Statistics and Probability Letters 97 (2015) 125–131

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

On residual inaccuracy of order statistics

Richa Thapliyal *, H.C. Taneja

Department of Applied Mathematics, Delhi Technological University, Bawana Road, Delhi 110042, India

ARTICLE INFO

Article history:
Received 28 May 2014
Received in revised form 28 October 2014
Accepted 9 November 2014
Available online 18 November 2014

Keywords:
Order statistics
Residual entropy
Hazard rate
Survival function

ABSTRACT

We propose the measure of residual inaccuracy of order statistics and prove a characterization result for it. Further we characterize some specific lifetime distributions using residual inaccuracy of the first order statistics. We also discuss some properties of the proposed measure.

1. Introduction

Let X and Y be two non-negative random variables with pdf respectively f and g. Shannon (1948) measure of uncertainty associated with X and Kullback (1959) measure of discrimination of X about Y are given by respectively

$$H(X) = H(f) = -\int_0^\infty f(x) \log f(x) dx, \tag{1}$$

and

$$H(f|g) = \int_0^\infty f(x) \log\left(\frac{f(x)}{g(x)}\right) dx. \tag{2}$$

Adding (1) and (2), we get

$$H(f) + H(f|g) = -\int_0^\infty f(x) \log g(x) dx, \tag{3}$$

which is Kerridge (1961) measure of inaccuracy associated with random variables X and Y. If we consider F as the actual distribution function then G can be interpreted as reference distribution function.

In survival analysis and life testing, the current age of the system under consideration is also taken into account. Thus, for calculating the remaining uncertainty of a system which has survived up to time t, the measures defined in (1)–(3) are not suitable. Ebrahimi (1996) considered a random variable $X_t = (X - t)|X > t,\ t \geq 0$ and defined uncertainty and discrimination of such a system, given by

$$H(f; t) = -\int_t^\infty \frac{f(x)}{\bar{F}(t)} \log\left(\frac{f(x)}{\bar{F}(t)}\right) dx, \tag{4}$$

* Corresponding author.
E-mail addresses: richa31aug@gmail.com (R. Thapliyal), hctaneja@rediffmail.com (H.C. Taneja).

http://dx.doi.org/10.1016/j.spl.2014.11.010

and

$$H(f|g;t) = \int_t^\infty \frac{f(x)}{\bar{F}(t)} \log\left(\frac{f(x) \setminus \bar{F}(t)}{g(x) \setminus \bar{G}(t)}\right) dx \tag{5}$$

respectively, where $\bar{F}(t) = 1 - F(t)$.

Clearly when $t = 0$, then (4) and (5) reduce respectively to (1) and (2).

Taneja et al. (2009) defined dynamic measure of inaccuracy associated with two residual lifetime distributions F and G corresponding to the Kerridge measure of inaccuracy given by

$$H(f,g;t) = -\int_t^\infty \frac{f(x)}{\bar{F}(t)} \log\left(\frac{g(x)}{\bar{G}(t)}\right) dx. \tag{6}$$

Clearly for $t = 0$, it reduces to (3).

In this communication we propose the measure of residual inaccuracy of order statistics. By the term order statistics we mean if $X_1, X_2, \ldots, X_n$ are n independent and identically distributed observations from a distribution F, where F is differentiable with a density f which is positive in an interval and zero elsewhere, then the order statistics of a sample is defined by the arrangement of $X_1, X_2, \ldots, X_n$ from the smallest to the largest denoted as $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$. These statistics have been used in a wide range of problems like detection of outliers, characterizations of probability distributions, testing strength of materials etc., for details refer to Arnold et al. (1992) and David and Nagaraja (2003). In reliability theory order statistics are used for statistical modeling, as the ith order statistics in a sample of size n corresponds to life length of a $(n-i+1)$-out-of-n system. The pdf of the ith order statistics $X_{i:n}$ is given by

$$f_{i:n}(x) = \frac{1}{B(i, n-i+1)} F(x)^{i-1}(1-F(x))^{n-i} f(x), \tag{7}$$

where $B(i, n-i+1) = \frac{\Gamma(i)\Gamma(n-i+1)}{\Gamma(n+1)}$ is the beta function with parameters i and $(n-i+1)$, for details refer to Arnold et al. (1992).

Several authors have worked on information theoretic aspects of order statistics, for details refer to Ebrahimi et al. (2004) and Zarezadeh and Asadi (2010). Recently Thapliyal and Taneja (2013) have introduced the concept of inaccuracy using order statistics. They have proposed the measure of inaccuracy between the ith order statistics and the parent random variable and proved a characterization result for the same. In this paper we extend the concept of inaccuracy of ordered random variables to dynamical system. The organization of this paper is as follows: In Section 2 we propose the measure of residual inaccuracy for the ith order statistics and explore some properties of it. Section 3 focuses on characterization results based on residual inaccuracy of order statistics. Some concluding remarks are mentioned in Section 4.

2. Measure of residual inaccuracy for $X_{i:n}$

Ebrahimi et al. (2004) studied some information theoretic measures based on order statistics using probability integral transformation and defined Shannon entropy and Kullback relative information measures, which are given by respectively

$$H_n(X_{i:n}) = H_n(f_{i:n}) = -\int_0^\infty f_{i:n}(x) \log f_{i:n}(x) dx = H_n(W_{i:n}) - E_{g_i}[\log f(F^{-1}(W_i))], \tag{8}$$

and

$$H_n(f_{i:n}; f) = \int_0^\infty f_{i:n}(x) \log\left(\frac{f_{i:n}(x)}{f(x)}\right) dx = -H_n(W_{i:n}), \tag{9}$$

where $W_{i:n}$ is the ith order statistics of uniformly distributed random variables $U_1, U_2, \ldots, U_n$ and

$$g_i(w) = \frac{1}{B(i, n-i+1)} w^{i-1}(1-w)^{n-i}, \quad 0 \le w \le 1,$$

is density function of $W_{i:n}$.

Thapliyal and Taneja (2013) defined the inaccuracy between the ith order statistics and the parent random variable as

$$I_n(f_{i:n}, f) = -\int_0^\infty f_{i:n}(x) \log(f(x)) dx. \tag{10}$$

Analogous to (10), we propose

$$I_n(f_{i:n}, f; t) = -\int_t^\infty \frac{f_{i:n}(x)}{\bar{F}_{i:n}(t)} \log\left(\frac{f(x)}{\bar{F}(t)}\right) dx, \tag{11}$$

Table 1
Residual inaccuracy for the first order statistics.

Distribution	P.d.f.	Residual inaccuracy
Uniform in $[0, b]$	$\frac{1}{b}$	$\log\left(\frac{b-t}{b}\right) + \log(b)$
Exponential	$\theta e^{-\theta x}$	$\frac{1}{n} - \log\theta$
Pareto	$\frac{ab^a}{(x+b)^{a+1}},\ a > 1, b > 0$	$-\log\left(\frac{a}{t+b}\right)+\left(\frac{a+1}{na}\right)$
Finite range	$a(1-x)^{a-1}$	$-\log\left(\frac{a}{1-t}\right)+\left(\frac{a-1}{na}\right)$

as the dynamic residual measure of inaccuracy associated with two residual lifetime distributions $F_{i:n}$ and F. Note that $\bar{F}_{i:n}(t) = 1 - F_{i:n}(t)$ is the survival function corresponding to $X_{i:n}$ given by

$$\bar{F}_{i:n}(t) = \frac{\bar{B}_{F(t)}(i, n-i+1)}{B(i, n-i+1)}$$

where

$$\bar{B}_x(a, b) = \int_x^1 u^{a-1}(1-u)^{b-1}du, \quad 0 < x < 1,\ a > 0,\ b > 0$$

is the incomplete beta function, for details, refer to David and Nagaraja (2003).

Note that when $t = 0$, (11) reduces to measure of inaccuracy as defined in (10).

In Table 1, we compute the residual inaccuracy measure for the first order statistics for some specific lifetime distributions which are applied widely in reliability and life testing of system.

Proposition 2.1. *Let $M = f(m) < \infty$ where $m = \sup\{x; f(x) \le M\}$ is the mode of the distribution. Then*

$$I_n(f_{i:n}, f; t) \ge \log(\bar{F}(t)) - \log M.$$

Proof. We have from (11)

$$\begin{aligned} I_n(f_{i:n}, f; t) &= -\int_t^\infty \frac{f_{i:n}(x)}{\bar{F}_{i:n}(t)} \log\left(\frac{f(x)}{\bar{F}(t)}\right) dx \\ &= \log \bar{F}(t) - \frac{1}{\bar{F}_{i:n}(t)} \int_t^\infty f_{i:n}(x) \log f(x) dx. \end{aligned}$$

As m is the mode of the distribution, hence $\log f(x) \le \log M$. Using this fact in above equation we get

$$I_n(f_{i:n}, f; t) \ge \log \bar{F}(t) - \log M. \tag{12}$$

Next we want to prove an important property of inaccuracy measure using some properties of stochastic ordering. For that we present the following definitions:

1. A random variable X is said to be less than Y in *likelihood ratio ordering* (denoted by $X \overset{lr}{\le} Y$) if $\frac{f_X(x)}{g_Y(x)}$ is non increasing in x.
2. A random variable X is said to be less than Y in the *stochastic ordering* (denoted by $X \overset{st}{\le} Y$) if $\bar{F}(x) \le \bar{G}(x)$ for all x, where $\bar{F}(x)$ and $\bar{G}(x)$ are the survival functions of X and Y respectively.

Theorem 2.1. *Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables representing the lifetime of a series system, that is when $i = 1$. Let F and f denote their distribution function and density function respectively. If f is decreasing in its support then corresponding inaccuracy is the decreasing function of n.*

Proof. We know that the random variable $\{X_{i:n} | X_{i:n} > t\}$ has density function

$$g_i(y) = \frac{1}{\bar{B}_{F(t)}(i, n)} y^{i-1}(1-y)^{n-1}, \quad F(t) \le y \le 1,$$

where $\bar{B}_{F(t)}(a, b) = \int_{F(t)}^1 x^{a-1}(1-x)^{b-1}dx$ is the incomplete beta function.

As f is decreasing in its support for a series system (that is for $i = 1$), hence

$$\frac{g_{n+1}(x)}{g_n(x)} = \frac{\bar{B}_{F(t)}(1, n)y}{\bar{B}_{F(t)}(1, n)}, \quad F(t) \le y \le 1$$

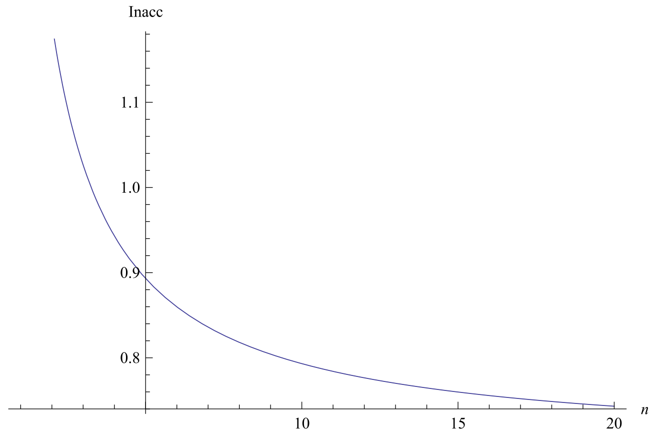

Fig. 1. Inaccuracy of the first order statistics of exponential distribution.

is a decreasing function. This implies that $X_{n+1} \overset{lr}{\leq} X_n$ which implies $X_{n+1} \overset{st}{\leq} X_n$, refer to Shaked and Shanthikumar (2007). Also it is given that $f(F^{-1}(x))$ is the decreasing function of x. Hence, for $i = 1$,

$$E_{g_1}[\log(f(F^{-1}(U_n)))] \leq E_{g_1}[\log(f(F^{-1}(U_{n+1})))].$$

Also from (11), the residual inaccuracy of the ith order statistics is

$$\begin{aligned} I_n(f_{i:n}, f; t) &= \log(\bar{F}(t)) - \frac{1}{\bar{F}_{i:n}(t)} \int_t^\infty f_{i:n}(x) \log f(x) dx \\ &= \log(\bar{F}(t)) - \int_{F(t)}^1 \frac{u^{i-1}(1-u)^{n-i} \log(f(F^{-1}(u)))du}{\bar{B}_{F(t)}(1, n)} \\ &= -E_{g_i}[\log(f(F^{-1}(U_n)))] + \log(\bar{F}(t)) \end{aligned}$$

using probability integral transformation, $U = F(X)$. Therefore, for $i = 1$ and $n \geq 1$ we have

$$\begin{aligned} I_n(f_{1:n}, f; t) - I_{n+1}(f_{1:n+1}, f; t) &= -E_{g_1}[\log(f(F^{-1}(U_n)))] + \log(\bar{F}(t)) + E_{g_1}[\log(f(F^{-1}(U_{n+1})))] - \log(\bar{F}(t)) \\ &\geq 0. \end{aligned}$$

This completes the proof.

In Fig. 1, we plot the inaccuracy of the exponential distribution with $\theta = 0.5$ of the first order statistics for $n = 1, 2, \ldots, 30$.

Remark. Note that Theorem 2.1 may not be true in general for all values of i. Let us consider $(n-2)$-out-of-n system, then lifetime of such a system is $X_{2:n}$. Consider a system with components having pdf $f(x) = \frac{2}{(x+1)^3}$, $x \geq 0$. Fig. 2 shows that inaccuracy function of $X_{2:n}$ for $t = 0.3$ and for $n = 1, 2, 3, \ldots, 10$ is an increasing function of n whereas f is decreasing. However, it can be easily checked that for the same $f(x)$ inaccuracy is a decreasing function of n for $i = 1$.

3. Characterization results

Baratpour et al. (2007, 2008) have shown that Shannon entropy and Renyi (1961) entropy of the ith order statistics characterize the distribution function uniquely. Baratpour (2010) proved that cumulative residual entropy (2004) of the first order statistics characterizes the underlying distribution function uniquely. Gupta et al. (2014) proved that the dynamic entropies of the ith order statistics characterize the distribution function uniquely. In this section we proposed measure of dynamic residual inaccuracy between the ith order statistics and parent random variable characterizes the distribution function uniquely using the sufficiency condition for the existence of unique solution of an IVP (initial value problem) given by

$$\frac{dy}{dx} = f(x, y), \quad y(x_0) = y_0,$$

where f is a function of two variables whose domain is a region $D \subset R^2$, (x_0, y_0) is a point in D and y is the unknown function. By the solution of the IVP on an interval $I \subset R$, we mean a function $\phi(x)$ such that (i) ϕ is differentiable on I, (ii) the growth of ϕ lies in D, (iii) $\phi(x_0) = y_0$ and (iv) $\phi'(x) = f(x, \phi(x))$, for all $x \in I$. The following theorem together with other results will help in proving our characterization result.

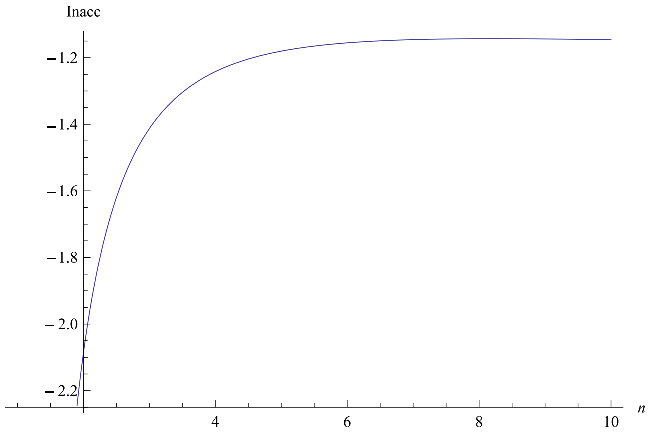

Fig. 2. Inaccuracy of the second order statistics, for $n = 1, 2, \ldots, 10$.

Theorem 3.1. *Let the function f be a continuous function defined in a domain $D \subset R^2$ and let f satisfy Lipschitz condition (with respect to y) in D, that is*

$$|f(x, y_1) - f(x, y_2)| \leq k|y_1 - y_2|, \quad k > 0, \tag{13}$$

for every point (x, y_1) and (x, y_2) in D. Then the function $y = \phi(x)$ satisfying the initial value problem $y' = f(x, y)$ and $\phi(x_0) = y_0,\ x \in I$, is unique.

We use the following lemma, refer to Gupta and Kirmani (2008), to present a sufficient condition which ensures that Lipschitz condition is satisfied in D.

Lemma 3.1. *Suppose that the function f is continuous in a convex region $D \subset R^2$. Suppose that $\frac{\partial f}{\partial y}$ exists and it is continuous in D. Then, f satisfies Lipschitz condition in D.*

Theorem 3.2. *Let X be a non-negative continuous random variable with distribution function $F(\cdot)$. Let the dynamic residual inaccuracy of the ith order statistics based on a random sample of size n be denoted by $I_n(f_{i:n}, f; t) < \infty,\ t \geq 0$. Then $I_n(f_{i:n}, f; t)$ characterizes the distribution.*

Proof. We know that

$$\begin{aligned} I_n(f_{i:n}, f; t) &= -\int_t^\infty \frac{f_{i:n}(x)}{\bar{F}_{i:n}(t)} \log\left(\frac{f(x)}{\bar{F}(t)}\right) dx \\ &= \log \bar{F}(t) - \frac{1}{\bar{F}_{i:n}(t)} \int_t^\infty f_{i:n}(x) \log f(x) dx. \end{aligned}$$

Taking derivative of both sides with respect to t, we have

$$\frac{d}{dt}[I_n(f_{i:n}, f; t)] = -\lambda_F(t) + \lambda_{F_{i:n}}(t) \left(I_n(f_{i:n}, f; t) + \log(\lambda_F(t))\right),$$

where $\lambda_F(t)$ and $\lambda_{F_{i:n}}(t)$ are the hazard rates of X and $X_{i:n}$ respectively.

Taking derivative with respect to t again and using the relation

$$\lambda_{F_{i:n}}(t) = c(t)\lambda_F(t),$$

where

$$c(t) = \left[\frac{(F(t))^{i-1}(1 - F(x))^{n-i+1}}{\bar{B}_{F(t)}(i, n - i + 1)}\right] \lambda_F(t),$$

we get

$$\lambda'_F(t) = \left(\frac{\lambda_F(t)\left[c(t)I'_n(f_{i:n}, f; t)\right] - \left[c'(t)I'_n(f_{i:n}, f; t) + c'(t)\lambda_F(t) + c^2(t)\lambda_F(t)I'_n(f_{i:n}, f; t)\right]}{c(t)\left[\lambda_F(t) + I'_n(f_{i:n}, f; t)\right]}\right). \tag{14}$$

Suppose that there are two functions F and F^* such that

$$I_n(f_{i:n}, f; t) = I_n(f^*_{i:n}, f; t) = h(t), \quad \text{say}.$$

Then for all t, from (14) we get

$$\lambda'_F(t) = \psi(t, \lambda_F(t)), \qquad \lambda'_{F^*}(t) = \psi(t, \lambda_{F^*}(t)),$$

where

$$\psi(t, y) = \left(\frac{y\left[c(t)h''(t) - \left(c'(t)h'(t) + c'(t)y + c''(t)yh'(t)\right)\right]}{c(t)\,(yc(t) + h'(t))} \right).$$

Using Theorem 3.1 and Lemma 3.1 we have, $\lambda_F(t) = \lambda_{F^*}(t)$, for all t. The fact that the hazard rate function characterizes the distribution function uniquely, we get the desired result.

Further by considering a relation between dynamic residual inaccuracy of the first order statistics and hazard rate function, we characterize some specific lifetime distributions. We give the following results:

Theorem 3.3. *Let X be a non-negative continuous random variable with distribution function $F(\cdot)$. Let the dynamic residual inaccuracy of the first order statistics based on a random sample of size n be denoted by $I_n(f_{1:n}, f; t) < \infty,\ t \geq 0$. Let $\lambda_F(t)$ be the hazard rate function of X and let*

$$I_n(f_{1:n}, f; t) = c - \log \lambda_F(t), \tag{15}$$

where c is a constant. Then X has

(i) *an exponential distribution iff $c = \frac{1}{n}$,*
(ii) *a Pareto distribution iff $c > \frac{1}{n}$,*
(iii) *a finite range distribution iff $c < \frac{1}{n}$.*

Proof. Let us assume that

$$I_n(f_{1:n}, f; t) = c - \log \lambda_F(t).$$

Taking derivative with respect to t on both sides of the above equation, we have

$$\frac{d}{dt}[I_n(f_{1:n}, f; t)] = -\lambda_F(t) + \lambda_{F_{1:n}}(t)\,(I_n(f_{1:n}, f; t) + \log(\lambda_F(t)))\,, \tag{16}$$

where $\lambda_F(t)$ and $\lambda_{F_{1:n}}(t)$ are the hazard rates of X and $X_{1:n}$ respectively. It is easy to see that $\lambda_{F_{1:n}}(t) = n\lambda_F(t)$. Using $I_n(f_{1:n}, f; t) = c - \log \lambda_F(t)$ and putting the value of $\lambda_{F_{1:n}}(t)$, (16) reduces to

$$-\lambda'_F(t) = (nc - 1)\lambda_F^2(t).$$

The solution of this differential equation is given by

$$\lambda_F(t) = \frac{1}{at + b}, \tag{17}$$

where $a = (nc - 1)$ and $b = \lambda_F 0$.

(i) If $c = \frac{1}{n}$, then $a = 0$ and from (17) $\lambda_F(t)$ turns out to be a constant, which is possible if X has exponential distribution.
(ii) If $c > \frac{1}{n}$, then $a > 0$, and (17) becomes the hazard rate function of the Pareto distribution.
(iii) If $c < \frac{1}{n}$, then $a < 0$ and (17) becomes the hazard rate function of the finite range distribution.
The only if part of this theorem is easy to prove.

4. Conclusion and comments

The concept of entropy as studied by Shannon (1948) in information theory plays a crucial role in many applications. For a system, which is observed at time t, the residual entropy measures the uncertainty about the remaining life of the distribution. Information measures based on order statistics are crucial for measuring uncertainty in statistical modeling. We have studied the dynamic measure of inaccuracy for the first and ith order statistics and have shown that these dynamic information measures uniquely determine the distribution function. The results studied in this paper can be useful for further exploring the concept of information measures based on order statistics.

Acknowledgments

Authors are thankful to the learned referees for their valuable suggestions. Also, the first author is thankful to Council of Scientific and Industrial Research, India, for providing financial assistance for this work under grant no. 08/133(0005)/2010-EMR-I.

References

Arnold, B.C., Balakrishnan, N., Nagaraja, H.N., 1992. A First Course in Order Statistics. John Wiley and Sons.
Baratpour, S., 2010. Characterizations based on cumulative residual entropy of first order statistics. Comm. Statist. Theory Methods 39 (20), 3645–3651.
Baratpour, S., Ahmadi, J., Arghami, N.R., 2007. Some characterizations based on entropy of order statistics and record values. Comm. Statist. Theory Methods 36, 47–57.
Baratpour, S., Ahmadi, J., Arghami, N.R., 2008. Characterizations based on Renyi entropy of order statistics and record values. J. Statist. Plann. Inference 138, 2544–2551.
David, H.A., Nagaraja, H.N., 2003. Order Statistics. Wiley, New York.
Ebrahimi, N., 1996. How to measure uncertainty in the residual lifetime distributions. Sankhyā Ser. A 58, 48–57.
Ebrahimi, N., Soofi, E.S., Zahedi, H., 2004. Information properties of order statistics and spacings. IEEE Trans. Inform. Theory 50, 177–183.
Gupta, R.C., Kirmani, S.N.U.A., 2008. Characterizations based on conditional mean function. J. Statist. Plann. Inference 138, 964–970.
Gupta, R.C., Taneja, H.C., Thapliyal, R., 2014. Stochastic comparisons based on residual entropy of order statistics and some characterization results. J. Stat. Theory Appl. 13 (1), 27–37.
Kerridge, D.F., 1961. Inaccuracy and inference. J. R. Stat. Soc. Ser. B 23, 184–194.
Kullback, S., 1959. Information Theory and Statistics. Wiley, New York.
Renyi, A., 1961. On measures of entropy and information. In: Proc. Fourth. Berkley Symp. Math. Stat. Prob. 1960, Vol. I. University of California Press, Berkley, pp. 547–561.
Shaked, M., Shanthikumar, J.G., 2007. Stochastic Orders. Springer Verlag.
Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423. and 623–656.
Taneja, H.C., Kumar, V., Srivastava, R., 2009. A dynamic measure of inaccuracy between two residual lifetime distributions. Int. Math. Forum 4 (25), 1213–1220.
Thapliyal, R., Taneja, H.C., 2013. A measure of inaccuracy in order statistics. J. Stat. Theory Appl. 12 (2), 200–207.
Zarezadeh, S., Asadi, M., 2010. Results on residual renyi entropy of order statistics and record values. Inf. Sci. 180 (21), 4195–4206.