Statistics and Probability Letters 79 (2009) 29–37

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Current k-records and their use in distribution-free confidence intervals

J. Ahmadi [a,*], M. Razmkhah [a], N. Balakrishnan [b]

[a] School of Mathematical Sciences, Ferdowsi University of Mashhad, P.O. Box 91775-1159, Mashhad, Iran

[b] Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada L8S 4K1

ARTICLE INFO

Article history:
Received 11 July 2008
Accepted 11 July 2008
Available online 25 July 2008

MSC:
primary 62G30
secondary 62G15

ABSTRACT

In a sequence of independent and identically distributed (iid) random variables, the kth largest (smallest) observation in a partial sample is well-known as the upper (lower) k-record value, when its value is greater (smaller) than the corresponding observation in the previous partial sample. In this paper, we consider the k-record statistics at the time when the nth k-record of any kind (either an upper or lower) is observed, termed as *current k-records*. We derive a general expression for the joint probability density function (pdf) of these current k-records and use it to construct distribution-free confidence intervals for population quantiles. It is shown that the expected width of these confidence intervals is decreasing in k and increasing in n. We also discuss the construction of tolerance intervals and limits in terms of current k-records. Finally, a numerical example is presented to illustrate all the methods of inference developed here.

1. Introduction

Let $\{X_i; i \geq 1\}$ be a sequence of iid continuous random variables. The first k observations in this sequence are called the first partial sample of size k. To achieve the next partial samples, we add the other observations of the interested sequence to the first partial sample one by one. The kth largest (smallest) observation in a partial sample is called the upper (lower) k-record statistics, when its value is greater (smaller) than the corresponding observation in the previous partial sample. Formally, let $X_{i:m}$ denote the ith order statistic from a random sample of size m. Then, the upper k-record times $T_{n,k}$ and the upper k-record values $U_{n,k}$ are defined as follows: $T_{1,k} = k$, $U_{1,k} = X_{1:k}$ 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 $U_{m,k} = X_{T_{m,k}-k+1:T_{m,k}}$. Lower k-record statistics can be defined analogously; see Arnold et al. (1998) for more details. In the special case when $k = 1$, we have the usual records. One can imagine situations wherein the largest and smallest observations are simultaneously recorded when a new record of either kind (upper or lower) occurs, such as in the case of weather data. These statistics are referred to as *current records* in the records literature.

Now, suppose $U'_{n,k}$ and $L'_{n,k}$ are the kth largest and kth smallest observations, respectively, when observing the nth k-record (upper or lower) from the sequence $\{X_n, n \geq 1\}$. We call such recent statistics *current k-records*. Of course, when new observations become available, new current k-records can arise. In infinite sequences, every new observation that is larger (smaller) than the recent upper (lower) current k-record will eventually become a current k-record. For example, let us consider the following sequence of observations:

$$3, 2, 2.5, 2.6, 1, 3.7, 2.2, 1.5, 2.7, 2.3, 0.5, \ldots.$$

* Corresponding author.
E-mail addresses: ahmadi-j@um.ac.ir (J. Ahmadi), razmkhah_m@yahoo.com (M. Razmkhah), bala@mcmaster.ca (N. Balakrishnan).

0167-7152/$ – see front matter
doi:10.1016/j.spl.2008.07.014

The current 3-records extracted from the above sequence are as follows:

n	1	2	3	4	5	6	7	8
$U'_{n,3}$	2	2.5	2.5	2.6	2.6	2.6	2.7	2.7
$L'_{n,3}$	3	2.6	2.5	2.5	2.2	2	2	1.5

Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ denote the order statistics of a sample of size n (with $n \geq 2k-1$). The spacing $W_{i,j:n} = X_{j:n} - X_{i:n}$ is a generalization of the sample range $W_n = X_{n:n} - X_{1:n}$. A special case of $W_{i,j:n}$ called the quasirange, is $W_{k:n} = X_{n-k+1:n} - X_{k:n}$; see Arnold et al. (1992). Let $R_{m,k}(m = k, k+1, \ldots)$ denote the mth usual record in the sequence of quasiranges $\{W_{k:n}, n \geq 2k-1\}$. Then, it can be seen that $R_{m,k}$ is the mth current k-record range in the $\{X_n\}$-sequence given by

$$R_{m,k} = U'_{m,k} - L'_{m,k}, \quad m = k, k+1, \ldots.$$

Clearly, $R_{k,k} = 0$ and $R_{k+1,k} = W_{k:2k}$. Intuitively, for fixed m, $R_{m,k+1} < R_{m,k}$ with probability one and for fixed k, $R_{m,k}$ is increasingly ordered. For fixed k, let $N_k(c)$ denote the stopping time such that $N_k(c) = \inf\{n \geq k; R_{n,k} > c\}$, where c is an arbitrary fixed number. Then, $N_k(c)$ is the waiting time until the k-record range of an iid sample exceeds a given value c. Some works have been done for the special case $k = 1$; see, for example, Basak (2000), Houchens (1984), Ahmadi and Balakrishnan (2004, 2005), and Raqab (in press). In this paper, we first develop the distribution theory for the current k-records and then use it to develop some nonparametric inferential procedures based on current k-records.

The population quantile ξ_p $(0 < p < 1)$ of the distribution function F is defined by $\xi_p = \inf\{x : F(x) \geq p\}$. Let p and q be any given real numbers such that $0 < p < q < 1$; then, $\left(\xi_p, \xi_q\right)$ is called a quantile interval and is given by $\{x | p \leq F(x) \leq q\}$. Several authors have discussed construction of confidence intervals for these quantiles and quantile intervals. Arnold et al. (1992, p. 183) have described how order statistics can be used to provide distribution-free confidence intervals for population quantiles; see also David and Nagaraja (2003). Ahmadi and Arghami (2003) obtained similar results based on record data. Ahmadi and Balakrishnan (2004, 2005) developed distribution-free confidence intervals for quantiles and quantile intervals on the basis of current records. Construction of distribution-free confidence intervals for quantiles and tolerance intervals on the basis of current k-records is the focus of this paper.

The rest of this paper is organized as follows. In Section 2, we derive some distributional results and in particular a general expression for the joint pdf of the nth upper and lower current k-records. Confidence intervals for quantiles in terms of current k-records are then derived in Section 3. In Section 4, we discuss the construction of tolerance intervals and limits on the basis of current k-records. Finally, in Section 5, we present a numerical example to illustrate all the methods of inference developed here.

2. Distributions of current k-records

Let $\{X_i; i \geq 1\}$ be a sequence of iid continuous random variables. To get the first current k-record, the first partial sample of size k, $\{X_1, \ldots, X_k\}$, from the above sequence is needed. The kth largest (smallest) observation among them is defined as the first upper (lower) current k-record. That is, $U'_{1,k} = X_{1:k}$ and $L'_{1,k} = X_{k:k}$. If $\{X_{k+1} < L'_{1,k}$ or $X_{k+1} > U'_{1,k}\}$, then X_{k+1} creates the second current k-record. Since $U'_{1,k} < L'_{1,k}$, the aforementioned event occurs with probability 1, and therefore X_{k+1} certainly induces the second current k-record. Formally, we have $U'_{2,k} = X_{2:k+1}$ and $L'_{2,k} = X_{k:k+1}$, and so $U'_{2,k} < L'_{2,k}$. Therefore, the second current k-record arises by adding only one observation to the first partial sample of size k. The nth current k-record arises in the same way when $n \leq k+1$. For example, the $(k-1)$th current k-record is then obtained from the first $2k-2$ observations as $U'_{k-1,k} = X_{k-1:2k-2}$ and $L'_{k-1,k} = X_{k:2k-2}$. Note that for $n < k$, we have $U'_{n,k} < L'_{n,k}$. Similarly, the kth current k-record arises from the first $2k-1$ observations such that the kth largest (smallest) observation is the kth upper (lower) current k-record, i.e., $U'_{k,k} = L'_{k,k} = X_{k:2k-1}$. Since the event $\{X_{2k} < L'_{k,k}$ or $X_{2k} > U'_{k,k}\}$ occurs with probability 1, X_{2k} induces the $(k+1)$th current k-record. That is, $U'_{k+1,k} = X_{k+1:2k}$ and $L'_{k+1,k} = X_{k:2k}$, and hence $L'_{k+1,k} < U'_{k+1,k}$. Summing up, we have

$$L'_{n,k} = X_{k:k+n-1} \quad \text{and} \quad U'_{n,k} = X_{n:k+n-1}, n \leq k+1. \tag{1}$$

Consequently, when $n \leq k+1$, each new observation after the first partial sample of size k induces a new current k-record, but this may be false thereafter. Note that if $L'_{k+1,k} < X_{2k+1} < U'_{k+1,k}$, then X_{2k+1} cannot induce a new current k-record and we then have to wait for an observation X_m $(m \geq 2k+1)$ such that $X_m < L'_{k+1,k}$ or $X_m > U'_{k+1,k}$.

Therefore, for $n \leq k+1$, the joint pdf of the nth value of the kth current records, $(L'_{n,k}, U'_{n,k})$, is readily obtained from the distributions of order statistics. But, for $n > k+1$, the distribution of the nth current k-record does not follow readily from the distributions of order statistics (see Lemma 1). We therefore had to come up with a new way to derive the joint pdf of $(L'_{n,k}, U'_{n,k})$ when $n > k+1$. We obtained an expression for this joint density function and this is what is presented in Theorem 1.

Lemma 1 (Arnold et al., 1992). *Let the ith order statistic from a random sample of size m from the uniform $U(0, 1)$ distribution be denoted by $X^*_{i:m}$. Then, the marginal cdf of $X^*_{i:m}$ is*

$$F_{X^*_{i:m}}(x) = \sum_{r=i}^{m} \binom{m}{r} x^r (1-x)^{n-r}, \quad 0 < x < 1. \tag{2}$$

For $n > k + 1$, we have the following theorem for the joint density of the upper and lower current k-records. It may be noted that for the special case $k = 1$, the result reduces to the known result for current records presented, for example, by Houchens (1984).

Theorem 1. *Let $\{X^*_i, i \geq 1\}$ be a sequence of iid $U(0, 1)$ random variables. Then, the joint density of the nth $(n > k + 1)$ lower and upper current k-records $(L^{*'}_{n,k}, U^{*'}_{n,k})$ is*

$$f_{n,k}(x, y) = \int_x^y \frac{1}{t+1-y} f_{n-1,k}(t, y)\mathrm{d}t + \int_x^y \frac{1}{x+1-s} f_{n-1,k}(x, s)\mathrm{d}s, \quad x < y. \tag{3}$$

Proof. Houchens (1984) proved this result for the usual current records (case $k = 1$). Now, for the nth $(n > k + 1)$ current k-record, we have

$$P(L^{*'}_{n,k} \leq x, U^{*'}_{n,k} > y | L^{*'}_{n-1,k} = t, U^{*'}_{n-1,k} = s) = \begin{cases} I_1, & x \geq t, y < s, \\ I_2, & x < t, y \geq s, \\ I_3, & x < t, y \leq s, \\ I_4, & x \geq t, y > s, \end{cases}$$

where $0 < x < y < 1$ and $0 < t < s < 1$. It is evident that $I_1 = 1$ and $I_2 = 0$. Let Z_1 be the first observation after $(L^{*'}_{n-1,k}, U^{*'}_{n-1,k})$ from the $U(0, 1)$ distribution. Then,

$$\begin{aligned} I_3 &= P(L^{*'}_{n,k} < x, U^{*'}_{n,k} > y | Z_1 < t, L^{*'}_{n-1,k} = t, U^{*'}_{n-1,k} = s)P(Z_1 < t) \\ &\quad + P(L^{*'}_{n,k} < x, U^{*'}_{n,k} > y | t < Z_1 < s, L^{*'}_{n-1,k} = t, U^{*'}_{n-1,k} = s)P(t < Z_1 < s) \\ &\quad + P(L^{*'}_{n,k} < x, U^{*'}_{n,k} > y | Z_1 > s, L^{*'}_{n-1,k} = t, U^{*'}_{n-1,k} = s)P(Z_1 > s). \end{aligned}$$

The event $\{Z_1 < t, L^{*'}_{n-1,k} = t, U^{*'}_{n-1,k} = s\}$ is equivalent to $\{L^{*'}_{n,k} < L^{*'}_{n-1,k}, U^{*'}_{n,k} = U^{*'}_{n-1,k}, L^{*'}_{n,k} \stackrel{\mathrm{d}}{=} U(0, t)\}$, where $\stackrel{\mathrm{d}}{=}$ means identical in distribution. Similarly, the event $\{Z_1 > s, L^{*'}_{n-1,k} = t, U^{*'}_{n-1,k} = s\}$ is equivalent to $\{L^{*'}_{n,k} = L^{*'}_{n-1,k}, U^{*'}_{n,k} > U^{*'}_{n-1,k}, U^{*'}_{n,k} \stackrel{\mathrm{d}}{=} U(s, 1)\}$. The event $\{t < Z_1 < s, L^{*'}_{n-1,k} = t, U^{*'}_{n-1,k} = s\}$ indicates the fact that the first observation after $(L^{*'}_{n-1,k}, U^{*'}_{n-1,k})$ is not a new current record. Therefore,

$$I_3 = \frac{x}{t} \times t + I_3 \times (s - t),$$

and so we obtain

$$I_3 = \frac{x}{t+1-s}.$$

Similarly, we can show that

$$I_4 = \frac{1-y}{t+1-s}.$$

Proceeding now on lines similar to those of Houchens (1984), the expression of the joint pdf of the nth lower and upper current k-records in Eq. (3) can be derived. □

For $n > k + 1$, using Theorem 1, the joint density of the nth upper and lower current k-records can be obtained for all n in a sequential manner. Though the algebraic calculations are generally too cumbersome, we present here the expressions for the special cases $n = k + 2$ and $n = k + 3$, in Lemmas 2 and 3, respectively, when the underlying distribution is $U(0, 1)$.

Lemma 2. *Under the assumptions of Theorem 1, we have the following:*

(i) *The joint pdf of $(L^{*'}_{k+2,k}, U^{*'}_{k+2,k})$ is, for $0 < x < y < 1$,*

$$f_{k+2,k}(x, y) = \frac{(2k)!}{[(k-1)!]^2} \left\{ (1-y)^{k-1} \varphi_1(x, y, k) + x^{k-1} \varphi_1(1-y, 1-x, k) \right\}, \tag{4}$$

where

$$\varphi_1(x,y,k) = -(y-1)^{k-1}\log(x+1-y) + \sum_{i=1}^{k-1}\binom{k-1}{i}(y-1)^{k-1-i}\left[\frac{1-(x+1-y)^i}{i}\right]. \tag{5}$$

(ii) *The marginal pdf of* $L^{*'}_{k+2,k}$ *is*

$$f_{L^{*'}_{k+2,k}}(x) = \Phi(x,k), \quad 0<x<1,$$

where

$$\begin{aligned}\Phi(x,k) = \frac{(2k)!}{[(k-1)!]^2}(-1)^{k-1}\Bigg\{ & x^{2k-2}\left[1-x(1-\log x)\right] \\ & + \sum_{i=0}^{2k-2}\binom{2k-2}{i}\frac{(-x)^{2k-i-2}}{(i+1)^2}\left[1-x^{i+1}(1-(i+1)\log x)\right] \\ & + \sum_{i=1}^{k-1}\binom{k-1}{i}\frac{(-1)^i}{i}\left[\frac{(1-x)^{2k-i-1}}{2k-i-1} - \sum_{j=0}^{i}\binom{i}{j}x^{i-j}\frac{(1-x)^{2k-i+j-1}}{2k-i+j-1}\right] \\ & + \sum_{i=1}^{k-1}\binom{k-1}{i}\frac{(-x)^{2k-i-2}}{i}\left[1-x+\frac{x^{i+1}-1}{i+1}\right]\Bigg\}. \end{aligned} \tag{6}$$

(iii) *The marginal pdf of* $U^{*'}_{k+2,k}$ *is*

$$f_{U^{*'}_{k+2,k}}(x) = \Phi(1-x,k), \quad 0<x<1,$$

where $\Phi(x,k)$ *is as defined in* (6).

Lemma 3. *Under the assumptions of Theorem 1, the joint pdf of* $(L^{*'}_{k+3,k}, U^{*'}_{k+3,k})$ *is, for* $0<x<y<1$,

$$f_{k+3,k}(x,y) = \frac{(2k)!}{[(k-1)!]^2}\left\{(1-y)^{k-1}\varphi_2(x,y,k) + x^{k-1}\varphi_2(1-y,1-x,k)\right\}, \tag{7}$$

where

$$\begin{aligned}\varphi_2(x,y,k) = {} & (y-1)^{k-1}(\log(x+1-y))^2 - \log(x+1-y)\sum_{i=1}^{k-1}\binom{k-1}{i}(y-1)^{k-1-i}\frac{[1+(-1)^i]}{i} \\ & - \sum_{i=1}^{k-1}\binom{k-1}{i}\frac{(y-1)^{k-1-i}}{i^2}[1-(x+1-y)^i] \\ & + \sum_{i=1}^{2k-2}\binom{2k-2}{i}\frac{(y-1)^{k-1-i}}{i^2}\left[1-(x+1-y)^i(1-i\log(x+1-y))\right] \\ & + \sum_{i=1}^{k-1}\binom{k-1}{i}\frac{(-1)^i}{i}\sum_{j=1}^{2k-2-i}\binom{2k-2-i}{j}(y-1)^{k-1-i-j}\frac{[1-(x+1-y)^j]}{j} \\ & - \sum_{i=1}^{k-1}\binom{k-1}{i}\frac{(-1)^i}{i}\sum_{j=0}^{2k-2-i}\binom{2k-2-i}{j}(y-1)^{k-1-i-j}\frac{[1-(x+1-y)^{i+j}]}{i+j}. \end{aligned}$$

The marginal densities can be readily obtained from (7) *by integration.*

For other values of n $(n \geq k+4)$, the joint density of the nth upper and lower current k-records can be obtained in a similar recursive manner.

3. Distribution-free confidence intervals for quantiles

As mentioned in Section 2, for n less than, equal to, or greater than k, $U^{*'}_{n,k}$ is less than, equal to, or greater than $L^{*'}_{n,k}$, respectively, with probability 1. Therefore, for $n<k$, a confidence interval for a quantile is in the form $(U^{*'}_{n,k}, L^{*'}_{n,k})$. For $n=k$, lower and upper current k-records coincide and consequently a confidence interval cannot be constructed in this case. For

$n > k$, a confidence interval is in the form $(L^{*'}_{n,k}, U^{*'}_{n,k})$. Next, the coverage probability of the confidence interval based on the nth current k-record (when $n \le k+1$) for ξ_p can be obtained easily from Eqs. (1) and (2). But, when $n > k+1$, we can utilize Theorem 1 (and Lemmas 2 and 3) for determining the required coverage probabilities.

Lemma 4. *Let $\{X_i, i \ge 1\}$ be a sequence of iid random variables with cdf F and pdf f.*
(i) *For $n < k$, the coverage probability of the event $\{U'_{n,k} \le \xi_p \le L'_{n,k}\}$ is*

$$\alpha(n, k; p) = \sum_{r=n}^{k-1} \binom{n+k-1}{r} p^r (1-p)^{n+k-1-r}; \tag{8}$$

(ii) *The coverage probability of the event $\{L'_{k+1,k} \le \xi_p \le U'_{k+1,k}\}$ is*

$$\alpha(k+1, k; p) = \binom{2k}{k} p^k (1-p)^k. \tag{9}$$

Proof. Let $X_i^* = F(X_i)$, then we immediately have

$$U^{*'}_{n,k} \left(L^{*'}_{n,k}\right) \quad \overset{\mathrm{d}}{=} \quad F(U'_{n,k}) \left(F(L'_{n,k})\right), \tag{10}$$

where $U^{*'}_{n,k} \left(L^{*'}_{n,k}\right)$ is the nth upper (lower) current k-record statistic associated with X_i^*'s which are iid uniform $U(0, 1)$ random variables. On the other hand, from Eq. (1), we have for $n < k$

$$\begin{aligned} P\{U'_{n,k} \le \xi_p \le L'_{n,k}\} &= P\{F(U'_{n,k}) \le p \le F(L'_{n,k})\} \\ &= P\{U^{*'}_{n,k} \le p \le L^{*'}_{n,k}\} \\ &= P\{U^{*'}_{n,k} \le p\} - P\{L^{*'}_{n,k} \le p\} \\ &= P\{X^*_{n:k+n-1} \le p\} - P\{X^*_{k:k+n-1} \le p\}. \end{aligned}$$

Thus, the expression in (8) follows readily from (2). Eq. (9) can be obtained similarly. □

For $n = k+2$, we obtain the following result.

Lemma 5. *Under the assumptions of Lemma 4, the coverage probability of the event $\{L'_{k+2,k} \le \xi_p \le U'_{k+2,k}\}$ is*

$$\alpha(k+2, k; p) = \vartheta_k(p; k+2) + \vartheta_k(1-p; k+2) - 1, \tag{11}$$

where

$$\begin{aligned} \vartheta_k(p; k+2) = {} & \frac{(2k)!}{[(k-1)!]^2} (-1)^{k-1} \left\{ \frac{p^{2k-1}}{2k-1} - \frac{p^{2k}}{2k} \left(\frac{2k+1}{2k} - \log p \right) \right. \\ & + \sum_{i=0}^{2k-2} \binom{2k-2}{i} \frac{(-1)^i}{i+1} \left[\frac{p^{2k-i-1}}{(i+1)(2k-i-1)} - \frac{p^{2k}}{2k} \left(\frac{1}{i+1} + \frac{1}{2k} - \log p \right) \right] \\ & + \sum_{i=1}^{k-1} \binom{k-1}{i} \frac{(-1)^i}{i} \left[\frac{p^{2k-i-1}}{2k-i-1} \frac{i}{i+1} - \frac{p^{2k-i}}{2k-i} + \frac{p^{2k}}{2k(i+1)} \right] \\ & + \sum_{i=1}^{k-1} \binom{k-1}{i} \frac{1}{i} \left[- \frac{p^{2k-i}}{(2k-i-1)(2k-i)} - \sum_{j=0}^{i} \binom{i}{j} \frac{(-1)^i}{2k-i+j-1} \right. \\ & \left. \left. \times \sum_{s=0}^{2k-i+j-1} \binom{2k-i+j-1}{s} (-1)^s \frac{p^{i-j+s+1}}{i-j+s+1} \right] \right\}. \end{aligned} \tag{12}$$

Proof. As in the proof of Lemma 4, we first write

$$P(L'_{n,k} \le \xi_p \le U'_{n,k}) = P(L^{*'}_{n,k} \le p) - P(U^{*'}_{n,k} \le p), \quad n > k.$$

Next, by Lemma 2, we have

$$P(U^{*'}_{k+2,k} \le p) = \int_0^p \Phi(1-x, k) \mathrm{d}x = 1 - \int_0^{1-p} \Phi(x, k) \mathrm{d}x = 1 - P(L^{*'}_{k+2,k} \le 1-p). \tag{13}$$

Table 1
Values of $\alpha(n, k; p)$ for some choices of n, k and p

p	k	n = 1	2	3	4	5	6	7	8
	1	–	0.180	0.323	0.486	0.647	0.782	0.878	0.938
0.1	2	0.181	–	0.015	0.176	0.337	0.515	0.681	0.811
	3	0.271	0.049	–	0.005	0.134	0.286	0.463	0.636
	1	–	0.320	0.546	0.726	0.855	0.932	0.972	0.989
0.2	2	0.318	–	0.082	0.385	0.600	0.773	0.888	0.951
	3	0.480	0.154	–	0.046	0.309	0.536	0.728	0.862
	1	–	0.420	0.694	0.856	0.940	0.978	0.993	0.998
0.3	2	0.418	–	0.185	0.569	0.779	0.902	0.963	0.988
	3	0.636	0.264	–	0.136	0.497	0.732	0.878	0.952
	1	–	0.480	0.779	0.920	0.975	0.993	0.998	0.999
0.4	2	0.478	–	0.276	0.692	0.879	0.956	0.988	0.997
	3	0.720	0.346	–	0.233	0.641	0.854	0.950	0.985
	1	–	0.500	0.807	0.940	0.984	0.997	0.999	0.999
0.5	2	0.500	–	0.312	0.735	0.912	0.976	0.995	0.999
	3	0.750	0.372	–	0.274	0.695	0.896	0.971	0.993

On the other hand, by using (6), we have

$$\begin{aligned} P(L'_{k+2,k} \le \xi_p) &= \int_0^p \Phi(x,k)\mathrm{d}x \\ &= \frac{(2k)!(-1)^k}{[(k-1)!]^2} \Bigg\{ \int_0^p x^{2k-2}\mathrm{d}x - \int_0^p x^{2k-1}\mathrm{d}x + \int_0^p x^{2k-1}\log x\mathrm{d}x \\ &\quad + \sum_{i=0}^{2k-2} \binom{2k-2}{i} \frac{(-1)^i}{(i+1)^2} \left[\int_0^p x^{2k-2-i}\mathrm{d}x - \int_0^p x^{2k-1}\mathrm{d}x + (i+1)\int_0^p x^{2k-1}\log x\mathrm{d}x \right] \\ &\quad + \sum_{i=1}^{k-1} \binom{k-1}{i} \frac{(-1)^i}{i} \left[\int_0^p x^{2k-2-i}\mathrm{d}x - \int_0^p x^{2k-1-i}\mathrm{d}x + \frac{\int_0^p x^{2k-1}\mathrm{d}x - \int_0^p x^{2k-2-i}\mathrm{d}x}{i+1} \right] \\ &\quad + \sum_{i=1}^{k-1} \binom{k-1}{i} \frac{(-1)^i}{i} \left[\frac{\int_0^p (1-x)^{2k-1-i}\mathrm{d}x}{2k-1-i} - \sum_{j=0}^{i} \binom{i}{j} \frac{\int_0^p x^{i-j}(1-x)^{2k-i+j-1}\mathrm{d}x}{2k-i+j-1}\mathrm{d}x \right] \Bigg\}, \end{aligned}$$

where

$$\int_0^p x^{2k-1}(-\log x)\mathrm{d}x = \int_{-\log p}^{\infty} z e^{-2kz}\mathrm{d}z = p^{2k}(1-2k\log p)$$

and

$$\int_0^p x^{i-j}(1-x)^{2k-i+j-1}\mathrm{d}x = \sum_{s=0}^{2k-i+j-1} \binom{2k-i+j-1}{s} (-1)^s \int_0^p x^{i-j+s}\mathrm{d}x.$$

So, after some simplification, we obtain

$$P(L'_{k+2,k} \le \xi_p) = \vartheta_k(p; k+2). \tag{14}$$

Finally, upon using (13), the expression in Eq. (11) is obtained. □

For other values of n ($\ge k+3$), the coverage probabilities of the confidence intervals based on the nth current k-record for ξ_p can be derived in a similar manner. Table 1 presents the values of $\alpha(n, k; p)$ for some choices of p, k and n up to 8.

From Eqs. (8), (9) and (11), it can be shown that for all possible $n(n \le k+2)$, $\alpha(n, k; p)$ is symmetric with respect to $p = 0.5$. For other values of n, this can be easily deduced. That is, for $n > k$,

$$P(L'_{n,k} \le \xi_p \le U'_{n,k}) = \gamma_0 \Leftrightarrow P(L'_{n,k} \le \xi_{1-p} \le U'_{n,k}) = \gamma_0;$$

a similar result can be deduced for $n < k$. For the special case when $k = 1$, this corresponds to the results of Ahmadi and Balakrishnan (2004).

From Table 1, the following points may be observed:

1. For fixed p and k, $\alpha(n, k; p)$ is decreasing for $n < k$ and increasing for $n > k$;

2. For fixed p and n, $\alpha(n, k; p)$ is decreasing for $k < n$ and increasing for $k > n$;
3. For fixed n and k, $\alpha(n, k; p)$ is increasing for $p < 0.5$ and decreasing for $p > 0.5$.

Remark 1. It may be noted that for fixed n, the expected width of the confidence interval, viz., $E(R_{n,k})$, decreases as k increases and for fixed k, it increases with n.

4. Tolerance intervals and limits

An interval (A, B) is said to be a $100\beta\%$ tolerance interval with probability level ν if

$$P\left(F(B) - F(A) > \beta\right) = \nu,$$

where ν is the tolerance coefficient and the end-points A and B are the tolerance limits. Letting $A = -\infty$ or $B = \infty$, we simply obtain the upper or lower tolerance limits, respectively. Here, in this section, we show how the current k-records can be used to construct tolerance intervals and lower and upper tolerance limits. As was already noted before, for the cases $n < k$ and $n = k + 1$, tolerance intervals can be readily obtained from the order statistics literature; see Arnold et al. (1992) for more details. For the case when $n > k + 1$, we can use Theorem 1 for this purpose. For example, for the special case of $n = k + 2$, the corresponding result is presented in Lemma 6. For other values of n (viz., $n \geq k + 3$), a similar result can be presented, although the ensuing algebraic calculations are quite cumbersome.

Lemma 6. *Let $\{X_i, i \geq 1\}$ be a sequence of iid random variables with cdf F; then, $(L'_{k+2,k}, U'_{k+2,k})$ is a $100\beta\%$ tolerance interval for population F with tolerance coefficient given by*

$$\begin{aligned}\nu(k+2, k) &= 2\frac{(2k)!(-1)^{k-1}}{[(k-1)!]^2}\left\{\frac{(1-\beta)^{2k}}{4k^2(2k-1)}\left[1 - 2k\log(1-\beta)\right]\right.\\ &\quad\left. + \sum_{i=1}^{k-1}\binom{k-1}{i}\frac{(-1)^i}{i(2k-i-1)}\left[\frac{(1-\beta)^{2k-i}}{2k-i} - \frac{(1-\beta)^{2k+1-i}}{2k+1-i}\right]\right\}. \end{aligned} \tag{15}$$

Proof. By using Eqs. (4) and (10), we have

$$\begin{aligned}\nu(k+2, k) &= P\left(F(U'_{n,k}) - F(L'_{n,k}) > \beta\right) = P\left(U^{*'}_{n,k} - L^{*'}_{n,k} > \beta\right)\\ &= \int_\beta^1 \int_0^{1-r} f_{k+2,k}(x, x+r)\mathrm{d}x\mathrm{d}r\\ &= \frac{(2k)!}{[(k-1)!]^2}\int_\beta^1\int_0^{1-r}\left\{(1-x-r)^{k-1}\varphi_2(x, x+r, k) + x^{k-1}\varphi_2(1-x-r, 1-x, k)\right\}\mathrm{d}x\mathrm{d}r,\end{aligned}$$

where $\varphi_2(x, y, k)$ is as defined in (5). Since

$$\int_\beta^1\int_0^{1-r}(1-x-r)^{k-1}\varphi_2(x, x+r, k)\mathrm{d}x\mathrm{d}r = \int_\beta^1\int_0^{1-r}x^{k-1}\varphi_2(1-x-r, 1-x, k)\mathrm{d}x\mathrm{d}r,$$

we get

$$\begin{aligned}\nu(k+2, k) &= 2\frac{(2k)!}{[(k-1)!]^2}\int_\beta^1\int_0^{1-r}(1-x-r)^{k-1}\varphi_2(x, x+r, k)\mathrm{d}x\mathrm{d}r\\ &= 2\frac{(2k)!}{[(k-1)!]^2}(-1)^k\int_\beta^1\int_0^{1-r}(1-x-r)^{2k-2}\log(1-r)\mathrm{d}x\mathrm{d}r\\ &\quad + \sum_{i=1}^{k-1}\binom{k-1}{i}(-1)^{k-1}\int_\beta^1\int_0^{1-r}(1-x-r)^{2k-2}\log(1-r)\mathrm{d}x\mathrm{d}r.\end{aligned}$$

Now, after some algebraic calculations, the expression in (15) is obtained. □

Corollary 1. *Under the assumptions of Lemma 6,*

(i) *$(L'_{k+2,k}, +\infty)$ is a lower tolerance limit, and*
(ii) *$(-\infty, U'_{k+2,k})$ is an upper tolerance limit*

Table 2
Current k-records extracted from the data set in Arnold et al. (1998, p. 180)

n	$L'_{n,1}$	$U'_{n,1}$	$L'_{n,2}$	$U'_{n,2}$	$L'_{n,3}$	$U'_{n,3}$	n	$L'_{n,2}$	$U'_{n,2}$	$L'_{n,3}$	$U'_{n,3}$
1	12.69	12.69	12.84	12.69	18.72	12.69	14	4.89	23.92	7.51	21.96
2	12.69	12.84	12.84	12.84	18.72	12.84	15	4.89	27.16	7.51	23.21
3	12.69	18.72	12.84	18.72	12.84	12.84	16	4.83	27.16	6.25	23.21
4	12.69	21.96	12.69	18.72	12.69	12.84	17	4.13	27.16	6.25	23.29
5	7.51	21.96	12.55	18.72	12.55	12.84	18	4.13	30.57	6.25	23.92
6	4.83	21.96	11.80	18.72	12.55	14.28	19	4.13	31.28	4.89	23.92
7	4.83	23.92	7.51	18.72	11.80	14.28	20	–	–	4.89	24.95
8	4.83	27.16	7.51	19.19	8.69	14.28	21	–	–	4.83	24.95
9	4.13	27.16	7.51	21.46	8.69	14.77	22	–	–	4.83	26.81
10	4.13	31.28	7.51	21.96	8.69	18.72	23	–	–	4.83	27.16
11	4.08	31.28	4.89	21.96	8.69	19.19	24	–	–	4.83	30.57
12	4.08	34.04	4.89	23.21	8.69	21.46	25	–	–	4.56	30.57
13	–	–	4.89	23.29	7.51	21.46	26	–	–	–	–

Table 3
Confidence intervals for ξ_p based on the current k-record data in Table 2, with confidence at least 95%

p	k	$(L'_{n,k}, U'_{n,k})$	$\alpha_1(n, k; p)$	p	k	$(L'_{n,k}, U'_{n,k})$	$\alpha_1(n, k; p)$
0.2	1	(4.83, 23.92)	0.972	0.4	1	(7.51, 21.96)	0.975
	2	(7.51, 19.19)	0.951		2	(11.80, 18.72)	0.956
	3	–	–		3	(11.80, 14.28)	0.950
0.3	1	(4.83, 21.96)	0.978	0.5	1	(7.51, 21.96)	0.984
	2	(7.51, 18.72)	0.963		2	(11.80, 18.72)	0.976
	3	(8.69, 14.28)	0.952		3	(11.80, 14.28)	0.971

for population F, whose tolerance coefficients are free of F and are given by

$$\nu^*(k+2, k; p) = \vartheta_k(1-\beta; k+2),$$

where $\vartheta_k(1-\beta; k+2)$ is as defined in (14).

Proof. Note that $(L'_{k+2,k}, +\infty)$ is a lower tolerance limit whose coefficient can be calculated as follows. From (14), we get

$$P\{1 - F(L'_{k+2,k}) > \beta\} = P(L'_{k+2,k} < \xi_{1-\beta}) = \vartheta_k(\beta; k+2).$$

Using (13), $(-\infty, U'_{k+2,k})$ is an upper tolerance limit with coefficient

$$P\{F(U'_{k+2,k}) > \beta\} = P(L'_{k+2,k} < \xi_{1-\beta}) = \vartheta_k(\beta; k+2).$$

Thus, the required result readily follows. □

5. Illustrative example

To illustrate the nonparametric inferential methods developed in the preceding sections, we use the following data which represent the records of the amount of annual rainfall in inches at the Los Angeles Civic Center during the 100-year period from 1890 until 1989; see Arnold et al. (1998, p. 180). The current k-records extracted from these data are tabulated in Table 2.

From Tables 1 and 2, the confidence intervals with confidence coefficient at least 95% for ξ_p are obtained for $p = 0.2(0.1)0.5$, and these are presented in Table 3.

Acknowledgments

The authors express their sincere thanks to an anonymous referee and the associate editor for their useful comments and constructive criticisms on the original version of this manuscript, which led to this considerably improved version. Partial support from Ordered and Spatial Data Center of Excellence of Ferdowsi University of Mashhad is acknowledged.

References

Ahmadi, J., Arghami, N.R., 2003. Nonparametric confidence and tolerance intervals from record values data. Statist. Papers 44, 455–468.
Ahmadi, J., Balakrishnan, N., 2004. Confidence intervals for quantiles in terms of record range. Statist. Probab. Lett. 68, 395–405.
Ahmadi, J., Balakrishnan, N., 2005. Distribution-free confidence intervals for quantile intervals based on current records. Statist. Probab. Lett. 75, 190–202.
Arnold, B.C., Balakrishnan, N., Nagaraja, H.N., 1992. A First Course in Order Statistics. John Wiley & Sons, New York.
Arnold, B.C., Balakrishnan, N., Nagaraja, H.N., 1998. Records. John Wiley & Sons, New York.

Basak, P., 2000. An application of record range and some characterization results. In: Balakrishnan, N. (Ed.), Advances on Theoretical and Methodological Aspects of Probability and Statistics. Gordon and Breach Science Publishers, Philadelphia, pp. 83–95.
David, H.A., Nagaraja, H.N., 2003. Order Statistics, third ed. John Wiley & Sons, Hoboken, NJ.
Houchens, R.L., (1984). Record Value, Theory and Inference, Ph. D. Dissertation, University of California, Riverside, CA.
Raqab, M.Z., (2008). Distribution free prediction intervals for the future current record statistics, Statist. Papers (in press).