GOODNESS-OF-FIT TESTS FOR THE BETA GOMPERTZ DISTRIBUTION

by

Hanaa ABU-ZINADAH [*] ***and Asmaa BINKHAMIS***

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

Original scientific paper
https://doi.org/10.2298/TSCI20S1069A

This article studied the goodness-of-fit tests for the beta Gompertz distribution with four parameters based on a complete sample. The parameters were estimated by the maximum likelihood method. Critical values were found by Monte Carlo simulation for the modified Kolmogorov-Smirnov, Anderson-Darling, Cramer-von Mises, and Lilliefors test statistics. The power of these test statistics founded the optimal alternative distribution. Real data applications were used as examples for the goodness of fit tests.

Key words: *Beta Gompertz distribution, Monte Carlo simulation, goodness-of-fit test, complete sampling*

Introduction

The goodness-of-fit test is an important step in statistical analysis for lifetime data to select the best distribution that sufficiently fits the data. The important idea in the test is testing the null hypothesis, H_0, about empirical distribution function (EDF) $F_n(x)$ when the data in H_0 comes from a cumulative distribution function (CDF) $F(x)$. For more details about this and the goodness of fit tests for the exponentiated Gompertz (EGpz) distribution, see Abu-Zinadah [1]. The H_0 is equivalent to the following.

The H_0: $F(x) = F_n(x)$, where the EDF is defined:

$$F_n\left(x\right)=\frac{\text{number of observation}\leq x}{n}\ ,\quad -\infty<x<\infty$$

where $x_1, x_2,..., x_n$ a random sample from the distribution of X.

The probability density function (PDF) of beta Gompertz (BGpz) distribution is given:

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

where B(. , .) is the beta function.

[*] Corresponding author, e-mail: hhabuznadah@uj.edu.sa

The CDF of BGpz distribution is given:

$$F\left(x;\alpha,\lambda,a,b\right)=\frac{1}{\beta\left(a,b\right)}\int\limits_{0}^{1-e^{-\lambda(e^{\alpha x}-1)}} w^{a-1}\left(1-w\right)^{b-1}dw;\ \ x>0\ \text{ and }\ a,b,\alpha,\lambda>0$$

where a, b, and λ are shape parameters and α is a scale parameter.

There are many different famous lifetime distributions that can be used to model reliability data, one example is the Gompertz distribution. Rasheed *et al.* [2] compared some estimators of parameters for a basic Gompertz distribution, such as maximum likelihood and Bayesian estimators under generalized weighted loss functions by using a gamma prior distribution and Soliman *et al.* [3, 4] discussed estimates of model parameters under the generalized censored scheme. El-Gohary *et al.* [5] introduced a new generalized distribution that is called the generalized Gompertz distribution. Abu-Zinadah and Aloufi [6] introduced the exponentiated Gompertz distribution and discussed some of its characterizations. Also, Abu-Zinadah and Aloufi [7] discussed the maximum likelihood and Bayes estimations for the three parameters of the generalized Gompertz distribution under three different types of loss functions based on type II censored samples. Moreover, Abu-Zinadah and Aloufi [8] explained different methods of estimation of the three parameters of the exponentiated Gompertz distribution based on a complete sample. Jafari *et al.* [9] introduced the BGpz distribution with some statistical properties and the maximum likelihood estimation of its parameters. Benkhelifa [10] proposed Beta generalization Gompertz distribution with five new parameters, studied some properties and estimated the parameters. Hajar and Abu-Zinadah [11] discussed some statistical properties for the Beta exponentiated Gompertz distribution. Abu-Zinadah and Hajar [12] considered a classical and Bayesian estimations for parameters of the Beta exponentiated Gompertz distribution with real data application based on complete samples under three different types of loss functions. Also, Abu-Zinadah and Hajar [13] studied statistical inference for parameters of the Beta exponentiated Gompertz distribution according to type-II censoring by using Bayesian and non-Bayesian estimation method. Bakoban and Abu-Zinadah [14] considered the Beta generalizes inverted exponential distribution with real data application.

On the other hand, Shawky and Bakoban [15] participated in modifying the goodness of fit test for an exponentiated gamma distribution with an unknown shape parameter based on complete and type II censored samples and found the power of it. Later on, Lenart and Missov [16] introduced the goodness of fit for the Gompertz distribution with four measures and presented the critical values by an empirical distribution of the test statistics. Furthermore, Badr [17] studied the goodness of fit tests for the compound Rayleigh distribution with application real data for complete and type II censored samples.

The main aim of this paper is to discuss the goodness of fit tests for the BGpz distribution. The maximum likelihood estimation was used to estimate the parameters of the BGpz distribution under complete samples. Critical values were obtained by using the Mon Carlo simulation via different test statistics for the BGpz distribution with four parameters based on complete samples.

Maximum likelihood estimation

Suppose a complete sample $\underline{x}=(x_1, x_2,..., x_n)$ where x_i is the i^{th} order statistics. In this case, the likelihood function can be written:

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

The natural logarithm of likelihood function is given:

$$L=\log l\left(\underline{x};\alpha,\lambda,a,b\right)\propto n\log\lambda+n\log\alpha-n\log\beta\left(a,b\right)+\alpha\sum_{i=1}^{n}x_i-\lambda b\sum_{i=1}^{n}\left(\mathrm{e}^{\alpha x_i}\right)+ +\lambda bn+\left(a-1\right)\sum_{i=1}^{n}\ln\left[1-e^{-\lambda\left(\mathrm{e}^{\alpha x_i}-1\right)}\right] \tag{1}$$

The maximum likelihood equations can be written:

$$\frac{\partial L}{\partial\alpha}=\frac{n}{\alpha}+\sum_{i=1}^{n}x_i-\lambda b\sum_{i=1}^{n}x_i\left(\mathrm{e}^{\alpha x_i}\right)+\left(a-1\right)\lambda\sum_{i=1}^{n}\frac{x_i\left(\mathrm{e}^{\alpha x_i}\right)\mathrm{e}^{-\lambda\left(\mathrm{e}^{\alpha x_i}-1\right)}}{1-\mathrm{e}^{-\lambda\left(\mathrm{e}^{\alpha x_i}-1\right)}}=0 \tag{2}$$

$$\frac{\partial L}{\partial\lambda}=\frac{n}{\lambda}-b\sum_{i=1}^{n}\left(\mathrm{e}^{\alpha x_i}\right)+bn+\left(a-1\right)\sum_{i=1}^{n}\frac{\left(\mathrm{e}^{\alpha x_i}-1\right)\mathrm{e}^{-\lambda\left(\mathrm{e}^{\alpha x_i}-1\right)}}{1-\mathrm{e}^{-\lambda\left(\mathrm{e}^{\alpha x_i}-1\right)}}=0 \tag{3}$$

$$\frac{\partial L}{\partial a}=n\psi\left(a+b\right)-n\psi\left(a\right)+\sum_{i=1}^{n}\log\left[1-\mathrm{e}^{-\lambda\left(\mathrm{e}^{\alpha x_i}-1\right)}\right]=0 \tag{4}$$

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

where $\Psi(.)$ is called the Psi function, see Jeffrey and Dai [18]. These equations do not have a closed form solution for α, λ, a, and b. Therefore, it can be solved using the eqs. (2)-(5) with a numerical technique such as the Newton-Raphson method.

In tabs. 1 and 2, show the performance estimates for the different parameters of the BGpz distribution that were studied using the Monte Carlo simulation. The simulation has been repeated 1000 times with different sample sizes n = (10, 30, 50, 100, 150) based on complete samples from BGpz (2, 2, 2, 2). The maximum likelihood estimates (MLE), the values of relative root mean square error (RRMSE), and absolute relative bias (ARBias) of the parameters presented in tabs. 1 and 2, where:

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

Table1. The MLE, RRMSE, and ARBias of the parameters *α* and *λ*

n	*MLE* of α			*MLE* of λ		
	$\hat{\alpha}$	*RRMSE*	*ARBias*	$\hat{\lambda}$	*RRMSE*	*ARBias*
10	3.47236	1.70319	0.73618	2.79316	1.29396	0.39658
30	2.47695	0.98330	0.23848	2.96019	1.17982	0.48009
50	2.34467	0.78042	0.17234	2.82635	1.05849	0.41318
100	2.16919	0.57814	0.08459	2.78581	0.99177	0.39291
150	2.15038	0.49051	0.07519	2.63474	0.85660	0.31737

Table 2. The MLE, RRMSE, and ARBias of the parameters *a* and *b*

n	*MLE* of a			*MLE* of b		
	$\hat{a}$	*RRMSE*	*ARBias*	$\hat{b}$	*RRMSE*	*ARBias*
10	2.82753	2.13808	0.41377	1.95475	0.55801	0.02263
30	2.15841	0.32863	0.07920	1.98878	0.47252	0.00561
50	2.12610	0.27306	0.06305	1.96771	0.43127	0.01615
100	2.06513	0.19978	0.03257	1.92770	0.37286	0.03615
150	2.05830	0.16744	0.02915	1.91059	0.34432	0.04470

From tabs. 1 and 2, the following were observed in the performance of all parameter estimates for the BGpz distribution:

– As the sample size increases, the estimates of the parameters are reduced to be close to real parameter value.
– In general, when the sample size increases, the RRMSE and ARBias of the parameters estimates decrease. This indicates that the maximum likelihood estimation supplies asymptotically normally distributed and consistent estimators for all of the parameters.
– The results for the RRMSE and ARBias of the parameter estimates had small values. This indicates that the MLE were suitable estimators for the four parameters of the BGpz distribution.

Goodness of fit tests based on the EDF statistics

Assume that a random variable $x_1, x_2, ..., x_n$ where x_i the i^{th} is order statistics, had a distribution function $F(x)$ for BGpz distribution on complete sampling with sample size n and unknown parameters. Four the goodness of fit test statistics considered based on the EDF:

The modified Kolmogorov-Smirnov statistic (KS):

$$\hat{D}^{+} = \begin{matrix} \max \\ 1 \le i \le n \end{matrix} \left\{ \frac{i}{n} - F\left(x_i;\hat{\psi}\right) \right\},\ \hat{D}^{-} = \begin{matrix} \max \\ 1 \le i \le n \end{matrix} \left\{ F\left(x_i;\hat{\psi}\right) - \frac{i-1}{n} \right\},$$
$$\hat{D} = \max\left(\hat{K}^{+}, \hat{K}^{-}\right) \tag{6}$$

The modified Cramer-von Mises statistic (CvM):

$$\hat{C} = \sum_{i=1}^{n} \left[F\left(x_i;\hat{\psi}\right) - \frac{\left(2i-1\right)}{2n} + \frac{1}{12n} \right] \tag{7}$$

The modified Anderson-Darling statistic (AD):

$$\hat{A} = -n - \frac{1}{n} \sum_{i=1}^{n} \left[\left(2i-1\right) \log F\left(x_i;\hat{\psi}\right) + \left(2n+1-2i\right) \log\left\{1 - F\left(x_i;\hat{\psi}\right)\right\} \right] \tag{8}$$

The lilliefors (LF) statistic that can be found using the following steps:

– Compute the z_i-score = $(x_i - m_i)/s_i$, where m_i is the sample mean and s_i – the standard deviation of the sample.
– Calculate the proportion of the score (smaller or equal) to the z_i value. This is called the frequency associated with the z_i -score and it is denoted by Gz_i.
– Calculate the probability associated with the z_i -score. If it comes from a standard normal distribution with a mean of 0 and a standard deviation of 1, denote this probability is Nz_i.
– The LF can be calculated by using:

$$\hat{L} = \max\left\{\left|\mathrm{G}z_{\mathrm{i}} - \mathrm{N}z_{\mathrm{i}}\right|,\ \left|\mathrm{G}z_{\mathrm{i}} - \mathrm{N}z_{\mathrm{i}-1}\right|\right\} \tag{9}$$

see Abdi and Molin [19]. Note that in our case $\hat{\Psi} = (\hat{\alpha}, \hat{\lambda}, \hat{a}, \hat{b})$ was the MLE of parameter $\Psi = (\alpha, \lambda, a, b)$ of the BGpz distribution.

The critical values for test statistics

In this part, the critical values were obtained using the Monte Carlo simulation of goodness of fit test statistics for the BGpz distribution with unknown parameters based on a complete sample. The simulation repeated 1000 times using the Monte Carlo method. In tab. 3, the results of the critical values for KS, CvM, AD, and LF with sample sizes (n = 10, 30, 50, 100, and 150) and levels of significance (υ = 0.20, 0.15, 0.10, 0.05, and 0.01) are shown. Note that the random samples come from.

Steps to obtain the critical values for test statistics:

- Determine H_0: $F(x; \Psi) = F_n(x)$, where $F(x; \Psi)$ is CDF of the BGpz distribution.
- Generate a random sample $\underline{x} = (x_1, x_2,..., x_n)$ that have $F(x; \Psi)$.
- Calculate the MLE of the parameters by solving system eqs. (2)-(5).
- Obtain the order statistics from $\underline{x}$.
- Compute the test statistics from eqs. (6)-(9).
- Repeat the previous steps (from 1-5) 1000 times.
- Determine the critical values with different significance levels υ = (0.20, 0.15, 0.10 ,0.05 ,0.01) via $P = P(T \le c/H_0) = 1 - \upsilon$ where T is a test statistic and c its critical value.

Table 3. Critical values of the KS, CvM, AD and LF test statistics based on complete samples

n	Statistics	Critical values				
		υ = 0.20	υ = 0.15	υ = 0.10	υ = 0.05	υ = 0.01
10	$\hat{D}$	0.21627	0.22748	0.23853	0.25738	0.28943
	$\hat{C}$	0.07684	0.08615	0.09664	0.11302	0.15521
	$\hat{A}$	0.46862	0.51534	0.57230	0.65821	0.891651
	$\hat{L}$	0.32905	0.33985	0.35619	0.37701	0.41278
30	$\hat{D}$	0.12647	0.13338	0.14082	0.15246	0.17922
	$\hat{C}$	0.07467	0.08297	0.09561	0.11793	0.16740
	$\hat{A}$	0.46898	0.50690	0.57557	0.70766	0.93141
	$\hat{L}$	0.19391	0.20031	0.20882	0.22505	0.25887
50	$\hat{D}$	0.10111	0.10515	0.11146	0.12203	0.14260
	$\hat{C}$	0.07501	0.08479	0.10017	0.12168	0.17185
	$\hat{A}$	0.46865	0.52260	0.59085	0.71727	1.00220
	$\hat{L}$	0.15856	0.16595	0.17269	0.18658	0.20560
100	$\hat{D}$	0.07029	0.07376	0.07899	0.08535	0.10151
	$\hat{C}$	0.07529	0.08463	0.09245	0.11574	0.16952
	$\hat{A}$	0.46358	0.51108	0.56858	0.67711	0.99392
	$\hat{L}$	0.12718	0.13097	0.13598	0.14649	0.16295
150	$\hat{D}$	0.05746	0.06053	0.06424	0.07003	0.08235
	$\hat{C}$	0.07307	0.08181	0.09186	0.11308	0.15780
	$\hat{A}$	0.44770	0.49843	0.56806	0.67232	1.01533
	$\hat{L}$	0.11177	0.11619	0.12151	0.12829	0.14491

From tab. 3, the following observations were noted:
- The critical values of KS, CvM, AD and LF test statistics decrease when the sample size increases.
- As the significant level decreases, the critical values increase for all test statistics.
- In all cases, the critical values of AD test statistics had the highest values among the others.

Power study

The power of test statistics for KS, CvM, AD and LF was found with different significance levels and sample sizes based on a complete sample and several common alternative distributions by using the Monte Carlo simulation. The log-normal, exponential, gamma, chi-square with one degree of freedom, Weibull and Frechet were considered as alternative distributions. Steps to get the power of test statistics: Determine alternative hypothesis H_1: H_1: $F_1(x) = F_n(x)$, where $F_1(x)$ is a CDF of an alternative common continuous distribution (selected); Repeat steps (2-6) from section IV to get the values of test statistics for KS, CvM, AD and LF; Calculate the power of test statistics, P, for KS, CvM, AD, and LF with significance levels $\upsilon = (0.20, 0.15, 0.10, 0.05, 0.01)$ via $P = P(T \leq c/H_1)$.

Table 4. Power of KS, CvM, AD, and LF test statistics based on complete samples where the log-normal distribution is the alternative distribution

n	Statistics	Power of the test				
		$\upsilon = 0.20$	$\upsilon = 0.15$	$\upsilon = 0.10$	$\upsilon = 0.05$	$\upsilon = 0.01$
10	$\hat{D}$	0.672	0.619	0.521	0.422	0.266
	$\hat{C}$	0.712	0.662	0.607	0.510	0.357
	$\hat{A}$	0.747	0.714	0.667	0.567	0.385
	$\hat{L}$	0.337	0.277	0.202	0.125	0.050
30	$\hat{D}$	0.709	0.616	0.556	0.446	0.231
	$\hat{C}$	0.733	0.680	0.617	0.511	0.283
	$\hat{A}$	0.777	0.733	0.670	0.574	0.311
	$\hat{L}$	0.501	0.418	0.319	0.201	0.067
50	$\hat{D}$	0.654	0.594	0.529	0.415	0.268
	$\hat{C}$	0.735	0.696	0.625	0.519	0.348
	$\hat{A}$	0.791	0.752	0.682	0.579	0.383
	$\hat{L}$	0.539	0.489	0.402	0.243	0.125
100	$\hat{D}$	0.674	0.606	0.532	0.434	0.219
	$\hat{C}$	0.738	0.681	0.624	0.508	0.351
	$\hat{A}$	0.794	0.756	0.698	0.557	0.369
	$\hat{L}$	0.690	0.603	0.532	0.390	0.183
150	$\hat{D}$	0.699	0.646	0.579	0.480	0.232
	$\hat{C}$	0.733	0.697	0.638	0.536	0.343
	$\hat{A}$	0.815	0.737	0.687	0.602	0.382
	$\hat{L}$	0.775	0.718	0.632	0.528	0.233

Table 5. Power of KS, CvM, AD, and LF test statistics based on complete samples where the exponential distribution is the alternative distribution

n	Statistics	Power of the test				
		$\upsilon = 0.20$	$\upsilon = 0.15$	$\upsilon = 0.10$	$\upsilon = 0.05$	$\upsilon = 0.01$
10	$\hat{D}$	0.616	0.572	0.513	0.422	0.263
	$\hat{C}$	0.685	0.638	0.584	0.497	0.325
	$\hat{A}$	0.740	0.700	0.648	0.563	0.396
	$\hat{L}$	0.447	0.372	0.307	0.212	0.075
30	$\hat{D}$	0.649	0.579	0.520	0.399	0.249
	$\hat{C}$	0.707	0.664	0.593	0.478	0.342
	$\hat{A}$	0.771	0.717	0.663	0.558	0.375
	$\hat{L}$	0.730	0.661	0.536	0.407	0.250
50	$\hat{D}$	0.642	0.596	0.516	0.422	0.249
	$\hat{C}$	0.726	0.673	0.581	0.491	0.380
	$\hat{A}$	0.790	0.736	0.656	0.546	0.436
	$\hat{L}$	0.853	0.786	0.734	0.613	0.355
100	$\hat{D}$	0.666	0.600	0.517	0.427	0.260
	$\hat{C}$	0.754	0.695	0.615	0.525	0.351
	$\hat{A}$	0.792	0.735	0.676	0.577	0.400
	$\hat{L}$	0.981	0.960	0.938	0.875	0.663
150	$\hat{D}$	0.680	0.614	0.547	0.416	0.190
	$\hat{C}$	0.735	0.694	0.620	0.512	0.314
	$\hat{A}$	0.803	0.745	0.657	0.542	0.366
	$\hat{L}$	0.996	0.996	0.987	0.968	0.860

Table 6. Power of KS, CvM, AD and LF test statistics based on complete samples where the chi-square distribution is the alternative distribution

n	Statistics	Power of the test				
		$\upsilon = 0.20$	$\upsilon = 0.15$	$\upsilon =0.10$	$\upsilon = 0.05$	$\upsilon = 0.01$
10	$\hat{D}$	0.715	0.646	0.552	0.468	0.296
	$\hat{C}$	0.741	0.679	0.630	0.513	0.342
	$\hat{A}$	0.789	0.740	0.683	0.575	0.366
	$\hat{L}$	0.979	0.969	0.651	0.893	0.764
30	$\hat{D}$	0.680	0.632	0.558	0.450	0.213
	$\hat{C}$	0.704	0.652	0.595	0.498	0.320
	$\hat{A}$	0.772	0.730	0.655	0.553	0.315
	$\hat{L}$	1.000	1.000	1.000	1.000	0.999
50	$\hat{D}$	0.687	0.635	0.558	0.414	0.213
	$\hat{C}$	0.746	0.690	0.631	0.538	0.302
	$\hat{A}$	0.797	0.745	0.674	0.579	0.325
	$\hat{L}$	1.000	1.000	1.000	1.000	1.000
100	$\hat{D}$	0.692	0.619	0.547	0.407	0.180
	$\hat{C}$	0.743	0.679	0.600	0.519	0.305
	$\hat{A}$	0.788	0.738	0.660	0.565	0.360
	$\hat{L}$	1.000	1.000	1.000	1.000	1.000
150	$\hat{D}$	0.709	0.674	0.595	0.489	0.283
	$\hat{C}$	0.772	0.727	0.668	0.584	0.374
	$\hat{A}$	0.836	0.792	0.734	0.637	0.404
	$\hat{L}$	1.000	1.000	1.000	1.000	1.000

Table 7. Power of KS, CvM, AD, and LF test statistics based on complete samples where the gamma distribution is the alternative distribution

n	Statistics	Power of the test				
		$\upsilon = 0.20$	$\upsilon = 0.15$	$\upsilon = 0.10$	$\upsilon = 0.05$	$\upsilon = 0.01$
10	$\hat{D}$	0.688	0.634	0.568	0.465	0.297
	$\hat{C}$	0.743	0.699	0.646	0.564	0.422
	$\hat{A}$	0.810	0.749	0.701	0.602	0.439
	$\hat{L}$	0.226	0.192	0.130	0.077	0.028
30	$\hat{D}$	0.676	0.621	0.551	0.423	0.295
	$\hat{C}$	0.753	0.703	0.634	0.539	0.350
	$\hat{A}$	0.801	0.747	0.668	0.572	0.375
	$\hat{L}$	0.262	0.185	0.120	0.062	0.020
50	$\hat{D}$	0.679	0.613	0.548	0.455	0.259
	$\hat{C}$	0.736	0.692	0.637	0.545	0.398
	$\hat{A}$	0.786	0.750	0.693	0.613	0.448
	$\hat{L}$	0.263	0.207	0.162	0.107	0.031
100	$\hat{D}$	0.669	0.613	0.557	0.404	0.264
	$\hat{C}$	0.757	0.706	0.643	0.489	0.345
	$\hat{A}$	0.826	0.786	0.706	0.562	0.398
	$\hat{L}$	0.300	0.229	0.174	0.114	0.022
150	$\hat{D}$	0.659	0.612	0.527	0.412	0.256
	$\hat{C}$	0.734	0.691	0.614	0.526	0.373
	$\hat{A}$	0.791	0.756	0.666	0.569	0.408
	$\hat{L}$	0.325	0.253	0.178	0.121	0.028

Table 8. Power of KS, CvM, AD, and LF test statistics based on complete samples where the Weibull distribution is the alternative distribution

n	Statistics	Power of the test				
		$\upsilon = 0.20$	$\upsilon = 0.15$	$\upsilon = 0.10$	$\upsilon = 0.05$	$\upsilon = 0.01$
10	$\hat{D}$	0.698	0.639	0.558	0.459	0.249
	$\hat{C}$	0.737	0.707	0.649	0.558	0.350
	$\hat{A}$	0.791	0.749	0.709	0.606	0.379
	$\hat{L}$	0.183	0.141	0.095	0.053	0.009
30	$\hat{D}$	0.654	0.594	0.523	0.421	0.224
	$\hat{C}$	0.729	0.681	0.616	0.509	0.319
	$\hat{A}$	0.791	0.728	0.679	0.569	0.396
	$\hat{L}$	0.171	0.122	0.075	0.032	0.007
50	$\hat{D}$	0.674	0.620	0.550	0.441	0.236
	$\hat{C}$	0.739	0.692	0.626	0.524	0.330
	$\hat{A}$	0.801	0.739	0.672	0.603	0.387
	$\hat{L}$	0.152	0.096	0.064	0.027	0.003
100	$\hat{D}$	0.721	0.651	0.594	0.482	0.275
	$\hat{C}$	0.780	0.723	0.658	0.569	0.383
	$\hat{A}$	0.834	0.788	0.721	0.592	0.385
	$\hat{L}$	0.127	0.096	0.065	0.033	0.006
150	$\hat{D}$	0.680	0.623	0.534	0.435	0.203
	$\hat{C}$	0.748	0.700	0.603	0.513	0.303
	$\hat{A}$	0.787	0.750	0.681	0.578	0.341
	$\hat{L}$	0.158	0.110	0.066	0.034	0.011

Table 9. Power of KS, CvM, AD and LF test statistics based on complete samples where the Frechet distribution is the alternative distribution

n	Statistics	Power of the test				
		$\upsilon = 0.20$	$\upsilon = 0.15$	$\upsilon = 0.10$	$\upsilon = 0.05$	$\upsilon = 0.01$
10	$\hat{D}$	0.650	0.582	0.507	0.398	0.217
	$\hat{C}$	0.722	0.654	0.576	0.463	0.275
	$\hat{A}$	0.781	0.731	0.642	0.534	0.338
	$\hat{L}$	0.497	0.449	0.367	0.245	0.108
30	$\hat{D}$	0.641	0.592	0.527	0.406	0.277
	$\hat{C}$	0.719	0.654	0.597	0.537	0.378
	$\hat{A}$	0.781	0.734	0.675	0.593	0.458
	$\hat{L}$	0.791	0.729	0.660	0.585	0.416
50	$\hat{D}$	0.677	0.627	0.543	0.436	0.278
	$\hat{C}$	0.706	0.653	0.604	0.511	0.346
	$\hat{A}$	0.782	0.732	0.665	0.543	0.338
	$\hat{L}$	0.897	0.870	0.823	0.732	0.567
100	$\hat{D}$	0.630	0.581	0.511	0.401	0.260
	$\hat{C}$	0.705	0.655	0.586	0.494	0.323
	$\hat{A}$	0.773	0.718	0.662	0.535	0.348
	$\hat{L}$	0.987	0.979	0.972	0.942	0.839
150	$\hat{D}$	0.700	0.635	0.569	0.562	0.235
	$\hat{C}$	0.760	0.695	0.636	0.535	0.372
	$\hat{A}$	0.809	0.773	0.702	0.594	0.434
	$\hat{L}$	0.998	0.996	0.994	0.988	0.957

Overall observations on tabs. 4-9:

- From tabs. 4, 7, and 8, the power of test statistics appeared as the following: the power of $\hat{A}$ > power of $\hat{C}$ > power of $\hat{D}$ > power of $\hat{L}$, when the log-normal, gamma and Weibull distributions were the alternative distributions.
- From tabs. 5 and 9, the exponential and Frechet distributions were the alternative distributions. Also, the AD test statistic was the most powerful in the whole when $n \le 30$ elsewhere the LF test statistic was the most powerful. Moreover, the power of CvM test statistics is greater than the power of KS test statistics in general.
- When the chi-square with one degree of freedom distribution was the alternative distribution as in tab. 6, the LF method had the greatest power of test statistics. From these observations it was found that: the power of $\hat{A}$ > power of $\hat{C}$ > power of $\hat{D}$.
- The relationship between the power of the test statistic and the significant level had a positive correlation at fixed n.
- The AD test statistics had the highest power in most of the cases among the different alternative distributions.
- The power of test statistic increased when the sample size n increased.
- Among the results, chi-square with one degree of freedom distribution was the best appropriate alternative distribution for the BGpz distribution. Also, the gamma and Weibull distributions were good alternative distributions with approximate similar behavior. Whereas, the log-normal, Frechet and exponential distributions were the worst alternative distributions for the BGpz distribution.

Real data analysis

In this section, various real data sets were employed to reveal the importance of using the BGpz distribution as a good lifetime model through comparing it by several known distributions such as the exponentiated Gompertz (EGpz), Gompertz (Gpz), and exponential (E) distributions. We obtained the MLE of unknown parameters for 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 to verify the MLE of unknown parameters. For more details about the model selection technique, see Whittaker and Furlow [20]. These models are defined:

$$\begin{cases} AIC = -2l\left(\hat{\theta}\right) + 2k \\ BIC = -2l\left(\hat{\theta}\right) + k\log\left(n\right) \\ CAIC = -2l\left(\hat{\theta}\right) + \dfrac{2kn}{n-k-1} \\ HQIC = -2l\left(\hat{\theta}\right) + 2k\log\left[\log\left(n\right)\right] \end{cases}$$

where $l(\hat{\theta})$ is the log-likelihood function evaluated at the maximum likelihood estimate in equation($\hat{\theta}$), k – the number of parameters, and n – the sample size. Also, the KS, CvM, AD, and LF test statistics were applied to compare between fit models for data. In general, the smallest values of these test statistics indicated in the best fit model for the data. All computations presented to analyze the data were carried out by MATHEMATICA 11.3.

The ten real lifetime data sets are provided in this section. The first, second and third data sets were presented in Abu-Zinadah and Hajar [12, 13]. Moreover, the fourth and sixth data sets have been taken from Badr [17], the fifth data set from Best *et al.* [21], the rest of the data sets from Shanker *et al.* [22] and the tenth data from Cakmakyapan and Ozal [23].

From tabs. 10 and 11, the following observations can be made:

– The MLE of unknown parameters for distributions were good estimates based on the AIC, BIC, HQIC, and CAIC measures.
– According to the KS, CvM, AD, and LF test statistics, the BGpz distribution was the best fit model most of the time, therefore, it can be used to analyze the lifetime of the data.

Conclusion

In this article, the maximum likelihood estimators of the four parameters for BGpz distribution were derived. Also, the critical values of Kolmogorov-Smirnov, Anderson-Darling, Cramer-von Mises, and Lilliefors test statistics for BGpz distribution were found at different significant levels and sample sizes. The power study of these test statistics for common alternative distribution such as: log-normal, exponential, gamma, chi-square with one degree of freedom, Weibull and Frechet, showed that the chi-square with one degree of freedom distribution was the best alternative for BGpz distribution. The application on nine real lifetime data sets presented the BGpz distribution as a good model to fit these data sets.

Table 10. The MLE of unknown parameters and the value of AIC, BIC, HQIC, and CAIC for different models

Data	Model	MLE				*AIC*	*BIC*	*HQIC*	*CAIC*
		$\hat{\alpha}$	$\hat{\lambda}$	$\hat{a}$	$\hat{b}$				
Data 1	BGpz	0.0116275	0.906177	0.72254	1.08773	474.647	482.295	477.559	475.536
	EGpz	0.0095522	1.3993	0.733429	–	474.742	480.478	476.927	475.264
	Gpz	0.0157953	0.755399	–	–	475.228	479.052	476.684	475.484
	E	–	0.0218885	–	–	484.179	486.091	484.907	484.263
Data 2	BGpz	0.0936911	3.5409	1.39831	1.64857	254.213	263.536	257.939	254.776
	EGpz	0.0918918	5.84745	1.43599	–	251.926	258.918	254.72	252.259
	Gpz	0.121567	3.38531	–	–	254.749	259.41	256.612	254.913
	E	-	0.510402	–	–	256.229	258.559	257.16	256.283
Data 3	BGpz	0.23718	1.77608	2.49739	1.06648	284.914	294.823	288.906	285.396
	EGpz	0.223399	2.04744	2.53055	–	283.787	291.219	286.782	284.073
	Gpz	0.0771897	4.72414	–	–	332.854	337.808	334.85	332.995
	E	–	0.389287	–	–	344.045	346.552	345.043	344.092
Data 4	BGpz	0.109753	4.69793	2.61098	2.01982	84.3877	89.9925	86.1808	85.9877
	EGpz	0.100208	9.56343	2.90999	–	82.2034	86.407	83.5482	83.1265
	Gpz	0.491394	0.617683	–	–	86.1523	88.9547	87.0488	86.5968
	E	–	0.597015	–	–	92.9488	94.35	93.397	93.0916
Data 5	BGpz	0.170084	4.03842	10.0583	0.439703	114.632	120.237	116.425	116.232
	EGpz	0.149224	2.72937	6.59323	–	114.752	118.955	116.097	115.675
	Gpz	0.0718961	3.01517	–	–	142.916	145.718	143.813	143.361
	E	–	0.239808	–	–	147.675	149.076	148.818	147.818
Data 6	BGpz	0.0767047	4.63528	1.02907	1.9683	125.45	132.765	128.19	126.426
	EGpz	1.0233	0.0847201	8.104	–	123.452	128.938	125.508	124.024
	Gpz	0l.122764	5.21572	–	–	121.525	125.182	122.895	121.804
	E	–	0.754903	–	–	119.867	121.696	120.552	119.958
Data 7	BGpz	0.231276	4.61693	5.42718	1.30997	81.9838	90.5563	85.3554	82.6734
	EGpz	2.02973	0.0715479	2.79666	–	36.1675	42.5969	38.6962	36.5743
	Gpz	0.0599799	10.8872	–	–	176.342	180.628	178.028	176.542
	E	–	0.663647	–	–	179.661	181.504	180.504	179.726
Data 8	BGpz	0.313472	0.793069	0.631881	0.79228	127.509	133.614	129.591	128.888
	EGpz	0.327902	0.489459	0.601237	–	125.455	130.034	127.016	126.255
	Gpz	0.0853083	5.08281	–	–	116.034	119.087	117.075	116.421
	E	-	0.532081	–	–	112.905	114.432	113.426	113.03
Data 9	BGpz	0.313297	8.48359	55.2793	0.315608	40.6701	44.653	41.4476	43.3368
	EGpz	0.364137	2.28184	5.95345	–	43.3132	46.3004	43.8963	44.8132
	Gpz	0.0974588	5.1427	–	–	66.6454	68.6369	67.0342	67.3513
	E	–	0.526316	–	–	67.6742	68.6699	67.8685	67.8964
Data 1D	BGpz	0.320427	1.18685	3.16675	1.16716	294.4	304.821	298.617	294.821
	EGpz	0.334333	1.23797	3.15175	–	292.403	300.218	295.566	292.653
	Gpz	0.0809659	4.51163	–	–	376.648	381.858	378.757	376.772
	E	–	0.388676	–	–	391.002	393.607	392.056	391.043

Table 11. The KS,CvM, AD and LF test statistics for different models

Data	n	Model	Test statistics			
			$\hat{D}$	$\hat{C}$	$\hat{A}$	$\hat{L}$
Data 1	50	BGpz	0.17894	0.41717	2.63940	0.16250
		EGpz	0.19842	0.51741	3.07940	0.16631
		Gpz	0.16417	0.44498	4.21039	0.17490
		E	0.19107	0.51886	3.65008	0.17570
Data 2	76	BGpz	0.12265	0.15560	0.88609	0.08512
		EGpz	0.12036	0.14776	0.84731	0.08575
		Gpz	0.12677	0.32172	1.80402	0.06979
		E	0.16633	0.57081	2.98811	0.07498
Data 3	88	BGpz	0.23712	0.18338	1.42541	0.27986
		EGpz	0.23455	0.20285	1.52585	0.28118
		Gpz	0.30027	2.38129	11.6804	0.28355
		E	0.30527	2.42237	12.1391	0.28324
Data 4	30	BGpz	0.05707	0.01588	0.12021	0.13067
		EGpz	0.05352	0.01440	0.10980	0.13089
		Gpz	0.11493	0.08359	0.64404	0.11581
		E	0.23520	0.45393	2.51401	0.09794
Data 5	35	BGpz	0.16761	0.13652	0.87297	0.13852
		EGpz	0.16740	0.15208	1.02475	0.31727
		Gpz	0.39993	1.24744	5.94506	0.14260
		E	0.40427	1.27653	6.20516	0.14605
Data 6	46	BGpz	0.12030	0.10206	0.63280	0.16283
		EGpz	0.12154	0.12154	0.64296	0.16346
		Gpz	0.12936	0.12936	0.71995	0.16720
		E	0.09325	0.06387	0.47098	0.15122
Data 7	63	BGpz	0.24696	1.25474	6.52174	0.18003
		EGpz	0.15531	0.23397	1.27456	0.09520
		Gpz	0.42569	3.95740	18.77320	0.20826
		E	0.41800	3.86222	18.42580	0.20918
Data 8	34	BGpz	0.15856	0.17714	1.11464	0.16565
		EGpz	0.17595	0.24846	1.35551	0.17134
		Gpz	0.13917	0.12174	0.65285	0.15836
		E	0.08896	0.04051	0.27196	0.15402
Data 9	20	BGpz	0.15962	0.07919	0.42562	0.16442
		EGpz	0.18558	0.13736	0.78091	0.20228
		Gpz	0.44120	0.97509	4.58052	0.19943
		E	0.43951	0.96296	4.60349	0.19848
Data 10	100	BGpz	0.06188	0.06798	0.40042	0.09461
		EGpz	0.06184	0.06820	0.40124	0.09462
		Gpz	0.30749	3.20718	15.9195	0.09878
		E	0.30677	3.21030	16.2985	0.10075

References

[1] Abu-Zinadah, H. H., Goodness-of-Fit Tests for the Exponentiated Gompertz Distribution, *International Journal of Innovation in Science and Mathematics*, *2* (2014), Jan., pp. 327-380
[2] Rasheed, H. A., *et al*., Comparing Some Estimators of the Parameter of Basic Gompertz Distribution, *Journal of Iraqi Al-Khwarizmi Society*, *2* (2018), Mar., pp. 178-187
[3] Soliman, A. A., *et al*., A Simulation Based Approach to the Study of Coefficient of Variation of Gompertz Distribution under Progressive First-Failure Censoring, *Indian Journal of Pure and Applied Mathematics*, *42* (2011), Oct., pp. 335-356
[4] Soliman, A. A., *et al*., Estimation of the Parameters of Life for Gompertz Distribution Using Progressive First-Failure Censoring Data, *Computational Statistics and Data Analysis*, *56* (2012), 8, pp. 2471-2485
[5] El-Gohary, *et al*., The Generalized Gompertz Distribution, *Applied Mathematical Modelling*, *37* (2013), 1-2, pp. 13-24
[6] Abu-Zinadah, H. H., Aloufi, A. S., Some Characterizations of the Exponentiated Gompertz Distribution, *International Mathematical Forum*, *9* (2014), Jan., pp. 1427-1439
[7] Abu-Zinadah, H. H., Aloufi, A. S., On Estimations of Parameters of Generalized Gompertz Distribution, *International Journal of Innovation in Science and Mathematics*, *3* (2015), Jan., pp 156-162
[8] Abu-Zinadah, H. H., Aloufi, A. S., Different Method Estimations of the Three Parameters of Exponentiated Gompertz Distribution, *Applied Mathematics and Information Sciences*. *10* (2016), Mar., pp. 705-710
[9] Jafari, A. A., *et al*., The Beta Gompertz Distribution, *Revista Colombiana de Estadistica*, *37* (2014), 1, pp. 139-156
[10] Benkhelifa, L., The Beta Generalized Gompertz Distribution, *Applied Mathematical Modelling, 52* (2017), Dec., pp. 341-357
[11] Hajar, N. M., Abu-Zinadah, H. H., Some Statistical Properties of the Beta Exponentiated Gompertz Distribution, *IOSR Journal of Mathematics*, *14* (2018), Feb., pp. 31-40
[12] Abu-Zinadah, H. H., Hajar, N. M., Classical and Bayesian Estimation of Parameters for The Beta Exponentiated Gompertz Distribution with Real Data, *Journal of Advanced Research in Dynamical and Control Syatem*, *11* (2019), Jan., pp. 446-453
[13] Abu-Zinadah, H. H., Hajar, N. M., Statistical Inference for Parameters of the Beta Exponentiated Gompertz Distribution According to Type-II Censoring, *Journal of Computational and Theoretical Nanoscience, 16* (2019), 7, pp. 3128-3135
[14] Bakoban, R. A., Abu-Zinadah, H. H., The Beta Generliazed Inverrted Exponential Distribution with Real Data Applications, *REVSTAT – Statistical Journal*, *15* (2017), 1, pp. 65-88
[15] Shawky, A. I., Bakoban, R. A., Modified Goodness-of-Fit Tests for Exponentiated Gamma Distribution with Unknown Shape Parameter, *InterStat*, *3* (2009), Jan., pp. 1-17
[16] Lenart, A., Missov, T. I., Goodness-of-Fit Tests for the Gompertz Distribution, *Taylor and Francis Group*, *45* (2016), 10, pp. 2920-2937
[17] Badr, M. M., Goodness-of-Fit Tests for the Compound Rayleigh Distribution with Application Real Data, *Heliyon*, *5* (2019), 8, pp. 1-6
[18] Jeffrey A., Dai, H., *Handbook of Mathematical Formulas and Integral*s, 4th ed., Academic Press, Cambridge, Mass., USA, 2008
[19] Abdi, H., Molin, P., Lilliefors/Van Soest's Test of Normality, *Encyclopedia of Measurement and Statistics*, (2007), Jan., pp. 540-544
[20] Whittaker, T. A., Furlow, C. F., The Comparison of Model Selection Criteria when Selection among Competing Hierarchical Linear Models, *Journal of Modern Applied Statistical Method*, *8* (2009), May, pp. 137-193
[21] Best, D. J., *et al*., Easily Applied Tests of Fit for the Rayleigh Distribution, *Synkhaya*, *72* (2010), Nov., pp. 254-263
[22] Shanker, R., *et al*., On Modelling of Lifetimes Data Using Exponential and Lindley Distributions, *Biometrics and Biostatistics International Journal*, *2* (2015), 5, pp. 140-147
[23] Cakmakyapan, S., Ozal, G., Lindley-Rayleigh Distribution with Application Lifetime Data, *Journal of Reliability and Statistical Studies*, *11* (2018), Jan., pp. 9-24

Paper submitted: May 1, 2020
Paper revised: May 28, 2020
Paper accepted: June 1, 2020