Statistics and Probability Letters 98 (2015) 115–122

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

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

Bing Xing Wang [a,*], Keming Yu [b], Frank P.A. Coolen [c]

[a] *Zhejiang Gongshang University, China*
[b] *Brunel University, UK*
[c] *Durham University, UK*

ARTICLE INFO

Article history:
Received 11 August 2014
Received in revised form 21 December 2014
Accepted 21 December 2014
Available online 26 December 2014

MSC:
62N05
62F25

Keywords:
Confidence interval
Proportional reversed hazard distribution
Record value
Sample size

ABSTRACT

This paper explores confidence intervals for the family of proportional reversed hazard distributions based on lower record values. The confidence intervals are validated as long as the sample is of size $n \geq 3$. 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.

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 last component breaks. Moreover, 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), Tavangar and Asadi (2011) and 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).

* Correspondence to: Department of Statistics, Zhejiang Gongshang University, China.
E-mail address: wangbingxing@163.com (B.X. Wang).

http://dx.doi.org/10.1016/j.spl.2014.12.019
0167-7152/

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 $\{X_n, n = 1, 2, \ldots\}$ 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 $\{R_n = X_{T_n}, n = 1, 2, \ldots\}$ 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

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

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

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

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

that the Jacobian of transformation is given by

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

and the pdf of $Y_1, \ldots, Y_n$ is given by

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

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

Lemma 2. *Suppose that $Y_1, Y_2, \ldots, Y_n$ are i.i.d. exponential random variables with mean θ. Let $S_i = Y_1 + \cdots + Y_i$, $i = 1, 2, \ldots, n$, then $S_1/S_2, (S_2/S_3)^2, \ldots, (S_{n-1}/S_n)^{n-1}, S_n$ are independent random variables. Also, $S_1/S_2, (S_2/S_3)^2, \ldots, (S_{n-1}/S_n)^{n-1}$ 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 $R_1, R_2 \ldots, R_n$ be the lower record values observed from the proportional reversed hazards family (1), then $F(R_1; \lambda, \alpha)$, $F(R_2; \lambda, \alpha), \ldots, F(R_n; \lambda, \alpha)$ are the lower record values observed from the standard uniform distribution $U(0, 1)$. Thus, we have from Lemma 1 that $Y_1 = -\log F(R_1; \lambda, \alpha)$, $Y_2 = \log F(R_1; \lambda, \alpha) - \log F(R_2; \lambda, \alpha), \ldots, Y_n = \log F(R_{n-1}; \lambda, \alpha) - \log F(R_n; \lambda, \alpha)$ are i.i.d. standard exponential random variables.