 0.$$

- exponentiated exponential distribution (EE) with PDF

$$f_{EE}(x;\alpha,\lambda) = \alpha\lambda(1 - e^{-\lambda x})^{\alpha-1} e^{-\lambda x}, \quad x > 0,$$

  see for example Gupta et al. (1998) or Gupta and Kundu (1999).

- weighted exponential distribution (WE) proposed by Gupta and Kundu (2009) with the PDF

$$f(x) = \frac{\alpha+1}{\alpha} \lambda e^{-\lambda x}(1 - e^{-\alpha\lambda x}), \quad x > 0.$$

We have computed the MLEs of α and λ and the associated log-likelihood values in all these cases. We have also obtained the Kolmogorov-Smirnov (K-S) distance between the empirical cumulative distribution function and the fitted distribution function in each case and the associated p value. The results are reported in Table 2.

Table 2: The maximum likelihood estimates and Kolmogorov-Smirnov statistics and p-values for coal-mining data.

The model	MLEs of the parameters	Log-likelihood	K-S statistic	p-value
gamma	$\hat{\alpha} = 0.8555, \hat{\lambda} = 0.0037$	-702.4007	0.0823	0.4517
Weibull	$\hat{\alpha} = 0.8848, \hat{\lambda} = 0.0046$	-701.7724	0.0784	0.5135
EE	$\hat{\alpha} = 0.8605, \hat{\lambda} = 0.0039$	-702.5524	0.0830	0.4402
WE	$\hat{\alpha} = 35.2748, \hat{\lambda} = 0.0045$	-705.1641	0.0836	0.4313
APE	$\hat{\alpha} = 0.2807, \hat{\lambda} = 0.0030$	-701.2132	0.0742	0.5852

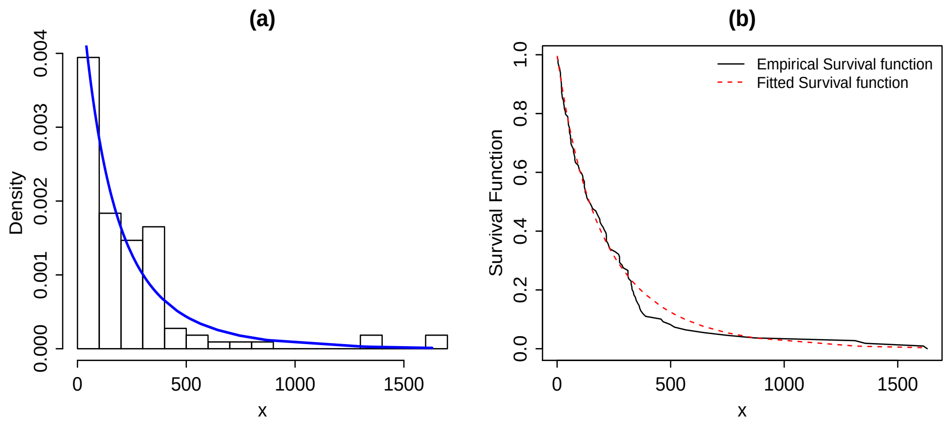

Figure 5: (a) the histogram and the fitted APE distribution. (b) the fitted APE survival function and empirical survival function for coal-mining data.

It is clear from the Table 2 that based on the log-likelihood value and also based on the K-S statistic, the proposed APE model provides a better fit than gamma, Weibull, EE, WE models to this specific data set. The relative histogram and the fitted APE distribution are plotted in Figure 5 (a). In order to assess if the model is appropriate, the plots of the fitted APE survival function and empirical survival function are displayed in Figure 5 (b). Although, it is not guaranteed that the proposed model always provides a better fit than the other models, but at least in certain cases it definitely can provide a better fit. Therefore, the APE model can be used as a possible alternative to the well known gamma, Weibull or EE models.

6 Three-Parameter APE Distribution

In this section we introduce three-parameter APE distribution and analyze one data set based on three-parameter APE distribution.

Definition 2: The random variable Y is said to have a three-parameter APE distribution, and will be denoted by APE(α, λ, μ), with the shape, scale and location parameters as $\alpha > 0$, $\lambda > 0$ and $-\infty < \mu < \infty$, if the PDF of Y, for $y > \mu$ is

$$f(y; \alpha, \lambda, \mu) = \begin{cases} \frac{\log\alpha}{\alpha-1}\lambda e^{-\lambda(y-\mu)}\alpha^{1-e^{-\lambda(y-\mu)}} & \text{if} \quad \alpha \neq 1 \\ \lambda e^{-\lambda(y-\mu)} & \text{if} \quad \alpha = 1. \end{cases} \tag{39}$$

The associated CDF becomes

$$F(y; \alpha, \lambda, \mu) = \begin{cases} \frac{\alpha^{1-e^{-\lambda(x-\mu)}}-1}{\alpha-1} & \text{if} \quad \alpha \neq 1 \\ 1 - e^{-\lambda(x-\mu)} & \text{if} \quad \alpha = 1. \end{cases} \tag{40}$$

Clearly, the shape of the PDF and the hazard function of APE(α, λ, μ) are same as APE(α, λ). Moreover, the generation of a random sample from APE(α, λ, μ) can be obtained similarly as APE(α, λ). Different moments of APE(α, λ, μ) distribution can be obtained from the different moments of APE(α, λ).

Now we discuss the maximum likelihood estimators of α, λ and μ based on a random sample, $\{y_1, \ldots, y_n\}$, of size n, from APE(α, λ, μ). The log-likelihood function can be written as

$$l(\alpha, \lambda, \mu) = n\log\alpha + n\log\left(\frac{\log\alpha}{\alpha-1}\right) + n\log\lambda - \lambda\sum_{i=1}^{n}(y_i - \mu) - (\log\alpha)\sum_{i=1}^{n} e^{-\lambda(y_i-\mu)}. \tag{41}$$

Therefore, the MLEs of α, λ and μ can be obtained by maximizing (41) with respect to the unknown parameters, and we denote them as $\widehat{\alpha}$, $\widehat{\lambda}$ and $\widehat{\mu}$, respectively. It is immediate from (41), that for $\alpha < 1$, $\widehat{\mu} = y_{(1)}$, the lowest order statistic of the sample. Hence $\widehat{\alpha}$ and $\widehat{\lambda}$ can be obtained by maximizing (32) based on the sample $y_1 - y_{(1)}, \ldots, y_n - y_{(1)}$. For $\alpha > 1$, the normal equations become

$$\frac{\partial l(\alpha,\lambda)}{\partial \alpha} = \frac{n}{\alpha} + \frac{n(\alpha-1-\alpha\log\alpha)}{\alpha(\alpha-1)\log\alpha} - \frac{1}{\alpha}\sum_{i=1}^{n} e^{-\lambda(y_i-\mu)} = 0, \quad (42)$$

$$\frac{\partial l(\alpha,\lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n}(y_i-\mu) + (\log\alpha)\sum_{i=1}^{n} y_i e^{-\lambda(y_i-\mu)} = 0, \quad (43)$$

$$\frac{\partial l(\alpha,\lambda)}{\partial \mu} = n\lambda - \lambda(\log\alpha)\sum_{i=1}^{n} y_i e^{-\lambda(y_i-\mu)} = 0. \quad (44)$$

The MLEs cannot be obtained in explicit forms, we need to solve the above non-linear equations to compute the MLEs. For $\alpha > 1$, the APE(α,λ,μ) is a regular family, hence as $n \to \infty$, $\sqrt{n}(\widehat{\alpha}-\alpha, \widehat{\lambda}-\lambda, \widehat{\mu}-\mu)$ converges to a trivariate normal distribution with mean vector $\mathbf{0}$ and the variance co-variance matrix $\boldsymbol{I}_3^{-1}$, where $\boldsymbol{I}_3 = ((I_{ij}))$ is the Fisher information matrix. The elements of $\boldsymbol{I}_3$ can be obtained similarly as $\boldsymbol{I}_2$, the details are avoided.

Now we provide the analysis of a data set to see how APE(α,λ,μ) behaves in a practical situation. The data set has been obtained from Bader and Priest (1982), and it represents the strength for the single carbon fibers and impregnated 1000-carbon fiber tows, measured in GPa. We report the data of single carbon fiber tested at gauge length 1mm. The data are presented below: 2.247, 2.64, 2.908, 3.099, 3.126, 3.245, 3.328, 3.355, 3.383, 3.572, 3.581, 3.681, 3.726 3.727, 3.728, 3.783, 3.785, 3.786, 3.896, 3.912, 3.964, 4.05, 4.063, 4.082, 4.111, 4.118, 4.141, 4.246, 4.251, 4.262, 4.326, 4.402, 4.457, 4.466, 4.519, 4.542, 4.555, 4.614, 4.632, 4.634, 4.636, 4.678, 4.698, 4.738, 4.832, 4.924, 5.043, 5.099, 5.134, 5.359, 5.473, 5.571, 5.684, 5.721, 5.998, 6.06.

The preliminary data analysis indicates that the data are skewed and the scaled TTT plot, see Figure 6 indicates that the empirical hazard function is an increasing function. Therefore, $\alpha > 1$. Finally we obtain the MLEs of the unknown parameters and they are as follows: $\widehat{\alpha} = 673.8379$, $\widehat{\lambda} = 1.1562$ and $\widehat{\mu} = 2.247$. The associated 95% confidence intervals are (659.2312,691.2256), (0.9312,1.3519), (1.916,2.247), respectively. We also the obtain the KS distance between the empirical CDF and the fitted CDF and it is 0.0925 and the associated

p value is 0.7243. The empirical and fitted survival functions are provided in Figure 7. The p-value of the KS statistic indicates that the three-parameter APE model provides a good fit to the data set.

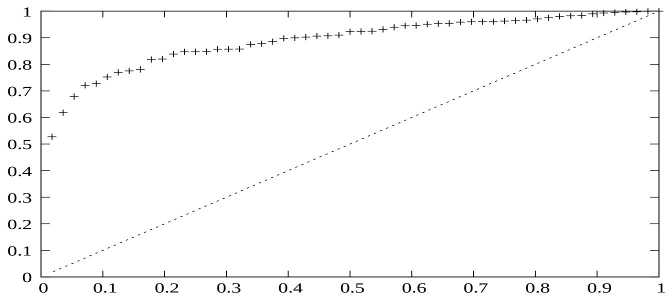

Figure 6: Scaled TTT plot of the single fiber strength data set.

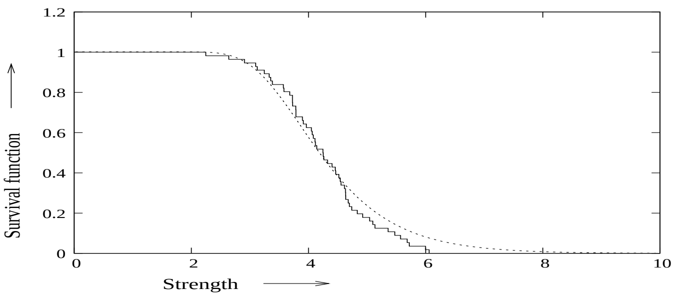

Figure 7: Empirical and fitted survival function for the single fiber strength data set.

7 Conclusion

A new APT method has been introduced to incorporate skewness to a family of distribution functions. We have used that method to the exponential family of distribution functions, and a new two-parameter APE distribution has been introduced. This proposed distribution has several desirable properties, and they are quite similar to the corresponding properties of the well known gamma or Weibull family. One data analysis has been performed, and it is observed that the proposed model provides a better fit than some of the existing models. Further we consider three-parameter APE distribution also, and discuss the maximum likelihood estimation procedure of the unknown parameters. One data analysis has been performed based on three-parameter APE distribution. It is observed that the three-parameter APE distribution provides a good fit to the data set.

It should be mentioned that although we have used the APT method to the exponential distribution function, similar method can be used to the normal distribution also. It will introduce a skewness parameter to a normal distribution similar to the skew-normal distribution of Azzalini (1985). It will be interesting to study the properties of this skew-normal distribution and also to develop the inferential procedures. The work is in progress, it will be reported later.

Acknowledgements:

The authors would like to thank the associate editor and two referees for their constructive comments.

References

M. A. Aljarrah, C. Lee, and F. Famoye. On generating tx family of distributions using quantile functions. *Journal of Statistical Distributions and Applications*, 1(1):1–17, 2014.

A. Alzaatreh, C. Lee, and F. Famoye. A new method for generating families of continuous distributions. *Metron*, 71(1):63–79, 2013.

B. C. Arnold and R. J. Beaver. Hidden truncation models. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 23–35, 2000.

A. Azzalini. A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, pages 171–178, 1985.

M. Bader and A. Priest. Statistical aspects of fibre and bundle strength in hybrid composites. *Progress in science and engineering of composites*, pages 1129–1136, 1982.

R. Bartow and F. Prochan. Statistical theory of reliability and life testing. *Rinehart and Winston Inc. NY (USA)*, 1975.

N. Eugene, C. Lee, and F. Famoye. Beta-normal distribution and its applications. *Communications in Statistics-Theory and Methods*, 31(4):497–512, 2002.

R. C. Gupta, P. L. Gupta, and R. D. Gupta. Modeling failure time data by lehman alternatives. *Communications in Statistics-Theory and Methods*, 27(4):887–904, 1998.

R. D. Gupta and D. Kundu. Theory & methods: Generalized exponential distributions. *Australian & New Zealand Journal of Statistics*, 41(2):173–188, 1999.

R. D. Gupta and D. Kundu. A new class of weighted exponential distributions. *Statistics*, 43(6):621–634, 2009.

N. L. Johnson, S. Kotz, and N. Balakrishnan. Continuous univariate distributions, vol. 2wiley. *New York*, 1995.

M. Jones. On families of distributions with shape parameters. *International Statistical Review*, 2014.

C. Lee, F. Famoye, and A. Y. Alzaatreh. Methods for generating families of univariate continuous distributions in the recent decades. *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(3):219–238, 2013.

B. A. Maguire, E. Pearson, and A. Wynn. The time intervals between industrial accidents. *Biometrika*, pages 168–180, 1952.

A. W. Marshall and I. Olkin. A new method for adding a parameter to a family of distributions with application to the exponential and weibull families. *Biometrika*, 84(3):641–652, 1997.

G. S. Mudholkar and D. K. Srivastava. Exponentiated weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*, 42(2):299–302, 1993.

G. P. Patil. *Weighted distributions*. Wiley Online Library, 2002.

G. P. Patil and C. R. Rao. Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, pages 179–189, 1978.

A. Rrnyi. On measures of entropy and information. In *Fourth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 547–561, 1961.

C. E. Shannon. Prediction and entropy of printed english. *Bell system technical journal*, 30 (1):50–64, 1951.