

The Tan-G Family of Distributions with Illustration to Data in the Health Sciences

Ampadu CB*

31 Carrolton Road, USA

***Corresponding author:** Clement Boateng Ampadu, 31 Carrolton Road, Boston, MA 02132-6303, USA, Email: drampadu@hotmail.com

Short Communication

Volume 3 Issue 3

Received Date: August 16, 2019

Published Date: August 27, 2019

Abstract

The Tan-G family of distributions is introduced, and we show a member of this new class of distributions is significant in fitting real-life data from the health sciences.

Keywords: Trigonometric functions; Statistical distributions; Breast cancer

Introduction

The modification of trigonometric functions to give new statistical distributions is becoming a popular technique in the literature, and for examples see [1-6]. In Kumar D, et al. [1], they transformed the sine function into a new statistical distribution called SS transformation with the following CDF

$$G(x) = \sin\left(\frac{\pi}{2} F(x)\right)$$

The new distribution is shown to be a better fit to data in the health sciences as compared to some well-known distributions available in the statistical literature through Akaike information criteria (AIC), Bayesian information criteria (BIC), -log-likelihood and the associated Kolmogorov-Smirnov (KS) test values. The sine function, as is, violates the properties of a distribution function, hence the argument of the sine function is modified accordingly taking into account any baseline CDF. The crucial properties can be easily checked, in particular it is enough to check that $G(-\infty) = 0$ and $G(\infty) = 1$. Since the baseline CDF is a distribution function, we know $F(-\infty) = 0$ and $F(\infty) = 1$, and since $\sin(0) = 0$ and $\sin(\frac{\pi}{2}) = 1$ we can see $G(x)$ is a proper distribution function. By this observation, the present paper proposes a new statistical distribution by modifying the tangent function.

The rest of this paper is organized as follows. In Section 2, we introduce the new family, and illustrate a member of this family in Section 3. In Section 4, we compare the new distribution with a modification of Ampadu-G [7] (the modification appearing in the parameter space). The last section is devoted to the conclusions.

The New Family

The CDF of Tan-G is given by

$$F(x; \xi) = \tan\left(\frac{\pi}{4} G(x; \xi)\right)$$

where $x \in \mathbb{R}$, and the baseline distribution has CDF $G(x; \xi)$ and PDF $g(x; \xi)$. The parameter space of ξ depends on the chosen baseline distribution. By differentiating the CDF, we get the PDF of Tan-G as follows

$$f(x; \xi) = \frac{1}{4} g(x; \xi) \pi \sec^2\left(\frac{\pi G(x; \xi)}{4}\right)$$

Practical Illustration

We assume the baseline distribution is Weibull with CDF

$$G(x; f, g) = 1 - e^{-\left(\frac{x}{g}\right)^f}$$

where $x, f, g > 0$, then the Tan-G family defined in the previous section implies the following

Definition 5.1. The CDF of Tan-Weibull is given by

$$F(x; f, g) = \tan\left(\frac{1}{4}\pi\left(1 - e^{-\left(\frac{x}{g}\right)^f}\right)\right)$$

where $x, f, g > 0$

Obviously, the PDF of Tan-Weibull can be obtained by differentiating the CDF

Remark 5.2. We write $J \sim TW(f, g)$, if J is a Tan-Weibull random variable (Figure 1).

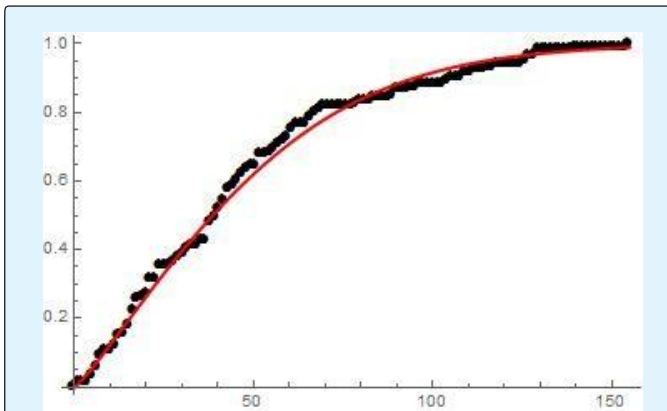


Figure 1: The CDF of $TW(1.23748, 42.4323)$ fitted to the empirical distribution of the data on patients with breast cancer, Section 5.2 [8].

Comparison with Modified Ampadu-G

At first we recall the following

Definition 6.1. Let $\lambda > 0, \xi > 0$ be a parameter vector all of whose entries are positive [7], and $x \in \mathbb{R}$. A random variable X will be said to follow the Ampadu-G family of distributions if the CDF is given by

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

and the PDF is given by

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

where the baseline distribution has CDF $G(x; \xi)$ and PDF $g(x; \xi)$

Remark 6.2. When $\lambda \in (-\infty, 0) \cup (0, \infty)$ and the parameter space of ξ depends on the baseline distribution with CDF G and PDF g , we call the distribution in the previous definition the Modified Ampadu-G Family of distributions

In the rest of this section we compare the Modified Ampadu-Weibull and the Tan Weibull family of distributions in fitting the breast cancer patients data, Section 5.2 [8]. However, we have the following

Proposition 6.3. The CDF of Modified Ampadu-Weibull is given by

$$F(x; f, g, \lambda) = \frac{1 - e^{-\lambda\left(1 - e^{-\left(\frac{x}{g}\right)^f}\right)^2}}{1 - e^{-\lambda}}$$

where $x, f, g > 0$ and $\lambda \in (-\infty, 0) \cup (0, \infty)$

Remark 6.4. We write $Q \sim MAW(f, g, \lambda)$, if Q is a Modified Ampadu-Weibull random variable (Figure 2).

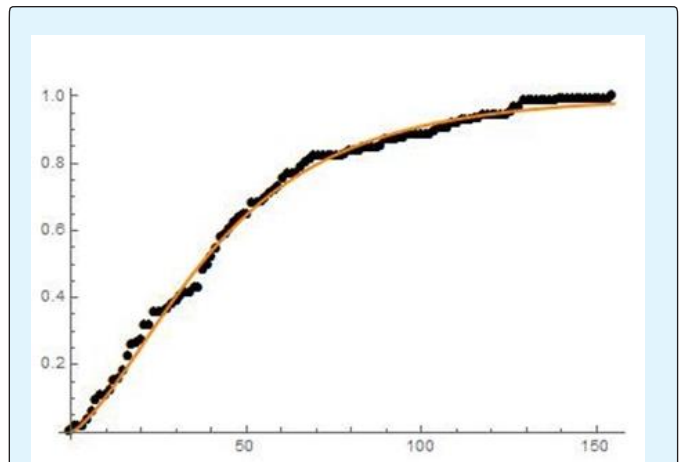


Figure 2: The CDF of $MAW(0.700744, 13.601, -2.45471)$ fitted to the empirical distribution of the data on patients with breast cancer, Section 5.2 [8].

Model	Parameter Estimate	Standard Error
TW(f, g)	(1.23748, 42.4323)	(0.0899789, 3.35782)
MAW(f, g, λ)	(0.700744, 13.601, -2.45471)	(0.167998, 9.26275, 2.03078)

Table 1: Estimated Parameters for the breast cancer patients data.

Model	-2(Log-likelihood)	AIC	AICC	BIC
TW(f, g)	1158.38	1162.38	1162.48	1167.97
MAW(f, g, λ)	1160.2	1166.2	1166.41	1174.59

Table 2: Criteria for Comparison.

In order to compare the two distribution models, we used the following criteria: -2(Log-likelihood) and AIC (Akaike information criterion), AICC (corrected Akaike information criterion), and BIC (Bayesian information criterion) for the data set. The better distribution corresponds to the smaller -2(Log-likelihood) AIC, AICC, and BIC values:

$$AIC = 2k - 2l$$

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

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

where k is the number of parameters in the statistical model, n is the sample size, and l is the maximized value of the log-likelihood function under the considered model.

From Table 2 above, it is clear that the TW distribution has the smallest values across all criteria considered, hence we see the newly discovered Tan-Weibull distribution is a better fit than the Modified Ampadu-Weibull distribution to the breast cancer patients data.

Concluding Remarks

In the present paper we introduced a so-called Tan-G family of distributions, and have shown fit to real-life data. The future interesting problem is to investigate some properties and application of this new class of statistical distributions.

References

1. Kumar D, Singh U, Singh SK (2015) A New Distribution Using Sine Function- Its Application To Bladder Cancer Patients Data. *J Stat Appl Pro* 4(3): 417-427.
2. Jamal F, Chesneau C (2019) A new family of polyno-expo-trigonometric distributions with applications.
3. Mahmood Z, Chesneau C (2019) A New Sine-G Family of Distributions: Properties and Applications.
4. Nadarajah S, Kotz S (2006) Beta trigonometric distributions. *Portuguese Economic Journal* 5(3): 207-224.
5. Al-Faris RQ, Khan S (2008) Sine square distribution: a new statistical model based on the sine function. *J Appl Probab Statist* 3(1): 163-173.
6. Kharazmi O, Saadatinik A, Jahangard S (2019) Odd Hyperbolic Cosine Exponential-Exponential (OHC-EE) Distribution. *Ann Data Sci* pp: 1-21.
7. Ampadu CB (2019) The Ampadu-G Family of Distributions with Application to the T-X(W) Class of Distributions. *Annal Biostat & Biomed Appl* 1(4): 1-5.
8. Moolath GB, Jayakumar K (2017) T-Transmuted X Family of Distributions. *STATISTICA* 77(3): 1-26.

