

Modeling by Beta Gompertz Distribution Based on K-S Test Statistic

Hanaa Abu-Zinadah ^{1*} and Asmaa Binkhamis ²

¹ Statistics Department, College of Science, University of Jeddah, Jeddah, Saudi Arabia.

² Statistics Department, College of Science, University of Jeddah, Jeddah, Saudi Arabia.

*Corresponding author email id: hhabuznadah@uj.edu.sa

Date of publication (dd/mm/yyyy): 30/12/2019

Abstract – The goodness of fit test for beta Gompertz (BGpz) distribution with two unknown parameters were studied by using the common goodness of fit test statistic (Kolmogorov-Smirnov (K-S)) based on a complete sample. A table of critical values obtained based the K-S test statistic for the BGpz distribution. Monte Carlo simulation have been carried out for power calculations to test different alternative distributions. Finally, two real data sets were used to test the proposed distribution with other well-known distributions.

Keywords – Beta Gompertz Distribution, Monte Carlo Simulation, Goodness of Fit Test, Complete Sampling.

I. INTRODUCTION

Jafari et al. [1] introduced a new generalization of Gompertz which is called the beta Gompertz (BGpz) distribution. The beta and Gompertz are the most popular distributions for analyzing lifetime data. Also, it is useful in different areas of statistics. The probability density function (PDF) and the cumulative distribution function (CDF) for the BGpz distribution are given respectively by:

$$f(x; \Phi) = \frac{\lambda \alpha}{\beta(a, b)} e^{\alpha x - \lambda (e^{\alpha x - 1})} (1 - e^{-\lambda (e^{\alpha x - 1})})^{a-1} [e^{-\lambda (e^{\alpha x - 1})}]^{b-1}; \quad x > 0 \text{ and } a, b, \alpha, \lambda > 0, \quad (1)$$

$$F(x; \Phi) = \frac{1}{\beta(a, b)} \int_0^{1 - e^{-\lambda (e^{\alpha x - 1})}} w^{a-1} (1 - w)^{b-1} dw; \quad (2)$$

$x > 0$ and $a, b, \alpha, \lambda > 0$,

where $\beta(\cdot, \cdot)$ is the Beta function, a, b and λ are shape parameters, α is a scale parameter and $\Phi = (a, b, \alpha, \lambda)^T$ is the parameters vector of the BGpz distribution.

Many authors adding one or more parameters to a distribution function increases the flexibility of the distribution for modeling the data. The beta normal distribution was introduced by Eugene et al. [2]. Lee et al. [3] considered the beta Weibull distribution and studied applications based on censored data. Benkhelifa [4] proposed the beta generalization Gompertz distribution with five new parameters. The exponentiated Gompertz distribution, proposed by Abu-Zinadah and Al-Oufi [5] addressed some of its characteristics and considered a particular estimations approach on a complete and Type-II censored samples. In addition, Ade et al. [6] presented the exponentiated generalized Gompertz Makeham distribution. Oguntunde et al. [7] introduced a new generalization of the Lomax distribution with Increasing, decreasing and constant failure rate. Hajar and Abu-Zinadah [8] discussed some statistical properties for the generalized exponentiated Gompertz distribution with five parameters that is called the beta exponentiated Gompertz Distribution.

On the other hand, many authors were interested about the goodness of fit tests. Hassan [9] studied the goodness

of fit tests for the generalized exponential distribution. Archer and Lemeshow [10] presented the goodness of fit tests for logistic regression model fitted using survey sample data. Abdelfattah [11] considered the goodness of fit tests for Weibull distribution with two parameters. Also, goodness of fit tests for the exponential distribution based on progressively Type-II censored sample introduced by Wang [12]. Shawky and Bakoban [13] participated in modifying the goodness of fit tests for an exponentiated gamma distribution with an unknown shape parameter based on complete and Type-II censored samples. Abd-Elfattah et al. [14] studied the goodness of fit tests for the generalized Frechet. Also, Abd-Elfattah [15] discussed the goodness of fit tests for the generalized Rayleigh distribution with unknown parameters. Abu-Zinadah [16] participated the goodness of fit tests for the exponentiated Gompertz distribution based on Type-II censored sampling. Later on, Lenart and Missov [17] introduced the goodness of fit tests for the Gompertz distribution. Al-Omari and Zamanzade [18] considered the different goodness of fit tests for Rayleigh distribution based on simple random sampling and ranked set sampling techniques. Zamanzade and Mahdizadeh [19] developed some goodness of fit tests for Rayleigh distribution based on Phi-divergence. Also, Badr [20] studied the goodness of fit tests for the compound Rayleigh distribution with application to real data for complete and Type-II censored samples.

The objective of this work is to study the goodness of fit test for beta Gompertz (BGpz) distribution on a complete sampling with two unknown parameters based on (Kolmogorov-Smirnov (K-S)) test statistic. The paper is arranged in the following order: the parameters were estimated by using the maximum likelihood estimation where it derives the unknown two parameters a and b based on complete sample in Section 2. The Monte Carlo technique was obtained for the critical values table of K-S test statistic with different sample sizes are provided in Section 3. For some alternative distributions, the power of the K-S test is evaluated based on complete sampling in Section 4. Finally, in Section 5 real data set examples are applied to presented the proposed distribution as a better than others distributions for modeling the life data sets.

II. ESTIMATION OF UNKNOWN PARAMETERS

Suppose a complete sample $\underline{x} = (x_1, x_2, \dots, x_n)$ where x_i the i^{th} is order statistics. The sample is obtained from BGpz distribution with PDF and CDF are given respectively, by equations (1) and (2). The maximum likelihood (ML) estimators of the two unknown parameters a and b from the parameters of BGpz (α, λ, a, b) can be obtained by using: The likelihood function of BGpz distribution:

$$l(\underline{x}; \Phi) = \left(\frac{\lambda\alpha}{\beta(a,b)}\right)^n \prod_{i=1}^n e^{\alpha x_i} \prod_{i=1}^n e^{-\lambda b(e^{\alpha x_i}-1)} \prod_{i=1}^n \left(1 - e^{-\lambda(e^{\alpha x_i}-1)}\right)^{a-1} \tag{3}$$

The logarithm of likelihood function is given by:

$$\log l(\underline{x}; \Phi) = n \log \lambda + n \log \alpha - n \log \beta(a,b) + \alpha \sum_{i=1}^n x_i - \lambda b \sum_{i=1}^n (e^{\alpha x_i}) + \lambda b n + (a - 1). \tag{4}$$

The ML estimators of a and b can be derived by solving the following likelihood equations:

$$\frac{\partial L}{\partial a} = n \psi(a + b) - n \psi(a) + \sum_{i=1}^n \log (1 - e^{-\lambda(e^{\alpha x_i} - 1)}) = 0 \tag{5}$$

$$\frac{\partial L}{\partial b} = n \psi(a + b) - n \psi(b) + \lambda n - \lambda \sum_{i=1}^n (e^{\alpha x_i}) = 0 \tag{6}$$

where $\psi(\cdot)$ is called the Psi function. Equations (5) and (6) should be solved by an iterative Newton Raphson method to find the solution of the system of nonlinear equations since the equations do not have a solution in a compact form.

In Table I, the Monte Carlo method were simulated 1000 times to estimate the unknown parameters a and b of $BGpz(\alpha, \lambda, a, b)$ for different sample sizes based on complete sample. The results of the estimators of the parameters has been evaluated in terms of their values of relative root mean square error (RRMSE) and absolute relative bias (ARBias), where

$$RRMSE(\hat{\theta}) = \frac{\sqrt{MSE(\hat{\theta} - \theta)}}{\theta} \quad \text{and} \quad ARBias(\hat{\theta}) = \left| \frac{\hat{\theta} - \theta}{\theta} \right|.$$

Table I. The ML estimates, RRMSE's and ARBias's of parameter a.

N	\hat{a}	RRMSE	ARBias
10	2.7580	0.885592	0.378998
30	2.19961	0.310203	0.0998064
50	2.12445	0.229917	0.0622228
100	2.0574	0.144887	0.0286992
150	2.04267	0.115974	0.0213364

Table II. The ML estimates, RRMSE's and ARBias's of parameter b.

N	\hat{b}	RRMSE	ARBias
10	2.77961	0.899857	0.389806
30	2.19892	0.310855	0.0994584
50	2.11504	0.217861	0.0575186
100	2.05997	0.147825	0.0299858
150	2.03824	0.114489	0.0191221

From Tables I and II, the following observations were made:

1. At large sample sizes, the values of RRMSE's and ARBias's decreases, which indicates that the estimators of a and b are consistent.
2. The estimates decrease when the sample size increase.

III. GOODNESS OF FIT TEST

The goodness of fit tests validates the closeness of the theoretical distribution function to the empirical distribution function (EDF). When these tests are designed for a null hypothesis H_0 about the distribution function $F(x)$ of the form $H_0: F(x) = F_n(x)$, where $F(x) = P(X < x)$ is a specified family of cumulative distribution functions and $F_n(x)$ is the EDF. Assume that a random sample x_1, x_2, \dots, x_n where x_i the i^{th} is order statistics had $F(x_i; \hat{\Phi})$ of BGpz distribution on complete sampling with sample size n and vector parameters $\hat{\Phi}$. The EDF $F_n(x)$, is defined as

$$F_n(x) = \begin{cases} 0, & x < X_{(1)} \\ \frac{i}{n}, & X_{(i)} \leq x < X_{(i-1)}, \quad i = 1, 2, \dots, n \\ 1, & X_{(n)} \leq x \end{cases}$$

Note that $F_n(x)$ is a step function that takes a step of height $\frac{1}{n}$ at each ordered sample observation. The common goodness of fit test statistic based on the EDF is K-S test statistic which defined by: $\hat{D} = \max(\hat{D}^+, \hat{D}^-)$, where $\hat{D}^+ = \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - F(x_i; \hat{\Phi}) \right\}$, $\hat{D}^- = \max_{1 \leq i \leq n} \left\{ F(x_i; \hat{\Phi}) - \frac{i-1}{n} \right\}$, and $\hat{\Phi}$ is the estimator of Φ .

Table II shows the critical values that were made by the Monte Carlo simulation of the K-S test statistics for the BGpz distribution with two unknown parameters and based on a complete sample of n (10, 30, 50, 100 and 150) and levels of significance ν (0.2, 0.15, 0.1, 0.05 and 0.01) In this research the following hypothesis test as:

H_0 : The complete data comes from a cumulative distribution function of the BGpz distribution.

H_1 : The complete data comes from a cumulative distribution function of another distribution.

Table III. Critical values of the K-S test statistic based on complete samples when a, b are unknown and $\alpha = \lambda = 2$.

n	Critical Values				
	$\nu = 0.20$	$\nu = 0.15$	$\nu = 0.10$	$\nu = 0.05$	$\nu = 0.01$
10	0.223074	0.235372	0.248753	0.269933	0.318714
30	0.135193	0.141561	0.149408	0.159222	0.184327
50	0.10616	0.111388	0.119506	0.132395	0.151572
100	0.0762922	0.0796378	0.0838155	0.0904139	0.101915
150	0.061361	0.0642331	0.0685038	0.0743557	0.0878589

IV. POWER STUDY

The power of the goodness of fit test statistic is defined as the probability that a statistic will lead to the rejection of the H_0 . When the H_0 is rejected, the best alternative distribution will be found for BGpz distribution. The power of K-S test statistic was obtained by generating a 1000 random sample of size n with different significance levels ν based on a complete sample by using the Monte Carlo simulation. The common alternative distributions are log-normal, exponential, gamma, Chi-square with one degree of freedom, Weibull and Frechet distribution are considered as alternative distributions.

Table IV. Power of K-S test statistic based on complete samples when a, b are unknown and $\alpha = \lambda = 2$.

Alternative Distribution	n	Power of the test				
		$v = 0.20$	$v = 0.15$	$v = 0.10$	$v = 0.05$	$v = 0.01$
Log- normal	10	0.616	0.544	0.482	0.390	0.206
	30	0.609	0.559	0.492	0.386	0.246
	50	0.609	0.549	0.482	0.409	0.253
	100	0.586	0.535	0.456	0.367	0.202
	150	0.596	0.523	0.438	0.337	0.187
Exponential	10	0.615	0.548	0.487	0.411	0.271
	30	0.615	0.555	0.476	0.388	0.234
	50	0.624	0.550	0.467	0.373	0.197
	100	0.587	0.543	0.463	0.377	0.217
	150	0.593	0.526	0.466	0.363	0.228
Chi-square	10	0.614	0.552	0.478	0.396	0.230
	30	0.620	0.556	0.478	0.380	0.204
	50	0.552	0.487	0.396	0.306	0.174
	100	0.613	0.560	0.494	0.388	0.238
	150	0.597	0.543	0.464	0.363	0.167
Gamma	10	1	1	1	1	0.996
	30	0.992	0.992	0.979	0.948	0.805
	50	0.919	0.896	0.830	0.708	0.463
	100	0.510	0.435	0.346	0.238	0.069
	150	0.185	0.146	0.092	0.04	0.005
Weibull	10	1	1	1	1	1
	30	1	0.999	0.998	0.993	0.958
	50	0.985	0.975	0.962	0.935	0.794
	100	0.832	0.780	0.712	0.619	0.388
	150	0.596	0.522	0.448	0.351	0.194
Frechet	10	1	1	1	1	1
	30	0.998	0.996	0.996	0.995	0.965
	50	0.985	0.976	0.964	0.934	0.781
	100	0.818	0.787	0.719	0.620	0.403
	150	0.595	0.533	0.459	0.366	0.184

From Table IV, the following notes were considered:

1. The Frechet distribution is the best alternative distribution in most of the cases.
2. The Weibull distribution is the best alternative distribution at small sample size ($n < 50$).
3. The exponential, chi-square with one degree of freedom distributions are a good alternative distributions at ($n = 150$).
4. The log-normal is worst alternative distribution.
5. The power of K-S test statistic decreased when the significant level decrease (positive relationship).

V. REAL DATA SET EXAMPLES

In this section, the real data sets were applied to present which distribution is a good model for fits these data sets. The goodness of fit measures are selected to present which distribution was a good model for these data sets. The comparison was made between the BGpz distribution with its sub models such as the Gompertz (Gpz) and exponential (E) distributions. The model selection technique such as Akaike information criterion (AIC), Bayesian information criterion (BIC), consistent Akaike information criterion (CAIC) and Hannan-Quinn information criterion (HQIC) were carried out. These models are defined as: $AIC = -2 \mathbf{l}(\hat{\Phi}) + 2k$, $BIC = -2 \mathbf{l}(\hat{\Phi}) + k \log(n)$, $CAIC = -2\mathbf{l}(\hat{\Phi}) + \frac{2kn}{n-k-1}$, $HQIC = -2\mathbf{l}(\hat{\Phi}) + 2k \log(\log(n))$, where $\mathbf{l}(\hat{\Phi})$ denotes the log-likelihood function evaluated at the maximum likelihood estimate, k is the number of unknown parameters and n is the sample size.

1. Example

Gross and Clark [21] used this data set represents the lifetime's data relating to relief times (in minutes) of 20 patients receiving an analgesic: 1.1, 1.4, 1.3, 1.7, 1.9, 1.8, 1.6, 2.2, 1.7, 2.7, 4.1, 1.8, 1.5, 1.2, 1.4, 3, 1.7, 2.3, 1.6, 2. The Results of Goodness of Fit Measures are Shown in Tables V and VI.

Table V. The ML estimates of unknown parameter for different models.

Model	$\hat{\alpha}$	$\hat{\lambda}$	\hat{a}	\hat{b}
BGpz	0.313297	8.48359	55.2793	0.31561
Gpz	0.0974588	5.1427	-	-
E	-	0.526316	-	-

Table VI. The K-S, AIC, BIC, HQIC and CAIC for different models.

Model	K-S	AIC	BIC	HQIC	CAIC
BGpz	0.159616	40.6701	44.653	41.4476	43.3368
Gpz	0.441197	66.6454	68.6369	67.0342	67.3513
E	0.439512	67.6742	68.6699	67.8685	67.8964

2. Example

One hundred data of observations on breaking stress of carbon fibres (in Gba) in Cakmakyapan and Ozal [22]: 3.7, 2.74, 2.73, 2.5, 3.6, 3.11, 3.27, 2.87, 1.47, 3.11, 4.42, 2.41, 2.41, 3.19, 3.22, 1.69, 3.28, 3.09, 1.87, 3.15, 4.9, 3.75, 2.43, 2.95, 2.97, 3.39, 2.96, 2.53, 2.67, 2.93, 3.22, 3.39, 2.81, 4.2, 3.33, 2.55, 3.31, 3.31, 2.85, 2.56, 3.56,

3.15, 2.35, 2.55, 2.59, 2.38, 2.81, 2.77, 2.17, 2.83, 1.92, 1.41, 3.68, 2.97, 1.36, 0.98, 2.76, 4.91, 3.68, 1.84, 1.59, 3.19, 1.57, 0.81, 5.56, 1.73, 1.59, 2, 1.22, 1.12, 1.71, 2.17, 1.17, 5.08, 2.48, 1.18, 3.51, 2.17, 4.7, 2.03, 1.8, 1.57, 1.08, 2.03, 1.61, 2.12, 1.61, 2.12, 1.89, 2.88, 2.82, 2.05, 3.65. The results of goodness of fit measures are shown in Tables VII and VIII.

Table VII. The ML estimates of unknown parameter for different models.

Model	$\hat{\alpha}$	$\hat{\lambda}$	\hat{a}	\hat{b}
BGpz	0.320427	1.18685	3.16675	1.16716
Gpz	0.080966	4.51163	-	-
E	-	0.388676	-	-

Table VIII. The K-S, AIC, BIC, HQIC and CAIC for different models.

Model	K-S	AIC	BIC	HQIC	CAIC
BGpz	0.061879	294.4	304.821	298.617	294.821
Gpz	0.30749	376.648	381.858	378.757	376.772
E	0.306769	391.002	393.607	392.056	391.043

From Tables (V –VIII), the following observations can be concluded:

1. The ML estimators of unknown parameters for distributions is good estimates based on the AIC, BIC, HQIC and CAIC measures.
2. In all cases, the AIC has the smallest results which makes him the best measure for these data.
3. The BGpz distribution is better than the other distributions in modeling lifetime data based on the goodness of fit measures.

REFERENCES

[1] A. A. Jafari, S. Tahmasebi, and M. Alizadeh, “The Beta Gompertz Distribution”, *Revista Colombiana de Estadística (Colombian Journal of Statistics)*. 37 (2014) 139 – 156.

[2] N. Eugene, C. Lee, and F. Famoye, “Beta-Normal distribution and its applications”, *Communications in Statistics-Theory and Methods*. 31 (2002) 497-512.

[3] C. Lee, F. Famoye, and O. Olumolade, “The Beta-Weibull Distribution: Some Properties and Applications Censored Data”, *J. Modern Applied Statistical Method*. 6 (2007) 173–186.

[4] L. Benkhelifa “The Beta Generalized Gompertz Distribution”, *Applied Mathematical Modelling*. 52 (2017) 341–357.

[5] H. H. Abu-Zinadah, and A. S. Aloufi, “Some Characterizations of the Exponentiated Gompertz Distribution”, *International Mathematical Forum*. 9 (2014) 1427-1439.

[6] O.A. Ade, F. Olayode, and A. Bamidele, “Performance Rating of the Exponentiated Generalized Gompertz Makeham Distribution: an Analytical Approach”, *American Journal of Theoretical and Applied Statistics*. 6 (2017) 228-235.

[7] P. E. Oguntunde, M. A. Khaleel, M. T. Ahmed, A. O. Adejumo, and O. AOdetunmbi, “ A New Generalization of the Lomax Distribution with Increasing, Decreasing and Constant Failure Rate”, *Hindawi, Forthcoming*. Volume (2017), Article ID 6043169 6 pages.

[8] N. M. Hajar, and H. H. Abu-Zinadah, “ Some Statistical Properties of The Beta Exponentiated Gompertz Distribution”, *IOSR Journal of Mathematics*. 14 (2018) 31-40.

[9] A. S. Hassan, “ Goodness-of-Fit for the Generalized Exponential Distribution”, *Inter Stat*, 1-15, 2005.

[10] J. K. Archer, and S. Lemeshow, “Goodness of Fit Test for Logistic Regression Model Fitted Using Survey Sample Data”, *The Stata Journal*. 6 (2006) 97-105.

[11] E. H. Abdelfattah, “ Goodness of Fit Tests for the Two Parameter Weibull Distribution”, *J. Statist. Theory Appl*. 7 (2008) 279–291.

[12] B. Wang, “ Goodness-of-Fit Test for The Exponential Distribution Based Progressively Type-II Censored Sample”, *Journal of Statistical Computation and Simulation*. 78 (2008) 125–132

[13] A. I. Shawky, and R. A. Bakoban, “Modified Goodness-of-Fit Tests for Exponentiated Gamma Distribution with Unknown Shape Parameter”, *InterStat*. 3 (2009 July) 1-17.

[14] A. M. Abd-Elfattah, H. A. Fergany, H. A. and A. M. Omima, “ Goodness of Fit Tests for The Generalized Frechet Distribution”, *Australian Journal of Basic and Applied Sciences*. 4 (2010) 286-301.

[15] A. M. Abd-Elfattah, “ Goodness of Fit Test for The Generalized Rayleigh Distribution”, *Journal of Statistical Computation and Simulation*. 81 (2011) 357–366.

[16] H. H. Abu-Zinadah “Goodness-of-Fit Tests for the Exponentiated Gompertz distribution”, *International Journal of Innovation in Science and Mathematics*. 2 (2014) 327–380.

- [17] A. Lenart, and T. I. Missov, "Goodness-of-Fit Tests for the Gompertz Distribution", Taylor and Francis group. 45 (2016) 2920-2937.
- [18] A. I. Al-Omari, and E. Zamanzade, "Different Goodness of Fit Tests for Rayleigh Distribution in Ranked Set Sampling", Pakistan Journal of Statistics and Operation Research. 12 (2016) 25-39.
- [19] E. Zamanzade, and M. Mahdizadeh, " Goodness-of-Fit Tests for Rayleigh Distribution Based on Phi-Divergence", Revista Colombiana de Estadística (Colombian Journal of Statistics). 40 (2017) 279-290.
- [20] M. M. Badr, "Goodness-of-Fit Tests for the Compound Rayleigh Distribution with Application to Real Data", Heliyon. 5 (2019) 1-6.
- [21] A. J. Gross, and V. A. Clark, "Survival Distributions: Reliability Applications in the Biometrical Sciences", John Wiley, New York, USA. (1975).
- [22] S. Cakmakyapan, and G. Ozal, "Lindley-rayleigh distribution with application to lifetime data", Journal of Reliability and Statistical Studies, 11 (2018) 09-24.

AUTHOR'S PROFILE



First Author

Hanaa H. Abu-Zinadah Associate Professor of Mathematical Statistics, Statistics Department, College of Science, University of Jeddah, Jeddah, Kingdom of Saudi Arabia, Born in Jeddah, Kingdom of Saudi Arabia on April 1976. 1996: Bachelor degree of Mathematic from Mathematic Department, Scientific Section, Girls College of Education in Jeddah, Kingdom of Saudi Arabia. 2001: Master degree of Mathematical Statistics from Mathematic Department, Scientific Section, Girls College of Education in Jeddah, Kingdom of Saudi Arabia. 2006: Doctorate degree of Mathematical Statistics from Mathematic Department, Girls College of Education in Jeddah, Scientific Section, King Abdulaziz University. Dr. Hanaa H. Abu-Zinadah the Head of Statistics Department, Faculty of Science - AL Faisaliah, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia from 2010 -2019 and Associate Professor of Mathematical Statistics, Statistics Department, College of Science, University of Jeddah, Jeddah, Kingdom of Saudi Arabia from 2019 until now. email id: hhabuznadah@uj.edu.sa.

Second Author

Asmaa Binkhamis Department of Statistics, Sciences Faculty, University of Jeddah, Jeddah, Saudi Arabia, Born in Jeddah, Kingdom of Saudi Arabia on July 1992. 2014: Bachelor degree of Science from the Faculty of Science in Statistics, King Abdulaziz University in Jeddah, Saudi Arabia. Asmaa K. Binkhamis the teaching assistant of Statistics Department, College of Science, University of Jeddah, Jeddah, Saudi Arabia.