http://communications.science.ankara.edu.tr

Commun.Fac.Sci.Univ.Ank.Ser. A1 Math. Stat.
Volume 72, Number 2, Pages 397–406 (2023)
DOI:10.31801/cfsuasmas.1179189
ISSN 1303-5991 E-ISSN 2618-6470

Research Article; Received: September 23, 2022; Accepted: December 18, 2022

A NEW TRANSMUTATION: CONDITIONAL COPULA WITH EXPONENTIAL DISTRIBUTION

Hüseyin ÜNÖZKAN[1] and Mehmet YILMAZ[2]

[1]Halic University, Faculty of Engineering, Department of Industrial Engineering, İstanbul, TÜRKİYE
[2]Ankara University, Faculty of Science, Department of Statistics, Ankara, TÜRKİYE

Abstract. In these days, many different techniques are implemented for generating distributions. The core aim in generating distribution, is better modeling capability. With generating new distribution more reliable and appropriate models are available for data sets. In this paper, a new distribution is gained by evaluating the conditional diagonal section of the bivariate Farlie-Gumbel-Morgenstern distribution with exponential marginals. Specifications and characteristics of this new distribution are studied. The statistical assessment and some reliability analyzes are carried out. The success of the new distribution on statistical modeling is detected by using data sets in literature. It is concluded that this new distribution suggests a model that can be used effectively in many different lifetime data sets.

1. Introduction

The exponential distribution is one of the most popular statistical distributions. This valuable distribution has been used widely in modeling time data sets ([11], [4], [12]). Exponential distribution has also been used in modeling other kinds of data sets (see [12]).

Although this distribution is very capable of modeling very different kinds of lifetime data sets, in some data sets, the modeling success rate may be lower. In some studies-to fix this situation-researchers add more parameters for better modeling ([10], [2], [8].

Exponential distribution has some specialties that this distribution can be used efficiently in industrial engineering and stochastic processes. ([11], [4]). The most

2020 *Mathematics Subject Classification.* 62E10, 62E15.
Keywords. Farlie-Gumbel-Morgenstern distribution, copula, reliability analysis, generating distribution, exponential distribution.
[1] huseyinunozkan@halic.edu.tr-Corresponding author; 0000-0001-9659-287X
[2] mehmetyilmaz@ankara.edu.tr; 0000-0002-9762-6688.

important and most known specialty of Exponential distribution is memoryless specialty. Exponential distribution also has a constant hazard rate.

In this study, main aim is generating an efficient statistical distribution which is more appropriate in some data sets than exponential distribution and other lifetime distributions.

We use Farlie-Gumbel-Morgenstern distribution and each marginal distribution in that copula function is Exponential distribution. A very similar technique was used in a study to gain a new distribution ([14]). In this study a different condition is carried out for achieving a new distribution. In this article, Exponential distribution gains better capability.

In this study, a new distribution for analyzing many different kinds of time data sets was suggested. This new distribution gains good results in modeling customer waiting times, time intervals in earthquakes and broken times in mechanic instruments. In our presentation at first new distribution is derived. And then the properties of new distribution are shown, and important characteristics are introduced. Section 4 illustrates the application of the new distribution on three data sets. There is the comparison of new distribution with most known lifetime distributions via data sets in the literature.

2. Material and Method

Theorem 1. ***(Sklar's Theorem):*** *Let F be a joint cumulative distribution function and H and G are continuous marginals, then there is a unique copula function C in R for every x and y ([13]).*

$$F(x,y) = C\left(H(x), G(y)\right).$$

Farlie-Gumbel-Morgenstern (FGM) copula with marginals has a formula as follows ([9])

$$C\left(u,v\right) = uv + \lambda uv\left(1-u\right)\left(1-v\right).$$

Two dimensional FGM with marginals $H\left(x\right)$ and $G\left(y\right)$ is as follows.

$$F\left(x,y\right) = H\left(x\right)G\left(y\right)\left[1+\lambda\bar{H}\left(x\right)\bar{G}\left(y\right)\right],$$

where $\bar{H}$ and $\bar{G}$ are the respective survival functions and $\lambda \in [-1,1]$ represents association parameter.

First, we assume that H and G are the same. Next, we will deal with the probability that the first component will fail in this range, when it is known that the second component fails in the range $(0,t]$. Then the conditional distribution function is as below.

$$\begin{aligned}Pr\left(X\leq t|Y\leq t\right)&=\frac{H\left(t\right)G\left(t\right)\left[1+\lambda\bar{H}\left(t\right)\bar{G}\left(t\right)\right]}{G\left(t\right)}\\&=H\left(t\right)\left[1+\lambda\bar{H}\left(t\right)\bar{G}\left(t\right)\right]\\&=H\left(t\right)\left[1+\lambda\bar{H}^{2}\left(t\right)\right]\\&=\left(1+\lambda\right)H\left(t\right)-\lambda H\left(t\right)\left(1-\bar{H}^{2}\left(t\right)\right).\end{aligned}$$

Thus, we obtain a univariate distribution. Let a random variable T distributed as above and F stand for this new distribution. Now, we explore what the association parameter means in this univariate case:
By taking $1+\lambda=2\delta$, where $\delta\in[0,1]$, we have

$$\begin{aligned}F(t)&=\left(1+\lambda\right)H\left(t\right)-\lambda H\left(t\right)\left(1-\bar{H}^{2}\left(t\right)\right)\\&=2\delta H\left(t\right)+(1-2\delta)H\left(t\right)\left(1-\bar{H}^{2}\left(t\right)\right)\\&=\delta\big[2H\left(t\right)-2H^{2}(t)+H^{3}(t)\big]+(1-\delta)H\left(t\right)\left(1-\bar{H}^{2}\left(t\right)\right).\end{aligned}$$

The expression in square brackets is actually a convex combination of two distribution functions as follows:

$$\frac{2}{3}\big(3H\left(t\right)-3H^{2}(t)+H^{3}(t)\big)+\frac{1}{3}H^{3}(t),$$

where components respectively represent the distributions of $min\{T_1,T_2,T_3\}$ and $max\{T_1,T_2,T_3\}$, when T_1, T_2 and T_3 are independently distributed as H. Accordingly, $H\left(t\right)\left(1-\bar{H}^2(t)\right)$ represents a distribution of $max\{T_1,min\{T_2,T_3\}\}$. Thus, F is a distribution function representing the convex combination of two distribution functions, while λ represents a transformed combination parameter.
Hence, probability density function (pdf) of this distribution is as below.

$$f\left(t\right)=h\left(t\right)\left(1+\lambda\overline{H}\left(t\right)\left(1-3H\left(t\right)\right)\right),$$

where $h\left(t\right)$ is a pdf of base distribution. In prospect of $H\left(t\right)=1-e^{-\theta t}$, we have

$$F\left(t\right)=\left(1-e^{-\theta t}\right)\left(1+\lambda e^{-2\theta t}\right),\tag{1}$$

and pdf of this distribution is as below.

$$f\left(t\right)=\theta e^{-\theta t}\left(1+\lambda\left(e^{-2\theta t}-2e^{-\theta t}\left(1-e^{-\theta t}\right)\right)\right),\tag{2}$$

where $\lambda\in[-1,1]$ and $\theta>0$. Plots of probability density function are as follows.

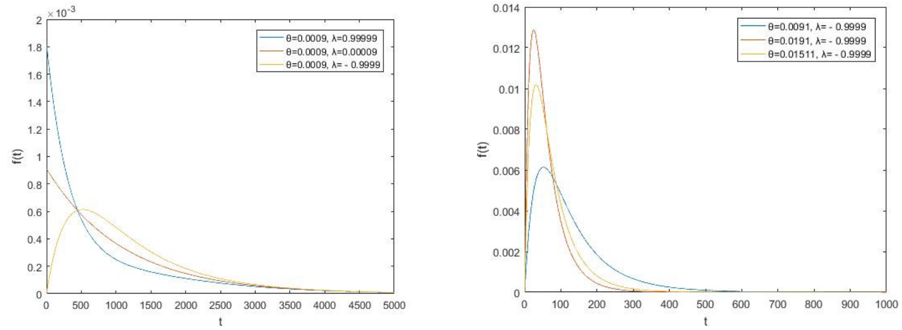

FIGURE 1. The pdf graphs for some parameters

According to plots in Figure 1, it was easily seen that parameter λ is the shape parameter and parameter θ is the location parameter. With the value of the parameter λ the shape of the probability density function changes significantly and this specialty gives us hope for this distribution to use in different kinds of data sets at the same time.

Survival and hazard rate functions of new distribution are as follows;

$$\bar{F}(t) = e^{-\theta t} - \lambda e^{-2\theta t} + \lambda e^{-3\theta t} = e^{-\theta t}\left(1 - \lambda e^{-\theta t}\left(1 - e^{-\theta t}\right)\right)$$

and

$$r(t) = \frac{f(t)}{\bar{F}(t)} = \theta\left[2 - \frac{1 - \lambda e^{-2\theta t}}{1 - \lambda e^{-\theta t}\left(1 - e^{-\theta t}\right)}\right].$$

If we want to calculate the risk in the starting point, we reach this value as below.

$$\lim_{t\to 0}\left(\theta\left[2 - \frac{1 - \lambda e^{-2\theta t}}{1 - \lambda e^{-\theta t}\left(1 - e^{-\theta t}\right)}\right]\right) = (1+\lambda)\theta.$$

If we want to calculate long term risk, we reach this value as below.

$$\lim_{t\to \infty}\left(\theta\left[2 - \frac{1 - \lambda e^{-2\theta t}}{1 - \lambda e^{-\theta t}\left(1 - e^{-\theta t}\right)}\right]\right) = \theta.$$

In Figures 1 and 2, it can be seen easily that parameter λ changes both the shapes of probability density function and hazard rate function. Therefore, we consider that this new distribution may be successful in analyzing different data sets which may have opposite kinds of risk in the same time.

When parameter λ is between $(0, 1]$, the shape of the hazard rate function becomes bathtub. With this there are decreasing starting deaths, and in the beginning some components rapidly break down. After that there is nearly a constant hazard rate for a while. At last in the final part, the components which complete life time, break down in increasing rate and the process completes.

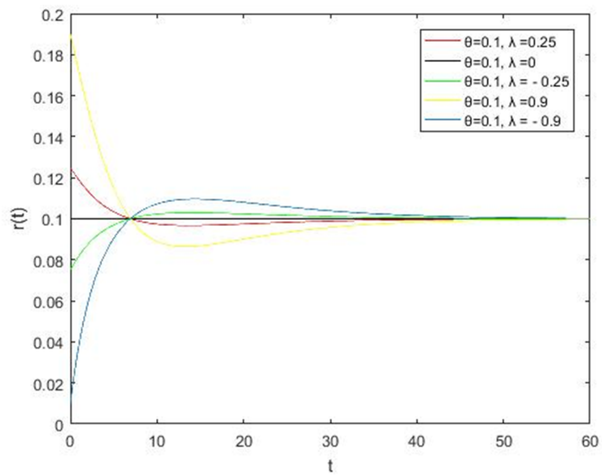

FIGURE 2. The plots of hazard rate function

When parameter λ is between $[-1, 0)$, the shape of the hazard rate function becomes the inverse position of the bathtub shape. This curve is symmetric to value of parameter θ which is the hazard rate of the exponential distribution. With this, there are increasing starting deaths, and at the beginning some components break down rapidly. After that there is a balance and nearly constant hazard rate. At last, the components which complete life time, break down in decreasing rate and the process completes. This shape calls upside-down bathtub or inverse bathtub.

3. Characteristics of Distribution

3.1. **Moment Generating Function (mgf).**

$$
\begin{aligned}
M_T(v) &= E\left(e^{vT}\right) \\
&= \int_0^\infty e^{vt}\left(\theta e^{-\theta t}\right)\left(1+\lambda\left(e^{-2\theta t}-2e^{-\theta t}\left(1-e^{-\theta t}\right)\right)\right)dt \\
&= \int_0^\infty \left(\theta e^{-t(\theta-v)} - 2\lambda\theta e^{-t(2\theta-v)} + 3\lambda\theta e^{-t(3\theta-v)}\right)dt \\
&= \int_0^\infty \theta e^{-t(\theta-v)}dt - 2\lambda\int_0^\infty \theta e^{-t(2\theta-v)}dt + 3\lambda\int_0^\infty \theta e^{-t(3\theta-v)}dt \\
&= \frac{\theta}{\theta-v} - \frac{2\lambda\theta}{2\theta-v} + \frac{3\lambda\theta}{3\theta-v},
\end{aligned}
$$

where $\theta > v$. This is a linear combination of mgf's of exponential distributions with three different means $\frac{1}{\theta}$, $\frac{1}{2\theta}$ and $\frac{1}{3\theta}$. In other words, for $Y_j \sim Exponential\left(\frac{1}{j\theta}\right)$, $j = 1, 2, 3$ this mgf can be represented as a mgf's of Y_j which are $M_{Y_j}(v) = \frac{j\theta}{j\theta - v}$.

$$M_T(v) = M_{Y_1}(v) - \lambda M_{Y_2}(v) + \lambda M_{Y_3}(v), \quad \theta > v. \tag{3}$$

3.2. **k. th Raw Moment.** We can provide raw moment easily by using (3) as follows:

$$E\left(T^k\right) = \frac{\Gamma(k+1)}{\theta^k} - \lambda\frac{\Gamma(k+1)}{2^k\theta^k} + \lambda\frac{\Gamma(k+1)}{3^k\theta^k} = \frac{k!}{\theta^k}\left[1 - \lambda\frac{1}{2^k} + \lambda\frac{1}{3^k}\right]$$

3.3. **Expected Value and Second Order Raw Moment.**

$$E(T) = \frac{1}{\theta} - \frac{\lambda}{2\theta} + \frac{\lambda}{3\theta} = \frac{6-\lambda}{6\theta},$$

$$E\left(T^2\right) = \frac{2}{\theta^2} - \frac{\lambda}{2\theta^2} + \frac{2\lambda}{9\theta^2} = \frac{36-5\lambda}{18\theta^2}.$$

3.4. **Variance.**

$$Var(T) = E\left(T^2\right) - E(T)^2 = \frac{36-5\lambda}{18\theta^2} - \left(\frac{6-\lambda}{6\theta}\right)^2 = \frac{36+2\lambda-\lambda^2}{36\theta^2}.$$

3.5. **Maximum Likelihood Estimation.** The log-likelihood function for a random sample $T_1, T_2, \cdots, T_n$ from (1) is:

$$\ell(\theta, \lambda\ ; \underline{t}) = \log(L(\theta, \lambda; \underline{t})) = n\log\theta - \theta\sum_{i=1}^{n} t_i + \sum_{i=1}^{n}\log\left(1 + \lambda\left(3e^{-2\theta t_i} - 2e^{-\theta t_i}\right)\right).$$

Now, by using Log-likelihood function, we get partial derivatives with respect to λ and θ as follows:

$$\frac{\partial}{\partial\lambda}\ell(\theta, \lambda\ ; \underline{t}) = \sum_{i=1}^{n}\frac{\left(3e^{-2\theta t_i} - 2e^{-\theta t_i}\right)}{1 + \lambda\left(3e^{-2\theta t_i} - 2e^{-\theta t_i}\right)} = 0, \tag{4}$$

$$\frac{\partial}{\partial\theta}\ell(\theta, \lambda\ ; \underline{t}) = \frac{n}{\theta} - \sum_{i=1}^{n} t_i + \sum_{i=1}^{n}\frac{2\lambda te^{-\theta t_i} - 6\lambda te^{-2\theta t_i}}{1 + \lambda\left(3e^{-2\theta t_i} - 2e^{-\theta t_i}\right)} = 0. \tag{5}$$

Equating these two expressions (4) and (5) to zero and solving them simultaneously yields the ML estimates of the θ and λ.

4. Results and Discussions

Now, we will compare our new distribution with most known lifetime distributions by some different kinds of data sets. While comparing distributions, we will use Kolmogorov-Smirnov test statistics. In using Kolmogorov-Smirnov statistics, least statistic value is appraised as best modeling. $p$value of Kolmogorov-Smirnov statistics informs us about plausibility of the conformity.

Data 1: This data sets represent waiting times of bank customers (see Table 1). Data set was first used by [6] and later it was evaluated by [1], [14]. We compare new distribution with Lindley and CFGMWEM, because these distributions were used in modeling before.

Table 1. Customer waiting times

0,1	0,2	0,3	0,7	0,9	1,1	1,2	1,8	1,9	2
2,2	2,3	2,3	2,3	2,5	2,6	2,7	2,7	2,9	3,1
3,1	3,2	3,4	3,4	3,5	3,9	4	4,2	4,5	4,7
5,3	5,6	5,6	6,2	6,3	6,6	6,8	7,3	7,5	7,7
7,7	8	8	8,5	8,5	8,7	9,5	10,7	10,9	11
12,1	12,3	12,8	12,9	13,2	13,7	14,5	16	16,5	28

Table 2. Customer waiting times test results

Model	K-S	p
Lindley	0,08	0,84
CFGMWEM	0,0618	0,9651
New Distribution	0,061	0,9689

Once examining Table 2, it is clear that this new distribution is capable in modeling waiting times and offer a strong model. According to Kolmogorov-Smirnov test statistics the most appropriate model is the new generated distribution.

Data 2: Second data set represent broken times of ventilation in airplanes (see Table 3). It was used by [7] and later [12] used for comparing distributions.

Table 3. Broken times of ventilation in airplanes

23	261	87	7	120	14	62	47	225	71
246	21	42	20	5	12	120	11	3	14
71	11	14	11	16	90	1	16	52	95

Table 4. Test results of broken times of ventilation in airplanes

Model	K-S	p
Exponential Poisson-Lindley	0,129	0,6531
Weibull	0,1531	0,4394
CFGMWEM	0,1528	0,4414
New Distribution	0,1157	0,7745

When Table 4 is examined it is clear that this new distribution is capable in modeling broken times and offer a strong model. According to Kolmogorov-Smirnov test statistics the most appropriate model is the new generated distribution.

Data 3: We evaluate the time intervals for earthquakes in Iran (see Table 5). These data were analyzed by [3]. This data set was also studied in Alpha-Power Transformed Lindley Distribution by [5].

Table 5. Time intervals of earthquakes in Iran

136	1187	117	944	24	70	716	1126	378	166
152	264	275							

Table 6. Test results of time intervals of earthquakes in Iran

Model	K-S	p
Exponential Lindley	0,1307	0,9585
Weibull	0,1527	0,8783
New Distribution	0,1241	0,9735

In Table 6, it is clear that this new distribution is capable in modeling time intervals and offer a strong model. According to Kolmogorov-Smirnov test statistics the most appropriate model is the new generated distribution.

5. Conclusion

Although there are many different and capable statistical distributions in use today, many new distributions may be needed with different data sets and better modeling opportunities. The new distribution which introduced in this study is capable in modeling time data sets. There are many lifetime distributions but this distribution may be very helpful in analyzing times more appropriately.

But why our new distribution is capable in modeling different kinds of data sets? In part two we showed that the value of parameter λ could change the structure of our new distribution. So, we consider that the values of this parameter in modeling may be important. In Table 7 there are maximum likelihood estimations of the parameters in modeling data1 to data 3.

Table 7. Values of Parameters in Models

Data	θ	λ
Customer waiting times	0,1715	-0,622
Broken times of ventilation	0,0141	1
Time intervals of earthquakes	0,0022	0,4041

We see that the new distribution fits the datasets better than the other distributions. According to test results in Table 2 to Table 6 we suggest that the new distribution can be used in many kinds of time data sets.

Authors Contribution Statement The authors contributed equally and significantly in this manuscript, and they read and approved the final manuscript.

Declaration of Competing Interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement We thank the two anonymous reviewers whose comments/ suggestions helped improve and clarify this manuscript.

References

[1] Al-Mutairi, D. K., Ghitany, M. E., Kundu, D., Inferences on stress-strength reliability from Lindley distributions, *Communications in Statistics, Theory and Methods,* 42(8) (2013), 1443–1463, https://dx.doi.org/10.1080/03610926.2011.563011.

[2] Alghamedi, A., Dey, S., Kumar, D., Dobbah, S. A., A new extension of extended exponential distribution with applications, *Annals of Data Science,* 7(1) (2020), 139–162, https://dx.doi.org/10.1007/s40745-020-00240-w.

[3] Barreto-Souza, W., Bakouch, H. S., A new lifetime model with decreasing failure rate, *Statistics,* 47(2) (2013), 465–476, https://dx.doi.org/10.1080/02331888.2011.595489.

[4] Devore, J. L., Probability and Statistics for Engineering and the Sciences, 8. ed., Internat. Student ed., Brooks/Cole, Cengage Learning, [Belmont, Calif. [u.a.]], 2012.

[5] Dey, S., Ghosh, I., Kumar, D., Alpha-power transformed lindley distribution: Properties and associated inference with application to earthquake data, *Sankhya,* 73 (2011), 623–650, https://dx.doi.org/10.1007/s40745-018-0163-2.

[6] Ghitany, M. E., Atieh, B., Nadarajah, S., Lindley distribution and its application, *Mathematics and computers in simulation,* 78(4) (2008), 493–506.

[7] Linhart, H., Zucchini, W., Model Selection, John Wiley, 1986.

[8] Nadarajah, S., Haghighi, F., An extension of the exponential distribution, *Statistics,* 45(6) (2011), 543–558, https://dx.doi.org/10.1080/02331881003678678.

[9] Nelsen, R. B., An Introduction to Copulas, Second Edition, @Springer Series in Statistics, Springer Science+Business Media, Inc, New York, NY, 2006.

[10] Rather, N. A., Rather, T. A., New generalizations of exponential distribution with applications, *Journal of Probability and Statistics,* 2017(2017), 1–9.
[11] Ross, S. M., Introduction to Probability Models, Elsevier, 2010.
[12] Shanker, R., Fesshaye, H., Selvaraj, S., On modeling of lifetimes data using exponential and lindley distributions, *Biometrics & Biostatistics International Journal,* 2(5) (2015), 1–9, https://dx.doi.org/10.15406/bbij.2015.02.00042.
[13] Sklar, A., Fonctions de repartition an dimensionset leurs marges, *Publ. Inst. Statis. Univ* (1959), 229–231.
[14] Ünözkan, H., Yilmaz, M., A new method for generating distributions: An application to flow data, *International Journal of Statistics and Applications,* 9(3) (2019), 92–99, https://dx.doi.org/10.5923/j.statistics.20190903.04.