Modeling survival times of the COVID-19 patients with a new distribution

A. Khosravi Tanak

Department of Statistics, Velayat University, Iranshahr, Iran

a.khosravi@velayat.ac.ir

***Abstract*— Over the past months, the spread of Coronavirus disease 2019 (COVID-19) epidemic has caused tremendous damage worldwide, and unstable many countries economically. Many studies have shown the potential of statistical distributions in modeling data in applied sciences, especially in medical science. In this paper, we propose a new continuous distribution and use it for modeling survival times of COVID-19 patients. We obtain the model parameters estimators using the maximum likelihood method. Finally, we demonstrate the effectiveness and utility of the proposed model by modeling COVID-19 patients data. The characteristics of the model in fitting to data are compared with the beta generalized Weibull family.**

Keywords—survival time, lifetime distribution, beta generalized Weibull distribution, hazard rate, maximum likelihood estimation.

I. Introduction

The modeling and analysis of lifetimes data play an important role in a wide variety of scientific and technological fields, especially medical science. In the last decades, a considerable amount of research was devoted to the creation of models that provide an adequate fit to real lifetime data, among them the Weibull distribution extensions have attracted a lot of attention (for example see [1, 2, 3, 4, 5]). Khosravi et al. [6] introduce a distribution for modeling income distribution using the maximum entropy principle. The probability distribution function of the proposed model is

$$F(x) = 1 - \left[1 + \alpha(\mathrm{e}^{\beta x} - 1)\right]^{-\frac{1}{\lambda-1}}, x > 0, \tag{1}$$

where α, β and λ are positive-valued parameters. In this paper, we consider a re-parameterized exponentiated form of (1) as

$$G(x) = \left(1 - \left[1 + \alpha\left(\mathrm{e}^{\beta x} - 1\right)\right]^{-\lambda}\right)^{\theta}, x > 0, \alpha, \beta, \lambda, \theta > 0. \tag{2}$$

We shall refer to the distribution given in (1) as the Exponentiated Life Time (ELT) distribution. If a random variable X has the ELT distribution, then we write $X \sim \mathrm{ELT}(\alpha, \beta, \lambda, \theta)$.

This paper is organized as follows. In Section 2, we present some properties of the ELT distribution and its failure rate function. Estimation of the parameters of ELT distribution by the method of maximum likelihood are studied in Section 3. In Section 4, we illustrate the potential of ELT distribution compared with other distributions by two real data applications of COVID-19 data.

II. Probability density function and failure rate

According to (2), the probability density function (pdf) of the ELT distribution takes the form

$$g(x) = \alpha\beta\lambda\theta e^{\beta x}\left[1 + \alpha\left(e^{\beta x} - 1\right)\right]^{-\lambda-1}\left(1 - \left[1 + \alpha\left(e^{\beta x} - 1\right)\right]^{-\lambda}\right)^{\theta-1}, \ x > 0, \tag{3}$$

where the parameters α, λ and θ are positive shape parameters of the distribution and $\beta > 0$ is the scale parameter. Some possible behaviors of the pdf of the ELT distribution for $\beta = 2$ are shown in Fig. 1. The plots are sketched for $\alpha = 0.8, \lambda = 0.9, \theta = 1.8$ (red-line), $\alpha = 0.02, \lambda = 1.2, \theta = 1.6$ (green-line), $\alpha = 0.1, \lambda = 0.5, \theta = 0.6$ (black-line), $\alpha = 1.8, \lambda = 0.1, \theta = 0.5$ (blue-line). The inverse of the ELT distribution function yields a quantile function given by

$$Q(u) = \frac{1}{\beta}\log\left\{\frac{1}{\alpha}\left[\left(1 - u^{\frac{1}{\theta}}\right)^{-\frac{1}{\lambda}} - 1\right] + 1\right\}, \ u \in (0, 1). \tag{4}$$

The above function facilitates ready quantile-based statistical modeling [7]. In addition, $Q(u)$ gives a trivial random variable generation: if $U \sim \mathcal{U}(0, 1)$, then

$$X = \frac{1}{\beta}\log\left\{\frac{1}{\alpha}\left[\left(1 - U^{\frac{1}{\theta}}\right)^{-\frac{1}{\lambda}} - 1\right] + 1\right\},$$

follows ELT distribution with parameters α, β, λ and θ. The median of ELT distribution can be derived from (4) by setting $u = 0.5$.

hazard rate function plays a key role in Applied Probability [8] and in the Theory of Reliability [9]. The hazard rate function of a random variable X is defined as

$$h(x) = \frac{f(x)}{\bar{F}(x)}.$$

Considering the random variable X as a lifetime random variable, the failure rate $h(x)$ represents the likelihood that X be realized right after time x, given that it was not realized up to time x. Some possible behaviors of the hazard rate function of the ELT distribution for $\beta = 2$ are shown in Fig. 2. The plots are sketched for $\alpha = 0.8, \lambda = 0.9, \theta = 1.8$ (red-line), $\alpha = 1.5, \lambda = 0.6, \theta = 0.6$ (green-line) and $\alpha = 0.05, \lambda = 1.2, \theta = 0.08$ (black-line). From the plots provided in Fig. 2, we can see that the proposed model captures different important behaviors of the hazard rate function such as increasing, decreasing and bathtub shapes.

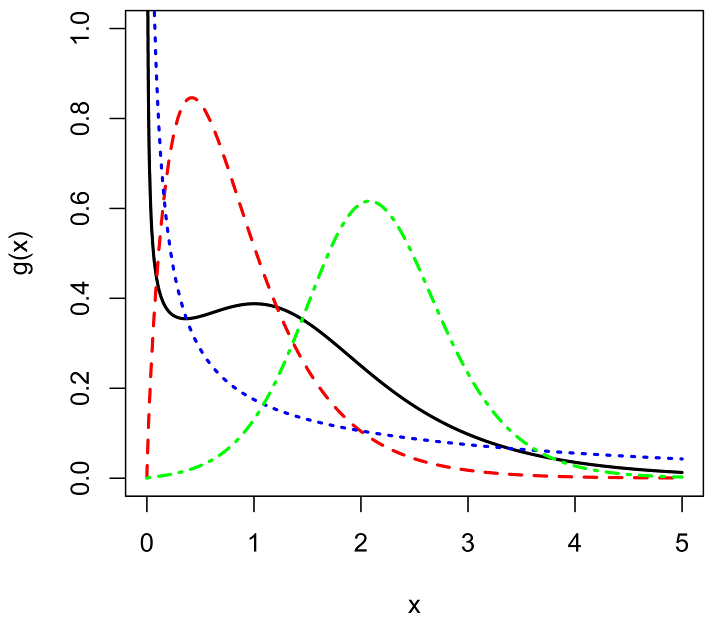

Fig. 1. Different probability density plots of the ELT distribution.

III. Maximum likelihood estimation

In this section, we consider the estimation of the parameters of the RT distribution by the method of maximum likelihood. Let $x_1, \dots, x_n$ be a random sample of size n of the RT distribution with unknown parameter vector $\boldsymbol{\theta} = (\alpha, \beta, \lambda, \theta)^T$. The log-likelihood function for $\boldsymbol{\theta}$ based on a given random sample is

$$\begin{aligned}\ell(\theta) =& n \log(\alpha\beta\lambda\theta) + \beta \sum_{i}^{n} x_i - (\lambda+1) \sum_{i=1}^{n} \log\left[1 + \alpha\left(e^{\beta x_i} - 1\right)\right] \\ &+ (\theta - 1) \sum_{i=1}^{n} \log\left(1 - \left[1 + \alpha\left(e^{\beta x_i} - 1\right)\right]^{-\lambda}\right).\end{aligned}$$

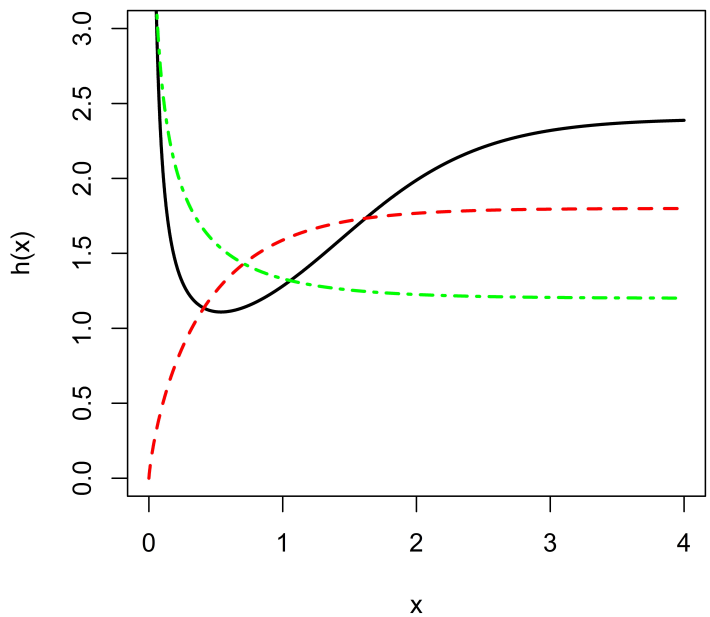

Fig. 2. Different hazard rate plots of the ELT distribution.

The maximum likelihood estimates (MLEs) of the unknown parameters are obtained by maximizing $\ell(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. The first partial derivatives of $\ell(\boldsymbol{\theta})$ with respect to the parameters are given by

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} = \frac{n}{\alpha} - (\lambda+1)\sum_{i=1}^{n} \frac{e^{\beta x_i}-1}{1+\alpha\left(e^{\beta x_i}-1\right)} + (\theta-1)\sum_{i=1}^{n} \frac{\lambda\left(e^{\beta x_i}-1\right)\left[1+\alpha\left(e^{\beta x_i}-1\right)\right]^{-\lambda-1}}{1-\left[1+\alpha\left(e^{\beta x_i}-1\right)\right]^{-\lambda}},$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta} = \frac{n}{\beta} + \sum_{i=1}^{n} x_i - (\lambda+1)\sum_{i=1}^{n} \frac{\alpha x_i e^{\beta x_i}}{1+\alpha\left(e^{\beta x_i}-1\right)} + (\theta-1)\sum_{i=1}^{n} \frac{\lambda \alpha x_i e^{\beta x_i}\left[1+\alpha\left(e^{\beta x_i}-1\right)\right]^{-\lambda-1}}{1-\left[1+\alpha\left(e^{\beta x_i}-1\right)\right]^{-\lambda}},$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} \log\left[1+\alpha\left(e^{\beta x_i}-1\right)\right] + (\theta-1)\sum_{i=1}^{n} \frac{\left[1+\alpha\left(e^{\beta x_i}-1\right)\right]^{-\lambda}\log\left(1-\left[1+\alpha\left(e^{\beta xi}-1\right)\right]^{-\lambda}\right)}{1-\left[1+\alpha\left(e^{\beta x_i}-1\right)\right]^{-\lambda}},$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \log\left(1-\left[1+\alpha\left(e^{\beta x_i}-1\right)\right]^{-\lambda}\right).$$

The MLE $\widehat{\boldsymbol{\theta}} = (\widehat{\alpha}, \widehat{\beta}, \widehat{\lambda}, \widehat{\theta})^T$ of $\boldsymbol{\theta} = (\alpha, \beta, \lambda, \theta)^T$ can be obtained by solving the following equations simultaneously:

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta} = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \lambda} = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta} = 0.$$

There is no closed-form expression for the MLEs, so nonlinear optimization algorithms such as Newton-Raphson iterative technique can be applied to solve the equations and obtain the estimate $\widehat{\boldsymbol{\theta}}$ numerically.

IV. Applications to COVID-19 data sets

The main interest of the derivation of the ELT distribution is its use in data analysis objectives, which makes it useful in many fields, particularly, in the fields dealing with lifetime analysis. Here, this feature is illustrated by taking two sets of data related to COVID-19 epidemic events. For the computation of the numerical results, we use the numerical Newton-Raphson iteration procedure with optim() R-function with the argument method =“BFGS” to estimate the model parameters.

We illustrate the best fitting power of the ELT as compared with well-known lifetime competitive distributions. The Beta Generalized Weibull (BGW) distribution introduced by Singla et al. [1] is a rich class of generalized distributions which has captured a considerable attention over the last years. The probability density function of the five-parameter BGW distribution is given by

$$f(x) = \frac{\alpha\beta\lambda^{\beta}x^{\beta-1}}{B(a,b)}\left(1-e^{-(\lambda x)^{\beta}}\right)^{\alpha a-1}\left\{1-\left(1-e^{-(\lambda x)^{\beta}}\right)^{\alpha}\right\}^{b-1}e^{-(\lambda x)^{\beta}}, \quad x>0, a, b, \alpha, \beta \text{ and } \lambda > 0.$$

Beta Generalized Exponential (BGE), Beta Weibull (BW) , Generalized or Exponentiated Weibull (GW or EW) , Generalized Rayleigh (GR) , Beta Exponential (BE), Generalized Exponential (GE), Weibull (W), Rayleigh (R) and Exponential (E) are well-known sub models of BGW. We show that the ELT distribution provides the best fit to the lifetime data related to the COVID-19 epidemic. The term “best fit” is used in the sense that the proposed model has smaller values of the goodness of fit criterion selected for comparison. These criterion consist of some discrimination measures. These measures are

- The AIC (Akaike information criterion)

$$\text{AIC} = 2k - 2\hat{\ell},$$

- The CAIC (Corrected Akaike information criterion)

$$\text{CAIC} = \frac{2nk}{n-k-1} - 2\hat{\ell},$$

- The BIC (Bayesian information criterion)

$$\text{BIC} = k\log(n) - 2\hat{\ell},$$

where $\hat{\ell}$ is the value of the log-likelihood function with the MLEs, k refers to the number of parameters of the model, and n is the sample size.

A. *Survival times of the COVID-19 patients data*

In this subsection, we consider the survival times (in days) of patients suffering from the COVID-19 epidemic in China. The considered data set representing the survival times of patients from the time admitted to the hospital until death. Among them, a group of fifty-three (53) COVID-19 patients were found in critical condition in hospital from January to February 2020. Among them, 37 patients (70%) were men and 16 women (30%). 40 patients (75%) were diagnosed with chronic diseases, especially including high blood pressure, and diabetes. 47 patients (88%) had common clinical symptoms of the flu, 42 patients

(81%) were coughing, 37 (69%) were short of breath, and 28 patients (53%) had fatigue. 50 (95%) patients had bilateral pneumonia showed by the chest computed tomographic scans. The data set is given by [5]: 0.054, 0.064, 0.704, 0.816, 0.235, 0.976, 0.865, 0.364, 0.479, 0.568, 0.352, 0.978, 0.787, 0.976, 0.087, 0.548, 0.796, 0.458, 0.087, 0.437, 0.421, 1.978, 1.756, 2.089, 2.643, 2.869, 3.867, 3.890, 3.543, 3.079, 3.646, 3.348, 4.093, 4.092, 4.190, 4.237, 5.028, 5.083, 6.174, 6.743, 7.274, 7.058, 8.273, 9.324, 10.827, 11.282, 13.324, 14.278, 15.287, 16.978, 17.209, 19.092, 20.083.

Table I gives the MLEs of the parameters and the values of the goodness of fit statistics for the considered models. From the values of these statistics, we infer that the four-parameter ELT distributions provides a better fit than other distributions for the first data set. From the values of these statistics, we

Table I. MLEs of the parameters for the models fitted to the first data set and the values of the AIC, CAIC and BIC.

Model	MLEs					AIC	CAIC	BIC
	a	b	λ	α	β			
5 parameters								
BGW	0.1441	0.1003	1.208	6.700	1.127	272.49	273.77	282.34
4 parameters								
ELT	$\theta = 4.590$	–	0.3367	124.1	0.5139	271.00	271.83	278.88
BGE	4.783	3.156	0.0377	0.2068	1	274.20	275.03	282.08
BW	0.8789	0.8031	0.2697	1	0.8532	274.77	275.60	282.65
3 parameters								
GW	1	1	0.2069	0.8793	0.8544	272.77	273.26	278.68
BE	0.6908	0.9597	0.1700	1	1	272.84	273.33	278.75

infer that the three-parameter RT distributions provides a better fit than other distributions for the real data set.

B. Second real data set

This data is reported in (https://www.worldometers.info/coronavirus/country/china/) which represents daily new deaths due to COVID-19 in China from 23 January to 28 March. The data are: 8, 16, 15, 24, 26, 26, 38, 43, 46, 45, 57, 64, 65, 73, 73, 86, 89, 97, 108, 97, 146, 121, 143, 142, 105, 98, 136, 114, 118, 109, 97, 150, 71, 52, 29, 44, 47, 35, 42, 31, 38, 31, 30, 28, 27, 22, 17, 22, 11, 7, 13, 10, 14, 13, 11, 8, 3, 7, 6, 9, 7, 4, 6, 5, 3, 5.

Table II gives the MLEs of the parameters and the values of the goodness of fit statistics for the considered models. From the values of these statistics, we infer that the four-parameter ELT distributions provides a better fit than other distributions for the second data set, too.

References

[1] N. Singla, K. Jain, and S.K. Sharma, “The beta generalized Weibull distribution: properties and applications,” Reliab. Eng. Syst. Saf. United Kingdom, vol. 102, pp. 5–15, 2012.

[2] W. Barreto-Souza, A. H. Santos, and G.M. Cordeiro, “The beta generalized exponential distribution,” J. Stat. Comput. Simul. United Kingdom, vol. 80, pp. 159–172, 2010.

[3] C. Lee, F. Famoye, and O. Olumolade, “Beta-Weibull distribution: some properties and applications to censored data,” J. Mod. Appl. Stat. Methods, United States, vol. 6, pp. 17, 2007.

Table II. MLEs of the parameters for the models fitted to the second data set and the values of the AIC, CAIC and BIC.

Model	MLEs					AIC	CAIC	BIC
	a	b	λ	α	β			
5 parameters								
BGW	0.2619	0.0885	0.1469	6.288	1.162	651.60	652.60	662.55
4 parameters								
ELT	$\theta = 0.8336$	$-$	0.0001	0.0043	4.410	648.21	648.87	656.97
BGE	0.2850	0.0934	0.2124	11.562	1	649.35	650.01	658.11
BW	1.687	0.0831	0.2391	1	1.0046	652.93	653.59	661.69
3 parameters								
GW	1	1	0.0184	0.9345	1.1362	652.96	653.35	659.53
BE	1.1511	0.7547	0.0288	1	1	653.04	653.42	659.61

[4] F.A. Bhatti, S. Mirzaei Salehabadi, G.G. Hamedani, "On Burr III-Inverse Weibull Distribution with COVID-19 Applications," J. Iran. Stat. Soc. Iran, vol. 20(1), pp. 101–121, 2021.

[5] X. Liu, Z. Ahmad, A.M. Gemeay, A.T. Abdulrahman, E.H. Hafez, and N. Khalil, "Modeling the survival times of the COVID-19 patients with a new statistical model: A case study from China", Plos one, vol. 16(7), e0254999, 2021.

[6] A. Khosravi Tanak, G. R. Mohtashami Borzadaran, and J. Ahmadi, "Entropy maximization under the constraints on the generalized Gini index and its application in modeling income distributions, "Physica A, Netherlands, vol. 438, pp. 657–666, 2015.

[7] W. G. Gilchrist, Statistical Modelling with Quantile Functions, Chapman and Hall/CRC, Boca Raton, LA. 2001.

[8] S. M. Ross, Introduction to Probability Models, Academic press, Boston, 2002.

[9] R. E. Barlow and F. Proschan, Mathematical Theory of Reliability. Society for Industrial and Applied Mathematics, Philadelphia, PA. 1996.