Ann. Inst. Statist. Math.
Vol. 56, No. 2, 383–396 (2004)

FISHER INFORMATION IN k-RECORDS

GLENN HOFMANN[1]* AND N. BALAKRISHNAN[2]

[1] *Departamento de Estadistica, Facultad de Ciencias Físicas y Matemáticas, Barrio Universitario, Casilla 160-C, Concepción, Chile,* e-mail: glennhofmann@yahoo.com
[2] *Department of Mathematics and Statistics, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1,* e-mail: bala@mcmail.cis.mcmaster.ca

(Received July 9, 2002; revised February 3, 2003)

Abstract. We derive some general results on the Fisher information (FI) contained in the upper (or lower) k-record values and associated k-record times generated from an i.i.d. sample of fixed size from a continuous distribution. We apply the results to obtain the FI in both upper and lower k-record data from an exponential distribution. We propose two estimators of the exponential mean, based on the upper and lower k-record data, and discuss their small sample properties. We also consider k-record data from an inverse sampling plan, and present general formulas for the FI contained in it.

Key words and phrases: Record values, record times, k-record values, Fisher information, exponential distribution, parameter estimation, inverse sampling.

1. Introduction

How much information is contained in record observations? This question was addressed by Ahmadi and Arghami (2001) and Hofmann (2003) by means of a comparison between the Fisher information (FI) in an i.i.d. sample and record data. They point out that for many distributions and parameters, the FI in the first m record values and record times is larger than the FI in m i.i.d. observations. The consideration of a fixed number of records is known as inverse sampling (Samaniego and Whitaker (1986)) and has been used for almost all known record based inference procedures (see Arnold *et al.* (1998), Chapter 5, and references therein). To our knowledge, only Samaniego and Whitaker (1986) and Hofmann and Nagaraja (2003) treat record based inference by fixing the number of observations rather than the number of records. Samaniego and Whitaker (1986) give an estimator based on lower record values and record times from an exponential distribution. Hofmann and Nagaraja (2003) establish its asymptotic efficiency, give small sample efficiencies and investigate the properties of the maximum likelihood estimator based on upper record statistics from an exponential distribution. Information measures have also been discussed for order statistics. Tukey (1965) introduced linear sensitivity to find out which order statistics are more important for linear estimation. Nagaraja (1994) considered this measure in detail. Balakrishnan and Chandrasekar (2002) presented a multivariate version of linear sensitivity. Fisher information in order statistics

*Supported in part by Fondo Nacional de Desarrollo Cientifico y Tecnologico (FONDECYT) grant # 1020479 of Chile.

has only recently been discussed by Park (1996) and Iyengar *et al.* (1999) among others.

In record value theory, while the inverse sampling considerations have given valuable insights, their practical implementation is greatly hindered by the sparsity of records, in fact, the expected waiting time is infinite for every record after the first. In this paper, we will combine the informational advantage of records with the practical necessity of fixing the sample size, and we will remedy the problem of sparsity by considering k-records instead. We will define this generalization of records in Section 2, and also give other preliminaries. We will show that the expected waiting time for k-records is finite ($k \geq 2$). In Section 3, we will present general expressions for the FI in k-record values and k-record times from a random sample of fixed size n. These results will be applied to the exponential distribution in Section 4. We will focus on an unbiased, lower k-record based estimator of the exponential parameter in Section 5. We will present the variance of this estimator, compare it with the Cramer-Rao lower bound, and with the variance of the estimator based on the whole sample. We will also discuss the MLE based on upper k-record data. Finally, in Section 6 we will state general expressions for the FI in inversely sampled k-records.

2. Preliminaries

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables with absolutely continuous cdf $F(x;\theta)$ and density $f(x;\theta)$, where θ is an unknown parameter. We are interested in the FI contained in the k-record data about the parameter θ. A k-record is basically the k-th largest observation in a partial sample. When new observations arrive, new k-records can occur. In infinite sequences, every new observation that is bigger than the current k-record will eventually become a k-record itself. A precise definition is given below. Under certain regularity conditions, the FI about the real parameter θ contained in a random variable X with density $f(x;\theta)$ is defined by (see, for example, Rao (1973), p. 329), $I_X = E_\theta(\frac{\partial \log f(X;\theta)}{\partial \theta})^2$. Traditionally, the FI has played a valuable role in statistical inference through the information (Cramer-Rao) inequality and its association with the asymptotic properties of the maximum likelihood estimators.

Let $X_{i:m}$ denote the i-th order statistic from a random sample of size m. We define the (upper) k-record times $T_{n,k}$ and the (upper) k-record values $R_{n,k}$ to be as follows: $T_{1,k} = k$, $R_{1,k} = X_{1:k}$ and for $m \geq 2$, $T_{m,k} = \min\{j : j > T_{m-1,k}, X_j > X_{T_{m-1,k}-k+1:T_{m-1,k}}\}$, and $R_{m,k} = X_{T_{m,k}-k+1:T_{m,k}}$. Let $\delta_m + 1 = T_{m+1,k} - T_{m,k}$ $(m \geq 1)$ be the k-interrecord times, and let $N_{m,k}$ be the number of k-record values in $X_1, \ldots, X_m$. Lower k-record statistics are defined similarly. Let $I_{RT}^U(n,k)$ be the FI contained in the upper k-record values and k-record times from a random sample of size n, and $I_R^U(n,k)$ the FI in just the k-record values. The corresponding notations for lower records are $I_{RT}^L(n,k)$ and $I_R^L(n,k)$. These k-records were introduced by Dziubdziela and Kopociński (1976) and have found acceptance in the literature (see, for example, Grudzién and Szynal (1985); Raqab and Amin (1997)). Arnold *et al.* (1998) call them *Type* 2 *k-records*. For $k = 1$, the usual records are recovered. Let

$$I_{m,k} = \begin{cases} 1 & \text{if } m \text{ is a } k\text{-record time} \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, $I_{1,k} = \cdots = I_{k-1,k} = 0$, $I_{k,k} = 1$ and $I_{k+1,k}, \ldots, I_{m,k}$ are independent Bernoulli random variables with $P(I_{i,k} = 1) = \frac{k}{i}$ $(i \geq k)$. We can then prove the following result.

THEOREM 2.1. *For $k \geq 2$, $ET_{m,k}$ is finite for all m.*

PROOF. The k-record time and the number of k-records are related in the following way. Let $m \geq 3$, $n \geq m+k-1$ ($P(T_{m,k} > m+k-2) = 1$), then

$$\begin{aligned} P(T_{m,k} > n) &= P(N_{n,k} < m) = P(I_{k+1,k} + \cdots + I_{n,k} < m-1) \\ &= P(\text{at least } n-k-m+2 I\text{'s are} = 0) \\ &\leq \frac{n-k}{n}\frac{n-1-k}{n-1}\cdots\frac{m-1}{k+m-1} = \frac{\binom{k+m-2}{k}}{\binom{n}{k}} \\ &\leq \binom{k+m-2}{k}\frac{k^2}{n^2} \quad \text{for} \quad k \geq 2. \end{aligned}$$

Hence, the expected k-record time $T_{m,k}$ can be expressed as

$$\begin{aligned} ET_{m,k} &= \sum_{i=0}^{\infty} P(T_{m,k} \geq i) = m+k+\sum_{i=m+k-1}^{\infty} P(T_{m,k} > i) \\ &\leq m+k+k^2\binom{k+m-2}{k}\sum_{i=m+k-1}^{\infty}\frac{1}{i^2} < \infty. \end{aligned}$$ □

Note that for $k = 1$, i.e. for records, $ET_{m,1} = \infty$ for $m \geq 2$ (Arnold *et al.* (1998), p. 26).

Dziubdziela and Kopociński (1976) showed that the sequence $\{R_{n,k}, n \geq 1\}$ from a cdf F is identical in distribution to a record ($k = 1$) sequence from the cdf $F_{1:k} = 1-(1-F)^k$. Using the joint density of records (see, for example, Arnold *et al.* (1998), p. 10), we readily have the joint density of the first m k-records $R_{1,k}, \dots, R_{m,k}$ as

$$f(r_1, \dots, r_m) = k^m \prod_{i=1}^{m} \frac{f(r_i)}{1-F(r_i)}(1-F(r_m))^k. \tag{2.1}$$

Note that such a distributional identity between F and $F_{1:k}$ cannot be found for the joint distribution of the k-record values and k-record times. However, a new k-record time $T_{m,k}$ only occurs when the corresponding observation $X_{T_{m,k}}$ is greater than the last k-record value $R_{m-1,k}$ (but not necessarily smaller than any following k-record value). Hence, given the k-record values, the k-interrecord times are conditionally independent and geometrically distributed with the following joint probability function:

$$\begin{aligned} &P(T_{1,k} = t_1, T_{2,k} = t_2, \dots, T_{m,k} = t_m, N_{n,k} = m \mid R_{1,k} = r_1, \dots, R_{m,k} = r_m) \\ &\quad = \prod_{i=1}^{m-1} F^{\delta_i}(r_i)(1-F(r_i))F^{\delta_m}(r_m) \end{aligned} \tag{2.2}$$

for $t_1 = k < t_2 < \cdots < t_m \leq n$, $r_1 < \cdots < r_m$, $m = 1, \dots, n-k+1$, where $\delta_i = t_{i+1} - t_i - 1$, and $t_{m+1} = n+1$. Combining (2.1) and (2.2), we obtain the joint likelihood of the k-record values and k-record times to be

$$f(r_1, \dots, r_m, t_1, \dots, t_m, N_{n,k} = m) = k^m \prod_{i=1}^{m} f(r_i)F^{\delta_i}(r_i)(1-F(r_m))^{k-1} \tag{2.3}$$

for $t_1 = k < t_2 < \cdots < t_m \leq n$, $r_1 < \cdots < r_m$, $m = 1, \dots, n-k+1$.

3. FI in finite sample k-record data

THEOREM 3.1. *The FI contained in all upper k-record values and k-record times from a random sample of size n is given by*

$$I_{RT}^{U}(n,k) = \int_{-\infty}^{\infty} kf(x)(1-F(x))^{k-1} \times \left\{ \sum_{i=0}^{n-k} \left[\binom{i+k-1}{k-1} \left(-\frac{\partial^2 \log f(x)}{\partial\theta^2} \right) + \binom{i+k-1}{k} \left(-\frac{\partial^2 \log F(x)}{\partial\theta^2} \right) \right] F^i(x) + (k-1)\binom{n}{k} F^{n-k}(x) \left(-\frac{\partial^2 \log(1-F(x))}{\partial\theta^2} \right) \right\} dx. \tag{3.1}$$

The FI $I_{RT}^{L}(n,k)$ can be obtained by replacing $F(x)$ with $1-F(x)$ in (3.1).

PROOF. To simplify notation, let us denote

$$\delta(m) = (\delta_1, \ldots, \delta_m), \qquad \text{and}$$

$$\Delta(m) = \left\{ (\delta_1, \ldots, \delta_m) : \sum_{i=1}^{m} \delta_i = n+1-k-m, \delta_i \geq 0, \delta_i \text{ integer } \forall i \right\}. \tag{3.2}$$

From (2.3) it then follows that

$$\begin{aligned} I_{RT}^{U}(n,k) &= E\left[\sum_{i=1}^{N_{n,k}} -\frac{\partial^2 \log f(R_{i,k})}{\partial\theta^2} \right] + E\left[\sum_{i=1}^{N_{n,k}} -\delta_i \frac{\partial^2 \log F(R_{i,k})}{\partial\theta^2} \right] \\ &\quad + (k-1)E\left[-\frac{\partial^2 \log(1-F(R_{N_{n,k},k}))}{\partial\theta^2} \right] \\ &= E_1 + E_2 + E_3 \qquad \text{(say)}, \end{aligned} \tag{3.3}$$

where the expectations are taken with respect to the joint density (2.3), i.e., with respect to $(R_{1,k}, \ldots, R_{m,k}, T_{1,k}, \ldots, T_{m,k}, N_{n,k} = m)$. We can write $E_1 = \sum_{m=1}^{n-k+1} E_1(m)$, where

$$E_1(m) = \int_{r_1<\cdots<r_m} \sum_{\delta(m)\in\Delta(m)} \left(-\sum_{i=1}^{m} \frac{\partial^2 \log f(r_i)}{\partial\theta^2} \right) k^m \times \prod_{i=1}^{m} f(r_i) F^{\delta_i}(r_i) (1-F(r_m))^{k-1} dr_1 \cdots dr_m.$$

Note that $\sum_{\delta(m)\in\Delta(m)} \prod_{i=1}^{m} F^{\delta_i}(r_i)$ is equal to the coefficient of $s^{n+1-k-m}$ in the Taylor series expansion of $\prod_{i=1}^{m} \frac{1}{1-F(r_i)s}$. Hence,

$E_1(m) =$ coefficient of $s^{n+1-k-m}$ in

$$\int_{r_1<\cdots<r_m} k^m \left(-\sum_{i=1}^{m} \frac{\partial^2 \log f(r_i)}{\partial\theta^2} \right) \prod_{i=1}^{m} \frac{f(r_i)}{1-F(r_i)s} (1-F(r_m))^{k-1} dr_1 \cdots dr_m.$$

The expression under the integral is symmetric in $r_1, \dots, r_{m-1}$. Therefore,

$$E_1(m) = \text{coefficient of } s^{n+1-k-m} \text{ in } \frac{k^m}{(m-1)!} \int_{-\infty}^{\infty} \frac{f(r_m)}{1-F(r_m)s}(1-F(r_m))^{k-1} \times \int_{-\infty}^{r_m} \cdots \int_{-\infty}^{r_m} -\sum_{i=1}^{m} \frac{\partial^2 \log f(r_i)}{\partial\theta^2} \prod_{i=1}^{m-1} \frac{f(r_i)}{1-F(r_i)s} dr_1 \cdots dr_m.$$

Since $\int_{-\infty}^{r_m} \frac{f(x)}{1-F(x)s} dx = -\frac{1}{s}\log(1-F(r_m)s)$, we obtain

$$\text{(3.4)} \quad E_1(m) = \text{coefficient of } s^{n-k} \text{ in } \frac{1}{(m-1)!} \int_{-\infty}^{\infty} -\frac{\partial^2 \log f(r_m)}{\partial\theta^2} \frac{kf(r_m)}{1-F(r_m)s} \times (1-F(r_m))^{k-1}(-k\log(1-F(r_m)s))^{m-1} dr_m + \text{coefficient of } s^{n-k-1} \text{ in } \frac{I_{\{m>1\}}}{(m-2)!} \int_{-\infty}^{\infty} \frac{kf(r_m)}{1-F(r_m)s}(1-F(r_m))^{k-1} \times (-k\log(1-F(r_m)s))^{m-2} \int_{-\infty}^{r_m} -\frac{\partial^2 \log f(x)}{\partial\theta^2} \frac{kf(x)}{1-F(x)s} dx dr_m,$$

where $I_{\{m>1\}}$ is the indicator function for $m > 1$. We now need to sum over m. Since for $l \geq n-k+1$, the coefficient of s^{n-k} in $(-k\log(1-F(x)s))^l$ is zero, we can extend the sums over m to ∞, and use $\sum_{i=0}^{\infty} \frac{(-k\log(1-F(r_m)s))^i}{i!} = (1-F(r_m)s)^{-k}$. Note further that

$$\text{(3.5)} \quad [\text{coefficient of } s^{n-k} \text{ in } (1-F(y)s)^{-(k+1)}] = \frac{1}{(n-k)!} \frac{\partial^{n-k}(1-F(y)s)^{-(k+1)}}{\partial s^{n-k}}\bigg|_{s=0} = \binom{n}{k} F^{n-k}(y)$$

and

$$\text{(3.6)} \quad \left[\text{coefficient of } s^{n-k-1} \text{ in } \frac{1}{1-F(x)s}\frac{1}{(1-F(y)s)^{k+1}}\right] = \sum_{i=0}^{n-k-1} \binom{n-i-1}{k} F^i(x) F^{n-1-k-i}(y).$$

Combining (3.5) and (3.6) with (3.4), we obtain

$$\text{(3.7)} \quad \sum_{m=1}^{n-k+1} E_1(m) = \int_{-\infty}^{\infty} -\frac{\partial^2 \log f(x)}{\partial\theta^2}(1-F(y))^{k-1} kf(y) \binom{n}{k} F^{n-k}(y) dy + A,$$

where

$$\text{(3.8)} \quad A = \int_{-\infty}^{\infty} kf(y)(1-F(y))^{k-1} \int_{-\infty}^{y} -\frac{\partial^2 \log f(x)}{\partial\theta^2} kf(x) \times \sum_{i=1}^{n-k-1} \binom{n-i-1}{k} F^i(x) F^{n-k-1-i}(y) dx dy.$$

Exchanging integrations with respect to x and y yields

$$(3.9)\qquad A = \int_{-\infty}^{\infty} -\frac{\partial^2 \log f(x)}{\partial\theta^2} kf(x) \sum_{i=0}^{n-k-1} F^i(x) \times \int_x^{\infty} \frac{(n-i-1)!}{(k-1)!(n-i-k-1)!} F^{n-k-i-1}(y)(1-F(y))^{k-1} f(y)dydx.$$

The inner integral is an incomplete Beta function which is equal to

$$P(X_{(n-i-k):(n-i-1)} > x) = \sum_{j=0}^{n-i-k-1} \binom{n-i-1}{j} F^j(x)(1-F(x))^{n-i-j-1}.$$

Hence,

$$A = \int_{-\infty}^{\infty} -\frac{\partial^2 \log f(x)}{\partial\theta^2} kf(x) \sum_{i=0}^{n-k-1} \sum_{j=0}^{n-i-k-1} \binom{n-i-1}{j} F^{i+j}(x)(1-F(x))^{n-i-j-1} dx.$$

By taking $l = i + j$, the expression simplifies to

$$A = \int_{-\infty}^{\infty} -\frac{\partial^2 \log f(x)}{\partial\theta^2} kf(x) \sum_{l=0}^{n-k-1} \binom{n}{l} F^l(x)(1-F(x))^{n-l-1} dx.$$

Now, it follows from (3.7) that

$$(3.10)\qquad \sum_{m=1}^{n-k+1} E_1(m) = \int_{-\infty}^{\infty} \left(-\frac{\partial^2 \log f(x)}{\partial\theta^2}\right) \frac{kf(x)}{1-F(x)} \sum_{l=0}^{n-k} \binom{n}{l} F^l(x)(1-F(x))^{n-l} dx.$$

Further, since

$$\begin{aligned}\sum_{l=0}^{n-k} \binom{n}{l} F^l(x)(1-F(x))^{n-l-k} &= \sum_{l=0}^{n-k} \sum_{i=0}^{n-l-k} \binom{n}{l} \binom{n-l-k}{i} (-1)^i F^{l+i}(x) \\ &= \sum_{j=0}^{n-k} \left[\sum_{i=0}^{j} \binom{n}{j-i} \binom{n-j+i-k}{i} (-1)^i \right] F^j(x) \\ &= \sum_{j=0}^{n-k} \binom{j+k-1}{k-1} F^j(x),\end{aligned}$$

we can write (3.10) as

$$(3.11)\qquad E_1 = \sum_{m=1}^{n-k+1} E_1(m) = \int_{-\infty}^{\infty} \left(-\frac{\partial^2 \log f(x)}{\partial\theta^2}\right) kf(x)(1-F(x))^{k-1} \sum_{i=0}^{n-k} \binom{i+k-1}{k-1} F^i(x) dx.$$

Note that

$$\sum_{\delta(m)\in\Delta(m)} \delta_i \prod_{j=1}^{m} F^{\delta_j}(r_j) = \sum_{\delta(m)\in\Delta(m)} \frac{F(r_i)}{f(r_i)} \frac{\partial}{\partial r_i} \prod_{j=1}^{m} F^{\delta_j}(r_j)$$
$$= \frac{F(r_i)}{f(r_i)} \frac{\partial}{\partial r_i} \left[\text{coefficient of } s^{n-m-k+1} \text{ in } \prod_{j=1}^{m} \frac{1}{1-F(r_j)s} \right]$$
$$= \text{coefficient of } s^{n-m-k+1} \text{ in } \frac{F(r_i)}{1-F(r_i)s} s \prod_{j=1}^{m} \frac{1}{1-F(r_j)s}.$$

By applying this relation and following the same arguments as for E_1, we can derive E_2 from (3.3) to have the form

$$E_2 = \int_{-\infty}^{\infty} \left(-\frac{\partial^2 \log F(x)}{\partial\theta^2}\right) kf(x)(1-F(x))^{k-1} \sum_{i=1}^{n-k} \binom{k+i-1}{k} F^i(x)dx. \tag{3.12}$$

Since $R_{N_{n,k},k} = X_{n-k+1:n}$, the last expectation in (3.3) can be written as

$$E_3 = (k-1) \int_{-\infty}^{\infty} - \frac{\partial^2 \log(1-F(x))}{\partial\theta^2} \frac{n!}{(k-1)!(n-k)!} F^{n-k}(x)(1-F(x))^{k-1} f(x)dx. \tag{3.13}$$

The result now readily follows from (3.3), (3.11), (3.12) and (3.13). □

For numerical calculation of $I_{RT}^U(n,k)$, it may be easier to use the following expression, which follows from (3.10) and a similar formula for E_2:

$$I_{RT}^U(n,k) = \int_{-\infty}^{\infty} \frac{kf(x)}{(1-F(x))} \left(-\frac{\partial^2 \log f(x)}{\partial\theta^2} - \frac{F(x)}{1-F(x)} \frac{\partial^2 \log F(x)}{\partial\theta^2}\right)$$
$$\times \left(1 - \sum_{l=n-k}^{n} \binom{n}{l} F^l(x)(1-F(x))^{n-l}\right) dx$$
$$+ \int_{-\infty}^{\infty} k\binom{n}{k} f(x)(1-F(x))^{k-1} F^{n-k}(x)$$
$$\times \left(-\frac{\partial^2 \log f(x)}{\partial\theta^2} - (k-1)\frac{\partial^2 \log(1-F(x))}{\partial\theta^2}\right) dx.$$

Remark. Hofmann and Nagaraja (2003) showed that $I_{RT}^U(n,1) = \sum_{i=1}^{n} \frac{1}{i} I(X_{i:i})$, where $I(X_{i:i})$ is the FI in $X_{i:i}$. Unfortunately, an equivalent relation of the type $I_{RT}^U(n,k) = \sum_{i=1}^{n-k+1} \alpha_i I(X_{i-k+1:i})$, where the α_i's are constants, does not hold for $k \geq 2$.

Let us now look at the information in only the k-record values. From (2.3), we have the joint likelihood of the upper k-record values $R_{1,k}, \ldots, R_{N_{n,k},k}$ as

$$f(r_1, \ldots, r_m, N_{n,k} = m) = k^m \prod_{i=1}^{m} f(r_i) \sum_{\delta(m)\in\Delta(m)} \prod_{i=1}^{m} F^{\delta_i}(r_i)(1-F(r_m))^{k-1} \tag{3.14}$$

for $r_1 < \cdots < r_m$, $m = 1, \ldots, n+1-k$, and $\delta(m)$, $\Delta(m)$ as given in (3.2). The FI contained in the upper k-record values can be expressed as

$$\begin{aligned}(3.15)\quad & I_R^U(n,k)\\ &= E\left[\left(-\frac{\partial \log f(R_{1,k},\ldots,R_{m,k},N_{n,k}=m)}{\partial\theta}\right)^2\right]\\ &= E\left[\left(\frac{\sum_{i=1}^{N_{n,k}} \frac{\partial \log F(R_{i,k})}{\partial\theta} \sum_{\delta(N_{n,k})\in\Delta(N_{n,k})} \delta_i \prod_{j=1}^{N_{n,k}} F^{\delta_j}(R_{j,k})}{\sum_{\delta(N_{n,k})\in\Delta(N_{n,k})} \prod_{j=1}^{N_{n,k}} F^{\delta_j}(R_{j,k})}\right.\right.\\ &\qquad \left.\left. + \sum_{i=1}^{N_{n,k}} \frac{\partial \log f(R_{i,k})}{\partial\theta} + (k-1)\frac{\partial \log(1-F(R_{m,k}))}{\partial\theta}\right)^2\right],\end{aligned}$$

where the expectation is taken with respect to (3.14). The expression does not seem to allow much algebraic simplification. However, we can generalize a method given in Hofmann and Nagaraja (2003) for record values ($k = 1$). It permits fast calculation of the term under the expectation, in the cases of location, scale and certain shape parameters. We can then obtain $I_R^U(n,k)$ by simulation.

THEOREM 3.2. *Let θ be a parameter of one of the following families:*
(i) *location*: $F(x) = F_0(x-\theta)$
(ii) *scale*: $F(x) = F_0(\theta x)$
(iii) *shape*: $F(x) = F_0(x^\theta)$,
where F_0 is free of θ, and the corresponding density is denoted by f_0. Then,

$$(3.16)\quad I_R^U(n,k) = h_1(\theta) E\left[\left(\frac{A(Y_1,\ldots,Y_{N_{n,k}},N_{n,k})}{B(Y_1,\ldots,Y_{N_{n,k}},N_{n,k})} + C(Y_1,\ldots,Y_{N_{n,k}},N_{n,k}) + (k-1)D(Y_{N_{n,k}})\right)^2\right],$$

where $Y_1,\ldots,Y_{N_{n,k}}$ are the k-record values from a sample of size n from the standard distribution F_0,

$$A(Y_1,\ldots,Y_m,m) = \text{coefficient of } s^{n-k-m} \text{ in } \left(\sum_{i=1}^m \frac{f_0(Y_i)h_2(Y_i)}{1-F_0(Y_i)s}\right)\left(\prod_{i=1}^m \frac{1}{1-F_0(Y_i)s}\right),$$

$$B(Y_1,\ldots,Y_m,m) = \text{coefficient of } s^{n+1-k-m} \text{ in } \left(\prod_{i=1}^m \frac{1}{1-F_0(Y_i)s}\right).$$

For the location, scale, and shape parameter cases, C, D, h_1, h_2 are given, respectively, by

(i) $C(Y_1,\ldots,Y_m,m) = \sum_{i=1}^m -\frac{f_0'(Y_i)}{f_0(Y_i)}$, $D(Y) = \frac{f_0(Y)}{1-F_0(Y)}$, $h_1(\theta) \equiv 1$, $h_2(Y_i) \equiv -1$

(ii) $C(Y_1,\ldots,Y_m,m) = m + \sum_{i=1}^m \frac{Y_i f_0'(Y_i)}{f_0(Y_i)}$, $D(Y) = -\frac{Y f_0(Y)}{1-F_0(Y)}$, $h_1(\theta) = \frac{1}{\theta^2}$, $h_2(Y_i) = Y_i$

(iii) $C(Y_1,\ldots,Y_m,m) = m + \sum_{i=1}^{m}(1 + \frac{Y_i f_0'(Y_i)}{f_0(Y_i)})\log Y_i$, $D(Y) = -\frac{Y f_0(Y)\log Y}{1-F_0(Y)}$, $h_1(\theta) = \frac{1}{\theta^2}$, $h_2(Y_i) = Y_i \log Y_i$.

The information $I_R^L(n,k)$ in the lower records is obtained upon replacing $F_0(x)$ by $1 - F_0(x)$, A by $-A$ and D by $-D$.

PROOF. The proof works by reduction of the expression under the expectation in (3.15), for the three cases. See Hofmann and Nagaraja ((2003), Theorem 2.2) for details. □

4. Information in exponential k-record data

Let us consider the k-records from an exponential random sample $X_1,\ldots,X_n$ with cdf $F(x) = 1-\exp(-\theta x)$. We will evaluate the FI in upper and lower k-record values and k-record times, as well as just in k-record values. Some straightforward simplification reduces (3.1) to the following formulas:

$$\theta^2 I_{RT}^L(n,k) = \begin{cases} \displaystyle\sum_{i=1}^{n}\frac{1}{i} & \text{for } k=1 \\ \displaystyle 2\sum_{i=2}^{n}\frac{1}{i} + 2n(n-1)\sum_{i=0}^{\infty}\frac{1}{(n+i)^3} & \text{for } k=2 \\ \displaystyle k\sum_{i=k}^{n}\frac{1}{i} + 2n(n-1)\binom{n-2}{k-2} & \\ \displaystyle\qquad\times\sum_{i=0}^{k-3}(-1)^i\frac{\binom{k-3}{i}}{(n-k+i+2)^3} & \text{for } k\geq 3, \end{cases} \tag{4.1}$$

$$\theta^2 I_{RT}^U(n,k) = \sum_{i=k}^{n}\frac{k}{i} + \sum_{j=0}^{\infty}\frac{2kI_{\{n>k\}}}{(k+1+j)^3} + \sum_{i=2}^{n-k}\binom{i+k-1}{k}\sum_{j=0}^{i-2}(-1)^j\binom{i-2}{j}\frac{2k}{(k+1+j)^3}, \tag{4.2}$$

where, as before, $I_{\{n>k\}}$ is the indicator function for $n>k$. Asymptotically, for $n\to\infty$, k fixed, this implies $\theta^2 I_{RT}^L(n,k) = k\log n + o(\log n)$, $\theta^2 I_{RT}^U(n,k) = \frac{k}{3}(\log\frac{n}{k})^3 + o((\log n)^3)$.

To calculate the information $I_R(n,k)$ in only the k-record values, Theorem 3.2 is used in conjunction with simulation. We used S-Plus to generate standard exponential random samples, and to extract the k-record values, and MAPLE to evaluate the expressions under the expectation in (3.16). We performed 500,000 simulations, which is sufficient for 3-leading-digit accuracy. The results are presented in Tables 1 and 2. Recall that the FI in a single exponential observation is $\theta^2 I_{X_1} = 1$.

We first note that upper k-records contain more FI than lower ones. Let us look at Table 1 for upper k-records. We can see that for small n ($n\leq 5$) the information decreases with increasing k. This is expected, since the mean number of k-records (E) also declines. For somewhat larger n ($n = 10, 20$), where we considered $k = 1,\ldots,\frac{n}{2}$, we observe that the information peaks for some k, although the expected number of k-records keeps increasing. For practical purposes, the most interesting case occurs when n is large and much greater than k, represented in our table by $n = 100, 200$. Here,

Table 1. FI in upper k-record data from an exponential distribution with mean $1/\theta$. The column labels R, RT and E stand for $\theta^2 I_R^U(n,k)$, $\theta^2 I_{RT}^U(n,k)$ and $EN_{n,k}$, respectively.

	$k=1$			$k=2$			$k=3$			$k=5$			$k=10$		
n	R	RT	E	R	RT	E	R	RT	E	R	RT	E	R	RT	E
2	1.90	1.90	1.50	1	1	1									
3	2.73	2.74	2.83	1.97	1.97	1.67	1	1	1						
4	3.48	3.52	2.08	2.91	2.92	2.17	1.99	1.99	1.75						
5	4.18	4.25	2.28	3.80	3.83	2.57	2.96	2.97	2.35	1	1	1			
10	7.10	7.43	2.93	7.82	8.00	3.86	7.48	7.57	4.29	5.92	5.92	4.23	1	1	1
20	11.4	12.4	3.60	14.2	15.0	5.20	15.1	15.6	6.29	14.8	15.0	7.57	10.9	10.9	7.69
50	19.5	22.8	4.50	27.3	30.4	7.00	31.3	34.2	9.00	35.5	37.6	12.1	37.1	37.9	16.7
100	27.6	34.4	5.19	40.7	48.3	8.37	49.5	56.7	11.1	59.9	66.3	15.5	71.4	75.3	23.6
200	37.2	49.9	5.88	58.0	73.2	9.76	72.5	88.5	13.1	93.9	108	19.0	121	134	30.5

Table 2. FI in lower k-record data from an exponential distribution with mean $1/\theta$. The column labels R, RT and E stand for $\theta^2 I_R^L(n,k)$, $\theta^2 I_{RT}^L(n,k)$ and $EN_{n,k}$, respectively.

	$k=1$			$k=2$			$k=3$			$k=5$			$k=10$		
n	R	RT	E	R	RT	E	R	RT	E	R	RT	E	R	RT	E
2	1.50	1.50	1.50	1.81	1.81	1									
3	1.76	1.83	2.83	2.59	2.59	1.67	2.50	2.50	1						
4	1.94	2.08	2.08	3.06	3.13	2.17	3.52	3.53	1.75						
5	2.06	2.28	2.28	3.40	3.54	2.57	4.17	4.22	2.35	3.66	3.66	1			
10	2.37	2.93	2.93	4.27	4.85	3.86	5.75	6.26	4.29	7.83	8.07	4.23	5.86	5.86	1
20	2.61	3.60	3.60	4.90	6.19	5.20	6.91	8.29	6.29	10.3	11.5	7.57	15.7	16.3	7.69
50	2.77	4.50	4.50	5.42	8.00	7.00	7.94	11.0	9.00	12.5	16.1	12.1	22.1	25.7	16.7
100	2.85	5.19	5.19	5.69	9.37	8.37	8.36	13.1	11.1	13.7	19.5	15.5	25.4	32.6	23.6
200	2.88	5.88	5.88	5.80	10.8	9.76	8.72	15.1	13.1	14.2	23.0	19.0	27.5	39.5	30.5

the FI increases with k. Consider, for example, $n = 200$, $k = 5$. The total FI in 200 i.i.d. observations is $\theta^2 n I_{X_1} = 200$. With an expected number of 19 5-records and record times, we can capture more than half of that information. For all n, the relative difference between $\theta^2 I_R$ and $\theta^2 I_{RT}$ decreases with increasing k. Similar conclusions can be drawn for lower k-records as well (Table 2). In this case, the relative gain in FI for increasing k is even more pronounced.

5. Parameter estimation for the exponential distribution

In the last section, we discussed the FI contained in k-record data. Let us now see how we can exploit that information for parameter estimation based on k-record values and k-record times. We will look at the lower and upper record cases separately.

Lower k-records can be of interest in a number of industrial situations where it is desired to identify the minimum value or lower empirical percentiles of a batch of manufactured items (Samaniego and Whitaker (1986)). By comparing a new item only with the previous lower k-record value, this can be achieved with a minimum of pairwise comparisons. It is especially relevant when actual measurements are expensive, time consuming or destructive. For example, in destructive stress testing, by applying stress levels only up to the last lower k-record value, fewer items will be destroyed. Moreover, the broken items are the weakest (least valuable) of the lot. Samaniego and Whitaker (1986) give an unbiased estimator for the mean $\mu = \frac{1}{\theta}$ based on lower record values

$(k = 1)$ and record times as

$$\hat{\mu}_{SW} = \frac{\sum_{i=1}^{n}(\delta_i + 1)R_{i,1}}{(N_{n,1} + 1)E\left[\frac{N_{n,1}}{N_{n,1} + 1}\right]}, \tag{5.1}$$

with variance

$$\operatorname{Var}(\hat{\mu}_{SW}) = \mu^2 \frac{E\left[\frac{1}{N_{n,1} + 1}\right]}{1 - E\left[\frac{1}{N_{n,1} + 1}\right]}.$$

Hofmann and Nagaraja (2003) showed that $\hat{\mu}_{SW}$ is asymptotically efficient. For general k, the joint likelihood of the lower k-record values and k-record times from an exponential distribution can be obtained from (2.3) as

$$f(r_1, \ldots, r_m, t_1, \ldots, t_m, N_{n,k} = m) = k^m \theta^m \exp\left(-\theta \sum_{i=1}^{m} (\delta_i + 1) r_i\right) (1 - \exp(-\theta r_m))^{k-1},$$

for $t_1 = k < t_2 < \cdots < t_m \leq n$, $r_1 > \cdots > r_m$, $m = 1, \ldots, n - k + 1$, where $\delta_i = t_{i+1} - t_i - 1$ and $t_{m+1} = n + 1$. Conditionally on $N_{n,k} = m$, the bivariate statistic $(\sum_{i=1}^{m}(\delta_i + 1)R_{i,k}, R_{m,k})$ is minimal sufficient. One could therefore hope to find a reasonable estimator of the form $a(N_{n,k}) \sum_{i=1}^{m}(\delta_i + 1)R_{i,k} + b(N_{n,k})R_{N_{n,k},k}$. Unfortunately, the derivation of theoretical properties of such estimators turns out to be very messy. Through Monte Carlo simulation, we found the following simple unbiased estimator which possesses good small sample behavior:

$$\hat{\mu}_L = \frac{\hat{\mu}_h}{\mu^{-1} E \hat{\mu}_h}, \quad \text{where} \quad \hat{\mu}_h = \frac{\sum_{i=1}^{N_{n,k}} (\delta_i + 1) R_{i,k}}{N_{n,k} + k}.$$

Note that $\mu^{-1} E \hat{\mu}_h$ is free of μ. For $k = 1$, $\hat{\mu}_L$ is identical to $\hat{\mu}_{SW}$. Table 3 gives the efficiency of $\hat{\mu}_L$, when compared to the Cramer-Rao lower bound, i.e., efficiency$(\hat{\mu}_L) = \frac{\text{C-R bound}}{\operatorname{Var}(\hat{\mu}_L)} = \frac{1}{\operatorname{Var}(\hat{\mu}_L) I_{RT}^{L}(n,k)}$, for various choices of n and k. For $k = 1$, the efficiencies are around 90%. With increasing k they rise significantly, getting close to 99% for $k = 10$ in the case of moderate and large n. Table 4 gives the variance of $\hat{\mu}_L$. We observe a large decrease in the variance for increasing k, nicely reflecting the effect of increasing information content in the k-records. For comparison, we also state the variance $\operatorname{Var}(\hat{\mu}_n) = \frac{\mu^2}{n}$ of the estimator $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ based on the whole sample. Of course, the k-record based estimator should only be used if it is costly or unpractical to obtain exact measurements of the whole sample. Both tables were obtained from $5 \cdot 10^6$ simulations to provide the given accuracy.

For upper k-record values and k-record times from an exponential distribution, the likelihood (2.3) simplifies to

$$\begin{aligned} f(r_1, \ldots, r_m, t_1, \ldots, t_m, N_{n,k} = m) \\ = k^m \theta^m \exp\left(-\theta \sum_{i=1}^{m} r_i\right) \prod_{i=1}^{m} (1 - \exp(-\theta r_i))^{\delta_i} e^{-\theta r_m (k-1)}. \end{aligned}$$

Table 3. Efficiencies (compared to Cramer-Rao lower bound) of the unbiased lower k-record based estimator $\hat{\mu}_L$ of the exponential mean.

	n						
k	5	10	20	50	100	500	1000
1	.890	.877	.874	.877	.882	.896	.901
2	.950	.938	.933	.933	.937	.945	.947
3	.966	.960	.956	.955	.957	.962	.964
5		.968	.974	.973	.974	.976	.977
10			.970	.986	.987	.988	.989

Table 4. Variances $\mu^{-2}\operatorname{Var}(\hat{\mu}_L)$ of the unbiased lower k-record based estimator of the exponential mean μ.

	n						
k	5	10	20	50	100	500	1000
1	.492	.389	.318	.253	.219	.164	.148
2	.297	.220	.173	.134	.114	.0841	.0756
3	.245	.166	.126	.0952	.0800	.0581	.0520
5		.128	.0890	.0639	.0526	.0372	.0330
10			.0631	.0395	.0311	.0208	.0182
$\mu^{-2}\operatorname{Var}(\hat{\mu}_n)$	.200	.100	.050	.020	.010	.002	.001

We will obtain the MLE $\hat{\mu}_U = \frac{1}{\hat{\theta}_U}$ and investigate its small sample properties. The likelihood equation is given by

$$g(\theta) = \sum_{i=1}^{m}(\delta_i + 1)r_i + (k-1)r_m - \frac{m}{\theta} - \sum_{i=1}^{m}\frac{\delta_i r_i}{1-\exp(-\theta r_i)} = 0. \tag{5.2}$$

It is easy to see that $g(\theta)$ is strictly increasing for $\theta > 0$, with $\lim_{\theta\to 0+} g(\theta) = -\infty$, $\lim_{\theta\to\infty} g(\theta) = \sum_{i=1}^{m} r_i + (k-1)r_m > 0$. Hence, (5.2) has a unique solution. We numerically calculated the estimator from simulated samples from a standard exponential distribution ($\theta = 1$). Table 5 gives the simulated values of the mean and variance of $\hat{\mu}_U = \frac{1}{\hat{\theta}_L}$ for various choices of n and k, as obtained from 1,000,000 samples each. It also provides the ratio of $(I_{RT}^U(n,k))^{-1}$ and the variance. The MLE appears to be very efficient for moderate sample sizes. In these cases, neither bias nor efficiency are affected much by k. Hence, the variances decrease with k, as long as n is much greater than k. For k's approaching the order of n, the opposite effect occurs.

6. Information in inversely sampled k-record data

Most record based inference procedures given in the literature are based on inversely sampled records. They may be of greater value when generalized to take advantage of the more frequent k-records. Knowing the FI that k-record values and k-record times could provide about the parameter, would then be useful. Therefore, we will briefly state the corresponding expressions. We define inverse sampling for k-records as taking

Table 5. Properties of the maximum likelihood estimator $\hat{\mu}_U$ based on upper k-record data from an exponential distribution. True mean $\mu = 1$. The column labels E, Var and r stand for $E\hat{\mu}_U$, $\text{Var}(\hat{\mu}_U)$ and the ratio $\frac{(I_{RT}^U(n,k))^{-1}}{\text{Var}(\hat{\mu}_U)}$, respectively.

	$k=1$			$k=2$			$k=3$			$k=5$			$k=10$		
n	E	Var	r	E	Var	r	E	Var	r	E	Var	r	E	Var	r
5	1.25	.59	.40	1.33	.92	.28	1.50	2.2	.15						
10	1.11	.20	.67	1.13	.20	.63	1.14	.23	.58	1.20	.37	.46			
20	1.052	.098	.82	1.056	.082	.81	1.059	.080	.80	1.067	.086	.78	1.100	.14	.68
50	1.017	.048	.91	1.020	.036	.92	1.021	.032	.92	1.022	.029	.92	1.025	.029	.91
100	1.006	.031	.93	1.009	.022	.96	1.0098	.018	.96	1.010	.016	.96	1.011	.014	.96
500	.996	.014	.93	.999	.0086	.97	1.0003	.0068	.98	1.001	.0053	.990	1.0018	.0040	.993
1000	.994	.010	.92	.998	.0062	.97	.9993	.0048	.98	1.0001	.0036	.991	1.0007	.0026	.996

observations until a fixed number m of k-records is reached. Let $J_R^U(m,k)$ be the FI in the first m (upper) k-record values and $J_{RT}^U(m,k)$ be the FI in the first m (upper) k-record values and k-record times.

Note that the quantities $I_{RT}^U(n,k)$ (Sections 2–4) and $J_{RT}^U(m,k)$ are quite different. Both denote the FI in k-record values and k-record times. However, $I_{RT}^U(n,k)$ is the FI in the random number of k-record values (and k-record times) that occur in a fixed number (n) of i.i.d. observations. On the other hand, $J_{RT}^U(m,k)$ is the FI in a fixed number (m) of k-record values (and k-record times). Here, the length of the i.i.d. sequence necessary to observe them is random. There is no direct mathematical connection between $I_{RT}^U(n,k)$ and $J_{RT}^U(m,k)$.

It follows directly from (2.1) that

$$\begin{aligned} J_R^U(m,k) = &\sum_{i=1}^{m-1} k^i \int_{-\infty}^{\infty} \left(-\frac{\partial^2 \log f(x)}{\partial\theta^2} + \frac{\partial^2 \log(1-F(x))}{\partial\theta^2}\right) \\ &\times \frac{(-\log(1-F(x)))^{i-1}}{(i-1)!}(1-F(x))^{k-1} f(x)dx \\ &+ k^m \int_{-\infty}^{\infty} \left(-\frac{\partial^2 \log f(x)}{\partial\theta^2} - (k-1)\frac{\partial^2 \log(1-F(x))}{\partial\theta^2}\right) \\ &\times \frac{(-\log(1-F(x)))^{m-1}}{(m-1)!}(1-F(x))^{k-1} f(x)dx. \end{aligned} \tag{6.1}$$

From (2.3), we can see that

$$\begin{aligned} J_{RT}^U(m,k) = &\sum_{i=1}^{m} E\left[-\frac{\partial^2 \log f(R_{i,k})}{\partial\theta^2}\right] + \sum_{i=1}^{m-1} E\left[-\delta_i \frac{\partial^2 \log F(R_{i,k})}{\partial\theta^2}\right] \\ &+ (k-1)E\left[-\frac{\partial^2 \log(1-F(R_{m,k}))}{\partial\theta^2}\right]. \end{aligned} \tag{6.2}$$

However, $(\delta_i + 1)$ has a geometric distribution with parameter $1 - F(R_{i,k})$ when conditioned on $R_{i,k}$. This implies that $E[\delta_i \mid R_{i,k}] = \frac{F(R_{i,k})}{1-F(R_{i,k})}$. Therefore,

$$E\left[-\delta_i \frac{\partial^2 \log F(R_{i,k})}{\partial\theta^2}\right] = E\left[-\frac{\partial^2 \log F(R_{i,k})}{\partial\theta^2} \frac{F(R_{i,k})}{1-F(R_{i,k})}\right]. \tag{6.3}$$

Further, recall that $R_{i,k}$ behaves like the i-th record from $F_{1:k}(x) = 1 - (1 - F(x))^k$ (Dziubdziela and Kopocińsiki (1976)) which implies that the density of $R_{i,k}$ is given by

$$f_i(x) = \frac{(-k\log(1-F(x)))^{i-1}}{(i-1)!} kf(x)(1-F(x))^{k-1}. \tag{6.4}$$

Combining (6.2) with (6.3) and (6.4), we obtain

$$\begin{aligned} J_{RT}^{U}(m,k) = &\sum_{i=1}^{m-1} k^i \int_{-\infty}^{\infty} \left(-\frac{\partial^2 \log f(x)}{\partial \theta^2} - \frac{\partial^2 \log F(x)}{\partial \theta^2} \frac{F(x)}{1-F(x)} \right) \\ &\times \frac{(-\log(1-F(x)))^{i-1}}{(i-1)!} (1-F(x))^{k-1} f(x)dx \\ &+ k^m \int_{-\infty}^{\infty} \left(-\frac{\partial^2 \log f(x)}{\partial \theta^2} - (k-1) \frac{\partial^2 \log(1-F(x))}{\partial \theta^2} \right) \\ &\times \frac{(-\log(1-F(x)))^{m-1}}{(m-1)!} (1-F(x))^{k-1} f(x)dx. \end{aligned} \tag{6.5}$$

The corresponding formulas for lower records are obtained from (6.1) and (6.5) by replacing $F(x)$ with $1 - F(x)$.

REFERENCES

Ahmadi, J. and Arghami, N. R. (2001). On the Fisher information in record values, *Metrika*, **53**, 195–206.

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

Balakrishnan, N. and Chandrasekar, B. (2002). On a multiparameter version of Tukey's linear sensitivity measure and its properties, *Annals of the Institute of Statistical Mathematics*, **54**, 796–805.

Dziubdziela, W. and Kopociński, B. (1976). Limiting properties of the kth record values, *Zastosowania Matematyki*, **15**, 187–190.

Grudzién, Z. and Szynal, D. (1985). On the expectd values of kth record values and associated characterizations of distributions, *Probability and Statistical Decision Theory*, Vol. A (eds. F. Konecny, J. Mogyoródi and W. Wertz), 119–127, Reidel, Dordrecht.

Hofmann, G. (2003). Comparing the Fisher information in record data and random observations, *Statistical Papers* (to appear).

Hofmann, G. and Nagaraja, H. N. (2003). Information in record data, *Metrika*, **57**, 177–193.

Iyengar, S., Kvam, P. and Singh, H. (1999). Fisher information in weighted distributions, *The Canadian Journal of Statistics*, **27**, 833–841.

Nagaraja, H. N. (1994). Tukey's linear sensitivity and order statistics, *Annals of the Institute of Statistical Mathematics*, **46**, 757–768.

Park, S. (1996). Fisher information in order statistics, *Journal of the American Statistical Association*, **91**, 385–390.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed., Wiley, New York.

Raqab, M. Z. and Amin, W. A. (1997). A note on reliability properties of k-record statistics, *Metrika*, **46**, 245–251.

Samaniego, F. J. and Whitaker, L. R. (1986). On estimating popular characteristics from record breaking observations, *Naval Research Logistics Quarterly*, **33**, 531–543.

Tukey, J. W. (1965). Which part of the sample contains the information?, *Proceedings of the National Academy of Sciences of the United States of America*, **53**, 127–134.