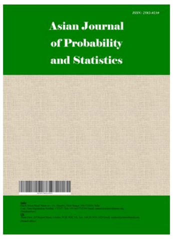


Asian Journal of Probability and Statistics

12(1): 1-19, 2021; Article no.AJPAS.67000
ISSN: 2582-0230

Generating New Flexible Lifetime Class of Distributions Based on Competing Risks and Frailty Models

M. M. E. Abd El-Monsef[1*] and M. M. El-Awady[1]

[1] *Faculty of Science, Tanta University, Egypt.*

Authors' contributions

This work was carried out in collaboration between both authors. Author AEM mainly contributed Sects. 4, 5 and 6 and revised the manuscript. Author MEA wrote the first draft of the manuscript and fully contributed Sects. 2, 3, and 7. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2021/v12i130277
Editor(s):
(1) Dr. Manuel Alberto M. Ferreira, Lisbon University, Portugal.
Reviewers:
(1) R. Thirumalai, N.G.P Institute of Technology, India.
(2) Sebastiano Pennisi, University of Cagliari, Italy.
(3) Muzhou Hou, Central South University, China.
Complete Peer review History: http://www.sdiarticle4.com/review-history/67000

Received: 26 January 2021
Accepted: 01 April 2021
Published: 11 April 2021

Original Research Article

Abstract

New classes of continuous distributions have been generated, in the last decad, based on a compounding procedure arises on a latent competing risks problem. This procedure assumes the homogeneity between the population individuals. In this paper, a new lifetime distribution is generated, assuming the heterogeneity at both population and individual levels, called Extended Gamma Gompertz (EGG) distribution. This distribution shows very desirable flexibility of its hazard function. Some properties of the proposed distribution are given. Maximum likelihood estimation technique is used to estimate the parameters. A simulation study is performed to examine the performance of the proposed model. Finally, application to a real data set is given to exemplify the utility of the EGG distribution.

Keywords: Gompertz distribution; gamma distribution; gamma Gompertz distribution; extended gamma Gompertz distribution; Marshal-Olken family of distributions.

2010 Mathematics Subject Classification: 60E05, 62E10, 62E15.

**Corresponding author: E-mail: mmezzat@science.tanta.edu.eg;*

1 Introduction

Gompertz distribution is a classical mathematical model, introduced by Gompertz [1], plays an important role in modeling survival times, human mortality and actuarial data. The probability density function (pdf) of Gompertz distribution is given by

$$f_G(x) = ae^{-\frac{a}{b}(e^{bx}-1)+bx}, x > 0, a > 0, b > 0, \tag{1.1}$$

and its cumulative distribution function (cdf) is

$$F_G(x) = 1 - e^{-\frac{a}{b}(e^{bx}-1)}, x > 0, a > 0, b > 0. \tag{1.2}$$

An extension of Gompertz distribution has been proposed by El-gohary et al. [2], called generalized Gompertz (GG) distribution, based on family of distributions given by Lehman alternatives (called exponentiated type family by Nadarajah and Kotz [3]) considered by Gupta et al. [4]. Jafari et al. [5] proposed beta Gompertz (BG) distribution using beta generator introduced by Eugene et al. [6]. Gompertz power series distributions by Jafri and Tahmasebi [7], using the technique of Marshall and Olkin [8]. Transimuted Gompertz (TG) distribution by Abdul-Moniem and Seham [9], who considered transmuted generator introduced by Show and Buckley [10]. Roozegar et al. introduced McDonald Gompertz (McG) distribution [11] based on McDonald generator introduced by Alexander et al. [12]. Eghwerido et al. [13] proposed alpha power Gompertz distribution based on alpha power transformation method by Mahdavi and Kundu [14].

This paper aims to utilize two key concepts in survival analysis, namely competing risks problem and frailty models, to propose new four-parameter distribution. The proposed distribution is called extended gamma Gompertz (EGG) distribution. An advantage of this model is to consider the heterogeneity that may appears among population individuals with their frailties described by the gamma distribution. Furthermore, the number of competing causes is modeled by the geometric distribution. The EGG distribution has more flexibility compared to Gompertz distribution and some of its extensions.

The paper is organized as follows: Section 2 discuses briefly the concepts of frailty models and competing risks problem. Section 3 proposes the EGG distribution. In section 4, Some statistical and reliability properties of the proposed model are provided. The Maximum Likelihood method is used to estimate the parameters of the proposed model in section 5. In section 6, Monte Carlo simulation study is performed to examine the average bias, mean square error, coverage probability and average confidence width of the maximum likelihood estimates. Finally, an application is given, to illustrate superiority of the EGG distribution, in section 7.

2 Preliminaries

2.1 Frailty models

Ordinary survival analysis deals with the case of independent and identically distributed data, this is based on the assumption that the study population is homogeneous, meaning all individuals have the same risk of death (same frailty). However, it is a basic observation of many areas that the individuals differ greatly (have different frailties). Thus, in the context of survival analysis, individuals that have more frail will fail earlier of that have lesser frail. The notion of frailty models is introduced to assess this heterogeneity in a nice way.

In its simplest form, frailty is an unobserved random factor that modifies the hazard function of an individual, or of related individuals. The most common frailty model is a model in which the hazard function is a product of random variable (frailty) and baseline hazard function which is common to all individuals. The individual hazard function is defined as

$$h_i(x) = z_i\ h_0(x)$$

where, z_i is a realization of the frailty Z and $h_0(x)$ is the conditional baseline hazard function for a subject with $z_i = 1$. The conditional survival function for subject i is then given by

$$S_i(x|z_i) = e^{-z_i \int_0^x h_0(s)\ ds} = e^{-z_i H_0(x)},$$

where H_0 is the base line cumulative Hazard function. In order to know how the hazard evolves over time in the population, we need to derive the population survival function by integrating out the frailty

$$S(x) = \int_0^\infty e^{-Z H_0(x)}\ f_Z(z)\, dz.$$

where $f_Z(z)$ is the pdf of the frailty random variable Z. For further details and discussions on frailty models in literature, see for example [15, 16, 17].

2.2 Competing risks

Another key concept arises in survival analysis is competing risks problem, [18]. Simplistically, we only observe the minimum component lifetime of a series system, which is the cause of failure for the system. In recent years, new classes of distributions have been generated based on a compounding procedure inherited from the competing risks problem with its simplest case, [8, 19, 20], etc, in which the components lifetimes $X_i, i = 1, \ldots, N$, are independent and identically distributed non-negative continuous random variables with common survival function $S(x)$. Also, N (the number of components) is zero truncated discrete random variable. Thus, the distribution function of the new family is the marginal distribution of the first order statistics $X_{(1)} = \min\{X_i : i = 1, \ldots, N\}$ and defined by

$$F(x) = \sum_{i=1}^{n} \{1 - S(x)^n\}\ P\{N = n\}. \tag{2.1}$$

The heterogeneity appears in this model at the system level through the unknown number N of causes of failure of the system. The Xs can be detected only after a component failure, in which case it is repaired perfectly. Thus, the homogeneity in the model appears at the component level, in which the components stay in the same condition even after a succession of failures and repairs.

The following section proposes the EGG distribution, on the latent of competing risks problem and frailty models, so that the compounding procedure (2.1) can accommodate heterogeneity at the components level, see section 2.

3 The Extended Gamma Gompertz Distribution

Consider $X_i,\ i = 1, \ldots, N$ are independent and identically distributed random variables represent lifetimes of a series systems components. Let $S_0(x) = e^{-\frac{a}{b}\left(e^{bx}-1\right)},\ a, b > 0$ be a common baseline Gompertz survival function of Xs, consequently $h_0(x) = ae^{bx}$ is the common baseline hazard rate function and $H_0(x) = \frac{a}{b}\left(e^{bx} - 1\right)$ is the baseline cumulative hazard function. Also, consider N as a geometric random variable with probability mass function

$$P(N = n) = p(1-p)^{n-1}. \tag{3.1}$$

In the case of a component failure, it can be repaired perfectly, repaired imperfectly or replaced with a more efficient component (replaced by a technologically more advanced component), [21, 22].

Therefore, after a succession of failures and repairs of the system's components, heterogeneity may appear between them where they exposed to failure differently (have different frailties). In this case, using the concepts of frailty models, the failure rate of the i_{th} component will be

$$h_i(x|z) = Z_i h_0\left(x\right)$$

where $Z_i,\ i = 1, \ldots, N$ are interpreted as independent and identically distributed random variables. The value of Z_i can be interpreted as: When $Z_i = 1$, the component is repaired perfectly. If $Z_i > 1$, the component is repaired imperfectly and if $Z_i < 1$, the component is replaced with a more efficient one. In frailty modelling, the typical choice for the distribution of the frailty Z_i is the one parameter gamma distribution, Gamma($1/\theta$,$1/\theta$), with pdf given by

$$f_{Z_i}\left(z_i\right) = \frac{1}{\Gamma\left(\frac{1}{\theta}\right)\theta^{\frac{1}{\theta}}}\ Z_i{}^{\frac{1}{\theta}-1}\ e^{-\frac{Z_i}{\theta}}\ , \tag{3.2}$$

where $E\left(Z_i\right) = 1$ and $Var\left(Z_i\right) = \theta$.

The conditional survival function for the i_{th} component is then given by

$$S_i\left(x|z_i\right) = e^{-Z_i\int\limits_0^x h_0(s)\ ds} = e^{-Z_i H_0(x)}.$$

Thus, the mixture survival for the i_{th} component is

$$S_i\left(x\right) = \int\limits_0^\infty S_i\left(x|z_i\right)\ f_{Z_i}\left(z_i\right)dz = L\left(H_0\left(x\right)\right)$$

where $L\left(s\right) = \left(1 + \theta\ s\right)^{-\frac{1}{\theta}}$ is Laplace transform for the gamma pdf (3.2).

Finally, the survival function for the i_{th} component, which is common to all components, is given by

$$S_i\left(x\right) = S\left(x\right) = \left(1 + \frac{a\theta}{b}\left(e^{bx} - 1\right)\right)^{-\frac{1}{\theta}}. \tag{3.3}$$

The survival function (3.3) corresponds to a Gamma Gompertz distribution with pdf given by,

$$g\left(x; a,\ b,\ \theta\right) = a\mathrm{e}^{bx}\left(1 + \frac{a\theta}{b}\left(e^{bx} - 1\right)\right)^{-\frac{1}{\theta}-1} \tag{3.4}$$

Note that, Gamma Gompertz distribution has been studied as a customer lifetime model with parameters $\eta = \frac{a\theta}{b}$, b and $s = \frac{1}{\theta}$, [23].

Using (3.1) and (3.3) in (2.1) we get a four parameter distribution, Extended Gamma Gompertz (EGG) distribution, with distribution function given by

$$F\left(x\right) = 1 - p\left\{p^* + \left[1 + \frac{a\theta}{b}\left(e^{bx} - 1\right)\right]^{\frac{1}{\theta}}\right\}^{-1}, x > 0 \tag{3.5}$$

where $a, b, \theta, p > 0$ and $p^* = p - 1$.

So that, the survival and probability density functions are respectively given by

$$\bar{F}\left(x\right) = p\left\{p^* + \left[1 + \frac{a\theta}{b}\left(e^{bx} - 1\right)\right]^{\frac{1}{\theta}}\right\}^{-1}, x > 0 \tag{3.6}$$

and

$$f(x) = \frac{pae^{bx}\left(1+\frac{a\theta}{b}\left(e^{bx}-1\right)\right)^{-1+\frac{1}{\theta}}}{\left(p^{*}+\left(1+\frac{a\theta}{b}\left(e^{bx}-1\right)\right)^{\frac{1}{\theta}}\right)^{2}}, x>0 \tag{3.7}$$

It is very complicated to study the shapes of the pdf is (1) analytically. Whereas, figure 3 shows the possible shapes of the pdf of EGG distribution. It can be decreasing, unimodal and decreasing-increasing-decreasing. It is also can be used to model left skewed, right skewed and symmetric data sets. The EGG distribution can be reduced to the following sub-models:

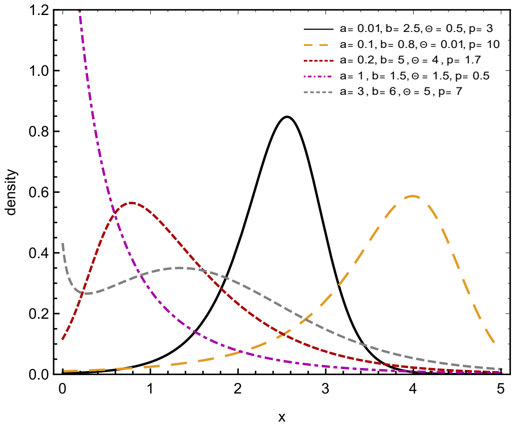


Fig. 1. The pdf of EGG distribution, for different values of parameters a, b, θ and p

- When $(p = 1)$, the Gamma Gompetz (GG) distribution is obtained.
- When $(a = b$ and $\theta = 1)$ or $(\theta \to 0^{+}$ and $b \to 0^{+}$), the Marshall-Olkin extended exponential (MOE) distribution is obtained, [8].
- When $(a = b$ and $\theta = p = 1)$ or $(p = 1,\ \theta \to 0^{+}$ and $b \to 0^{+})$, the exponential (E) distribution is obtained.
- When $\theta \to 0^{+}$, the Marshall-Olkin extended Gompertz (MOG) distribution is obtained, [7].
- When p=1 and $\theta \to 0^{+}$, the Gompertz (G) distribution is obtained.

It worth noted that Mazucheli et al. [24] used the negative exponential function transformation to proposed unit-Gompertz distribution for modeling data on the unit-interval, $(0, 1)$.

4 Properties

This section investigates some statistical and reliability properties of the EGG distribution.

4.1 Quantiles and mode

The quantile function of EGG is obtained in an explicit form as

$$Q(u) = \frac{1}{b}\ln\left[\frac{b}{a\ \theta}\left(\frac{u(p-1)+1}{1-u}\right)^{\theta} - \left(\frac{b-a\ \theta}{a\ \theta}\right)\right];\quad 0<u<1 \tag{4.1}$$

To get the mode of the EGG distribution, first we have to differentiate its pdf with respect to x

$$f'(x) = \frac{f(x)}{b+a\ \theta\left(e^{b\ x}-1\right)}\ \varphi(x)$$

where

$$\varphi(x) = (p-1)\left(b + a\left(e^{b\ x} - \theta\right)\right) + \left(1 + \frac{a\theta}{b}\left(e^{b\ x} - 1\right)\right)^{\frac{1}{\theta}}\ \left(b - a\ \left(e^{b\ x} + \theta\right)\right).$$

Since $\frac{f(x)}{b+a\ \theta\left(e^{b\ x}-1\right)} > 0$, then the mode is the solution of the equation $\varphi(x) = 0$ with respect to x. We have to find that solution numerically using a mathematical package.

4.2 Hazard and reversed rate function

The hazard and reverse hazard functions of the model are given by

$$h(x) = \frac{ae^{bx}(1+\theta\ \frac{a}{b}\left(e^{bx}-1\right))^{\frac{1}{\theta}-1}}{p^{*} + (1+\theta\ \frac{a}{b}\left(e^{bx}-1\right))^{\frac{1}{\theta}}} \tag{4.2}$$

and

$$r(x) = \frac{ae^{bx}p(1+\theta\ \frac{a}{b}\left(e^{bx}-1\right))^{\frac{1}{\theta}-1}}{(-1+(1+\theta\ \frac{a}{b}\left(e^{bx}-1\right))^{\frac{1}{\theta}})(-1+p+(1+\theta\ \frac{a}{b}\left(e^{bx}-1\right))^{\frac{1}{\theta}})} \tag{4.3}$$

where $a, b, \theta, p > 0$ and $p^{*} = p - 1$. The limiting behavior of the hazard function can be readily established as follows $\lim_{x\to 0} h(x) = \frac{a}{p}$ and $\lim_{x\to\infty} h(x) = \frac{b}{\theta}$.

Fig. 2. shows a very desirable flexibility of the hazard functions in which it shows various shapes including increasing, decreasing, unimodal, bathtub and increasing-constant shapes which make the model useful in the real life applications.

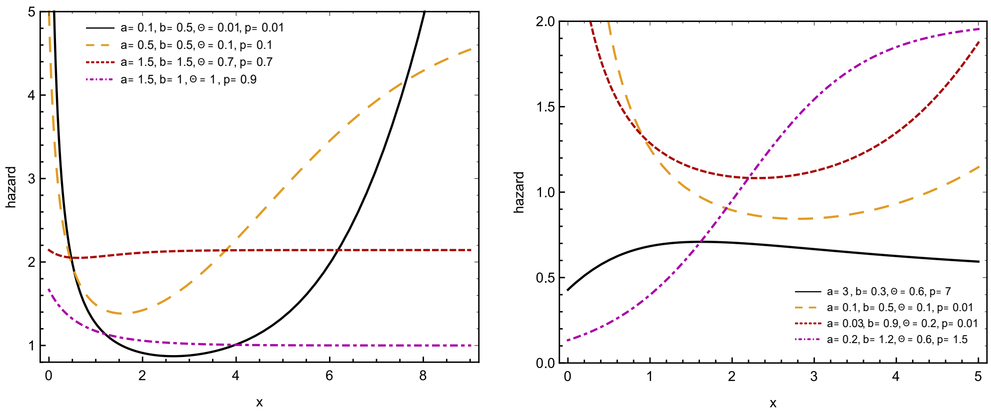


Fig. 2. Plots of the hazard function for different values of the parameters

4.3 Series representation for survivor and probability density functions

This section presents expansions of the survivor function and pdf of the EGG distribution. To do so, the following power series identity is considered

$$(x+a)^{\nu} = \sum_{j=0}^{\infty} \binom{\nu}{j} x^{j} a^{\nu-j},$$

where ν is a real number. This bower series converges for $\nu \geq 0$ an integer(in this case the index j stopped at ν) , or $\left|\frac{x}{a}\right| < 1$. This general form is given by Graham et al. [25].

The pdf of EGG distribution can be written as an infinite mixture of GG distribution with parameters a, b and θ. The pdf in (1) can be rewritten as

$$f(x) = \frac{p\ a e^{bx}\left(1 + \frac{a\theta}{b}\left(e^{bx} - 1\right)\right)^{-\frac{1}{\theta}-1}}{\left(1 - (1-p)\left(1 + \frac{a\theta}{b}\left(e^{bx} - 1\right)\right)^{-\frac{1}{\theta}}\right)^2}$$

Because $\left|(1-p)\left(1 + \frac{a\theta}{b}\left(e^{bx} - 1\right)\right)^{-\frac{1}{\theta}}\right| < 1$ given that $p \in (0,1)$, the pdf of the EGG distribution is reduced to

$$f(x) = p\ a e^{bx} \sum_{j=0}^{\infty} (j+1)(1-p)^j \left(1 + \frac{a\theta}{b}\left(e^{bx} - 1\right)\right)^{-\frac{j+1}{\theta}-1}.$$

Using the pdf of the GG distribution (3.4), the pdf of the EGG distribution can be written as

$$f(x) = p \sum_{j=0}^{\infty} (1-p)^j g\left(x; a(j+1),\ b,\ \frac{\theta}{(j+1)}\right); p \in (0,1) \tag{4.4}$$

Various mathematical properties of the EGG distribution can be obtained from Equation (4.4) using the corresponding properties of the GG distribution.

Furthermore, the survival function (3.6) could be expanded as follows

Case I: for $p \in (0,1)$, the survival function (3.6) can be written as

$$\bar{F}(x) = p \sum_{k=0}^{\infty} (-1)^k\ (p-1)^k \left(\left(1 - \frac{a\theta}{b}\right) + \frac{a\theta}{b} e^{bx}\right)^{-\omega}; \omega = \frac{k+1}{\theta}$$

$$\bar{F}(x) = p \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} (-1)^{k+j} \binom{\omega + j - 1}{j}\ (p-1)^k \left(\frac{b}{a\theta} - 1\right)^j \left(\frac{a\theta}{b}\right)^{-\omega} e^{-b(\omega+j)x}, \tag{4.5}$$

where $p \in (0,1), \omega = \frac{k+1}{\theta}$ and $\frac{a\theta}{b} > 0.5$.

Case II: for $p > 1$, the survival function (3.6) can be written as

$$\begin{aligned}
\bar{F}(x) =& \left(1 + \frac{a\theta}{b}\left(e^{bx} - 1\right)\right)^{-\frac{1}{\theta}} \left(1 - \left(\frac{p-1}{p}\right) \left[1 - \left(1 + \frac{a\theta}{b}\left(e^{bx} - 1\right)\right)^{-\frac{1}{\theta}}\right]\right)^{-1} \\
=& \left(1 + \theta\ \frac{a}{b}\left(e^{bx} - 1\right)\right)^{-\frac{1}{\theta}} \sum_{k=0}^{\infty} \left(\frac{p-1}{p}\right)^k \left[1 - \left(1 + \frac{a\theta}{b}\left(e^{bx} - 1\right)\right)^{-\frac{1}{\theta}}\right]^k \\
=& \sum_{k=0}^{\infty} \sum_{j=0}^{k} (-1)^j \binom{k}{j} \left(\frac{p-1}{p}\right)^k \left(\left(1 - \frac{a\theta}{b}\right) + \frac{a\theta}{b} e^{bx}\right)^{-\psi}; \qquad \psi = \frac{j+1}{\theta}
\end{aligned}$$

Finally,

$$\bar{F}(x) = \sum_{k=0}^{\infty} \sum_{j=0}^{k} \sum_{l=0}^{\infty} (-1)^{j+l} \binom{k}{j} \binom{\psi + l - 1}{l} \left(\frac{p-1}{p}\right)^k \left(\frac{b}{a\theta} - 1\right)^l \left(\frac{a\theta}{b}\right)^{-\psi} e^{-b(\psi+l)x} \tag{4.6}$$

where $p > 1$, $\psi = \frac{j+1}{\theta}$ and $\frac{a\theta}{b} > 0.5$.

Similarly, the pdf could be represented as

$$f(x) = p \sum_{k=0}^{\infty}\sum_{j=0}^{\infty}(-1)^{k+j}\binom{\omega+j-1}{j}(p-1)^k\left(\frac{b}{a\theta}-1\right)^j\left(\frac{a\theta}{b}\right)^{-\omega} b(\omega+j)\ e^{-b(\omega+j)x}, \quad (4.7)$$

where $p \in (0,1), \omega = \frac{k+1}{\theta}$ and $\frac{a\theta}{b} > 0.5$.
and

$$f(x) = \sum_{k=0}^{\infty}\sum_{j=0}^{k}\sum_{l=0}^{\infty}(-1)^{j+l}\binom{k}{j}\binom{\psi+l-1}{l}\left(\frac{p-1}{p}\right)^k\left(\frac{b}{a\theta}-1\right)^l\left(\frac{a\theta}{b}\right)^{-\psi} b(\psi+l)\,e^{-b(\psi+l)x} \quad (4.8)$$

where $p > 1$, $\psi = \frac{j+1}{\theta}$ and $\frac{a\theta}{b} > 0.5$.

4.4 Moments

Many interesting characteristics and features of a distribution can be studied through its moments. Let X be a random variable following the EGG distribution. The r_{th} ordinary moments of the random variable X, denoted by $\mu_r^{'}$, is the expected value of X^r

$$\mu_r^{'} = \int_0^{\infty} r\ x^{r-1}\ \bar{F}(x)\ dx.$$

When $p \in (0,1)$, the r_{th} ordinary moments of the random variable X is

$$\mu_r^{'} = p\ r \sum_{k=0}^{\infty}\sum_{j=0}^{\infty}(-1)^{k+j}\binom{\omega+j-1}{j}(p-1)^k\left(\frac{b}{a\theta}-1\right)^j\left(\frac{a\theta}{b}\right)^{-\omega}\int_0^{\infty} x^{r-1}\ e^{-b(\omega+j)x}\ dx.$$

That is,

$$\mu_r^{'} = \sum_{k=0}^{\infty}\sum_{j=0}^{\infty}(-1)^{k+j}\binom{\omega+j-1}{j}(p-1)^k\left(\frac{b}{a\theta}-1\right)^j\left(\frac{a\theta}{b}\right)^{-\omega}\frac{p\ r\ \Gamma(r)}{(b(\omega+j))^r}, \quad (4.9)$$

where $p \in (0,1)$, $\omega = \frac{k+1}{\theta}$ and $\Gamma(r) = \int_0^{\infty} u^{r-1}\ e^{-u}\ du$, $u = b(\omega+j)x$.

Similarly, when $p > 1$, the r_{th} ordinary moments of the random variable X is

$$\mu_r^{'} = \sum_{k=0}^{\infty}\sum_{j=0}^{k}\sum_{l=0}^{\infty}(-1)^{j+l}\binom{k}{j}\binom{\psi+l-1}{l}\left(\frac{p-1}{p}\right)^k\left(\frac{b}{a\theta}-1\right)^l\left(\frac{a\theta}{b}\right)^{-\psi}\frac{r\ \Gamma(r)}{(b(\psi+l))^r} \quad (4.10)$$

where $p > 1$ and $\psi = \frac{j+1}{\theta}$.

The r_{th} moment of the EGG distribution could be expressed in terms of the generalized hypergeometric function ${}_{\mathrm{P}}\mathrm{F}_{\mathrm{Q}}\left[\{a_1,\ldots,a_{\mathrm{P}}\},\{b_1,\ldots,b_{\mathrm{Q}}\},z\right] = \sum_{j=0}^{\infty}\frac{(a_1)_j\ldots(a_{\mathrm{P}})_j}{(b_1)_j\ldots(b_{\mathrm{Q}})_j}\frac{z^j}{j!}$, Abramowitz et al. [26], as

$$\mu_r^{'} = \sum_{k=0}^{\infty}(-1)^k\ (p-1)^k\left(\frac{a\theta}{b}\right)^{-\omega}\frac{p\ \theta^r\ \Gamma(r)}{b^r(k+1)^r}\ GHG(\omega,a,b,\theta) \quad (4.11)$$

where $p \in (0,1)$ and $GHG(\omega,a,b,\theta) = {}_{(r+1)}\mathrm{F}_r\left(\left\{\frac{\omega+1}{\theta},\ldots,\frac{\omega+1}{\theta}\right\},\left\{1+\frac{\omega+1}{\theta},\ldots,1+\frac{\omega+1}{\theta}\right\},\left(1-\frac{b}{a\theta}\right)\right)$.

Similarly, for $p > 1$

$$\mu_r^{'} = \sum_{k=0}^{\infty}\sum_{j=0}^{k}(-1)^j\binom{k}{j}\left(\frac{p-1}{p}\right)^k\left(\frac{a\theta}{b}\right)^{-\psi}\frac{r\ \Gamma(r)}{b^r(1+j)^r}\ GHG(j,a,b,\theta), \quad (4.12)$$

where $GHG(j,a,b,\theta) = {}_{(r+1)}\mathrm{F}_r\left(\left\{\frac{j+1}{\theta},\ldots,\frac{j+1}{\theta}\right\},\left\{1+\frac{j+1}{\theta},\ldots,1+\frac{j+1}{\theta}\right\},\left(1-\frac{b}{a\theta}\right)\right)$.

Table 1. lists the first four moments, variance, skewness, and kurtosis of the EGG distribution for various values of the parameter vector $\boldsymbol{\theta} = (a, b, \theta, p)$. The results we have using the series representations (4.9)–(4.12) are the same as using the numerical integrations.

Table 1. The first four moments, variance, skewness, and kurtosis of EGG distribution for different values of parameter vector $\boldsymbol{\theta} = (a, b, \theta, p)$

$\boldsymbol{\theta}$	μ_1'	μ_2'	μ_3'	μ_4'	Variance	skewness	kurtosis
(1.6,0.9,1.4,0.8)	0.9088	2.2972	9.919	59.8826	1.4711	2.8902	15.3242
(2,5,2,3)	0.8442	1.0192	1.5544	2.8622	0.3065	1.0397	4.7652
(5,8,5,7)	1.2828	2.433	5.8343	16.8652	0.7874	0.9918	4.5588
(0.25,1.3,0.08,7)	1.87395	3.8408	8.2985	18.6451	0.3291	-0.7015	3.4298

4.5 Mean residual life

Another aging property for EGG distribution is the mean residual life (mrl). Its defined simply as the expected additional lifetime given that a component has survived until time t is a function of t. More specifically, if the random variable X represents the life of a component, then the mean residual life is given by $mrl(t) = E(X - t|X > t)$, [27].

The mean residual life of the EGG distribution

$$mrl(t) = E(X-t|X>t) = \frac{1}{\bar{F}(t)}\int_t^\infty \bar{F}(x)\ dx, \quad t>0$$

$$= \frac{p}{\bar{F}(t)}\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}(-1)^{k+j}\binom{\omega+j-1}{j}(p-1)^k\left(\frac{b}{a\theta}-1\right)^j\left(\frac{a\theta}{b}\right)^{-\omega}\int_t^\infty e^{-b(\omega+j)x}dx$$

$$= \frac{p}{\bar{F}(t)}\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}(-1)^{k+j}\binom{\omega+j-1}{j}(p-1)^k\left(\frac{b}{a\theta}-1\right)^j\left(\theta\ \frac{a}{b}\right)^{-\omega}\frac{e^{-b(\omega+j)\ t}}{b(\omega+j)}$$

The quantity $\binom{\omega+j-1}{j}$ can be written as $\frac{(\omega)_j}{j!}$, where $(\omega)_j$ is a rising factorial power (Pochhammer symbol). Thus,

$$mrl(t) = \frac{p}{\bar{F}(t)}\sum_{k=0}^{\infty}(-1)^k(p-1)^k\left(\frac{a\theta}{b}\right)^{-\omega}e^{-b\omega t}\sum_{k=0}^{\infty}\frac{(\omega)_j}{j!}\left(\left(1-\frac{b}{a\theta}\right)e^{-b\ t}\right)^j\frac{1}{b(\omega+j)}.$$

The quantity $\frac{1}{b(\omega+j)}$ can be written as $\frac{\theta}{b(k+1)}\ \frac{(\omega)_j}{(\omega+1)_j}$.

Thus, when $p \in (0,1)$, the mean residual life of EGG distribution can be written as

$$mrl(t) = \left(p^* + \left(1+\frac{a\theta}{b}\left(\mathrm{e}^{bt}-1\right)\right)^{\frac{1}{\theta}}\right)\sum_{k=0}^{\infty}(-1)^k\frac{\theta(p-1)^k}{b(k+1)}\left(\frac{a\theta}{b}\right)^{-\omega}e^{-b\omega t}\ HG(\omega,a,b,\theta) \quad (4.13)$$

where $p \in (0,1)$, $HG(\omega,a,b,\theta) = \text{HyperGeometric2F1}\left(\omega,\omega,\omega+1,\left(1-\frac{b}{a\theta}\right)e^{-b\ t}\right)$ and $\frac{a\theta}{b} > 0.5$, Abramowitz et al. [26].

Similarlly, when $p>1$, the mean residual life can be written as

$$mrl\,(t)=\frac{p^{*}+\left(1+\frac{a\theta}{b}\left(\mathrm{e}^{bt}-1\right)\right)^{\frac{1}{\theta}}}{p}\sum_{k=0}^{\infty}\sum_{j=0}^{k}\frac{(-1)^{j}e^{-b\psi t}\left(\frac{p-1}{p}\right)^{k}\left(\frac{b}{a\theta}\right)^{\psi}\theta k!}{b(j+1)j!(k-j)!}\;HG(\psi,a,b,\theta) \qquad (4.14)$$

where $p>1$, $HG(\psi,a,b,\theta)=\text{HyperGeometric2F1}\left(\psi,\psi,\psi+1,\left(1-\frac{b}{a\theta}\right)e^{-bt}\right)$ and $\frac{a\theta}{b}>0.5$.

4.6 Distribution of order statistics

The pdf of the i_{th} order statistics, for $p\in(0,1)$, is given by

$$\begin{aligned}
f_{i:n}\,(x)=&\frac{n!f(x)}{(i-1)!(n-i)!}\;\left[1-\bar{F}\,(x)\right]^{i-1}\;\left[\overline{F}\,(x)\right]^{n-i}\\
=&\frac{n!f(x)}{(i-1)!(n-i)!}\sum_{q=0}^{i-1}(-1)^{q}\binom{i-1}{q}\left[\overline{F}\,(x)\right]^{q+n-i}\\
=&\frac{n!ae^{bx}\left(1+\theta\frac{a}{b}\left(e^{bx}-1\right)\right)^{-1+\frac{1}{\theta}}}{(i-1)!(n-i)!}\sum_{q=0}^{i-1}(-1)^{q}\binom{i-1}{q}p^{\nu+1}\left\{p^{*}+\left(1+\frac{a\theta}{b}\left(e^{bx}-1\right)\right)^{\frac{1}{\theta}}\right\}^{-(\nu+2)},
\end{aligned}$$

$$\begin{aligned}
f_{i:n}\,(x)=&\frac{an!}{(i-1)!(n-i)!}\sum_{q=0}^{i-1}\sum_{k,j=0}^{\infty}(-1)^{q+k+j}p^{\nu+1}\binom{i-1}{q}\binom{\nu+k+1}{k}\binom{\delta+j-1}{j}\\
&\times(p-1)^{k}\left(\frac{b}{a\theta}-1\right)^{j}\left(\frac{a\theta}{b}\right)^{-\delta}e^{-b(\delta+j-1)x}
\end{aligned}$$

for $p\in(0,1)$, where $\nu=q+n-i$, $\delta=\frac{\nu+k+1}{\theta}+1$ and $\frac{a\theta}{b}>0.5$.

For $p>1$, the pdf of the i_{th} order statistics is given by

$$\begin{aligned}
f_{i:n}(x)=&\frac{n!}{(i-1)!(n-i)!}\frac{\frac{1}{p}ae^{bx}\left(1+\frac{a\theta}{b}\left(e^{bx}-1\right)\right)^{-\frac{1}{\theta}-1}}{\left\{1-\left(\frac{p-1}{p}\right)\left[1-\left(1+\frac{a\theta}{b}\left(e^{bx}-1\right)\right)^{-\frac{1}{\theta}}\right]\right\}^{2}}\sum_{q=0}^{i-1}(-1)^{q}\binom{i-1}{q}\\
&\times\left(1+\frac{a\theta}{b}\left(e^{bx}-1\right)\right)^{-\frac{\nu}{\theta}}1-\left(\frac{p-1}{p}\right)\left[1-\left(1+\frac{a\theta}{b}\left(e^{bx}-1\right)\right)^{-\frac{1}{\theta}}\right]^{-\nu}\\
=&\frac{n!}{(i-1)!\,(n-i)!}\frac{a}{p}e^{bx}\sum_{q=0}^{i-1}(-1)^{q}\binom{i-1}{q}\left(1+\frac{a\theta}{b}\left(e^{bx}-1\right)\right)^{-\left(\frac{\nu+1}{\theta}+1\right)}\\
&\times\left\{1-\left(\frac{p-1}{p}\right)\left[1-\left(1+\frac{a\theta}{b}\left(e^{bx}-1\right)\right)^{-\frac{1}{\theta}}\right]\right\}^{-(\nu+2)}\\
=&\frac{n!}{(i-1)!(n-i)!}\frac{a}{p}e^{bx}\sum_{q=0}^{i-1}\sum_{k=0}^{\infty}(-1)^{q}\binom{i-1}{q}\binom{\nu+k+1}{k}\left(\frac{p-1}{p}\right)^{k}\\
&\times\left(1+\frac{a\theta}{b}\left(e^{bx}-1\right)\right)^{-\left(\frac{\nu+1}{\theta}+1\right)}\left[1-\left(1+\frac{a\theta}{b}\left(e^{bx}-1\right)\right)^{-\frac{1}{\theta}}\right]^{k}\\
=&\frac{n!}{(i-1)!\,(n-i)!}\frac{a}{p}e^{bx}\sum_{q=0}^{i-1}\sum_{k=0}^{\infty}\sum_{j=0}^{k}(-1)^{q+j}\binom{k}{j}\binom{i-1}{q}\binom{\nu+k+1}{k}\left(\frac{p-1}{p}\right)^{k}\\
&\times\left((1-\frac{a\theta}{b})+\frac{a\theta}{b}e^{bx}\right)^{-\vartheta}
\end{aligned}$$

Finally,

$$f_{i:n}(x) = \frac{n!}{(i-1)!\,(n-i)!}\,\frac{a}{p}\sum_{q=0}^{i-1}\sum_{k=0}^{\infty}\sum_{j=0}^{k}\sum_{l=0}^{\infty}(-1)^{q+j+l}\binom{i-1}{q}\binom{\nu+k+1}{k}\binom{k}{j}\binom{\vartheta+l-1}{l}$$
$$\times\left(\frac{p-1}{p}\right)^{k}\left(\frac{b}{a\theta}-1\right)^{l}\left(\frac{a\theta}{b}\right)^{-\vartheta}e^{-b(\vartheta+l-1)x}.$$

for $p > 1$, where $\nu = q+n-i$, $\vartheta = \frac{\nu+j+1}{\theta}+1$ and $\frac{a\theta}{b} > 0.5$.

4.7 Entropy

The entropy is a measure of uncertainty associated with a probability distribution of a random variable X. Several measures of entropy have been studied in the literature. Here, we consider Rényi entropy [28] and survival entropy [29]. Rényi entropy of order δ is defined by

$$H_\delta(X) = -\frac{1}{\delta-1}log\int_0^\infty f^\delta(x)\,dx, \quad \forall\delta>0,\ \delta\neq 1.$$

Rényi entropy tends to Shannon entropy as $\delta \to 1$. Similarly, using the survival function, the survival entropy of order δ is defined by

$$SE_\delta(X) = -\frac{1}{\delta-1}log\int_0^\infty \bar{F}^\delta(x)\,dx, \quad \forall\delta>0,\ \delta\neq 1.$$

Let $X \sim EGG(a,b,\theta,p)$, the corresponding Rényi entropy is obtained as

$$H_\delta(X) = \frac{p^\delta}{1-\delta}\log\left\{\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}(-1)^{k+j}\binom{\delta+k-1}{j}\binom{\omega^*+j-1}{j}(p-1)^k\left(\frac{b}{a\theta}-1\right)^j\left(\frac{a\theta}{b}\right)^{-\omega^*}\right\},$$

where $p\in(0,1)$, $\omega^* = \frac{k+\delta}{\theta}$ and $\frac{a\theta}{b} > 0.5$, or

$$H_\delta(X) = \frac{1}{1-\delta}\log\left\{\sum_{k=0}^{\infty}\sum_{j=0}^{k}\sum_{l=0}^{\infty}(-1)^{j+l}\binom{\delta+k-1}{j}\binom{k}{j}\binom{\psi^*+l-1}{l}\left(\frac{p-1}{p}\right)^k\left(\frac{b}{a\theta}-1\right)^l\left(\frac{a\theta}{b}\right)^{-\psi^*}\right\},$$

where $p > 1$, $\psi^* = \frac{j+\delta}{\theta}$ and $\frac{a\theta}{b} > 0.5$.

The survival entropy can be obtained using

$$SE_\delta(X) = \frac{p^\delta}{1-\delta}\log\left\{\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\frac{(-1)^{k+j}}{b(\omega^*+j)}\binom{\delta+k-1}{j}\binom{\omega^*+j-1}{j}(p-1)^k\left(\frac{b}{a\theta}-1\right)^j\left(\frac{a\theta}{b}\right)^{-\omega^*}\right\},$$

where $p\in(0,1)$, $\omega^* = \frac{k+\delta}{\theta}$ and $\frac{a\theta}{b} > 0.5$, or

$$SE_\delta(X) = \frac{1}{1-\delta}\log\left\{\sum_{k=0}^{\infty}\sum_{j=0}^{k}\sum_{l=0}^{\infty}\frac{(-1)^{j+l}}{b(\psi^*+l)}\binom{\delta+k-1}{j}\binom{k}{j}\binom{\psi^*+l-1}{l}\left(\frac{p-1}{p}\right)^k\left(\frac{b}{a\theta}-1\right)^l\left(\frac{a\theta}{b}\right)^{-\psi^*}\right\},$$

where $p > 1$, $\psi^* = \frac{j+\delta}{\theta}$ and $\frac{a\theta}{b} > 0.5$.

4.8 Stochastic order

This section presents the stochastic orders for EGG distribution. Let X and Y be two random variables distributed according to EGG distribution, that is $X \sim MOGG(a_1, b_1, \theta_1, p_1)$ and $Y \sim MOGG(a_2, b_2, \theta_2, p_2)$, with corresponding cdfs F and G, respectivel, and $f,\ g$ their respective pdfs. We say that X is stochastically smaller than Y in the likelihood ratio order ($X \leq_{lr} Y$) if $f(x)/g(x)$ is decreasing function of x. Likelihood ratio order implies hazard rate order ($X \leq_{hr} Y$) which in turn implies usual stochastic order ($X \leq_{st} Y$), for further detail on stochastic orders, see (Shaked and Shanthikumar, 1995).

If $a_1 = a_2 = a,\ b_1 = b_2 = b$ and $\theta_1 = \theta_2 = \theta$, then ($X \leq_{lr} Y$) if and only if $p_1 < p_2$, this can be proved as follows, let

$$K\left(x\right) = \frac{f(x)}{g(x)} = \frac{p_2(1-(1-p_1)\left(1+\ \frac{a\theta}{b}\left(e^{bx}-1\right)\right)^{-\frac{1}{\theta}})^2}{p_1(1-(1-p_2)\left(1+\ \frac{a\theta}{b}\left(e^{bx}-1\right)\right)^{-\frac{1}{\theta}})^2}$$

and is such that

$$K'\left(x\right) = \frac{2ae^{bx}(p_2-p_1)p_2\left(1+\ \frac{a\theta}{b}\left(e^{bx}-1\right)\right)^{-1-\frac{1}{\theta}}(1-(1-p_1)\left(1+\ \frac{a\theta}{b}\left(e^{bx}-1\right)\right)^{-\frac{1}{\theta}})}{p_1(1-(1-p_2)\left(1+\ \frac{a\theta}{b}\left(e^{bx}-1\right)\right)^{-\frac{1}{\theta}})^3} < 0,$$

if and only if $p_1 < p_2$.

5 Estimation

This section consider the maximum likelihood estimations (MLEs) to estimate and derive the asymptotic confidence intervals of the unknown parameter vector $\boldsymbol{\theta} = (a, b, \theta, p)$ of the EGG distribution. Let $x_1, \ldots, x_n$ be a random sample of size n from EGG distribution with pdf (1), then the log-likelihood function can be written as

$$\begin{aligned}\ell(\boldsymbol{\theta}) =& n\log(ap) + b\sum_{i=1}^{n} x_i + \left(\frac{1}{\theta}-1\right)\sum_{i=1}^{n}\log(1+\frac{a\theta}{b}\left(e^{bx_i}-1\right)) \\ &-2\sum_{i=1}^{n}\log\left[p-1+\left(1+\frac{a\theta}{b}\left(e^{bx_i}-1\right)\right)^{\frac{1}{\theta}}\right]\end{aligned} \tag{5.1}$$

The maximum likelihood estimates (MLEs) of $\boldsymbol{\theta} = (a, b, \theta, p)$, say $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{b}, \hat{\theta}, \hat{p})$, are obtained from maximizing the log-likelihood function (5.1). The log-likelihood function is maximized by solving the nonlinear system $\frac{\partial \ell}{\partial a} = 0,\ \frac{\partial \ell}{\partial b} = 0,\ \frac{\partial \ell}{\partial \theta} = 0,\ \frac{\partial \ell}{\partial p} = 0$, simultaneously. These equations can not be solved analytically and must be solved numerically via iterative methods such as Newton Raphson technique via mathematical software, i.e. Mathematica (FindRoot function), R (rootSolve package).

Alternatively, the log-likelihood functions (5.1) can be maximized directly using a mathematical software, i.e. Mathematica (NMaximize and FindMaximum functions), R (optim and MaxLik functions).

Confidence intervals (CIs) for the parameters were based on asymptotic normality. Its well known that the distribution of $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ can be approximated by a 4−variate normal distribution with zero means and covariance matrix $I^{-1}(\hat{\boldsymbol{\theta}})$, where $I(\boldsymbol{\theta})$ is the observed information matrix defined by

$$I(\boldsymbol{\theta}) = -\begin{pmatrix} \mathrm{L}_{aa} & \mathrm{L}_{ab} & \mathrm{L}_{a\theta} & \mathrm{L}_{ap} \\ \mathrm{L}_{ba} & \mathrm{L}_{bb} & \mathrm{L}_{b\theta} & \mathrm{L}_{bp} \\ \mathrm{L}_{\theta a} & \mathrm{L}_{\theta b} & \mathrm{L}_{\theta\theta} & \mathrm{L}_{\theta p} \\ \mathrm{L}_{pa} & \mathrm{L}_{pb} & \mathrm{L}_{p\theta} & \mathrm{L}_{pp} \end{pmatrix},$$

where the components of $I(\boldsymbol{\theta})$ are the second derivative of the log-likelihood function with respect to $a, b,\ \theta$ and p. The asymptotic $100(1-\alpha)\%$ confidence interval for the parameters a, b, θ and p are $\hat{a} \pm \mathcal{Z}_{\frac{\alpha}{2}} \sqrt{Var(\hat{a})}$, $\hat{b} \pm \mathcal{Z}_{\frac{\alpha}{2}} \sqrt{Var(\hat{b})}$, $\hat{a} \pm \mathcal{Z}_{\frac{\alpha}{2}} \sqrt{Var(\widehat{\theta})}$ and $\hat{a} \pm \mathcal{Z}_{\frac{\alpha}{2}} \sqrt{Var(\hat{p})}$ respectively, where $Var\left(\hat{a}\right),\ Var\left(\hat{b}\right),\ Var(\widehat{\theta})$ and $Var(\hat{p})$ are the diagonal elements of $I^{-1}\left(\widehat{\boldsymbol{\theta}}\right)$ corresponding to each parameter and $\mathcal{Z}_{\frac{\alpha}{2}}$ is the upper $(\frac{\alpha}{2})$ percentile of standard normal distribution.

6 Simulation

In this section, a simulation study is performed to generate random samples with different sizes from the EGG distribution and investigate the performance of the EGG distribution using Mont-Carlo Method.

Table 2. Monte carlo simulation results: MSE, AB, CP and AW

n	$a = 0.2$				$b = 2.5$				$\theta = 0.25$				$p = 1.7$			
	AB	MSE	CP	AW	AB	MSE	CP	AW	AB	MSE	CP	AW	AB	MSE	CP	AW
25	0.0032	0.0019	0.9752	0.7847	0.0112	0.0400	0.9404	6.0352	-0.0172	0.0229	0.9798	4.6655	0.1223	0.571	0.931	11.640
50	0.0026	0.0010	0.9494	0.8214	0.0107	0.0215	0.9564	4.0340	-0.012	0.0118	0.966	2.4624	0.0634	0.1976	0.942	10.059
75	0.0021	0.0007	0.942	0.8960	0.0094	0.0146	0.947	3.2880	-0.0097	0.0077	0.958	1.7208	0.037	0.1446	0.944	10.349
100	0.0014	0.0005	0.945	0.826	0.0067	0.011	0.951	2.7142	-0.0073	0.0059	0.948	1.3442	0.0323	0.0949	0.956	9.4725
n	$a = 1.6$				$b = 0.9$				$\theta = 1.4$				$p = 0.8$			
	AB	MSE	CP	AW	AB	MSE	CP	AW	AB	MSE	CP	AW	AB	MSE	CP	AW
25	0.1000	0.396	0.924	2.3940	0.1407	0.203	0.989	4.9706	0.0224	0.1903	0.927	5.7504	0.0128	0.2232	0.98	2.2785
50	0.0448	0.2065	0.925	1.8892	0.081	0.1037	0.976	3.8913	0.0150	0.1035	0.938	3.7940	0.0274	0.0409	0.982	1.3550
75	0.0308	0.1368	0.92	1.7384	0.0593	0.0675	0.972	3.4194	0.0145	0.0701	0.932	3.0480	0.0170	0.0317	0.988	1.0363
100	0.0193	0.0990	0.922	1.6778	0.0446	0.0487	0.968	3.0427	0.0103	0.0539	0.939	2.9136	0.0152	0.021	0.988	1.0697
n	$a = 5$				$b = 8$				$\theta = 5$				$p = 7$			
	AB	MSE	CP	AW	AB	MSE	CP	AW	AB	MSE	CP	AW	AB	MSE	CP	AW
25	0.2442	4.4493	0.9816	23.8131	0.1842	1.0143	0.968	28.0202	-0.0469	0.3003	0.988	21.0524	0.0738	2.5601	0.986	18.249
50	0.2309	4.369	0.943	20.6315	0.1016	0.495	0.954	21.336	-0.0268	0.1538	0.978	16.2769	0.0827	1.4583	0.952	12.5711
75	0.2056	4.116	0.918	15.7889	0.0572	0.3333	0.972	16.3805	-0.0109	0.1065	0.974	12.6537	0.0892	1.1177	0.943	9.7176
100	0.1754	3.725	0.8902	19.0455	0.0423	0.2635	0.964	9.2058	-0.0065	0.0853	0.962	11.4626	0.0854	0.9262	0.931	6.9367

The simulation study is repeated $N = 5000$ times each with sample size $n = 25,\ 50,\ 75$ and 100. The following scenarios of the parameter vector $\boldsymbol{\theta} = (a, b, \theta, p)$ are considered: (0.2, 2.5, 0.25, 1.7), (1.6, 0.9, 1.4, 0.8) and (5, 8, 5, 7). These selected values of $\boldsymbol{\theta}$ gives unimodal, decreasing and decreasing-increasing-decreasing shapes for pdf of the EGG distribution. Four Quantities were examined in this Mont-Carlo study:

i) Average bias (AB) of the MLE $\widehat{\vartheta}$ of the parameter $\vartheta = a, b, \theta, p$:

$$\frac{1}{N} \sum_{i=1}^{N} (\widehat{\vartheta} - \vartheta)$$

ii) Mean square error (MSE) of the MLE $\widehat{\vartheta}$ of the parameter $\vartheta = a, b, \theta, p$:

$$\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\vartheta}-\vartheta\right)^2$$

iii) Coverage probability (CP) of 95% confidence intervals of the parameter $\vartheta = a, b, \theta, p$, i.e., the percentage of intervals containing the true value of ϑ.

iv) Average width (AW) of 95% confidence intervals of the parameter $\vartheta = a, b, \theta, p$.

Table 2. presents the results of Mont-Carlo simulation study for EGG distribution. From the results, we can verify that as the sample size increases, the MSEs decay toward zero. We also observe that for all the parametric values, the ABs decrease as the sample size n increases. Also, the table shows that the coverage probabilities of the confidence intervals are quite close to the nominal level of 95% and that the average widths of the confidence intervals decrease as the sample size increases. Consequently, the MLEs and their asymptotic results can be used for estimating and constructing confidence intervals even for reasonably small sample sizes.

7 Application

This section is devoted to demonstrate the usefulness of EGG distribution by fitting a real data set and comparing with its sub-models and other well competitive models. The pdf of the distributions used with this comparison is as follows:

- Gamma Gompertz distribution, $GG(a, b, \theta)$: $ae^{bx}\left(\frac{(a\theta)\left(e^{bx}-1\right)}{b}+1\right)^{-\frac{\theta+1}{\theta}}$
- Marshal Olken extended Gompertz distribution, $MOG(a, b, p)$: $\frac{ape^{bx}e^{-\frac{a\left(e^{bx}-1\right)}{b}}}{\left(1-(1-p)e^{-\frac{a\left(e^{bx}-1\right)}{b}}\right)^2}$
- Marshal Olken extended exponential distribution, $MOE(\lambda, p)$: $\frac{\lambda pe^{-\lambda x}}{\left(1-(1-p)e^{-\lambda x}\right)^2}$
- Gompertz distribution, $G(a, b)$: $ae^{bx}e^{-\frac{a\left(e^{bx}-1\right)}{b}}$
- Exponential distribution, $E(\lambda)$: $\lambda(e^{(}-\lambda x))$
- Marshal Olken extended generalized linear exponential distribution,

  $MOGLE(a, b, p, \theta)$: $\frac{p\theta\lambda(ax+b)e^{-\left(\frac{ax^2}{2}+bx\right)^{\theta}}\left(\frac{ax^2}{2}+bx\right)^{\theta-1}}{\left(1-(1-p)e^{-\left(\frac{ax^2}{2}+bx\right)^{\theta}}\right)^2}$
- Marshal Olken extended Weibull distribution, [30], $MOW(\lambda, \alpha, p)$: $\frac{\alpha\lambda px^{\alpha-1}e^{-\lambda x^{\alpha}}}{\left(1-(1-p)e^{-\lambda x^{\alpha}}\right)^2}$
- Alpha power Gompertz distribution, $APG(a, b, \alpha)$: $\frac{a\log(\alpha)}{\alpha-1}e^{bx-\frac{a\left(e^{bx}-1\right)}{b}}\alpha^{1-e^{-\frac{a\left(e^{bx}-1\right)}{b}}}$

where $x > 0; a, b, \alpha, \lambda, \theta, p > 0$.

The data set, considered here, represents the lifetimes of 50 devices, [31]. In order to identify the hazard shape for the given data set, the total time on test (TTT) plot is used, Aarset [31]. The TTT plot is drawn by plotting

$$T(i/n) = \frac{\sum_{k=1}^{i} X_{k:n} + (n-i)X_{i:n}}{\sum_{k=1}^{n} X_{k:n}}$$

against

$$i/n$$

Fig. 3. showed that the TTT plot is first convex and then concave, therefor the hazard rate is bathtub-shaped. So, the EGG distribution is suitable for Aarset data.

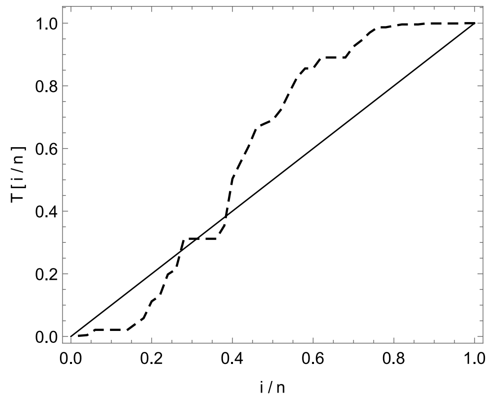


Fig. 3. TTT plot for Aarset data

To select the best distribution for the used data set, the following criteria are calculated: −log-likelihood, Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC), Bayesian Information Criterion (BIC), Anderson-Darling statistic (A^*), Cramer-von Mises statistic (W^*), Kolmogorov-Smirnov (K-S) distances between the empirical distribution function and the fitted distribution function and corresponding P-value.

In addition, three graphical goodness of fit methods were used: *i*) Empirical and fitted pdfs curves; *ii*) Kaplan-Meier and fitted survival curves; *iii*) quantile-quantile (QQ) plots for the fitted distributions.

Table 3. presents estimates of the parameters of the fitted distributions and the statistics used fr the comparison. The results in this table show that the EGG distribution has the smallest AIC, BIC, CAIC, A^*, W^* and K-S values, which prove the superiority of the EGG distribution.

Fig. 4. shows the empirical and fitted pdfs in the left panel, whereas the right panel of this figure shows Kaplan-Meier curve and the estimated survival functions. The fitted pdf and survival function of EGG distribution are the closest to the empirical density and Kaplan-Meier curve, respectively.

These result are further investigated by showing the quantile-quantile (QQ) plots for the fitted distributions in fig. 5. The plotted points of the QQ plot for the EGG distribution are closer to the diagonal line than of the other models.

The likelihood ratio test (LRT) is performed to demonstrate the superiority of EGG to its sub-models. We tested the null hypotheses, presented in table 5, against the alternative hypothesis $H_a = EGG(a, b, \theta, p)$. Furthermore, table 5. shows the likelihood ratio test statistics (Λ) and the corresponding P-values which show that all the null hypotheses are rejected.

The asymptotic variance covariance matrix of MLEs for the EGG model parameters is given by

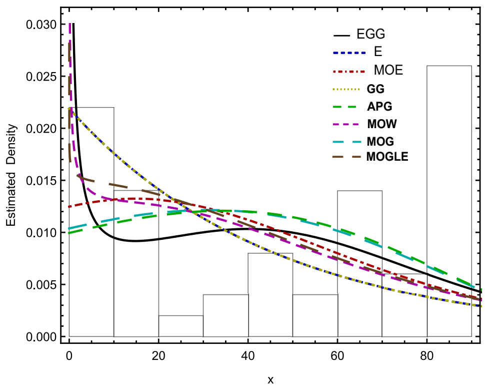
$$I^{-1}\left(\hat{a}, \hat{b}, \hat{p}, \hat{\theta}\right) = \begin{pmatrix} 187.517 & -0.2711 & 421.941 & -8.7704 \\ -0.2712 & 0.0029 & -1.2798 & 0.0701 \\ 421.941 & -1.2798 & 1310.15 & -39.7641 \\ -8.7704 & 0.0701557 & -39.7641 & 1.8775 \end{pmatrix}$$

and the 95% two sided asymptotic confidence interval for a, b, p and θ are given by 6.1802 ± 26.8396, 0.1238 ± 0.1065, 26.4448 ± 70.9442 and 3.1343 ± 2.6856, respectively.

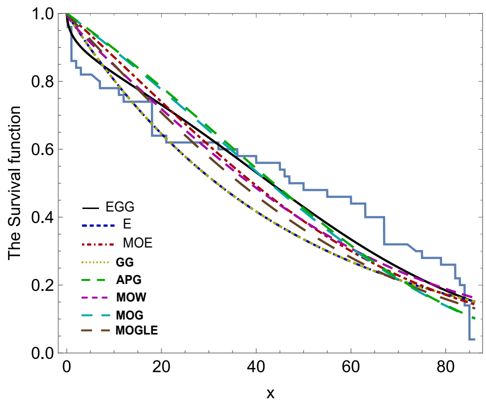
The profile of the log-likelihood function of the parameters a, b, p and θ is plotted in fig. 6. to show that the likelihood equations have a unique solution.

Table 3. AIC, BIC, CAIC, A^* , W^* and K-S and its corresponding P-Value

Model	Estimates	$-\ell$	AIC	BIC	CAIC	A^*	W^*	K-S	P-Value
$EGG(a,b,p,\theta)$	$\hat{a}$ =6.1802; $\hat{b}$ =0.1238; $\hat{p}$ =26.4448; $\hat{\theta}$ =3.1343	230.541	469.083	476.731	469.972	1.7134	0.2451	0.1515	0.2013
$GG(a,b,\theta)$	$\hat{a}$ =0.0219; $\hat{b}$ =0.05486; $\hat{\theta}$ =2.5062	241.09	488.179	493.915	488.701	3.6501	0.5189	0.191	0.052
$MOG(a,b,p)$	$\hat{a}$ =0.0077; $\hat{b}$ =0.0224; $\hat{p}$ =0.741	235.241	476.483	482.219	477.005	4.6178	0.4688	0.1598	0.1557
$MOE(\lambda,p)$	$\hat{\lambda}$ =2.6215; $\hat{p}$ =0.0320	239.554	483.108	486.93	483.363	4.0525	0.4331	0.163	0.1402
$G(a,b)$	$\hat{a}$ =0.0097; $\hat{b}$ =0.0203	235.331	474.662	478.486	474.917	4.8035	0.4675	0.1697	0.1123
$E(\lambda)$	$\hat{\lambda}$ =0.0219	241.09	486.179	490.003	486.435	3.6501	0.5188	0.191	0.052
$MOGLE(a,b,p,\theta)$	$\hat{a}$ =0.00026; $\hat{b}$ =0.0099; $\hat{p}$ =0.9558; $\hat{\theta}$ =0.7364	238.017	484.034	491.683	484.923	3.7793	0.4654	0.1809	0.0759
$MOW(\lambda,\alpha,p)$	$\hat{\lambda}$ =0.1585; $\hat{\alpha}$ =0.6992; $\hat{p}$ =6.6973	237.723	481.447	487.183	481.968	2.9561	0.3720	0.1626	0.1421
$APG(a,b,\alpha)$	$\hat{a}$ =0.0089; $\hat{b}$ =0.0211; $\hat{\alpha}$ =0.8058	-235.299	476.597	482.333	477.119	4.7607	0.4692	0.1661	0.1267



(a)



(b)

Fig. 4. The estimated densities and survival curves for Aarset data

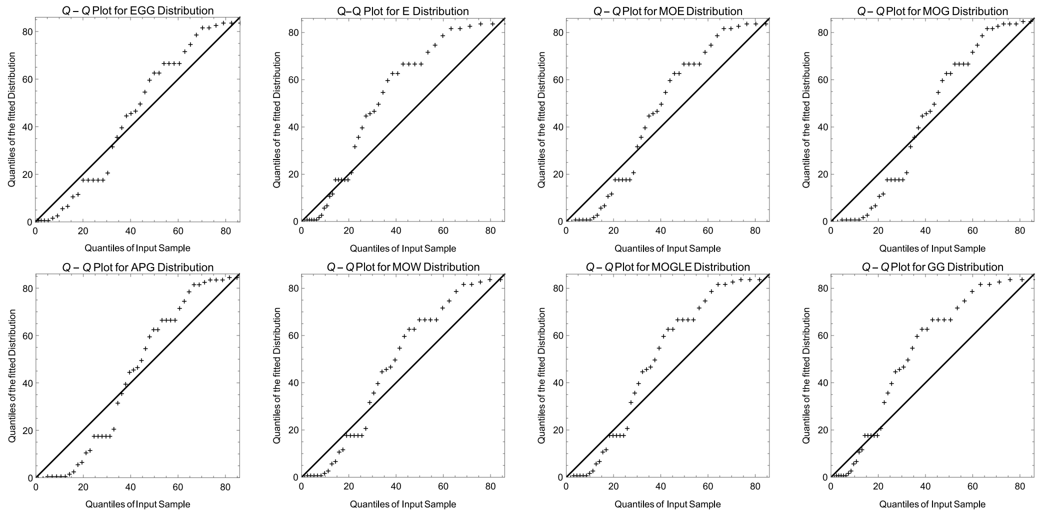


Fig. 5. The QQ plots for the fitted distributions

Tabe 5. Likelihood ratio test statistics and corresponding P-Value

The model	H_0	Λ	d.f.	P-Value
GG	$EGG(a, b, \theta, 1)$	21.089	1	< 0.0000
MOE	$EGG(b, b, 1, p)$	18.026	2	< 0.0000
E	$EGG(b, b, 1, 1)$	21.089	3	< 0.0000

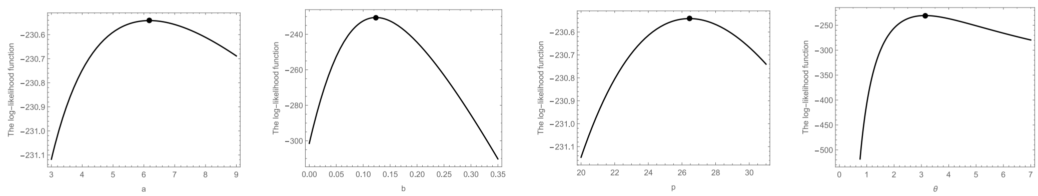


Fig. 6. The profile of the log-likelihood function

8 Concluding Remarks and Future Research

This paper proposed a new distribution called Extended Gamma Gompertz distribution based on two key concepts in survival analysis, frailty models and competing risks problem. The new distribution considers the possible heterogeneity in the data set. Furthermore, the number of competing causes of failure is modeled by the geometric distribution. Some properties have been derived for the EGG distribution. The maximum likelihood method is used to estimate the unknown parameters. A simulation study is performed to examine the accuracy of the maximum likelihood estimates. A real data set is used to explore the superiority of the proposed model to its sub-models and other distributions.

As part of future research, different distributions for the frailty random variable can be used, i.e. positive stable distribution, [15]. Furthermore, the number of competing causes of failure can be modeled by the power series distribution which absorbs as particular cases the Poisson, logarithmic, geometric, binomial and negative binomial allowing a wide range of models to be considered. Further discussion on the impact of the parameters on the performance of the proposed models will be proposed.

Competing Interests

The authors declare that no competing interests exist.

References

[1] Gompertz B. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Philosophical transactions of the Royal Society of London. 1825;(115):513-583.

[2] El-Gohary A, Alshamrani A, Al-Otaibi A. The generalized Gompertz distribution. Applied Mathematical Modelling. 2013;37(1-2):13-24.

[3] Nadarajah S, Kotz S. The exponentiated type distributions. Acta Applicandae Mathematica. 2006;92(2):97-111.

[4] Gupta RC, Gupta P, Gupta R. Modeling failure time data by lehman alternatives. Communications in Statistics-Theory and methods. 1998;27(4):887-904.

[5] Jafari A, Tahmasebi S, Alizadeh M. The beta-Gompertz distribution. Revista Colombiana de Estadistica. 2014;37(1):141-158.

[6] Eugene N, Lee C, Famoye F. Beta-normal distribution and its applications. Communications in Statistics-Theory and methods. 2002;31(4):497-512.

[7] Jafari AA, Tahmasebi S. Gompertz-power series distributions. Communications in Statistics - Theory and Methods. 2016;45(13):3761-3781.

[8] Marshall AW, Olkin I. A new method for adding a parameter to a family of distributions with application to the exponential and weibull families. Biometrika. 1997;84(3):641-652.

[9] Abdul-Moniem I, Seham M. Transmuted gompertz distribution. Comput Appl Math J. 2015;1(3):88-96.

[10] Shaw W, Buckley I. The alchemy of probability distributions: Beyond gram-charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map; 2009. arXiv preprint arXiv:0901.0434

[11] Roozegar R, Tahmasebi S, Jafari A. The mcdonald Gompertz distribution: Properties and applications. Communications in Statistics-Simulation and Computation. 2017;46(5):3341-3355.

[12] Alexander C, Cordeiro G, Ortega E, Sarabia J. Generalized beta-generated distributions. Computational Statistics & Data Analysis. 2012;56(6):1880-1897.

[13] Eghwerido JT, Nzei LC, Agu FI. The alpha power gompertz distribution: Characterization, properties, and applications. Sankhya A. 2020;83:1-27.

[14] Mahdavi A, Kundu D. A new method for generating distributions with an application to exponential distribution. Communications in Statistics-Theory and Methods. 2017;46(13):6543-6557.

[15] Hougaard P. Analysis of multivariate survival data. Springer Science & Business Media; 2012.

[16] Duchateau L, Janssen P. The frailty model. Springer Science & Business Media; 2007.

[17] Finkelstein M, Esaulova V. On asymptotic failure rates in bivariate frailty competing risks models. Statistics & Probability Letters. 2008;78(10):1174-1180.

[18] Basu AP, Klein JP. Some recent results in competing risks theory. Lecture Notes-Monograph Series. 1982;2:216-229

[19] Adamidis K, Loukas S. A lifetime distribution with decreasing failure rate. Statistics & Probability Letters. 1998;39(1):35-42.

[20] Kuş C. A new lifetime distribution. Computational Statistics & Data Analysis. 2007;51(9):4497-4509.

[21] Borgonovo E, Marseguerra M, Zio E. A monte carlo methodological approach to plant availability modeling with maintenance, aging and obsolescence. Reliability Engineering & System Safety. 2000;67(1):61-73.

[22] Nguyen T, Yeung T, Castanier B. Optimal maintenance and replacement decisions under technological change with consideration of spare parts inventories. International Journal of Production Economics. 2013;143(2):472-477.

[23] Bemmaor AC, Glady N. Modeling purchasing behavior with sudden death: A flexible customer lifetime model. Management Science. 2012;58(5):1012-1021.

[24] Mazucheli J, Menezes AF, Dey S. Unit-gompertz distribution with applications. Statistica. 2019;79(1):25-43.

[25] Graham R, Knuth D, Patashnik O, Liu S. Concrete mathematics: A foundation for computer science. Computers in Physics. 1989;3(5):106-107.

[26] Abramowitz M, Stegun IA. Handbook of mathematical functions: With formulas, graphs, and mathematical tables. National Bureau of Standards Applied Mathematics Series 55. Tenth Printing. ERIC; 1972.

[27] Gupta RC, Bradley DM. Representing the mean residual life in terms of the failure rate. Mathematical and Computer Modelling. 2003;37(12-13):1271-1280.

[28] Rényi A. On measures of entropy and information. Tech. rep, Hungarian Academy of Sciences. Budapest Hungary; 1961.

[29] Zografos K, Nadarajah S. Survival exponential entropies. IEEE Transactions on Information Theory. 2005;51(3):1239-1246.

[30] Barreto-Souza W, de Morais AL, Cordeiro GM. The weibull-geometric distribution. Journal of Statistical Computation and Simulation. 2011;81(5):645-657.

[31] Aarset MV. How to identify a bathtub hazard rate. IEEE Transactions on Reliability. 1987;36(1):106-108.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://www.sdiarticle4.com/review-history/67000