Original Research Article

A Weibull-Gompertz Makeham Distribution with Properties and Application to Cancer Data

Abstract

The article presents an extension of the Gompertz Makeham distribution using the Weibull-G family of continuous probability distributions proposed by Tahir *et al*. (2016a). This new extension generates a more flexible model called Weibull-Gompertz Makeham distribution. Some statistical properties of the distribution which include the moments, survival function, hazard function and distribution of order statistics were derived and discussed. The parameters were estimated by the method of maximum likelihood and the distribution was applied to a bladder cancer data. Weibull-Gompertz Makeham distribution performed best (AIC = -6.8677, CAIC = -6.3759, BIC = 7.3924) when compared with other existing distributions of the same family to model bladder cancer data.

Keywords: Gompertz-Makeham Distribution, Weibull-Gompertz Makeham Distribution, hazard function, survival function, cancer.

1 Introduction

The Gompertz-Makeham distribution (GMD) was introduced by Makeham in 1860 (Makeham (1860)). It is an extended model of the Gompertz probability distribution that was introduced by Gompertz in 1825 (Gompertz (1825)). The GMD is a continuous probability distribution that has been widely used in survival analysis, modelling human mortality, constructing actuarial tables and growth models. It has been recently used in many fields of sciences including actuaries, biology, demography, gerontology, and computer science.

A comprehensive review of the history and theory of the GMD can be found in Marshall and Olkin (2007). Golubev (2004) emphasizes the practical importance of this probability distribution. Detailed information about the GM distribution, its mathematical and statistical properties, and its applications can be found in Johnson et al. (1995) and Dey et al. (2018).
The cumulative distribution function (cdf) and probability density function (pdf) of the Gompertz-Makeham distribution are defined as:

$$G(x)=1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)} \qquad (1)$$

and

$$g(x) = \left[\theta + \alpha e^{\beta x}\right] e^{-\theta x - \frac{\alpha}{\beta}\left(e^{\beta x} - 1\right)} \quad (2)$$

respectively.

For $x > 0, \alpha, \beta, \theta > 0,$ where θ is the scale parameter and α and β are the shape parameters of Gompertz-Makeham distribution.

There are families of distributions proposed by different researchers that are used in extending other distributions to produce compound distributions with better performance. These families among others include the beta generalized family (Beta-G) by Eugene et al. (2002), the Kumaraswamy-G by Cordeiro and de Castro (2011), Transmuted family of distributions by Shaw and Buckley (2007), Exponentiated T-X by Alzaghal et al. (2013), Exponentiated-G (EG) by Cordeiro et al. (2013), Logistic-G by Torabi and Montazari (2014), Logistic-X by Tahir et al. (2016b), Weibull-X by Alzaatreh et al. (2013), Weibull-G by Bourguignon et al. (2014), a new Weibull-G family by Tahir et al. (2016a), a Lomax-G family by Cordeiro et al. (2014), a new generalized Weibull-G family by Cordeiro et al. (2015) and Beta Marshall-Olkin family of distributions by Alizadeh et al. (2015) etc.

Recently, many authors have extended the Gompertz-Makeham distribution. Chukwu and Ogunde (2016) introduced and studied the Kumaraswamy Gompertz Makeham distribution. El-Bar (2017) used the quadratic rank transmutation map by Shaw and Buckley (2007) to defined and study the transmuted Gompertz Makeham distribution with useful discussions as well as applications.

Hence, the aim of this article is to introduce another extension of the Gompertz Makeham model, a new continuous distribution called Weibull-Gompertz Makeham distribution (WGMD) from the proposed family by Tahir et al. (2016a). The rest of this article is arranged as follows: the definition of the new distribution will be presented with its plots and some properties. These are followed by the reliability functions, the order statistics for the distribution and the maximum likelihood estimates (MLEs) of the unknown parameters. The last part involves the application of the proposed model with other models to a lifetime dataset and the conclusion.

2 Materials and Method

2.1 Construction of Weibull-Gompertz Makeham Distribution (WGMD)

This section defines the cdf and pdf of the Weibull-Gompertz Makeham distribution (WGMD) using the family of distributions proposed by Tahir et al. (2016a), which has been used by other authors including Ieren et al. (2018). According to Tahir et al. (2016a), the function for defining the cdf and pdf of any Weibull-based continuous distribution is given as:

$$F(x)=\int_{0}^{-\log\left[G(x)\right]} abt^{b-1}e^{-at^{b}}dt=e^{-a\left\{-\log\left[G(x)\right]\right\}^{b}} \qquad (3)$$

and

$$f(x)=ab\frac{g(x)}{G(x)}\left\{-\log\left[G(x)\right]\right\}^{b-1}e^{-a\left\{-\log\left[G(x)\right]\right\}^{b}} \qquad (4)$$

respectively, where $g(x)$ and $G(x)$ are the pdf and cdf of any continuous distribution to be generalized respectively. The parameters, α and β are the two additional new parameters responsible for the scale and shape of the distribution respectively.

Using equation (1) and (2) in (3) and (4) and simplifying, the cdf and pdf of the WGMD of a random variable X can be obtained as:

$$F(x)=e^{-a\left\{-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right\}^{b}} \qquad (5)$$

and

$$f(x)=\frac{ab\left(\theta+\alpha e^{\beta x}\right)e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{b-1}}{\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]e^{a\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{b}}} \qquad (6)$$

respectively.

For $x>0; a,b,\theta,\alpha,\beta>0$; where a, b, θ, α and β are the parameters of the WGMD.

The following is a graphical representation of the pdf and cdf of the WGMD using arbitrary values of the parameters a, b, θ, α and β.

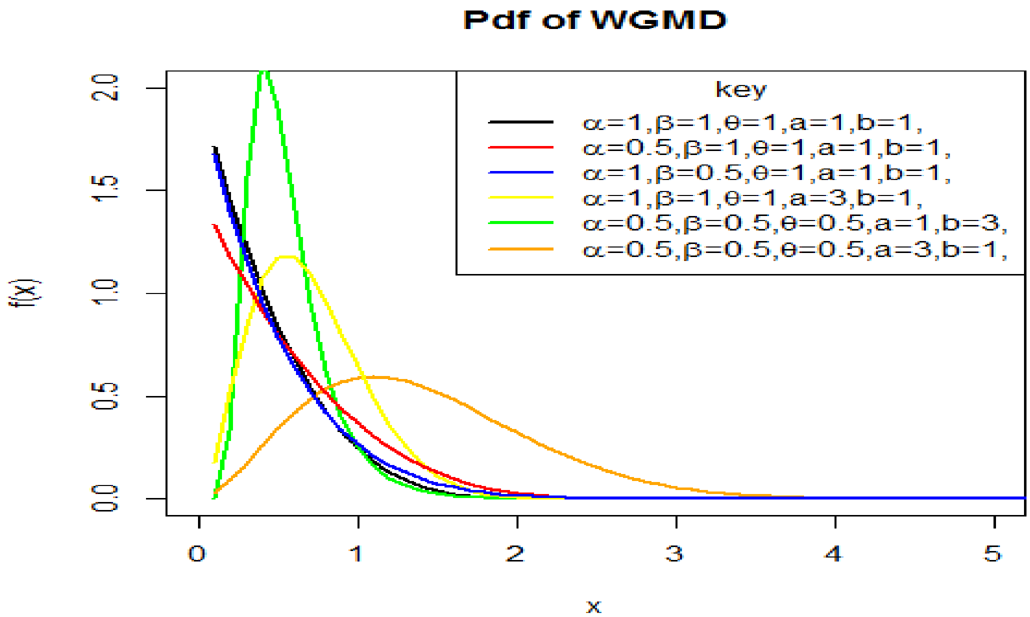

Figure 1: A plot of PDF of the WGMD for varying parameter values

It is observed in Figure 1 that the WGMD is a positively skewed distribution and can take various forms. This means that distribution can be very useful for datasets that are skewed.

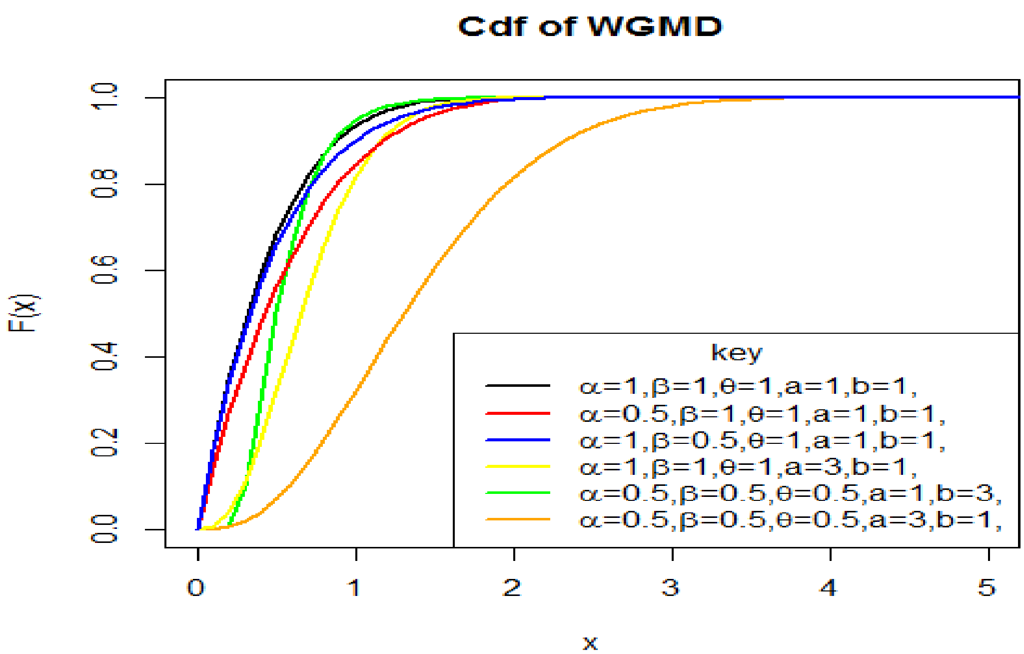

Figure 2: A plot of CDF of the WGMD for varying parameter values

From the above cdf plot, the cdf increases when X increases, and approaches 1 when X becomes large or tends to infinity as expected.

3 Properties

In this section, we defined and discuss some properties of the WGMD distribution.

3.1 Moments

Let X denote a continuous random variable, the nth moment of X is given by;

$$\mu_n^{'} = E\left[X^n\right] = \int_0^\infty x^n f(x)dx \tag{7}$$

Considering f(x) to be the pdf of the WGMD as given in equation (6)

$$\mu_n^{'} = E\left[X^n\right] = \int_0^\infty x^n f(x)dx$$

Recall that from equation (6),

$$f(x) = \frac{ab\left(\theta + \alpha e^{\beta x}\right)e^{-\theta x - \frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\left(-\log\left[1-e^{-\theta x - \frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{b-1}}{\left[1-e^{-\theta x - \frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]e^{a\left(-\log\left[1-e^{-\theta x - \frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^b}} \tag{8}$$

To simplify the pdf in (8) above, we carryout the following operations:

Let

$$A = e^{-a\left(-\log\left[1-e^{-\theta x - \frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^b}$$

Then, using a power series expansion for A, we can write A as:

$$A = e^{-a\left(-\log\left[1-e^{-\theta x - \frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^b} = \sum_{i=0}^{\infty}\frac{(-1)^i a^i}{i!}\left(-\log\left[1-e^{-\theta x - \frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{bi}$$

Substituting for the expansion above in equation (8), we have;

$$f(x) = \frac{ab\left(\theta + \alpha e^{\beta x}\right)e^{-\theta x - \frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\sum_{i=0}^{\infty}\frac{(-1)^i a^i}{i!}\left(-\log\left[1-e^{-\theta x - \frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{bi}}{\left[1-e^{-\theta x - \frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\left(-\log\left[1-e^{-\theta x - \frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{-(b-1)}}$$

$$f(x)=\frac{ab\left(\theta+\alpha e^{\beta x}\right)e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\sum_{i=0}^{\infty}\frac{(-1)^{i}a^{i}}{i!}\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{bi+(b-1)}}{\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]} \tag{9}$$

$$B=\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{bi+(b-1)}$$

Also, let

Now, considering the following formula from Tahir et al. (2016b) and Ieren et al. (2018) which holds for B for i≥1, then B can be written as follows:

$$\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{bi+(b-1)}=\sum_{k,l=0}^{\infty}\sum_{j=0}^{k}\frac{(-1)^{j+k+l}\left(b(i+1)\right)}{\left(b(i+1)-1-j\right)}\binom{k-\left(b(i+1)-1\right)}{k}\binom{k}{j}\binom{\left(b(i-1)+1\right)+k}{l}P_{j,k}\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]^{l} \tag{10}$$

Where for (for j ≥ 0) $P_{j,0}=1$ and (for k = 1, 2, …..)

$$P_{j,k}=k^{-1}\sum_{m=1}^{k}(-1)^{m}\frac{\left[m(j+1)-k\right]}{(m+1)}P_{j,k-m} \tag{11}$$

Combining equation (10) and (11) and inserting the above power series in equation (9) and simplifying, it gives:

$$f(x)=ab\sum_{i=0}^{\infty}\frac{(-1)^{i}a^{i}}{i!}\sum_{k,l=0}^{\infty}\sum_{j=0}^{k}\frac{(-1)^{j+k+l}\left(b(i+1)\right)}{\left(b(i+1)-1-j\right)}\binom{k-\left(b(i+1)-1\right)}{k}\binom{k}{j}\binom{\left(b(i+1)-1\right)+k}{l}P_{j,k}\left(\theta+\alpha e^{\beta x}\right)e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]^{l-1}$$

$$f(x)=b\sum_{i=0}^{\infty}\sum_{k,l=0}^{\infty}\sum_{j=0}^{k}\frac{(-1)^{i+j+k+l}a^{i+1}\left(b(i+1)\right)}{i!\left(b(i+1)-1-j\right)}\binom{k-\left(b(i+1)-1\right)}{k}\binom{k}{j}\binom{\left(b(i+1)-1\right)+k}{l}P_{j,k}\left(\theta+\alpha e^{\beta x}\right)e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]^{l-1} \tag{12}$$

Now, if l is a positive non-integer, we can expand the last term in (12) as:

$$\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]^{l-1}=\sum_{m=0}^{\infty}(-1)^{m}\binom{l-1}{m}\left[e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]^{m} \tag{13}$$

Therefore, f(x) becomes:

$$f(x)=b\sum_{m=0}^{\infty}\sum_{i=0}^{\infty}\sum_{k,l=0}^{\infty}\sum_{j=0}^{k}\frac{(-1)^{i+j+k+l+m}a^{i+1}\left(b(i+1)\right)}{i!\left(b(i+1)-1-j\right)}\binom{l-1}{m}\binom{k-\left(b(i+1)-1\right)}{k}\binom{k}{j}\binom{\left(b(i+1)-1\right)+k}{l}P_{j,k}\left(\theta+\alpha e^{\beta x}\right)e^{-(m+1)\left(\theta x+\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)\right)}$$

$$f\left(x\right)=b\sum_{m=0}^{\infty}\sum_{i=0}^{\infty}\sum_{k,l=0}^{\infty}\sum_{j=0}^{k}\frac{\left(-1\right)^{i+j+k+l+m}a^{i+1}\left(b\left(i+1\right)\right)}{e^{-\frac{\alpha}{\beta}(m+1)}i!\left(b\left(i+1\right)-1-j\right)}\binom{l-1}{m}\binom{k-\left(b\left(i+1\right)-1\right)}{k}\binom{k}{j}\binom{\left(b\left(i+1\right)-1\right)+k}{l}P_{j,k}\left(\theta+\alpha e^{\beta x}\right)e^{-\theta(m+1)x}e^{-\frac{\alpha}{\beta}(m+1)e^{\beta x}} \quad (14)$$

Using power series expansion on the last term in equation (14), we have

$$e^{-\frac{\alpha}{\beta}(m+1)e^{\beta x}}=\sum_{r=0}^{\infty}\frac{\left(-1\right)^{r}\alpha^{r}\left(m+1\right)^{r}}{r!\beta^{r}}e^{r\beta x} \quad (15)$$

Now, substituting equation (15), the power series expansion in equation (14) above, one gets:

$$f\left(x\right)=b\sum_{r=0}^{\infty}\sum_{m=0}^{\infty}\sum_{i=0}^{\infty}\sum_{k,l=0}^{\infty}\sum_{j=0}^{k}\frac{\left(-1\right)^{i+j+k+l+m+r}\alpha^{r}a^{i+1}\left(m+1\right)^{r}\left(b\left(i+1\right)\right)}{e^{-\frac{\alpha}{\beta}(m+1)}r!\beta^{r}i!\left(b\left(i+1\right)-1-j\right)}\binom{l-1}{m}\binom{k-\left(b\left(i+1\right)-1\right)}{k}\binom{k}{j}\binom{\left(b\left(i+1\right)-1\right)+k}{l}P_{j,k}\left(\theta+\alpha e^{\beta x}\right)e^{-\left[\theta(m+1)-r\beta\right]x}$$

Now, let

$$\eta_{i,j,k,l,m,r}=b\sum_{r=0}^{\infty}\sum_{m=0}^{\infty}\sum_{i=0}^{\infty}\sum_{k,l=0}^{\infty}\sum_{j=0}^{k}\frac{\left(-1\right)^{i+j+k+l+m+r}\alpha^{r}a^{i+1}\left(m+1\right)^{r}\left(b\left(i+1\right)\right)}{e^{-\frac{\alpha}{\beta}(m+1)}r!\beta^{r}i!\left(b\left(i+1\right)-1-j\right)}\binom{l-1}{m}\binom{k-\left(b\left(i+1\right)-1\right)}{k}\binom{k}{j}\binom{\left(b\left(i+1\right)-1\right)+k}{l}P_{j,k}$$

This implies that:

$$f(x)=\eta_{i,j,k,l,m,r}\left\{\theta e^{-\left[\theta(m+1)-r\beta\right]x}+\alpha e^{-\left[\theta(m+1)-\beta(1+r)\right]x}\right\} \quad (16)$$

Hence,

$$\mu_{n}^{'}=E\left(X^{n}\right)=\int_{0}^{\infty}x^{n}f(x)dx=\int_{0}^{\infty}x^{n}\left(\eta_{i,j,k,l,m,r}\left\{\theta e^{-\left[\theta(m+1)-r\beta\right]x}+\alpha e^{-\left[\theta(m+1)-\beta(1+r)\right]x}\right\}\right)dx$$

$$\mu_{n}^{'}=\int_{0}^{\infty}x^{n}f(x)dx=\eta_{i,j,k,l,m,r}\left[\theta\int_{0}^{\infty}x^{n}e^{-\left[\theta(m+1)-r\beta\right]x}dx+\alpha\int_{0}^{\infty}x^{n}e^{-\left[\theta(m+1)-\beta(1+r)\right]x}dx\right] \quad (17)$$

Also, using integration by substitution method in equation (17) gives the following:

Let $u_1=\left[\theta\left(m+1\right)-r\beta\right]x\Rightarrow x=\frac{u_1}{\theta\left(m+1\right)-r\beta}$; $\frac{du_1}{dx}=\theta\left(m+1\right)-r\beta$ and $dx=\frac{du_1}{\theta\left(m+1\right)-r\beta}$

Let $u_2 = \left[\theta(m+1) - \beta(1+r)\right]x \Rightarrow x = \dfrac{u_2}{\theta(m+1) - \beta(1+r)}$; $\dfrac{du_2}{dx} = \theta(m+1) - \beta(1+r)$ and

$$dx = \frac{du_2}{\theta(m+1) - \beta(1+r)}$$

Substituting for u, x and dx in equation (17) and simplifying gives:

$$\mu_n' = \eta_{i,j,k,l,m,r}\left[\frac{\theta}{\left(\theta(m+1) - r\beta\right)^{n+1}}\int_0^\infty u_1^{n+1-1}e^{-u_1}du_1 + \frac{\alpha}{\left(\theta(m+1) - \beta(1+r)\right)^{n+1}}\int_0^\infty u_2^{n+1-1}e^{-u_2}du_2\right] \quad (18)$$

Again recall that $\int_0^\infty t^{n-1}e^{-t}dt = \Gamma(n)$ and that $\int_0^\infty t^{n}e^{-t}dt = \int_0^\infty t^{n+1-1}e^{-t}dt = \Gamma(n+1)$

Thus, the nth ordinary moment of X for the WGMD is given as follows:

$$\mu_n' = \eta_{i,j,k,l,m,r}\left[\frac{\theta\Gamma(n+1)}{\left(\theta(m+1) - r\beta\right)^{n+1}} + \frac{\alpha\Gamma(n+1)}{\left(\theta(m+1) - \beta(1+r)\right)^{n+1}}\right] \quad (19)$$

3.2 The Mean

The mean of the WGMD can be obtained from the nth moment of the distribution when n = 1 as follows:

$$\mu_1' = \eta_{i,j,k,l,m,r}\left[\frac{\theta}{\left(\theta(m+1) - r\beta\right)^{2}} + \frac{\alpha}{\left(\theta(m+1) - \beta(1+r)\right)^{2}}\right] \quad (20)$$

3.3 The Variance

The nth central moment or moment about the mean of X, say μ_n, can be obtained as

$$\mu_n = E\left[X - \mu_1'\right]^n = \sum_{i=0}^{n}(-1)^i\binom{n}{i}\mu_1'^{i}\mu_{n-i}' \quad (21)$$

The variance of X for WGMD is obtained from the nth central moment when n = 2, that is, the variance of X is the nth central moment of order two (n = 2) and is given as follows:

$$Var(X) = E[X^2] - \{E[X]\}^2 \quad (22)$$

$$Var(X) = \mu_2^{'} - \{\mu_1^{'}\}^2$$

$$Var(X) = \eta_{i,j,k,l,m,r}\left[\frac{2\theta}{\left(\theta(m+1)-r\beta\right)^3} + \frac{2\alpha}{\left(\theta(m+1)-\beta(1+r)\right)^3}\right] - \left\{\eta_{i,j,k,l,m,r}\left[\frac{\theta}{\left(\theta(m+1)-r\beta\right)^2} + \frac{\alpha}{\left(\theta(m+1)-\beta(1+r)\right)^2}\right]\right\}^2 \quad (23)$$

The coefficients variation, skewness and kurtosis measures can also be calculated from the non-central moments using some well-known relationships.

3.4 Moment Generating Function

The mgf of a random variable X can be obtained by

$$M_x(t) = E\left[e^{tx}\right] = \int_0^\infty e^{tx} f(x)dx \quad (24)$$

Using power series expansion in equation (24) and simplifying the integral gives;

$$M_x(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!}\mu_n^{'} = \sum_{n=0}^{\infty} \frac{t^n}{n!}\left\{\eta_{i,j,k,l,m,r}\left[\frac{\theta\Gamma(n+1)}{\left(\theta(m+1)-r\beta\right)^{n+1}} + \frac{\alpha\Gamma(n+1)}{\left(\theta(m+1)-\beta(1+r)\right)^{n+1}}\right]\right\} \quad (25)$$

where n and t are constants, t is a real number and μ_n' denotes the n^{th} ordinary moment of X .

3.5 Characteristic Function

The characteristic function of a random variable X is given by;

$$\varphi_x(t) = E\left[e^{itx}\right] = E\left[\cos(tx) + i\sin(tx)\right] = E\left[\cos(tx)\right] + E\left[i\sin(tx)\right] \quad (26)$$

Simple algebra and power series expansion proves that

$$\phi_x(t) = \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n}}{(2n)!}\mu_{2n}^{'} + i\sum_{n=0}^{\infty} \frac{(-1)^n t^{2n+1}}{(2n+1)!}\mu_{2n+1}^{'} \quad (27)$$

Where μ_{2n}' and $\mu_{2n+1}'$are the moments of X for n=2n and n=2n+1 respectively and can be obtained from μ_n' in equation (20).

4 Some Reliability Functions

In this section, the survival and hazard functions from the WGMD are presented with adequate plots and their discussions.

4.1 The Survival Function

The survival function as the name implies describes the probability that a component or an individual will not fail after a given time. It is mathematically given as:

$$S(x) = 1 - F(x) \tag{28}$$

Taking F(x) to be the cdf of the WGMD, substituting and simplifying (28) above, we get the survival function for the WGMD as:

$$S(x) = 1 - e^{-a\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{b}} \tag{29}$$

The following is a plot of the survival function for arbitrary parameter values.

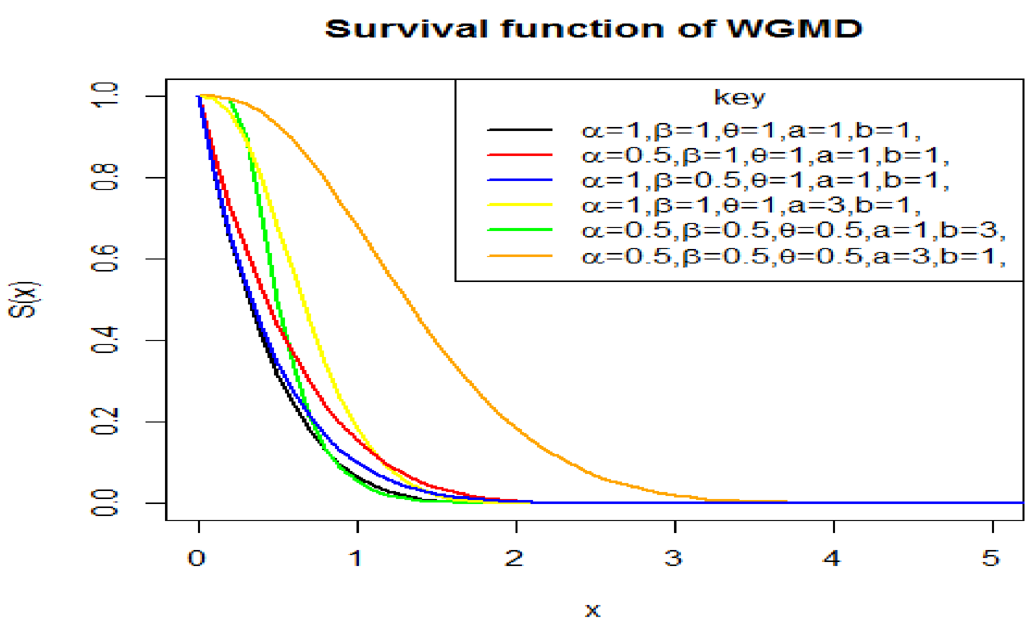

Figure 3: A plot of the survival function of WGMD.

The figure above reveals that the probability of survival for any random variable following a WGMD which decreases as the time increases, that is, as time goes on, probability of life decreases as it is expected. This shows that the WGMD would be useful for modeling most real life situations.

4.2 The Hazard Function

Hazard function is also called failure or risk function. It describes the probability failure for a component given an interval of time. The hazard function is defined mathematically as;

$$h(x)=\frac{f(x)}{1-F(x)}=\frac{f(x)}{S(x)} \quad (30)$$

Considering f(x) and F(x) to be the pdf and cdf of the proposed WGMD given previously, we obtain the hazard function as:

$$h(x)=\frac{ab\left(\theta+\alpha e^{\beta x}\right)e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{b-1} e^{-a\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{b}}}{\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\left(1-e^{-a\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{b}}\right)} \quad (31)$$

The following is a plot of the hazard function at chosen parameter values.

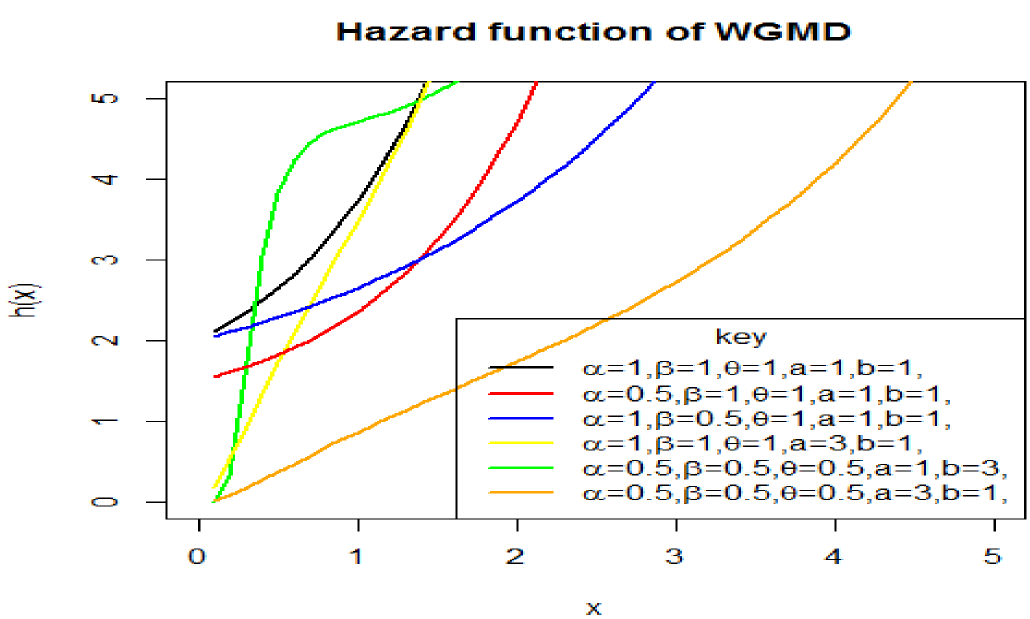

Figure 4: A plot of the hazard function of the WGMD.

Figure 4 above shows the behaviour of hazard function of the WGMD. It means that the probability of failure for any WGM random variable increases as the time or age of a subject increases, that is, as time goes on, the probability of failure or death increases.

5 Order Statistics

Suppose $X_1, X_2, ..., X_n$ is a random sample from a distribution with pdf, f(x), and $X_{1:n} < X_{2:n} < ...... < X_{i:n}$ denote the corresponding order statistic obtained from this sample. Then the pdf, $f_{i:n}(x)$ of the i^{th} order statistic can be defined as;

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!} \sum_{k=0}^{n-i} (-1)^k \binom{n-i}{k} f(x) F(x)^{k+i-1} \tag{32}$$

where f(x) and F(x) are the pdf and cdf of the Weibull Gompertz Makeham distribution respectively.

Using (5) and (6), the pdf of the i^{th} order statistics $X_{i:n}$, can be expressed from (32) as;

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!} \sum_{k=0}^{n-i} (-1)^k \binom{n-i}{k} \left[\frac{ab\left(\theta+\alpha e^{\beta x}\right)\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{b-1}}{\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right] e^{a\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{b}} e^{\theta x+\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}} \right] * \left[e^{-a\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{b}} \right]^{i+k-1} \tag{33}$$

Hence, the pdf of the minimum order statistic $X_{(1)}$ and maximum order statistic $X_{(n)}$ of the WGMD are given by;

$$f_{1:n}(x) = n \sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k} \left[\frac{ab\left(\theta+\alpha e^{\beta x}\right)\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{b-1}}{\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right] e^{a\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{b}} e^{\theta x+\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}} \right] * \left[e^{-a\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{b}} \right]^{k} \tag{34}$$

and

$$f_{n:n}(x)=n\left[\frac{ab\left(\theta+\alpha e^{\beta x}\right)\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{b-1}}{\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]e^{a\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{b}}e^{\theta x+\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}}\right]\left[e^{-a\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{b}}\right]^{n-1} \tag{35}$$

respectively.

6 Estimation of unknown Model Parameters using Maximum Likelihood Method

Let $X_1, X_2, ..., X_n$ be a sample of size 'n' independently and identically distributed random variables from the WGMD with unknown parameters $a, b,$ α, β and θ defined previously. The pdf of the WGMD is given from (6) as:

$$f(x)=\frac{ab\left(\theta+\alpha e^{\beta x}\right)e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{b-1}e^{-a\left(-\log\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]\right)^{b}}}{\left[1-e^{-\theta x-\frac{\alpha}{\beta}\left(e^{\beta x}-1\right)}\right]}$$

The likelihood function is given by;

$$L\left(\underline{X}\mid a,b,\alpha,\beta,\theta\right)=\frac{(ab)^{n}\prod_{i=1}^{n}\left(\left(\theta+\alpha e^{\beta x_i}\right)e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right)\prod_{i=1}^{n}\left(-\log\left[1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right]\right)^{b-1}e^{-a\sum_{i=1}^{n}\left(-\log\left[1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right]\right)^{b}}}{\prod_{i=1}^{n}\left[1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right]} \tag{36}$$

Let the log-likelihood function, $l=\log L\left(\underline{X}\mid a,b,\alpha,\beta,\theta\right)$, therefore

$$l=n\log a+n\log b+\sum_{i=1}^{n}\log\left(\theta+\alpha e^{\beta x_i}\right)-\theta\sum_{i=1}^{n}x_i-\tfrac{\alpha}{\beta}\sum_{i=1}^{n}\left(e^{\beta x_i}-1\right)+(b-1)\sum_{i=1}^{n}\log\left(-\log\left[1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right]\right)$$
$$-a\sum_{i=1}^{n}\left(-\log\left[1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right]\right)^{b}-\sum_{i=1}^{n}\left(1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right) \tag{37}$$

Differentiating **l** partially with respect to $a, b,$ α, β and θ respectively gives;

$$\frac{\partial l}{\partial a}=\frac{n}{a}-\sum_{i=1}^{n}\left(-\log\left[1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right]\right)^{b} \tag{38}$$

$$\frac{\partial l}{\partial b}=\frac{n}{b}+\sum_{i=1}^{n}\log\left(-\log\left[1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right]\right)-a\sum_{i=1}^{n}\left(-\log\left[1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right]\right)^{b}\log\left(-\log\left[1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right]\right) \tag{39}$$

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

$$-\frac{1}{\beta}\sum_{i=1}^{n}\left\{\frac{\left(e^{\beta x_i}-1\right)e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}}{\left(1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right)}\right\}+ab\sum_{i=1}^{n}\left\{\frac{\left(-\log\left[1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right]\right)^{b-1}\left(e^{\beta x_i}-1\right)e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}}{\left(1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right)}\right\} \quad (40)$$

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

$$-\frac{\alpha}{\beta}\sum_{i=1}^{n}\left\{\frac{\left(x_i e^{\beta x_i}-\beta^{-1}\left(e^{\beta x_i}-1\right)\right)e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}}{\left(1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right)}\right\}+\frac{ab\alpha}{\beta}\sum_{i=1}^{n}\left\{\frac{\left(-\log\left[1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right]\right)^{b-1}\left(x_i e^{\beta x_i}-\beta^{-1}\left(e^{\beta x_i}-1\right)\right)e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}}{\left(1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right)}\right\} \quad (41)$$

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

$$-\sum_{i=1}^{n}\left\{\frac{x_i e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}}{\left(1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right)}\right\}+ab\sum_{i=1}^{n}\left\{\frac{\left(-\log\left[1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right]\right)^{b-1}x_i e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}}{\left[1-e^{-\theta x_i-\frac{\alpha}{\beta}\left(e^{\beta x_i}-1\right)}\right]}\right\} \quad (42)$$

Equating equation (38), (39), (40), (41) and (42) to zero and solving for the solution of the non-linear system of equations gives the maximum likelihood estimates of the parameters a, b, α, β and θ respectively. However the solution cannot be obtained analytically except numerically with the aid of suitable statistical software like Python, R, SAS, etc., when data sets are given.

7 Application to a real life dataset

This section presents a dataset on the remission times (in months) of a random sample of 128 bladder cancer patients with its descriptive statistics and application to some selected extensions of the Gompertz-Makeham distribution together with the classical Gompertz distribution. The performance of the Weibull Gompertz-Makeham distribution (WGMD) is compared to some families of Makeham distribution such as Kumaraswamy Gompertz Makeham distribution (KGMD), Transmuted Gompertz-Makeham distribution (TGMD), Gompertz-Makeham distribution (GMD) and the Gompertz distribution (GD).

The performance of the above listed models is ranked using some criteria such as the AIC (Akaike Information Criterion), CAIC (Consistent Akaike Information Criterion) and BIC (Bayesian Information Criterion). It is considered that the model with the smallest values of these statistics will be the best model to fit the data.

Data set: This data set represents the remission times (in months) of a random sample of 128 bladder cancer patients. It has previously been used by Ieren and Chukwu (2018), Ieren and Kuhe (2018) and Abdullahi et al. (2018). It is summarized as follows:

Table 1: Summary Statistics for the dataset

Parameter	n	Min	Q_1	Median	Q_3	Mean	Max	Var	Skew	Kurt
Values	128	0.0800	3.348	6.395	11.840	9.366	79.05	110.425	3.3257	19.1537

n-sample size, Min–Minimum, Q_1-First Quartile, Q_3-Third Quartile, Max-Maximum, Var-Variance, Skew-Skewness, Kurt-Kurtosis

From the descriptive statistics in table 1, it is observed that the data set is positively skewed with a very high coefficient of kurtosis and therefore suitable for flexible and skewed distributions.

Table 2: The strength of the selected models using the AIC, CAIC and BIC values of the models evaluated from the maximum likelihood estimations based on the bladder cancer data.

Distributions	Parameter estimates	AIC	CAIC	BIC	Ranks of models
WGMD	$\alpha = 0.006233$ $\beta = 0.005620$ $\theta = 0.006119$ $a = 0.006070$ $b = 0.004811$	-6.8677	-6.3759	7.3924	**1**
KGMD	$\alpha = 1.4996$ $\beta = 0.0008286$ $\theta = 3.2157$ $a = 0.1934$ $b = 0.09415$	17.3139	17.8057	31.5741	**3**
TGMD	$\alpha = 9.7913$ $\beta = 9.6327$ $\theta = 9.4246$ $\lambda = 0.8682$	51.5307	51.8559	62.9388	**5**
GMD	$\alpha = 2.04152$ $\beta = 7.8924$ $\theta = 7.9969$	29.2259	29.4194	37.7820	**4**
GD	$\alpha = 4.5814$ $\beta = 4.5809$	12.8378	12.9338	18.5418	**2**

From Table 2**,** comparing the values of the AIC, CAIC and BIC for each model, the WGMD has the best performance compared to the KGMD, TGMD GMD and GD. This is due to the decision rule which says that the distribution or model with the smallest values of the test statistics (AIC, CAIC and BIC) is taken as the most adequate or efficient model. These values also agree with the fact that generalizing any continuous distribution provides a compound distribution with a better fit than the baseline distribution (Koleoso et al. (2019)).

8 Conclusion

This article proposed a new distribution called Weibull-Gompertz Makeham distribution. The statistical properties of the distribution have been derived and studied extensively. The model parameters were estimated using maximum likelihood method. The distribution (WGMD) has the best fit compared to the other four models considered in this study when applied to real life time data.

Competing Interests

Authors have declared that no competing interests exist.

References

1. Abd El-Bar AMT. An extended Gompertz-Makeham distribution with application to lifetime data. Communications in Statistics-Simulation and Computation. 2017;1-48. DOI:10.1080/03610918.2017.1348517

2. Abdullahi J, Abdullahi UK, Ieren TG, Kuhe DA, Umar AA. On the properties and applications of transmuted odd generalized exponential-exponential distribution. Asian Journal of Probability and Statistics. 2018;1(4):1-14. DOI: 10.9734/AJPAS/2018/44073

3. Alizadeh M, Cordeiro GM, de Brito E, Demetrio CGB. The Beta Marshall-Olkin family of distributions. Journal of Statistical Distributions and Applications. 2015;2(4):1–18.

4. Alzaatreh A, Famoye F, Lee C. Weibull-Pareto distribution and its applications. Communications in Statistics-Theory and Methods 2013;42(9):1673–1691.

5. Alzaghal A, Famoye F., Lee C. Exponentiated T-X family of distributions with some applications. International Journal of Statistics and Probability 2013;2(3):31–49.

6. Bourguignon M, Silva RB, Cordeiro GM. The Weibull-G family of probability distributions. Journal of Data Science. 2014;12:53-68.

7. Chukwu AU, Ogunde AA. On Kumaraswamy Gompertz Makeham distribution. American Journal of Mathematics and Statistics. 2016;6(3):122-127.

8. Cordeiro GM, Ortega EMM, da Cunha DCC. The Exponentiated generalized class of distribution. Journal of Data Science. 2013;11:1–27.

9. Cordeiro GM, Ortega EMM, Ramires TG. A new generalized Weibull family of distributions: mathematical properties and applications. Journal of Statistical Distributions and Applications. 2015;2(13):1-25. DOI 10.1186/s40488-015-0036-6

10. Dey S, Moala FA, Kumar D. Statistical properties and different methods of estimation of Gompertz distribution with application. Journal of Statistics and Management Systems. 2018;21(5):839-876.

11. Eugene N, Lee C, Famoye F. Beta-Normal distribution and its applications. Communications in Statistics-Theory and Methods. 2002;31(4):497-512.

12. Golubev A. Does Makeham make sense? Biogerontology. 2004;5:159-167.

13. 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.

14. Ieren TG, Chukwu AU. Bayesian analysis of a shape parameter of the Weibull-Frechet distribution. Asian Journal of Probability and Statistics. 2018;2(1):1-19. DOI: 10.9734/AJPAS/2018/44184

15. Ieren TG, Kuhe DA. On the properties and applications of Lomax-Exponential distribution. Asian Journal of Probability and Statistics. 2018;1(4):1-13. DOI: 10.9734/AJPAS/2018/42546

16. Ieren TG, Oyamakin SO, Chukwu AU. Modeling lifetime data with Weibull-Lindley distribution. Biometrics and Biostatistics International Journal. 2018;7(6):532–544.

17. Johnson NL, Kotz S, Balakrishnan N. Continuous Univariate Distributions, volume 2. John Wiley & Sons, New York;1995.

18. Koleoso PO, Chukwu AU, Bamiduro TA. A Three-Parameter Gompertz-Lindley Distribution: Its Properties and Applications. Journal of Mathematical Theory and Modeling. 2019;9(4):29-42. DOI: 10.7176/MTM

19. Lawless JF. Statistical Models and Methods for Lifetime Data, 2nd Edition, Wiley, New Jersey; 2003.

20. Makeham WM. On the law of mortality and the construction of annuity tables. Journal of the Institute of Actuaries; 1860;8(6):301-310.

21. Marshall A, Olkin I. Life Distributions. Structure of Nonparametric, Semiparametric, and Parametric Families. Springer, New York; 2007.

22. Shaw WT, Buckley IRC. The alchemy of probability distributions: beyond Gram-Charlier expansions and a skew-kurtotic-normal distribution from a rank transmutation map. Research report; 2007.

23. Tahir MH, Cordeiro GM, Alzaatreh A, Mansoor M, Zubair M. The logistic-X family of distributions and its applications. Communications in Statistics-Theory and Methods. 2016;45(24):7326-7349. DOI: 10.1080/03610926.2014.980516

24. Tahir MH, Zubair M, Mansoor M, Cordeiro GM, Alizadeh M, Hamedani GG. A new Weibull-G family of distributions. Hacettepe Journal of Mathematics and Statistics. 2016;45(2):629-647.

25. Torabi H, Montazeri NH. The Logistic-Uniform distribution and its application. Communications in Statistics-Simulation and Computation. 2014;43(10):2551–2569. DOI: 10.1080/03610918.2012.737491