International Journal of Research and Innovation in Applied Science (IJRIAS) | Volume IV, Issue XII, December 2019|ISSN 2454-6194

Exponential-Gamma Distribution and Its Applications

Ayeni Taiwo Michael, Ogunwale Olukunle Daniel, Adewusi Oluwasesan Adeoye
Department of Statistics, Ekiti State University, Nigeria

Abstract: - **Researchers in medical sciences have applied different statistical methodologies in modeling diverse prevailing diseases. In this study we apply the new Exponential-Gamma distribution in modeling patients with remission of Bladder Cancer and survival time of Guinea pigs infected with tubercle bacilli. We compared the performance of the results with other existing statistical distributions. The result shows that the Exponential-Gamma distribution performed better than the existing distributions.**

Keywords: **Exponential-Gamma distribution, Bladder Cancer, Akaike Information Criterion, Bayesian Information Criterion.**

I. INTRODUCTION

Bladder cancer is the ninth most frequently diagnosed malignancy world-wide [1] and one of the most prevalent, representing 3% of cancers diagnosed globally [2]. Bladder cancer accounts for an estimated 386,000 new diagnoses and 150,000 related deaths annually [1]. Early detection of bladder cancer remains one of the most urgent issues in many researches [3]. Although, there are some improvements in imaging and surgical techniques, the overall mortality of patients with bladder cancers has not been unchanged [4] and outcomes for patients remain suboptimal [2]. Currently, morphologic and pathologic criteria such as histology, stage, and grade are used for conventional diagnosis of bladder cancer, which play an important role in determining treatment [5]. However, even though essential prognostic information is provided by these clinical criteria, they have inadequate power to predict patient outcome precisely [5] and there remains significant variability in the prognosis of patients with similar characteristics [2]. Therefore, Reister*et al.,* in [2] examined additional tumor characteristics that predict clinical behavior is highlighted in patients with bladder cancer.

The high susceptibility of guinea pigs to human tuberculosis is well recognized. Even an infection initiated with a few virulent tubercle bacilli will lead to progressive disease and death, in spite of the fact that guinea pigs do acquire resistance, as first demonstrated by Koch nearly 70 years ago. The nature of the mechanisms constituting native and acquired resistance to tuberculosis is still obscure; the effect on the tubercle bacilli of these mechanisms, however, is generally accepted as inhibiting multiplication and restricting dissemination, though in guinea pigs the infection is rarely arrested completely [13],[14],[15],[16],[17]and [18].

In [12], the study reveals that the first animal to die in each regimen is assumed to have the least resistance; the last to die, the most resistance. Within each study, animals with the same order of survival in the different regimens are considered equal with regard to their inherent, constitutional capacities to cope with a tuberculous infection.

Statistical distributions play a crucial role in parametric inferences and applications to describe real world phenomena. Recently, studies has revealed that some real life data that cannot be modeled sufficiently by existing standard distributions are occasionally discovered to follow distributions of some combined form of two or more random variables with known probability distributions. In view of their flexibility and variety in function and performance, statistical distributions are of particular interest to many researchers such as; [6],[7] and [9].

Therefore this study aimed to investigate the performance of the new Exponential-Gamma distribution to other exiting probability distributions using the data on remission time of patients with Bladder Cancer and survival time of Guinea pigs using the model selection criteria like the Akaike information criterion (AIC), Bayesian information criterion (BIC) and the log likelihood function (l).

II. METHODS

The Exponential-Gamma distribution was developed by [10] and its pdf is defined as

$$f(x;\alpha,\lambda)=\frac{\lambda^{\alpha+1}x^{\alpha-1}e^{-2\lambda x}}{\Gamma(\alpha)},x,\lambda,\alpha>0 \qquad (1)$$

With the mean and variance of the form;

$$\mu=\frac{\alpha}{2^{\alpha+1}} \qquad (2)$$

and $$V(x)=\frac{\alpha\left(\alpha 2^{\alpha}-\lambda\alpha+2^{\alpha}\right)}{\lambda\left(2^{2(\alpha+1)}\right)} \qquad (3)$$

The cumulative distribution function is defined as

$$F(x)=\frac{\lambda\gamma(\alpha,x)}{2^{\alpha}\Gamma(\alpha)} \qquad (4)$$

The survival function for the distribution defined by $S(x)=1-F(x)$ was obtained as;

$$S(x)=1-\frac{\lambda\gamma(\alpha,x)}{2^{\alpha}\Gamma(\alpha)} \qquad (5)$$

While the corresponding hazard function defined by $h(x)=\dfrac{f(x)}{S(x)}$ was obtained as;

$$h(x)=\frac{\lambda^{\alpha+1}x^{\alpha-1}e^{-2\lambda x}2^{\alpha}}{2^{\alpha}\Gamma(\alpha)-\lambda\gamma(\alpha,x)} \tag{7}$$

The cumulative hazard function for distribution defined by $H(x)=\int_0^x h(x)dx$ was obtained as;

$$H(x)=\frac{\lambda\gamma(\alpha,x)}{2^{\alpha}\Gamma(\alpha)-\lambda\gamma(\alpha,x)} \tag{8}$$

A. Maximum Likelihood Estimator

Let $X_1, X_2, ..., X_n$ be a random sample of size n from Exponential-Gamma distribution. Then the likelihood function is given by;

$$L(\alpha,\lambda;x)=\left(\frac{\lambda^{\alpha+1}}{\Gamma(\alpha)}\right)^{n}\prod_{i=1}^{n}x_i^{\alpha-1}\exp\left(-2\lambda\sum_{i=1}^{n}x_i\right) \tag{9}$$

by taking logarithm of (9), we find the log likelihood function as;

$$\log(L)=\alpha n\log\lambda+n\log\lambda-n\log\Gamma(\alpha)+(\alpha+1)\sum_{i=1}^{n}\log x_i-2\lambda\sum_{i=1}^{n}x_i \tag{10}$$

Therefore, the MLE which maximizes (10) must satisfy the following normal equations;

$$\frac{\partial\log L}{\partial\alpha}=n\log\lambda-\frac{n\Gamma^{'}(\alpha)}{\Gamma(\alpha)}+\sum_{i=1}^{n}\log x_i \tag{11}$$

$$\frac{\partial\log L}{\partial\lambda}=\frac{\alpha n}{\lambda}+\frac{n}{\lambda}-2\sum_{i=1}^{n}x_i \tag{12}$$

The solution of the non-linear system of equations is obtained by differentiating (10) with respect to (α,λ) gives the maximum likelihood estimates of the model parameters. The estimates of the parameters can be obtained by solving (11) and (12) simultaneously as it cannot be done analytically, therefore a numerical technique will be adopted. The solution can also be obtained directly by using python software using the data sets.

In this study, we applied the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and the log likelihood function (l) to compare the new developed Exponential-Gamma distribution with the existing probability distributions such as the Exponential and the Gamma distributions.

The AIC is defined by;

$$AIC=2k-2In(\hat{\ell}) \tag{13}$$

where$k$ is the number of the estimated parameter in the model

$\hat{\ell}$ is the maximized value of the model.

The BIC is defined by;

$$In(n)k-2In(\hat{\ell}) \tag{14}$$

where k is the number of the estimated parameter in the model

n is the number of observations

$\hat{\ell}$ is the maximized value of the model.

The approach in (13) and (14) above is used when comparing the performance of different distributions to determine the best fit model. To select the appropriate model by considering the number parameters and maximum likelihood function; the AIC, BIC and likelihood function are examined; consequently an acceptable model has smaller AIC and BIC value while the log likelihood value is expected to be greater. The Python software was used for the comparison of the performance of the Exponential-Gamma, Exponential and Gamma distributions.

III. RESULTS

The analysis of the different data set was carried out by using python

The first data set used presents the remission times (in months) of a random sample of 128 Bladder Cancer patients. It was previously used by Lee and Wang in [19] and Oguntunde *et al.,*in [20].

The data is as follows:

0.08, 2.09, 3.48, 4.87, 6.94, 8.66, 13.11, 23.63, 0.20, 2.23, 3.52, 4.98,6.97, 9.02, 13.29, 0.40, 2.26, 3.57, 5.06, 7.09, 9.22, 13.80, 25.74, 0.50,2.46, 3.64, 5.09, 7.26, 9.47, 14.24, 25.82, 0.51, 2.54, 3.70, 5.17, 7.28,9.74, 14.76, 26.31, 0.81, 2.62, 3.82, 5.32, 7.32, 10.06, 14.77, 32.15, 2.64,11.79, 18.10, 1.46, 4.40, 5.85, 8.26, 11.98, 19.13, 1.76, 3.25, 4.50, 6.25,8.37, 12.02, 2.02, 3.31, 4.51, 6.54, 8.53, 12.03, 20.28, 2.02, 3.36, 6.76,12.07, 21.73, 2.0, 3.36, 6.93, 8.65, 12.63, 22.69.3.88, 5.32, 7.39, 10.34, 14.83, 34.26, 0.90, 2.69, 4.18, 5.34, 7.59, 10.66,15.96, 36.66, 1.05, 2.69, 4.23, 5.41, 7.62, 10.75, 16.62, 43.01, 1.19, 2.754.26, 5.41, 7.63, 17.12, 46.12, 1.26, 2.83, 4.33, 5.49, 7.66, 11.25, 17.14,79.05, 1.35, 2.87, 5.62, 7.87, 11.64, 17.36, 1.40, 3.02, 4.34, 5.71, 7.93.

Table 1: Summary of data (bladder cancer)

Parameters	Values
n	128
min	0.080
max	79.050
mean	9.365
variance	110.45
skewness	3.287
kurtosis	15.483

From the above table it was observed that the distribution of the data is skewed to the right with skewness 3.287. This shows that Exponential-Gamma has the ability to fit a right skewed data.

Table 2: Estimates and performance of the distributions (bladder cancer)

Distribution	Parameters	(l)	AIC	BIC
Exponential-Gamma	$\hat{\alpha} = \mathbf{0.830}$ $\hat{\lambda} = \mathbf{0.006}$	**-27.94**	**62.0667**	**70.4292**
Exponential	$\hat{\lambda} = 9.286$	-413.24	830.5837	836.1917
Gamma	$\hat{\alpha} = 0.804$ $\hat{\lambda} = 0.080$	-859.33	1724.852	1733.214

The estimates of the parameters, log-likelihood, Akaike information criterion (AIC) and Bayesian information criterion (BIC) for the data on remission times of a random sample of 128 bladder Cancer patients is presented in Table 2. It was obvious that Exponential-Gamma provides a better fit as compared to Exponential and Gamma distributions since it has highest value of log-likelihood (l) and the lowest value of Akaike information criterion (AIC) and Bayesian information criterion (BIC). Hence, the Exponential-Gamma distribution performed better than other distribution compared.

The second data set represents the survival times (in days) of 72 guinea pigs infected with virulent tubercle bacilli reported by Bjerkedal in [12] and used by Tahir*et al.*, [11]. The data is as follows:

10, 33, 44, 56, 59, 72, 74, 77, 92, 93, 96, 100, 100, 102, 105, 107, 107, 108, 108, 108, 109, 112, 113, 115, 116,120, 121, 122, 122, 124, 130, 134, 136, 139, 144, 146, 153, 159, 160, 163, 163, 168, 171, 172, 176,183, 195, 196, 197, 202, 213, 215, 216, 222, 230,231, 240, 245, 251, 253, 254, 255, 278, 293, 327,342, 347, 361, 402, 432, 458, 555.

Table 3: Summary of data (survival times of guinea pig)

Parameters	Value
n	72
min	10
max	555
mean	176.833
variance	10556.417
skewness	1.341284
kurtosis	1.988524

From the above table it was observed that the distribution of the data is skewed to the right with skewness 1.341284. This shows that Exponential-Gamma has the ability to fit a right skewed data.

Table 4: Estimates and performance of the distributions (survival times of guinea pig)

Distribution	Parameters	(l)	AIC	BIC
Exponential-Gamma	$\hat{\alpha} = \mathbf{0.336}$ $\hat{\alpha} = \mathbf{1.000}$	**-11.36**	**28.7183**	**35.54826**
Exponential	$\hat{\lambda} = 1.000$	-440.42	884.8473	889.4007
Gamma	$\hat{\alpha} = 0.336$ $\hat{\lambda} = 1.2$	-2546.23	5098.470	5105.3

The estimates of the parameters, log-likelihood, Akaike information criterion (AIC) and Bayesian information criterion (BIC) for the data set on the survival times (in days) of 72 guinea pigs infected with virulent tubercle bacilli is presented in Table 4. It was obvious that Exponential-Gamma provides a better fit as compared to Exponential and Gamma distributions since it has highest value of log-likelihood (l) and the lowest value of Akaike information criterion (AIC) and Bayesian information criterion (BIC). Hence, the Exponential-Gamma distribution performed better than other distribution compared.

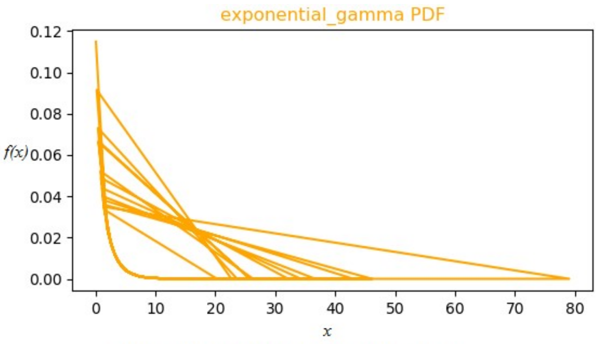

figure 1: EGD pdf plot for bladder cancer date set

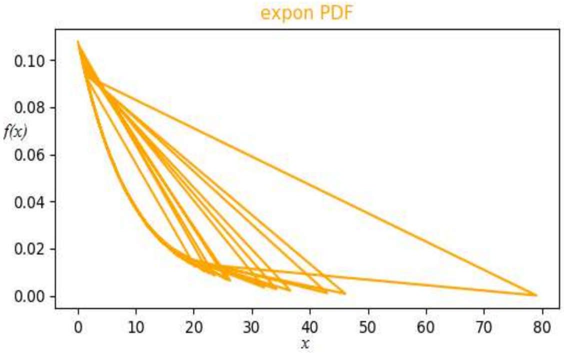

figure 2: ED pdf plot for bladder cancer data set

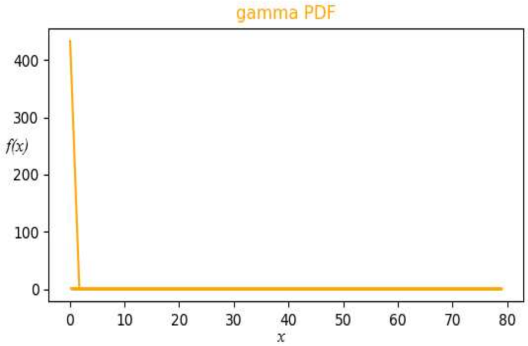

figure 3: GD pdf plot for bladder cancer data set

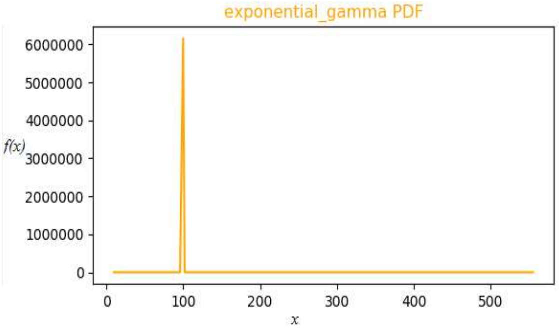

figure 4: EGD pdf plot for survival time of guniea pig data set

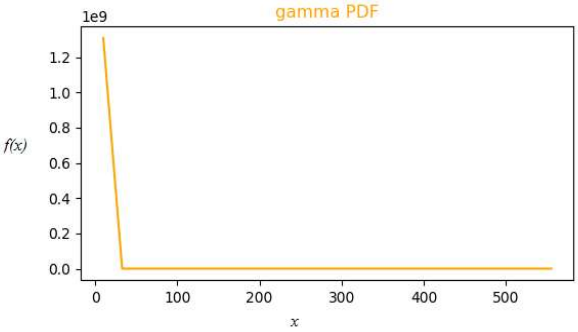

figure 5: GD pdf plot for survival time of guniea pig data set

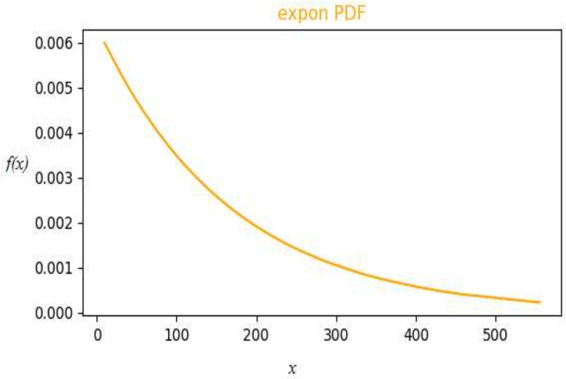

figure 6: ED pdf plot for survival time of guniea pig data set

IV. CONCLUSION

Statistical distributions play a crucial role in parametric inferences and applications to fit real world phenomena. In this study we apply the new Exponential-Gamma distribution in modeling patients with remission of Bladder Cancer and survival time of Guinea pigs infected with tubercle bacilli. We compared the performance of the results with other existing statistical distributions. The result shows that the Exponential-Gamma distribution performed better than the existing distributions.

REFERENCES

[1]. Gurung, P., Veerakumarasivam, A., Williamson, M., Counsell, N., Douglas, J., Tan, W.S., Feber, A., Crabb, S.J., Short S.C., Freeman, A., *Loss of expression of the tumour suppressor gene AIMP3 predicts survival following radiotherapy in muscle-invasive bladder cancer.*Int J Cancer, **136**(3), pp 709-720, Feb. 2015.

[2]. Riester, M., Taylor, J.M., Feifer, A., Koppie, T., Rosenberg, J.E., Downey, R.J., Bochner, B.H., Michor, F.,*Combination of a novel gene expression signature with a clinical nomogram improves the prediction of survival in high-risk bladder cancer*. Clin Cancer Res, **18**(5), pp1323-1333. Jan. 2012

[3]. Rosser, C.J., Liu, L., Sun, Y., Villicana, P., McCullers, M., Porvasnik, S., Young, P.R., Parker, A.S., Goodison, S.,*Bladder Cancer–Associated Gene Expression Signatures Identified by Profiling of Exfoliated Urothelia.*Cancer Epidemiol Biomarkers Prev., **18**(2), pp444-453. 2009

[4]. Meeks, J.J., Bellmunt, J., Bochner, B.H., Clarke, N.W., Daneshmand, S., Galsky, M.D., Hahn, N.M., Lerner, S.P., Mason, M., Powles, T.,*A systematic review of neoadjuvant and adjuvant chemotherapy for muscle-invasive bladder cancer.*EurUrol, **62**(3) pp523-533,Sep. 2012.

[5]. Blaveri, E., Simko,JP.,Korkola, J.E, Brewer, J.L., Baehner, F., Mehta, K., DeVries, S., Koppie, T., Pejavar,S., Carroll, P., *Bladder cancer outcome and subtype classification by gene expression.*Clin Cancer Res, **11**(11) pp4044-4055.June 2005.

[6]. Baghestani, A.R, Gohari, M.R., Orooji, A., Pourhoseingholi, M.A., Zali, M.R.,*Evaluation of parametric models by the prediction error in colorectal cancer survival analysis.*GastroenterolHepatol Bed Bench. **8**(3) pp183–7. 2015.

[7]. Baghestani, A.R., Hajizadeh, E., Fatemi, S.R.,*Parametric model to analyse the survival of gastric cancer in the presence of interval censoring.*Tumori.**96**(3), pp433–7 June. 2010.

[8]. Moein, Y., Ahmad, R.B., Naghmeh, K., Mohamad, P.,Alireza, A., and Azin, K, *Survival Analysis of Colorectal Cancer Patients Using ExponentiatedWeibull Distribution*Int J Cancer Manag. **11**(3). pp 1-6. Feb. 2018.

[9]. Ogunwale, O.D., Adewusi, O.A. and Ayeni, T.M. (2019), *Exponential-Gamma Distribution;* international journal of emerging technology and advanced engineering; **9**(10); pp245-249 ISSN:2250-2459

[10]. Tahir, M., Cordeiro, G., Alzaatreh, A., Mansoor, M. and Zubair, M. *The Logistic-X Family of Distributions and Its Applications* Communication in Statistics- Theory and Methods **45**(24), pp 1-26. Oct. 2014

[11]. Bjerkedal, T., Acquisition *of resistance in guinea pigs infected with different doses of virulent tubercle bacilli*. American Journal of Hygiene **72**:pp130–148. 1960

[12]. Canetti, G. *The Tubercle Bacillus in the Pulmonary Lesion of Man*. New York:Springer Publishing Company, Inc., 1955. Chapter 4, pp. 127-150.

[13]. Hirsch, J. G. *Immunity to infectious diseases: review of some concepts of Metchnikoff.*Bact. Rev., **23**(2),pp48-60. Jun. 1959.

[14]. Pinner, M*., Pulmonary Tuberculosis in the Adult; its Fundamental Aspects.* Springfield,111.: Charles C Thomas, Publisher, Chapter 3, pp. 90-151.1945

[15]. Raffel, S.,*Immunity, Hypersensitivity, Serology.* New York: Appleton-Century- Crofts, Inc., 1953. Chapter 23, pp.321-353.

[16]. Rich, A. R., *The Pathogenesis of Tuberculosis.*Second Edition. Springfield, Illinois: Charles C Thomas, Publisher,1951. Chapters 13, 15 and 17, pp571-613 and 667-675.

[17]. Seibert, F. B. A theory of immunity in tuberculosis. Perspectives in Biol. And Med., Winter, **3;** pp264-281. 1960.

[18]. Lee, E.T. and Wang, J.W., *Statistical Methods for Survival Data Analysis.* 3rd Edn.,JohnWiley and Sons, New York, ISBN: 9780471458555, pp 534, 2003.

[19]. Oguntunde, P.E., Owoloko, E.A. and Balogun, O.S.,*On A New Weighted Exponential Distribution: Theory and Application;* Asian Journal of Applied Sciences **9** (1): 1-12, Dec. 2015.