Journal of Statistical Research *ISSN 0256 - 422 X*
2008, Vol. 42, No. 2, pp. 89-98
Bangladesh

A NOTE ON KULLBACK-LEIBLER DISCRIMINATION INFORMATION PROPERTIES OF RECORD VALUES

M. Shakil
Department of Mathematics, Miami-Dade College, Hialeah Campus, Florida 33012, USA
Email: mshakil@mdc.edu

J. N. Singh
Department of Mathematics and Computer Science, Barry University
N. E. 2nd Ave., Miami Shores, FL 33161,USA
Email: jsingh@mail.barry.edu

B. M. Golam Kibria
Department of Statistics, Florida International University, Miami, FL 33199, USA
Email: kibriag@fiu.edu

summary

The Kullback-Leibler (K-L) discrimination information measure has been widely used for many years in various fields of studies, for example, economics, engineering, hydrology, order statistics, physics, psychology, etc. In this paper, some of the K-L information properties of record value distributions have been discussed. It has been shown that K-L discrimination of record value distribution is distribution-free irrespective of the parent distribution from which the record values are obtained.

Keywords and phrases: Entropy; Digamma Function; Kullback-Leibler Discrimination Information; Record Value; Weibull Distribution

AMS Classification: 94A15, 62B10, 62B15

1 Introduction

Let $\{X_n : n = 1, 2, 3, \ldots\}$ be a sequence of independent and identically distributed (iid) random variables from an absolutely continuous distribution function $F(x)$, with a probability density function $f(x)$. Then a record occurs at index time j, if $X_j > X_i, \forall 1 \le i \le j-1$, that is, X_j is larger than each of the previous values $X_1, X_2, \ldots, X_{j-1}$. Thus, the first observation is always a record, and if a record occurs at time t, then X_t will be called a record value. Let $R(j)$ denote the time (index) at which the jth record value is observed. Since the first observation is always a record value, we have $R(1) = 1, \ldots, R(j+1) = \min\left\{i : X_i > X_{R(j)}\right\}$, where we define $R(0) = 0$. Let $\Delta(j) = R(j+1) - R(j)$ denote the inter-record time between jth and $(j+1)th$ record values, and let the jth record value $X_{R(j)}$ be denoted by $X(j)$ for simplicity. Then the joint probability density function of the record values $X(1), X(2) \ldots X(j)$ is given by

$$g_{X(1),\ldots,X(j)}\left(x_1,\ldots,x_j\right) = [f(x_j)] \prod_{r=1}^{j-1} \frac{f(x_r)}{1-F(x_r)}, \tag{1.1}$$

and the marginal probability density function (pdf) of the jth record value $X(j)$ is given by

$$g_{X(j)}(x) = \frac{\left[-\ln\left\{1-F(x)\right\}\right]^{(j-1)} f(x)}{(j-1)!}, \tag{1.2}$$

(see, for example, Qasem (1996), Gulati and Padgett (2003), Arnold et al. (1998), and Ahsanullah (2004), among others). There are applications of record values in many areas such as climatology, medicine, science, sports, traffic, industry, etc. The development of the general theory of statistical analysis of record values began with the work of Chandler (1952). Further developments continued with the contributions of Foster and Stuart (1954), Arnold et al. (1998), and Ahsanullah (1995, 2004), among others. It appears from the literature that, in spite of the extensive work on record values, very little attention has been paid to Shannon's entropy or K-L information properties of record values. For details on Kullback-Leibler discrimination information function and its properties, please visit Kullback (1959), Kapur (1989), Soofi et al. (1995), Ebrahimi and Kirmani (1996), Alwan et al. (1998), and references therein. Recently, Park (1999) has discussed a goodness-of-fit test for normality based on the sample entropy of order statistics. Raqab and Awad (2001) have discussed the characterizations of the Pareto and related distributions based on Shannon's entropy of k-record statistics. Analogous to record value distributions, Ebrahimi et al. (2004) have explored the properties of entropy, Kullback-Leibler information, and mutual information for order statistics. A goodness of fit test statistic based on the Kullback-Leibler information with the type II censored data is discussed in Park (2005). The entropies of record value distributions obtained from some commonly used continuous probability models have been discussed in Zahedi and Shakil (2006). For information properties of record values corresponding to Weibull and normal distributions, see, for example, Shakil (2005, 2006), among others. For recent work on entropy of record value distributions, see, for example, Baratpour et al. (2007a, 2007b). In this paper, some of the Kullback-Leibler information properties of record value distributions have been discussed.

The organization of this paper is as follows. Section 2 discusses some of the K-L information properties of record value distributions. Some concluding remarks are presented in section 3.

2 K-L Discrimination Information Function Between the Distribution of jth Record Value and the Parent Distribution

This section discusses some results on Kullback-Leibler discrimination information function between the distribution of jth record value and parent distribution.

2.1 Some Preliminaries

Definition 2.1. (Entropy of Record Value Distributions) Let X be an absolutely continuous random variable with a distribution function $F(x\,;\theta)$ and a probability density function $f(x\,;\theta)$. Then entropy of random variable X is given by

$$H_X\left[f(x)\right] = -\int_{-\infty}^{+\infty} \left(f(x)\right) \ln\left(f(x)\right)\, dx \tag{2.1}$$

(see Shannon, 1948). Let $H_{(j)}$ denote entropy of jth record value $X(j)$. Then, from (1.2) and (2.1), it follows that

$$H_{(j)} \;=\; \ln(\Gamma(j)) - (j-1)\,\psi(j) - \frac{1}{\Gamma(j)} \int_{-\infty}^{+\infty} \left[-\ln(1-F_X(x)\right]^{j-1} f_X(x)\, \ln(f_X(x))\, dx, \tag{2.2}$$

where $j = 1, 2, 3, \ldots$, and $\psi(j)$ is digamma function. For example, it is easy to see from (2.2) that entropy of jth record value from a standard normal distribution $N(0,1)$ is given by

$$\begin{aligned} H_{(j)} \; &= \ln\left\{\sqrt{2\pi}\Gamma(j)\right\} - (j-1)\,\psi(j) \\ &+ \tfrac{1}{2\sqrt{2\pi}\,\Gamma(j)} \int_{-\infty}^{+\infty} x^2\, e^{-x^2/2} \left[\ln\left\{\tfrac{2}{1-erf\left(x/\sqrt{2}\right)}\right\}\right]^{j-1} dx \end{aligned}, \tag{2.3}$$

where $erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2} du = \Pr\left\{|Y| \le x\right\}$, $Y \sim N\left(0, \frac{1}{2}\right)$, and $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^0 e^{-u^2/2} du = \frac{1}{2}$. Also, in view of (2.2), entropy of jth record value from a two-parametric Weibull $W\left(\lambda, \beta\right)$ parent distribution can be expressed as

$$H_{(j)} = \ln\left[\Gamma(j)\right] - \ln(\beta) - \frac{\ln(\lambda)}{\beta} - \left[j - \frac{1}{\beta}\right] \psi(j) + j, \quad j = 1, 2, 3, \ldots, \tag{2.4}$$

where $\lambda > 0$ and $\beta > 0$. For details on these, see for example, Shakil (2005, 2006), and Zahedi and Shakil (2006), among others.

Definition 2.2. Kullback-Leibler Discrimination Information Function0 Given any two distributions $F_1(x)$ and $F_2(x)$, with probability density functions $f_1(x)$ and $f_2(x)$, respectively, the discrepancy (also known as distance, divergence, or information number) between the distributions $F_1(x)$ and $F_2(x)$ is measured by Kullback-Leibler discrimination information function defined by

$$K(f_1(x), f_2(x)) = E\left(\ln\left[\frac{f_1(x)}{f_2(x)}\right]\right) = H_1(X) - \int f_1(x)\ln\left(f_2(x)\right)dx, \tag{2.5}$$

and

$$K(f_2(x), f_1(x)) = E\left(\ln\left[\frac{f_2(x)}{f_1(x)}\right]\right) = H_2(X) - \int f_2(x)\ln\left(f_1(x)\right)dx, \tag{2.6}$$

where $H_1(X)$ and $H_2(X)$ denote entropies of the distributions $F_1(x)$ and $F_2(x)$ respectively. Note that $K(f_1(x),f_2(x))$ is entropy of $f_1(x)$ relative to $f_2(x)$; $K(f_1(x),f_2(x)) \geq 0$, and the equality sign holds if and only if $f_1(x) = f_2(x)$; $K(f_1(x),f_2(x)) \neq K(f_2(x),f_1(x))$; $K(f_1(x),f_2(x))$ is a convex function of the pair $(f_1(x),f_2(x))$, and, hence, $K(f_1(x),f_2(x))$ can be minimized with respect to $f_1(x)$ or $f_2(x)$; $K(f_1(x),\ f_2(x))$ is invariant under 1-1 transformation of x; and the symmetric divergence function between the distributions $F_1(x)$ and $F_2(x)$ is defined by

$$\begin{aligned} J(f_1(x), f_2(x)) &= J(f_2(x), f_1(x)) = K(f_1(x), f_2(x)) + K(f_2(x), f_1(x)) \\ &= \int [f_1(x) - f_2(x)] \ln\left[\frac{f_1(x)}{f_2(x)}\right] dx. \end{aligned}$$

2.2 K-L Discrimination Information Function Between the Distribution of Record Values

Let $f(x)$ denote the probability density function of the parent distribution, with distribution function as $F(x)$. Let $f_j(x)$ denote the probability density function of jth record value, with distribution function as $F_j(x)$, obtained from parent distribution.

Theorem 1. *Let $f_X(x)$ denote the probability density function of the parent distribution, with distribution function as $F(x)$. Let $f_{X(j)}(x)$ denote the probability density function of the jth record value, with distribution function as $F_{X(j)}(x)$. Let $K(f_X(x),\ f_{X(j)}(x))$ denote the Kullback-Leibler discrimination information function between the parent distribution $F(x)$ and distribution $F_{X(j)}(x)$ of jth record value, and $K(f_{X(j)}(x),\ f_X(x))$ denote the Kullback-Leibler discrimination information function between the distribution $F_{X(j)}(x)$ of jth record value and parent distribution $F(x)$. Let $K(f_X(x),\ f_{X(l)}(x))$ denote the Kullback-Leibler discrimination information function between the parent distribution and distribution of lth record value, where $l > j$. Let $l = j + k$,where $k \geq 1$ is a positive integer. Then, we have*

(i) $K(f_X(x), f_{X(j)}(x)) = \ln[\Gamma(j)] + (j-1)\gamma$, *and* $K(f_{X(j)}(x), f_X(x)) = (j-1)\psi(j) - \ln(\Gamma(j), \forall j \geq 1$, *where* $\gamma = \lim_{j\to\infty}\left[\sum_{k=1}^{j-1}\frac{1}{k} - \ln j\right] \approx 0.57721566490\ldots$ *is Euler's constant, and* $\psi(j)$ *denotes digamma function;*

(ii) $K(f_X(x), f_{X(j)}(x))$ *is a monotone increasing function in* j, $\forall j \geq 2$*;*

(iii) $K(f(x), f_{j+k}(x)) = K(f_X(x), f_{X(j)}(x)) + \sum_{i=j}^{j+k-1} \ln(i) + k\gamma, \quad k = 1, 2, 3, \ldots;$

(iv) $\Delta(l, j) = K(f(x), f_l(x)) - K(f(x), f_j(x)) = \ln\left[\frac{\Gamma(l)}{\Gamma(j)}\right] + (l - j).\gamma$, $l > j$;

(v) $\Delta^k(j) = K(f(x), f_{j+k}(x)) - K(f(x), f_j(x)) = \ln[\Gamma(k)] - \ln(B(j,k)) + k.\gamma$; *where* $B(j,k)$ *denotes beta function between* j *and* k*; and*

(vi) K-L discrimination of record value distribution is distribution-free and is a function of index j *only.*

Proof. The Kullback-Leibler discrimination information function between the distribution of jth record value and parent distribution is given by

$$\begin{aligned} K(f(x), f_j(x)) &= \int_{-\infty}^{+\infty} f(x) \ln\left[\frac{f(x)}{f_j(x)}\right] dx \\ &= \int_{-\infty}^{+\infty} f(x) \ln[f(x)]\, dx - \int_{-\infty}^{+\infty} f(x) \ln[f_j(x)] \\ &= \int_{-\infty}^{+\infty} f(x) \ln[f(x)]\, dx - \int_{-\infty}^{+\infty} f(x) \ln\left[\frac{\{-\ln(1 - F(x))\}^{j-1} f(x)}{\Gamma(j)}\right] dx \\ &= \ln[\Gamma(j)] - \int_{-\infty}^{+\infty} f(x) \ln[\{-\ln(1 - F(x))\}^{j-1}] dx. \end{aligned} \tag{2.7}$$

Proof of (i) easily follows by substituting $-\ln(1 - F(x)) = u$ in the integral of equation (2.7), and noting that $\int_0^\infty e^{-u} \ln u du = -\gamma$, where γ is Euler's constant (see, for example, Gradshteyn and Ryzhik (1980), Equation 8.367 (2.2), p. 946, or Abramowitz and Stegun (1970), among others). The proof of second part in (i) can be similarly established. Proofs of (ii) and (iii) easily follow from (i). Proof of (iv) easily follows from part (i), and noting that $\ln[\Gamma(l)] - \ln[\Gamma(j)] = \ln\left[\frac{\Gamma(l)}{\Gamma(j)}\right]$. Since $l = j + k$, that is, $k = l - j$, proof of (v) follows easily by substituting $l - j = k$ in (i), and using the definition of beta function $B(j, k)$. Part (vi) is obvious from the above results (i) – (v), that is, K-L discrimination of record value distribution is distribution-free. The behaviors of the functions in equations (iv) and (v) are illustrated below in figures 1 and 2 respectively. □

Theorem 2. *The average Kullback-Leibler discrimination information discrepancies between the parent distribution* $F(x)$ *and distribution* $F_{X(j)}(x)$ *of jth record value are given by*

$$\overline{K}(f_X(x), f_{X(j)}(x)) = \frac{1}{n} \sum_{j=1}^{n} \ln[\Gamma(j)] + \frac{1}{2}(n - 1)\gamma,$$

and

$$\overline{K}(f_{X(j)}(x), f_X(x)) = \frac{1}{n} \sum_{j=1}^{n} (j - 1)\psi(j) - \frac{1}{n} \sum_{j=1}^{n} \ln[\Gamma(j)].$$

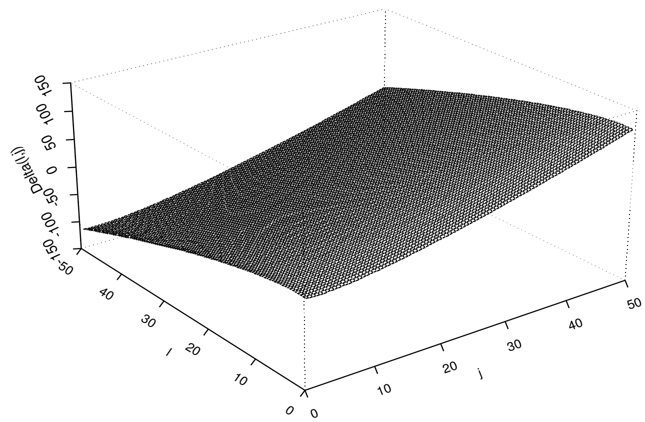

Figure 1: Figure 1: Plot of the function in equation (iv) in theorem 2.2.1

Proof. The proof of Theorem 2.2.2 easily follows from (i) of Theorem 2.2.1. Obviously, Theorem 2.2.2 is distribution-free and is a function of index j only. □

Corollary 2.1. The symmetric divergence between the parent distribution $F(x)$ and distribution $F_{X(j)}(x)$ of jth record value is given by

$$J(f_X(x), f_{X(j)}(x)) = (j-1)\left[\gamma + \psi(j)\right],$$

and is a monotone increasing function in j, $\forall\ j \geq 1$.

Proof. The proof easily follows from part (i) Theorem 2.2.1. □

Corollary 2.2. The average symmetric divergence between the distribution $F_{X(j)}(x)$ of jth record value and parent distribution $F(x)$ is given by

$$\overline{J}(g_{X(j)}(x), f_X(x)) = \frac{1}{n}\sum_{j=1}^{n}(j-1)\psi(j) + \frac{1}{2}(n-1)\gamma.$$

Proof. The proof easily follows from part (i) of Theorem 2.2.1. □

We note that the K-L discrimination of record value distribution is distribution-free which has been illustrated by using the Weibull as the parent distribution and provided them in the following theorem.

Theorem 3. *Let $g_{X(j)}(x)$ and $g_{X(k)}(x)$ denote the probability density functions of jth and kth record values, with distribution functions as $G_{X(j)}(x)$ and $G_{X(k)}(x)$, respectively, obtained from Weibull parent distribution. Let $K(g_{X(j)}(x), g_{X(k)}(x))$ denote Kullback-Leibler*

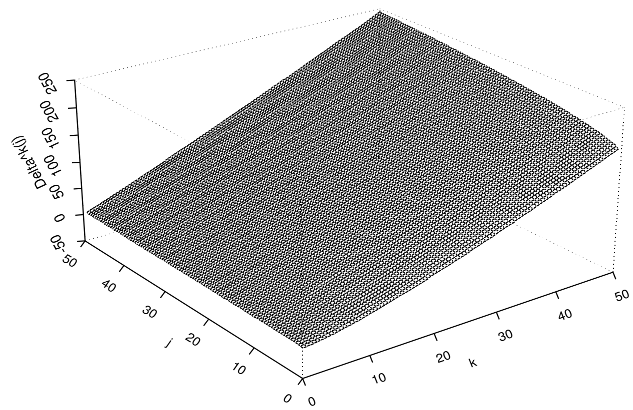

Figure 2: Figure 2: Plot of the function in equation (v) in theorem 2.2.1

discrimination information function between the distribution functions $G_{X(j)}(x)$ *and* $G_{X(k)}(x)$ *of the jth and kth record values respectively. Then*

(i) $K(g_{X(j)}(x), g_{X(k)}(x)) = \ln\left(\frac{\Gamma(k)}{\Gamma(j)}\right) + (j-k)\psi(j),\ j = 1,2,3,\ldots, k = 1,2,3,\ldots,$ *where* $\psi(j)$ *denotes the digamma function;*

(ii) $K(g_{X(j)}(x), g_{X(j+1)}(x)) = \ln(j) - \psi(j)$*, and is a monotone decreasing function in* j*,* $\forall j \geq 1$*; and*

(iii) $K(g_{X(j+1)}(x), g_{X(j)}(x)) = \psi(j+1) - \ln(j)$*, and is a monotone decreasing function in* $j, \forall j \geq 1$*.*

Proof. The Kullback-Leibler discrimination information function between the distribution functions $G_{X(j)}(x)$ and $G_{X(k)}(x)$ of the jth and kth record values, respectively, is given by

$$\begin{aligned} K(g_{X(j)}(x), g_{X(k)}(x)) &= \int_0^\infty g_{X(j)}(x) \ln\left[\frac{g_{X(j)}(x)}{g_{X(k)}(x)}\right] dx \\ &= \int_0^\infty g_{X(j)}(x) \ln\left[g_{X(j)}(x)\right] dx - \int_0^\infty g_{X(j)}(x) \ln\left[g_{X(k)}(x)\right] dx \\ &= -H_{X(j)}(g_{X(j)}(x)) - \int_0^\infty \frac{\lambda^j \beta x^{\beta j-1} e^{-\lambda x^\beta}}{\Gamma(j)} \ln\left[\frac{\lambda^k \beta x^{\beta k-1} e^{-\lambda x^\beta}}{\Gamma(k)}\right] dx, \end{aligned} \tag{2.8}$$

where $H_{X(j)}(g_{X(j)}(x))$ denotes the entropy of jth record value from parent Weibull (λ, β) in $(0, \infty)$ distribution. In order to evaluate the integral in (2.8), substituting $x^\beta = u$, and noting again that $\int_0^\infty u^j e^{-\lambda u} du = \frac{\Gamma(j+1)}{\lambda^{j+1}}$, and $\int_0^\infty u^{j-1} e^{-\lambda u} \ln u du = \frac{\Gamma(j)}{\lambda^j}\left[\psi(j) - \ln(\lambda)\right]$, it is easy to see that

$$\int_0^\infty \frac{\lambda^j \beta x^{\beta j-1} e^{-\lambda x^\beta}}{\Gamma(j)} \ln\left[\frac{\lambda^k \beta x^{\beta k-1} e^{-\lambda x^\beta}}{\Gamma(k)}\right] dx = \ln\left[\frac{\lambda^k \beta}{\Gamma(j)}\right] + \frac{(\beta k-1)}{\beta}\left[\psi(j) - \ln\lambda\right] - j. \tag{2.9}$$

Hence, using the expression (2.4) for $H_{X(j)}(g_{X(j)}(x))$, and the above expression (2.9) in (2.8), we obtained as

$$K(g_{X(j)}(x), g_{X(k)}(x)) = \ln\left(\frac{\Gamma(k)}{\Gamma(j)}\right) + (j-k)\psi(j), j = 1, 2, 3, \ldots, k = 1, 2, 3, \ldots, \tag{2.10}$$

which completes the proof of part (i). The proofs of parts (ii) and (iii) easily follow from (2.10). □

Corollary 2.3. The symmetric divergence between the distribution functions $G_{X(j+1)}(x)$ and $G_{X(j)}(x)$ of the $(j+1)th$ and jth record values respectively is given by $J(g_{X(j+1)}(x), g_{X(j)}(x)) = \frac{1}{j}$, and is a monotone decreasing function in $j, \forall j \geq 1$.

Proof. The proof easily follows from parts (ii) and (iii) of Theorem 2.2.3. □

3 Concluding Remarks

This paper has investigated the Kullback-Leibler discrimination information function between the distribution of jth record value and parent distribution. It has been observed that K-L discrimination of record value distribution is distribution-free irrespective of the parent distribution from which the record values are obtained. This has been illustrated by using Weibull as parent distribution. We hope that the findings of the paper would be very useful for the applied as well as theoretical researchers.

Acknowledgements

The authors are thankful to the referees for their valuable comments/suggestions which improved the quality and presentation of the paper greatly.

References

[1] Abramowitz, M., and Stegun, I.A. (1970). *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*. Dover: New York.

[2] Ahsanullah, M. (1995). *Record Statistics. Nova Science Publishers*: New York.

[3] Ahsanullah, M. (2004). *Record Values - Theory and Applications*. University Press America: Lanham, MD.

[4] Alwan, L.C., Ebrahimi, N., and Soofi, E.S. (1998). Information theoretic framework for process control - Theory and Methodology. *European Journal of Operational Research*, **111**, 526-542.

[5] Arnold, B.C., Balakrishnan, N., and Nagaraja, H.N. (1998). *Records*. John Wiley & Sons: N. Y.

[6] Baratpour, S., Ahmadi, J., and Arghami, N. (2007a). Entropy Properties of record statistics. *Statistical Papers*, **48**, 197-213.

[7] Baratpour, S., Ahmadi, J., and Arghami, N. (2007b). Some Characterizations Based on Entropy of Order Statistics and Record Values. *Communications in Statistics-Theory & Methods*, **36**, 47-57.

[8] Chandler, K.M. (1952). The Distribution and Frequency of Record Values. *Journal of the Royal Statistical Society, series B*, **14**, 220-28.

[9] Ebrahimi, N., and Kirmani, S.N.U.A. (1996). A Characterization of the Proportional Hazards Model through a Measure of Discrimination Between Two Residual Life distributions. *Biometrika*, **83**, 233-35.

[10] Ebrahimi, N., Soofi, E.S., and Zahedi, H. (2004). Information Properties of Order Statistics and Spacings. *IEEE Transactions on Information Theory*, **50**, 177-183.

[11] Foster, F.C., and Stuart, A. (1954). Distribution – Free Test in Time Series Based on the Breaking of Records. *Journal of the Royal Statistical Society*, Series B, **16**, 1-2.

[12] Gulati, S., and Padgett, W.J. (2003). Parametric and Nonparametric Inference from Record-Breaking Data. Springer- Verlag: New York.

[13] Kapur, J.N. (1989). Maximum-Entropy Models in Science and Engineering. Wiley-Eastern: New Delhi.

[14] Gradshteyn, I., and Ryzhik, I. (1980). Tables of Integrals, Series and Products. Academic Press: New York.

[15] Kullback, S (1959). *Information Theory and Statistics*. John Wiley & Sons: N. Y.

[16] Park, S. (1999). A goodness-of-fit test for normality based on the sample entropy of order statistics.it Statistics and Probability Letters., **44**, 359-363.

[17] Park, S. (2005). Testing exponentiality based on the Kullback-Leibler information with the type II censored data. *IEEE Transactions on Reliability*, **54**(1), 22-6.

[18] Qasem, A.A. (1996). Estimation via Record Values. *Journal of Information & Optimization Sciences*, **17**, 541-48.

[19] Raqab, M.Z., and Awad, A.M. (2001). A Note on Characterization Based on Shannon Entropy of Record Statistics. *Statistics*, **35**, 411-413.

[20] Shakil, M. (2005). Entropies of Record Values Obtained From The Normal Distribution and Some of their Properties. *Journal of Statistical Theory and Applications*,**4**(4), 371-386.

[21] Shakil, M. (2006). Entropy and Information Properties of Record Values from a Weibull Distribution. *Journal of Statistical Theory and Applications*, **5**(2), 105-126.

[22] Shannon, C.E. (1948). A Mathematical Theory of Communication. *Bell System Tech. Journal*, **27**, 379-23.

[23] Soofi, E.S., Ebrahimi, N., and Habibullah, M. (1995). Information Distinguishability with Application to Analysis of Failure Data. *Journal of the American Statistical Association*, **90**, 657-68.

[24] Zahedi, H. and Shakil, M. (2006). Properties of Entropies of Record Values in Reliability and Life Testing Context. *Communications in Statistics-Theory & Methods*, **35**, 997-1010.