Volume 66, Issue 1, 2022
Journal of Scientific Research
of
The Banaras Hindu University

Model Discrimination Based on Record Values

Ayush Tripathi*[1], Umesh Singh[2], Sanjay Kumar Singh[3]

*[1]Department of Mathematics and Statistics, Banasthali Vidyapith, Rajasthan-304022, India, tripathiayush0@gmail.com
[2]Department of Statistics, Banaras Hindu University, Varanasi-221005, U.P., India, usingh_52@yahoo.co.in
[3]Department of Statistics, Banaras Hindu University, Varanasi-221005, U.P., India, singhsk64@gmail.com

***Abstract*—In this study, the problem of model discrimination is discussed when only record data are available. The models considered are inverse Rayleigh and inverted exponential distribution. The ratio of the maximized likelihoods is used for discriminating between the two distribution functions. The probability of correct selection and power of the test are obtained through Monte Carlo simulation method for different sample sizes of record values. The problem is discussed for both lower as well as upper records. Real datasets are considered to illustrate the applicability of the proposed method.**

***Index Terms*—Inverse Rayleigh (IR) distribution, lower and upper record values, inverted exponential (IE) distribution, likelihood ratio (LR) test, probability of correct selection (PCS) and power of the test.**

I. Introduction

In many fields like weather, economics, sports, hydrology, seismology, mining, industry and life testing studies, it is quite natural to observe record values only; e.g. in sports, record-breaking performances are of interest; in weather studies, records of highest or lowest rainfall, highest flood level, highest and lowest temperature are often analyzed for further predictions and other related inferences. Usually, in such situations the researcher is left with only record-breaking observations which are a subset of complete observations; hence, the sample size considerably reduces in comparison to the size of complete observations. Record values have great importance in industry and reliability statistics; particularly in those situations where measurements of failure times are made sequentially but only the record values (lower or upper; as the case may be) are noted. (Chandler, 1952) was the first, who introduced the basic theory of records. After that, a considerable number of authors have studied records and associated inferential problems. For more details related to the problems associated with record values, readers may refer (Galambos, 1978), (Nevzorov, 1987), (Bunge & Nagaraja, 1992), (Arnold, Balakrishnan, & Nagaraja, 1998), (Balakrishnan & Chan, 1998), (Raqab & Ahsanullah, 2001), (Ahsanullah & Nevzorov, 2015), (Shafay, Balakrishnan, & Ahmadi, 2017), (Tripathi, Singh, & Singh, 2019), (Tripathi, Singh, & Kumar Singh, 2021).

Mostly, the problems discussed regarding record values are generalizations, characterization and point estimation problems; but it seems that model discrimination based on record values has not been attempted. No doubt, the problem of model discrimination on the basis of a given sample of random observations is very old and has been discussed by various authors including (Cox, 1961), (Cox, 1962), (Jackson, 1968), (Atkinson, 1969), (Atkinson, 1970) and (Dyer, 1973) and many more. The consequences of choosing the wrong model have been studied by (Cox, 1961). The problem of model discrimination between Weibull and generalized exponential was attempted by (Gupta & Kundu, 2003, 2004). It is well known that a random sample of large size provides a small number of records. Thus, it is natural to be curious to know whether we can choose between the two considered models based on record values that fits better than the other one for the data in hand. The simple answer to this query would be affirmative but it would be of interest to see what is the probability of correct model selection associated with the procedure developed for this purpose. Motivated by this thought, we shall try to address this problem by considering the availability of record values coming from either of the two distribution functions belonging to the same family of distributions and then wish to decide to which model the given sample of record values are expected to belong.

The specific aim of this paper is to propose the procedure of model discrimination between inverse Rayleigh (IR) and inverted exponential (IE) models, both of which are the members of the inverse family of continuous distributions and are special cases of inverse Weibull distribution. It is further aimed to study the effect of the type of records on PCS and power of the test. For the purpose of choosing between the considered models, use of the ratio of maximized likelihoods based on records (lower as well as upper) is proposed. The property of the proposed method is studied through a simulation study.

It may be worthwhile to mention here that the models considered in this study belong to the inverse Weibull family of distributions. If we take shape parameter equal to 1 in the probability distribution function (pdf) of inverse Weibull distribution (IWD), we get the pdf of IE distribution. Similarly,

DOI: 10.37398/JSR.2022.660144

when we take shape parameter equal to 2, we get the pdf of the IR distribution. It may also be noted that IR distribution is one of the lifetime distributions which gained enough popularity in recent years, in reliability studies and survival analysis. (Voda, 1972) studied this distribution and commented that in a large number of situations, the lifetimes of experimental units can be approximated by the IR distribution. Bayesian prediction bounds for the s^{th} future record value has been suggested by (AL-Hussaini & Ahmad, 2003). (Dey, 2005) obtained Bayes estimator of the unknown parameter and reliability function and also constructed the HPD intervals for the parameter and reliability function. Bayesian and non-Bayesian estimation of the parameter of the IR distribution and Bayesian prediction based on lower record values has been discussed by (Soliman, Amin, & Abd-El Aziz, 2010). (Dey, 2012) presented the Bayes estimators of an IR distribution under squared error loss and linear exponential loss functions. The other distribution function considered by us is IE distribution which has also wide applicability in life testing and reliability theory. IE distribution was introduced by (Keller, Kamath, & Perera, 1982). (Lin, Duran, & Lewis, 1989) suggested that the IE distribution may be used as appropriate lifetime distribution for those situations in which hazard rate is non-monotone type. (Dey, 2007) has discussed the Bayes estimator of IE distribution under squared error and LINEX loss functions. (Singh, Singh, & Kumar, 2013) have developed the Bayes estimators of the parameters and reliability estimation procedure under the general entropy loss function for complete as well as type-I and type-II censored sample for IE distribution. IE distribution has no finite moments i.e. expectation and variance of the IE distribution do not exist.

The rest of the paper is organized as follows. Section II discusses the ML estimation method for the parameters of the considered models based on lower and upper records. The model discrimination procedure for assumed models based on lower as well as upper records through the likelihood ratio (LR) test is described in Section III. This section also contains simulation results regarding PCS and power of the test. Real datasets are considered for illustrative purposes in Section IV. Section V contains some important conclusions.

II. Maximum Likelihood Estimation Based on Record Values

The distributions under consideration here are IR and IE. The probability density function (pdf) and cumulative distribution function (cdf) of the IR distribution with scale parameter θ, are given in Eq. (1) and (2) respectively.

$$f(x) = \left(\frac{2\theta}{x^3}\right) exp\left(-\frac{\theta}{x^2}\right) \qquad x, \theta > 0, \tag{1}$$

$$F(x) = exp\left(-\frac{\theta}{x^2}\right) \qquad x, \theta > 0. \tag{2}$$

and the probability density function (pdf) $f(x)$ and cumulative distribution function (cdf) $F(x)$ of IE model are given as

$$f(x) = \frac{1}{\delta x^2} exp\left(-\frac{1}{\delta x}\right) \qquad x, \delta > 0, \tag{3}$$

$$F(x) = exp\left(-\frac{1}{\delta x}\right) \qquad x, \delta > 0, \tag{4}$$

The pdf and hazard functions of these distribution are plotted in Fig.1 and Fig.2 respectively. These figures indicate that the behavior of these models are quite similar. The shape of hazard functions of both of the models shows similar pattern, initially the failure rate is increasing and after some point it starts declining. So, these models are considered here to test the suitability of the discriminating procedure when only records are observed.

Fig. 1: Probability Density Functions of IR and IE Models

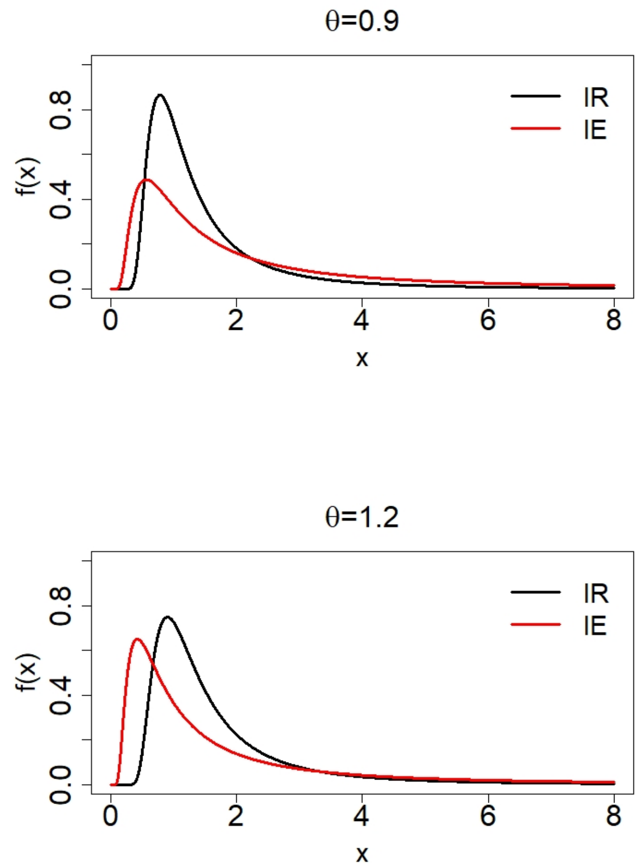
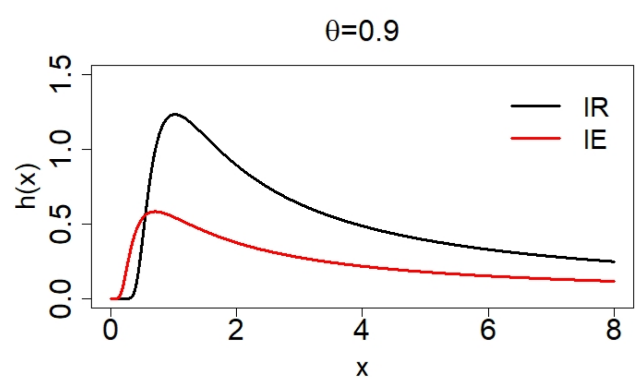
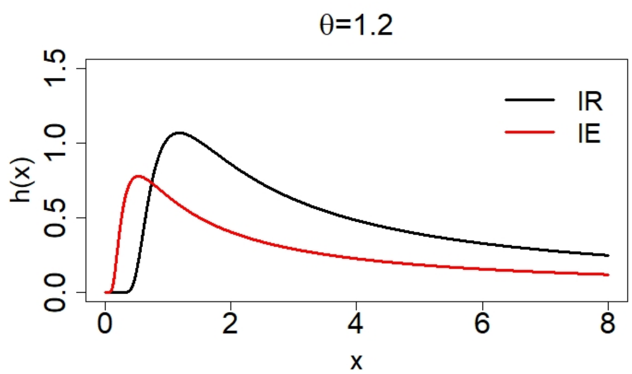


Fig. 2: Hazard Functions of IR and IE Models





Let X_1, X_2, X_3, ...be a sequence of independent observations of a random variable X with probability density function (pdf) $f(x)$ and cumulative distribution function (cdf) $F(x)$. An observation X_j is called a record value (or simply a record) if its value is smaller than or greater than all the preceding observations. In particular, if it is smaller than all the preceding observations, it is called lower record. Hence X_j is a lower record if $X_j < X_i$ for every $i < j$. An analogous definition can be provided for the upper record. Let $r_{l(1)}, r_{l(2)}, r_{l(3)}, \ldots, r_{l(m)}$ be the m lower records, then the likelihood function based on these, is given by (see (Arnold et al., 1998)).

$$L(\theta; r) = f(r_{l(m)}) \prod_{i=1}^{m-1} \frac{f(r_{l(i)})}{F(r_{l(i)})}. \tag{5}$$

Similarly, suppose $R_{u(1)}, R_{u(2)}, \ldots, R_{u(n)}$ is the set of n upper records, then the likelihood function based on these is given by

$$L(\theta; R) = f(R_{u(n)}) \prod_{i=1}^{n-1} \frac{f(R_{u(i)})}{1 - F(R_{u(i)})}. \tag{6}$$

A. *Parameter Estimation for Inverse Rayleigh Distribution*

In this section, we consider the parameter estimation for IR distribution. We obtain the ML estimates of the parameter based on lower as well as upper records. The procedure is explained in the following subsections.

1) Maximum Likelihood Estimation Based on Lower Record Values: Let $r_{l(1)}, r_{l(2)}, r_{l(3)}, \ldots, r_{l(m)}$ be the m lower records, arising from the IR distribution having probability density function (1) and distribution function (2). Hence likelihood function can be easily obtained by using Eq. (5), (1) and (2) as

$$L(\theta; r) = \left(\frac{2\theta}{r^3_{l(m)}}\right) exp\left(\frac{-\theta}{r^2_{l(m)}}\right) \prod_{i=1}^{m-1} \left(\frac{2\theta}{r^3_{l(i)}}\right). \tag{7}$$

Taking the logarithm of the likelihood function, we get

$$ln(L(\theta; r)) = ln(2\theta) - 3ln(r_{l(m)}) - \left(\frac{\theta}{r^2_{l(m)}}\right) + \sum_{i=1}^{m-1} \left(ln\left(\frac{2\theta}{r^3_{l(i)}}\right)\right),$$

now differentiating it with respect to the parameter θ and equating to zero, we get the likelihood equation which can be easily solved for θ to get the maximum likelihood (ML) estimate, say $\hat{\theta}_l$, as

$$\hat{\theta}_l = (m) r^2_{l(m)}. \tag{8}$$

2) Maximum Likelihood Estimation Based on Upper Record Values: Let $R_{u(1)}, R_{u(2)}, \ldots, R_{u(n)}$ be the set of upper records of size n from IR distribution. Likelihood function can be obtained by using Eq. (6), (1) and (2) as follows:

$$L(\theta; R) = (2\theta/R^3_{u(n)}) exp(-\theta/R^2_{u(n)}) \prod_{i=1}^{n-1} \left(\frac{(2\theta/R^3_{u(i)}) exp(-\theta/R^2_{u(i)})}{(1 - exp(-\theta/R^2_{u(i)}))}\right). \tag{9}$$

After taking the natural logarithm of the likelihood function (9), we get the log-likelihood as

$$ln(L(\theta; R)) = (n) ln(2\theta) - 3\sum_{i=1}^{n} ln(R_{u(i)}) - \sum_{i=1}^{n}\left(\frac{\theta}{R^2_{u(i)}}\right) - \sum_{i=1}^{n-1} ln(1 - exp(-\theta/R^2_{u(i)})), \tag{10}$$

differentiating Eq. (10) with respect to the parameter θ and equating it to the zero, we get the likelihood equation for θ as

$$\frac{n}{\theta} - \sum_{i=1}^{n}\left(\frac{1}{R^2_{u(i)}}\right) - \sum_{i=1}^{n-1}\left(\frac{exp(-\theta/R^2_{u(i)})}{(1 - exp(-\theta/R^2_{u(i)}))R^2_{u(i)}}\right) = 0. \tag{11}$$

Since Eq. (11) does not provide a closed form solution for θ, numerical techniques can be used to solve it. The solution, thus, obtained is the maximum likelihood estimate of the parameter, say $\hat{\theta_u}$, on the basis of n upper record values. It is to be noted here that the estimator based on lower record values is in closed form while exact expression for the estimator based on upper records does not exist.

B. *Parameter Estimation for Inverted Exponential Distribution*

This section deals with the ML estimation of the parameter of IE distribution based on lower and upper records. The methodology for the same is described below.

1) Maximum Likelihood Estimation Based on Lower Record Values: Suppose $r_{l(1)}, r_{l(2)}, r_{l(3)}, \ldots, r_{l(m)}$ are m lower records from IE distribution function. So the likelihood function of the given m lower records is given by

$$L(\delta; r) = \left(\frac{1}{\delta r_{l(m)}^2}\right) exp\left(\frac{-1}{\delta r_{l(m)}}\right) \prod_{i=1}^{m-1} \left(\frac{1}{\delta r_{l(i)}^2}\right). \quad (12)$$

After taking the logarithm of the likelihood function, we get

$$\ln(L(\delta; r)) = -ln(\delta) - 2ln(r_{l(m)}) - \left(\frac{1}{\delta r_{l(m)}}\right) + \sum_{i=1}^{m-1} \left(ln\left(\frac{1}{\delta r_{l(i)}^2}\right)\right), \quad (13)$$

on differentiating $\ln(L(\delta; r))$ with respect to the parameter δ and equating to zero, we get the maximum likelihood equation; the solution of which for δ, say $\hat{\delta}_l$, called maximum likelihood estimator based on m lower record values is obtained as

$$\hat{\delta}_l = \left(\frac{1}{m r_{l(m)}}\right). \quad (14)$$

2) Maximum Likelihood Estimation Based on Upper Record Values: Suppose $R_{u(1)}, R_{u(2)}, \ldots, R_{u(n)}$ are n upper records from the IE distribution. Likelihood function based on these is given by

$$L(\delta; R) = \left(\frac{1}{\delta R_{u(n)}^2}\right) exp\left(\frac{-1}{\delta R_{u(n)}}\right) \prod_{i=1}^{n-1} \left(\frac{\frac{1}{\delta R_{u(i)}^2} exp\left(\frac{-1}{\delta R_{u(i)}}\right)}{1 - exp\left(\frac{-1}{\delta R_{u(i)}}\right)}\right). \quad (15)$$

After taking the logarithm of the above equation, we can write it as

$$ln(L(\delta; R)) = -(n)ln(\delta) - 2\sum_{i=1}^{n} ln(R_{u(i)}) - \sum_{i=1}^{n} \left(\frac{1}{\delta R_{u(i)}}\right) - \sum_{i=1}^{n-1} ln\left(1 - exp\left(\frac{-1}{\delta R_{u(i)}}\right)\right), \quad (16)$$

differentiating Eq. (16) with respect to the parameter δ and equating it to zero to get the likelihood equation as

$$\frac{-n}{\delta} + \sum_{i=1}^{n} \left(\frac{1}{\delta^2 R_{u(i)}}\right) + \sum_{i=1}^{n-1} \left(\frac{exp\left(\frac{-1}{\delta R_{u(i)}}\right)}{\left(1 - exp\left(\frac{-1}{\delta R_{u(i)}}\right)\right) \delta^2 R_{u(i)}}\right) = 0. \quad (17)$$

After solving above non-linear likelihood equation, we get the ML estimate $\hat{\delta}_u$ based on upper record values. Needless to mention here again that the above equation is analytically unsolvable, however, numerical solutions can be obtained. It is interesting to note here that for both IR and IE distributions, the maximum likelihood (ML) estimators of the parameters based on lower records are in nice closed form and depends only on the lowest lower records whereas the maximum likelihood estimators of the parameters based on upper records are not obtainable in nice closed form but these utilize all the upper records. Now, we are interested in developing the model discrimination procedure for assumed models based on lower as well as upper records. For this purpose, we propose the use of ratio of the maximized likelihoods (RML), which is explained in the next section.

III. Proposed Ratio of the Maximized Likelihoods Test

In this section, we wish to develop a test procedure to decide whether the data (upper or lower records) in hand come from IR distribution or IE distribution i.e. we want to test the null hypotheses;

$$H_0 : \text{data come from IR distribution}$$
vs
$$H_1 : \text{data come from IE distribution.}$$

We know that the likelihood function is the joint probability density function of the data under the assumed model (functional form specified but the parameter treated as arbitrary constant). Thus, replacing the parameter with its MLE, we get the maximized likelihood. Suppose that L_{IR} and L_{IE} represent the maximized likelihoods for IR and IE distributions. Naturally, if L_{IR} is greater than L_{IE}, one may consider that probably the data are from IR distribution and vice-a-versa. For a similar argument see (Dumonceaux, Antle, & Haas, 1973).

Suppose $r_{l(1)}, r_{l(2)}, r_{l(3)}, \ldots, r_{l(m)}$ is the set of m lower records. We have already obtained the ML estimates of the parameters as in Eq. (8) and (14), hence T_l, the ratio of maximized likelihoods (RML) for given set of m lower records, is defined as

$$T_l = \frac{L_{IR}(\hat{\theta}_l; r)}{L_{IE}(\hat{\delta}_l; r)}. \quad (18)$$

It reduces to the following from by using Eq. (7) and (12):

$$T_l = \frac{\left(\frac{2\hat{\theta}_l}{r_{l(m)}^3}\right) exp\left(\frac{-\hat{\theta}_l}{r_{l(m)}^2}\right) \prod_{i=1}^{m-1} \left(\frac{2\hat{\theta}_l}{r_{l(i)}^3}\right)}{\left(\frac{1}{\hat{\delta}_l r_{l(m)}^2}\right) exp\left(\frac{-1}{\hat{\delta}_l r_{l(m)}}\right) \prod_{i=1}^{m-1} \left(\frac{1}{\hat{\delta}_l r_{l(i)}^2}\right)}. \quad (19)$$

Substituting the ML estimates of the parameters from Eq. (8) and (14), in Eq (19), the simplified form of the ratio of the maximized likelihoods is obtained as follows

$$T_l = 4 \times r_{l(m)} \prod_{i=1}^{m-1} \left(\frac{1}{r_{l(i)}}\right). \quad (20)$$

It is to be noted that the ratio of maximized likelihoods (RML) is independent of the parameters and depends only on the records (lower). Hence T_l can be used as the test statistics and the following test procedure can be applied. If test statistic $T_l > 1$, we choose IR distribution: otherwise, we choose IE distribution as the preferred model.

Now, we discuss the procedure based on upper records for the hypotheses as mentioned earlier. Suppose $R_{u(1)}, R_{u(2)}, R_{u(3)}, \ldots, R_{u(n)}$ are the n upper records available to us. ML estimates ($\hat{\theta_u}$, $\hat{\delta_u}$) of the parameters (θ, δ) are obtained by solving the non-linear Eq. (11) and (17) respectively and the ratio of the two maximized likelihoods can be written as

$$T_u = \frac{L_{IR}(\hat{\theta}_u; R)}{L_{IE}(\hat{\delta}_u; R)}. \tag{21}$$

Using Eq. (9) and (15), it can be written as follows:

$$T_u = \frac{(2\hat{\theta}_u/R^3_{u(n)})exp(-\hat{\theta}_u/R^2_{u(n)})}{\left(\frac{1}{\hat{\delta}_u R^2_{u(n)}}\right) exp\left(\frac{-1}{\hat{\delta}_u R_{u(n)}}\right)} \times \frac{\prod_{i=1}^{n-1}\left(\frac{(2\hat{\theta}_u/R^3_{u(i)})exp(-\hat{\theta}_u/R^2_{u(i)})}{(1-exp(-\hat{\theta}_u/R^2_{u(i)}))}\right)}{\prod_{i=1}^{n-1}\left(\frac{\frac{1}{\hat{\delta}_u R^2_{u(i)}} exp\left(\frac{-1}{\hat{\delta}_u R_{u(i)}}\right)}{1-exp\left(\frac{-1}{\hat{\delta}_u R_{u(i)}}\right)}\right)}. \tag{22}$$

As mentioned earlier, the ML estimates for the parameters of IR and IE distributions based on upper records are not expressible in closed form; hence, the exact expression for RML (T_u) based on upper records can not be obtained. However, numerical methods can be used to obtain its value for given values of the records. The test procedure for testing the hypotheses can be proposed parallel to that based on lower record i.e. if the resulting test statistic $T_u > 1$, select IR distribution; otherwise, select IE distribution as the parent distribution.

A. Simulation Study

In this section, a simulation study is done to study the behavior of the proposed RML (T) test statistic for different choices of parameters and sample (records sample) sizes. The probability of correct selection (PCS) and power of the test are obtained based on lower as well as upper records. Numerically generated record data, from the distributions under consideration are used to calculate PCS which is defined as probability of accepting H_0 when H_0 is true. It is well known that probability of Type-I error (α) is defined as probability of rejecting H_0 when H_0 is true. Therefore, α = 1 - PCS. Probability of Type-II error (β) is defined as probability of accepting H_0 when H_0 is false and power of the test is defined as probability of rejecting H_0 when H_0 is false i.e. 1 - P[Type-II error]. If PCS is observed to be high, it can be concluded that we are making correct decision for choosing the appropriate model in the light of the data. In order to check whether the number of records have any effect on PCS and power of the test, the study has been carried out for varying number of records. For all the computations, the codes are developed on statistical software package R.

For the computation of the simulated value of PCS and power of the proposed test procedure, 30,000 independent complete samples are generated from IR distribution for considered values of the parameter θ (= 0.4, 0.6, 1.1), then lower records are generated from these. A similar procedure is followed for generating the upper records. The above mentioned procedure is also followed for generating the lower and upper records from IE distribution. The values of the parameter δ considered by us are 0.4, 0.6 and 1.1. From the samples obtained through the method described above, The MLE of the parameters based on lower records are calculated from expression given in Eq. (8) and (14). Then the proposed ratio of maximized likelihoods T_l is calculated. Now, as mentioned earlier, PCS is obtained as the ratio of number of times $T_l > 1$ to the total cases when the samples are generated from IR distribution. For calculating power of the test, samples generated from IE distribution are considered.

The above explained process is repeated for different sample sizes of records (k). Various values of k considered here are 7, 8, 9, 10 and 11. A similar procedure is followed for upper records also with the only difference that the MLEs for upper records are calculated numerically as mentioned in subsections II-A2 and II-B2.

Further, ML estimators of the parameter ($\theta = 0.5$) with their MSEs and biases for IR and IE distributions are recorded in Table I and Table II respectively. From both the tables, it can be seen that lower records provides better estimate in comparison to the upper records for both the distributions in terms of the MSEs and biases. It is also noted that as the number of records (k) increases, MSEs and biases of the ML estimator decreases for the considered distributions.

Table III provides PCS and Power of the test for different values of θ and δ and varying number of records (k) based on lower as well as upper records. From the Table III, it is observed that the PCS as well as power of the test increases as the number of records increases. For the test based on lower records, the PCS increases from approximately 92% at 7 records to 93% at 11 records, whereas, for upper records the PCS increase from approximately 98% to 99%. We observed that the PCS and power of the test do not show much variation with the change in values of θ and δ. This shows that the choice of the parameter does not affect the suitability of the method for model discrimination. This method provides approximately similar values of PCS and power of the test for particular number of records in the sample irrespective of the values of parameter.

TABLE I: Avg. Estimates, MSEs and Biases for IR distribution when $\theta = 0.5$.

	Lower			Upper		
No. of Records (k)	Avg. Estimate	MSE	Bias	Avg. Estimate	MSE	Bias
7	0.7925	0.1685	0.2945	0.8835	2.1590	0.5828
8	0.7724	0.1314	0.2732	0.8818	2.1296	0.5812
9	0.7546	0.1076	0.2550	0.8809	2.1110	0.5803
10	0.7383	0.0911	0.2388	0.8806	2.1070	0.5800
11	0.7119	0.0757	0.2155	0.8806	2.1041	0.5799

TABLE II: Avg. Estimates, MSEs and Biases for IE Distribution when $\theta = 0.5$.

	Lower			Upper		
No. of Records (k)	Avg. Estimate	MSE	Bias	Avg. Estimate	MSE	Bias
7	0.3436	0.0327	0.1583	0.6497	2.1720	0.4491
8	0.3459	0.0304	0.1548	0.6493	2.1691	0.4488
9	0.3486	0.0286	0.1518	0.6490	2.1666	0.4487
10	0.3545	0.0263	0.1462	0.6489	2.1658	0.4486
11	0.3684	0.0236	0.1360	0.6488	2.1656	0.4486

TABLE III: PCS and Power of the Test

		No. of Records (k)	PCS $\theta = 0.4$	PCS $\theta = 0.6$	PCS $\theta = 1.1$
Type of Records	Lower	7	0.9214	0.9199	0.9192
		8	0.9256	0.9237	0.9241
		9	0.9300	0.9271	0.9281
		10	0.9329	0.9309	0.9313
		11	0.9350	0.9330	0.9344
	Upper	7	0.9863	0.9814	0.9734
		8	0.9975	0.9945	0.9937
		9	0.9995	0.9992	0.9984
		10	0.9998	0.9999	0.9997
		11	0.9999	1.0000	0.9999
			Power $\delta = 0.4$	Power $\delta = 0.6$	Power $\delta = 1.1$
Type of Records	Lower	7	0.6443	0.6479	0.6429
		8	0.6814	0.6875	0.6795
		9	0.7141	0.7194	0.7122
		10	0.7433	0.7463	0.7435
		11	0.7728	0.7740	0.7709
	Upper	7	0.5374	0.5298	0.5043
		8	0.5816	0.5763	0.5524
		9	0.6373	0.6328	0.6145
		10	0.7204	0.7167	0.7041
		11	0.8384	0.8390	0.8337

IV. Illustrative Examples

Two real datasets are considered for illustration purpose in this section. We have considered first dataset which is the March precipitation data (in inches) for Minneapolis/St. Paul over a period of 30 years. This dataset is analyzed by (Hinkley, 1977). The complete dataset is given below.

TABLE IV: Dataset 1

0.77	1.74	0.81	1.2	1.95	1.2	0.47	1.43
3.37	2.2	3	3.09	1.51	2.1	0.52	1.62
1.31	0.32	0.59	0.81	2.81	1.87	1.18	1.35
4.75	2.48	0.96	1.89	0.9	2.05		

K-S test is applied to check the fitting of IR and IE distributions. K-S distance and p-value for IR distribution are 0.23969 and 0.06368 respectively. It indicates that the IR distribution is the appropriate model for the given Dataset 1. Fig.3 shows that the IR distribution fits to the considered real dataset. The lower and upper records are extracted from the above complete Dataset 1, reported in the Table V.

TABLE V: Generated Records from Dataset 1

Lower records:	0.77	0.47	0.32		
Upper records:	0.77	1.74	1.95	3.37	4.75

We consider another real dataset of failure times of the air-conditioning system of an airplane. This data set is considered by (Linhart & Zucchini, 1986). Complete dataset is reproduced below.

TABLE VI: Dataset 2

23	261	87	7	120	14	62	47	225	71
246	21	42	20	5	12	120	11	3	14
71	11	14	11	16	90	1	16	52	95

To check the fitting of the considered distributions, we have applied K-S test. The numerical values of K-S distance and p-value for IE distribution are 0.23296 and 0.07706 respectively. Therefore it is suggested that IE distribution is the suitable model for the Dataset 2. Fig.4 shows the fitting of the IE distribution for the given Dataset 2. The observed lower and upper records from the Dataset 2 are as,

TABLE VII: Generated Records from Dataset 2

Lower records:	23	7	5	3	1
Upper records:	23	261			

After analyzing the two real datasets, our primary objective is to select a suitable model for both datasets when we have only record values. So we apply a proposed likelihood ratio test for both the datasets. We have obtained RML by considering both the distributions for lower as well as upper records. Results are reported in Table VIII. From Table VIII, it is observed that the values of RML is greater than 1 when it is assumed that records are coming from IR distribution. This supports our assumption of IR being the null distribution. Similarly, for Dataset 2, RML test suggests IE as the preferred distribution (to calculate RML, we take IE as the null distribution). This leads to the conclusion that IR distribution is the appropriate model for Dataset 1 and IE distribution is the preferred model for Dataset 2.

TABLE VIII: Ratio of Maximized Likelihood (RML)

			Likelihood Ratio $\left(\frac{L_1}{L_2}\right)$	Decision
Dataset 1	Null: IR	Lower	$2.2636\ (> 1)$	IR model is preferred
		Upper	$8.2274\ (> 1)$	IR model is preferred
Dataset 2	Null: IE	Lower	$75.4687\ (> 1)$	IE model is preferred
		Upper	$2.5152\ (> 1)$	IE model is preferred

Fig. 3: Dataset 1

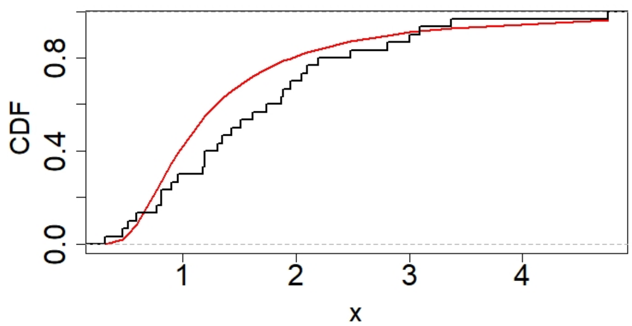
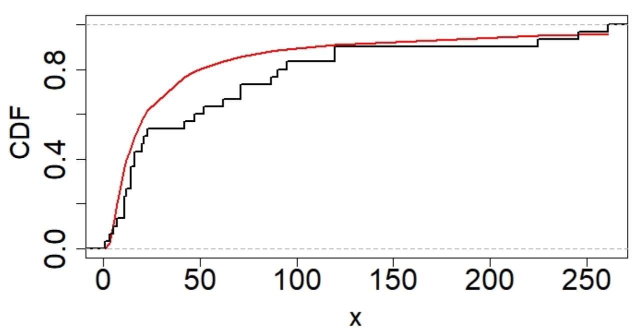


Fig. 4: Dataset 2



V. Conclusion

The aim of this article is to develop a discrimination procedure between two distribution functions on the basis of the availability of the record values. The proposed procedure is based on the ratio of maximized likelihoods which selects a model having larger maximized likelihood. The two distribution functions of inverse family distribution namely: inverse Rayleigh and inverted exponential are considered. From the discussion of the results mentioned above, we may conclude that the proposed model discrimination procedure can be used for discriminating between the IR and IE distributions on the basis of record values (lower as well as upper record), because the PCS are found to be satisfactorily high. ML estimation for the parameter of IR and IE model based on upper record values is attempted for the first time in this paper and we observe that the PCS and power of the proposed test procedure are larger for the upper record than those based on lower record. Therefore, if both types of records are available, it is better to use upper records.

References

Ahsanullah, M., & Nevzorov, V. B. (2015). *Records via probability theory* (Vol. 6). Atlantis Press.

AL-Hussaini, E. K., & Ahmad, A. E.-B. A. (2003). On bayesian predictive distributions of generalized order statistics. *Metrika*, *57*(2), 165 -176.

Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (1998). Records. *Wiley Series in Probability and Statistics*, 313 -319.

Atkinson, A. C. (1969). A test for discriminating between models. *Biometrika*, *56*(2), 337 -347.

Atkinson, A. C. (1970). A method for discriminating between models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 323 -353.

Balakrishnan, N., & Chan, P. S. (1998). On the normal record values and associated inference. *Statistics & probability letters*, *39*(1), 73 -80.

Bunge, J. A., & Nagaraja, H. N. (1992). Exact distribution theory for some point process record models. *Advances in applied probability*, *24*(01), 20 -44.

Chandler, K. N. (1952). The distribution and frequency of record values. *Journal of the Royal Statistical Society. Series B (Methodological)*, 220 -228.

Cox, D. R. (1961). Tests of separate families of hypotheses. In *Proceedings of the fourth berkeley symposium on mathematical statistics and probability* (p. 105 -123).

Cox, D. R. (1962). Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society. Series B (Methodological)*, 406 -424.

Dey, S. (2005). Bayesian intervals for parameter and reliability of inverse rayleigh distribution. *Journal of Ravishankar University*, *18(B)*, 57 -68.

Dey, S. (2007). Inverted exponential distribution as a life distribution model from a bayesian viewpoint. *Data Science Journal*, *6*, 107 -113.

Dey, S. (2012). Bayesian estimation of the parameter and reliability function of an inverse rayleigh distribution. *Malaysian Journal of Mathematical Sciences*, *6*(1), 113 -124.

Dumonceaux, R., Antle, C. E., & Haas, G. (1973). Likelihood ratio test for discriminagon between two models with unknown location and scale parameters. *Technometrics*, *15*(1), 19 -27.

Dyer, A. R. (1973). Discrimination procedures for separate families of hypotheses. *Journal of the American Statistical Association*, *68*(344), 970 -974.

Galambos, J. (1978). *The asymptotic theory of extreme order statistics* (Vol. 352). Wiley New York.

Gupta, R. D., & Kundu, D. (2003). Discriminating between weibull and generalized exponential distributions. *Computational statistics & data analysis*, *43*(2), 179 -196.

Gupta, R. D., & Kundu, D. (2004). Discriminating between gamma and generalized exponential distributions. *Journal of Statistical Computation & Simulation*, *74*(2), 107 -121.

Hinkley, D. (1977). On quick choice of power transformation. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *26*(1), 67–69.

Jackson, O. (1968). Some results on tests of separate families of hypotheses. *Biometrika*, 355 -363.

Keller, A. Z., Kamath, A. R. R., & Perera, U. D. (1982). Reliability analysis of cnc machine tools. *Reliability Engineering*, *3*(6), 449 -473.

Lin, C. T., Duran, B. S., & Lewis, T. O. (1989). Inverted gamma as a life distribution. *Microelectronics Reliability*, *29*(4), 619 -626.

Linhart, H., & Zucchini, W. (1986). *Model selection*. John Wiley.

Nevzorov, V. B. (1987). Two characterizations using records. In *Stability problems for stochastic models* (p. 79 -85). Springer.

Raqab, M. M., & Ahsanullah, M. (2001). Estimation of the location and scale parameters of generalized exponential distribution based on order statistics. *Journal of Statistical Computation and Simulation*, *69*(2), 109 -123.

Shafay, A. R., Balakrishnan, N., & Ahmadi, J. (2017). Bayesian prediction of order statistics with fixed and random sample sizes based on k-record values from pareto distribution. *Communications in Statistics-Theory and Methods*, *46*(2), 721–735.

Singh, S. K., Singh, U., & Kumar, D. (2013). Bayes estimators of the reliability function and parameter of inverted exponential distribution using informative and non-informative priors. *Journal of Statistical Computation and Simulation*, *83*(12), 2258 -2269.

Soliman, A. A., Amin, E. A., & Abd-El Aziz, A. A. (2010). Estimation and prediction from inverse rayleigh distribution based on lower record values. *Applied Mathematical Sciences*, *4*(62), 3057 -3066.

Tripathi, A., Singh, U., & Kumar Singh, S. (2021). Estimation of P (X< Y) for gompertz distribution based on upper records. *International Journal of Modelling and Simulation*, 1–12.

Tripathi, A., Singh, U., & Singh, S. K. (2019). Inferences for the dus-exponential distribution based on upper record values. *Annals of Data Science*, 1–17.

Voda, V. G. (1972). On the inverse rayleigh distributed random variable. *Reports of Statistical Application Research*, *19*, 13–21.