Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume25, Issue1 (2023) ISSN$_{Print}$: 1816-9171
Manuscript ID:

The survival function for COVID-19 patients using the XGD model estimated by Moments method and Maximum Likelihood method

Mohammed H. AL- Sharoot[1] and Karrar R. AL-Sayeab[2]
[1]*Professor Doctor of Statistic at the Faculty of Faculty of Administration and Economics of Al-Qadisiyah University, Iraq*
Email: mohammed.alsharoot@qu.edu.iq
[2]Master Student of the Faculty of Faculty of Administration and Economics of Al-Qadisiyah University, Iraq.
Email: karrar.alsayeab@jmu.edu.iq

Corresponding Author: Karrar R. AL-Sayeab *Affiliation:* University of Al-Qadisiyah *Email:* karrar.alsayeab@jmu.edu.iq

Abstract: *This paper studies some properties of the Xgamma probability distribution, which results from mixing the exponential distribution and the gamma distribution with specific weights. Then a new form of the distribution was proposed by replacing the mixing weights, one with the other for the XG distribution in the mixing function, to obtain the proposed distribution Xgamma2 (XG2D), after which the mathematical properties and survival functions of the proposed model was studied, then some properties of the exponential distribution were studied for comparison with the previous two distributions. The Moments method and the Maximum Likelihood function were used to estimate parameter for studied models. In addition, by using simulations, we found that the Maximum Likelihood method is the best in estimating the survival function of the proposed distribution. Finally, We found The data are better fit by Xgamma than the other of studied models when analyzing patients infected with Covid-19 virus in Najaf, because it achieved the lowest criteria (AIC, AICc, BIC) from other distributions, followed by the proposed Xgamma2 distribution, followed by the exponential distribution.*

Keywords- *Xgamma distribution, Xgamma2 distribution, Survival function, Moments estimation, Maximum Likelihood estimation.*

I. INTRODUCTION

The human need to continue living better by knowing the period of his survival when contracting a particular disease made him a basic motive to delve into the study of survival analysis, as the study of the time extending from (the beginning of the event), i.e. the beginning of the infection, up to the designated end point (the point of occurrence of the event recovery or death), and the term survival function is usually used in medical and life studies, and in the field of medicine, the patient's health status is still the most important thing that is studied in statistical analysis applications, through estimating the patient's health status. Based on the foregoing, the interest in this paper was to study the survival function for one of the distributions used in modeling survival data, which is the one-parameter Xgamma distribution, which is one of the important distributions that has proven its advantage in estimating the survival function, Subhradev Sen was the first to study this distribution in 2016, from a combination of exponential and gamma distributions with specified weights. Then he derived many mathematical and structural properties like moments, measures of kurtosis and skewness, after which important survival characteristics like risk, and applied it in estimating the survival function to Survival times of patients rested (in hours) for 20 patients receiving analgesics. The Xgamma distribution was better fit to the data than the other distributions was compared (Sen et al., 2016). In 2017, Subhradev Sen and others studied the weighted XGD (WXGD), then the length biased version of XGD (LBXG) was acquired as a special state of the density function of the XGD distribution. WXGD. The distribution characteristics (LBXG) and survival characteristics were studied. Finally, the (LBXG) distribution was compared with other distributions by applying the stress lifetime data for 23 deep-groove ball bearings to prove that (LBXG) was the best among the rest of the comparative models (Sen et al., 2017). In 2018, Altun, Emrah, and G. G. Hamedani introduced a new one-parameter distribution called "log-xgamma" by the random replacing variable of the Xgamma distribution with a natural logarithm function for the random variable. Two sets of real data for the log-xgamma distribution were analyzed compared with the Beta, Kumaraswamy and Topp-Leone distributions, and it was concluded the log-xgamma distribution has provided a better fit for the real data (Altun & Hamedani, 2018). In (2019) Yadav et al. derive a new probability distribution, whose was the Inverted Xgamma (IXG) distribution. Survival characteristics, reverse moments, reverse life expectancy and ordered statistics of the proposed distribution were studied. A data set of failure times (in minutes) for a sample of 15 electronic components was used to prove the applicability of the IXGD to real life and to compare it with other inverse distributions. It was concluded that the IXG distribution is the most appropriate in matching the real data used on the basis of comparison criteria (AIC, BIC) (Yadav et al., 2019).

II. Xgamma distribution:

"Sen et al., 2016" proposed special mixture of EXP (θ) and gamma (3,θ) distributions, denoted XGD, an assuming that the two parameters of mixing are $\pi_1 = \frac{\theta}{1+\theta}$, $\pi_2 = 1-\pi_1 = \frac{1}{1+\theta}$. He derived the survival function, as well as inferences of the XGD (Elshahhat & Elemary, 2021).

T is a random variable then the XGD is define as follow:

$$f\,(\mathrm{t}) = \pi_1 f_1(t;\theta) + \pi_2 f_2(t;\theta)$$

$$f\,(\mathrm{t}) = \frac{\theta}{1+\theta}\left(\theta e^{-\theta t}\right) + \frac{1}{1+\theta}\left(\frac{\theta^3 t^2}{\rceil 3} e^{-\theta t}\right)$$

$$f\,(\mathrm{t};\theta) = \frac{\theta^2}{(1+\theta)}\left(1+\frac{\theta}{2}t^2\right)e^{-\theta t} \quad ,t>0,\theta>0 \qquad (1)$$

This is represented by T ~ xgamma(θ).

The cumulative density function (CDF) of T is given by

$$F(\mathrm{t}) = 1 - \frac{\left(1+\theta+\theta t+\frac{\theta^2 t^2}{2}\right)}{(1+\theta)}e^{-\theta t} \quad ,t>0,\theta>0 \qquad (2)$$

The survival function S(t) and the hazard function h(t) are defined a follow

$$S(t) = \frac{\left(1+\theta+\theta t+\frac{\theta^2 t^2}{2}\right)}{(1+\theta)}e^{-\theta t} \quad ,t>0,\theta>0 \qquad (3)$$

$$h(t) = \frac{\left(1+\frac{\theta t^2}{2}\right)\theta^2}{(1+\theta+\theta t+\frac{\theta^2 t^2}{2})} \qquad (4)$$

The moments and the variance and the other properties are defined a follow:

$$\mu'_r = \frac{r![2\theta+(\mathrm{r}+1)(\mathrm{r}+2)]}{2\theta^r(1+\theta)} \qquad for \quad r=1,2,3,\ldots \qquad (5)$$

$$\mu'_1 = \frac{(\theta+3)}{\theta(1+\theta)} \qquad (6)$$

$$\mathrm{var(t)} = \frac{\theta^2+8\theta+3}{\theta^2(1+\theta)^2} \qquad (7)$$

$$Mode = \begin{cases} \dfrac{1+\sqrt{1-2\theta}}{\theta}, & 0<\theta\le 0.5 \\ 0, & otherwise \end{cases} \quad (8)$$

$$Skewness\,(\Upsilon_1) = \frac{2(\theta^3+15\theta^3+9\theta+3)}{\left(\theta^2+8\theta+3\right)^{3/2}} \quad (9)$$

$$Kurtosis\,(\Upsilon_2) = \frac{3(3\theta^4+64\theta^3+102\theta^2+72\theta+15)}{\left(\theta^2+8\theta+3\right)^2} \quad (10)$$

III. Xgamma2 distribution (XG2):

We propose a new form for the Xgamma distribution, by mixing the exponential distribution with the parameter (θ) and the gamma distribution with the two parameters (3,θ) and switching the mixing weights in the mixing equation for the XG distribution, one in place of the other, to become $\pi_1 = \frac{1}{1+\theta}$ and $\pi_2 = 1-\pi_1 = \frac{\theta}{1+\theta}$; Let's get the proposed new format as follows:

$$f\,(t;\theta) = \pi_1 f_1(t) + \pi_2 f_2(t)$$

Where

$$\pi_1 = \frac{1}{1+\theta}, \quad and \quad \pi_2 = 1-\pi_1 = \frac{\theta}{1+\theta}$$

$$f_1(t) = \theta e^{-\theta t} \quad and \quad f_2(t) = \frac{t^2\theta^3}{\rceil 3} e^{-\theta t}$$ then:

$$f\,(t;\theta) = \frac{1}{1+\theta}\left(\frac{1}{\theta}e^{-\theta t}\right) + \frac{\theta}{1+\theta}\left(\frac{t^2\theta^3}{\rceil 3}e^{-\theta t}\right)$$

$$f\,(t) = \frac{\theta}{1+\theta}e^{-\theta t}\left(1+\frac{\theta^3 t^2}{2}\right), \quad t>0, \quad \theta>0 \quad (11)$$

This is represented by T ~ Xgamma2 (θ), and is denoted a XG2D.

The corresponding cumulative distribution function (C.D.F) of the XG2D is:

$$F\,(t;\theta) = 1-\frac{e^{-\theta t}}{(1+\theta)}\left[1+\frac{t^2\theta^3}{2}+t\theta^2+\theta\right] \quad (12)$$

The survival S(t) function and hazard function h(t) are given by

$$S\,(t) = \frac{e^{-\theta t}}{(1+\theta)}\left(1+\theta+\theta^2 t+\frac{\theta^3 t^2}{2}\right) \quad ,t>0,\theta>0 \quad (13)$$

$$h(t) = \frac{\theta(1+\frac{\theta^3 t^2}{2})}{1+\theta+\theta^2 t+\frac{\theta^3 t^2}{2}} \quad (14)$$

The rth moments of x about zero is:

$$\mu_r' = \frac{r!}{(1+\theta)\theta^r}\left(1+\frac{\theta(r+2)(r+1)}{2}\right) \quad for\ r=1,2,... \tag{15}$$

The mean and the variance and the other properties are defined a follow:

$$\mu_1' = \frac{(1+3\theta)}{\theta(1+\theta)} \tag{16}$$

$$\text{var}(x) = \frac{(3\theta^2+8\theta+1)}{\theta^2(1+\theta)^2} \tag{17}$$

$$Mode = \begin{cases} \dfrac{\theta+\sqrt{\theta(\theta-2)}}{\theta^2}, & 2\le\theta<\infty \\ 0, & otherwise \end{cases} \tag{18}$$

$$Skewness\,(\Upsilon_1) = \frac{2(3\theta^3+9\theta^3+15\theta+1)}{\left(3\theta^2+8\theta+1\right)^{2/3}} \tag{19}$$

$$Kurtosis\,(\Upsilon_2) = \frac{3(15\theta^4+72\theta^3+102\theta^2+64\theta+3)}{\left(3\theta^2+8\theta+1\right)^2} \tag{20}$$

$$M_x(t) = \frac{\theta}{(1+\theta)}\left(\frac{1}{(\theta-t)}+\frac{\theta^3}{(\theta-t)^3}\right); t\in R \tag{21}$$

IV. The Exponential distribution (EXP):

The exponential distribution has proven to be one of the distinct continuous distributions over time, probably due to its simple and useful nature in modeling life-time data and various problems with waiting-time events. The probability density function and mathematical, survival properties are given below (Okoli et al., 2016) (Gu, 2014) (Triana & Purwadi, 2019):

$$f(t;\theta) = \theta e^{-\theta t} \quad ,t>0,\theta>0 \tag{22}$$

$$F(\text{t}) = 1-e^{-\theta t} \quad ,\text{t}>0,\theta>0 \tag{23}$$

$$S(t) = e^{-\theta t} \quad ,t>0,\theta>0 \tag{24}$$

$$h(t) = \theta \tag{25}$$

$$\mu_1' = \frac{1}{\theta} \tag{26}$$

$$\text{var(t)} = \frac{1}{\theta^2} \tag{27}$$

V. Estimation of the Parameter

V.I. The maximum Likelihood Estimation:

The maximum likelihood method is widely used in statistical estimation. There are differing opinions about who proposed the method first. While Fisher invented the name of maximum

likelihood, who spread the use of it widely, and demonstrated its optimality properties.(Le Cam, 1990)
We calculate it for the xgamma2 distribution as follows:
Assuming that (T1, T2, ..., Tn) is a random sample of size n from the XG2D, the estimate can be found by the MLE for the parameter θ as follows:

$$L = \prod_{i=1}^{n} f\,(t_i;\theta)$$

$$L = \prod_{i=1}^{n} \frac{\theta}{1+\theta}\left(1+\frac{\theta^3}{2}t_i^2\right)e^{-\theta t_i}$$

$$LnL\,(\theta;\mathrm{t}) = n\ln\theta - n\ln(1+\theta) + \sum_{i=1}^{n}\ln\left(1+\frac{\theta^3}{2}t_i^2\right) - \theta\sum_{i=1}^{n} t_i$$

Let $l\,(\theta;t) = \frac{\partial\,\mathrm{lnL}}{\partial\theta} = 0$

$$\frac{n}{\theta} - \frac{n}{1+\theta} + \sum_{i=1}^{n}\frac{3\theta^2 t_i^2}{\left(2+\theta^3 t_i^2\right)} - \sum_{i=1}^{n} t_i = 0 \qquad (28)$$

Since (28) is a non-linear equation, it cannot be solved analytically; hence, numerical method is used, like Newton-Raphson .
Let the initial solution is:

$$\theta_0 = \frac{n}{\sum_{i=1}^{n} t_i}$$

$$\theta^{(i)} = \theta^{(i-1)} - \frac{l\,(\theta|x)}{l'(\theta|x)}$$

When $\theta(i) \cong \theta(i-1)$, we choose $\theta_{mle} = \theta(i)$.
Then the survival function and the hazard rate function by MLE:

$$\hat{S}\,(t) = \frac{e^{-t\hat{\theta}_{MLE}}}{(1+\hat{\theta}_{MLE})}\left(1+\hat{\theta}_{MLE} + t\,\hat{\theta}_{MLE}^2 + \frac{t^2\hat{\theta}_{MLE}^3}{2}\right) \qquad (29)$$

$$\hat{h}(t) = \frac{\hat{\theta}_{MLE}\,(1+\frac{t^2\hat{\theta}_{MLE}^3}{2})}{\left(1+\hat{\theta}_{MLE} + t\,\hat{\theta}_{MLE}^2 + \frac{t^2\hat{\theta}_{MLE}^3}{2}\right)} \qquad (30)$$

V.II. Method of Moments (MoM):

This method is based on finding N of the community moments and equating it with n of the sample moments to get a number of equations equal to the number of parameters to be estimated and by solving them we get the estimators (Sen & Chandra, 2017).

Assuming that (t1,t2,…,tn) a random sample of size n was drawn from the Xgamma2 distribution to estimate the parameter θ by moment method.

We equate the sample mean with the original population mean (the first moment around the origin of the Xgamma2 distribution) to find the parameter estimator.

$$E\,(t) = \frac{(1+3\theta)}{\theta(\theta+1)}$$

$$\bar{t} = \frac{\left(1+3\hat{\theta}\right)}{\hat{\theta}\left(\hat{\theta}+1\right)}$$

$$\hat{\theta}_{mom} = \frac{3-\bar{t}-\sqrt{\bar{t}^{\,2}-2\bar{t}+9}}{2\bar{t}} \tag{31}$$

And so we can compute the Moments estimation for the survival and hazard rate functions for XG2D:

$$\hat{S}(t) = \frac{e^{-t\hat{\theta}_{mom}}}{(1+\hat{\theta}_{mom})}\left(1+\hat{\theta}_{mom}+t\hat{\theta}^2_{mom}+\frac{t^2\hat{\theta}^3_{mom}}{2}\right) \tag{32}$$

$$\hat{h}(t) = \frac{\hat{\theta}_{mom}\,(1+\frac{t^2\hat{\theta}^3_{mom}}{2})}{\left(1+\hat{\theta}_{mom}+t\hat{\theta}^2_{mom}+\frac{t^2\hat{\theta}^3_{mom}}{2}\right)} \tag{33}$$

VI. Simulation study

For examining the behavior of selected estimators and survival functions of the xgamma2 distribution, a Monte Carlo simulation was conducted with M = 1000 iterations. A sample size (10, 50, 100) was assumed, and (θ) were assumed (0.5, 1, 5), and we use the IMSE measure to select the appropriate one Calculated measure include:

$$IMSE\left(\hat{S}(t)\right) = \frac{1}{M}\sum_{i=1}^{M}\left[\frac{1}{n_i}\sum_{j=1}^{n_i}\left(\hat{S}_i(t_j)-S_i(t_j)\right)^2\right] \tag{34}$$

VI.I. generation random data

VI.I.I. For generation of random data based on the xgamma, see (Sen et al., 2016).

VI.I.II. For generation of random data T_j, j = 1, 2,..., n based on the xgamma2, use this algorithm:

a. Generate $A_j \sim uniform(0,\ 1)$, j = $1,\ldots,n$

b. Generate $V_j \sim EXP(\theta)$, j = $1,\ldots,n$

c. Generate $W_j \sim gamma(3,\theta)$, j = $1,\ldots,n$

d. Generate $T_j \sim Xgamma2(\theta)$, j = $1,\ldots,n$ by the following:

If $A_j \le \frac{1}{(1+\theta)}$, then set $T_j = V_j$, Otherwise, set $T_j = W_j$.

Table1. Simulation results when θ=0.5, 1, 5

n	θ	Dis.	XG		XG2		EXP	
			mle	mom	mle	mom	mle	mom
10	0.5	$\hat{\theta}$	0.527387	0.5251378	0.6247297	0.6178786	0.5549105	0.5549105
		S_{real}	0.5009148		0.5128728		0.5004917	
		$\hat{S}$	0.489425	0.4926569	0.487079	0.4953837	0.4876831	0.4876831
		IMSE	**0.008186**	0.0082596	**0.0054646**	0.0056707	0.0077705	0.0077705
	1	$\hat{\theta}$	1.055763	1.049523	1.045013	1.120585	1.106996	1.106996
		S_{real}	0.4966129		0.5224438		0.4987454	
		$\hat{S}$	0.487108	0.49043	0.5213259	0.515708	0.4868873	0.4868873
		IMSE	0.008216	**0.0081979**	**0.0032401**	0.0046299	0.0077864	0.0077864

	5	$\hat{\theta}$	5.434126	5.426155	4.849746	5.267983	5.450326	5.450326
		S_{real}	0.4975784		0.5007694		0.4949617	
		$\hat{S}$	0.486922	0.4874187	0.5162514	0.4950024	0.486179	0.486179
		IMSE	**0.007918**	0.0080036	0.0201164	**0.0080628**	0.0077338	0.0077338
50	0.5	$\hat{\theta}$	0.504589	0.5044612	0.5256333	0.5151045	0.5123625	0.5123625
		S_{real}	0.5005795		0.4971599		0.5020491	
		$\hat{S}$	0.498732	0.4991479	0.4918952	0.495862	0.4984967	0.4984967
		IMSE	**0.001585**	0.0015987	0.0013622	**0.0013112**	0.0014929	0.0014929
	1	$\hat{\theta}$	1.01288	1.011593	1.010142	0.9994652	1.018877	1.018877
		S_{real}	0.4999777		0.4941066		0.5012358	
		$\hat{S}$	0.497281	0.4979977	0.4934746	0.496713	0.4989847	0.4989847
		IMSE	**0.001554**	0.0015598	**0.0007832**	0.0010584	0.0014736	0.0014736
	5	$\hat{\theta}$	5.082211	5.080838	5.050899	5.048531	5.088361	5.088361
		S_{real}	0.5005785		0.4997756		0.4990417	
		$\hat{S}$	0.498277	0.4983546	0.4975292	0.4986164	0.4972008	0.4972008
		IMSE	**0.001481**	0.0014882	**0.0013495**	0.0015373	0.0015560	0.0015560
100	0.5	$\hat{\theta}$	0.502194	0.5017607	0.5086084	0.5077258	0.5042744	0.5042744
		S_{real}	0.4990645		0.5001627		0.4989836	
		$\hat{S}$	0.49822	0.4987255	0.4985285	0.4992819	0.4980696	0.4980696
		IMSE	**0.000786**	0.0007990	**0.0004518**	0.0005105	0.0007157	0.0007157
	1	$\hat{\theta}$	1.003654	1.002944	1.010779	1.012146	1.008111	1.008111
		S_{real}	0.498844		0.500671		0.499392	
		$\hat{S}$	0.498457	0.4988406	0.4993821	0.4998774	0.4985079	0.4985079
		IMSE	**0.000725**	0.0007349	**0.0004585**	0.0005520	0.0006919	0.0006919
	5	$\hat{\theta}$	5.053385	5.052163	5.019285	5.013858	5.052812	5.052812
		S_{real}	0.500506		0.498914		0.500804	
		$\hat{S}$	0.498565	0.4986364	0.4981753	0.4989987	0.4995244	0.4995244
		IMSE	**0.000720**	0.0007236	**0.0007473**	0.0008258	0.0007639	0.0007639

Remarks:

WE note from Tables (1) following

1- The XG2D distribution is the best among the studied distributions, because it achieved the lowest integral error rate (IMSE) in the simulation experiments.
2- We clearly see that the MLE method for XG2D was better than the MOM method, although the MOM method was good at estimating.

VII. **Application**

Data were collected from patients infected by Coronavirus at Iraq - Najaf Al-Ashraf - Al Amal Hospital for Infectious Diseases, representing the times of survival (in days) until death or recovery due to Coronavirus infection, in February 2022 with a total of (61) patients, (4,3,3,7,5,7,1,6,2,3,22,1,16, 15,9,10,24,3,10,3,1,9,1,1,6,15,3,44,3,3,7,7,1,17,6,4,3,6,6,8,3,12,11,6,1,7,6,1,8,7,7,9,3,3,5,4,5,19,1,1,4), The stay times were rounded to the nearest integer number of days due to the lack of data on the hours of stay without staying overnight (a patient is not counted as a day as long as he did not spend the night).The following results were obtained by analyzing the data using X_c^2 statistics for good fit using the R language:

Table2. Results of the data fit test for the xgamma2 distribution

Dis.	df	X_c^2	X_t^2	α	Decision
Xgamma2	4	0.39344	9.49	0.05	Accept H_0

Remarks:

As can be seen from Table (2), X_c^2 is calculated as (0.39344) and is less than the value of X_t^2 tabular. As a result, the null hypothesis is accepted, meaning the real data are distributed according to xgamma2.

For the purpose of determining which distribution was best when applied to real data, xgamma2 distribution, xgamma distribution, and Exponential distribution were compared. AIC, AICc and BIC were used for model selection; the results are presented in Table (3):

Table 3. Results of the goodness of fit.

Distribution	**AIC**	**AICc**	**BIC**
Xgamma	361.9931	362.0609	364.104
Xgamma2	366.2431	366.3009	367.354
Exponential	366.6943	366.7542	368.7973

In Table 3, it can be seen that AIC, AICc and BIC values of XGD are smaller than the other distributions (xgamma2, Exponential), so the XGD is the best model, followed by XG2D and followed by EXP, as a result, the Xgamma2 distribution was a good model in Fit the real data used after the original XG model.

VIII. Conclusion

An exponential-gamma mixture, known as Xgamma2, was derived. The distribution was analyzed from various mathematical and structural perspectives. We derived and discussed important survival properties such as the hazard rate. An algorithm for simulation and stochastic ordering was also proposed. Xgamma2 distribution was observed to have added flexibility with regard to certain important properties. We proposed two methods for estimating parameters: maximum likelihood and Moments. An application of the Xgamma2 distribution was demonstrated through a simulation study. We also compared the distribution to the xgamma distribution and to the Exponential distribution using a real data set. According to the results, the Xgamma2 distribution is an adequate fit to the data set after XG distribution. Xgamma2 may be used in future applications under different types of censoring strategies.

IX. REFERENCES

Elshahhat, A., & Elemary, B. R. (2021). Analysis for Xgamma parameters of life under Type-II adaptive progressively hybrid censoring with applications in engineering and chemistry. *Symmetry*, *13*(11), 2112.

Gu, J. (2014). *Statistical partition problem for exponential populations and statistical surveillance of cancers in Louisiana* University of New Orleans].

Le Cam, L. (1990). Maximum likelihood: an introduction. *International Statistical Review/Revue Internationale de Statistique*, 153-171.

Okoli, O., Osuji, G., Nwosu, D., & Njoku, K. (2016). On the modified extended generalized exponential distribution. *European Journal of Statistics and Probability*, *4*(4), 1-11.

Sen, S., & Chandra, N. (2017). The quasi xgamma distribution with application in bladder cancer data. *Journal of data science*, *15*(1), 61-76.

Sen, S., Maiti, S. S., & Chandra, N. (2016). The xgamma distribution: statistical properties and application. *Journal of Modern Applied Statistical Methods*, *15*(1), 38.

Triana, Y., & Purwadi, J. (2019). Exponential Distribution Parameter Estimation with Bayesian SELF Method in Survival Analysis. Journal of Physics: Conference Series,