computer sciences and mathematics forum

MDPI

Proceeding Paper

On Univariate and Bivariate Log-Toppleone Distribution Using Censored and Uncensored Datasets †

Abubakar Usman [1,*], Aliyu Ismail Ishaq [1], Ahmad Abubakar Suleiman [2,3], Mahmod Othman [2,*], Hanita Daud [3] and Yakubu Aliyu [1]

1 Department of Statistics, Ahmadu Bello University, Zaria 810107, Nigeria; binishaq05@gmail.com (A.I.I.); aliyuyakubu40@gmail.com (Y.A.)
2 Department of Statistics, Kano University of Science and Technology, Wudil 713281, Nigeria; ahmadabubakar31@kustwudil.edu.ng
3 Fundamental and Applied Sciences Department, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Perak Darul Ridzuan, Malaysia; hanita_daud@utp.edu.my
* Correspondence: abubakarusman28@gmail.com (A.U.); mahmod.othman@utp.edu.my (M.O.)
† Presented at the 1st International Online Conference on Mathematics and Applications; Available online: https://iocma2023.sciforum.net/.

Abstract: The univariate Toppleone distribution introduced by [12] with close forms of cumulative distribution function i.e [0, 1] is extended to unbounded limit called Log-Toppleone distribution, the shapes of the hazard function can be increase, decrease or constant, therefore, can serve as an alternative distribution to the gamma, Weibull and exponential distributions. Bivariate of this proposed distribution is introduced by joining probability density function using three distinct copulas. The MLE, IFM and Bayesian method of estimation were employed to estimate the parameters, the Plackett copula regarded as the best based on MLE and IFM method of estimation while Clayton copula regarded as the best using Bayesian method.

Keywords: Log-Toppleone distribution; Farlie- Gumbel- Morgenstern; Clayton copula; Plackett copula

Citation: Usman, A.; Ishaq, A.I.; Suleiman, A.A.; Othman, M.; Daud, H.; Aliyu, Y. On univariate and bivariate Log-Toppleone distribution using censored and uncensored datasets. *Comput. Sci. Math. Forum* **2023**, *2*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s):

Published: 1 December 2022

1. Introduction

Development and the use of statistical distributions is not a new matter in statistics. Generating statistical distribution began with the used of system of differential equation approach as in [1], the method for generating system of frequency curves and the quantile method by [2]. Since then, the trend change to adding parameter(s) to an existing distributions as in [3] or combining existing standard distributions as in [4]. Other methods are beta generated method and transformed-transformer method respectively proposed by [5,6].

Many real life phenomena Such as Engineering, Science, Economics and so on, present datasets inform of bivariate in which one component may influence the lifetime of the other component ie in Science one may study the age and resting heart rate for an individuals. To model these datasets, several bivariate distributions were introduced by [7–11] and many more.

2. Methods

This section, provides the structural form of proposed unbounded univariate and bivariate Log-Toppleone distribution using three different copulas functions.

The probability distribution and the density function of Toppleone distribution introduced by [12] are respectively given by:

$$F(x/\alpha) = x^{\alpha}(2-x)^{\alpha} \qquad 0 < x < 1; \quad \alpha > 0 \tag{1}$$

$$f(x/\alpha) = 2\alpha x^{\alpha-1}(1-x)(2-x)^{\alpha-1} \qquad 0 < x < 1; \quad \alpha > 0 \tag{2}$$

where $\alpha > 0$ is the shape parameter.

2.1. Log- Toppleone Distribution

The new propose log- Toppleone distribution is introduce by transforming $X = -\log(1-t)$ in Equation (1) if T serve as a random variable denote the time to the occurrence of an event of interest:

$$F(t/\alpha) = (1-e^{-t})^{\alpha}(2-(1-e^{-t}))^{\alpha} = (1-e^{-2t})^{\alpha}, \qquad t > 0; \quad \alpha > 0 \tag{3}$$

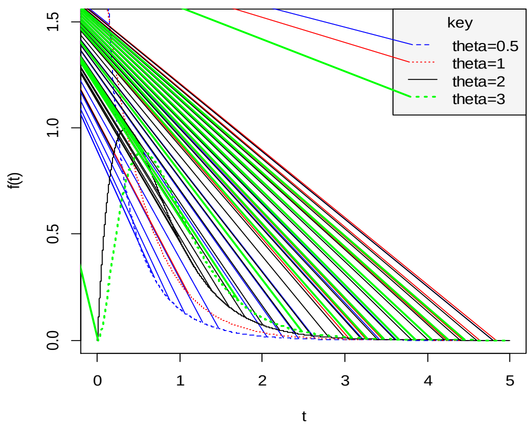
the parameter α will maintain its status as a shape parameter. The corresponding pdf is obtain by differentiating Equation (3) as:

$$f(t/\alpha) = 2\alpha e^{-2t}(1-e^{-2t})^{\alpha-1}, \qquad t > 0; \quad \alpha > 0 \tag{4}$$

The survival and hazard function of the log-Toppleone distribution are respectively given by:

$$S(t/\alpha) = 1-(1-e^{-2t})^{\alpha} \tag{5}$$

The shapes of the hazard function can be increase, decrease or constant, therefore, can serve as an alternative distribution to the gamma, Weibull and exponential distributions.



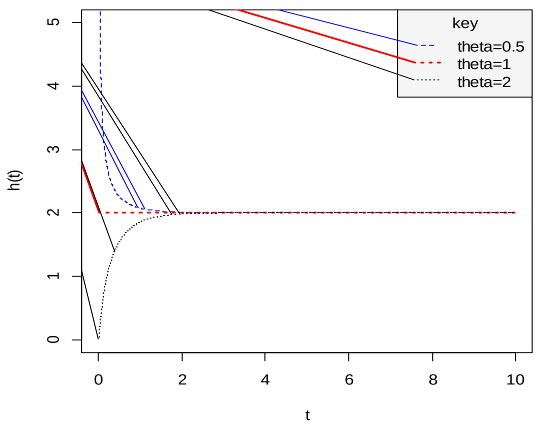


Figure 1. Plots of the pdf (1st) and hazard function (2nd) of Log- Toppleone distribution for some values of parameters.

2.2. Copula

Sklar [13] was first introduced the Copula function to connect the multivariate distribution function with their individual marginal

2.2.1. The Model Based on Farlie- Gumbel- Morgenstern Copula

The joint survival function based on FGM copula for T_1 and T_2 is given by

$$S(t_1, t_2) = S(t_1)S(t_2)\{1+\phi F(t_1)F(t_2)\} \tag{6}$$

where $\phi \in [-1,1]$

2.2.2. The Model Based on Clayton Copula

The joint survival function based on Clayton copula for T_1 and T_2 is given by

$$S(t_1,t_2) = \left(S(t_1)^{-\phi} + S(t_2)^{-\phi} - 1\right)^{-\frac{1}{\phi}} \tag{7}$$

where ϕ is the dependence parameter, takes values in the interval $(0,\infty)$

2.2.3. The Model Based on Plackett Copula

The joint survival function based on Plackett copula for T_1 and T_2 is given by

$$S(t_1,t_2) = \frac{1+(\theta-1)\left(S(t_1)+S(t_2)\right) - \sqrt{\left[1+(\theta-1)S(t_1)+S(t_2)\right]^2 - 4S(t_1)S(t_2)\theta(\theta-1)}}{2(\theta-1)} \tag{8}$$

where $\theta \in (0,\infty)$.

2.3. Inference Methods

This section provide the parameter estimates of bivariate log-Toppleone distribution using the MLE, IFM and Bayesian method of estimation.

Bayesian Method of Estimation

This section look at, when both t_{1i} and t_{2i} are censored and uncensored observations.

(a) **When both t_{1i} and t_{2i} are censored observations**

Let assume, such that $\omega_{ji} = 1$ if t_{ji} are observations, for $j = 1,2$ and $i = 1,2,.....n$, then the likelihood function is given by:

$$\prod_{i=1}^{n}\left[\frac{\partial^2 S(t_{1i},t_{2i})}{\partial t_{1i}\partial t_{2i}}\right]^{\omega_{1i}\omega_{2i}}\left[\frac{-\partial S(t_{1i},t_{2i})}{\partial t_{1i}}\right]^{\omega_{1i}(1-\omega_{2i})}\left[\frac{-\partial S(t_{1i},t_{2i})}{\partial t_{2i}}\right]^{\omega_{2i}(1-\omega_{1i})}\left[S(t_{1i},t_{2i})\right]^{(1-\omega_{1i})(1-\omega_{2i})} \tag{9}$$

ω_{1i} and ω_{2i} be two indicator variables and and $i = 1,2,.....n$

(b) **When both t_{1i} and t_{2i} are uncensored or complete observations**

When both t_{1i} and t_{2i} are uncensored or complete observations, i.e., $\omega_{ji} = 1$, then the likelihood function in Equation (8) will reduce to:

$$\prod_{i=1}^{n}\left[\frac{\partial^2 S(t_{1i},t_{2i})}{\partial t_{1i}\partial t_{2i}}\right] \tag{10}$$

2.4. Deviance Information Criteria

Deviance Information Criteria (DIC) proposed by [14] is defined as:

$$D(\mathrm{Z}) = D(\hat{\mathrm{Z}}) - 2n_p$$

where $D(\hat{\mathrm{Z}})$ is the deviance and $n_p = \bar{D} - D(\hat{\mathrm{Z}})$ and $\bar{D}$ is the posterior deviance.

3. Results and Discussion

This section provides the goodness of fit results for all the copulas

From Table 1, all the results of p-value for the copulas are statistically significant, this proved the suitability of all the copulas for the dataset. The goodness of fit measures i.e., AIC and BIC are employed to select the best model, the model with least values of AIC and BIC is regarded as the best model. The results from the Plackett copula is the least for all the criterion, therefore Plackett copula is the best over FGM copula. For the estimation method, the MLE estimates are better than that of IFM estimates for the two models, based on standard error values.

Table 1. Standard error, p-value and goodness of fit measures results for the copula.

Copula.	Methods	SE	p-Value	Dependence Parameter	AIC	BIC
FGM Copula		0.0000	0.0000	37.0010	6177.690	6179.106
Plackett Copula	MLE	2.0970	0.0000	41.1350	6168.544	6169.960
FGM Copula		2.0970	0.0000	28.8580	6200.704	6202.120
Plackett Copula	IFM	2.9660	0.0000	29.9265	6196.408	6197.824

From Tables 2 and 3, The joint posterior distribution is obtained by combining the likelihood function with joint prior distribution to have some information of interest, this information were derived by generating different gibbs sample for each parameter. The different sample generated helped for observing the value of DIC as sample size increases and its clearly showed that this process needs large sample size for small value of DIC and batter selection of model. Here the Clayton copula with least value of DIC for different sample size is regarded as best model over FGM for the censored and uncensored cases.

Table 2. Posterior summary statistics for censored dataset using FGM and Clayton copula function.

Gibbs Samples for Parameters		FGM COPULA			CLAYTON COPULA		
	Par.	Mean	MC Error	95% CI	Mean	MC Error	95% CI
1000	α_1	0.9738	0.0040	(0.8889, 0.9998)	0.8720	0.0204	(0.6556, 0.9986)
	α_2	61.000	5.9670	(5.1140, 94.960)	13.450	2.4130	(1.5550, 36.860)
	ϕ	8.7960	0.8389	(0.9910, 14.340)	17.850	3.3990	(1.4890, 49.380)
		DIC = 4409			DIC = 2981		
10,000	α_1	0.9734	0.0021	(0.8894, 0.9994)	0.7746	0.0052	(0.6479, 0.9887)
	α_2	45.180	1.1470	(9.7790, 84.0200)	25.700	0.5973	(1.8310, 33.680)
	ϕ	69.660	2.1100	(1.5960, 92.9900)	44.270	1.0790	(1.7960, 57.560)
		DIC = 3868			DIC = 2502		
100,000	α_1	0.9747	0.0004	(0.9063, 0.9994)	0.7624	0.0010	(0.6425, 0.8983)
	α_2	43.650	0.1995	(34.740, 53.680)	26.850	0.1072	(20.940, 33.500)
	ϕ	76.340	0.3931	(60.640, 93.550)	46.940	0.1988	(36.750, 58.350)
Gibbs samples for Parameters		DIC = 3786			DIC = 2448		

Table 3. Posterior summary statistics for uncensored dataset using FGM and Clayton copula function.

Gibbs Samples for Parameters	FGM COPULA				CLAYTON COPULA		
	Par.	Mean	MC Error	95% CI	Mean	MC Error	95% CI
1000	α_1	0.9224	0.0154	(0.6035, 0.9993)	0.7164	0.0113	(0.5797, 0.8807)
	α_2	15.750	1.2350	(5.1590, 29.210)	8.3210	0.4135	(0.9622, 11.610)
	ϕ	22.100	2.9410	(0.9919, 44.240)	22.370	0.9948	(4.8590, 31.990)
			DIC = 4489			DIC = 2679	
10,000	α_1	0.9665	0.0030	(0.8747, 0.9992)	0.7255	0.0022	(0.6162, 0.8484)
	α_2	14.540	0.2097	(11.080, 21.480)	9.8050	0.0897	(7.0100, 12.130)
	ϕ	36.100	0.7086	(5.9070, 45.080)	21.360	0.1585	(16.650, 27.690)
			DIC = 4188			DIC = 2631	
100,000	α_1	0.9681	0.0006	(0.8864, 0.9991)	0.7267	0.0006	(0.6159, 0.8497)
	α_2	14.400	0.0223	(11.640, 17.500)	9.9510	0.0173	(7.8550, 12.230)
	ϕ	37.610	0.1258	(30.400, 45.410)	21.240	0.0238	(16.730, 26.390)
			DIC = 4156			DIC = 2631	

4. Conclusions

The univariate Toppleone distribution introduced by [12] with close forms of cumulative distribution function i.e [0, 1] is extended to unbound limit called Log-Toppleone distribution, the shapes of the hazard function can be increase, decrease or constant, therefore, can serve as an alternative distribution to the gamma, Weibull and exponential distributions. Bivariate of this proposed distribution is introduced by joining probability density function using three distinct copulas, first, two models are studied based on FGM and Plackett copula, the parameters were estimated using the MLE and IFM estimation method and the Plackett copula with least values of AIC and BIC for all two methods of estimation fit the dataset very well compare to FGM copula. Another two copulas: Clayton and FGM were implemented using the Bayesian method of estimation, this method is based on Markov Chain Monte Carlo simulation technique and the criteria used is Deviance information criteria (DIC). The Clayton copula with least value of DIC for different sample size is regarded as best model over FGM for the censored and uncensored cases.

Acknowledgments: The authors would like to thank the Universiti Teknologi PETRONAS for providing support to this project.

References

1. Pearson, K. Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Philos. Trans. R Soc. Lond A* **1895**, *186*, 343–414.
2. Hastings, J.C.; Mosteller, F.; Tukey, J.W.; Windsor, C. Low moments for small samples: A comparativestudy of order statistics. *Ann. Stat.* **1947**, *18*, 413–426.
3. Gupta, R.D.; Kundu, D. Generalized exponential distributions. *Aust. N. Zealand J. Stat.* **1998**, *41*, 173–188.
4. Usman, A.; Doguwa, S.I.S.; Alhaji, B.B.; ; Imam, A.T. A New Generalized Weibull Odd Frechet Family of Distributions: Statistical Properties and Applications. *Asian J. Probab. Stat.* **2020**, *9*, 25–43.
5. Eugene, N.; Lee, C.; Famoye, F. The beta-normal distribution and its applications. *Commun. Stat. Theory Methods* **2002**, *31*, 497–512.
6. Alzaatreh, A.; Famoye, F.; Lee, C. A new method for generating families of continuous distributions. *Metron* **2013**, *71*, 63–79.
7. Morgenstern, D. Einfache beispiele zweidimensionaler verteilungen. *Mitteilings Bl. Fur Math. Stat.* **1956**, *8*, 234–235.
8. Farlie, D.J. The performance of some correlation coe_cients for a general bivariate distribution. *Biometrika* **1960**, *47*, 307–323.
9. Gumbel, E.J. Bivariate exponential distributions. *J. Am. Stat. Assoc.* **1960**, *55*, 698–707.
10. Plackett, R.L. A class of bivariate distributions. *J. Am. Stat. Assoc.* **1965**, *60*, 516–522. https://doi.org/10.2307/2282685.

11. Clayton, D.G. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **1978**, *65*, 141–151.
12. Topp, C.W.; Leone, F.C. A Family of J-Shaped Frequency Functions. *J. Am. Stat. Assoc.* **1955**, *50*, 209–219.
13. Sklar, A. Fonctions de r'epartition `a *n* dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **1959**, *8*, 229–231.
14. Spiegelhalter, D.; Best, N.; Carlin, B.; van der Linde, A. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B* **2002**, *64*, 583–639.