Applied Mathematics & Information Sciences
An International Journal

Estimation of the Weibull parameters by Kullback-Leibler divergence of Survival functions

Gholamhossein Yari[1], *Alireza Mirhabibi*[2] *and Abolfazl Saghafi*[3,4]

[1] School of Mathematics, Iran University of Science and Technology, Tehran, Iran
[2] Center of Excellence for Advanced Materials and Processing, Iran University of Science and Technology, Tehran, Iran
[3] School of Mathematics, Iran University of Science and Technology, Tehran, Iran
[4] Department of Mathematics, University of Patras, Patras, Greece

Received: 15 Jan. 2012; Revised 12 Jun. 2012; Accepted 17 Jul. 2012
Published online: 1 Jan. 2013

Abstract: Recently, a new entropy based divergence measure has been introduced which is much like Kullback-Leibler divergence. This entropy measures the distance between an empirical and a prescribed survival function and is a lot easier to compute in continuous distributions than the K-L divergence. In this paper we show that this distance converges to zero with increasing sample size and we apply it to estimate Weibull parameters. Detailed simulations show a higher performance of the new estimation method than the commonly used maximum likelihood and linear regression methods in Weibull scale parameter estimation. Using unbiasing factors provided in this paper for Weibull shape parameter estimation, one can obtain unbiased estimation for Weibull modulus.

Keywords: Kullback-Leibler divergence, Survival function, Weibull distribution, Reliability, Simulation.

1. Introduction

Since the Kullback-Leibler divergence has been defined as a measure of distance between two probability distributions[1], it has been extensively utilized in numerous applications of science and engineering. Although, it is linked with statistical subjects such as model selection and parameter estimation, it has been applied in many other analytical and experimental concepts. The reader is recommended to see [2,3] for a list of applications.

However, there are some limitations and difficulties in utilizing the K-L divergence for continuous distributions. For two continuous random variables X and Y with probability density functions f and g respectively, the K-L divergence of f relative to g is defined by:

$$D(f||g) = \int_{\mathbb{R}} f(x) \ln \frac{f(x)}{g(x)}\, dx. \qquad (1)$$

This definition is based on the density of two random variables which in general may or may not exist [3]. Moreover, if we suppose the existence of the densities, it is always difficult to properly estimate them from sample data. Above all, even by increasing the sample size, there is no guarantee that the estimated density would converge to its true measure. To overcome the mentioned problems, different K-L estimation methods have been proposed by researchers, e.g. methods defined in [4–6]. Also, several alternative measures have been defined in the literature, e.g. in [7,8].

The above problems still remain when one wants to measure the distance between a set of sample data and a probability density function using the K-L divergence. Recently, Liu [9] has defined a new divergence measure between sample data and a probability distribution which is based on the survival function of the random variable X, namely $F(x) = P(X > x)$, instead of its density function $f(x)$. The survival function is more regular than the density function because it always exists, can be easily estimated from sample data, and its estimation is convergent by the law of large numbers. Moreover, in practice what is of interest and/or measurable in reliability theory is the survival function. For example, in describing the life span of a ceramic under a uniaxial tensile stress, the aim of interest is not whether the life span equals x, but whether

* Corresponding author: e-mail: asaghafi@iust.ac.ir

it exceeds x. The introduced divergence measures the distance between an empirical and a prescribed survival function and Liu has used it to estimate the parameters of exponential and uniform distributions. But in spite of the good results, the new measure has not been used seriously in parameter estimation sine then.

In this paper we have used the new divergence measure, called Kullback-Leibler divergence of Survival functions (KLS), to estimate the parameters of a Weibull distribution. We chose the two-parameter Weibull distribution since its survival function has a simple form. Also, our results would generalize the previous work of Liu for exponential distribution. Besides, the Weibull distribution has been vastly utilized in many scientific disciplines especially materials science. For a list of applications of the Weibull family of distributions see [10, 11].

The rest of the paper is organized as follows: an outline of the Weibull distribution is mentioned in Section 2. In Section 3 the KLS definition is restated and the Weibull parameter estimation using the KLS is described. Section 4 provides a simulation study to show the advantages of the KLS parameter estimation over the alternative methods and in Section 5 conclusions are made.

2. The Weibull distribution

Weibull distribution has become a well-established modeling tool in many scientific areas such as biology, environment, health, material and social science. The statistical variation in such scientific measurements can be described properly with the two-parameter Weibull survival function:

$$F(x) = \exp\left[-\left(\frac{x}{\sigma}\right)^m\right], \quad x \geq 0,\ m, \sigma > 0, \tag{2}$$

where m and σ are the shape and scale parameters respectively. $F(x)$ is also called probability of success in reliability theory and is widely used in materials science when fracture strength of brittle materials is measured. This is because of the well-known ”weakest link property” stating that the minimum of independent, identically distributed random variables (not necessarily Weibull distributed) has an approximate Weibull distribution, subject to some mild conditions concerning the distribution of such random variables. Bearing in mind that a piece of material can be viewed as a concentration of many smaller material cells, each of which has its random breaking strength X_i when subjected to stress. Thus the strength of the concentrated total piece is the strength of its weakest link, that is $min(X_1, \cdots, X_n)$, i.e., approximately Weibull.

In practice, it is usually necessary to fit a set of experimentally measured data into the Weibull equation given in (2), i.e. to estimate the two parameters m and σ. There have been several approaches for the estimation of these two parameters in the literature. Maximum likelihood and linear regression methods are the most widely used ones due to their precision and simple computations, respectively. These methods are not discussed in this paper and the reader is recommended to see [10, 12] for detailed methodology. The wide range of the Weibull applications in engineering and materials science motivated the authors to apply the new divergence measure in Weibull parameters estimation and compare the results with commonly used estimation methods. The results are provided by simulation and gathered in Section 4.

3. Kullback-Leibler divergence of Survival functions

The key idea to the KLS definition is to use survival functions in place of density functions in the Kullback-Leibler divergence and add a new term to make sure the new measure is always positive. This definition also represents the well established principle that the logarithm of the probability of an event should represent the information content in the event. First, recall a basic definition.

Definition 1. *Let $X_1, X_2, \cdots$ be a sequence of positive, independent and identically distributed (i.i.d) random variables from a non-increasing survival function $F(x, \Theta) = P_\Theta(X > x)$ with support S_x and vector of parameters Θ. Define the empirical survival function of a random sample of size n by*

$$G_n(x) = \sum_{i=0}^{n-1}(1 - \frac{i}{n})\, I_{[X_{(i)}, X_{(i+1)})}(x), \tag{3}$$

where I is the indicator function and $(0 = X_{(0)} \leq) X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ are the ordered sample.

Below we restated the KLS definition.

Definition 2. *Let $F(x, \Theta)$ be the true survival function with unknown parameters Θ and $G_n(x)$ be the empirical survival function of a random sample of size n from $F(x, \Theta)$. Define the Kullback-Leibler divergence of Survival functions G_n and F by*

$$KLS(G_n||F) = \int_0^\infty G_n(x) \ln\frac{G_n(x)}{F(x)} - [G_n(x) - F(x)]\, dx. \tag{4}$$

The two following theorems show that the KLS is a divergence measure which converges to zero with increasing sample size.

Theorem 1. *$KLS(G_n||F) \geq 0$ for all $n \in \mathbb{N}$, equality holds if and only if $G_n = F$.*

proof. It follows from the log-sum inequality and the inequality $x \ln\frac{x}{y} \geq x - y, \forall x > 0, y > 0$.

Theorem 2. *If $\int_0^\infty F(x) \ln F(x)\, dx < \infty$, the introduced measure converges to zero as n tends to infinity.*

proof. Note that G_n and F are integrable and (4) can be simplified to

$$KLS(G_n||F) = \int_0^\infty G_n(x) \ln \frac{G_n(x)}{F(x)}\, dx - [\bar{X}_n - E(X_1)]. \tag{5}$$

By the law of large numbers, $\bar{X}_n - E(X_1)$ converges to zero almost surely with increasing sample size, so we need only to show that the integral part in (5) converges to zero. In Theorem 9 of [13] it has been stated that

$$\left| \int_0^\infty G_n(x) \ln G_n(x)\, dx - \int_0^\infty F(x) \ln F(x)\, dx \right| \overset{n\to\infty}{\longrightarrow} 0, \tag{6}$$

so, it suffices to show that

$$\left| \int_0^\infty G_n(x) \ln F(x)\, dx - \int_0^\infty F(x) \ln F(x)\, dx \right| \to 0, \tag{7}$$

as $n \to \infty$. For a fixed n, using the definition of $G_n(x)$ given in (3) we can write

$$\int_0^\infty G_n(x) \ln F(x)\, dx = \sum_{i=0}^{n-1} (1 - \frac{i}{n}) \int_{x_{(i)}}^{x_{(i+1)}} \ln F(x)\, dx. \tag{8}$$

Define $h(x) := \int_0^x \ln F(t)\, dt$ for $x \in S_x$. Notice that h is well-defined because F is monotone on its support S_x and $h(0) = 0$. Then, (8) gives

$$\begin{aligned}
\int_0^\infty G_n(x) \ln F(x)\, dx &= \sum_{i=0}^{n-1} (1 - \frac{i}{n}) [h(x_{(i+1)}) - h(x_{(i)})] \\
&= \sum_{i=0}^{n-1} [h(x_{(i+1)}) - h(x_{(i)})] - \frac{1}{n} \sum_{i=0}^{n-1} i\, [h(x_{(i+1)}) - h(x_{(i)})] \\
&= [h(x_{(n)}) - h(x_{(0)})] - \frac{1}{n} \sum_{i=0}^{n-1} [(i+1) h(x_{(i+1)}) - i\, h(x_{(i)})] \\
&\quad + \frac{1}{n} \sum_{i=0}^{n-1} h(x_{(i+1)}) \\
&= h(x_{(n)}) - h(x_{(0)}) - \frac{1}{n} (n h(x_{(n)})) - 0) + \frac{1}{n} \sum_{i=1}^{n} h(x_{(i)}) \\
&= \frac{1}{n} \sum_{i=1}^{n} h(x_i).
\end{aligned}$$

Since $x_1, x_2, \cdots, x_n$ are random samples from F, if we tend n to infinity $\sum_{i=1}^n h(x_i)/n$ converges to $E[h(X)]$ by the law of large numbers. On the other hand, using integration by parts one can obtain

$$E[h(X)] = \int_0^\infty h\, d(1 - F) = \int_0^\infty F(x) \ln F(x)\, dx. \tag{9}$$

and the proof is complete.

The KLS divergence is not a metric in the mathematical sense. It is not symmetric and the triangle inequality does not hold for it. But it is still good enough for our purposes. Consider the following definition.

Definition 3. *Let $\hat{\Theta} = \arg\min_\Theta KLS(G_n||F)$. Call $\hat{\Theta}$ the KLS estimator of Θ.*

In order to use the KLS to estimate the parameters of a Weibull distribution, we just have to put the Weibull survival function given in (2) instead of F in (4). After simplifying we get

$$\begin{aligned}
KLS(G_n||F) &= \sum_{i=1}^{n-1} (1 - \frac{i}{n}) \ln(1 - \frac{i}{n}) \Delta x_{i+1} \\
&\quad + \frac{\sum_{i=1}^n x_i^{m+1}}{n(m+1)\sigma^m} - [\, \bar{x}_n - \sigma \Gamma(1 + \frac{1}{m})\,],
\end{aligned} \tag{10}$$

where $\Delta x_{i+1} = x_{i+1} - x_i$, $x_0 = 0$ and Γ is the gamma function defined as $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x}\, dx$ for $t > 0$. Equation (10) should be minimized for the values of m and σ to yield the KLS estimations. Setting the partial derivative with respect to σ equal to zero, yields

$$\sigma^{m+1} = \frac{m}{n(m+1)} \frac{\sum_{i=1}^n x_i^{m+1}}{\Gamma(1 + \frac{1}{m})}. \tag{11}$$

After substituting (11) into (10) and after some rearrangements, we get

$$\begin{aligned}
g(m) &= \sum_{i=1}^{n-1} (1 - \frac{i}{n}) \ln(1 - \frac{i}{n}) \Delta x_{i+1} - \bar{x}_n \\
&\quad + \frac{m^{-m/(m+1)} + m^{1/(m+1)}}{(n(m+1))^{1/(m+1)}} \sqrt[m+1]{\Gamma^m(1 + \frac{1}{m}) \sum_{i=1}^n x_i^{m+1}},
\end{aligned} \tag{12}$$

which is only a function of m. Thus, $g(m)$ in (12) should be minimized for m by a non-derivative based optimization method. Subsequently, σ is estimated from (11). Obtaining a closed-form solution of (12) for m is not possible and it must be solved numerically. This is easily done by MATLAB's `fminsearch` function.

4. Weibull parameters estimation

A series of MATLAB codes was written to generate values $x_1, x_2, \cdots, x_n$ from a Weibull distribution with $m = 10$ and $\sigma = 1$. For convenience we assumed $m_{\text{true}} = 10$ and $\sigma_{\text{true}} = 1$ throughout this study. This set of simulated values was used to estimate the Weibull parameters using the KLS, maximum likelihood (ML) and linear regression (LR) methods. The procedure was then repeated 10,000 times. Consequently, a total of 10,000 samples were generated and 10,000 Weibull parameter estimations were produced using each method. Then the mean values $\bar{m}$ and $\bar{\sigma}$, and sample variances S_m^2 and S_σ^2 were computed using:

$$\bar{m} = \sum_{j=1}^{10^4} \frac{m_j}{10^4}, \quad \bar{\sigma} = \sum_{j=1}^{10^4} \frac{\sigma_j}{10^4}, \tag{13}$$

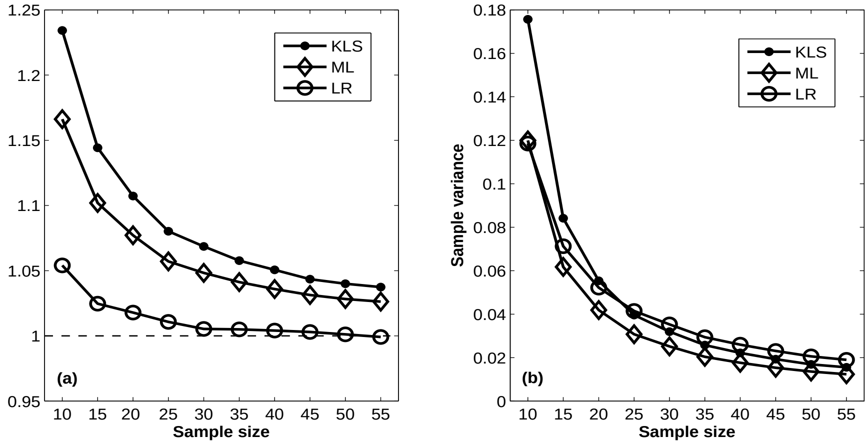

Figure 1: (a) $\bar{m}/m_{\text{true}}$ (b) S_m^2 as functions of sample size

$$S_m^2 = \sum_{j=1}^{10^4} \frac{(m_j - \bar{m})^2}{10^4 - 1}, \quad S_\sigma^2 = \sum_{j=1}^{10^4} \frac{(\sigma_j - \bar{\sigma})^2}{10^4 - 1}, \quad (14)$$

where m_j and σ_j represent the estimated Weibull shape and scale parameters from the jth sample. Clearly, the method which makes $\bar{m}/m_{\text{true}}$ and $\bar{\sigma}$ be equal to one provides a better estimation.

To illustrate the effect of the sample size, random samples of different sizes $n = 10, 15, 20, 25, 30, 35, 40, 45$ and 50 were used. For the linear regression method, we have used the probability index $P_i = \frac{i-0.5}{n}$, as it has been shown to provide the best estimations [12].

Although the above simulation was carried out for arbitrary-chosen values of $m_{\text{true}} = 10$ and $\sigma_{\text{true}} = 1$, its results are valid for any value of m_{true} and σ_{true}, as previous studies and numerical calculations have shown that the values of $\bar{m}/m_{\text{true}}$ and $\bar{\sigma}/\sigma_{\text{true}}$, and their distributions are independent of the prescribed values of m_{true} and σ_{true} [12,14].

Figure 1 (a) shows normalized mean Weibull modulus ($\bar{m}/m_{\text{true}}$) of the three different parameter estimation methods in each sample size. As it can be seen, all three methods actually overestimate the Weibull modulus. The KLS method leads to more biased estimations than the other two methods. Though, its bias decreases more rapidly with increasing sample size. Obviously the LR method provides a better estimation from this point of view. However, the estimation precision is related not only to the bias but also to the deviation from the true value of the parameter.

Figure 1 (b) shows sample variances of the three different methods. Its value for the KLS method is more than those of the ML and the LR methods for small sample sizes. For sample sizes more than 30, the sample variances of the KLS lies between that of the maximum likelihood and the linear regression. Together, it seems maximum likelihood estimation for Weibull modulus is more precise. The results of the maximum likelihood and linear regression methods which are presented here are in agreement with previous studies.

The estimators of the Weibull shape parameter are always biased for the maximum likelihood and linear regression methods [12,15]. Simulations provided here verified that also showing that the KLS method provides biased estimations, too. A general method to correct the statistical bias of the Weibull modulus estimation is to multiply it by an appropriate unbiasing factor [14]. Theses factors are computed for the KLS estimation method using a different simulation with a total of 100,000 samples and provided in Table 1. We show the unbiased estimations of the KLS method by UKLS. To make comparisons fair, we have derived unbiased estimations of the maximum likelihood method (UML) using unbiassing factors provided in ASTM Standard [15]. For the linear regression method we applied the method described in Wu et. al[16]. Figure 2 (a) shows the normalized unbiased mean Weibull modulus estimations with the KLS method along with that of the ML and the LR methods. The improvement in UKLS estimations is obvious. It seems that all three estimation methods perform equitable. Figure 2 (b) provides estimated mean square errors. Comparison of the three methods reveals almost equal performance of modulus estimation by the three methods.

It has been shown in previous studies that the estimated scale parameter from maximum likelihood and linear regression methods is approximately unbiased. Figure 3 (a) shows the same results and also demonstrates that the KLS estimator provides the least biased estimations for every sample size. Figure 3 (b) illustrates that the sample variances for the three methods are almost the same.

The parameter σ is in the exponent's argument, $(x/\sigma)^m$, in the Weibull equation. Hence its dispersion may still exert an important effect on the form of the Weibull survival function despite the small variation from its true value.

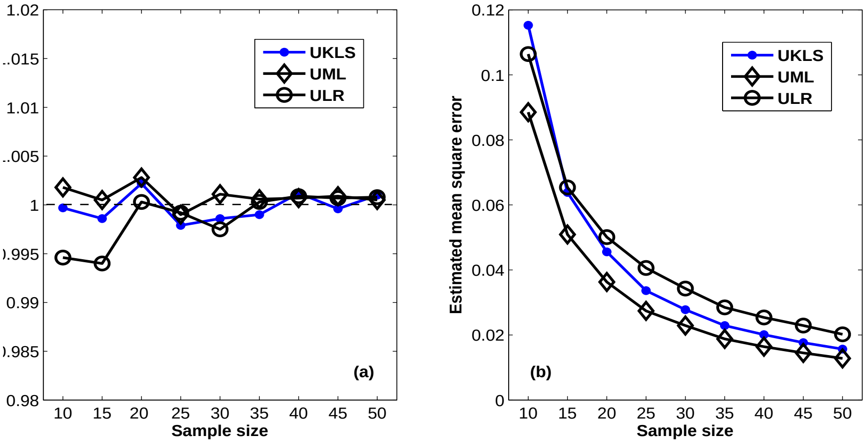

Figure 2: (a) $\bar{m}/m_{\text{true}}$ (b) $M\hat{S}E_m$ as functions of sample size

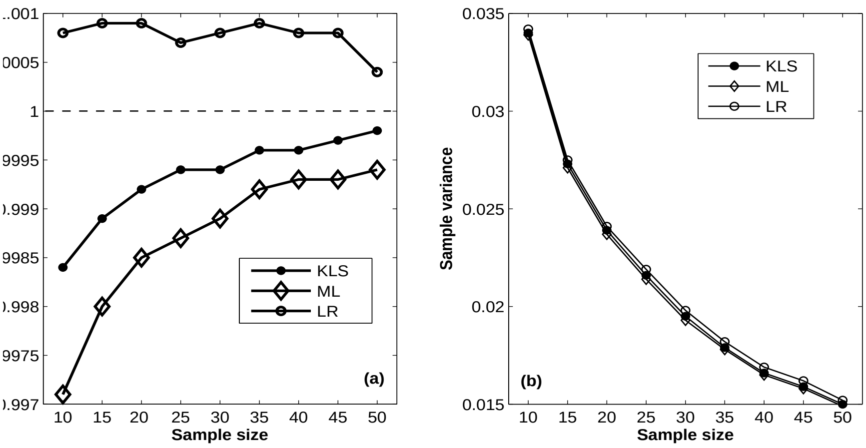

Figure 3: (a) $\bar{\sigma}/\sigma_{\text{true}}$ (b) S^2_σ as functions of sample size

Table 1: Unbiasing factors for Weibull modulus estimation using the DCE method

sample size n	Unbiasing factor	sample size n	Unbiasing factor
10	0.810	35	0.943
15	0.871	40	0.953
20	0.907	45	0.956
25	0.922	50	0.963
30	0.933		

So, based on these results, it is recommended to use the KLS method at least to estimate the Weibull scale parameter. Moreover, using unbiasing factors one can estimate the Weibull modulus by the KLS as good as the ML and LR methods.

Remark. If we set $m = 1$ from the beginning, the problem reduces to estimating the scale parameter σ of an exponential distribution with survival function $F(x) = \exp(-x/\sigma)$, $x \geq 0$. The estimation using the KLS would be $\hat{\sigma} = \left(\frac{1}{2n}\sum_{i=1}^n x_i^2\right)^{1/2}$ which is unique, consistent and its square is the unbiased estimator of σ^2. Given the ML estimation $\hat{\sigma} = \frac{1}{n}\sum_{i=1}^n x_i$, it's obvious that the KLS estimation converges in mean square error to the true value of σ faster than ML estimation. This especial case is studied in Liu [9].

5. Conclusion

This paper developed a new technique (KLS) for estimation of the Weibull parameters. The new method minimizes an entropy based distance between the empirical survival function and the Weibull survival function. Detailed results show that the KLS estimator for the scale parameter provides the least biased estimations for all sample sizes. Using unbiasing factors one can estimate the Weibull shape parameter as reliable as the maximum likelihood and the linear regression methods.

Applying the KLS in goodness-of-fit tests and modeling censored experimental data is a future work that will bring this measure more into attention since survival function is more easily estimated in case of censored samples. Furthermore, characteristics of the estimations by this method are a challenge and the objective of the authors.

References

[1] S. Kullback and R.A. Leibler, On Information and Sufficiency, Annals of Mathematical Statistics. **22**(1), 79-86 (1951).

[2] R. Abramov, A. Majda and R. Kleeman, Information theory and predictability for low-frequency variability, Journal of the Atmospheric Sciences. **62**(1), 65-87 (2005).

[3] T.M. Cover and J.A. Thomas, Elements of information theory. (Wiley, 2nd ed., 2006).

[4] Y.K. Lee and B.U. Park, Estimation of Kullback-Leibler divergence by local likelihood, Annals of the Institute of Statistical Mathematics. **58**(2), 327-340 (2006).

[5] Q. Wang, S. Kulkarni and S. Verdú, A nearest-neighbor approach to estimating divergence between continuous random vectors, IEEE International Symposium on Information Theory, Seattle, USA. (2006).

[6] F. Pérez-Cruz, Kullback-Leibler divergence estimation of continuous distributions, IEEE International Symposium on Information Theory, Toronto, Canada. (2008).

[7] J. Aczel and Z. Daroczy, On measures of information and their characterization.(Academic Press, New York, 1975).

[8] B. Forte and W. Hughes, The maximum entropy principle: a tool to define new entropies, Reports of mathematical physics. **26**(2), 227-238 (1988).

[9] J. Liu, Information theoretic content and probability. (Ph.D. Thesis, University of Florida, USA, pp. 25-26, 2007).

[10] H. Rinne. The Weibull Distribution: A handbook. (CRC Press, pp. 275, 2009).

[11] W. WEIBULL, A statistical distribution function of wide applicability, Journal of Applied Mechanics. **18**, 293-297 (1951).

[12] A. Khalili and K. Kromp, Statistical properties of Weibull estimators, Journal of Materials Science. **26**: 6741-6752 (1991).

[13] M. Rao, Y. Chen, B.C. Vemuri and F. Wang, Cumulative Residual Entropy: A new measure of Information, IEEE Transactions on Information Theory. **50**(6), 1220-1228 (2004).

[14] R. Ross, Bias and standard deviation due to Weibull parameter estimation for small data sets, IEEE Transactions on Dielectrics and Electrical Insulation. **3**(1), 28-42 (1996).

[15] ASTM standard, C1239-00 Standard practice for reporting uniaxial strength data and estimating Weibull distribution parameters for advanced ceramics, American Society for Testing and Materials, Philadelphia, PA., (2003).

[16] D. Wu, J. Zhou and Y. Li, Unbiased estimation of Weibull parameters with the linear regression method, Journal of the European Ceramic Society. **26**, 1099-1105 (2006).

Dr. Gholamhossein Yari is an assistant professor in Iran University of Science and Technology-(IUST) since 1988. Dr. Yari has received his Ph.D. in Applied Mathematics and Statistics in 2003 from IUST. His main research interests are entropy, information theory and stochastic processes. He has published over 15 scientific papers in national and international journals, over 17 conference papers, and written two books with several notable contributions to the field of entropy. He has been the head of the School of Mathematics of IUST since 2007.

Dr. Alireza Mirhabibi is an associate professor of Materials Sciences in Iran University of Science and Technology (IUST), member of Center of Excellence for Advanced Materials and Processes, and visiting professor in Institute for Materials Research (IMR), Leeds University, UK. He has published more that 50 refereed journal papers, more than 200 conference papers, 15 books, and 65 patents. Dr. Mirhabibi's main research area is ceramic processing, ceramic coating and colorant, refractories and carbon materials. His research results has been awarded many times in national and international societies.

Abolfazl Saghafi is a Ph.D. student in Iran University of Science and Technology working on cumulative entropy under supervision of Dr. Gholamhossein Yari. Mr. Saghafi has several assistantship, research cooperation and advisor experiences in his career and has published over six scientific and conference papers in national and international journals. Currently he is a visitor at University of Patras, Greece, and working with professor Andreas Philippou, a leader in the theory and applications of distributions of order k.