

An Application of Generalized Entropy Optimization Methods in Survival Data Analysis

Aladdin Shamilov¹, Cigdem Kalathilparmbil^{1*}, Sevda Ozdemir²

¹Faculty of Science, Department of Statistics, Anadolu University, Eskişehir, Turkey

²Ozalp Vocational School, Accountancy and Tax Department, Yuzuncu Yil University, Van, Turkey

Email: asamilov@anadolu.edu.tr, *cgiriftinoglu@anadolu.edu.tr, sevdaoazdemir@yyu.edu.tr

How to cite this paper: Shamilov, A., Kalathilparmbil, C. and Ozdemir, S. (2017) An Application of Generalized Entropy Optimization Methods in Survival Data Analysis. *Journal of Modern Physics*, 8, 349-364.

<https://doi.org/10.4236/jmp.2017.83024>

Received: August 5, 2016

Accepted: February 25, 2017

Published: February 28, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this paper, survival data analysis is realized by applying Generalized Entropy Optimization Methods (GEOM). It is known that all statistical distributions can be obtained as MaxEnt distribution by choosing corresponding moment functions. However, Generalized Entropy Optimization Distributions (GEOD) in the form of MinMaxEnt, MaxMaxEnt distributions which are obtained on basis of Shannon measure and supplementary optimization with respect to characterizing moment functions, more exactly represent the given statistical data. For this reason, survival data analysis by GEOD acquires a new significance. In this research, the data of the life table for engine failure data (1980) is examined. The performances of GEOD are established by Chi-Square criteria, Root Mean Square Error (RMSE) criteria and Shannon entropy measure, Kullback-Leibler measure. Comparison of GEOD with each other in the different senses shows that along of these distributions $(\text{MinMaxEnt})_4$ is better in the senses of Shannon measure and of Kullback-Leibler measure. It is showed that, $(\text{MinMaxEnt})_3, ((\text{MaxMaxEnt})_4)$ is more suitable for statistical data among $(\text{MinMaxEnt})_m, m = 1, 2, 3, 4, ((\text{MaxMaxEnt})_m, m = 1, 2, 3, 4)$. Moreover, $(\text{MinMaxEnt})_3$ is better for statistical data than $(\text{MaxMaxEnt})_4$ in the sense of RMSE criteria. According to obtained distribution $(\text{MinMaxEnt})_3, ((\text{MaxMaxEnt})_4)$ estimator of Probability Density Function $\hat{f}(t)$, Cumulative Distribution Function $\hat{F}(t)$, Survival Function $\hat{S}(t)$ and Hazard Rate $\hat{h}(t)$ are evaluated and graphically illustrated. The results are acquired by using statistical software MATLAB.

Keywords

Survival Function, Censored Observation, Generalized Entropy Optimization Methods, MaxEnt, MinMaxEnt, MaxMaxEnt Distributions

1. Introduction

Entropy Optimization Methods (EOM) have important applications, especially in statistics, economy, engineering and so on. There are several examples in the literature that known statistical distributions do not conform to statistical data; however, the entropy optimization distributions conform well. Generalized Entropy Optimization Methods (GEOM) have suggested distributions in the form of MinMaxEnt which is the closest to statistical data, and MaxMaxEnt which is the furthest from mentioned data in the sense of information theory [1] [2], respectively. For this reason, GEOM can be more successfully applied in Survival Data Analysis.

Different aspects and methods of investigations of survival data analysis are considered in [3]-[8].

In particular in the paper [6], it is investigated several problems of hazard rate function estimation based on the maximum entropy principle. The potential applications include developing several classes of the maximum entropy distributions which can be used to model different data-generating distributions that satisfy certain information constraints on the hazard rate function.

In order to represent the results of our investigations, we give some auxiliary concepts and facts first.

2. Survival Analysis

Survival time can be defined broadly as the time to the occurrence of a given event. This event can be the development of a disease, response to a treatment, relapse or death [9].

Censoring: The techniques for reducing experimental time are known as censoring. In survival analysis, the observations are lifetimes, which can be indefinitely long. So quite often the experiment is so designed that the time required for collecting the data is reduced to manageable levels.

Let T be a continuous, non-negative valued random variable representing the lifetime of a unit. This is the time for which an individual (or unit) carries out its appointed task satisfactorily and then passes into “failed” or “dead” state thereafter [10].

The probabilistic properties of the random variable are studied through its cumulative distribution function F or other equivalent functions defined below [9]:

Cumulative Distribution Function:

$$F(t) = P\{T < t\} = \int_0^t f(u)du, \quad 0 < t < \infty$$

Survival Function: This function is denoted by $S(t)$, is defined as the probability that an individual survives longer than t :

$$S(t) = 1 - F(t)$$

$$S(t) = P\{T > t\} = \int_t^{\infty} f(u)du, \quad 0 < t < \infty$$

Probability Density Function: Like any other continuous random variable, the survival time T has a probability density function defined as the limit of the

probability that an individual fails in the short interval $t + \Delta t$ per unit width Δt , or simply the probability of failure in a small interval per unit time. It can be expressed as

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}$$

Hazard Rate: This function is defined as the probability of failure during a very small time interval, assuming that the individual has survived to the beginning of the interval, or as the limit of the probability that an individual fails in a very short interval, $t + \Delta t$, given that the individual has survived to time t :

$$h(t) = \frac{f(t)}{S(t)}.$$

3. Generalized Entropy Optimization Methods (GEOM)

Entropy Optimization Problem (EOP) [11] and Generalized Entropy Optimization problem (GEOP) [10] can be formulated in the following form.

EOP: Let $f^{(0)}(x)$ be given probability density function (p.d.f.) of random variable X , L be an entropy optimization measure and g be a given moment vector function generating m moment constraints. It is required to obtain the distribution corresponding to g , which gives extreme value to L .

GEOP: Let $f^{(0)}(x)$ be given probability density function of random variable X , L be an entropy optimization measure and K be a set of given moment vector functions. It is required to choose moment vector functions $g^{(1)}, g^{(2)} \in K$ such that $g^{(1)}$ defines entropy optimization distribution $f^{(1)}(x)$ closest to $f^{(0)}(x)$, $g^{(2)}$ defines entropy optimization distribution $f^{(2)}(x)$ furthest from $f^{(0)}(x)$ with respect to entropy optimization measure L . If L is taken as Shannon entropy measure, then $f^{(1)}(x)$ is called the MinMaxEnt distribution, and $f^{(2)}(x)$ is called the MaxMaxEnt distribution [1] [2] [12] [13] [14].

The method of solving GEOP is called as GEOM.

3.1. MaxEnt Functional

The problem of maximizing entropy function

$$H(p) = -\sum_{i=1}^n p_i \ln p_i, \quad p = (p_1, \dots, p_n) \quad (1)$$

subject to constraints

$$\sum_{i=1}^n p_i = 1 \quad \sum_{i=1}^n p_i g_j(x_i) = \mu_j, \quad j = 0, 1, 2, \dots, m \quad (2)$$

where $\mu_0 = 1$, $g_0(x) = 1$, $p_i \geq 0$, $i = 1, 2, \dots, n$; $m+1 < n$,

has solution

$$p_i = e^{-\sum_{j=0}^m \lambda_j g_j(x_i)}, \quad i = 1, 2, \dots, n \quad (3)$$

where λ_j ($j = 0, 1, 2, \dots, m$) are Lagrange multipliers. Finding the distribution (p_1, p_2, \dots, p_n) which maximizes function (1) subject to constraints generated by equations in (2) is an optimization problem. In the literature, there have been numerous studies that have calculated these multipliers [1]. In this study, we use

the MATLAB program to calculate Lagrange multipliers.

If (3) is substituted into (1), the maximum entropy value is obtained:

$$H_{\max}(p) = -\sum_{i=1}^n e^{-\sum_{j=0}^m \lambda_j g_j(x_i)} \left(-\sum_{j=0}^m \lambda_j g_j(x_i) \right) = \sum_{j=0}^m \lambda_j \mu_j. \tag{4}$$

If distribution $p^{(0)} = (p_1^{(0)}, \dots, p_n^{(0)})$ is calculated from the data, the moment vector value $\mu = (1, \mu_1, \dots, \mu_m)$ can be obtained for each moment vector function $g(x) = (1, g_1(x), \dots, g_m(x))$. Thus, H_{\max} is considered as a functional of $g(x)$ and called the MaxEnt functional. Therefore, we use the notation $U(g)$ to denote the maximum value of H corresponding to $g(x) = (1, g_1(x), \dots, g_m(x))$.

3.2. MinMaxEnt and MaxMaxEnt Distributions

Let K be the compact set of moment vector functions $g(x)$. $U(g)$ reaches its least and greatest values in this compact set, because of its continuity property. For this reason,

$$\min_{g \in K} U(g) = U(g^{(1)}); \quad \max_{g \in K} U(g) = U(g^{(2)}).$$

Consequently,

$$U(g^{(1)}) \leq U(g^{(2)}).$$

Distributions $p^{(1)} = (p_1^{(1)}, \dots, p_n^{(1)})$ and $p^{(2)} = (p_1^{(2)}, \dots, p_n^{(2)})$ corresponding to the $g^{(1)}(x)$ and $g^{(2)}(x)$, respectively, are called MinMaxEnt and MaxMaxEnt distributions [1].

MinMaxEnt (MaxMaxEnt) method for a finite set of characterizing moment functions can be defined in following form.

Let $K_0 = \{g_1, \dots, g_r\}$ be the set of characterizing moment vector functions and all combinations of r elements of K_0 taken m elements at a time be $K_{0,m}$. We note that, each element of $K_{0,m}$ is vector g with m components.

Solving the MinMaxEnt and the MaxMaxEnt problems require to find vector functions $(g_0, g^{(1)}(x))$, $(g_0, g^{(2)}(x))$, where $g_0(x) = 1$, $g^{(1)} \in K_{0,m}$, $g^{(2)} \in K_{0,m}$ minimizing and maximizing $U(g)$ accordingly with respect to Shannon entropy measure. It should be noted that $U(g)$ reaches its minimum value subject to constraints generated by function $g_0(x)$ and all m -dimensional vector functions $g(x), g \in K_{0,m}$. In other words, minimum value of $U(g)$ is least value of values H_{\max} corresponding to $g(x), g \in K_{0,m}$. If $(g_0, g^{(1)})$ gives the minimum value to $U(g)$ then distribution $p^{(1)} = (p_1^{(1)}, \dots, p_n^{(1)})$ corresponding to $(g_0, g^{(1)})$ is called the MinMaxEnt distribution. MinMaxEnt method represents probability distribution in the form of MinMaxEnt distribution. In a similar way, $U(g)$ reaches its maximum value subject to constraints generated by function $g_0(x)$ and all m -dimensional vector functions $g(x), g \in K_{0,m}$. In other words, maximum value of $U(g)$ is greatest value of values H_{\max} corresponding to $g(x), g \in K_{0,m}$. If $(g_0, g^{(2)})$ gives the maximum value to $U(g)$ then distribution $p^{(2)} = (p_1^{(2)}, \dots, p_n^{(2)})$ corresponding to $(g_0, g^{(2)})$ is called the

MaxMaxEnt distribution. MaxMaxEnt method represents probability distribution in the form of MaxMaxEnt distribution. It should be noted that both distributions can be applied in solving proper problems in survival data analysis.

4. Application of MinMaxEnt and MaxMaxEnt Methods to Survival Data

4.1. MinMaxEnt and MaxMaxEnt Distributions for Finite Set of Characterizing Moment Functions

In the present research, the data of the life table for engine failure data (1980) given in **Table 1** is considered [10].

In our investigation, the experiment is planned for 200 numbers of patients surviving at beginning of interval but the presence of censoring from the planning patients 97 individuals stay out the experiment. This situation is taken into account in **Table 2**.

It should be noted that, the presence of censoring in the survival times leads to a situation where the sum of observation probabilities stands less than 1 for the

Table 1. The data of the life table for engine failure data (1980).

Survival Time (year) t	Working at the beginning of interval n_i	Failed during the interval d_i	Censored during the interval c_i
0 - 1	200	5	0
1 - 2	195	10	1
2 - 3	184	12	5
3 - 4	167	8	2
4 - 5	157	10	0
5 - 6	147	15	6
6 - 7	126	9	3
7 - 8	114	8	1
8 - 9	105	4	0
9 - 10	101	3	1

Table 2. Observed and corrected probabilities.

t	n_i	d_i	c_i	Observed probabilities p_i	Corrected probabilities p_i^*
0 - 1	200	5	0	0.0485	0.0485
1 - 2	195	10	1	0.0971	0.1068
2 - 3	184	12	5	0.1165	0.1650
3 - 4	167	8	2	0.0777	0.0971
4 - 5	157	10	0	0.0971	0.0971
5 - 6	147	15	6	0.1456	0.2039
6 - 7	126	9	3	0.0874	0.1165
7 - 8	114	8	1	0.0777	0.0874
8 - 9	105	4	0	0.0388	0.0388
9 - 10	101	3	1	0.0291	0.0388

survival data. For this reason, in solving many problems, it is required to supplement the sum of observation probabilities up to 1. Since the sum of observed probabilities p_i in **Table 2** is 0.8155, according to the number of censoring, supplementary probability $1 - 0.8155 = 0.1845$ is uniformly distributed to each censoring data and corrected probabilities p_i^* are obtained.

As we noted that above, MinMaxEnt and MaxMaxEnt distributions can be applied in solving proper problems in survival data analysis. In our investigation as components of K_0 characterizing moment vector function

$g_1(x) = x, g_2(x) = x^2, g_3(x) = \ln x, g_4(x) = (\ln x)^2, g_5(x) = \ln(1 + x^2)$ are chosen. The set of moment functions is chosen from the characteristic moments which are mostly used in Statistics.

Consequently, $K_0 = \{g_1, \dots, g_5\}$. For example, if $m = 3$ then

$$(g_0, g^{(1)}) = (1, x, x^2, (\ln x)^2), \quad g^{(1)} \in K_{0,3}$$

gives the least value to $U(g)$ and

$$(g_0, g^{(2)}) = (1, x^2, \ln x, \ln(1 + x^2)), \quad g^{(2)} \in K_{0,3}$$

gives the greatest value to $U(g)$.

The MaxEnt distributions corresponding to $(g_0, g), g_0(x) = 1, g \in K_{0,m}, m = 1, 2, \dots, 4$ and H_{\max} values are shown in **Tables 3-6**. In these tables, MinMaxEnt and MaxMaxEnt distributions corresponding to H_{MinMax} and H_{MaxMax} are represented with bold font. By virtue of these tables are also obtained $(\text{MinMaxEnt})_m, (\text{MaxMaxEnt})_m, m = 1, 2, 3, 4$ distributions which are shown in **Table 7** and **Table 8**.

In order to obtain the performance of the mentioned distributions, we use various criteria as Root Mean Square Error (RMSE), Chi-Square, entropy values of distributions. The acquired results are demonstrated in **Table 9** and **Table 10**.

All $(\text{MinMaxEnt})_m, (\text{MaxMaxEnt})_m, m = 1, 2, 3, 4$ distributions are acceptable to survival data in the sense of Chi - Square criteria.

In the sense of RMSE criteria each $(\text{MinMaxEnt})_m (m = 1, 2, 3, 4)$ distribution is better than corresponding $(\text{MaxMaxEnt})_m (m = 1, 2, 3, 4)$ distribution. Moreover, $(\text{MinMaxEnt})_1$ is nearer to statistical data than $(\text{MaxMaxEnt})_1$ and

Table 3. The predicted probabilities for the MaxEnt distribution corresponding to $(g_0, g), g_0(x) = 1, g \in K_{0,1}$ and H_{\max} values.

(g_0, g)	(g_0, g_1)	(g_0, g_2)	(g_0, g_3)	(g_0, g_4)	(g_0, g_5)
H_{\max}	3.3084	3.2854	3.3219	3.3040	3.3204
$p^{(0)}$ for MaxEnt Distribution	0.1229	0.1269	0.0989	0.1202	0.1093
	0.1171	0.1249	0.0995	0.1238	0.1058
	0.1116	0.1210	0.0998	0.1161	0.1030
	0.1064	0.1153	0.1000	0.1083	0.1010
	0.1015	0.1081	0.1001	0.1014	0.0994
	0.0967	0.0997	0.1002	0.0954	0.0981
	0.0922	0.0905	0.1003	0.0901	0.0971
	0.0879	0.0809	0.1004	0.0855	0.0962
	0.0838	0.0711	0.1004	0.0814	0.0954
	0.0799	0.0615	0.1005	0.0777	0.0947

Table 4. The predicted probabilities for the MaxEnt distribution corresponding to $(g_0, g), g_0(x) = 1, g \in K_{0,2}$ and H_{\max} values.

(g_0, g)	(g_0, g_1, g_2)	(g_0, g_1, g_3)	(g_0, g_1, g_4)	(g_0, g_1, g_5)	(g_0, g_2, g_3)
H_{\max}	3.2042	3.2140	3.2921	3.2106	3.2041
$p^{(0)}$ for MaxEnt Distribution	0.0619	0.0429	0.0863	0.0543	0.0483
	0.0921	0.1137	0.1449	0.0939	0.1036
	0.1218	0.1451	0.1299	0.1351	0.1363
	0.1434	0.1490	0.1126	0.1529	0.1493
	0.1502	0.1377	0.0999	0.1480	0.1456
	0.1400	0.1194	0.0914	0.1290	0.1296
	0.1160	0.0994	0.0861	0.1046	0.1068
	0.0856	0.0803	0.0833	0.0804	0.0820
	0.0562	0.0634	0.0824	0.0593	0.0589
	0.0328	0.0492	0.0832	0.0424	0.0397
(g_0, g)	(g_0, g_2, g_4)	(g_0, g_2, g_5)	(g_0, g_3, g_4)	(g_0, g_3, g_5)	(g_0, g_4, g_5)
H_{\max}	3.2408	3.2057	3.2305	3.2500	3.2237
$p^{(0)}$ for MaxEnt Distribution	0.1101	0.0579	0.0400	0.0423	0.0403
	0.0825	0.0928	0.1294	0.1425	0.1205
	0.1058	0.1281	0.1481	0.1432	0.1478
	0.1285	0.1484	0.1403	0.1278	0.1454
	0.1396	0.1499	0.1251	0.1134	0.1310
	0.1344	0.1353	0.1091	0.1018	0.1133
	0.1150	0.1108	0.0944	0.0924	0.0961
	0.0877	0.0830	0.0816	0.0848	0.0809
	0.0598	0.0573	0.0706	0.0785	0.0679
	0.0366	0.0365	0.0613	0.0733	0.0570

Table 5. The predicted probabilities for the MaxEnt distribution corresponding to $(g_0, g), g_0(x) = 1, g \in K_{0,3}$ and H_{\max} values.

(g_0, g)	(g_0, g_1, g_2, g_3)	(g_0, g_1, g_2, g_4)	(g_0, g_1, g_2, g_5)	(g_0, g_1, g_3, g_4)	(g_0, g_1, g_3, g_5)
H_{\max}	3.2024	3.2000	3.2042	3.2083	3.2100
$p^{(0)}$ for MaxEnt Distribution	0.0537	0.0489	0.0625	0.0515	0.0503
	0.0972	0.1034	0.0920	0.0974	0.0988
	0.1291	0.1314	0.1211	0.1359	0.1382
	0.1471	0.1458	0.1427	0.1519	0.1526
	0.1488	0.1466	0.1501	0.1472	0.1458
	0.1355	0.1342	0.1405	0.1290	0.1268
	0.1118	0.1117	0.1167	0.1051	0.1033
	0.0838	0.0844	0.0859	0.0808	0.0802
	0.0573	0.0578	0.0561	0.0593	0.0601
	0.0357	0.0358	0.0324	0.0420	0.0438
(g_0, g)	(g_0, g_2, g_3, g_5)	(g_0, g_2, g_3, g_4)	(g_0, g_2, g_3, g_5)	(g_0, g_2, g_3, g_5)	(g_0, g_2, g_3, g_5)
H_{\max}	3.2104	3.2035	3.2039	3.2050	3.2155
$p^{(0)}$ for MaxEnt Distribution	0.0522	0.0518	0.0502	0.0533	0.0501
	0.0964	0.0987	0.1007	0.0968	0.0986
	0.1369	0.1322	0.1343	0.1324	0.1416
	0.1530	0.1487	0.1493	0.1497	0.1543
	0.1468	0.1480	0.1469	0.1483	0.1436
	0.1277	0.1330	0.1313	0.1325	0.1227
	0.1038	0.1093	0.1079	0.1085	0.0999
	0.0802	0.0827	0.0822	0.0822	0.0792
	0.0598	0.0579	0.0584	0.0580	0.0619
	0.0433	0.0376	0.0388	0.0383	0.0480

Table 6. The predicted probabilities for the MaxEnt distribution corresponding to $(g_0, g), g_0(x) = 1, g \in K_{0,4}$ and H_{\max} values.

	(g_0, g)	$(g_0, g_1, g_2, g_3, g_4)$	$(g_0, g_1, g_2, g_3, g_5)$	$(g_0, g_1, g_2, g_4, g_5)$	$(g_0, g_1, g_3, g_4, g_5)$	$(g_0, g_2, g_3, g_4, g_5)$
H_{\max}	3.1937	3.1935	3.1936	3.1932		3.1934
$p^{(0)}$ for MaxEnt Distribution	0.0477	0.0476	0.0477	0.0475		0.0476
	0.1198	0.1203	0.1201	0.1217		0.1209
	0.1221	0.1216	0.1218	0.1192		0.1207
	0.1306	0.1302	0.1303	0.1292		0.1298
	0.1397	0.1402	0.1400	0.1422		0.1408
	0.1395	0.1400	0.1398	0.1421		0.1407
	0.1230	0.1230	0.1230	0.1225		0.1229
	0.0921	0.0917	0.0919	0.0898		0.0911
	0.0570	0.0568	0.0568	0.0560		0.0565
	0.0284	0.0287	0.0286	0.0298		0.0290

Table 7. Distributions of $(\text{MinMaxEnt})_m, m = 1, 2, 3, 4$.

t	d_i	c_i	p_i^*	$(\text{MinMaxEnt})_1$	$(\text{MinMaxEnt})_2$	$(\text{MinMaxEnt})_3$	$(\text{MinMaxEnt})_4$
1	5	0	0.0485	0.1269	0.0483	0.0489	0.0475
2	10	1	0.1068	0.1249	0.1036	0.1034	0.1217
3	12	5	0.1650	0.1210	0.1363	0.1314	0.1192
4	8	2	0.0971	0.1153	0.1493	0.1458	0.1292
5	10	0	0.0971	0.1081	0.1456	0.1466	0.1422
6	15	6	0.2039	0.0997	0.1296	0.1342	0.1421
7	9	3	0.1165	0.0905	0.1068	0.1117	0.1225
8	8	1	0.0874	0.0809	0.0820	0.0844	0.0898
9	4	0	0.0388	0.0711	0.0589	0.0578	0.0560
10	3	1	0.0388	0.0615	0.0397	0.0358	0.0298

Table 8. Distributions of $(\text{MaxMaxEnt})_m, m = 1, 2, 3, 4$.

t	d_i	c_i	p_i^*	$(\text{MaxMaxEnt})_1$	$(\text{MaxMaxEnt})_2$	$(\text{MaxMaxEnt})_3$	$(\text{MaxMaxEnt})_4$
1	5	0	0.0485	0.0989	0.0863	0.0501	0.0477
2	10	1	0.1068	0.0995	0.1449	0.0986	0.1198
3	12	5	0.1650	0.0998	0.1299	0.1416	0.1221
4	8	2	0.0971	0.1000	0.1126	0.1543	0.1306
5	10	0	0.0971	0.1001	0.0999	0.1436	0.1397
6	15	6	0.2039	0.1002	0.0914	0.1227	0.1395
7	9	3	0.1165	0.1003	0.0861	0.0999	0.1230
8	8	1	0.0874	0.1004	0.0833	0.0792	0.0921
9	4	0	0.0388	0.1004	0.0824	0.0619	0.0570
10	3	1	0.0388	0.1005	0.0832	0.0480	0.0284

$(\text{MaxMaxEnt})_2$ distributions; each $(\text{MinMaxEnt})_m (m = 2, 3, 4)$ is better than all of $(\text{MaxMaxEnt})_m (m = 1, 2, 3, 4)$ distributions. From these results follows that among of distributions $(\text{MinMaxEnt})_m, (m = 1, 2, 3, 4)$ the distribution

$(\text{MinMaxEnt})_3$ is more suitable and among of distributions $(\text{MaxMaxEnt})_m$, $(m = 1, 2, 3, 4)$ the distribution $(\text{MaxMaxEnt})_4$ is more convenient for statistical data. These results are also corroborated by graphical representation (see **Figures 1-4**). Consequently, we shall consider Probability Density Function $\hat{f}(t)$, Cumulative Distribution Function $\hat{F}(t)$, Survival Function $\hat{S}(t)$ and Hazard Rate $\hat{h}(t)$ for only $(\text{MinMaxEnt})_3$ and $(\text{MaxMaxEnt})_4$ distributions.

Although the distribution with the largest number of moment functions tends to fit better, it should be noted that in some cases, the set of moment functions with fewer elements is more informative than a different set of moment functions with more number of elements.

Table 9. The obtained results for $(\text{MinMaxEnt})_m$, $m = 1, 2, 3, 4$.

Distribution of MinMaxEnt	H	Calculated value of Chi-Square	Probability of Chi-Square value	Table value of Chi-Square	RMSE
$(\text{MinMaxEnt})_1$	3.2854	4.4310	0.8163	$\chi_{8,\alpha}^2 = 15.51$	0.3158
$(\text{MinMaxEnt})_2$	3.2041	0.6512	0.9987	$\chi_{7,\alpha}^2 = 14.07$	0.1873
$(\text{MinMaxEnt})_3$	3.2000	1.7787	0.9389	$\chi_{6,\alpha}^2 = 12.59$	0.1799
$(\text{MinMaxEnt})_4$	3.1932	1.6161	0.8993	$\chi_{5,\alpha}^2 = 11.07$	0.1830

Table 10. The obtained results for $(\text{MaxMaxEnt})_m$, $m = 1, 2, 3, 4$.

Distribution of MinMaxEnt	H	Calculated value of Chi-Square	Probability of Chi-Square value	Table value of Chi-Square	RMSE
$(\text{MaxMaxEnt})_1$	3.3219	5.3820	0.7161	$\chi_{8,\alpha}^2 = 15.51$	0.3492
$(\text{MaxMaxEnt})_2$	3.2921	4.9233	0.6693	$\chi_{7,\alpha}^2 = 14.07$	0.3492
$(\text{MaxMaxEnt})_3$	3.2155	2.2804	0.8922	$\chi_{6,\alpha}^2 = 12.59$	0.2104
$(\text{MaxMaxEnt})_4$	3.1937	1.6383	0.8966	$\chi_{5,\alpha}^2 = 11.07$	0.1888

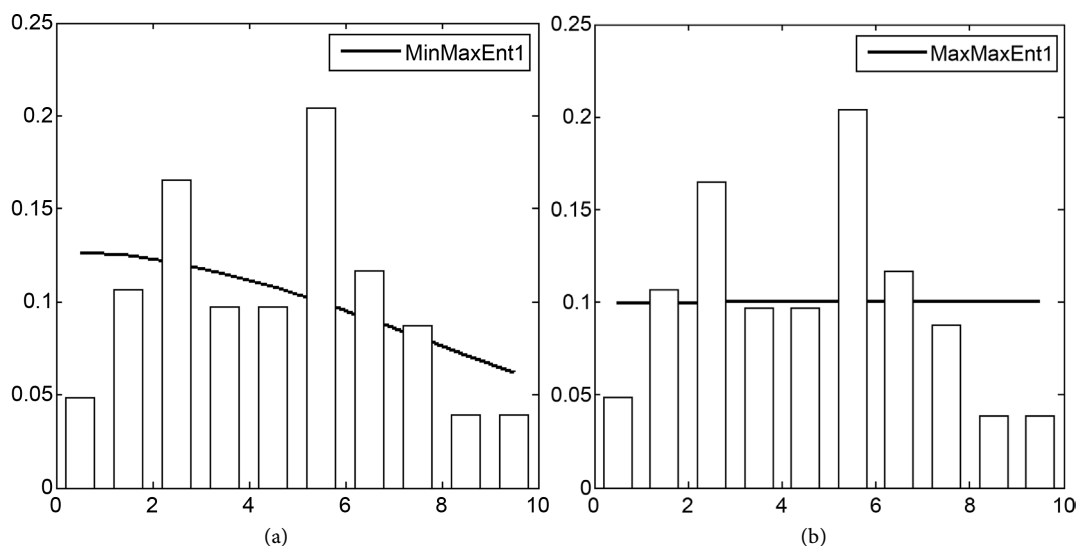


Figure 1. Graphic of $(\text{MinMaxEnt})_1$ and $(\text{MaxMaxEnt})_1$ distributions.

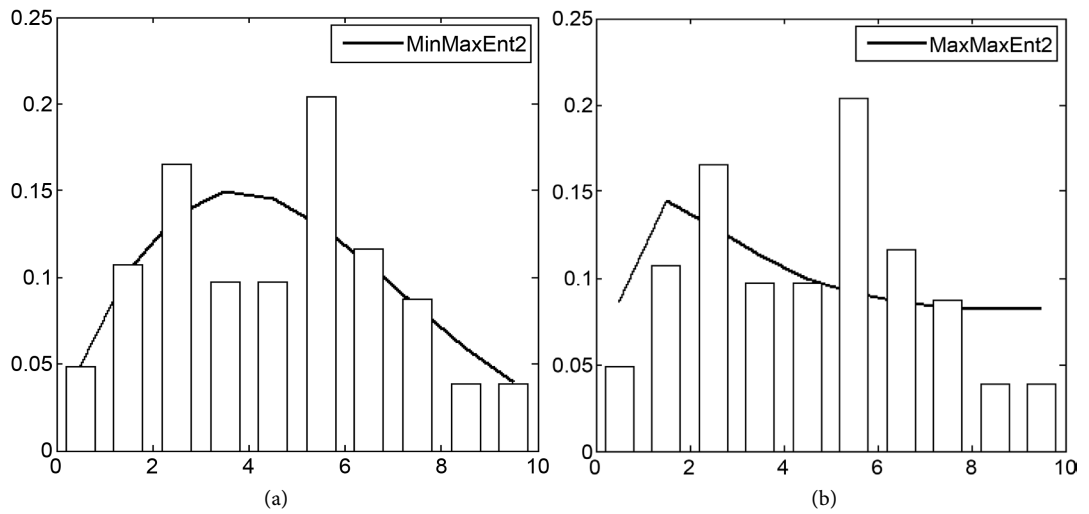


Figure 2. Graphic of $(\text{MinMaxEnt})_2$ and $(\text{MaxMaxEnt})_2$ distributions.

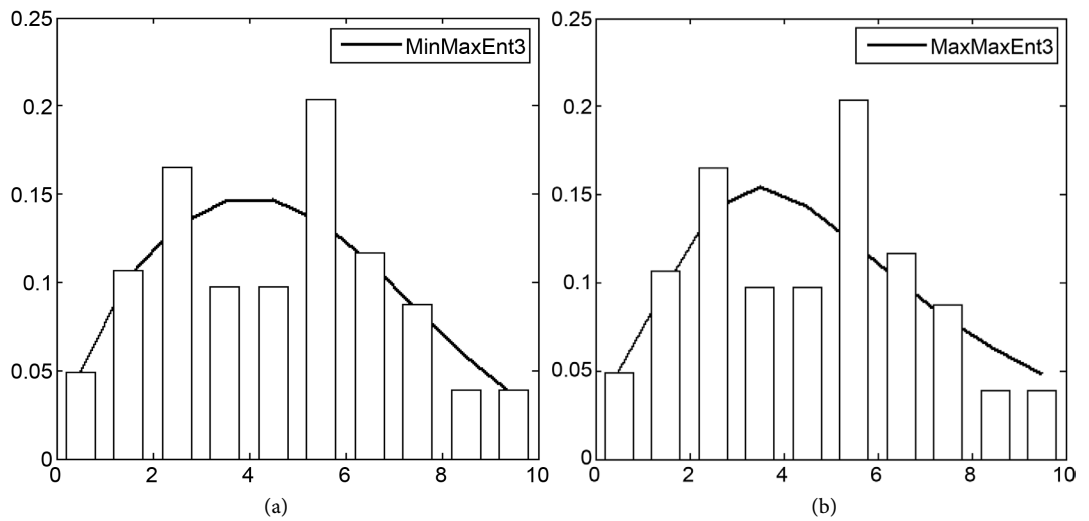


Figure 3. Graphic of $(\text{MinMaxEnt})_3$ and $(\text{MaxMaxEnt})_3$ distributions.

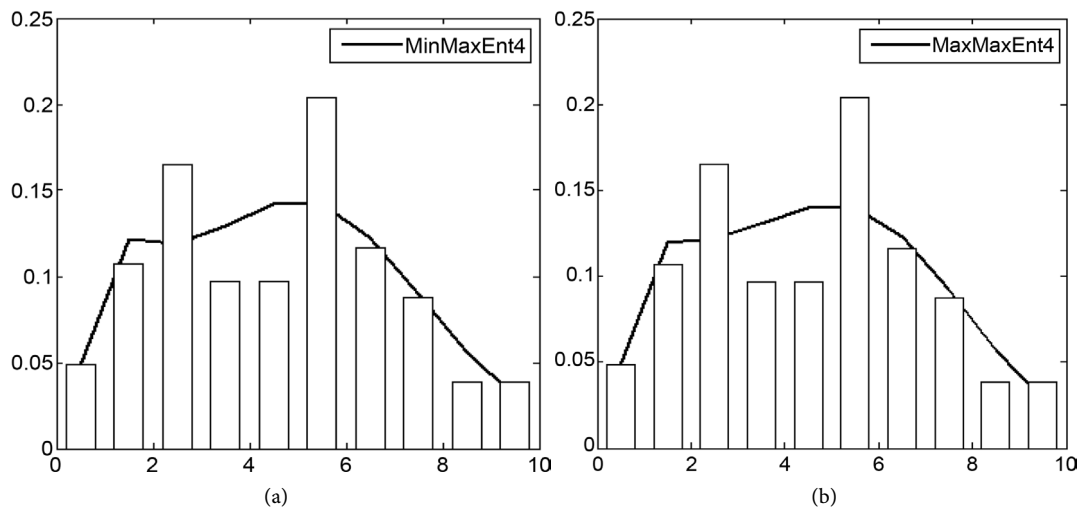


Figure 4. Graph of $(\text{MinMaxEnt})_4$ and $(\text{MaxMaxEnt})_4$ distributions.

4.2. Availability of GEOD to Survival Data in the Sense of Shannon Measure

In order to establish availability of GEOD to survival data in the sense of Shannon measure it is required to consider entropy values of GEOD.

From **Table 3** it is seen that the MinMaxEnt (the MaxMaxEnt) distribution is realized by vector function $(g_0, g_2) = (1, x^2)$ ($(g_0, g_3) = (1, \ln x)$) and $H((\text{MinMaxEnt})_1) = 3.2854$ ($H((\text{MaxMaxEnt})_1) = 3.3219$).

From **Table 4** it is seen that the MinMaxEnt (the MaxMaxEnt) distribution is realized by vector function

$(g_0, g_2, g_3) = (1, x^2, \ln x)$ ($(g_0, g_1, g_4) = (1, x, (\ln x)^2)$) and $H((\text{MinMaxEnt})_2) = 3.2041$ ($H((\text{MaxMaxEnt})_2) = 3.2921$).

From **Table 5** it is seen that the MinMaxEnt (the MaxMaxEnt) distribution is realized by vector function

$(g_0, g_1, g_2, g_4) = (1, x, x^2, (\ln x)^2)$ ($(g_0, g_2, g_3, g_5) = (1, x^2, \ln x, \ln(1+x^2))$) and $H((\text{MinMaxEnt})_3) = 3.2000$ ($H((\text{MaxMaxEnt})_3) = 3.2155$).

From **Table 6** it is seen that the MinMaxEnt (the MaxMaxEnt) distribution is realized by vector function

$(g_0, g_1, g_3, g_4, g_5) = (1, x, \ln x, (\ln x)^2, \ln(1+x^2))$ ($(g_0, g_1, g_2, g_3, g_4) = (1, x, x^2, \ln x, (\ln x)^2)$) and $H((\text{MinMaxEnt})_4) = 3.1932$ ($H((\text{MaxMaxEnt})_4) = 3.1937$).

Comparison of GEOD with each other in the sense of Shannon measure shows that along of these distributions $(\text{MinMaxEnt})_4$ is better.

The results of our investigation according to using known characterizing moment vector functions from $K_{0,m}$ are summarised in the form of following Corollary.

Corollary 1. *If by $(\text{MaxMaxEnt})_m$ ($(\text{MinMaxEnt})_m$) denote the MaxMaxEnt (the MinMaxEnt) distribution corresponding to m moment conditions generated by moment functions $g(x), g \in K_{0,m}$, then inequality*

$$\begin{aligned} H((\text{MaxMaxEnt})_{m_1}) &> H((\text{MaxMaxEnt})_{m_2}) \\ H((\text{MinMaxEnt})_{m_1}) &> H((\text{MinMaxEnt})_{m_2}) \end{aligned}$$

is fulfilled, when $m_1 < m_2$. In other words, entropy value of the MaxMaxEnt (the MinMaxEnt) distribution depending on the number m of moment conditions decreases.

Moreover for any m the inequality

$$H((\text{MaxMaxEnt})_m) > H((\text{MinMaxEnt})_m)$$

takes place.

4.3. Availability of GEOD to Survival Data in the Sense of Kullback-Leibler Measure

Now, we calculate the distance between observed distribution

$p_i = (p_1, p_2, \dots, p_{10})$ given in **Table 2** and distributions

$p_{\min}^{(i)} = (\text{MinMaxEnt})_i$, $p_{\max}^{(i)} = (\text{MaxMaxEnt})_i$, $i = 1, 2, 3, 4$ given in **Table 7** and **Table 8** respectively.

It is known that the Kullback – Leibler distance between distributions $p = (p_1, p_2, \dots, p_n)$ and $q = (q_1, q_2, \dots, q_n)$ is obtained by formula

$$D(p; q) = \sum_{i=1}^n p_i \ln \frac{p_i}{q_i}.$$

By starting these formula Kullback-Leibler measures for the distance between observed distribution $p_i = (p_1, p_2, \dots, p_{10})$ and distributions $p_{\min}^{(i)} = (\text{MinMaxEnt})_i, p_{\max}^{(i)} = (\text{MaxMaxEnt})_i, i = 1, 2, 3, 4$ are given in **Table 11** and **Table 12** respectively.

From **Table 11** and **Table 12** follows that along of GEOD $(\text{MinMaxEnt})_4$ is better in the sense of Kullback-Leibler measure.

The results of our investigation according to using known characterizing moment vector functions from $K_{0,m}$ are summarised in the form of following Corollary.

Corollary 2. If $p_i = (p_1, p_2, \dots, p_{10})$ observed distribution and $p_{\min}^{(i)} = (\text{MinMaxEnt})_i, i = 1, 2, 3, 4$ ($p_{\max}^{(i)} = (\text{MaxMaxEnt})_i, i = 1, 2, 3, 4$) denote the MinMaxEnt (the MaxMaxEnt) distribution corresponding to i moment conditions generated by moment functions $g(x), g \in K_{0,m}$, then inequality

$$D(p_{\min}^{(m_1)}; p_i) > D(p_{\min}^{(m_2)}; p_i) \\ \left(D(p_{\max}^{(m_1)}; p_i) > D(p_{\max}^{(m_2)}; p_i) \right)$$

is fulfilled, when $m_1 < m_2$. In other words, Kullback-Leibler value of the MinMaxEnt (the MaxMaxEnt) distribution depending on the number m of moment conditions decreases.

Moreover for any i the inequality

$$D(p_{\max}^{(i)}; p_i) > D(p_{\min}^{(i)}; p_i)$$

takes place.

Table 11. Kullback-Leibler measure of $(\text{MinMaxEnt})_m, i = 1, 2, 3, 4$ distributions.

$(\text{MinMaxEnt})_m, i = 1, 2, 3, 4$ distributions	$D(p_{\min}^{(i)}; p_i), i = 1, 2, 3, 4$
$(\text{MinMaxEnt})_1$	0.3938
$(\text{MinMaxEnt})_2$	0.3348
$(\text{MinMaxEnt})_3$	0.3300
$(\text{MinMaxEnt})_4$	0.3193

Table 12. Kullback-Leibler measure of $(\text{MaxMaxEnt})_m, i = 1, 2, 3, 4$ distributions.

$(\text{MaxMaxEnt})_m, i = 1, 2, 3, 4$ distributions	$D(p_{\max}^{(i)}; p_i), i = 1, 2, 3, 4$
$(\text{MaxMaxEnt})_1$	0.4441
$(\text{MaxMaxEnt})_2$	0.4009
$(\text{MaxMaxEnt})_3$	0.3457
$(\text{MaxMaxEnt})_4$	0.3198

4.4. Survival Expression of Distributions $(\text{MinMaxEnt})_3$, $(\text{MaxMaxEnt})_4$

In this section survival data analysis is conducted by $(\text{MinMaxEnt})_3$, $(\text{MaxMaxEnt})_4$ distribution since the above acquired investigations $(\text{MinMaxEnt})_3$, $(\text{MaxMaxEnt})_4$ is more presentable for survival data among $(\text{MinMaxEnt})_m$, $(\text{MaxMaxEnt})_m$, $m = 1, 2, \dots, 4$ distributions.

$(\text{MinMaxEnt})_3$ and $(\text{MaxMaxEnt})_4$ estimations of Probability Density Function $\hat{f}(t)$, Cumulative Distribution Function $\hat{F}(t)$, Survival Function $\hat{S}(t)$ and Hazard Rate $\hat{h}(t)$ are given in **Table 13** & **Table 14**, respectively.

On basis of the results given in **Table 13** & **Table 14**, graphs of $\hat{f}(t)$, $\hat{F}(t)$, $\hat{S}(t)$ and $\hat{h}(t)$ are demonstrated in **Figures 5(a)-(c)** & **Figures 6(a)-(c)**.

Table 13. Survival analysis by $(\text{MinMaxEnt})_3$.

t	n_i	d_i	c_i	$\hat{f}(t) = (\text{MinMaxEnt})_3$	$\hat{F}(t)$	$\hat{S}(t)$	$\hat{h}(t) = \frac{\hat{f}(t)}{\hat{S}(t)}$
1	200	5	0	0.0489	0.0489	0.9511	0.0514
2	195	10	1	0.1034	0.1523	0.8477	0.1220
3	184	12	5	0.1314	0.2837	0.7163	0.1834
4	167	8	2	0.1458	0.4295	0.5705	0.2556
5	157	10	0	0.1466	0.5761	0.4239	0.3458
6	147	15	6	0.1342	0.7103	0.2897	0.4632
7	126	9	3	0.1117	0.8220	0.1780	0.6275
8	114	8	1	0.0844	0.9064	0.0936	0.9017
9	105	4	0	0.0578	0.9642	0.0358	1.6145
10	101	3	1	0.0358	1.0000	0.0000	--

Table 14. Survival analysis by $(\text{MaxMaxEnt})_4$.

t	n_i	d_i	c_i	$\hat{f}(t) = (\text{MaxMaxEnt})_4$	$\hat{F}(t)$	$\hat{S}(t)$	$\hat{h}(t) = \frac{\hat{f}(t)}{\hat{S}(t)}$
1	200	5	0	0.0477	0.0477	0.9523	0.0501
2	195	10	1	0.1198	0.1675	0.8325	0.1439
3	184	12	5	0.1221	0.2896	0.7104	0.1719
4	167	8	2	0.1306	0.4202	0.5798	0.2253
5	157	10	0	0.1397	0.5599	0.4401	0.3174
6	147	15	6	0.1395	0.6994	0.3006	0.4641
7	126	9	3	0.1230	0.8224	0.1776	0.6926
8	114	8	1	0.0921	0.9145	0.0855	1.0772
9	105	4	0	0.0570	0.9715	0.0285	2.0000
10	101	3	1	0.0284	0.9999	0.0001	--

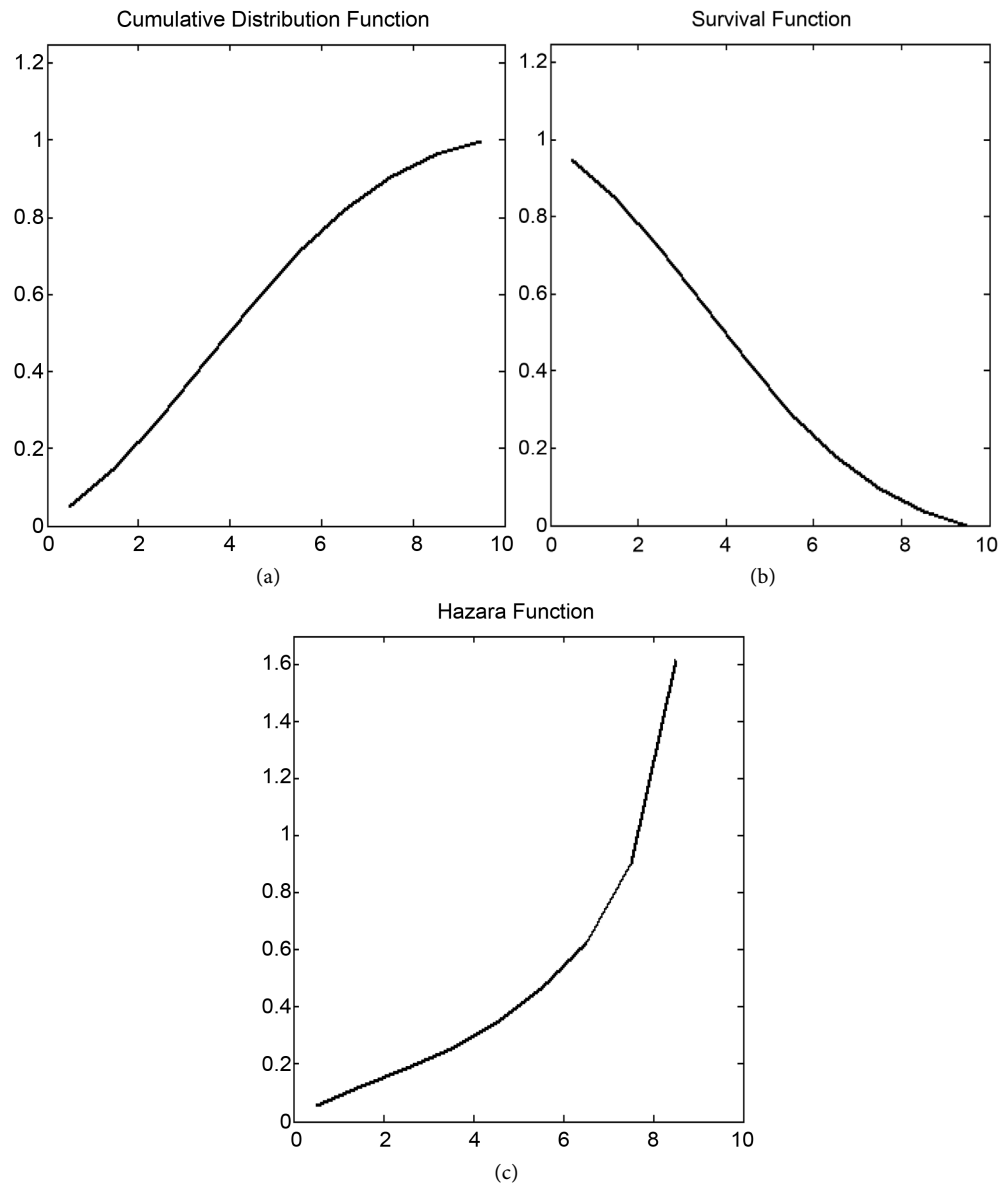


Figure 5. Survival expression of distribution $(\text{MinMaxEnt})_3$.

5. Conclusion

In this study, it is established that survival data analysis is realized by applying Generalized Entropy Optimization Methods (GEOM). Generalized Entropy Optimization Distributions (GEOD) in the form of MinMaxEnt, MaxMaxEnt distributions which are obtained on basis of Shannon measure and supplementary optimization with respect to characterizing moment functions, more exactly represent the given statistical data. For this reason, survival data analysis by GEOD acquires a new significance. The performances of GEOD are established by Chi-Square criteria, Root Mean Square Error (RMSE) criteria and Shannon entropy measure, Kullback-Leibler measure. Comparison of GEOD with each other in the different senses shows that along of these distributions $(\text{MinMaxEnt})_4$ is better in the senses of Shannon measure and of Kullback-Leibler measure. It

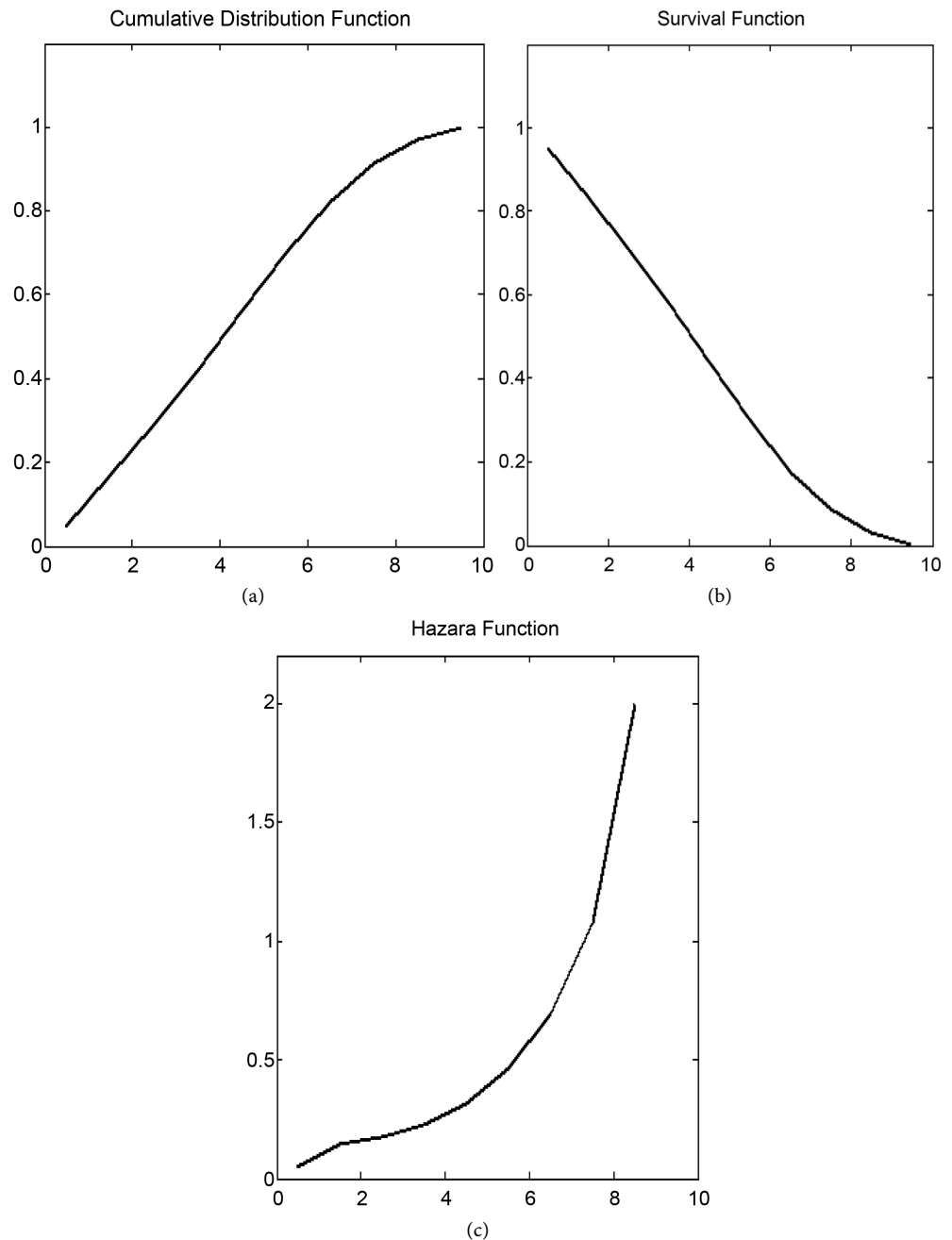


Figure 6. Survival expression of distribution $(\text{MaxMaxEnt})_4$.

is showed that, $(\text{MinMaxEnt})_3((\text{MaxMaxEnt})_4)$ is more suitable for statistical data among $(\text{MinMaxEnt})_m, m = 1, 2, 3, 4((\text{MaxMaxEnt})_m, m = 1, 2, 3, 4)$. Moreover, $(\text{MinMaxEnt})_3$ is better for statistical data than $(\text{MaxMaxEnt})_4$ in the sense of RMSE criteria. According to obtained distribution $(\text{MinMaxEnt})_3((\text{MaxMaxEnt})_4)$ estimator of Probability Density Function $\hat{f}(t)$, Cumulative Distribution Function $\hat{F}(t)$, Survival Function $\hat{S}(t)$ and Hazard Rate $\hat{h}(t)$ are evaluated and graphically illustrated. These results are also corroborated by graphical representation. Our investigation indicates that GEOM in survival data analysis yields reasonable results.

References

- [1] Shamilov, A. (2006) A Development of Entropy Optimization Methods. *Wseas Transactions on Mathematics*, **5**, 568-575.
- [2] Shamilov, A. (2007) Generalized Entropy Optimization Problems and the Existence of Their Solutions. *Physica A: Statistical Mechanics and Its Applications*, **382**, 465-472. <https://doi.org/10.1016/j.physa.2007.04.014>
- [3] Kaminski, D. and Geisler, C. (2012) Survival Analysis of Faculty Retention in Science and Engineering by Gender. *Science*, **335**, 864-866. <https://doi.org/10.1126/science.1214844>
- [4] Reingold, E.M., Reichle, E.D. and Glaholt, M.G. (2012) Heather Sheridan, Direct Lexical Control of Eye Movements in Reading: Evidence from a Survival Analysis of Fixation Durations. *Cognitive Psychology*, **65**, 177-206.
- [5] Wang, H. and Dai, H.S. (2012) Accelerated Failure Time Models for Censored Survival Data under Referral Bias. *Biostatistics*, **14**, 313-326.
- [6] Ebrahimi, N. (2000) The Maximum Entropy Method for Lifetime Distributions. *Sankhyā: The Indian Journal of Statistics, Series A*, **62**, 236-243.
- [7] Guyot, P., Ades, A., Ouwens, M.J. and Welton, N.J. (2012) Enhanced Secondary Analysis of Survival Data: Reconstructing the Data from Published Kaplan-Meier Survival Curves. *BMC Medical Research Methodology*, **12**, 9. <https://doi.org/10.1186/1471-2288-12-9>
- [8] Joly, P., Gerds, T.A., Qvist, V., Commenges, D. and Keiding, N. (2012) Estimating Survival of Dental Fillings on the Basis of Interval-Censored Data and Multi-State Models. *Statistics in Medicine*, **31**, 11-12.
- [9] Lee, E.T. and Wang, J.W. (2003) *Statistical Methods for Survival Data Analysis*. Wiley-Interscience, Oklahoma.
- [10] Deshpande, J.V. and Purohit, S.G. (2005) *Life Time Data: Statistical Models and Methods*, Series on Quality. Vol. 11, Reliability and Engineering Statistics, India.
- [11] Kapur, J.N. (1992) *Kesavan, Entropy Optimization Principles with Applications*.
- [12] Shamilov, A. (2009) *Entropy, Information and Entropy Optimization*. T.C. Anadolu University Publication, Eskisehir.
- [13] Shamilov, A. (2010) Generalized Entropy Optimization Problems with Finite Moment Functions Sets. *Journal of Statistics and Management Systems*, **13**, 595-603. <https://doi.org/10.1080/09720510.2010.10701489>
- [14] Shamilov, A., Giriftinoglu, C., Usta, I. and Mert Kantar, Y. (2008) A New Concept of Relative Suitability of Moment Function Sets. *Applied Mathematics and Computation*, **206**, 521-529. <https://doi.org/10.1016/j.amc.2008.05.063>

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact jmp@scirp.org