Journal of Economics and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq
DOI: https://doi.org/10.33095/jeas.v28i133.2360

Choosing the best method for estimating the survival function of inverse Gompertz distribution by using Integral mean squares error (IMSE)

Mustafa A. Hashim [1]
Woman Studies Center, Baghdad, Iraq
mostafa.adnan1201a@coadec.uobaghdad.edu.iq

Suhail N. Abood[2]
Department of Statistics, Baghdad, Iraq
suhnaj2005@coadec.uobaghdad.edu.iq

Received:28/4/2022 Accepted: 30/5/2022 Published: September / 2022

Abstract

In this research , we study the inverse Gompertz distribution (IG) and estimate the survival function of the distribution , and the survival function was evaluated using three methods (the Maximum likelihood, least squares, and percentiles estimators) and choosing the best method estimation ,as it was found that the best method for estimating the survival function is the squares-least method because it has the lowest IMSE and for all sample sizes.

Paper type: Research paper

Keywords: Inverse Gompertz Distribution, Maximum Likelihood Method, Least Squares Method, Method of Percentiles estimators, simulation, IMSE

[1]Corresponding author, University of Baghdad, Women's Studies Center
[2]Assistant Prof., University of Baghdad, College of Administration and Economics, Department of Statistics, Iraq.

1. Introduction

Survival analysis is a branch of statistics for analyzing the expected duration of time until one event occurs ;such data describe the length of time from a time origin to an endpoint of interest. For example, individuals might be followed from birth to the onset of some disease, or the survival time after the diagnosis of some disease might be studied. Survival analysis methods are usually used to analyze data collected prospectively in time, such as data from a prospective cohort study or data collected for a clinical trial"9". The main motive at the beginning of studies and research related to survival analysis is the human need to continue living in a better way .

Survival analysis studies are concerned with knowing the duration of survival for a person who develops a specific disease, where survival analysis is defined as

a survey of the time extended from (the beginning of the injury), that is, the beginning of the study to the endpoint (the point at which the tumor or disease developed). And through this research, we will estimate the survival function of the Inverse Gompertz Distribution, one of the extensions of the Gompertz Distribution.

2. Inverse Gompertz Distribution

Eliwa, M. S and El-Morshedy"7" in 2019 They have introduced the (inverse Gompertz distribution).

Let y be the random variable of the (Gompertz distribution) GD with shape parameter α and scale parameter β. The probability density function (pdf) for the random variable y is:

$$g(y) = \alpha e^{\frac{-\alpha}{\beta}(e^{\beta y}-1)+\beta y} \quad \dots (1)$$

And the absolute value of the Jacobian effect of substituting the variable y for the variable x is:

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^{n} |J| = |-\frac{1}{x^2}| = \frac{1}{x^2} > 0 \quad \dots (2)$$

So the probability density function for the (IGD) is:

$$f(x) = g\left(\frac{1}{x}\right) * |J|$$

$$f(x) = \frac{\alpha}{x^2} e^{\frac{-\alpha}{\beta}\left(e^{\frac{\beta}{x}}-1\right)+\frac{\beta}{x}} \quad \dots (3)$$

As for the cumulative distribution function (c.d.f) for the (IGD), it is as follows:

$$F(x) = p(X \le x)$$

$$F(x) = \int_0^x e^{\frac{-\alpha}{\beta}\left(e^{\frac{\beta}{t}}-1\right)} \frac{\alpha}{t^2} e^{\frac{\beta}{t}}\, dt$$

When we substitute the limits of the integration, we find that the integration is defective of the first type, that is, concerning the lower limit.

$$F(x) = e^{\frac{-\alpha}{\beta}\left(e^{\frac{\beta}{x}}-1\right)} - lim[\frac{1}{e^{\frac{\alpha}{\beta}\left(e^{\frac{\beta}{t}}-1\right)}}]$$

$$F(x) = e^{\frac{-\alpha}{\beta}\left(e^{\frac{\beta}{x}}-1\right)} \quad , x>0 \qquad ... (4)$$

It can be verified that the function f(x) is the probability density function when the following two conditions are met:
- The function f(x) is positive for all values of x, α, β .
- The integral of this function over its domain is integer one

$$\int_0^{\infty} f(x)dx = \int_0^{\infty} \frac{\alpha}{x^2} e^{\frac{-\alpha}{\beta}\left(e^{\frac{\beta}{x}}-1\right)+\frac{\beta}{x}} \, dx$$

This integral is an improper Integral of the third type concerning the upper limit as well as the lower limit, equaling one.

The survival function of IGD IS

$$S(x) = 1 - e^{\frac{-\alpha}{\beta}\left(e^{\frac{\beta}{x}}-1\right)} \qquad ... (5)$$

This function is continuous, differentiable, monotonic, decreasing, and its values lie between zero and one.

2.1 Estimation Method:

2.1.1 Maximum Likelihood Method:

The scientist Johann Carl Friedrich Gauss (30 April 1777 – 23 February 1855) who a German mathematician and physicist was the first to formulate the method of the maximum likelihood function"10" . The extracted estimators ,according to this method, are characterized by being efficient estimators and having the property of least possible variance as well as very important property, which is the property of stability, and it is more accurate when the sample size is increased (n). It depends on the idea of finding the estimator that makes the likelihood function of the observations at its great end "3" .
Let (x1,x2,….,xn) be a random sample of size (n) drawn from a population whose items are distributed according to the probability density function of the inverse Gompertz distribution with two parameters (α,β).The likelihood function for the observations is according to the following formulas:

$$\ell(\alpha,\beta; x1, \dots, xn) = \prod_{i=1}^{n} f(xi)$$

$$\ell(\alpha,\beta; x1, \dots, xn) = \frac{\alpha^n}{\prod_{i=1}^{n} xi^2} e^{\frac{-\alpha}{\beta}\sum_{i=1}^{n}\left(e^{\frac{\beta}{xi}}-1\right)} + \beta \sum_{i=1}^{n} \frac{1}{xi}$$

By taking the natural logarithm of the likelihood function, we get the two maximum- likelihood estimators for the two parameters α,β by maximizing the function ln(L) concerning α, β:

$$Ln\,\ell = nln\alpha - 2\sum_{i=1}^{n}\ln(xi) - \frac{\alpha}{\beta}\sum_{i=1}^{n}\left(e^{\frac{\beta}{xi}} - 1\right) + \beta\sum_{i=1}^{n}\frac{1}{xi}$$

By partial derivation concerning to α we get:

$$\frac{\partial\,\ell}{\partial\alpha} = \frac{n}{\alpha} - \frac{1}{\beta}\sum_{i=1}^{n}\left(e^{\frac{\beta}{xi}} - 1\right) \qquad \ldots(6)$$

$$\frac{n}{\hat{\alpha}} - \frac{1}{\hat{\beta}}\sum_{i=1}^{n}\left(e^{\frac{\hat{\beta}}{xi}} - 1\right) = 0$$

$$n\hat{\beta} - \hat{\alpha}\sum_{i=1}^{n}\left(e^{\frac{\hat{\beta}}{xi}} - 1\right) = 0 \qquad \ldots(7)$$

From equation (7) we find the value of ^

$$\hat{\alpha} = \frac{\sum_{i=1}^{n}\left(e^{\frac{\hat{\beta}}{xi}} - 1\right)}{n\hat{\beta}} \qquad \ldots(8)$$

By partial derivation concerning to β we get :

$$\frac{\partial\,\ell}{\partial\beta} = -\frac{\alpha}{\beta}\left(\sum_{i=1}^{n}\frac{1}{xi}e^{\frac{\beta}{xi}}\right) + \frac{\alpha}{\beta^{2}}\sum_{i=1}^{n}\left(e^{\frac{\beta}{xi}} - 1\right) + \sum_{i=1}^{n}\frac{1}{xi} \qquad \ldots(9)$$

$$= -\frac{\hat{\alpha}}{\hat{\beta}}\left(\sum_{i=1}^{n}\frac{1}{xi}e^{\frac{\hat{\beta}}{xi}}\right) + \frac{\hat{\alpha}}{\hat{\beta}^{2}}\sum_{i=1}^{n}\left(e^{\frac{\hat{\beta}}{xi}} - 1\right) + \sum_{i=1}^{n}\frac{1}{xi} = 0 \qquad \ldots(10)$$

By solving the above equation (8) and (10) using one of the numerical methods, we find the estimators $\hat{\beta}, \hat{\alpha}$ for the two parameters α, β and by substituting these two estimators we get the estimator of the survival function:

$$\hat{S}_{(x)mle} = 1 - e^{-\frac{\hat{\alpha}}{\hat{\beta}}\left(e^{\frac{\hat{\beta}}{x}} - 1\right)} \qquad \ldots(11)$$

2.1.2 Least Squares (LS) Method

The Least - squares method is one of the methods in estimating the parameters of the probability distribution because it works to find estimators by minimizing the sum of squares of errors (the difference) between the cumulative distribution function CDF for the studied distribution and one of the nonparametric estimators, for the cumulative probability function "1".

Let $x_1,\ldots, x_n$ be a random sample of size n from the IGD in increasing order. The LS estimates can be obtained by minimizing the following expression.

$$S(\alpha, \beta) = \sum_{i=1}^{n}\left(F(xi) - E\{F_{(Xi:n)}\}\right)^{2} \qquad \ldots(12)$$

$$= \sum_{i=1}^{n}\left(e^{\frac{-\alpha}{\beta}\left(e^{\frac{\beta}{xi}} - 1\right)} - \frac{i}{n+i}\right)^{2} \qquad \ldots(13)$$

$$\frac{\partial S}{\partial\alpha} = 2\sum_{i=1}^{n}\left[e^{\frac{-\alpha}{\beta}\left(e^{\frac{\beta}{xi}} - 1\right)} - \frac{i}{n+i}\right]\left[e^{\frac{-\alpha}{\beta}\left(e^{\frac{\beta}{xi}} - 1\right)}\left(-\frac{1}{\beta}\left(e^{\frac{\beta}{xi}} - 1\right)\right)\right] \qquad \ldots(14)$$

$$\sum_{i=1}^{n}\left[e^{-\frac{\hat{\alpha}}{\hat{\beta}}\left(e^{\frac{\hat{\beta}}{xi}} - 1\right)} - \frac{i}{n+i}\right]\left[e^{\frac{\hat{\beta}}{xi}} - 1\right]\left[e^{-\frac{\hat{\alpha}}{\hat{\beta}}\left(e^{\frac{\hat{\beta}}{xi}} - 1\right)}\right] = 0 \qquad \ldots(15)$$

By partial differentiation of equation (13) concerning β, we get:

$$\sum_{i=1}^{n}\left[e^{-\frac{\hat{}}{\hat{}}\left(e^{\frac{\hat{}}{xi}}-1\right)}-\frac{i}{n+i}\right]\left[\left(\frac{1}{\hat{}\,xi}\,e^{\frac{\hat{}}{xi}}\right)-\frac{1}{\hat{}^{2}}\left(e^{\frac{\hat{}}{xi}}-1\right)\right]e^{-\frac{\hat{}}{\hat{}}\left(e^{\frac{\hat{}}{xi}}-1\right)}=0 \qquad ...(16)$$

By solving equations (15) and (16) using one of the numerical methods, we get the estimator of least squares $\hat{\alpha}$, $\hat{\beta}$ for the two parameters α, β and by substituting these two estimators, we get the estimator of the survival function:

$$\hat{S}_{(x)LS}=1-e^{-\frac{\hat{\alpha}}{\hat{\beta}}\left(e^{\frac{\hat{\beta}}{x}}-1\right)} \qquad \ldots(17)$$

<u>2.1.3 method of percentiles estimators</u>

The percentile estimator is a statistical method used to estimate the parameters by comparing the sample points with the theoretical issues. Kao (1958, 1959)"4 " originally suggested this method depends on the cumulative distribution function ,assuming that (pi) is the estimator of the distribution function $F_{(xi)}$ by finding the estimators that make the function

$$\sum_{i=1}^{n}\left(pi-F_{(xi)}\right)^{2}$$ **at its lower end.**

To estimate the parameter α, we assume that β is known.

$$\mathrm{F}(xi)=e^{\frac{-\alpha}{\beta}\left(e^{\frac{\beta}{xi}}-1\right)}$$

After equating the estimator (pi) with the cumulative distribution function, we get the following:

$$Pi=e^{\frac{-\alpha}{\beta}\left(e^{\frac{\beta}{xi}}-1\right)} \qquad \ldots(18)$$

To simplify, we take the logarithm of both sides.

$$lnPi=\frac{-\alpha}{\beta}\left(e^{\frac{\beta}{xi}}-1\right) \qquad \ldots(19)$$

$$\sum_{i=1}^{n}\left[ln(p_i)+\frac{\alpha}{\beta}\left(e^{\frac{\beta}{xi}}-1\right)\right]^{2}=0 \qquad \ldots(20)$$

Differentiating the parameter α and dividing by 2, we get:

$$\sum_{i=1}^{n}\left[ln(p_i)+\frac{\alpha}{\beta}\left(e^{\frac{\beta}{xi}}-1\right)\right]\frac{1}{\beta}\left(e^{\frac{\beta}{xi}}-1\right)=0 \qquad ...(21)$$

The nonparametric estimator pi has the following form:

$$pi=\frac{i-0.3}{n+0.25}$$

To simplify, we assume that:

$$K1=\frac{1}{\beta}\left(e^{\frac{\beta}{xi}}-1\right)$$

Substituting into equation (20), we get :

$$\sum_{i=1}^{n}k1\,ln(p_i)+\sum_{i=1}^{n}k1\frac{\alpha}{\beta}\left(e^{\frac{\beta}{xi}}-1\right)=0 \qquad \ldots(22)$$

By partial differentiation concerning β, we get:

$$\sum_{i=1}^{n}[ln(p_i)+\frac{\alpha}{\beta}\left(e^{\frac{\beta}{xi}}-1\right)](\frac{\alpha}{\beta xi}e^{\frac{\beta}{xi}}-\frac{\alpha}{\beta^2}e^{\frac{\beta}{xi}}+\frac{\alpha}{\beta^2})=0 \quad \dots (23)$$

To simplify, we assume that:

$$K2=\frac{\alpha}{\beta xi}e^{\frac{\beta}{xi}}-\frac{\alpha}{\beta^2}e^{\frac{\beta}{xi}}+\frac{\alpha}{\beta^2}$$

$$\sum_{i=1}^{n}k2\,ln(p_i)+\sum_{i=1}^{n}k2\frac{\alpha}{\beta}\left(e^{\frac{\beta}{xi}}-1\right)=0 \quad \dots (24)$$

By solving equations (22) and (24) using one of the numerical methods, we get the estimator of the percentiles method $\hat{\alpha}$, $\hat{\beta}$ for the two parameters α, β and by substituting these two estimators , we get the estimator of the survival function:

$$\hat{S}_{(x)PER}=1-e^{-\frac{\hat{\alpha}}{\hat{\beta}}\left(e^{\frac{\hat{\beta}}{x}}-1\right)} \quad \dots (25)$$

3. Results and Discussion

In this section, the simulation method is dealt with to generate IGD data with different sample sizes and to choose default values for the parameters to compare the other estimation methods to estimate the survival function to know the preference of these methods.

1- Choosing the default values for the two distribution parameters (α=0.5, 0.9, 1.5 and β=0.3, 1.2, 1.5) for different sample sizes n=[20,50,75, 100,150] from N= 1000 replications .

2- Generating the random variable that is distributed according to the distribution (IGD) with the two parameters (α,β).

$$x=\frac{\beta}{Ln(1-\frac{\beta}{\alpha}LnU)} \quad \dots (26)$$

3- Estimation of the survival functions for the IGD distribution using the above estimation methods.

4- Comparison between the methods according to the statistical comparison standard (IMSE)

$$IMSE(\hat{S}(x))=\frac{1}{q}\sum_{i=1}^{q}\left[\frac{1}{n_i}\sum_{j=1}^{ni}\left(\hat{S}_i(x_i)-S(x_i)\right)^2\right] \quad \dots (27)$$

$i=1,2\dots\dots q$ **, q= Replication**

$\hat{S}_{(x)}=$ **Estimator of the survival function according to the estimation method**

$S_{(X)}=$ **The survival function of initial values, n_i= The number of values for x**

Here, we review the simulation results. The table below shows simulation experiment results to estimate the survival function of the IGD distribution according to the estimation methods (MLE, LS, and Percentile).

Table (1) Value of IMSE for the Estimated Survival Function by MLE, LS & Percentile

α=0.5 & β=0.3				
N	Method			Best
	IMSE $\widehat{S}(x)$ MLE	IMSE $\widehat{S}(x)$ LS	IMSE $\widehat{S}(x)$ Percentile	
20	0.0035000	0.0000110	0.0009670	LS
50	7.8090984	0.0000318	0.0003977	LS
75	1.6028029	0.0000533	0.0002357	LS
100	0.0667542	0.0000548	0.0001704	LS
150	0.0022265	0.0000630	0.0001300	LS
α=0.9 & β=1.2				
20	0.0018308	0.0000375	0.0008699	LS
50	0.0008461	0.0000812	0.0003714	LS
75	0.0003429	0.0000859	0.0002230	LS
100	0.0002252	0.0000783	0.0001635	LS
150	1.5383098	0.0000782	0.0001262	LS
α=1.5 & β=1.5				
20	0.0132407	0.0000162	0.0007196	LS
50	0.0011878	0.0000794	0.0003710	LS
75	0.0004009	0.0000861	0.0002227	LS
100	0.0002530	0.0000786	0.0001633	LS
150	0.0001652	0.0000786	0.0001260	LS

The simulation experiment results in the above table, at default parameter values and for different sample sizes, in the first model , we found the LS method is better than the ML and percentile method for all sample sizes. To possess less (IMSE), in the second and third models the preference was given to the LS method in all sample sizes.

4. Conclusion

The best method to estimate the survival function of the Inverse Gompertz Distribution according to the parameter values and sample sizes assumed in this research is the (Least – squares) method .

References

1. Abu-Zinadah, H.H., 2014. Six methods of estimations for the shape parameter of exponentiated Gompertz distribution. Applied Mathematical Sciences, 8(88).
2. Anis, M.Z., 2020. The Unit-Gompertz Distribution: Characterizations and Properties. arXiv preprint arXiv:2010.04347.
3. Aslam, M., Afzaal, M. and Bhatti, M.I., 2021. A study on exponentiated Gompertz distribution under Bayesian discipline using informative priors. Statistics in Transition New Series, 22(4), pp.101-119.
4. Dey, S., Moala, F.A. and Kumar, D., 2018. Statistical properties and different methods of estimation of Gompertz distribution with the application. Journal of Statistics and Management Systems, 21(5), pp.839-876.
5. El-Bassiouny, A.H., El-Damcese, M., Mustafa, A. and Eliwa, M.S., 2017. Exponentiated generalized Weibull-Gompertz distribution with application in survival analysis. J. Stat. Appl. Probably, 6(1), pp.7-16.
6. El-Gohary, A., Alshamrani, A. and Al-Otaibi, A.N., 2013. The generalized Gompertz distribution. Applied mathematical modeling, 37(1-2), pp.13-24.
7. Eliwa, M.S., El-Morshedy, M. and Ibrahim, M., 2019. Inverse Gompertz distribution: properties and different estimation methods with application to complete and censored data. Annals of data science, 6(2), pp.321-339.
8. Elshahhat, A., Aljohani, H.M. and Afify, A.Z., 2021. Bayesian and Classical Inference under Type-II Censored Samples of the Extended Inverse Gompertz Distribution with Engineering Applications. Entropy, 23(12), p.1578.
9. Kartsonaki, C., 2016. Survival analysis. Diagnostic Histopathology, 22(7), pp.263-270.
10. Lenart, A., 2014. The moments of the Gompertz distribution and maximum likelihood estimation of its parameters. Scandinavian Actuarial Journal, 2014(3), pp.255-277.

Journal of Economics and Administrative Sciences Vol.28 (NO. 133) 2022, pp. 149-157

اختيار أفضل طريقة لتقدير دالة البقاء لتوزيع Gompertz Inverse باستخدام متوسط مربعات الخطأ التكاملي (IMSE)

الباحث/ مصطفى عدنان هاشم[1]
مركز دراسات المرأة ، بغداد ، العراق
mostafa.adnan1201a@coadec.uobaghdad.edu.iq

أ.م. سهيل نجم عبود[2]
قسم الاحصاء ، بغداد ، العراق
suhnaj2005@coadec.uobaghdad.edu.iq

Received:28/4/2022 Accepted: 30/5/2022 Published: September / 2022

مستخلص البحث:

في هذا البحث قمنا بدراسة توزيع (IG) (Inverse Gompertz) وتقدير دالة البقاء على قيد الحياة للتوزيع، وتم تقدير دالة البقاء باستعمال ثلاث طرق وهي (طريقة الامكان الاعظم والمربعات الصغرى وطريقة المقدرات التجزيئية) وتم اجراء المحاكاة والمقارنة بين الطرق باستعمال متوسط مربعات الخطأ التكاملي (IMSE) واختيار أفضل طريقة للتقدير حيث تبين ان أفضل طريقة لتقدير دالة البقاء هي طريقة المربعات الصغرى وذلك لامتلاكها اقل IMSE ولجميع احجام العينات.

نوع البحث: ورقة بحثية.

المصطلحات الرئيسة للبحث: توزيع Inverse Gompertz ، طريقة الإمكان الأعظم، طريقة المربعات الصغرى، طريقة المقدرات التجزيئية ،المحاكاة ،متوسط مربعات الخطأ التكاملي .

*البحث مستل من رسالة ماجستير

(1) احصائي في مركز دراسات المرأة /جامعة بغداد
(2) استاذ مساعد في قسم الاحصاء / كلية الادارة والاقتصاد /جامعة بغداد