Available online at www.sciencedirect.com
SCIENCE DIRECT®

Statistics & Probability Letters 76 (2006) 537–545

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

On the Fisher information in record data

N. Balakrishnan[a,*], A. Stepanov[b]

[a]Department of Mathematics and Statistics, McMaster University, 1280 Main Street West, Hamilton, Ont., Canada L8S 4K1
[b]Department of Mathematics, Kaliningrad State Technical University, Sovietsky Prospect 1, Kaliningrad, 236000 Russia

Received 25 February 2005; received in revised form 25 July 2005
Available online 6 September 2005

Abstract

The Fisher information contained in records, weak records and numbers of records are discussed in this paper. In the case when the initial distribution belongs to the exponential family, the Fisher information contained in record values as well as in record values and record times are found analytically. A new inverse sampling plan (ISP-II) is considered next and some results on Fisher information in record statistics obtained from ISP-II are then derived. We also finally propose some new estimators based on records and weak records.

MSC: primary 60G70; 62G30

Keywords: Fisher information; Records; Weak records; Inverse sampling plan; Sufficient statistics; Estimators based on records; Exponential families; Geometric distribution; Pareto distribution; Weibull distribution

1. Introduction

Let $X_1, X_2, \ldots$ be independent and identically distributed (i.i.d.) random variables with absolutely continuous cdf $F(x, \theta)$, where θ is an unknown parameter. The Fisher information (FI) about the real parameter θ contained in record values and weak record values have been discussed quite a bit recently in the literature; see, for example, Ahmadi and Arghami (2001), Hofmann and Nagaraja (2003), Stepanov et al. (2003), Hofmann and Balakrishnan (2004), Hofmann (2004), and Hofmann et al. (2005).

The FI contained in the first n record values and in n i.i.d. random variables have been compared in some of the above mentioned papers. Hofmann and Nagaraja (2003), Hofmann and Balakrishnan (2004), and Hofmann (2004) have also compared the FI contained in the first n record values and record times with the FI contained in n i.i.d. observations. In many cases, the FI in the first n record values is less than the FI in n i.i.d. observations, but the FI in the first n record values and record times exceeds the FI in n i.i.d. observations. One example is the normal mean.

It should be noted that n i.i.d. observations produce approximately [log n] record values, where [ ] denotes the integral part of the number; see Arnold et al. (1998). Thus, even if the FI contained in the first n record values and record times is greater than the FI in n i.i.d. observations, it does not mean that record values and

*Corresponding author. Tel.: +1 905 5259140; fax: +1 905 5221676.
E-mail addresses: bala@univmail.cis.mcmaster.ca (N. Balakrishnan), alexei@step.koenig.ru (A. Stepanov).

0167-7152/$ - see front matter
doi:10.1016/j.spl.2005.08.021

record times are more attractive for statistical inferential procedures. It is intuitively clear and it will be formally shown in Section 2 (see Theorem 2.1) that i.i.d. observations $X_1, \dots, X_n$ contain FI that is more than the FI in record values and record times induced by this sample.

Sometimes, only record values are available. The consideration of a fixed number of records was introduced by Samaniego and Whitaker (1986), who termed it as an inverse sampling plan. We shall denote this inverse sampling plan by ISP-I. One situation where this model arises is in industrial stress and life testing wherein measurements are made sequentially and only values larger than all previous values are recorded. For some other examples, see Galambos (1984), Hofmann and Nagaraja (2003), and Hofmann (2004). The other case when record samples could be used instead of traditional ones is when we have samples consisting of too many observations.

In some situations, even record values may not be available. Suppose during an experiment we get information only about the numbers of record values belonging to different intervals. Naturally, we can study the FI contained in these numbers of records. In the discrete case, when X_1 is defined on $0, 1, \dots,$ we may be interested in the FI contained in the numbers of weak record values registered at the points $0, 1, \dots, n$. This is a new inverse sampling plan as compared to the one of Samaniego and Whitaker (1986), and we denote it by ISP-II.

The sequences of upper record values $X^u(n)$ and upper record times $L^u(n)$ are defined as follows:

$$\begin{aligned} &L^u(1) = 1, \\ &L^u(n+1) = \min\{j : j > L^u(n), X_j > X_{L^u(n)}\}, \\ &X^u(n) = X_{L^u(n)} \quad (n \geqslant 1). \end{aligned} \tag{1.1}$$

If in (1.1) the condition $X_j > X_{L^u(n)}$ is replaced by $X_j \geqslant X_{L^u(n)}$, then the above gives the definitions of weak record values $X^w(n)$ and weak record times $L^w(n)$. Observe that there is no difference between upper records and weak records in the case when the initial $F(x) = P\{X_1 \leqslant x\}$ is continuous; see Arnold et al. (1998) and Nevzorov (2000). Moreover, if in (1.1) the inequality in $X_j > X_{L^u(n)}$ is reversed, then the above become the definitions of lower record values $X^l(n)$ and lower record times $L^l(n)$.

Let us now use I_n, $I_n(R^u)$, $I_n(R^uT^u)$, and $I_n(M)$ to denote the FI contained in i.i.d. observations $X_1, \dots, X_n$, in all upper record values belonging to this sample, in all upper record values and upper record times belonging to the sample, and in the maximum M_n of the sample, respectively. We will replace u by l to denote the corresponding quantities for lower records. It needs to be mentioned that in all notation of the FI, we suppress the parameter θ.

For dealing with inverse sampling plan ISP-I, we will use $J_n(R^u)$, $J_n(R^uT^u)$ and $J_n(W)$ to denote the FI contained in n upper record values, in n upper record values and upper record times, and in n weak record values, respectively. For discussing the inverse sampling plan ISP-II, let us introduce a non-negative integer variable $\mu^u(a,b)$ $(-\infty \leqslant a < b \leqslant \infty)$ to denote the number of upper record values belonging to the interval (a,b). For the discrete case, we similarly define random variables ξ_i^w by $\xi_i^w = k$ if k weak records occur at point i, $i = 1, 2, \dots, k \geqslant 0$. Denote by $I(\mu^u(a,b))$ the FI contained in $\mu^u(a,b)$, i.e. the FI in the number of upper record values belonging to the interval (a,b). Finally, let $I(\xi_n^w)$ be the FI contained in ξ_n^w and $I(\overline{\xi_n^w})$ be the FI contained in $(\xi_0^w, \dots, \xi_n^w)$.

We shall assume throughout that appropriate regularity conditions are satisfied.

The rest of the paper is organized as follows. The FI contained in records and in the numbers of record values is discussed in Section 2 for the continuous case. Section 2.1 presents theoretical results while some examples are presented in Section 2.2. In Section 3, we discuss the FI contained in weak record values and in the numbers of weak record values for the discrete case. Section 3.2 presents some examples for the discrete case.

2. Fisher information contained in record values and in the numbers of record values

2.1. Main results

Let $X_1, X_2, \dots$ be i.i.d. random variables with absolutely continuous cdf $F(x,\theta)$ and pdf $p(x,\theta) = F'_x(x,\theta)$. As has been pointed out in the introduction, the i.i.d. variables $X_1, \dots, X_n$ contain FI more than the record