ISSN: 2641-6336 DOI: 10.33552/ABBA.2019.02.000541

Annals of Biostatistics & Biometric Applications

Iris Publishers

Short Communication

Maximum Likelihood Estimation in an Alpha-Power

Clement Boateng Ampadu*[1] and Abdulzeid Yen Anafo[2]

[1]*Department of Biostatistics, USA*

[2]*African Institute for Mathematical Sciences (AIMS), Ghana*

***Corresponding author:** Clement Boateng Ampadu, Department of Biostatistics, USA.

Received Date: May 26, 2019
Published Date: June 03, 2019

Abstract

In this paper a new kind of alpha-power transformed family of distributions (APTA-F for short) is introduced. We discuss the maximum likelihood method of estimating the unknown parameters in a sub-model of this new family. The simulation study performed indicates the method of maximum likelihood is adequate in estimating the unknown parameters in the new family. The applications indicate the new family can be used to model real life data in various disciplines. Finally, we propose obtaining some properties of this new family in the conclusions section of the paper.

Keywords: Maximum likelihood estimation; Monte carlo simulation; Alpha-power transformation; Guinea Big data

Introduction

Mahdavi and Kundu [1] proposed the celebrated APT-*F* family of distributions with CDF

$$G(x;\alpha,\xi)=\frac{\alpha^{F}(x;\xi)-1}{\alpha-1}$$

where $1\neq\alpha>0, x\in\mathbb{R}$ and ξ is a vector of parameters all of whose entries are positive. By modifying the above CDF, the Zubair-G family of distributions appeared in [2] with the following CDF

$$F(x;\alpha,\xi)=\frac{e^{\alpha G(x;\xi)^2}-1}{e^{\alpha}-1}$$

where $\alpha>0, x\in\mathbb{R}$ and ξ is a vector of parameters all of whose entries are positive. Subsequently in [3] we introduced the Ampadu-G family of distributions by modifying the above CDF, to give the following CDF

$$F(x;\lambda,\xi)=\frac{1-e^{-\lambda G(x;\xi)^2}}{1-e^{-\lambda}}$$

where $\lambda>0, x\in\mathbb{R}$, and ξ is a vector of parameters all of whose entries are positive. Consequently, we introduced new variants of the APT-F family of distributions, which are recorded in the table below (Table 1).

Table 1: Variants of the Alpha Power Transformed Family of Distributions.

Year	Name of Distribution	Author
2019	Ampadu APT -G	Ampadu [4]
2019	$\left(\frac{1}{e}\right)^{\alpha} PT-G$	Ampadu [4]
2019	$\left(\frac{1}{e}\right)^{\alpha} PT-F$ Within Quantile Distribution (New)	Ampadu and to appear in [5]
2019	APTA-F (Current Manuscript)	Ampadu

The rest of this paper is organized as follows. In Section 2 we introduce and illustrate the new family. In Section 3 we derive the quantile function of the APTA-F family of distributions. In Section 4, we discuss approximate random sampling from the APTAF family of distributions, and maximum likelihood stimation in Section 5. Section 6 presents simulation results associated with the APTA-F family of distributions, assuming F follows the Weibull distribution as defined in Section 2. Section 7 discusses usefulness of the APTA-F family of distributions in fitting real-life data. Section 8 is devoted to the conclusions.

The New Family

Definition 2.1. Let $x\in\mathbb{R}, \alpha\in\mathbb{R}$, and $F(x;\xi)$ be a baseline CDF all of whose parameters are in the vector ξ. We say a random variable X follows the alpha-power transformed distribution of the Ampadu type (APTA-F for short) if the CDF is given by

$$\frac{F(x;\xi)e^{-\alpha F(x;\xi)}}{e^{-\alpha}}$$

If the baseline distribution is given by the Weibull distribution with CDF

$$1-e^{-\left(\frac{x}{b}\right)^a}$$

for $x,a,b>0$, then we have the following from the above definition

Proposition 2.2. The CDF of the APTA-Weibull distribution is given by

$$\left(1-e^{-\left(\frac{x}{b}\right)^a}\right)e^{\alpha e^{-\left(\frac{x}{b}\right)^a}}$$

Where $x,a,b>0$ and $\alpha\in\mathbb{R}$

Notation 2.3. If a random variable V follows the APTA-Weibull distribution, we write

$$V \sim APTAW(a,b,\alpha) \quad \text{(Figure 1)}$$

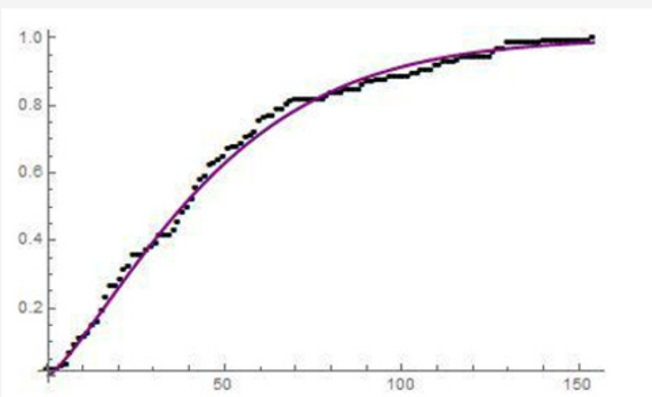

Figure 1: The CDF of APTAW (1.28352800, 48.41833872, -0.07850413) fitted to the empirical distribution of the data on patients with breast cancer, Section 5.2 [6].

By differentiating the CDF of the APTA-F distribution we have the following

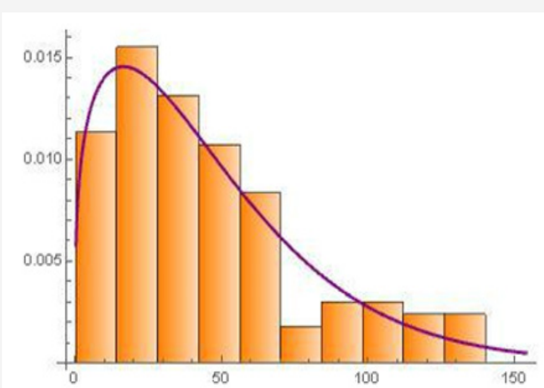

Figure 2: The PDF of APTAW (1.28352800, 48.41833872, -0.07850413) fitted to the histogram of the data on patients with breast cancer, Section 5.2 [6].

Proposition 2.4. The PDF of the alpha-power transformed distribution of the Ampadu type is given by

$$e^{\alpha-\alpha F(x;\xi)}\left(1-\alpha F(x;\xi)\right)f(x;\xi)$$

where $x\in\mathbb{R},\alpha\in\mathbb{R}$ and $F(x;\xi)$ and $f(x;\xi)$ are the CDF and PDF, respectively, of some basline distribution all of whose parameters are in the vector ξ (Figure 2)

The Quantile

Theorem 3.1. *Let* $Q_F(\cdot)=F^{-1}(\cdot)$ denote the quantile of some baseline distribution with CDF F and PDF f, $\alpha\in\mathbb{R}$ and $0<u<1$. The quantile function of the alpha-power transformed distribution of the ampadu type is given by

$$Q_F\left[-\frac{W\left(-\alpha e^{-\alpha}u\right)}{\alpha}\right]$$

Where $W(z):=\mathrm{Pr}\,oductLog(z)$ gives the principal solution for w in $z=we^w$

Proof. Let $0<u<1$. Since $Q_F(\cdot)=F^{-1}(\cdot)$ the quantile function can be obtained by solving the following equation for y

$$u=\frac{F(y)e^{-\alpha F(y)}}{e^{-\alpha}}$$

Approximate Random Number Generation

It is well known that the principal solution for w in $z=we^w$, denoted $W(z):=\mathrm{Pr}\,oductLog(z)$ admit the following Taylor series expansion [7]

$$\sum_{n=1}^{\infty}\frac{(-n)^{n-1}}{n!}z^n=z-z^2+\frac{3}{2}z^3+\frac{8}{3}z^4+\frac{125}{24}z^5-\ldots$$

Using the first term of this series to approximate $W(z):=\mathrm{Pr}\,oductLog(z)$, gives us a way to approximate the random sample from the APTA-F family of distributions.

In particular if $u\sim U(0,1)$, that is, u is a uniform random variable, then an (approximate) random sample from the APTA-*F* family of distributions can be obtained via

$$X=Q_F\left(e^{-\alpha}u\right)$$

where $Q_F(\cdot)=F^{-1}(\cdot)$ denote the quantile of some baseline distribution with CDF *F* and PDF *f*, $\alpha\in\mathbb{R}$

Parameter Estimation

In this section, we obtain the maximum likelihood estimators (MLEs) for the parameters of the APTA-*F* family of distributions. For this, let $X_1,X_2,\ldots X_n$ be a random sample of size *n* from the APTA-*F* family of distributions. The likelihood function from Proposition 2.4 is given by

$$L=\prod_{i=1}^{n}\left\{e^{-\alpha-\alpha F(x_i;\xi)}\left(1-\alpha F(x_i;\xi)\right)f(x_i;\xi)\right\}$$

Citation: Clement Boateng Ampadu, Abdulzeid Yen Anafo. Maximum Likelihood Estimation in an Alpha-Power. Annal Biostat & Biomed Appli. 2(3): 2019. ABBA.MS.ID.000541. DOI: 10.33552/ABBA.2019.02.000541.

From the above the log-likelihood function is given by

$$\ln L=\sum_{i=1}^{n}\alpha\left(1-F\left(x_i;\xi\right)\right)+\sum_{i=1}^{n}\ln\left(1-\alpha F\left(x_i;\xi\right)\right)+\sum_{i=1}^{n}\ln f\left(x_i;\xi\right)$$

The MLE's of ξ and can be obtained by maximizing the equation immediately above.

The derivatives of the equation immediately above with respect to the unknown parameters, are given as follows

$$\ln L=\sum_{i=1}^{n}\alpha\left(1-F\left(x_i;\xi\right)\right)+\sum_{i=1}^{n}\ln\left(1-\alpha F\left(x_i;\xi\right)\right)+\sum_{i=1}^{n}\ln f\left(x_i;\xi\right)$$

$$\frac{\partial \ln L}{\partial \alpha}=\sum_{i=1}^{n}\ln\left(1-F\left(x_i;\xi\right)\right)+\sum_{i=1}^{n}\frac{F\left(x_i;\xi\right)}{\alpha F\left(x_i;\xi\right)-1}$$

$$\frac{\partial \ln L}{\partial \xi}=-\sum_{i=1}^{n}\alpha\frac{\partial F\left(x_i;\xi\right)}{\partial \xi}+\sum_{i=1}^{n}\frac{\alpha\frac{\partial F\left(x_i;\xi\right)}{\partial \xi}}{\alpha F\left(x_i;\xi\right)-1}+\sum_{i=1}^{n}\frac{\frac{\partial f\left(x_i;\xi\right)}{\partial \xi}}{f\left(x_i;\xi\right)}$$

Now solving the system below for _ and _ gives the maximum likelihood estimators, $\hat{\alpha}$ and $\hat{\xi}$ of the unknown parameters:

$$\frac{\partial \ln L}{\partial \alpha}=0 \text{ and } \frac{\partial \ln L}{\partial \xi}=0$$

Simulation Studies

In this section, a Monte Carlo simulation study is carried out to assess the performance of the estimation method, when the baseline distribution is defined as in Section 2.

Simulation Study One

Approximate samples of sizes 200, 350, 500, and 700, are drawn from the APTA-Weibull distribution. The approximate samples have been drawn for (a,b,α) = (1.3, 48.4, 0) using

$$X=b\left(-\log\left(1-e^{-\alpha}u\right)\right)^{1/a}$$

Table 2: Result of Simulation Study.

Parameter a		
Sample Size	**Average Estimate**	**MSE**
200	1.276885	0.037847
350	1.269485	0.019242
500	1.261209	0.021077
700	1.268349	0.016317
Parameter b		
Sample Size	**Average Estimate**	**MSE**
200	48.55782	198.7113
350	48.3844	150.1065
500	48.73268	140.3203
700	48.10531	123.0901
Parameter α		
Sample Size	**Average Estimate**	**MSE**
200	-0.10479	0.377691
350	-0.16417	0.436073
500	-0.08719	0.311987
700	-0.06227	0.213638

and the maximum likelihood estimators for the parameters *a*, *b*, and α are obtained. The procedure has been repeated 200 times and the mean and mean square error for the estimates are computed, and the results are summarized in Table 2 below.

From Table 2 above, we find that the simulated estimates are close to the true values of the parameters and hence the estimation method is adequate. We have also observed that the estimated mean square errors (MSEs) consistently decrease with increasing sample size as seen in the Figures below.

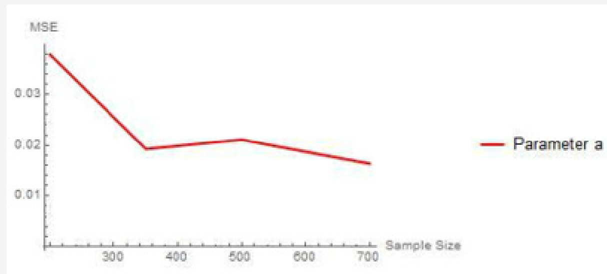

Figure 3: Decreasing MSE of a for increasing sample size.

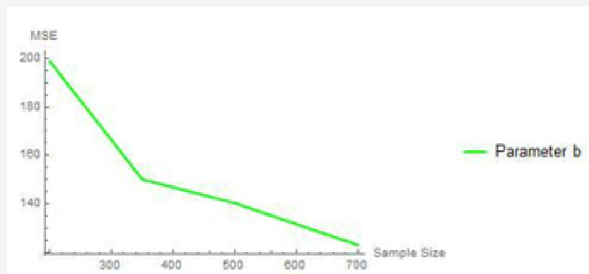

Figure 4: Decreasing MSE of b for increasing sample size.

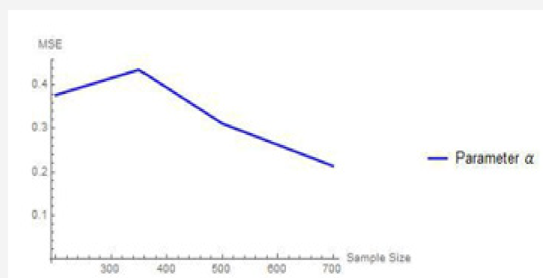

Figure 5: Decreasing MSE of α for increasing sample size.

Simulation study two

Approximate samples of sizes 400, 550, 700, and 900, are drawn from the APTA-Weibull distribution. The approximate samples have been drawn for (a,b,α) = (1.3, 1.3, 0) using

$$X=b\left(-\log\left(1-e^{-\alpha}u\right)\right)^{1/a}$$

and the maximum likelihood estimators for the parameters *a*, *b*, and _ are obtained. The procedure has been repeated 400 times and the bias and root mean square error for the estimates are computed, and the results are summarized in Table 3 below.

From Table 3 above, we find that the estimated root mean square errors (RMSEs) consistently decrease with increasing sample size as seen in the Figures below. We also find that the bias is consistently around zero, hence estimating the parameters in the distribution via the method of maximum likelihood is adequate.

Table 3: Result of Simulation Study.

Parameter a		
Sample Size	**Bias**	**RMSE**
400	-0.01948298	0.1563376
550	-0.02780227	0.1355823
700	-0.01273009	0.1193164
900	-0.01177412	0.1100411
Parameter b		
Sample Size	**Bias**	**RMSE**
400	0.02225872	0.3380576
550	-0.004792135	0.3157712
700	0.007635633	0.2872791
900	0.009964405	0.2334487
Parameter α		
Sample Size	**Bias**	**RMSE**
400	-0.1131752	0.5872724
550	-0.07089181	0.5248752
700	-0.09093898	0.505457
900	-0.03949046	0.430892

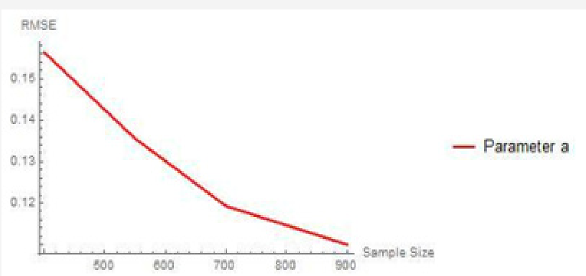

Figure 6: Decreasing RMSE of a for increasing sample size.

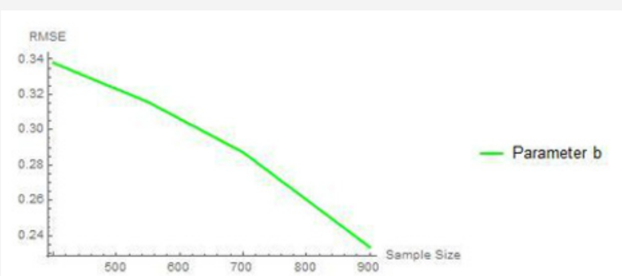

Figure 7: Decreasing RMSE of b for increasing sample size.

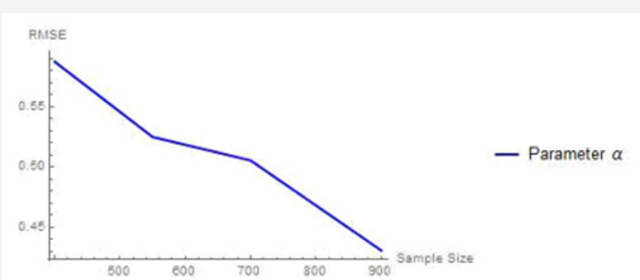

Figure 8: Decreasing RMSE of α for increasing sample size.

Application

In this section, we illustrate the usefulness of the APTA-F family of distributions in modelling real-life data. We compare the fit of the APTA-Weibull distribution, with the proposed Weibull distribution appearing in [5] to the dataset on the survival times (in days) of 72 guinea pigs infected with virulent tubercle bacilli. The dataset can be found in [8]. The estimates of the unknown parameters in both distributions are obtained by the maximum likelihood method using the R language software. The measures of goodness of fit considered included Akaike information criterion (AIC), consistent Akaike information criterion (CAIC), Bayesian information criterion (BIC) and Hannan-Quinn information criterion (HQIC) statistics and they are defined as follows:

$$AIC = 2l + 2k$$

$$BIC = 2l + k\log(n)$$

$$CAIC = AIC + \frac{2k(k+1)}{n-k-1}$$

$$HQIC = 2\log\left[\log(n)(k-2l)\right]$$

where k is the number of parameters in the statistical model, n is the sample size, and $l(\cdot)$ is the maximized value of the log-likelihood function under the considered model. The proposedWeibull distribution appearing in [5], which we denote, Proposed Weibull (a,b,α) has CDF given by

$$\frac{1-e^{-\alpha\left(1-e^{-\left(\frac{x}{b}\right)^{a}}\right)}}{1-e^{-\alpha}}$$

Where $\alpha \neq \frac{1}{e}, \alpha > \frac{1}{e}, a, b > 0$, and $x > 0$

Since the Proposed Weibull (a,b,α) distribution has the smallest AIC, BIC, CAIC, and HQIC values compared to the other distribution in Table 4, it can be considered a better fit to the survival times (in days) of 72 guinea pigs infected with virulent tubercle bacilli The asymptotic variance-covariance matrix of the MLEs under the $APTAW(a,b,\alpha)$, and the Proposed Weibull (a,b,α) distributions, respectively, are given by

Table 4: Estimated Parameters for the dataset on the survival times (in days) of 72 guinea pigs infected with virulent tubercle bacilli.

Model	Parameter Estimate	Standard Error
APTAW (a, b, α)	$(\hat{a},\hat{b},\hat{\alpha})$=(2.0665258, 265.3997306, 0.7428645)	(0.1801981, 37.1849579, 0.2413815)
Proposed Weibull (a, b, α)	$(\hat{a},\hat{b},\hat{\alpha})$ =(2.208316, 310.356991, 3.239273)	(0.1968281, 70.0390762,1.7622610)

$$\begin{bmatrix} 3.247136e-02 & -0.6552024 & -3.845612e-05 \\ -6.552024e-01 & 1382.7210969 & 7.889494e+00 \\ -3.845612e-05 & 7.8894941 & 5.826503e-02 \end{bmatrix}$$

$$\begin{bmatrix} 0.03874132 & -0.7365994 & 0.03693494 \\ -0.73659941 & 4905.4721882 & 116.48517435 \\ 0.03693494 & 116.4851744 & 3.10556397 \end{bmatrix}$$

Hence the approximate 95% confidence intervals for the parameters under the $APTAW(a,b,\alpha)$ and the Proposed Weibull distributions, respectively, are given in Tables 5-7.

Table 5: Criteria for Comparison.

Model	AIC	BIC	CAIC	HQIC
AP T AW (a, b, α)	857.8961	864.7261	858.2491	16.40831
Proposed Weibull (a, b, α)	856.7445	863.5745	857.0974	16.40561

Table 6: APTAW (a, b, α) distribution.

Parameter	CI
a	(2.055226, 2.077825)
b	(263.0680, 267.7315)
α	(0.7277283,0.7580008)

Table 7: Proposed Weibull (a, b, α) distribution.

Parameter	CI
a	(2.195974 2.220659)
b	(305.9651, 314.7489)
α	(3.128767 3.349778)

Concluding Remarks

In this paper a new kind of alpha power transformed family of distributions is introduced, and members of this family are shown to be effective in fitting real life data. As a further development we propose obtaining some mathematical and statistical properties of the APTA-*F* family of distributions.

Acknowledgement

None.

Conflict of Interest

No conflict of interest.

References

1. Mahdavi A, Kundu D (2017) A new method for generating distributions with an application to exponential distribution. Communications in Statistics-Theory and Methods 46(13): 6543-6557.
2. Ahmad Z (2018) The Zubair-G Family of Distributions: Properties and Applications. Annals of Data Science 5: 1-14.
3. Clement Boateng Ampadu (2019) The Ampadu-G Family of Distributions with Application to the T-X(W) Class of Distributions. Annal Biostat & Biomed Appli 1(4).
4. Clement Boateng Ampadu (2019) The Ampadu APT qT –X Family of Distributions Induced by V with an Illustration to Data in the Health Sciences. Annal Biostat & Biomed Appli 2(1).
5. Abdulzeid Yen Anafo (2019) The New Alpha Power Transform: Properties and Ap-plications, Master of Science in Mathematical Sciences Essay, African Institute for Mathematical Sciences (Ghana). Unpublished Manuscript, In Preparation.
6. Girish Babu Moolath, Jayakumar K (2017) T-Transmuted X Family of Distributions. Statistica 3.
7. Wikipedia contributors (2019) “Lambert W function.” Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia.
8. T Bjerkedal (1960) Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli. American Journal of Hygiene 72(1): 130-148.