Survival Function Estimation for Fuzzy Gompertz Distribution with neutrosophic data

Alan Adham Bibani*[1], Zakariya Yahya Algamal[2]

[1]Department of Statistics and Informatics, University of Mosul, Mosul, Iraq

[2]Department of Statistics and Informatics, University of Mosul, Mosul, Iraq

Emails: alanadham0@gmail.com; zakariya.algamal@uomosul.edu.iq

Abstract

A relatively recent area of research known as neutrosophic statistics deals with data that are ambiguous, indeterminate, and inconsistent. By embracing the idea of neutrosophy, which denotes the existence of three components in a statement: truth, falsity, and indeterminacy, it broadens the application of classical statistics. One of the significant offshoots of statistics is life-time data analysis. Traditional statistical methods only account for variation within the data and calculate lifetime observations as accurate numbers. Actually, there are two different kinds of uncertainty in data: fluctuation between observations and fuzziness. As a result, analysis techniques that solely employ precise lifetime data and ignore fuzziness use incomplete information and produce false results. This paper sought to generalize hazard rates, survival functions, and parameter estimates for fuzzy Gompertz Distribution. Simulation studies are implemented to examine the performance of the fuzzy Gompertz Distribution. The results show that the fuzzy Gompertz Distribution has better flexibility in handling over the standard Gompertz Distribution.

Keywords: Gompertz Distribution; fuzzy numbers, neutrosophic statistics; survival analysis; hazard function.

1. Introduction

The problem of analyzing time to event data arises in a number of applied fields, such as medicine, biology, public health, and epidemiology [1, 2]. Analysis of data, estimate of probability distributions, and stochastic models are all topics covered in statistics. As a result, statistics requires a quantitative explanation of the data. Data are typically thought of as being numbers, vectors, or conventional functions in normal statistics. However, real data used in applications is typically more or less imprecise, sometimes known as fuzzy, rather than precise numbers or vectors [3-8].

The analysis of survival distributions is a component of survival analysis. By this, we mean the lifespans of individuals, cancer patients, industrial robots, parts, cogs, and software.

A set of data or a portion of it that is somewhat ambiguous is referred to as neutrosophic statistics, as are the techniques used to analyze the data [9]. The difference between neutrosophic statistics and classical statistics is that all data are determined in classical statistics. Neutosophic statistics often overlaps with classical statistics when indeterminacy is zero. The neutrosophic metric can be used to quantify the uncertain data.

We can interpret and arrange neutrosophic data, which may contain certain ambiguities, using neutrosophic statistical approaches in order to identify underlying patterns [10].

DOI: https://doi.org/10.54216/IJNS.210313
Received: February 18, 2023 Revised: May 21, 2023 Accepted: June 24, 2023

2. Neutrosophic statistics

The concept of neutrosophic probability as a function $NP :\to [0,\ 1]^3$ was originally presented by [11], where U is a neutrosophic sample space and defined the probability mapping to take the form $NP(S) = (ch(S),\ ch(neut\ S),\ ch(anti\ S)) = (\eta,\ \beta,\ \tau)$ $where\ 0 \le \eta,\ \beta,\ \tau \le 1\ and\ 0 \le \eta + \beta + \tau \le 3$. Additionally, a variety of neutrosophic probability models, including Poisson, binomial, exponential, uniform, normal, Weibull, Kumaraswamy, extended Pareto, Maxwell, Lognormal, and Gamma, have been explored by numerous academics [10, 12-17]. The term Ψ represents the set of sample space, R represents the set of real numbers, and Υ denotes a sample space event, X_N and Y_N denote neutrosophic r.v. Furthermore, we demonstrate certain renowned definitions and characteristics of neutrosophic probability and logic that will be important in creating this neutrosophic probability model.

Definition 1 Consider the real-valued crisp r.v. X, which has the following definition: $X : \Psi \to R$ where Ψ is the event space and X_N neutrosophic r.v. as follows:

$X_N : \Psi \to R(I)$

and

$X_N = X + I$

The term I represents indeterminacy.

Theorem1: Let the neutrosophic r.v. $X_N = X + I$ and the CDF of X is $F_X(x) = P(X \le x)$ [14]. The following assertions are correct:

$FX_N(x) = F_X(x - I)$,

$f_{X_N}(x) = f_X(x - I)$,

where F_{X_N} and f_{X_N} are the CDF and PDF of a neutrosophic r.v. X_N, respectively.

Theorem 2: Let $X_N = X + I$, is the neutrosophic r.v., then the expected value and variance can be derived as follows: $E(X_N) = E(X) + I$ and $V(X_N) = V(X)$ [14].

3. Fuzzy Gompertz Distribution

One frequently uses the exponential, Gompertz, generalized exponential distributions while examining lifetime data. It is well known that the hazard function for an exponential distribution can only be constant, whereas hazard functions for a Gompertz distribution and a generalized exponential distribution can only be monotonic (growing for a Gompertz distribution and increasing or decreasing for a generalized exponential distribution) [18]. These distributions are widely used in reliability and medical research to simulate lifetime data.

Gompertz distribution is considered one of the most popular distribution when analyzing survival data [19, 20]. We know that the form of Gompertz distribution is:

$$f(t;\lambda,\theta) = \lambda e^{\theta t} \exp\left(\frac{\lambda}{\theta}\left(1-e^{\theta t}\right)\right) \text{for } t \ge 0, \quad (1)$$

where $\lambda,\theta > 0$ is the scale and shape parameters. The CDF is

$$F(t;\lambda,\theta) = 1-\exp\left(\frac{\lambda}{\theta}\left(1-e^{\theta t}\right)\right) \quad (2)$$

The survival function and hazard function are defined as

$$S(t) = \exp\left\{\frac{\lambda}{\theta}\left(1-e^{\theta t}\right)\right\} \quad (3)$$

$$h(t;\lambda,\theta) = \lambda e^{\theta t} \quad (4)$$

Depending on the maximum likelihood method, the distribution parameters can be estimated efficiently.

Regarding the fuzzy dataset, some definitions are needed.

DOI: https://doi.org/10.54216/IJNS.210313
Received: February 18, 2023 Revised: May 21, 2023 Accepted: June 24, 2023

Definition 2 [21]:

A membership function mapping the elements of nonempty set X to the interval $[0, 1]$ defines a fuzzy set $\tilde{A}$. $\mu_{\tilde{A}}: X \to [0,1]$, is called the degree of membership function of the fuzzy set $\tilde{A}$. $\mu_{\tilde{A}}(x)(x)$ is called the membership value of $x \in X$ in the fuzzy set $\tilde{A}$

Definition 3 [22]:

A fuzzy number $\tilde{A}$ is a fuzzy set on real line $\mathbb{R}$, if satisfy the following conditions:

1- $\mu_{\tilde{A}}(z)$ is piecewise continuous, where $z \in \mathbb{R}$.

2- $\exists z \in \mathbb{R}$ such that $\mu_{\tilde{A}}(z) = 1$ ($i.e.$ $\tilde{A}$ is said to be normal)

3- $\tilde{A}$ is convex $(i.e.\ \mu_{\tilde{A}}(\alpha_{x_1} + (1-\alpha)_{x_2}) \geq \min\{\mu_{\tilde{A}}(x_1), \mu_{\tilde{A}}(x_2)\}$, for all $x_1, x_2 \in \mathbb{R}$, and all $\alpha \in [0, 1]$, [5]).

Definition 4 [23]:

The $\alpha-cut$ set of a fuzzy set $\tilde{A}$ is crisp set defined by $A_\alpha = \{x \in X \mid \mu_{\tilde{A}}(x) \geq \alpha\}$.

A fuzzy set is convex $\leftrightarrow$ all its $\alpha-$ $cuts$ are convex.

Regarding to this, the fuzzy Gompertz distribution is defined as:

$$\begin{aligned}\tilde{F}\left(\tilde{t}_{A(\alpha)}\right) &= \int_0^{\tilde{t}_{A(\alpha)}} f(u)du \\ &= \int_0^{\tilde{t}_{A(\alpha)}} \lambda e^{\theta t} \exp\left(\frac{\lambda}{\theta}\left(1-e^{\theta t}\right)\right)du \qquad (5) \\ &= 1-\exp\left(\frac{\lambda}{\theta}\left(1-e^{\theta \tilde{t}_{A(\alpha)}}\right)\right) = \tilde{F}\left(\tilde{t}_{A(\alpha)}\right)\end{aligned}$$

and

$$\tilde{f}\left(\tilde{t}_{A(\alpha)}\right) = \lambda e^{\theta \tilde{t}_{A(\alpha)}} \exp\left(\frac{\lambda}{\theta}\left(1-e^{\theta \tilde{t}_{A(\alpha)}}\right)\right) \qquad (6)$$

The fuzzy survival and fuzzy hazard functions are, respectively

$$\begin{aligned}\tilde{S}\left(\tilde{t}_{A(\alpha)}\right) &= 1-\tilde{F}\left(\tilde{t}_{A(\alpha)}\right) \\ &= \exp\left(\frac{\lambda}{\theta}\left(1-e^{\theta \tilde{t}_{A(\alpha)}}\right)\right) \qquad (7)\end{aligned}$$

$$\tilde{h}\left(\tilde{t}_{A(\alpha)}\right) = \frac{\lambda e^{\theta \tilde{t}_{A(\alpha)}} \exp\left(\frac{\lambda}{\theta}\left(1-e^{\theta \tilde{t}_{A(\alpha)}}\right)\right)}{\exp\left(\frac{\lambda}{\theta}\left(1-e^{\theta \tilde{t}_{A(\alpha)}}\right)\right)} \qquad (8)$$

4. Simulation Results

In this section, we conduct simulation to assess how well the proposed approach. The Akaike information criterion (AIC), Bayesian information criteria (BIC), and the mean time to failure (MTTF) are assessed. The sample size, n, in the simulation is considered: 30, 50, 150, respectively. We considered the initial values of the Gompertz distribution parameters, (λ, θ), as: (1,0.5) and (1,5). Further, the value of the cut for the fuzzy set is: 0.1, 0.5, and 0.7. The simulation is repeated 1000 times for each sample size. We reported all results in Tables 1–6.
Based on the obtained results Tables 1–6, several observations can be obtained as follows:

1- Generally, the behavior of the fuzzy Gompertz distribution is better than that of Gompertz distribution, regarding the survival and hazard functions.

DOI: https://doi.org/10.54216/IJNS.210313
Received: February 18, 2023 Revised: May 21, 2023 Accepted: June 24, 2023

2- Clearly, in terms of AIC and BIC, fuzzy Gompertz distribution improved the performance of the survival and hazard functions compared to Gompertz distribution in all the cases of simulation by getting the small values of AIC and BIC.

3- In terms of MTTF, fuzzy Gompertz distribution again improved the performance of the survival and hazard functions compared to Gompertz distribution in all the cases of simulation in yielding the high MTTF.

4- In terms of the membership (mem) values, there is increasing in the AIC and BIC values regardless the value of n with the priority for fuzzy Gompertz distribution.

5- With respect to the value of n , the AIC and BIC values decrease when n increases, regardless the value of the membership degree (mem).

Table 1: The results when $(\lambda = 1,\ \theta = 0.5),\ mem = 0.1$

Method	n=30		n=50		n=150	
	Real Time	Fuzzy Time	Real Time	Fuzzy Time	Real Time	Fuzzy Time
AIC	144.9961	140.7697	143.9861	138.7457	142.9861	136.7354
BIC	142.3173	140.1029	142.2482	138.0369	141.0068	135.3619
MTTF	3.97221	3.985389	3.98221	3.985389	3.9864	4.005389

Table 2: The results when $(\lambda = 1,\ \theta = 0.5),\ mem = 0.5$

Method	n=30		n=50		n=150	
	Real Time	Fuzzy Time	Real Time	Fuzzy Time	Real Time	Fuzzy Time
AIC	143.1633	197.7739	143.6742	197.7739	144.7728	197.7739
BIC	141.7621	196.6384	141.7621	196.6384	141.7621	196.6384
MTTF	2.73821	2.810379	2.73821	2.810379	2.73821	2.810379

Table 3: The results when $(\lambda = 1,\ \theta = 0.5),\ mem = 0.7$

Method	n=30		n=50		n=150	
	Real Time	Fuzzy Time	Real Time	Fuzzy Time	Real Time	Fuzzy Time
AIC	146.4941	142.2677	145.4941	140.2437	144.4941	138.2634
BIC	143.8953	141.5999	143.7462	139.5349	142.5048	136.8599
MTTF	5.47021	5.483389	5.48021	5.483389	5.4844	5.503389

Table 4: The results when $(\lambda = 1,\ \theta = 5),\ mem = 0.1$

Method	n=30		n=50		n=150	
	Real Time	Fuzzy Time	Real Time	Fuzzy Time	Real Time	Fuzzy Time
AIC	583.9849	515.6699	582.9849	514.5599	647.38	464.4243
BIC	583.3861	513.6535	580.8969	511.6755	642.3907	463.5598
MTTF	4.561402	4.576941	4.573402	4.596517	4.868814	5.08339

Table 5: The results when $(\lambda = 1,\ \theta = 5),\ mem = 0.5$

Method	n=30		n=50		n=150	
	Real Time	Fuzzy Time	Real Time	Fuzzy Time	Real Time	Fuzzy Time
AIC	583.4559	630.2176	647.85103	68.99937	647.85103	68.99937
BIC	584.8571	633.0298	649.76305	69.77195	650.86167	69.77195
MTTF	5.032402	5.089107	5.339814	5.083757	5.339814	5.083757

DOI: https://doi.org/10.54216/IJNS.210313
Received: February 18, 2023 Revised: May 21, 2023 Accepted: June 24, 2023

Table 6: The results when $(\lambda = 1,\ \theta = 5),\ mem = 0.7$

Method	n=30		n=50		n=150	
	Real Time	Fuzzy Time	Real Time	Fuzzy Time	Real Time	Fuzzy Time
AIC	583.931	528.9816	648.32613	30.59019	648.32613	30.59019
BIC	585.3322	531.351	650.23815	30.89278	651.33677	30.89278
MTTF	5.507502	5.504525	5.814914	5.349018	5.814914	5.349018

5. Conclusion

Based on the results shown, the proposed approach, fuzzy Gompertz distribution, yields better results for a very important estimation for both survival and hazard functions. Due to the power and accuracy of this proposed approach, it will be possible to extend the application to other neutrosophic data.

References

[1] M. Cockeran, S. G. Meintanis, and J. S. Allison, "Goodness-of-fit tests in the Cox proportional hazards model," *Communications in Statistics - Simulation and Computation*, pp. 1-12, 2019.

[2] T. Emura, Y. H. Chen, and H. Y. Chen, "Survival prediction based on compound covariate under Cox proportional hazard models," *PLoS One*, vol. 7, p. e47627, 2012.

[3] A. Bouchet, M. Sesma-Sara, G. Ochoa, H. Bustince, S. Montes, and I. Díaz, "Measures of embedding for interval-valued fuzzy sets," *Fuzzy Sets and Systems*, vol. 467, p. 108505, 2023/09/15/ 2023.

[4] T. Tan and T. Zhao, "A data-driven fuzzy system for the automatic determination of fuzzy set type based on fuzziness," *Information Sciences*, vol. 642, p. 119173, 2023/09/01/ 2023.

[5] A. A. Ewees, L. Abualigah, D. Yousri, Z. Y. Algamal, M. A. A. Al-qaness, R. A. Ibrahim, *et al.*, "Improved Slime Mould Algorithm based on Firefly Algorithm for feature selection: A case study on QSAR model," *Engineering with Computers*, vol. 38, pp. 2407-2421, 2022/08/01 2022.

[6] Z. Y. Algamal, M. H. Lee, and A. M. J. J. o. C. Al-Fakih, "High-dimensional quantitative structure–activity relationship modeling of influenza neuraminidase a/PR/8/34 (H1N1) inhibitors based on a two-stage adaptive penalized rank regression," vol. 30, pp. 50-57, 2016.

[7] A. F. Lukman, Z. Y. Algamal, B. G. Kibria, K. J. C. Ayinde, C. Practice, and Experience, "The KL estimator for the inverse Gaussian regression model," vol. 33, p. e6222, 2021.

[8] Z. J. E. J. o. A. S. A. Algamal, "An efficient gene selection method for high-dimensional microarray data based on sparse logistic regression," vol. 10, pp. 242-256, 2017.

[9] F. Smarandache, "Neutrosophic Statistics is an extension of Interval Statistics, while Plithogenic Statistics is the most general form of statistics (second version)," *International Journal of Neutrosophic Science*, vol. 19, pp. 148-165, 2022.

[10] Z. Khan, A. Amin, S. A. Khan, M. J. N. S. Gulistan, and Systems, "Statistical development of the neutrosophic Lognormal model with application to environmental data," vol. 47, p. 1, 2021.

[11] F. Smarandache, *Introduction to neutrosophic statistics*: Infinite Study, 2014.

[12] M. J. N. S. S. Ahsan-ul-Haq, "Neutrosophic Kumaraswamy distribution with engineering application," vol. 49, pp. 269-276, 2022.

[13] W.-Q. Duan, Z. Khan, M. Gulistan, A. Khurshid, and Z. Stevic, "Neutrosophic Exponential Distribution: Modeling and Applications for Complex Data Analysis," *Complexity*, vol. 2021, pp. 1-8, 2021.

[14] C. J. H. J. o. M. Granados and Statistics, "Some discrete neutrosophic distributions with neutrosophic parameters based on neutrosophic random variables," vol. 51, pp. 1442-1457, 2022.

[15] Z. Khan, M. M. A. Almazah, O. Hamood Odhah, H. M. Alshanbari, and T. Mehmood, "Generalized Pareto Model: Properties and Applications in Neutrosophic Data Modeling," *Mathematical Problems in Engineering*, vol. 2022, pp. 1-11, 2022.

[16] G. S. Rao, "Neutrosophic Log-Logistic Distribution Model in Complex Alloy Metal Melting Point Applications," *International Journal of Computational Intelligence Systems*, vol. 16, 2023.

[17] F. Shah, M. Aslam, Z. Khan, M. M. A. Almazah, F. S. Alduais, and M. Gulzar, "On Neutrosophic Extension of the Maxwell Model: Properties and Applications," *Journal of Function Spaces*, vol. 2022, pp. 1-9, 2022.

[18] A. El-Gohary, A. Alshamrani, and A. N. Al-Otaibi, "The generalized Gompertz distribution," *Applied Mathematical Modelling*, vol. 37, pp. 13-24, 2013/01/01/ 2013.

DOI: https://doi.org/10.54216/IJNS.210313
Received: February 18, 2023 Revised: May 21, 2023 Accepted: June 24, 2023

[19] K. Lee and J.-I. Seo, "Different Approaches to Estimation of the Gompertz Distribution under the Progressive Type-II Censoring Scheme," *Journal of Probability and Statistics,* vol. 2020, pp. 1-7, 2020.
[20] M. A. Khaleel, N. H. Al-Noor, M. K. J. P. o. E. Abdal-Hameed, and N. Sciences, "Marshall Olkin exponential Gompertz distribution: Properties and applications," vol. 8, pp. 298-312, 2020.
[21] L. A. J. I. Zadeh and control, "Fuzzy sets," vol. 8, pp. 338-353, 1965.
[22] T. Pathinathan, K. J. A. o. P. Ponnivalavan, and A. Mathematics, "Reverse order triangular, trapezoidal and pentagonal fuzzy numbers," vol. 9, pp. 107-117, 2015.
[23] I. Alkanani, F. A. J. M. T. Adnan, and Modeling, "Ranking function methods for solving fuzzy linear programming problems," vol. 4, pp. 65-72, 2014.

DOI: https://doi.org/10.54216/IJNS.210313
Received: February 18, 2023 Revised: May 21, 2023 Accepted: June 24, 2023