The Beta Generalized Weibull distribution: Properties and applications

Neetu Singla *, Kanchan Jain, Suresh Kumar Sharma

Department of Statistics, Panjab University, Chandigarh 160014, India

A R T I C L E   I N F O

Article history:
Received 28 June 2011
Received in revised form
9 February 2012
Accepted 12 February 2012
Available online 25 February 2012

Keywords:
Weibull
Beta Weibull distribution
Generalized Weibull distribution
Maximum likelihood estimation (MLE)
Information matrix

A B S T R A C T

A five-parameter distribution called Beta Generalized Weibull (BGW) distribution is introduced. Beta Generalized Exponential (BGE), Beta Weibull (BW), Generalized or Exponentiated Weibull (GW or EW), Generalized Rayleigh (GR), Beta Exponential (BE), Generalized Exponential (GE), Weibull, Rayleigh and Exponential are its sub models. The cumulative distribution function (cdf) and the probability density function (pdf) have been expressed as mixtures of the Generalized Weibull cdfs and pdfs. The kth order moment has been derived. The non-linear equations for deriving the maximum likelihood estimators function (pdf) have been expressed as mixtures of the Generalized Weibull cdfs and pdfs. The kth order moment has been derived. The non-linear equations for deriving the maximum likelihood estimators

1. Introduction

For complex electronic and mechanical systems, the failure rate often exhibits non-monotonic (bathtub or upside-down bathtub (unimodal) shaped) failure rates ([Lai and Xie [1]]). Distributions with such failure rates have attracted a considerable attention of researchers in reliability engineering. In software reliability, bathtub shaped failure rate is encountered in firmware, an embedded software in hardware devices. Firmware plays an important role in functioning of hard drives of large computers, spacecraft and high performance aircraft control systems, advanced weapon systems, safety critical control systems used for monitoring the industrial process in chemical and nuclear plants (Zhang et al. [2]). The upside down bathtub shaped failure rate is used in data of motor bus failures (Mudholkar et al. [3]), for optimal burn-in decisions (Block and Savits [4], Chang [5]) and for ageing properties in reliability (Gupta and Gupta [6], Jiang et al. [7]).

Weibull distribution introduced by Weibull [8] is a popular distribution for modeling phenomenon with monotonic failure rates. But this distribution does not provide a good fit to data sets with bathtub shaped or upside-down bathtub shaped (unimodal) failure rates, often encountered in reliability, engineering and biological studies. Hence a number of new distributions modeling the data in a better way have been constructed in literature as ramifications of Weibull distribution. Pham and Lai [9] present a review of some of the generalizations or modifications of Weibull distribution.

Xie and Lai [10] introduced the additive Weibull model obtained by adding two Weibull survival functions. They showed that the failure rate function of the proposed model is expressible as the sum of two Weibull failure rate functions. They demonstrated the graphical estimation technique based on conventional Weibull plot technique. This model is applicable in case of bathtub shaped failure rate function. Xie et al. [11] studied a generalization of the Weibull distribution, the modified Weibull extension, for modeling the systems with a bathtub shaped failure rate function. They explored the distributional properties and found the maximum likelihood estimates of the parameters. They used two lifetime data sets for illustrating the applications of new model. Beta Exponential (BE) distribution, generated from the logit of a Beta random variable, was introduced by Nadarajah and Kotz [12]. They provided mathematical properties of BE model and also derived expressions for the mgf, first four moments, variance, skewness and kurtosis. They provided estimates of the parameters using the methods of moments and maximum likelihood estimate and gave expressions for the Fisher Information matrix. Bebbington et al. [13] proposed a new distribution called the new flexible Weibull distribution. They characterized its failure rate function and considered parametric estimation. They found the new distribution to be quite flexible for modeling IFR, IFRA and modified bathtub (MBT). Jiang et al. [7] studied the ageing properties of unimodal failure rate models. Jiang and Murthy [14] studied three different models, each involving two parameter Weibull distributions, to analyze the failure data and discussed parameter estimation. Bucar et al. [15] proved that the reliability of an arbitrary system can be approximated well by a finite Weibull mixture with
positive component weights. They estimated the unknown parameters and the weights of the Weibull mixture. Calculation of AICs was done for comparing the fitted models to the empirical ones.

Some other variants of Weibull distribution are Generalized or Exponentiated Weibull (GW or EW) [Mudholkar et al. [16,17]], the modified Weibull (MW) distribution [Lai et al. [18], Nadarajah and Kotz [19]], the Beta Weibull (BW) distribution [Lee et al. [20]] and the Generalized Modified Weibull (GMW) distribution [Carrasco et al. [21]].

Beta Generalized (Beta G), introduced by Singh et al. [22] is a rich class of generalized distributions. This class has captured a considerable attention over the last few years.

Sepanski and Kang [23] applied the Beta G distribution to model the size distribution of income. This distribution has been studied in literature for various forms of G. The distributions that have been explored are the Beta Normal (BN) [Eugene et al. [24]], the Beta Gumbel (BGu) distribution [Nadarajah and Kotz [25]], the Beta Fréchet (BFr) distribution [Nadarajah and Gupta [26]], the Beta Exponential (BE) distribution [Nadarajah and Kotz [12]], the Beta Weibull (BW) distribution [Gambo et al. [27], Lee et al. [20] and Cordeiro et al. [28]]. Souza et al. [29] introduced the Beta Generalized Exponential (BGE) distribution and the Beta Modified Weibull (BMW) distribution was discussed by Silva et al. [30]. Beta Inverse Weibull (BIW) [Khan [31]], Beta Generalized Pareto (BGP) [Mahmoudi [32]] and Beta Dagum (BDA) (Domma and Contino [33]) are some new extensions of the Beta G class of distributions.

The cumulative distribution function (cdf) of the Beta G distribution has the form

\[ F(x) = \int_0^{G(x)} w^{-1} (1-w)^{b-1} dw = \frac{B_G(a,b)}{B(a,b)} = I_G(a,b), a,b > 0, \]

(1)

where \( G(x) \) is an arbitrary parent/baseline cdf of a random variable; \( B_G(a,b) = \int_0^x w^{-1} (1-w)^{b-1} dw \) is the incomplete beta function with \( B_G(a,b) = B_1(a,b) \) and \( I_G(a,b) = (B_1(a,b)/B(a,b)) \) is the incomplete beta function ratio.

The density corresponding to (1) is written as

\[ f(x) = \frac{1}{B(a,b)} G(x)^{a-1} (1-G(x))^{b-1} g(x), \]

(2)

where \( g(x) = (dG(x)/dx) \) is the probability density function (pdf) of the parent/baseline distribution.

In this paper, we introduce a new five-parameter distribution called Beta Generalized Weibull (BGW) distribution by taking \( G(x) \) to be the cumulative distribution function of Generalized Weibull (GW). The GW family was applied for analyzing bathtub failure data for bus motor failure data, head and neck cancer clinical trial data (Mudholkar and Srivastava [3] and flood data (Mudholkar and Hutson [34]). The BGW distribution unifies many existing distributions as its sub models with applications in reliability engineering. These distributions are Beta Generalized Exponential (BGE), Beta Weibull (BW), Generalized or Exponentiated Weibull (GW or EW), Generalized Rayleigh (GR) (Kundu and Rakab [35]), Beta Exponential (BE), Generalized Exponential (GE) (Gupta and Kundu [36]), Weibull, Rayleigh and Exponential distributions. The addition of new parameters helps in controlling the skewness and the tail weight. This new distribution exhibits increasing, decreasing, bathtub and upside down bathtub shaped failure rate functions for different parametric combinations. The motive behind introducing this new distribution is its flexibility in modeling a wide spectrum of real data sets in reliability and biological sciences. Although other variants of Weibull or Beta G distributions also possess various shapes of the failure rate function, it has been observed through two data sets and simulations that the proposed distribution provides a better fit than all its sub models. Hence, the new distribution is expected to provide a more general and flexible framework having wider applications in reliability engineering and survival studies.

The paper is organized as follows. In Section 2, the proposed BGW distribution is introduced. In Section 3, we derive some alternative forms of the cdf and the pdf of the BGW distribution. It is shown that the BGW pdf can be expressed as an infinite (or finite) weighted linear combination of the pdfs of the GW random variables. The pdf of order statistics is also derived. Section 4 derives the moments through as well as Laplace transformation. In Section 5, the non-linear equations for deriving the maximum likelihood estimators and the elements of the observed information matrix are presented. The importance of the proposed distribution is further emphasized in Section 6 by fitting it to two real data sets. It is observed that the BGW provides a better fit than BGE, BW, GW, BE, GE and Weibull distributions. Concluding remarks are presented in Section 7. Some proofs and mathematical expressions are included in the appendix.

2. Beta Generalized Weibull distribution

In this section, we introduce the five-parameter Beta Generalized Weibull (BGW) distribution by assuming \( G(x) \) to be the cdf of the Generalized Weibull (GW) distribution.

The Generalized Weibull distribution with two shape parameters \( \alpha, \beta > 0 \) and a scale parameter \( \lambda > 0 \) was introduced by Mudholkar et al. [16]. The cdf of the GW is given by

\[ G(x) = \left(1 - e^{-\lambda x^{\alpha}}\right)^{1-\beta}, \quad x > 0, \beta > 0, \lambda > 0. \]

(3)

The pdf corresponding to (3) is

\[ g(x) = \beta \lambda x^{\alpha-1} \left(1 - e^{-\lambda x^{\alpha}}\right)^{1-\beta} e^{-\lambda x^{\alpha}}, \quad x > 0, \beta > 0, \lambda > 0. \]

(4)

Using (3) in (1), the cdf of the BGW distribution can be written as

\[ F(x) = \frac{1}{B(a,b)} \int_0^{(1-\beta)\lambda x^{\alpha-1}} w^{-1} (1-w)^{b-1} dw, \quad x > 0, a,b,\alpha, \beta \]

(5)

The pdf and the failure rate function of the new distribution take the form

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

(6)

and

\[ h(x) = \frac{\beta \lambda x^{\alpha-1}}{B(a,b) \left\{1 - \left(1 - e^{-\lambda x^{\alpha}}\right)^{1-\beta}\right\}^{b-1}} \left(1 - e^{-\lambda x^{\alpha}}\right)^{-\beta} \left\{1 - \left(1 - e^{-\lambda x^{\alpha}}\right)^{1-\beta}\right\}^{b-1} e^{-\lambda x^{\alpha}}. \]

(7)

If \( X \) is a random variable with pdf (6), we use the notation \( X \sim BGW (a, \beta, \lambda, \alpha, \beta) \).

**Special cases:**

1. For \( \beta = 1 \), we get the Beta Generalized Exponential (BGE (\( a, \beta, \lambda, \alpha \)) distribution.
2. If \( a = b = 1 \), then (6) reduces to the Generalized Weibull (GW) distribution. If in addition \( \beta = 2 \), we obtain Generalized Rayleigh (GR) distribution.
3. Beta Weibull (BW) distribution arises as a special case of BGW by taking \( \alpha = 1 \).
4. Applying \( a = b = \beta = 1 \), we can obtain Generalized Exponential (GE) distribution.
5. With $x = \beta = 1$ in (6), Beta Exponential distribution can be obtained.

6. For $a = b = x = 1$, we obtain the Weibull distribution with parameters $\lambda$ and $\beta$. If in addition $\beta = 2$, the BGW distribution becomes Rayleigh distribution.

The above special cases are also depicted in pictorial form in Fig. 1.

The following result helps in simulating observations from the BGW distribution.

If $V$ follows Beta distribution with parameters $a$ and $b$, then

$$X = G^{-1}(V) = \frac{[-\log(1-V^{1/2})]^{1/\beta}}{\lambda}$$

follows BGW distribution with parameters $a$, $b$, $\lambda$, $x$ and $\beta$.

This follows easily from the result of Jones [37] which states that if $Y$ follows Beta distribution with parameters $a$ and $b$, then $X = G^{-1}(Y)$ follows Beta Generalized distribution given by (1).

---

Fig. 1. Sub models of BGW.

Fig. 2. Plot for comparison of exact and empirical cdf of BGW to simulated random numbers.

Fig. 3. Plots for BGW densities for simulated data sets: (a) $a = 8$, $b = 0.5$, $x = 1$, $\beta = 1$ and (b) $a = 0.5$, $b = 0.5$, $x = 1.5$, $\beta = 4.5$. 
For checking the correctness of the procedure for simulating a data set from BGW distribution, we plot, the exact and the empirical cdf of BGW in Fig. 2 using a pseudo random sample of size 1000. The histograms for two generated data sets and the exact BGW density plots are displayed in Fig. 3. These plots show that the simulated values are consistent with the BGW distribution. For some values of the parameters, the plots of the density and the failure rate function are shown in Figs. 4 and 5 respectively. It is depicted by Fig. 5 that the failure rate function of the BGW distribution can take monotonic, bathtub and upside-down bathtub shapes for different parametric combinations.

3. Some results for the cumulative distribution and density function

For any device/equipment, the interest might be in determining the survival probabilities which can be obtained directly or through the cdf. The probability of failure of a device within an interval of infinitesimal length $dx$ is given by $f(x)dx$. Hence, we present various representations of the cdf, the survival function and the pdf which can be helpful in numerical applications.

The mathematical relation given below will be used in Theorem 1 and Section 4.

$$\frac{(1-z)^{b-1}}{z_0} = \sum_{j=0}^{\infty} \left( \frac{(-1)^j \Gamma(b)}{\Gamma(b-j)\Gamma(j+1)} \right)^{p_j}$$

when $b$ is a positive real non-integer and $|z| < 1$.

The following theorem shows that the cdf (pdf) of the BGW can be expressed as an infinite (finite for integer $b$) weighted sum of the cdfs (pdfs) of the GW distribution. Alternative expressions for the pdf and cdf are given. The density function of the $i$th order statistic is also presented.

**Theorem 1.** Let $G_{\beta,b,\alpha+j}(x)$ and $g_{\beta,b,\alpha+j}(x)$ be the cdf and pdf of the GW distribution with scale parameter $\alpha+j$ and shape parameters $\beta$ and $\alpha+j$ and

$$w_j = \frac{(-1)^j \Gamma(a+b)}{\Gamma(a)\Gamma(b-j)\Gamma(j+1)\Gamma(a+j)}$$

be the simple weights.

Then
(A) For real and non-integer \( b > 0 \),

(i) \( F(x) = \sum_{j=0}^{\infty} \binom{\alpha}{b} G_{i,b,nx+b}(x) \).

(ii) \( f(x) = \frac{\partial F(x)}{\partial x} = \sum_{j=0}^{\infty} \binom{\alpha}{b} G_{i,b,nx+b}(x) \).

(iii) \( f(x) = \frac{\partial f(x)}{\partial x} = \sum_{j=0}^{\infty} \binom{\alpha}{b} G_{i,b,nx+b}(x) \).

(B) For general \( a \) and \( b \) and

\[ f_{X_1}(x;\alpha,\beta;\gamma;\lambda) = \sum_{m=0}^{\infty} g_m((x)(\beta)/\gamma), \]

the hypergeometric function (Gradsteyn and Ryzhik [38]), the cdf helps us in writing

\[ F(x) = \frac{1}{b(a,b)} \sum_{i=0}^{\infty} \left[ (1-y)^i \right] (1-e^{-y})^i dy \]

\[ = \frac{\Gamma(b)}{B(a,b)} \left[ \sum_{i=0}^{\infty} \left( -\frac{1}{i+b-j} \right)^i e^{-y} \right] \left( 1-e^{-y} \right)^i dy \]

where \( \delta = \delta(a+j) \).

Using binomial expansion and writing

\[ a_i = \frac{\delta(\delta-1) \ldots (\delta-i+1)}{i!} (-y)^i, \]

\[ E(x^k) = \frac{\Gamma(b)}{B(a,b) \gamma^k} \sum_{i=0}^{\infty} \left[ \left( -\frac{1}{i+b-j} \right)^i e^{-y} \right] \left( 1-e^{-y} \right)^i dy \]

\[ = \frac{\Gamma(b)}{B(a,b) \gamma^k} \sum_{i=0}^{\infty} \left[ \left( -\frac{1}{i+b-j} \right)^i e^{-y} \right] \left( 1-e^{-y} \right)^i \gamma^i \]

\[ = \frac{\Gamma(b)}{B(a,b) \gamma^k} \sum_{i=0}^{\infty} \left[ \left( -\frac{1}{i+b-j} \right)^i e^{-y} \right] \left( 1-e^{-y} \right)^i \gamma^i \]

\[ = \frac{\Gamma(b)}{B(a,b) \gamma^k} \sum_{i=0}^{\infty} \left[ \left( -\frac{1}{i+b-j} \right)^i e^{-y} \right] \left( 1-e^{-y} \right)^i \gamma^i \]

Eq. (10) gives the general expression for the kth moment of BGW distribution.

If \( x \) is a positive integer, then the last expression in the parentheses is

\[ 1 + \sum_{i=0}^{x-1} \left( (i+1)^{-x} \right) \gamma^i \]

where \( \gamma^{x+1} = \gamma \).

Substituting in (10), this gives

\[ E(x^k) = \frac{\Gamma(b) \Gamma(k/b+1)}{B(a,b) \gamma^k} \sum_{i=0}^{\infty} \left[ \left( -\frac{1}{i+b-j} \right)^i e^{-y} \right] \left( 1-e^{-y} \right)^i \gamma^i \]

\[ = \frac{\Gamma(b) \Gamma(k/b+1)}{B(a,b) \gamma^k} \sum_{i=0}^{\infty} \left[ \left( -\frac{1}{i+b-j} \right)^i e^{-y} \right] \left( 1-e^{-y} \right)^i \gamma^i \]

Special cases:

- For \( a=b=1 \),

\[ E(x^k) = 2 \int_{0}^{\infty} \frac{\alpha}{\beta} \gamma^k \left( 1-e^{-\alpha \gamma} \right)^{k-1} \gamma^k \left( 1-e^{-\alpha \gamma} \right)^{k-1} \gamma^k \]

which gives the \( k \)th order moment of the Generalized Weibull distribution given by Mudholkar et al. [3,16].

For \( a=b=1 \), \( E(x^k) = (1+x)^{-1} \) is the \( k \)th order moment of the Weibull distribution.

If \( b \) is a positive non-integer and \( x=1 \), then

\[ E(x^k) = \frac{\Gamma(b) \Gamma(k/b+1)}{B(a,b) \gamma^k} \sum_{i=0}^{\infty} \left[ \left( -\frac{1}{i+b-j} \right)^i e^{-y} \right] \left( 1-e^{-y} \right)^i \gamma^i \]

gives the \( k \)th moment of the Beta Weibull distribution.

Cordeiro et al. [28] found the moments of the Beta Weibull when \( a \) is a positive non-integer.

The general expression for the \( k \)th moment of BGW can also be found using the Laplace transformation technique as discussed below.

The Laplace transform of the BGW is

\[ L(s) = \int_{0}^{\infty} \alpha \beta \gamma^k \left( 1-e^{-\alpha \gamma} \right)^{k-1} \gamma^k \left( 1-e^{-\alpha \gamma} \right)^{k-1} \gamma^k \]

\[ = \int_{0}^{\infty} (s-\alpha \log(\gamma))^{k-1} \left( 1-e^{-\alpha \gamma} \right)^{k-1} \gamma^k \]

Letting \( \nu = \alpha \gamma \), we get

\[ L(s) = \int_{0}^{\infty} (s-\alpha \log(\gamma))^{k-1} \left( 1-e^{-\alpha \gamma} \right)^{k-1} \gamma^k \]

where \( \delta = \delta(a+j) \).
Interchanging the summation and integration and using binomial theorem,
\[ L(s) = \frac{2\Gamma(b)}{B(a,b)} \sum_{j=0}^{\infty} \left( -1 \right)^j \left( \frac{\Gamma(b-j)}{\Gamma(b+j)} \right)^a \frac{1}{\Gamma(j+1)} \int_0^1 \left( 1 + \sum_{i=1}^n a_i^j \right) (\log v)^j dv \]
where \( a_i \) is defined in (9).

Writing \( L = \sum_{i=1}^n a_i^j \left( \log v \right)^j dv \) and substituting \( e^{-y} = v \), we get
\[ l = \int_0^\infty ( -y e^{-y} e^d y ) \sum_{i=1}^n a_i^j \left( \log v \right)^j dv = ( -1 )^{\frac{1}{\beta}} \left( \int_0^\infty ( -y e^{-y} e^d y ) \sum_{i=1}^n a_i^j \left( \log v \right)^j dv \right) \]
\[ = ( -1 )^{\frac{1}{\beta}} \left( \left[ \Gamma \left( \frac{b}{\beta} + 1 \right) \right] \sum_{i=1}^n a_i^j (i+1)^{\left( \frac{b}{\beta} - 1 \right)} \right) \]

Hence,
\[ L(s) = \frac{2\Gamma(b)}{B(a,b)} \sum_{j=0}^{\infty} \left( -1 \right)^j \left( \frac{\Gamma(b-j)}{\Gamma(b+j)} \right)^a \frac{1}{\Gamma(j+1)} \left( \left[ \Gamma \left( \frac{b}{\beta} + 1 \right) \right] \sum_{i=1}^n a_i^j (i+1)^{\left( \frac{b}{\beta} - 1 \right)} \right) \times \left( 1 + \sum_{i=1}^n a_i^j (i+1)^{\left( \frac{b}{\beta} - 1 \right)} \right) \]

This gives
\[ E \left( X_k^j \right) = ( -1 )^\left( \frac{1}{\beta} \right) \left( \frac{\Gamma(b)}{B(a,b)} \right)^{\frac{1}{\beta}} \left( \frac{1}{\Gamma(b+j)} \right) \left( \sum_{i=1}^n a_i^j (i+1)^{\left( \frac{b}{\beta} - 1 \right)} \right) \]
\[ \times \left[ \Gamma \left( \frac{b}{\beta} + 1 \right) \right] \sum_{i=1}^n a_i^j (i+1)^{\left( \frac{b}{\beta} - 1 \right)} \]
\[ \left( 1 + \sum_{i=1}^n a_i^j (i+1)^{\left( \frac{b}{\beta} - 1 \right)} \right) \]

which is same as (10).

Hence, the moments can be determined using any one of the two approaches.

5. Estimation and inference

In this section, we derive the non-linear equations for finding the maximum likelihood estimates (MLEs) of the parameters. It is assumed that \( X \) follows BGW \((a, b, \lambda, \alpha, \beta)\) and \( \theta=(a, b, \lambda, \alpha, \beta) \) denotes the parameter vector. The log likelihood for a single observation \( x \) of \( X \) can be written as
\[ l = \ln(b,a,\lambda,\alpha,\beta) \]
\[ = \ln x + \ln \beta + \beta \log x + (\beta - 1) \log (x - (ax - 1) \log (1 - e^{-ix}) + (b - 1) \log (1 - e^{-ix})), \quad x > 0. \]

Let \( \psi(.) \) denote the digamma function, the logarithmic derivative of the gamma function. Then the components of the unit score vector \( U=((\partial / \partial a),(\partial / \partial b),(\partial / \partial \lambda),(\partial / \partial \alpha),(\partial / \partial \beta)) \) are given by
\[ \frac{\partial l}{\partial a} = -\psi(a)+\psi(a+b)+2\log(1-e^{-ia}) \]
\[ \frac{\partial l}{\partial b} = -\psi(b)+\psi(a+b)+2\log(1-e^{-ib}) \]
\[ \frac{\partial l}{\partial \lambda} = \frac{\beta}{\lambda} + \frac{(ax-1)e^{-ia}b/(2b-x)2^{1-x}}{(1-e^{-ib})^2} - \frac{(b-1)x\beta(1-e^{-ia})^2-1-e^{-ia}b/(2b-x)2^{1-x}}{(1-e^{-ib})^2} \]
\[ \frac{\partial l}{\partial \alpha} = \frac{1}{\alpha} + a\log(1-e^{-ia}) - \frac{(b-1)(1-e^{-ia})^2\log(1-1-e^{-ia})}{1-e^{2ia}} \]
\[ \frac{\partial l}{\partial \beta} = \frac{1}{\beta} + \log \beta + \log x + \frac{(az-1)e^{-ia}b/(2b-x)2^{1-x}}{(1-e^{-ib})^2} \]

The expected value of the unit score vector vanishes leading to the following equations:
\[ E \left( \log(1-e^{-ia}) \right) = \psi(a) - \psi(a+b) \]
\[ E \left( \log \left( 1 - (1-e^{-ia})^2 \right) \right) = \psi(b) - \psi(a+b) \]
\[ E \left( \frac{(1-e^{-ia})^2\log(1-e^{-ia})}{1-1-e^{-ia})^2} \right) = \frac{a(\psi(a) - \psi(a+b)) + 1}{\alpha(\beta-1)} \]

For a random sample \( x=(x_1,\ldots,x_n) \) of size \( n \), the total log-likelihood is written as
\[ l_n = l_n(a,b,\lambda,\alpha,\beta) = \sum_{i=1}^{n} l_i \]
where \( l_i \) denotes the log-likelihood for the \( i \)th observation \((i=1,\ldots,n)\).

The total score function is \( U_n = \sum_{i=1}^{n} U_i \) where \( U_i \) is the score vector for the \( i \)th observation. The MLE \( \hat{\theta} \) of \( \theta \) can be obtained numerically from the non-linear equations \( U_n=0 \).

For interval estimation and testing of hypotheses for the parameters in \( \theta \), the \( 5 \times 5 \) information matrix is obtained as
\[ K = K(\theta) = \left( \begin{array}{cc} K_{aa} & K_{ab} & K_{a\lambda} & K_{a\alpha} & K_{a\beta} \\ K_{ba} & K_{bb} & K_{b\lambda} & K_{b\alpha} & K_{b\beta} \\ K_{a\lambda} & K_{b\lambda} & K_{\lambda\lambda} & K_{\lambda\alpha} & K_{\lambda\beta} \\ K_{a\alpha} & K_{b\alpha} & K_{\alpha\lambda} & K_{\alpha\alpha} & K_{\alpha\beta} \\ K_{a\beta} & K_{b\beta} & K_{\beta\lambda} & K_{\beta\alpha} & K_{\beta\beta} \end{array} \right) \]

The expressions for the elements of \( K \) are given in the appendix. The total information matrix is \( K_n = K_n(b) = nK(\theta) \).

Under certain regularity conditions which include that the true parameter values must lie in the interior of the parameter space, the expressions in the appendix hold.

<table>
<thead>
<tr>
<th>n</th>
<th>Parameter</th>
<th>Mean</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>( \hat{a} )</td>
<td>7.9999</td>
<td>.00250</td>
</tr>
<tr>
<td></td>
<td>( \hat{b} )</td>
<td>5.020</td>
<td>.00672</td>
</tr>
<tr>
<td></td>
<td>( \hat{\lambda} )</td>
<td>1.0018</td>
<td>.00859</td>
</tr>
<tr>
<td></td>
<td>( \hat{\alpha} )</td>
<td>.4996</td>
<td>.00384</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta} )</td>
<td>1.5480</td>
<td>.08640</td>
</tr>
<tr>
<td>100</td>
<td>( \hat{\lambda} )</td>
<td>1.0012</td>
<td>.003681</td>
</tr>
<tr>
<td></td>
<td>( \hat{\alpha} )</td>
<td>.4997</td>
<td>.001691</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta} )</td>
<td>1.5342</td>
<td>.05969</td>
</tr>
<tr>
<td></td>
<td>( \hat{a} )</td>
<td>7.9999</td>
<td>.000045</td>
</tr>
<tr>
<td></td>
<td>( \hat{b} )</td>
<td>5.005</td>
<td>.001338</td>
</tr>
<tr>
<td>200</td>
<td>( \hat{\lambda} )</td>
<td>1.0005</td>
<td>.001418</td>
</tr>
<tr>
<td></td>
<td>( \hat{\alpha} )</td>
<td>.4998</td>
<td>.000702</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta} )</td>
<td>1.5229</td>
<td>.03927</td>
</tr>
<tr>
<td></td>
<td>( \hat{a} )</td>
<td>7.9999</td>
<td>.000032</td>
</tr>
<tr>
<td></td>
<td>( \hat{b} )</td>
<td>5.003</td>
<td>.001134</td>
</tr>
<tr>
<td>300</td>
<td>( \hat{\lambda} )</td>
<td>1.0003</td>
<td>.001180</td>
</tr>
<tr>
<td></td>
<td>( \hat{\alpha} )</td>
<td>.4999</td>
<td>.00050</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta} )</td>
<td>1.5186</td>
<td>.03394</td>
</tr>
</tbody>
</table>
the asymptotic distribution of
\[ \sqrt{n}(\hat{\theta} - \theta) \sim N(0, K(\theta)^{-1}). \]

This provides a basis for the construction of tests of hypotheses and confidence regions. An asymptotic confidence interval with significance level \( \gamma \) for each parameter \( \theta_i \) is given by
\[ \left( \hat{\theta}_i - z_{\gamma/2} \sqrt{k^{i,i}}, \hat{\theta}_i + z_{\gamma/2} \sqrt{k^{i,i}} \right), \]
where \( k^{i,i} \) is the \( i \)th diagonal element of \( K(\theta)^{-1} \) for \( i = 1, 2, 3, 4, 5 \) and \( z_{\gamma/2} \) is the upper \( \gamma/2 \) point of standard normal distribution.
To select the best model among a range of models, the criteria used are Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and the second order Akaike information criterion (AICc). For $k$ as the number of free parameters in the model and sample size $n$, these are defined as

\[
\text{AIC} = 2k - 2\ln(\text{likelihood})
\]

\[
\text{BIC} = k\ln(n) - 2\ln(\text{likelihood})
\]

\[
\text{AICc} = \text{AIC} + \frac{2k(k+1)}{n-k-1}
\]

The likelihood can be increased by adding parameters, but doing so may result in overfitting. The BIC resolves this problem by introducing a penalty term for the number of parameters in the model. AICc, the second order information criterion, takes into account the sample size by increasing the relative penalty for model complexity with small data sets.

The best model is the one with least values of AIC, BIC and AICc. In case of small sample size or large no. of parameters, AICc is preferred over AIC.

For partitioned $\theta = (\theta_1^0, \theta_2^0)^T$, testing for

\[
H_0: \theta_1 = \theta_1^0 \text{ versus } H_1: \theta_1 \neq \theta_1^0
\]

can be carried out using one of the three well known asymptotically equivalent test statistics-namely, the likelihood ratio (LR), Rao score ($S_R$) and Wald (W) statistics. For testing the validity of some sub models of the BGW distribution for fitting to some available data, the maximum values of the restricted and unrestricted likelihoods are used for construction of the LR statistic. This statistic can be written as

\[
D = -2\ln\left(\frac{\text{likelihood under null hypothesis}}{\text{likelihood under alternative hypothesis}}\right)
\]

Under $H_0$, $D$ follows approximately Chi-square distribution with $d$ degrees of freedom, where $d$ is the dimension of vector $\theta_1$ of interest.

We apply the likelihood ratio test to a real data set in the next section.

### 6. Simulations and applications

For validating the theoretical results reported in Section 5, we carry out simulations by generating $n$ observations from BGW distribution with parametric values $a = 8$, $b = .5$, $\lambda = 1$, $\alpha = .5$ and $\beta = 1.5$. The estimated values of the parameters are found using the BFGS quasi-newton method in R package. The considered sample sizes are $n = 50, 100, 200$ and $300$ and $10000$ is the number of repetitions. The estimates for different sample sizes are reported in Table 1. The values in Table 1 substantiate the theoretical results reported in Section 5. It is observed that for $n = 300$, the estimated parameters are quite close to the assumed parametric values.

We demonstrate the superiority of the new distribution over its sub models using two real data sets arising in reliability engineering.

Data set 1: This data set gives the time-to-failure ($10^3$ h) of turbocharger of one type of engine given in Xu et al. [39]. The data

![Fig. 8. TTT plot to time-to-failure of components of Aarset data.](image)

<table>
<thead>
<tr>
<th>Models</th>
<th>Value of LR Statistics</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BGE versus BGW</td>
<td>6.745</td>
<td>.0093</td>
</tr>
<tr>
<td>BW versus BGW</td>
<td>8.697</td>
<td>.0031</td>
</tr>
<tr>
<td>GW versus BGW</td>
<td>25.298</td>
<td>.0000</td>
</tr>
<tr>
<td>BE versus BGW</td>
<td>19.368</td>
<td>.0000</td>
</tr>
<tr>
<td>GE versus BGW</td>
<td>24.724</td>
<td>.0000</td>
</tr>
</tbody>
</table>

Table 4

Estimates of the parameters for some models fitted to Aarset data set and the values of the AIC, BIC and AICc.

<table>
<thead>
<tr>
<th>Model</th>
<th>MLEs</th>
<th>AIC</th>
<th>BIC</th>
<th>AICc</th>
</tr>
</thead>
<tbody>
<tr>
<td>BGW</td>
<td>.587</td>
<td>.315</td>
<td>.016</td>
<td>.136</td>
</tr>
<tr>
<td>BGE</td>
<td>.221</td>
<td>.043</td>
<td>.418</td>
<td>3.41</td>
</tr>
<tr>
<td>BW</td>
<td>.124</td>
<td>.313</td>
<td>.015</td>
<td>1</td>
</tr>
<tr>
<td>GW</td>
<td>1</td>
<td>.013</td>
<td>.088</td>
<td>5.63</td>
</tr>
</tbody>
</table>

Fig. 8. TTT plot to time-to-failure of components of Aarset data.
set is:
1.6, 2.0, 2.6, 3.0, 3.5, 3.9, 4.5, 4.6, 4.8, 5.0, 5.1, 5.3, 5.4, 5.6, 5.8, 6.0, 6.1, 6.3, 6.5, 6.5, 
6.7, 7.0, 7.1, 7.3, 7.3, 7.7, 7.7, 7.8, 7.9, 8.0, 
8.1, 8.3, 8.4, 8.4, 8.5, 8.7, 8.8, 9.0.

As displayed in Fig. 6, the TTT plot for the above data has concave shape implying that this data set possesses an increasing failure rate.

The superiority of BGW is established using AIC, BIC and the AICc, LR statistic and plots of the histogram and the estimated densities. Table 2 displays the values of the MLEs of the parameters and the values of the AIC, BIC and AICc for six fitted models. It is seen that the values of AIC, BIC and AICc for the BGW model are the lowest. Hence our model can be labeled as the best model.

The histogram of the data set and the plots of the estimateddensities of all six distributions are shown in Fig. 7. This figure also depicts that the BGW distribution produces a better fit than all its sub models.

The values of the LR statistic for testing the goodness of fit of BGW in comparison with its sub models BGE, BW, GW, GR, BE and GE are given in Table 3. For example, to test BGE versus BGW, the hypotheses are 

H0: \( \theta_1: \beta = 1 \) versus H1: \( \theta_1: \beta \neq 1 \) and for testing GW versus 
BGW, the hypotheses are 

H0: \( \theta_1: (a,b) = 1 \) versus H1: \( \theta_0 \) is not true.

The values in Table 3 indicate that the null hypothesis will be rejected in all the cases. This helps in concluding that the proposed distribution fits the data set more closely as compared to some of its sub models.

Data set 2: We consider another data set from Aarset [40], on lifetimes of 50 components. The data set is:

0.10,0.20,1.1,1.1,1.1,1.2,3.6,7.11,12,18,18,18,18,21,32,36,40,
45,46,47,47,50,55,60,63,63,67,67,67,67,72,75,79,82,82,83,
84,84,84,85,85,85,85,85,85,86,86.

The TTT plot in Fig. 8 has first a convex shape and then a 
concave shape. It depicts that the data set possesses the bathtub 
shaped failure rate. Table 4 displays the values of the MLEs of the 
parameters and the values of the AIC, BIC and AICc for BGE, 
BGW, BW and GW models. On the basis of tabulated values, 
we conclude that BGW provides the best fit as compared to its sub models.

7. Conclusions

We propose a new five-parameter distribution, named as the 
Beta Generalized Weibull (BGW) distribution, which is a general-
ization of the BGE, BW, GW, GR, BE, GE, Weibull, Rayleigh and 
Exponential distributions. As the proposed distribution can have 
monotone, bathtub shaped and unimodal failure rate for different 
parametric combinations, it is expected that this generalization 
will be widely applicable in reliability engineering and biological 
sciences. It is shown that the BGW density can be expressed as 
a mixture of GW densities. Some expansions for the cdf and pdf 
are provided and the kth order moment of the BGW distribution 
has been derived. It is also observed that the method of maximum 
likelihood estimation can be applied by solving non-linear equa-
tions through numerical methods. For testing goodness of fit of 
the BGW distribution versus some of its special cases, the like-
lihood ratio test is applied. An application of the BGW distribu-
tion to two real data sets is provided to illustrate that this distribu-
tion provides a better fit than its sub-models.

Acknowledgment

The authors are grateful to the anonymous referees for their 
valuable suggestions that helped in improving the paper to a great 
extent. The first author is also thankful to University Grants Commiss-
ion, Government of India, for providing financial support for 
this work.

Appendix A

Proof of Theorem 1.

(A) (i) Using (8) in (5), the cdf of the BGW distribution is

\[ F(x) = \frac{1}{B(a,b)} \int_{0}^{1} e^{-\frac{1-e^{-\theta x^a}}{\theta x^a}} \left(1-1_{y=1}w^{b-1} \right) \, dw \]

\[ = \frac{1}{B(a,b)} \sum_{j=0}^{\infty} \left( -1 \right)^j \left( \frac{1}{Gamma(b-j)} \frac{w^{b+j-1}}{(a+j)} \right) \]

\[ = \frac{1}{B(a,b)} \sum_{j=0}^{\infty} \left( -1 \right)^j \left( 1-1_{y=1}w^{b+j} \right) \frac{Gamma(b-j)}{(a+j)} \frac{1}{Gamma(a+b)} \]

\[ Hence \ F(x) = \sum_{j=0}^{\infty} \left( \theta_j(x) \right) \]

(ii) The result follows from A (i).

(iii) \[ f(x) = \frac{\theta^2 x^{\theta_2-1}}{B(a,b)} e^{-\theta x^a} \left( 1-1_{y=1}w^{b-1} \right) \left( 1-1_{y=1}w^{b+j} \right) \frac{Gamma(b-j)}{(a+j)} \frac{1}{Gamma(a+b)} \]

(B) (i) For general a and b, using in (1), the relation

\[ \int_{0}^{G(x)} w^{a-1}(1-w)^{b-1} \, dw = \frac{(G(x))^a}{a} F_1(a,1-b,a+1;G(x)) \]

we get

\[ F(x) = \frac{1}{B(a,b)} \left( \frac{G(x)}{a} \right)^a \frac{1}{a} F_1(a,1-b,a+1;G(x)) \]

where \( F_1(a,1-b,a+1;G(x)) \) is the hypergeo-

metric function and \( (z)_i = z(z+1) \ldots (z+i-1) \) denotes 
the ascending factorial.

(ii) The survival function

\[ S(x) = 1 - F(x) = \frac{1}{B(a,b)} \frac{1}{b} \frac{1}{b} F_1(b,1-a,b+1,1-G(x)) \]

(iii) \[ S(x) = 1 - F(x) = 1 - \frac{1}{b} F_1(b,1-a,b+1,1-G(x)) \]

(C) It is well known that the density function of ith order statistic 
\( X_{i:n} \) can be written as

\[ f_{i:n}(x) = \frac{1}{B(i,n-i+1)} \left( x \right)^{i-1} \left( 1-x \right)^{n-i} \]

where \( i = 1, \ldots, n \).

Using (6) and the results of (B) (i) and (ii),

\[ f_{i:n}(x) = \frac{1}{B(i,n-i+1)} \left( \frac{x^{i-1}}{a} \right)^{\alpha_i} G(x)_{i:a} \]

\[ = \frac{1}{B(a,b)} \left( \frac{G(x)}{a} \right)^a \frac{1}{a} F_1(a,1-b,a+1;G(x)) \]

\[ \left( G(x)_{i:a} \right)^{1-i} \]
A.1. Expressions for elements of information matrix K

For $i, j, k, l, m, n \in \{0, 1, 2\}$ and $V$ following Beta distribution with parameters $a$ and $b$, we define

$$T_{ij,k,l,m,n} = E[(1-V)^{-a}(1-V)^{b}(\log(1-V^{1/3}))^{j}] 	imes (\log V)^{k} \log(-\log(1-V^{1/3}))^{1/\beta}].$$

If $\psi(.)$ denotes the derivative of the digamma function, then the elements of information matrix $K$ are given by

$$K_{a,b} = -E\left[ \frac{\partial^2 l}{\partial a \partial b} \right] = \psi(a) - \psi(a+b),$$

$$K_{a,c} = \psi(a+b),$$

$$K_{a,i} = -E\left[ \frac{\partial^2 l}{\partial a \partial \alpha} \right] = -E\left[ \frac{\log(1-e^{-x})}{1-e^{-x}} \right] = \psi(a+b) - \psi(a),$$

$$K_{a,b} = -E\left[ \frac{\partial^2 l}{\partial a \partial b} \right] = \psi(a+b) - \psi(a),$$

$$K_{b,b} = \psi(b) - \psi(a+b),$$

$$K_{b,i} = E\left[ \frac{\partial^2 l}{\partial b \partial \alpha} \right] = -E\left[ \frac{\log(1-e^{-x})}{1-e^{-x}} \right] = \psi(a+b) - \psi(a),$$

$$K_{b,b} = \psi(b) - \psi(a+b),$$

$$K_{a,i} = E\left[ \frac{\partial^2 l}{\partial a \partial \alpha} \right] = -E\left[ \frac{\log(1-e^{-x})}{1-e^{-x}} \right] = \psi(a+b) - \psi(a),$$

$$K_{b,b} = \psi(b) - \psi(a+b),$$

$$K_{a,b} = -E\left[ \frac{\partial^2 l}{\partial a \partial b} \right] = \psi(a+b) - \psi(a),$$

$$K_{a,b} = \psi(a+b) - \psi(a).$$

References


