Estimating the survival function by Bayes and MLE methods for a proposed new form of Xgamma distribution

Mohammed H. AL- Sharoot1 and Karrar R. AL-Sayeab2

[1] Professor Doctor of Statistics at the Faculty of Administration and Economics of Al-Qadisiyah University, Iraq.
Email: mohammed.alsharoot@qu.edu.iq
[2] Master Student of the Faculty of Administration and Economics of Al-Qadisiyah University, Iraq.
Email: karrar.alsayeab@jmu.edu.iq

Abstract

In this research, we have studied a new form of the Xgamma (XG) distribution, resulting from mixing the two distributions exponential with parameter (λ=1/θ), and gamma with the two parameters (3, λ=1/θ), with the substitution of the mixing weights in the mixing equation for the distribution XG, one in place of the other, which produced a new form for which we proposed the name Xgamma3 distribution (XG3D). Then we derive some mathematical properties and some survival properties of the new proposed shape XG3D, We estimate the survival function of the proposed shape by estimating its parameter with two estimation methods (the Standard Bayesian and maximum likelihood estimation methods).

A simulation study was carried out to find out which of the two estimation methods is the best by comparing them by Integrated Mean square Error, which showed that the standard Bayes method is the best.

And using real data (for patients infected with the Covid-19 virus in the city of Najaf, Iraq), the proposed shape was compared with the original form of the XG distribution (some of its mathematical and survival properties were presented and its survival function was estimated using the same methods above) and the comparison was made using the criteria (AIC, AICc, BIC), which concluded that the original form of distribution was better than the proposed form because it achieved the lowest comparative criteria, but the proposed form remains a competitive model.

Keywords: Xgamma distribution, Xgamma3 distribution, Survival function, Standard Bayes estimation, Maximum Likelihood estimation.

INTRODUCTION

Since ancient times, human has been interested in trying to increase their chances of survival, and in recent years, interest has increased in the study of the survival function because it represents an important detail in knowing the period of human survival when he suffers from a disease. The study of survival extends from the beginning of the event (infection with the disease) until it ends with recovery or death; The survival function expression is used in studies that concern the health aspect of the patient, as the most important thing that is studied in the applications of statistical analysis is the health aspect of humans.

As a result, we developed a new form of the Xgamma probability distribution, which he suggested proposed by Subhradev Sen in 2016, derived from mixing two distributions with specific weights, which are the exponential distribution and the gamma distribution, followed by deriving its survival properties and mathematical properties. Then estimating its parameter using the Moments method and the Maximum Likelihood Estimation method. It used real data for patients receiving analgesics to calculate the probability of resting time for them for 20 patients, and it was found that the distribution was better than others in representing the data used when compared to another distribution [1]. In 2017, the same researcher, along with others, combined an additional parameter for the Xgamma distribution with one parameter and called it the quasi xgamma distribution as a generalization or extension of the xgamma distribution with one parameter. The different mathematical and statistical properties were studied, along with some important survival properties of the distribution. Estimating model parameters using the method of greatest possibility and method of moments. The distribution was compared with some other distributions by applying to an unsupervised data set of resting times (in months) for a random sample of 128 patients with bladder cancer and it appeared that the generalized model of the Xgamma distribution provides a better fit than the other models taking into account the real data set [2]. In the year 2018, the researcher (S.Sen) generalized the Xgamma distribution by adding a parameter, deriving the two-parameter Xgamma distribution, which added additional flexibility in modeling real-life data. To compare methods in the estimation process. Two sets of data were analyzed, they studied the failure times of 18 electronic devices and a data set of the age of 50 devices, and compared the proposed model with some other two-parameter lifetime models [3]. In 2021 (Elshahhat, Ahmed, and Berihan R. Elemary) studied the one-parameter xgamma distribution. Then the unknown parameter and some survival properties of the xgamma model were estimated based on the controlled samples. A simulation study was conducted to compare the different estimates. They showed that the Bayes estimates are better than the classical estimates. Two data sets related to engineering and chemical experiments were analyzed (a data set of failure times for a sample of 25 pieces of yarn tested at a certain stress level with a length of (100 cm) and a data set for 34 BAL observations (mg / l).) for vinyl chloride obtained from clean gradient control wells) [4].

1. Xgamma distribution:

In 2016 proposed by Sen et al., a particular hodgepodge of EXP (θ) & gamma (3,θ) distributions, indicated XGD, and the mixing's two parameters were assumed as that $\pi_1 = \frac{\theta}{1+\theta}$, $\pi_2 = 1-\pi_1 = \frac{1}{1+\theta}$. According to him, he had derived the survival function and inferred the distribution of the XGD [4].

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

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

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

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

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 as 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 as follow:

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

$$\mu_1' = \frac{(\theta+3)}{\theta(1+\theta)} \tag{6}$$

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

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

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

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

2. Xgamma3 distribution (XG3D):

The new formula we propose in this section is referred to as the Xgamma3 distribution and denoted as XG3D. We call it the new formula the Xgamma3 distribution. We suppose we have exponential (θ) and gamma (3,λ=1/θ) distributions, and by making the first weight the place of the second and versa in the mixing equation of the XG distribution, we get the new form of the following mixing equation:

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

Where

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

$$f_1(t) = \frac{1}{\theta} e^{-\frac{t}{\theta}} \quad and \quad f_2(\text{t}) = \frac{t^2}{\theta^3\Gamma 3} e^{-\frac{t}{\theta}} \qquad \text{then:}$$

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

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

The corresponding cumulative distribution function (C.D.F) of the XG3D is:

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

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

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

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

The rth moments of x about zero is:

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

The mean and the variance and the other properties are defined as follows:

$$\mu_1^{'} = \frac{\theta(1 + 3\theta)}{(1+\theta)} = Mean(t) \quad (16)$$

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

$$Mode = \begin{cases} -\theta + \theta\sqrt{1 + 2\theta}, & 0 < \theta < \infty \\ 0, & otherwise \end{cases} \quad (18)$$

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

$$\Upsilon_2 = \frac{3\theta^4(75\theta^4 + 218\theta^3 + 216\theta^2 + 94\theta + 5)}{\left(\theta^2 + 8\theta + 3\right)^2} \quad (20)$$

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

4. Estimation of the Parameter [5]

4.1 The maximum Likelihood Estimation:

Statistical estimation often uses maximum likelihood methods. In terms of who came up with the method first, there are differing opinions. Maximum likelihood was invented by Fisher, but its properties of optimality were demonstrated through its wide application [6].

It is calculating it in the XG3D as follows:

We Assume that (T1, T2, ..., Tn) is a random sample of size n from the XG2D,

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

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

Let $l(\theta;t) = \frac{\partial \ln L}{\partial \theta} = 0$

$$\sum_{i=1}^{n} \frac{-t_i^2}{\theta\left(2\theta + t_i^2\right)} - \frac{n}{1+\theta} - \frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} t_i = 0 \quad (22)$$

The non-linear equation (22) cannot be solved analytically; the numerical method (Newton-Raphson) is used .

Let the initial solution is:

Then the survival function and the hazard rate function by MLE:

$$s(t) = e^{\frac{-t}{\hat{\theta}_{MLE}}} \left(1 + \frac{\left(2\hat{\theta}_{MLE} + t\right)t}{2\hat{\theta}_{MLE}(1+\hat{\theta}_{MLE})}\right) \quad (23)$$

$$h(t) = \frac{(2 + \frac{t^2}{\hat{\theta}_{MLE}})}{2\hat{\theta}_{MLE}(1+\hat{\theta}_{MLE}) + \left(2\hat{\theta}_{MLE} + t\right)t} \quad (24)$$

4.2 Standard Bayesian Estimation method:

The parameter XG3D will be estimated using the standard Bayesian method in this section. We suppose the gamma distribution is the prior distribution for the parameter θ is defined as that:

$$p(\theta) = \frac{\theta^{\alpha-1}}{\beta^{\alpha}\Gamma\alpha} e^{-\frac{\theta}{\beta}}; \quad \alpha, \beta > 0$$

And let a squared error loss function, then we can compute the $\widehat{\theta}_{Bayes}$ as follows:

$$f\ (\theta|t_1,t_2,...,t_n) = \frac{L(t_1,t_2,...,t_n|\theta)\,\mathrm{P}(\theta)}{\int_0^{\infty} L(t_1,t_2,...,t_n|\theta)\,\mathrm{P}(\theta)d\,\theta}$$

$$f\ (\theta\ /\ t_1,t_2,...,t_n) = \frac{\dfrac{\theta^{\alpha-1-n}}{(1+\theta)^n} e^{-\frac{1}{\theta}\left(\frac{\theta^2}{\beta}+\sum_{i=1}^{n} t_i\right)} \prod_{i=1}^{n}\left(1+\dfrac{t_i^2}{2\theta}\right)}{\int_0^{\infty} \dfrac{\theta^{\alpha-1-n}}{(1+\theta)^n} e^{-\frac{1}{\theta}\left(\frac{\theta^2}{\beta}+\sum_{i=1}^{n} t_i\right)} \prod_{i=1}^{n}\left(1+\dfrac{t_i^2}{2\theta}\right) d\,\theta} \tag{25}$$

$$Risk\ = E\,(\widehat{\theta}-\theta)^2$$

$$\frac{\partial Risk}{\partial \theta} = 0$$

$$\widehat{\theta}_{Bayes} = E\,(\theta|T\,)$$

$$\hat{\theta}_{Bayes} = \frac{\dfrac{\theta^{\alpha-n}}{(1+\theta)^n} e^{-\frac{1}{\theta}\left(\frac{\theta^2}{\beta}+\sum_{i=1}^{n} t_i\right)} \prod_{i=1}^{n}\left(1+\dfrac{t_i^2}{2\theta}\right)}{\int_0^{\infty} \dfrac{\theta^{\alpha-1-n}}{(1+\theta)^n} e^{-\frac{1}{\theta}\left(\frac{\theta^2}{\beta}+\sum_{i=1}^{n} t_i\right)} \prod_{i=1}^{n}\left(1+\dfrac{t_i^2}{2\theta}\right) d\,\theta} \tag{26}$$

we will use Lindley's approximation method to reckon the integral for getting the Bayes estimator to the parameter θ of XG3D because the equation (26) isn't a closed formulation, thus cannot compute it numerically.

And the survival function and the hazard rate function for XG3D by Bayes as follows:

$$s(t) = e^{\frac{-t}{\hat{\theta}_{Bayes}}}\left(1+\frac{\left(2\hat{\theta}_{Bayes}+t\right)t}{2\hat{\theta}_{Bayes}(1+\hat{\theta}_{Bayes})}\right) \tag{27}$$

$$h(t) = \frac{(2+\dfrac{t^2}{\hat{\theta}_{Bayes}})}{2\hat{\theta}_{Bayes}(1+\hat{\theta}_{Bayes})+\left(2\hat{\theta}_{Bayes}+t\right)t} \tag{28}$$

5. Simulation study

A Monte Carlo simulation was conducted with M = 1000 iterations to examine the behavior of selected estimators and survival functions of XG3D. Based on the assumed sample size (20, 80) and assume () is (0.5,1,5), and by IMSE measures, we calculate the following:

$$IMSE\left(\hat{S}(t)\right)=\frac{1}{M}\sum_{i=1}^{M}\left[\frac{1}{n_i}\sum_{j=1}^{n_i}\left(\hat{S}_i(t_j)-S_i(t_j)\right)^2\right] \qquad (29)$$

5.1 generation of random data

5.1.1 For generation of random data based on the xgamma, see [1].

5.1.2 We used the below algorithm to generate a random sample T_j, j = 1, 2,..., n from XG3D.

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

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

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

d. Generate $T_j \sim Xgamma3(\theta), j= 1,\ldots,n$ 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		XG3	
			mle	Bay	mle	Bay
20	0.5	$\hat{\theta}$	0.5117965	0.5112623	0.52693	0.5075526
		S_{real}	0.5002135		0.4722976	
		$\hat{S}$	0.4955356	0.4957594	0.4905018	0.4774799
		IMSE	0.00386554	0.00370174	0.000391633	0.000317844
	1	$\hat{\theta}$	1.018832	0.9927622	1.033494	1.00763
		S_{real}	0.4967301		0.5000317	
		$\hat{S}$	0.4944526	0.5035647	0.5118085	0.5026429
		IMSE	0.00347596	0.003170633	0.000810366	0.000668730
	5	$\hat{\theta}$	5.244611	3.297329	5.282231	5.245344
		S_{real}	0.5001982		0.4782431	
		$\hat{S}$	0.4934614	0.6227599	0.4979265	0.4950749
		IMSE	0.00426099	0.01824733	0.005778552	0.005371075
80	0.5	$\hat{\theta}$	0.5026795	0.5026452	0.4770932	0.481711
		S_{real}	0.4996549		0.480159	
		$\hat{S}$	0.4985668	0.4985835	0.4633047	0.466753
		IMSE	0.00094456	0.000934995	0.00329888	0.002087301
	1	$\hat{\theta}$	1.004509	0.9985339	0.9891171	0.9955127
		S_{real}	0.4988837		0.5061915	
		$\hat{S}$	0.4985176	0.5006833	0.5018582	0.5041941

		IMSE	0.00092917	0.000910025	0.000442676	0.0004224679
	5	$\hat{\theta}$	5.055853	4.633148	4.981749	4.975727
		S_{real}	0.4997219		0.4942596	
		$\hat{S}$	0.4980056	0.5232483	0.4921971	0.4916633
		IMSE	0.00092418	0.001377478	0.000206111	0.000206397

Remarks:

From Table (1), we note that

The Bayes method is the best for XG3D, and the MLE method was kind as well at estimating for XG3D.

6. Application

From Iraq - Najaf Al-Ashraf - Al Amal Hospital, we possessed data on the times of survival (in days) to death or recovery for (40) patients who were stomachic by Coronavirus in March 2022. (15, 8, 13, 9, 6, 1, 2, 7, 1, 8, 1, 3, 5, 9, 24, 14, 20, 6, 2, 1, 44, 2, 18, 14, 7, 10, 9, 7, 10, 35, 5, 3, 10, 4, 10, 53, 10, 10, 4,6) and because of the insufficiency of data on hours of stay without staying nightly, the stay times were rounded to the nearest number.

In using the R language, we analyzed the data for good fit using X_c^2 test and obtained the following Outcome:

Table2. The outcome of the data fit test for XG3D

Dis.	df	X_c^2	X_t^2	α	Decision
XG3D	4	0.5	9.49	0.05	Accept H_0

Remarks:

The null hypothesis is agreeable because the X_t^2 tabular is greater than the value calculated shown in Table (2). So the real data distributed according to XG3D.

We use the AIC, AICc, and BIC to choose the best distribution between XG and XG3D when applied to real data. The Outcomes are shown in Table (3):

Table 3. The outcome for choosing the best distribution.

Distribution	**AIC**	**AICc**	**BIC**
XG	406.4321	406.4999	408.5429
XG3D	279.0095	279.1147	280.6983

The XG3D is the best to fit the real data used because the values of AIC, AICc, and BIC for XG3D are smaller than the values for XG, shown in Table.

Conclusion

A new form of the Xgamma distribution has been derived and the name Xgamma3 has been proposed for it. The mathematical and survival characteristics were studied and his parameter was estimated using MLE & Bayes methods to estimate his survival function by estimating his parameter with these two methods. Bayes estimation method was better than MLE, after comparing them using simulation. A comparison was made between the original form of distribution and the proposed form in dealing with real data through comparison standards AIC, AICc, and BIC It turned out that the proposed form was better at dealing with the used data. Therefore, we recommend increasing experiments on it and proving its efficiency to be used in the future instead of the original form.

REFERENCES

1. Sen, S., S.S. Maiti, and N. Chandra, *The xgamma distribution: statistical properties and application.* Journal of Modern Applied Statistical Methods, 2016. **15**(1): p. 38.
2. Sen, S. and N. Chandra, *The quasi xgamma distribution with application in bladder cancer data.* Journal of data science, 2017. **15**(1): p. 61-76.
3. Sen, S., N. Chandra, and S.S. Maiti, *On Properties and Applications of a Two-Parameter Xgamma Distribution.* J. Stat. Theory Appl., 2018. **17**(4): p. 674-685.
4. Elshahhat, A. and B.R. Elemary, *Analysis for Xgamma parameters of life under Type-II adaptive progressively hybrid censoring with applications in engineering and chemistry.* Symmetry, 2021. **13**(11): p. 2112.
5. Casella, G. and R.L. Berger, *Statistical inference*. 2021: Cengage Learning.
6. Le Cam, L., *Maximum likelihood: an introduction.* International Statistical Review/Revue Internationale de Statistique, 1990: p. 153-171.