JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-19, No.-11, November (2024) pp 191 - 206 ISSN (Print) 0973-

ESTIMATION OF ONE-AND-FIVE DIMENSIONAL SURVIVAL FUNCTIONS FOR CATEGORICAL DATA USING ENTROPY

Hasanain Jalil Neamah Alsaedi

University of information technology and communications, Iraq.

Acharya Nagarjuna University / India / state of Andhra Pradesh. College of sciences department of Statistics

Email : Hasanien.1975@uoitc.edu.iq

https://doi.org/10.26782/jmcms.2024.11.00013

(Received: September 03, 2024; Revised: October 27, 2024; Accepted: November 06, 2024)

Abstract

Life tables are used in many fields in demographic and health research They represent an important indicator of death in society. There are two types of life tables; complete life tables are based on the age at death based on single-age categories and are obtained using a comprehensive survey method. The second type is the abbreviated life tables which are based on the age at death of five-year age groups and are obtained by the sample survey method. In this research, the survival function was estimated for the data obtained from the Central Statistical Organization, social and Economic Survey of the Family in Iraq (IHSES II) using parametric methods (the principle of Maximizing Entropy method (POME), and maximum likelihood method (MLE)), as well as the use of A non-parametric approach, the kernel smoothing method (KS), the compared between the estimation methods using (RMSE) and (MAPE). One of the most important conclusions was the emergence of a preference for the (POME) method for the five-age groups, but in the case of the single-age groups, the (KS) method is the best.

Keywords: *life tables, the principle of maximum entropy method, kernel smoothing method.*

I. Introduction

When the population census for some countries is absent, obtaining an accurate calculation of the probability of survival through the survival function within the life tables of single-age groups would extremely be difficult, let alone the problems of disturbance that require mathematical treatment according to an appropriate statistical probability distribution that is compatible with the phenomenon of life or death. The entropy method is a measure of disturbances that affect a specific phenomenon. This method comes within the framework of information related to the probability distribution itself. This method is used in many different scientific fields. Entropy can be defined as a method for estimating probability functions within specific constraints

that can be expressed by integrating the probability function itself. This principle in entropy is called the Principle of Maximizing Entropy (POME) (Cropper, 2004).
Demographic data often suffer from behavioral instability and volatility, especially that related to the survival function of the Iraqi population according to age groups. The research problem appears in the presence of data with only five-year age groups that are dealt with and analyzed by the Central Statistical Organization in Iraq. To add more, there is no data on the survival function based on single-age groups, let alone the existence of a theoretical analytical problem that hinders the issue of estimating the Generalized Gamma Distribution with three parameters since it involves the incomplete gamma integral when finding the survival function, which makes the formulation of the inferential processes more complicated, the traditional method of estimation is very difficult. (Jowitt, 1979).

In the present research, we will estimate the parameters of the generalized gamma distribution for a sample of data obtained from the Central Statistical Organization in Iraq from the survey (IHSES-II 2012), using the maximum likelihood method (MLE), the principle of maximum entropy function (POME), and the kernel bootstrap method (Ks). After that, the survival functions for the population for each method and compare the three methods will be found using the statistical measures of root mean square error (RMSE), the mean absolute relative error (MAPE), and the Sprague Multipliers will be used to convert the five-age categories of the population in Iraq, which are (18) ones, into single-age categories, which will result in (81) single categories.

I.i. General Gamma Distribution

The three-parameter generalized gamma distribution, abbreviated as GG, was found by Stacy in 1962. He assumed that the location parameter (μ=0). The probability function of the three-parameter GG distribution can be expressed in the following formula:

$$f(t,\theta,\beta,k) = \frac{\beta}{\Gamma(k)\,\theta^{\beta k}}\, t^{\beta k-1}\, e^{-(\frac{t}{\theta})^{\beta}} \,, t \geq 0, k \geq 0, \theta \geq 0, \beta \geq 0 \quad (1)$$

Where the symbol (t) represents the random variable (age), (θ) represents the scale parameter, and the symbols (k,β) represent the shape parameters.

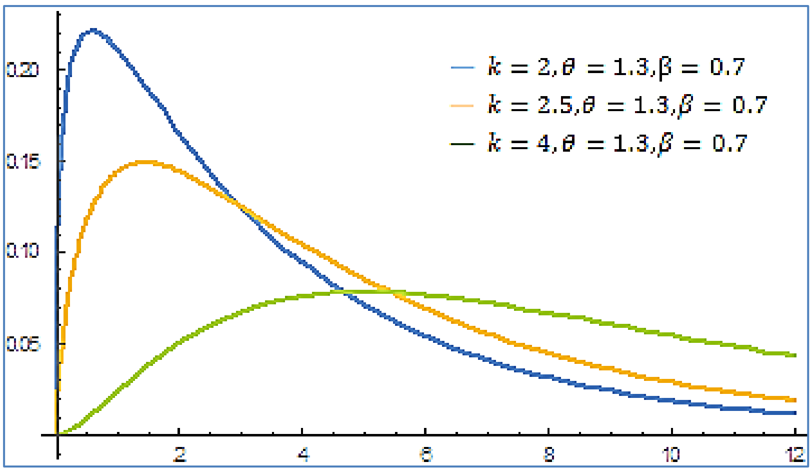

Fig.1. Probability density function of the general three-parameter gamma distribution for several cases

$$F(t) = \int_o^t \frac{\beta}{\Gamma(k)\theta^{\beta k}} x^{\beta k-1} e^{-\left(\frac{x}{\theta}\right)^{\beta}} dx$$

$$= \frac{1}{\Gamma(k)} \int_0^{\left(\frac{t}{\theta}\right)^{\beta}} u^{k-1} e^{-u} du = \frac{IG((\frac{t}{\theta})^{\beta}, k)}{\Gamma(k)} \tag{2}$$

Since IG(S, K)=1/(Γ(k)) ∫_0^s 〖u^(k-1) e^(-u) (24&du)〗 is the incomplete function which can be expressed in the following figure (2):

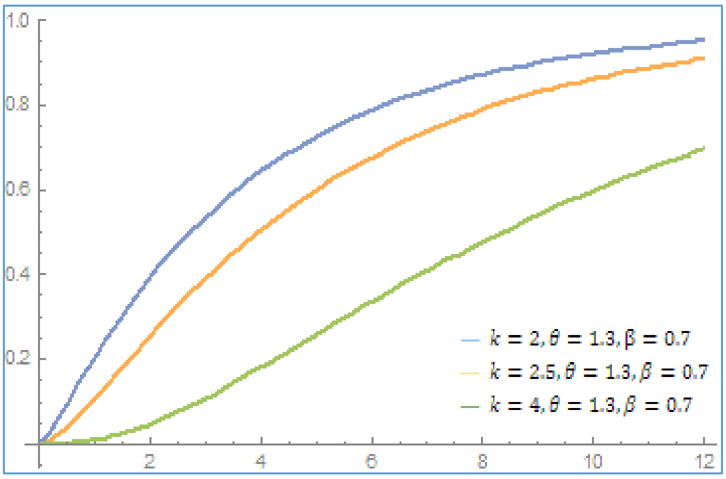

Fig. 2. represents the cumulative distribution function of the general three-parameter gamma distribution for several cases.

The mean value of the variable (t) of the rank (β) is:

$$E(t^{\beta}) = \frac{\theta^{\beta}\Gamma(k+1)}{\Gamma(k)} = \theta^{\beta} k \tag{3}$$

The average logarithm of the variable (t) is:

$$E(\ln t) = \frac{\Psi(k)}{\beta} + \ln\theta \tag{4}$$

Since:

$$\Psi(k) = \frac{d}{dk}\ln(\Gamma(k)) = \frac{\Gamma'(k)}{\Gamma(k)} \tag{5}$$

I.ii. Survival Function Analysis

Survival function analysis tackles analyzing the time required before the occurrence of one or more events such as a breakdown in the case of machines or death in the case of living organisms. Through survival function analysis, we try to an answer to such several questions (What is the probability of the population that will survive to a certain time? How can multiple causes of death be determined? Which age groups can survive? How do certain conditions reduce or increase the probability of survival?).

To answer these questions, we must define a (time of life) in the case of biological survival, and death is a fact, but mechanical failures and reliability may not be clear-cut in the industrial aspect, for example, and some events (such as the failure of an organ system) may have the same ambiguity (Suleiman and Farhan, 2014).

The survival function can be defined as the probability that an event (death) has not occurred yet according to time (t), thus (T) denotes the time until death, and the symbol s (t) denotes the probability of survival, since:

$$(t) = 1 - F(t) = P\ (T\ >\ t)\ , \forall\ t\ >\ 0 \qquad (6)$$

Let us assume that the probability of the injured person remaining alive at time (0) is equal to one, i.e. S (0) = 1, and that the survival function is non-increasing and continuous on the right side, i.e. :

$$S(u) \leq S(t) \quad , if\ \ u > t$$

The probability of survival approaches zero as the organism's age increases.

$$S(t) \rightarrow 0 \quad as \quad t \rightarrow \infty$$

That is, the survival time cannot be negative, and the survival function for a continuous distribution is:

$$(t) = p(T > t) = \int_t^{\infty} p(t)\ dt \qquad (9)$$

Thus, the survival function of the generalized gamma distribution can be expressed by the following formula (Qamruz & Karl, 2011):

$$S_{GG}(t) = 1 - \frac{IG((\frac{t}{\theta})^{\beta},k)}{\Gamma(k)} \qquad (10)$$

I.iii. Maximizing the Place Function

The maximum likelihood method is considered one of the important and commonly used methods in estimation since it includes good properties, including consistency, and inversion, and is often unbiased. The principle of this method lies in finding an estimation of the parameters that make the likelihood function at its maximum limit, in addition to the fact that it keeps pace with all samples. Since the maximum likelihood estimators are characterized by the property of stability, i.e. if T=g(θ) is a function in terms of the variable (θ), let us assume that (θˆ_n) is the maximum likelihood estimator of the variable (θ), then Tˆ_n=g(θˆ_n) is the maximum likelihood estimator of the function (T). Thus, the likelihood function of the general gamma distribution with three parameters (θ, k, β) can be expressed by the following equation: (Chen & lio, 2009):

$$L = \prod_{i=1}^{n} f(t_i, k, \beta, \theta) \quad (11)$$

$$L = \left(\frac{\beta}{\Gamma(k)\theta^{\beta k}}\right)^n \prod_{i=1}^{n} {t_i}^{\beta k-1} \ e^{\frac{-\sum_{i=1}^{n} {t_i}^{\beta}}{\theta^{\beta}}} \quad (12)$$

To convert the possibility function in equation (10) to linear form, we take its natural logarithm:

$$\ln L = n \ln \beta - n\beta k \ln \theta - n \ln \Gamma(k) + (\beta k - 1)\sum_{i=1}^{n} \ln t_i - \frac{\sum_{i=1}^{n} {t_i}^{\beta}}{\theta^{\beta}}$$

$$= n \ln \beta - n\beta k \ \ln \theta - n \ln\Gamma(k) + \beta k \sum_{i=1}^{n} \ln t_i - \sum_{i=1}^{n} \ln t_i - \frac{\sum_{i=1}^{n} {t_i}^{\beta}}{\theta^{\beta}} \quad (13)$$

By finding the partial derivatives of the parameters in equation (11), and setting each partial derivative equal to zero, we obtain the estimated values of the parameters (θ, k, β) that maximize the possibility function (Tan & Drosses, 1975)

$$\hat{k} = \frac{\hat{\theta}^{-\hat{\beta}} \sum_{i=1}^{n} {t_i}^{\hat{\beta}}}{n} \quad , when\ \theta\ \&\ \beta\ are\ constant \quad (14)$$

$$\hat{\theta} = (exp(-\Psi(\hat{k}) + \frac{\hat{\beta}\sum_{i=1}^{n} \ln t_i}{n}))^{\frac{1}{\hat{\beta}}} \quad , when\ k\ \&\ \beta\ are\ constant$$

$$\Psi(k) = \frac{d}{dk} \ln\big(\Gamma(k)\big) = \frac{\Gamma'(k)}{\Gamma(k)} \quad (15)$$

The estimated value of the parameter (β) is extracted using the Newton–Raphson method, where we obtain the estimator of the survival function for the general three-parameter gamma distribution through the stability property as in the following formula:

$$\hat{s}_{GG}(t) = 1 - \frac{IG\left(\left(\frac{t}{\hat{\theta}}\right)^{\hat{\beta}}, \hat{k}\right)}{\Gamma(\hat{k})} \quad \text{(Reshi et al., 2014)} \quad (16)$$

I.iv. Entropy Maximization Principle

The Shannon Entropy function is defined for a continuous random variable (t) with a probability distribution f(t,θ), where the symbol (θ) represents the vector of distribution parameters, and it can be mathematically formulated as follows:

$$H(f) = -\int_{-\infty}^{\infty} f(t, \theta) \ln f(t, \theta)\, dt \quad (17)$$

The symbol H(f) represents the entropy function of the function f(t,θ) and it is fulfilled with the following condition:

$$\int_{-\infty}^{\infty} f(t, \theta) dt = 1 \quad (18)$$

The Entropy function can be written as follows:

$$H(f) = E(-\ln f(t, \theta)) \quad (19)$$

The entropy function is characterized by that in case of maximizing it for a given distribution, it will produce a less biased estimate of that distribution's function, and this is called the Principle of Maximizing Entropy. In other words, by maximizing H (f), an estimate of the probability distribution will be extracted (Jaynes, 1959).

The principle of maximum entropy (POME) is also regarded as a testing method. It tests the probability function f(t,θ) after subjecting it to several constraints from which a distribution estimation method can be derived to best represent the behavior of the variable (t). One of these constraints is the use of Lagrange multipliers when estimating the parameters as found in the case of checking the optimization problem. (Jowitt, 1979).

Suppose that there are m independent linear constraints C_i such that i= 1,2,…m

$$ci = \int w_i \, f(t,\theta)\, dt \qquad (20)$$

Since (w_i) are known functions defined on f(t,θ), the constraints C_i will be applied to maximize the function H(f) which will produce the following probability function:

$$f(t,\theta) = e^{-a_0-\sum_{i=1}^{m} a_i w_i(t)} \;\; , \; \forall i = 1,2,\dots m \qquad (21)$$

Since (a_i) are Lagrange multiples, by substituting equation (20) into equation (18) we get:

$$H(f) = a_0 + \sum_{i=1}^{m} a_i c_i \qquad (22)$$

Thus, the probability density function can be formulated in terms of entropy using the following formula:

$$f(t,\theta) = e^{-H(f)} \qquad (23)$$

By substituting the probability function of the general gamma distribution as in equation (1) in equation (16), the constraints are extracted, where: (Singh & Guo, (1995):

$$a_0 = \ln \Gamma(k) - \ln \beta + k\beta \ln \theta$$

$$a_1 = \frac{1}{\theta^{\beta}}$$

$$a_2 = -(k\beta - 1) = 1 - k\beta$$

According to equation (20), the probability function in terms of Lagrange multipliers will be:

$$f(t) = e^{-a_0 - a_1 t^{\beta} - a_2 \ln t}\,) \qquad (24)$$

By solving equation (23), we obtain the survival function f(〖t,a〗 _1,a_2,β):

$$S(t,a_1,a_2,\beta) = 1 - \frac{IG\left(\frac{t^{\beta}}{a_1{}^{-1}}, \frac{1-a_2}{\beta}\right)}{\Gamma\left(\frac{1-a_2}{\beta}\right)} \qquad (25)$$

When estimating the parameters (a_1,a_2,β) we estimate the survival function according to the formula: (Singh & Fiorentino, 1992):

$$\hat{S}(t,\hat{a}_1,\hat{a}_2,\hat{\beta}) = 1 - \frac{IG\left(\frac{t^{\hat{\beta}}}{\hat{a}_1{}^{-1}}, \frac{1-\hat{a}_2}{\hat{\beta}}\right)}{\Gamma\left(\frac{1-\hat{a}_2}{\hat{\beta}}\right)} \qquad (26)$$

I.v. Core Smoothing Function

All statistical applications are represented by being acquainted with the special distribution and its characteristics for the community to be studied in order to represent the community well by using common statistical methods. Each probability distribution has a probability density function f(t) with a random variable (t), when collecting a random sample containing (n) observations taken from the study community, as that community is expressed by the probability density function f(t), which is used to identify the characteristics of the studied society. The probability density function represents a basic concept in statistics, and determining the function provides us with a description of the distribution since the probabilities associated with (t) can be determined through the following relationship:

$$p_r(a \le t \le b) = \int_a^b f(t)\, dt \qquad , \forall \ a, b \in R \tag{27}$$

Assuming that a set of data points that are assumed to represent a sample drawn from a probability density function is available, then, the goal is to estimate a function from the sample observations 〖〖{t〗_i}〗_(i=1)^n which can be considered as independent random variables with an identical distribution (i.i.d). A kernel estimator is used to estimate the nonparametric probability density function f(t) for the random variable (t) in the following form: (Lagos et al., 2011)

$$\hat{f}(\underline{t}) = \frac{1}{n\,\hat{h}} \sum_{i=1}^{n} k\left(\frac{t-t_i}{h}\right) \tag{28}$$

Where (k) represents the kernel order, and k(u) represents the kernel function that satisfies the following conditions:

1-A continuous function with real nonnegative values.

2-The function k(u) is a probability density function, i.e.: ∫ 〖k(u)du=1

3-It is a symmetric function about zero, which leads to the value of the first moment being equal to zero, while the value of the second moment being finite,

$$\mu_2(k) = \int u^2 k(u) du < \infty$$

Many of the kernel functions satisfy the above conditions, including the Gaussian function, which will be used in this research.

$$k(u) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{u^2}{2}} \quad I(-\infty, \infty) \tag{29}$$

Where $u = \frac{t-t_i}{h}$, and (h) is the smoothing parameter or bandwidth, which has a significant effect on the estimation process, as the variance decreases and the bias increases with increasing bandwidth and vice versa. Therefore, this parameter affects the degree of smoothing of the curve and its closeness to the true curve. The bandwidth according to the normal scale distribution rule is:

$$\hat{h} = 1.06\,(s)\,n^{-1/5} \tag{30}$$

Where the symbol (s) represents the standard deviation. (Rao, 1983).

I.vi. The Golden Ratio Method for Estimating the Bandwidth

The golden ratio method is based on two basic ideas. The first idea is that every two numerical values achieve the golden ratio if the ratio between the larger of the two numbers and the smaller of them equals the ratio between the sum of these two numbers and the larger of them, and the first to determine the golden ratio was the scientist Leonardo Fibonacci through a sequence (Fibonacci number). The golden ratio is represented by the number (1.618034). It is calculated in two ways (Calot & Sardon, 2004):

-The first way:

$$\varphi = \frac{1+\sqrt{5}}{2}\,) \tag{31}$$

-The second way:

$$\varphi = 2\cos\,(36^0) \tag{32}$$

When dividing the value of h ̂ as in the equation ((29) by the golden ratio, we get (h) as an ideal and as follows:

$$\hat{h}_{GR} = \frac{\hat{h}}{1.618034} \tag{33}$$

After estimating the probability density function, the survival probability is calculated using the following formula (Qiao & Tsokos, 1994):

$$\hat{s}(t) = p(T > t) = \int_x^{\infty} \hat{f}\,(T)\,dT = \frac{1}{n}\sum_{i=1}^{n}\int_{\frac{t-t_i}{h}}^{\infty} k(u)du \tag{34}$$

I.vii. Comparison criteria

1-Root mean square error (RMSE):

It is one of the statistical criteria used to measure the accuracy of the resulting estimators to estimate a parameter. If (θ) is a parameter of distribution and ((θ_(1)) ̂,(θ_2) ̂,(θ_3,) ̂…,(θ_n) ̂) are (n) estimators of the parameter (θ), where (R) represents the number of iterations of the generated sample size and the general formula for the root mean square error (RMSE):

$$RMSE(\hat{\theta}) = \sqrt{\frac{1}{R}\sum_{i=1}^{n}\left(\theta_i - \hat{\theta}_i\right)^2} \tag{35}$$

2- Mean absolute relative error (MAPE):

This error can be calculated by calculating the absolute average of the difference between the actual value achieved and the predicted value. The general formula for (MAPE) is:

$$MAPE(\hat{\theta}) = \frac{1}{R}\sum_{i=1}^{n}\left|\frac{\theta_i-\hat{\theta}_i}{\theta_i}\right| \tag{36}$$

II. Research-Data Description

Research data under study were obtained from the Iraqi Central Statistical Organization - Household Social and Economic Survey (IHSES II 2012), which provides detailed information about the living conditions of every Iraqi citizen and family. To add more, it also presents an objective picture of the development of living conditions in various fields such as spending and income, demographic status, housing, etc. The sample was made of about (25488) families from all Iraqi governorates, with (216) families from each district. It is worth mentioning here that the whole number of districts in Iraq is (118) ones. The number of clusters reached (2832) ones, and each cluster includes (9) families distributed among the districts and governorates and in the rural and urban environments.

The mortality numbers for the survey (IHSES II 2012) were not collected by single age groups since the data were found in the form of five-year age groups for the number of deaths. Since we need the number of deaths and the population, a mathematical method was used that leads to the results based on Sprague multipliers in order to derive the death toll or population numbers by single age within given five-year age groups using the death toll or population numbers in these age groups and adjacent five-year groups. The survival function will be estimated using five-year and single-year data, and the survival function will be compared using the three methods in the case of five-year and single-year data (Iraq Household Social and Economic Survey, 2012).

II.i. Testing Research Data

The five-class data was tested using the statistical program (Easy fit). It was found that the data had followed the general gamma distribution ((GG) and because the general gamma distribution is one of the exponential distributions used to describe the human mortality rate, the general gamma distribution was used to estimate the survival function.

● Test hypothesis

H_0: The data follow the general gamma distribution (GG)
H1: The data do not follow the general gamma distribution (GG)
The results obtained from the data were:

$$\chi^2 = \sum_{i=1}^{n}\frac{(O_i-E_i)^2}{E_i} = 0.71613, \quad p - value = 0.69903$$

So, the null hypothesis is accepted, that is, the population data follows the general gamma distribution. When testing the single-class data, the p-value reached (0.31866), which indicates that the null hypothesis is accepted, that is, the population data follows the general gamma distribution.

II.ii. Analysis of Survival Function Estimates for Five-year Age Groups

The demographic data resulting from dividing the number of deaths by the population m(x,n) for each age group for the year (2011-2012) were input to extract the survival function for the five-year groups. The methods for estimating survival functions were compared, the maximum likelihood (mle), with the entropy method (ent), and the kernel bootstrap method (ks), and the following table (1) shows the results: (United Nations, 2013)

Table 1: Results of survival function estimation methods for the five-category IHSES II survey data - urban and rural

Class	$\hat{S}_{ks}$	$\hat{S}_{ent}$	$\hat{S}_{mle}$
1-0	0.729	0.997	0.933
4-1	0.728	0.997	0.929
9-5	0.720	0.995	0.908
14-10	0.710	0.992	0.886
19-15	0.702	0.990	0.870
24-20	0.701	0.990	0.866
29-25	0.693	0.987	0.851
34-30	0.686	0.985	0.836
39-35	0.654	0.973	0.781
44-40	0.622	0.961	0.732
49-45	0.503	0.904	0.576
54-50	0.498	0.902	0.571
59-55	0.312	0.763	0.346
64-60	0.260	0.699	0.275
69-65	0.231	0.657	0.234
74-70	0.122	0.388	0.067
79-75	0.100	0.348	0.052
80 فاكثر	0.028	0.146	0.008

The behavior of estimating the survival function when using the three methods can be seen in table (1) above. This behavior starts with the probability of survival begins to drop below (0.5) at the true function at age (80 years and above), while the maximum likelihood method begins to drop at the age group (55-59), and the entropy method begins to drop at the age group (70-74), and finally the nonparametric method at the age group (50-54), as the method is the best the older the person gets and the probability of his survival is higher than (0.5).

Table (2) and Figure (3) as well show that the entropy method is the best among the other methods in estimating the population survival function - urban and rural - and for the two criteria (RMSE, MAPE):

Table 2: Statistical measures for the estimation methods of the survival function for the five-category (IHSES II) survey data - urban and rural

Methods	RMSE	MAPE
MLE	0.3758	0.3863
Kernel smoothing	0.4258	0.4795
Entropy	0.1506	0.1319

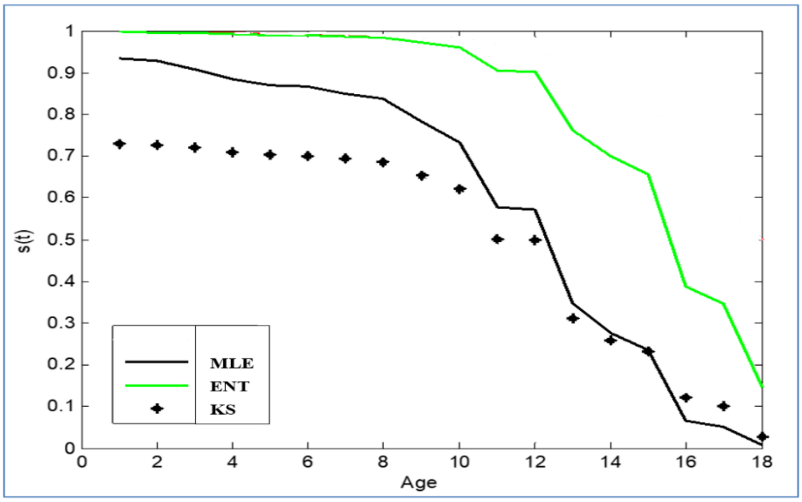

Fig.3. The estimation methods diagram of the survival function for the five-category (IHSES II) survey data- urban and rural

II.iii. Analysis of survival function estimates for single age groups

The survival function for single groups was found out, and the methods for estimating survival functions were compared, the kernel bootstrap method (ks), the entropy method (ent), and the maximum likelihood (mle), and the results were as follows:

Table 3: Results of survival function estimation methods for single-category (IHSES II) survey data - urban and rural

age	$\hat{S}_{ks}$	$\hat{S}_{ent}$	$\hat{S}_{mle}$	age	$\hat{S}_{ks}$	$\hat{S}_{ent}$	$\hat{S}_{mle}$
0	0.767	1.000	0.962	40	0.606	0.967	0.715
1	0.763	0.999	0.953	41	0.602	0.966	0.710
2	0.762	0.999	0.950	42	0.601	0.965	0.709
3	0.753	0.998	0.925	43	0.596	0.964	0.704
4	0.750	0.998	0.919	44	0.591	0.963	0.699
5	0.739	0.996	0.897	45	0.577	0.959	0.683
6	0.739	0.996	0.897	46	0.544	0.949	0.651

7	0.737	0.996	0.894	47	0.530	0.945	0.637
8	0.737	0.996	0.892	48	0.498	0.934	0.606
9	0.733	0.995	0.885	49	0.451	0.916	0.561
10	0.733	0.995	0.885	50	0.446	0.915	0.557
11	0.729	0.994	0.878	51	0.424	0.905	0.535
12	0.725	0.994	0.872	52	0.424	0.905	0.535
13	0.723	0.993	0.867	53	0.396	0.892	0.507
14	0.719	0.993	0.862	54	0.385	0.887	0.497
15	0.716	0.992	0.856	55	0.380	0.884	0.492
16	0.713	0.991	0.851	56	0.374	0.881	0.485
17	0.710	0.991	0.846	57	0.371	0.880	0.483
18	0.707	0.990	0.842	58	0.349	0.867	0.458
19	0.705	0.990	0.839	59	0.347	0.866	0.457
20	0.704	0.990	0.837	60	0.316	0.844	0.421
21	0.702	0.989	0.835	61	0.268	0.799	0.355
22	0.701	0.989	0.832	62	0.229	0.742	0.287
23	0.697	0.988	0.828	63	0.203	0.693	0.237
24	0.695	0.988	0.823	64	0.192	0.670	0.217
25	0.692	0.987	0.820	65	0.190	0.667	0.214
26	0.689	0.987	0.816	66	0.187	0.659	0.208
27	0.686	0.986	0.811	67	0.186	0.656	0.205
28	0.684	0.986	0.808	68	0.174	0.629	0.184
29	0.683	0.985	0.807	69	0.148	0.555	0.134
30	0.682	0.985	0.806	70	0.124	0.456	0.083
31	0.681	0.985	0.804	71	0.123	0.449	0.080
32	0.680	0.985	0.804	72	0.106	0.374	0.052
33	0.676	0.984	0.798	72	0.091	0.307	0.033
34	0.667	0.982	0.786	74	0.073	0.251	0.021
35	0.659	0.980	0.775	75	0.073	0.251	0.021
36	0.653	0.979	0.769	76	0.052	0.210	0.014
37	0.640	0.976	0.753	77	0.038	0.189	0.011
38	0.625	0.972	0.736	78	0.035	0.183	0.010
39	0.613	0.969	0.722	79	0.029	0.174	0.009
				80+	0.006	0.087	0.002

The results given in Table (3) show that the behavior of estimating the survival function using the entropy method begins with the probability of survival falling below (0.5) for the true function at age (51), while the entropy method begins to fall at age (70), the maximum likelihood method at age (54), and the kernel function method at age (48).

The following Table (4) shows the statistical measures (RMSE, MAPE) to compare the estimation methods for the data under study.

Table 4 : Statistical measures for the estimation methods of the survival function for the single-category IHSES II survey data - urban and rural

Methods	RMSE	MAPE
MLE	0.0618	0.1616
Kernel smoothing	0.0596	0.1222
Entropy	0.3075	1.1178

We note that the nonparametric method, the pulmonary method, is the best for the criterion (RMSE, MAPE) despite the low probability of survival in newborns.

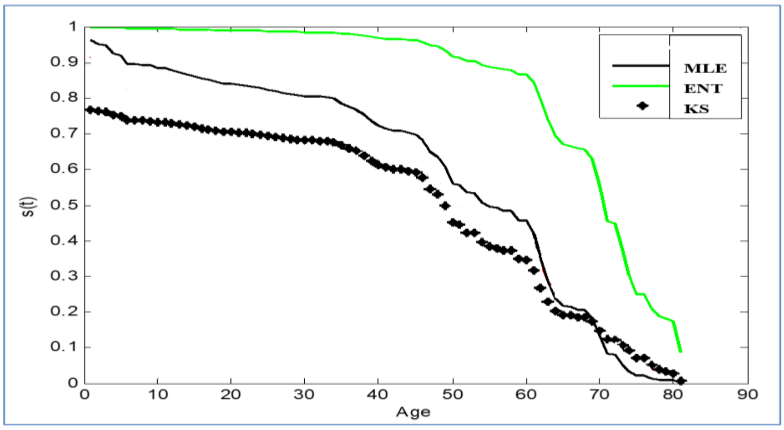

Fig. 4. The estimation methods diagram of the survival function for the single-category IHSES II survey data - urban and rural

An application for one-and-five dimensional survival functions for categorical data using entropy can significantly enhance performance as well as reliability for microstrip devices, cybersecurity protocols, mathematics and amplifiers. In microstrip devices, these survival functions can be utilized to analyze failure rates of various components under different operational conditions, thereby optimizing design parameters and improving longevity. Via assessing the entropy of categorical data related to device performance, engineers can identify critical factors that contribute to reliability, leading to more robust designs. In cybersecurity, an application for these survival functions allows for a better understanding of threat patterns and vulnerabilities over time. Via categorizing cyber incidents and analyzing their survival probabilities, security analysts can prioritize resources and develop strategies to mitigate risks effectively. For amplifiers, employing one-and-five dimensional survival functions can aid in evaluating the performance degradation over time due to various environmental factors or usage patterns. This analysis can guide a selection for materials as well as design choices that maximize efficiency and minimize failure rates [XV-XXIV].

III. Conclusions

● Via the application of the theoretical formulas of the methods used in estimating survival probabilities using data with five categories, it was concluded that it is better to use the principle of maximum entropy function (POME) estimation according to both comparison criteria (RMSE, MAPE) despite the sensitivity of the latter criterion.

● Comparing the estimation methods using single-category data, paves the way for the result that the kernel bootstrap method (ks) is the best in estimating survival functions based on the comparison criteria (RMSE, MAPE)

● Concerning the five-year age groups, the survival function rate is observed to be lower than the probability (0.5) for the (mle) method, as the ages range between (55-59), while for the entropy (POME) method, the ages range between (74-70), and for the bootstrap (ks) method, the ages range between (49-45), (54-50), (59-55). As far as the single-year age groups are concerned, the survival function rate is observed to be lower than the probability (0.5) for the principle of the greatest entropy function method at ages (69), (70), (71), as the highest survival probability that a human being reaches through the entropy method.

IV. Recommendations

● It is advisable to use the bootstrap (ks) method for estimating the survival function for single-age groups instead of five-age groups because it is more sensitive to which age group is most at risk of mortality and at the sex level. It is also recommended to use the POME method for data for five-age groups.

● It is also recommended that the management of the survey implementation should take into consideration the single-age categories, as they can be combined and become five-age categories instead of the method of breaking up the five-age data and converting it to one-age, which reduces the accuracy of the results.

● What is also recommended is the extension of the search throughout obtaining a new entropy function restricted to other distributions by introducing new constraints, then optimization problems can be employed to estimate the parameters and then estimate the survival functions.

Conflicts of interest

The author declares that they have no conflicts of interest.

References

I. Calot, Gérard, and Jean-Pierre Sardon. *Calculation of Eurostat's Demographic Indicators.* 2004.

II. Central Statistical Organization, Ministry of Planning. *Iraqi Household Social and Economic Survey (IHSES II, 2012) Tables Report.* Central Statistical Organization Press, 2014. www.cost.gov.iq.

III. Chen, Dong-Guk, and Ying-Chung Lio. "A Note on the Maximum Likelihood Estimation for the Generalized Gamma Distribution Parameters under Progressive Type-II Censoring." *International Journal of Intelligent Technologies and Applied Statistics,* vol. 2(2), 2009, pp. 145-152.

IV. Cropper, William H. *Great Physicists: The Life and Times of Leading Physicists from Galileo to Hawking*. Oxford UP, 2004. "The Road to Entropy: Rudolf Clausius."

V. Jaynes, E. T. *Probability Theory in Science and Engineering.* No. 4, Socony Mobil Oil Company Field Research Laboratory, 1959.

VI. Jowitt, P. W. "The Extreme-Value Type-1 Distribution and the Principle of Maximum Entropy." *Journal of Hydrology,* vol. 42, no. 1-2, 1979, pp. 23-38. 10.1016/0022-1694(79)90004-0

VII. Lagos Álvarez, B., Ferreira, G., and Valenzuela Hube, M. "A Proposed Reparameterization of Gamma Distribution for the Analysis of Data of Rainfall-Runoff Driven Pollution." *Proyecciones (Antofagasta),* vol. 30, no. 3, 2011, pp. 415-439. https://www.scielo.cl/pdf/proy/v30n3/art09.pdf

VIII. Qamruz, Z., and Karl, P. "Survival Analysis Medical Research." *InterStat,* 2011, http://interstat.statjournals.net/YEAR/2011/abstracts/1105005.php.

IX. Qiao, H., and C. P. Tsokos. "Nonparametric Approach to Reliability Analysis." *Proceedings of SOUTHEASTCON'94,* April 1994, pp. 231-235. IEEE.

X. Rao, B. L. S. P. *Nonparametric Functional Estimation.* Academic Press, 1983.

XI. Reshi, J. A., Ahmed, A., and Mir, K. A. "Some Important Statistical Properties, Information Measures, and Estimations of Size Biased Generalized Gamma Distribution." *Journal of Reliability and Statistical Studies*, 2014, pp. 161-179.

XII. Roshani, S., Yahya, S. I., Mezaal, Y. S., Chaudhary, M. A., Al-Hilali, A. A., Mojirleilani, A., & Roshani, S. (2023). Design of a compact quad-channel microstrip diplexer for L and S band applications. *Micromachines, 14*(3), 553.

XIII. Roshani, S., Yahya, S. I., Alameri, B. M., Mezaal, Y. S., Liu, L. W., & Roshani, S. (2022). Filtering power divider design using resonant LC branches for 5G low-band applications. Sustainability, 14(19), 12291.

XIV. Singh, V. P., and Fiorentino, M. "A Historical Perspective of Entropy Applications in Water Resources." *Entropy and Energy Dissipation in Water Resources*, 1992, pp. 21-61. https://link.springer.com/chapter/10.1007/978-94-011-2430-0_2

XV. Singh, V. P., and Guo, H. "Parameter Estimation for 3-Parameter Generalized Pareto Distribution by the Principle of Maximum Entropy (POME)." *Hydrological Sciences Journal*, vol. 40, no. 2, 1995, pp. 165-181. 10.1080/02626669509491402

XVI. Sulaiman, Abbas Najm, and Ebtihal Hussein Farhan. "Estimation of the Survival Function for Complete Real Data of Lung Cancer Patients." *Ibn Lahitham Journal of Pure and Applied Sciences*, vol. 27, no. 3, 2014, pp. 531-541. https://jih.uobaghdad.edu.iq/index.php/j/article/view/318

XVII. S. A. AbdulAmeer *et al.*, "Cyber Security Readiness in Iraq: Role of the Human Rights Activists," *International Journal of Cyber Criminology*, vol. 16, no. 2, pp. 1–14-1–14, 2022.

XVIII. Tarrad , K. M. *et al.*, "Cybercrime Challenges in Iraqi Academia: Creating Digital Awareness for Preventing Cybercrimes," *International Journal of Cyber Criminology*, vol. 16, no. 2, pp. 15–31-15–31, 2022.

XIX. Tan, P., and C. Drossos. "Invariance Properties of Maximum Likelihood Estimators." *Mathematics Magazine*, vol. 48, no. 1, 1975, pp. 37-41.

XX. United Nations. *MORT PAK for Windows*, version 4.3, New York, 2013.

XXI. Y. S. Mezaal, & Ali, J. K. (2016). Investigation of dual-mode microstrip bandpass filter based on SIR technique. PLoS one, 11(10), e0164916.

XXII. Yahya, Salah I., et al. "A New Design method for class-E power amplifiers using artificial intelligence modeling for wireless power transfer applications." *Electronics* 11.21 (2022): 3608.

XXIII. Y. S. Mezaal, K. Al-Majdi, A. Al-Hilalli, A. A. Al-Azzawi, and A. A. Almukhtar, "New miniature microstrip antenna for UWB wireless communications," Proceedings of the Estonian Academy of Sciences, vol. 71, no. 2, pp. 194-202, 2022.

XXIV. Y. S. Mezaal , H. A. Hussein, and B. M. Alameri, "Miniaturized microstrip diplexer based on fr4 substrate for wireless communications," *Elektronika Ir Elektrotechnika*, vol. 27, no. 5, pp. 34-40, 2021.