Interval estimation for proportional reversed hazard family based on lower record values

Bing Xing Wang *

Zhejiang Gongshang University, China

Keming Yu

Brunel University, UK

Frank P.A. Coolen

Durham University, UK

Abstract

This paper explores confidence intervals for the family of proportional reversed hazard distributions based on lower record values. The proposed procedure can be extended to the family of proportional hazard distributions based on upper record values. Numerical results show that the method is promising.

Key words: Confidence interval, proportional reversed hazard distribution, record value, sample size.

MR Subject Classification: 62N05, 62F25

1 Introduction

An important topic in survival and reliability analyses is the study of parametric probability distributions in order to model the faults in a product or the lifetime of a product or entity. Many lifetime distributions are related to extreme values, e.g. a series system stops working when the first component breaks while a parallel system stops working when the

---

Address for correspondence: Department of Statistics, Zhejiang Gongshang University, China.
E-mail: wangbingxing@163.com

last component breaks. Moveover, in big data scenarios, which are becoming more and more relevant these days, there may be a specific interest in record values only, such as extreme weather events, and no other aspects of the data may be stored or reported.

Since Chandler (1952) introduced the topic of record values and studied their basic properties, a substantial literature has appeared devoted to record values, for example see Glick (1978), Smith (1988), Carlin and Gelfand (1993), Feuerverger and Hall (1996), Chan (1998), Sultan et al. (2008), Wong and Wu (2009), Tavangara and Asadia (2011), Cramer and Naehrig (2012). Record statistics are widely used in many real life application areas, such as weather forecast (Chandler, 1952; Coles and Tawn, 1996), maximum water levels in hydrology (Katz et al., 2002), sports and economics (Balakrishnan et al., 1993; Robinson and Tawn, 1995; Balakrishnan and Chan, 1998; Raqab, 2002; Einmahl and Magnus, 2008), life-tests (Soliman et al., 2006; Ahmadi et al., 2009), stock markets (Wergen, 2014) and so on. Due to the commonality and the importance, there has been a number of literature on probabilistic modeling and statistical inference for record data. For a book-length account on this topic, see Arnold et al. (1998) and Ahsanullah (2004).

Sample size is an important issue in statistical testing and confidence intervals because it has such a significant impact on the validity of analytic results and is so often misunderstood. Without a sufficiently large sample, a statistical test or confidence interval may not have the targeted statistical properties, if the derivation of the test procedure or interval depends on assumptions which are only asymptotically justified. Therefore, as data sets consisting of record values often lack sufficient data for statistical inference based on asymptotically justified methods, it is important to develop exact inferential methods which apply for any sample size. This paper presents a new method of exact inference for interval estimation for a family of proportional reversed hazard distributions based on data consisting of lower record values.

Let $\{X_n, n = 1, 2, ...\}$ be a sequence of independent and identically distributed (i.i.d.) random variables with cumulative distribution function (cdf) $F(x)$ and probability density function (pdf) $f(x)$. An observation X_j is called a lower record value if its value is less than the values of all of previous observations, so if $X_j < X_i$ for each $i < j$. Then the record times sequence $\{T_n, n \geq 1\}$ is defined in the following manner: $T_1 = 1$ (with probability 1) and for $n \geq 2$, $T_n = \min\{j : X_j < X_{T_{n-1}}\}$. The sequence $\{R_n = X_{T_n}, n = 1, 2, ...\}$ is called the sequence of lower record values of the original sequence.

In this paper, new exact interval estimation is presented based on record values for the following family of probability distributions, which provides a flexible family to model lifetime variables. Let $F(x; \lambda, \alpha)$ denote the cdf of a probability distribution with parameters λ and α. Consider parameter estimation for the family of probability distributions specified by

$$F(x; \lambda, \alpha) = [G(x; \lambda)]^{\alpha}, \ x > 0, \tag{1}$$

where $G(\cdot; \lambda)$ is a cdf dependent only on λ. These families of distributions — without necessarily confining attention to a one-parameter G — are discussed by Marshall and Olkin (2007, Section 7.E. & ff.). They call (1) a 'resilience parameter' or 'proportional reversed hazard' family. When α is an integer, (1) is the distribution function of the maximum of a random sample of size α from the distribution $G(\cdot; \lambda)$. Examples of families (1) include the inverse Weibull distribution, and generalized exponential distribution (Gupta and Kundu, 1999). The latter can be used as an alternative to gamma or Weibull distributions in many situations and has attracted much attention in the literature recently, it arises when $G(x; \lambda) = 1 - e^{-x/\lambda}$ in family (1).

In Section 2 of this paper exact interval estimation for the parameters λ and α is presented, as well as some characteristics of $F(x; \lambda, \alpha)$. In Section 3 the results of a simulation study in order to investigate the performance of the proposed method are presented, while an example with data from the literature is presented in Section 4.

2 Interval estimation

In this section new methods for interval estimation for the proportional inversed hazards family are presented. In order to do so, the following lemmas are needed.

Lemma 1 *Let $R_1, R_2, ..., R_n$ be the lower record values observed from the standard uniform distribution $U(0, 1)$, then $-\log(R_1), \log(R_1) - \log(R_2), ..., \log(R_{n-1}) - \log(R_n)$ are i.i.d. standard exponential random variables.*

Proof: Let $Y_1 = -\log(R_1), Y_2 = \log(R_1) - \log(R_2), ..., Y_n = \log(R_{n-1}) - \log(R_n)$. Notice that the pdf of $R_1, R_2, ..., R_n$ is given by

$$f(r_1, r_2, ..., r_n) = f(r_n) \prod_{i=1}^{n-1} f(x_i)[F(r_i)]^{-1} = \prod_{i=1}^{n-1} r_i^{-1}, \ \ 0 < r_n < ... < r_1 < 1,$$

that the Jacobian of transformation is given by

$$\frac{\partial(R_1, ..., R_n)}{\partial(Y_1, ..., Y_n)} = e^{-nY_1 - (n-1)Y_2 - ... - Y_n},$$

and the pdf of $Y_1, ..., Y_n$ is given by

$$f(y_1, ..., y_n) = e^{-y_1 - y_2 - ... - y_n}, \ y_1 > 0, ..., y_n > 0.$$

Therefore, $Y_1, ..., Y_n$ are i.i.d. standard exponential random variables.

Lemma 2 *Suppose that $Y_1, Y_2, ..., Y_n$ are i.i.d. exponential random variables with mean θ. Let $S_i = Y_1 + ... + Y_i$, $i = 1, 2, ..., n$, then $S_1/S_2, (S_2/S_3)^2, ..., (S_{n-1}/S_n)^{n-1}, S_n$ are independent random variables. Also, $S_1/S_2, (S_2/S_3)^2, ..., (S_{n-1}/S_n)^{n-1}$ have standard uniform distributions and S_n has gamma distribution with shape parameter n and scale parameter 1, denoted by $\Gamma(n, 1)$. (see Wang et al., 2010)*

2.1 Interval estimation of λ

Let $R_1, R_2 ..., R_n$ be the lower record values observed from the proportional reversed hazards family (1), then $F(R_1; \lambda, \alpha), F(R_2; \lambda, \alpha), ..., F(R_n; \lambda, \alpha)$ are the lower record values observed from the standard uniform distribution $U(0, 1)$. Thus, we have from Lemma 1 that $Y_1 = -\log F(R_1; \lambda, \alpha), Y_2 = \log F(R_1; \lambda, \alpha) - \log F(R_2; \lambda, \alpha), ..., Y_n = \log F(R_{n-1}; \lambda, \alpha) - \log F(R_n; \lambda, \alpha)$ are i.i.d. standard exponential random variables.

Notice that $-\log F(R_i; \lambda, \alpha) = -\alpha \log G(R_i; \lambda)$, and $Y_1 + ... + Y_i = -\alpha \log G(R_i; \lambda)$, so we have from Lemma 2 that $U_1, ..., U_{n-1}, U_n = -\alpha \log G(R_n; \lambda)$ are independent random variables. Also, $U_1, ..., U_{n-1}$ have standard uniform distributions and U_n has gamma distribution $\Gamma(n, 1)$, where

$$U_i = \left(\frac{\log G(R_i; \lambda)}{\log G(R_{i+1}; \lambda)} \right)^i , i = 1, 2, ..., n - 1.$$

Therefore,

$$W_1(\lambda) = -2 \sum_{i=1}^{n-1} \log(U_i) = 2 \sum_{i=1}^{n-1} \log \left(\frac{\log G(R_n; \lambda)}{\log G(R_i; \lambda)} \right) \sim \chi^2(2n - 2). \tag{2}$$

If $W_1(\lambda)$ is a strictly increasing or decreasing function of λ, which can be shown case-by-case, then, for any $0 < \beta < 1$,

$$\left[W_1^{-1}(\chi_{\beta/2}^2(2n - 2)), W_1^{-1}(\chi_{1-\beta/2}^2(2n - 2)) \right] \text{ or } \left[W_1^{-1}(\chi_{1-\beta/2}^2(2n - 2)), W_1^{-1}(\chi_{\beta/2}^2(2n - 2)) \right]$$

is a $1 - \beta$ confidence interval for λ, where $\chi^2_\beta(v)$ is the β percentile of the χ^2 distribution with v degrees of freedom and, for $t > 0$, $W_1^{-1}(t)$ is the solution in λ of the equation $W_1(\lambda) = t$. These exact confidence intervals could be simplified for each distribution of the family (1).

When $G(x; \lambda)$ is a scale distribution family, the following lemma gives a sufficient condition in which $W(\lambda)$ is a strictly decreasing function of λ.

Lemma 3 *Let $G(x; \lambda) = G_1(x/\lambda)$ and $g_1(x) = G_1'(x)$, where $G_1(x)$ and $g_1(x)$ are the known continuous functions. If the reversed failure rate function $\rho(x, \lambda) = \frac{\partial G(x;\lambda)}{\partial x} / G(x; \lambda)$ for the distribution family $G(x; \lambda)$ is a strictly decreasing function of the scale parameter λ in $(0, \infty)$, $W_1(\lambda)$ defined as (2) is then a strictly decreasing function of λ.*

Proof.   Let

$$h(t) = \frac{t g_1(t)}{G_1(t)} \quad \text{and} \quad y(\lambda) = \frac{\log G(R_n, \lambda)}{\log G(R_i, \lambda)}.$$

Notice that for the scale family $G(x; \lambda) = G_1(x/\lambda)$, we have

$$\rho(x, \lambda) = \frac{g_1(x/\lambda)}{\lambda G_1(x/\lambda)} = \frac{1}{x} \cdot h(x/\lambda). \tag{3}$$

Thus we obtain from (3) that when $\rho(x, \lambda)$ is a strictly decreasing function of λ, $h(t)$ is a strictly increasing function of t.   Notice that $G_1(t)$ is an increasing function, thus $h(t)/(-\log G_1(t))$ is a strictly increasing function. For the function $y(\lambda)$, we have

$$y'(\lambda) = \frac{\log G_1(R_n/\lambda)}{\lambda \log G_1(R_i/\lambda)} \left[\frac{h(R_n/\lambda)}{-\log G_1(R_n/\lambda)} - \frac{h(R_i/\lambda)}{-\log G_1(R_i/\lambda)} \right] < 0 \tag{4}$$

Hence $y(\lambda)$ is a strictly decreasing function of λ. Therefore, we have from (2) that $W_1(\lambda)$ is a strictly decreasing function of λ.

Example 1: The Inverse Weibull distribution

The cdf of the inverse Weibull distribution $IW(\lambda, \alpha)$ is

$$F(x; \lambda, \alpha) = e^{-(\alpha/x)^\lambda}, \quad x > 0,$$

where $\lambda > 0$ is the shape parameter and $\alpha > 0$ is the scale parameter. Note that, in the inverse Weibull case, α in (1) is reparameterized as α^λ. In this case, $W_1(\lambda)$ is equal to

$$W_1(\lambda) = 2\lambda \sum_{i=1}^{n-1} \log\left(\frac{R_i}{R_n}\right). \tag{5}$$

It is obvious that $W_1(\lambda)$ is an increasing function of λ, thus

$$\left[\frac{\chi^2_{\beta/2}(2n - 2)}{2\sum_{i=1}^{n-1} \log(R_i/R_n)}, \frac{\chi^2_{1-\beta/2}(2n - 2)}{2\sum_{i=1}^{n-1} \log(R_i/R_n)} \right]$$

is a $1 - \beta$ confidence interval for the parameter λ.

Remark 1: Similar to the proof in Wang and Ye (2015), the statistics $(\sum_{i=1}^{n-1} \log(R_i/R_n), R_n)$ are complete.

Example 2: The generalized exponential distribution

The cdf of the generalized exponential distribution $GE(\lambda, \alpha)$ is

$$F(x; \lambda, \alpha) = (1 - e^{-x/\lambda})^\alpha, \quad x > 0, \tag{6}$$

where $\alpha > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter. In this case, $W_1(\lambda)$ is equal to

$$W_1(\lambda) = 2 \sum_{i=1}^{n-1} \log \left(\frac{\log(1 - e^{-R_n/\lambda})}{\log(1 - e^{-R_i/\lambda})} \right).$$

Notice that the reversed failure rate for the generalized exponential distribution is given by

$$\rho(x, \lambda) = \frac{(x/\lambda)e^{-x/\lambda}}{x(1 - e^{-x/\lambda})}$$

and that $\frac{xe^{-x}}{1-e^{-x}}$ is a strictly decreasing function of x in $(0, \infty)$, thus we know from lemma 3 that $W_1(\lambda)$ is a strictly decreasing on $(0, \infty)$. Furthermore, we have

$$\lim_{\lambda \to 0+} W_1(\lambda) = \infty \quad \text{and} \quad \lim_{\lambda \to \infty} W_1(\lambda) = 0.$$

Thus $\left[W_1^{-1}(\chi^2_{1-\beta/2}(2n - 2)), W_1^{-1}(\chi^2_{\beta/2}(2n - 2)) \right]$ is a $1 - \beta$ confidence interval for the parameter λ. This confidence interval is the same as one based on $Q_{k,n}$ proposed by Raqab and Sultan (2014).

Remark 2: Notice that $W_1(\lambda)$ and U_n are independent, thus the $1 - \beta$ joint confidence region of (λ, α) is obtained by

$$P\left(\chi^2_{\frac{1-\sqrt{1-\beta}}{2}}(2n - 2) < W_1(\lambda) < \chi^2_{\frac{1+\sqrt{1-\beta}}{2}}(2n - 2), \chi^2_{\frac{1-\sqrt{1-\beta}}{2}}(2n) < 2U_n < \chi^2_{\frac{1+\sqrt{1-\beta}}{2}}(2n) \right) = 1 - \beta.$$

2.2 Interval estimation of α and other quantities

Below we derive generalized confidence intervals for the parameter α, mean, quantiles and reliability function of the proportional reversed hazard family.

Suppose that $W_1(\lambda)$ is a strictly monotone function of λ. Let $g(W, R)$ be the unique solution of $W_1(\lambda) = W$, where $R = (R_1, R_2, ..., R_n)$ and $W \sim \chi^2(2n - 2))$. Notice that $V_1 = 2U_n = -2\alpha \log G(R_n; \lambda)$ has the χ^2 distribution with $2n$ degrees of freedom, therefore we have that

$$\alpha = -\frac{V_1}{2 \log G(R_n; \lambda)}.$$

6

Using the substitution method presented by Weerahandi (2004), we substitute $g(W, R)$ for λ in the expression for α and obtain the following generalized pivotal quantity for α:

$$W_2 = -\frac{V_1}{2 \log G(r_n; g(W, r))} \tag{7}$$

$$= \frac{\alpha \log G(R_n; g(W, R))}{\log G(r_n; g(W, r))}, \tag{8}$$

where $r = (r_1, r_2, ..., r_n)$ is the observed value of $R = (R_1, R_2, ..., R_n)$.

It is obvious from (7) that the distribution of W_2 is free of any unknown parameters. It is also obvious from (8) that W_2 reduces to α when $R = r$. Thus W_2 is a generalized pivotal quantity. The cdf of W_2 is given by

$$\begin{aligned} F_{W_2}(w) &= \int_0^\infty P(W_2 \le w | W = x) f_W(x) dx \\ &= 1 - \int_0^\infty F_{\chi^2(2n)} \left(-2w \log G(r_n; g(x, r))\right) f_{\chi^2(2n-2)}(x) dx, \end{aligned} \tag{9}$$

where $F_{\chi^2(v)}(x)$ and $f_{\chi^2(v)}(x)$ are the cdf and the pdf of the χ^2 distribution with v degrees of freedom, respectively. Percentiles of the generalized pivotal quantity W_2 can be obtained from the cdf (9). Another way to obtain the percentiles is based on the following simulation algorithm. For a given data set (n, r), generate $W \sim \chi^2(2n - 2)$ and $V_1 \sim \chi^2(2n)$, independently. Using these values, we compute the values of W_2 in (7). This process of generating the value of W_2 is repeated $m(\ge 10,000)$ times for the fixed values of (n, r). Based on the generated values of W_2, the percentiles of W_2 can be estimated. Let $W_{2,\beta}$ denote the β percentile of W_2, then $[W_{2,\beta/2}, W_{2,1-\beta/2}]$ is a $1 - \beta$ generalized confidence interval for α.

Notice that the mean, pth quantile $(0 < p < 1)$ and reliability function of the proportional inverse hazards family are respectively given by $\mu = \int_0^\infty x dF(x; \lambda, \alpha) \equiv h(\lambda, \alpha)$, $x_p = G^{-1}[p^{1/\alpha}, \lambda]$ and $R(x_0) = [G(x_0; \lambda)]^\alpha$, where $h(\cdot, \cdot)$ is a known function, and $G^{-1}(t, \lambda)$ is the solution of $G(x; \lambda) = t$. Similar to the derivation of W_2 for α, we obtain the following generalized pivotal quantities W_3, W_4 and W_5 for μ, x_p and $R(x_0)$ respectively:

$$\begin{aligned} W_3 &= h\left(g(W, r), W_2\right), \\ W_4 &= G^{-1}\left(p^{1/W_2}, g(W, r)\right), \\ W_5 &= [G(x_0; g(W, r))]^{W_2}. \end{aligned}$$

Similar to W_2, the cdf of W_4 is given by

$$F_{W_4}(w) = 1 - \int_0^\infty F_{\chi^2(2n)} \left(-2G(w; g(x, r)) \log G(r_n; g(x, r)) / \log(p)\right) f_{\chi^2(2n-2)}(x) dx.$$

7

Let $W_{3,\beta}, W_{4,\beta}$ and $W_{5,\beta}$ denote the β percentiles of W_3, W_4 and W_5, respectively. Then $W_{3,\beta}, W_{4,\beta}, W_{5,\beta}$ are the $1 - \beta$ lower confidence limits for μ, x_p and $R(x_0)$, respectively. Just as in the case of W_2, the percentiles of W_3, W_4, W_5 can also be obtained by Monte Carlo simulations.

Example 1 continued

For the inverse Weibull distribution, $\mu = \alpha\Gamma(1 - \lambda^{-1})$ when $\lambda > 1$, $x_p = \alpha(-\log p)^{-1/\lambda}$ and $R(x_0) = 1 - e^{-(\alpha/x_0)^\lambda}$, thus we have

$$W_2 = \left(\frac{V_1}{2}\right)^{\frac{2\sum_{i=1}^{n-1}\log(r_i/r_n)}{W}} r_n,$$

$$W_3 = W_2 \cdot \Gamma\left(1 - \frac{2\sum_{i=1}^{n-1}\log(r_i/r_n)}{W}\right),$$

$$W_4 = W_2(-\log p)^{-\frac{2\sum_{i=1}^{n-1}\log(r_i/r_n)}{W}},$$

$$W_5 = 1 - \exp\left[-\left(\frac{W_2}{x_0}\right)^{\frac{W}{2\sum_{i=1}^{n-1}\log(r_i/r_n)}}\right].$$

Remark 3: Let $T_1 = \frac{n}{\sum_{i=1}^{n-1}\log(r_i/r_n)}\log(W_2/r_n)$. Then $T_1 = \frac{2n}{W}\log(V_1/2)$. Hence T_1 is a pivotal quantity. Similarly, $T_2 = \frac{n}{\sum_{i=1}^{n-1}\log(r_i/r_n)}\log(W_4/r_n)$ is also a pivotal quantity.

Remark 4: Let $T_{1,\gamma}$ is the β percentile of T_1. Then a $1 - \beta$ confidence interval for α is given by

$$\left[R_n e^{\frac{T_{1,\beta/2}\sum_{i=1}^{n-1}\log(R_i/R_n)}{n}}, \ R_n e^{\frac{T_{1,1-\beta/2}\sum_{i=1}^{n-1}\log(R_i/R_n)}{n}}\right].$$

Using (5) and $V_1 = 2(\alpha/R_n)^\lambda \sim \chi^2(2n)$, the average length of the confidence interval for α is given by

$$\frac{\alpha\Gamma(n - \lambda^{-1})}{\Gamma(n)}\left[\left(1 - \frac{T_{1,1-\beta/2}}{n\lambda}\right)^{-n} - \left(1 - \frac{T_{1,\beta/2}}{n\lambda}\right)^{-n}\right].$$

It is obvious that for small λ, the average length of the confidence interval for α may be infinite. For the confidence interval of x_p, there is the similar results.

Remark 5: $1 - \frac{2\sum_{i=1}^{n-1}\log(r_i/r_n)}{W}$ may be less than 0 due to $W \sim \chi^2(2n - 2)$, so we ignore W_3 in this case.

Example 2 continued

For the generalized exponential distribution $GE(\lambda, \alpha)$, $\mu = [\psi(\alpha + 1) - \psi(1)]/\lambda$, $x_p = -\lambda^{-1}\log(1 - p^{1/\alpha})$ and $R(x_0) = 1 - (1 - e^{-\lambda x_0})^\alpha$, where $\psi(\cdot)$ is the digamma function. Thus we have

$$W_2 = -\frac{V_1}{2\log[1 - \exp(-g(W,r)r_n)]},$$

8

$$
\begin{aligned}
W_3 &= [\psi(W_2 + 1) - \psi(1)]/g(W, r), \\
W_4 &= -[g(W, r)]^{-1} \log(1 - p^{1/W_2}), \\
W_5 &= 1 - (1 - e^{-g(W,r)\,x_0})^{W_2}.
\end{aligned}
$$

Remark 6: Via simulations we find that for small α or n, the average length of the confidence intervals for α and μ may be infinite.

Because the coverage probabilities of their generalized confidence intervals may depend on nuisance parameters, we study the performance of coverage probabilities of these confidence intervals via simulations in Section 3.

2.3 Prediction interval

While inference for characteristics of the underlying probabilit distributions in case of record value observations is of interest, it is also important to derive inferential methods for prediction as one is often interested in the value of the next record(s) based on the current record values. Providing a prediction interval with good frequentist properties is a challenging issue for many existing methods but such a prediction is most useful and highly expected. Below we present a new method to predict the $(n + k)$th lower record value R_{n+k} based on n existing lower record values $R = (R_1, R_2, ..., R_n)$.

To derive a prediction interval for R_{n+k}, we continue to use the earlier definition of $Y_1, \cdots, Y_n$ and we further define

$$
Y_{n+1} = \alpha[\log G(R_n; \lambda) - \log G(R_{n+1}; \lambda)], ..., Y_{n+k} = \alpha[\log G(R_{n+k-1}; \lambda) - \log G(R_{n+k}; \lambda)].
$$

Then $Y_1, Y_2, \cdots, Y_{n+k}$ are i.i.d. standard exponential random variables. Recall that we have defined $U_n = \sum_{i=1}^{n} Y_i = -\alpha \log G(R_n; \lambda)$, and that $2U_n \sim \chi^2(2n)$. So $U_{n+k} = U_n + \sum_{i=n+1}^{n+k} Y_i$, where $2 \sum_{i=n+1}^{n+k} Y_i \sim \chi^2(2k)$ and it is independent of U_n. Therefore

$$
V_2 \equiv U_n/U_{n+k} = \frac{\log G(R_n; \lambda)}{\log G(R_{n+k}; \lambda)}
$$

follows the beta distribution $Beta(n, k)$ with pdf

$$
f(v) = \frac{\Gamma(n + k)}{\Gamma(n)\Gamma(k)} v^{n-1} (1 - v)^{k-1}, \quad 0 < v < 1.
$$

Thus we have

$$
R_{n+k} = G^{-1}([G(R_n; \lambda)]^{1/V_2}; \lambda).
$$

We now define the generalized pivotal prediction quantity

$$W_6 = G^{-1}([G(r_n; g(W, r))]^{1/V_2}; g(W, r)).$$

Similar to W_2, the cdf of W_6 is given by

$$F_{W_6}(w) = \int_0^\infty F_{Beta(n,k)} \left(\log G(r_n; g(x, r)) / \log G(w; g(x, r))\right) f_{\chi^2(2n-2)}(x) dx, \qquad (10)$$

where $F_{Beta(n,k)}(x)$ is the cdf of the $Beta$ distribution $Beta(n, k)$.

In particular, for the inverse Weibull distribution, we have

$$W_6 = r_n V_2^{\frac{2\sum_{i=1}^{n-1} \log(r_i/r_n)}{W}},$$

and for the generalized exponential distribution, we have

$$W_6 = -[g(W, r)]^{-1} \log\left[1 - (1 - \exp\{-g(W, r)r_n\})^{1/V_2}\right].$$

Let $W_{6,\beta}$ denote the β percentile of W_6. Then $W_{6,\beta}$ is the $1 - \beta$ lower prediction limit for R_{n+k}. Just as in the case of W_2, the percentiles of W_6 can be obtained from the cdf (10) or by Monte Carlo simulations. The performance of the generalized prediction interval, based on the coverage probabilities, will again be investigated by simulations in Section 3.

Remark 7: Let $T_3 = \frac{n}{\sum_{i=1}^{n-1} \log(r_i/r_n)} \log(W_6/r_n)$ in the inverse Weibull distribution case. Then $T_3 = \frac{2n \log(V_2)}{W}$. Hence T_3 is a pivotal prediction quantity.

Remark 8: Notice that, for the inverse Weibull distribution, W_2 can be rewritten as

$$W_2 = \left[\left(\frac{V_1}{2}\right)^{\frac{2\sum_{i=1}^{n-1} \log(r_i^\lambda/r_n^\lambda)}{W}} r_n^\lambda\right]^{1/\lambda},$$

where r_i^λ is the record value from the inverse Weibull distribution with shape parameter 1, thus the coverage probability of the generalized confidence interval for α does not depend on the shape parameter λ. Similar results hold for the generalized confidence interval of x_p and the prediction interval of Y_{n+k}. However, the coverage probability of the generalized confidence interval for $R(x_0)$ does depend on the shape parameter λ in this case.

Remark 9: Let $R_1, R_2 ..., R_n$ be the upper record values observed from the proportional hazards family with the cdf $F(x; \lambda, \alpha) = 1 - [1 - G(x; \lambda)]^\alpha$, $x > 0$, where $G(x; \lambda)$ is a cdf dependent only on λ. Then $R_1^{-1}, R_2^{-1} ..., R_n^{-1}$ be the lower record values observed from the proportional reverse hazards family with the cdf $F(x; \lambda, \alpha) = [1 - G(x^{-1}; \lambda)]^\alpha$, $x > 0$. Therefore, it is obvious that the proposed procedure is extended to the proportional hazards family for the upper record values.

3 Simulation study

To assess the performance of the proposed generalized confidence intervals and prediction intervals, we performed a simulation study with lower record values generated via various scenarios. Because the proposed generalized confidence intervals and prediction intervals are scale equivariant and invariant, we take, without loss of generality, $\alpha = 1$ for the inverse Weibull distribution and $\lambda = 1$ for the generalized exponential distribution in our simulation study. For each scenario, 10,000 replicates of the lower record values were generated from the inverse Weilbull distribution or the generalized exponential distribution. The quantiles of W_i are obtained by Monte Carlo methods with $m = 10,000$. The simulation results are reported in Tables 1 and 2. The corresponding interval lengths are provided in parentheses.

Table 1: The coverage probabilities of the generalized confidence intervals for the inverse Weibull distribution.

λ	n	α 90%	α 95%	$x_{0.1}$ 90%	$x_{0.1}$ 95%	$R(1)$ 90%	$R(1)$ 95%	R_{n+1} 90%	R_{n+1} 95%
2	5	0.9032	0.9496	0.8956	0.9480	0.9035	0.9485	0.8994	0.9470
		(14.7993)	(71.2156)	(1.7440)	(3.3927)	(0.6990)	(0.7775)	(0.1639)	(0.2070)
	10	0.8984	0.9463	0.8967	0.9492	0.8968	0.9467	0.9026	0.9496
		(3.5945)	(5.9931)	(1.1441)	(1.6692)	(0.6702)	(0.7503)	(0.0523)	(0.0659)
3	5	0.9033	0.9496	0.8956	0.9480	0.9035	0.9485	0.8994	0.9470
		(3.7094)	(8.5065)	(0.9912)	(1.6069)	(0.6990)	(0.7775)	(0.1500)	(0.1939)
	10	0.8984	0.9463	0.8967	0.9492	0.8968	0.9467	0.9026	0.9496
		(1.6653)	(2.4256)	(0.7256)	(0.9903)	(0.6702)	(0.7503)	(0.0520)	(0.0661)
7	5	0.9031	0.9496	0.8956	0.9480	0.9035	0.9485	0.8994	0.9470
		(0.8119)	(1.2440)	(0.3889)	(0.5564)	(0.6990)	(0.7775)	(0.0936)	(0.1248)
	10	0.8984	0.9463	0.8967	0.9492	0.8968	0.9467	0.9026	0.9496
		(0.5283)	(0.6942)	(0.3078)	(0.3970)	(0.6702)	(0.7503)	(0.0353)	(0.0454)

Table 2: The coverage probabilities of the generalized confidence intervals for the generalized exponential distribution.

α	n	α 90%	α 95%	μ 90%	μ 95%	$x_{0.1}$ 90%	$x_{0.1}$ 95%	$R(1)$ 90%	$R(1)$ 95%	R_{n+1} 90%	R_{n+1} 95%
1	5	0.8971	0.9475	0.8964	0.9480	0.8989	0.9504	0.9024	0.9520	0.8987	0.9529
		(∞)	(∞)	(∞)	(∞)	(5.5037)	(46.5421)	(0.7770)	(0.8566)	(0.0279)	(0.0305)
	10	0.8994	0.9507	0.8990	0.9463	0.9055	0.9490	0.9012	0.9475	0.8972	0.9501
		(1.3279)	(1.5924)	(∞)	(∞)	(11.1331)	(64.5724)	(0.7851)	(0.8620)	(0.0008)	(0.0009)
3	5	0.8975	0.9473	0.8975	0.9486	0.8992	0.9508	0.9009	0.9504	0.8982	0.9529
		(∞)	(∞)	(35.6392)	(188.8644)	(2.7610)	(5.3734)	(0.6367)	(0.7135)	(0.1981)	(0.2286)
	10	0.9006	0.9506	0.9004	0.9467	0.9067	0.9497	0.9030	0.9497	0.8958	0.9499
		(5.0104)	(6.0964)	(12.0362)	(23.1565)	(2.5522)	(4.1923)	(0.6361)	(0.7117)	(0.0347)	(0.0404)
7	5	0.8981	0.9490	0.8970	0.9483	0.8993	0.9516	0.8993	0.9505	0.8997	0.9528
		(∞)	(∞)	(12.0561)	(24.8339)	(2.7345)	(4.2736)	(0.3310)	(0.3982)	(0.4107)	(0.4975)
	10	0.8986	0.9500	0.9009	0.9473	0.9047	0.9497	0.9022	0.9501	0.8954	0.9497
		(18.1846)	(23.2668)	(6.3463)	(9.2714)	(2.3266)	(3.2333)	(0.3145)	(0.3742)	(0.1264)	(0.1537)

Tables 1 and 2 show the 90% and 95% generalized confidence intervals for the parameters and further quantities considered in Section 2 of these two probability distributions, as well as prediction interval for R_{n+1}. However, we do not include μ for the inverse Weibull distribution here due to the possibly negative value of $1 - \frac{2 \sum_{i=1}^{n-1} \log(r_i/r_n)}{W}$. Clearly, this simulation study shows that, even if the number of observed record values is very small, such as $n = 5, 10$, the coverage probabilities of the 90% and 95% confidence intervals resulting from the method presented in this paper are close to 0.9 and 0.95, respectively.

4 An illustrative example

Madi and Raqab (2007) presented a data analysis of the amount of annual rainfall (in inches) recorded at the Los Angeles Civic Center for 127 years, from 1878 to 2005 (season July 1 – June 30). They found that these data can be fitted well by the generalized exponential distribution. They used six lower record values up to 1959 as follows: 11.35, 10.40, 9.21, 6.73, 5.59, 5.58, and they presented Bayesian prediction for the next two lower records. The actually observed next two lower records were set in 1960 and 2001, and were 4.85 and 4.35, respectively. We use the method presented in this paper to re-analyze these data, using only this small set of six lower record values. The results are presented in Table 3. Compared with the Bayesian interval estimation in Madi and Raqab (2007), the Bayesian interval for α has shorter interval length than the generalized confidence interval, but the Bayesian interval for λ has longer interval length than the generalized confidence interval. In particular, we obtain 95% prediction intervals for the next two lower record values, R_7 and R_8, as (3.2836, 5.5676) and (2.3595, 5.4770), respectively, using $m = 1,000,000$ in the simulations. It is interesting to compare these prediction intervals to the 95% Bayesian prediction intervals presented by Madi and Raqab (2007) for R_7 and R_8, which are (1.4824, 5.5383) and (0.6788, 5.2447), respectively. Both the Bayesian prediction intervals and the new prediction intervals based on our method contain the true record values, but our prediction intervals have substantially shorter length due to much larger lower limits of the intervals. As the actual number of observations is small, any Bayesian method for such prediction will be influenced by the choice of the prior distribution, which will typically be rather difficult to assess meaningfully due to the complex nature of record values. From this perspective, the presented method in this paper has the advantage of not requiring additional input beyond the data.

Table 3: The confidence intervals and prediction intervals based on the Los Angeles rainfall data.

	λ	α	μ	$x_{0.1}$	R_7	R_8
0.90	[1.6917, 11.4370]	[4.8600, 164.9770]	[9.0279, 28.2269]	[5.9390, 14.1814]	[3.7793, 5.5548]	[2.9102, 5.4276]
0.95	[1.5080, 15.9885]	[3.6350, 250.5091]	[8.5560, 35.9129]	[5.4487, 16.7153]	[3.2836, 5.5676]	[2.3595, 5.4770]

Acknowledgements

References

1. Ahmadi J., Jozani M.J., Marchand É. and Parsian A. (2009) Prediction of k-records from a general class of distributions under balanced type loss functions, *Metrika*, 70, 19–33.

2. Ahsanullah M. (2004) *Record Values - Theory and Applications*. University Press of America.

3. Arnold B.C., Balakrishnan N. and Nagaraja H.N. (1998) *Records*. Wiley, New York.

4. Balakrishnan N. and Chan P.S. (1998) On the normal record values and associated inference. *Statistics & Probability Letters*, 39, 73–80.

5. Balakrishnan N., Chan P.S., and Ahsanullah M. (1993) Recurrence relations for moments of record values from generalized extreme value distribution. *Communications in Statistics-Theory and Methods*, 22, 1471–1482.

6. Carlin B.P. and Gelfand A.E. (1993) Parametric likelihood inference for record breaking problems. *Biometrika*, 80, 507–515.

7. Chan P.S. (1998) Interval estimation of location and scale parameters based on record values. *Statistics & Probability Letters*, 37, 49-58.

8. Chander K.N. (1952) The distribution and frequency of record values. *Journal of the Royal Statistical Society, Series B*, 14, 220–228.

9. Coles S.G. and Tawn J.A. (1996) A Bayesian analysis of extreme rainfall data. *Applied Statistics*. 45, 463–478.

10. Cramer E. and Naehrig G. (2012) Laplace record data. *Journal of Statistical Planning and Inference*, 142, 2179–2189.

11. Einmahl, J.H.J. and Magnus, J.R. (2008) Records in athletics through extreme-value theory. *Journal of the American Statistical Association*, 103, 1382–1389.

12. Feuerverger, A. and Hall, P. (1996) On distribution-free inference for record-value data with trend. *The Annals of Statistics*, 24, 2655–2678.

13. Glick N., (1978) Breaking records and breaking boards. *The American Mathematical Monthly*, 85, 2–26.

14. Gupta R.D. and Kundu D. (1999) Generalized exponential distributions. *Australian and New Zealand Journal of Statistics*, 41, 173–188.

15. Katz R.W., Parlange M B., and Naveau P. (2002) Statistics of extremes in hydrology. *Advances in Water Resources*, 25, 1287–1304.

16. Madi M.T. and Raqab, M. (2007) Bayesian prediction of rainfall records using the generalized exponential distribution. *Environmetrics*, 18, 541–549.

17. Marshall A.W. and Olkin I. (2007) *Life distributions; structure of nonparametric, semi-parametric, and parametric families*, Springer.

18. Raqab M.Z. (2002) Inferences for generalized exponential distribution based on record statistics. *Journal of Statistical Planning and Inference*, 104, 339–350.

19. Raqab M.Z. and Sultan K.S. (2014) Generalized exponential records: existence of maximum likelihood estimates and its comparison with transforming based estimates. *Metron*, 72, 65–76.

20. Robinson M.E. and Tawn J.A. (1995) Statistics for exceptional athletics records. *Applied Statistics*, 44, 499–511.

21. Smith R.L. (1988) Forecasting records by maximum likelihood. *Journal of the American Statistical Association*, 83, 331–338.

22. Soliman A.A., Abd Ellah A., and Sultan K. (2006) Comparison of estimates using record statistics from Weibull model: Bayesian and non-Bayesian approaches. *Computational Statistics & Data Analysis*, 51, 2065–2077.

23. Sultan K., Al-Dayian G., and Mohammad H. (2008) Estimation and prediction from gamma distribution based on record values, *Computational Statistics & Data Analysis*, 52, 1430–1440.

24. Tavangar M. and Asadi M. (2011) Some results on conditional expectations of lower record values. *Statistics*, 45, 237–255.

25. Wang B.X., Yu K.M. and Jones M.C. (2010) Inference under progressively type II right-censored sampling for certain lifetime distributions. *Technometrics*, 52, 453–460.

26. Wang B.X. and Ye Z.S. (2015) Inference on the Weibull distribution based on record values, *Computational Statistics & Data Analysis*, 83, 26-36.

27. Weerahandi S. (2004) *Generalized Inference in Repeated Measures: Exact Methods in MANOVA and Mixed Models*. Wiley, New York.

28. Wergen G. (2014) Modeling record-breaking stock prices. *Physica A*, 396, 114C133.

29. Wong A. and Wu Y.Y. (2009) A note on interval estimation of $P(X < Y)$ using lower record data from the generalized exponential distribution. *Computational Statistics & Data Analysis*, 53, 3650–3658.