Original Research Paper

Prediction Intervals for Future Observations Based on Samples of Random Sizes

[1,2]**Mohammad Z. Raqab** and [3]**Haroon M. Barakat**

[1]*Department of Mathematics, The University of Jordan, Amman 11942, Jordan*
[2]*King Abdulaziz University, Jeddah, Saudi Arabia*
[3]*Department of Mathematics, Faculty of Science, Zagazig University, Zagazig, Egypt*

Abstract: On the basis of failure times of an informative sample (*X*-sample) of random size *M* of iid continuous random variables, we consider the prediction problem of failure times of another independent future sample (*Y*-sample) of random size *N* of iid variables from the same distribution. In this paper, we derive various exact prediction intervals for future failure times from the *Y*-sample based on the failure times from the *X*-sample. Specifically, prediction intervals for individual failure times as well as outer and inner prediction intervals are derived based on *X*-order statistics. Prediction intervals for the increments of order statistics are also investigated. Exact expressions for the coverage probabilities of these intervals are also derived and computed numerically. A practical example on a biometric data set is used to illustrate the results developed here.

Keywords: Order Statistics, Probability Coverage, Random Sample Size, Prediction Intervals, Outer and Inner Intervals

Introduction

Prediction of future order statistics is of natural interest in many practical problems. Considerably work has been done on the one-sample and two-sample prediction problems, and both parametric and nonparametric inferential methods have been developed in this regard. Interested readers may refer to Gulati and Padgett (2003) for details on related developments. An excellent review of development on prediction problems till late 90s can be found in Kaminsky and Nelson (1998). Using the Bayesian approach, Raqab and Madi (2002) considered the prediction of the total time on test using doubly censored Rayleigh data. Basak *et al*. (2006) considered the maximum likelihood predictors for different lifetime distributions, when the data are progressively censored.

The prediction intervals are used extensively in reliability theory, survival studies and industrial applications for predicting the number of defective items to be produced during future production process. Distribution-free confidence and prediction intervals have been discussed rather extensively in the context of order statistics. While Wilks (1962; Krewski, 1976) discussed the construction of outer and inner confidence intervals for quantile intervals based on order statistics,

concise reviews on this topic may be found in the books by David and Nagaraja (2003; Arnold *et al*., 2008; Ahmadi and Balakrishnan, 2011). Barakat *et al*. (2011) obtained prediction intervals for future exponential lifetimes based on random generalized order statistics. El-Adll (2011) discussed the problem of predicting future lifetimes based on three-parameter Weibull distribution. Barakat *et al*. (2014) obtained prediction intervals of future observations for a sample of random size from a continuous distribution. For similar prediction problems, one also may refer to Balakrishnan *et al*. (2005; Sultan and Ellah, 2006; Asgharzadeh and Valiollahi, 2010; Abdel-Hamid and AL-Hussaini, 2014; El-Adll and Aly, 2016). Prediction intervals based on two-sample prediction problems is discussed, for example, in Fligner and Wolfe (1976; 1979) for order statistics. Recently, Basiri *et al*. (2016) developed various non-parametric prediction intervals of order statistics from a future sample based on observed order statistics. Barakat *et al*. (2016) constructed prediction intervals for future two-parameter exponential lifetimes based on a random number of generalized order statistics under a general set-up.

Let $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$ be order statistics of a random sample with fixed sample size *n* from an absolutely continuous cumulative distribution function

(cdf) $F(x)$ and probability density function (pdf) $f(x)$: In reliability theory, $X_{s:n}$ represents the life length of a $(n-s+1)$-out-of-n system made up of n identical components with independent life lengths. When $s = n$; it is better known as the parallel system. For detailed discussion in this respect, see Barlow and Proschan (1981). In biological, agriculture and some quality control problems, we often come across situations where the sample size is a random variable (rv) and it is almost impossible to have a fixed sample size because either some observations get lost for various reasons or the size of the target population and its representative sample cannot be determined well. Let us consider the following practical example. Suppose that fish biologist would like to study the weight data of a specific rare fish living in the red sea over one-year period. The fish weight will vary depending on age, sex, season and recent feeding activity. Clearly, the number of fishes to be caught on daily or yearly basis is rv. The weight outcomes of this study is the ordered data, $X_{1:m} \leq X_{2:m} \leq \ldots \leq X_{m:m}$; where $X_{i:m}$ is the i-th ordered weight and m is the observed value of a rv M: Although the distribution of M is unknown but we can estimate the empirical distribution of M by considering its frequency table over one-year period. Based on the observed ordered weights over one-year period and under similar conditions, we aim at providing non-parametric prediction intervals for the fish weights to be caught in the next year. It is often reasonable to assume that the data sample and future sample are independent and the random number of fishes to be caught is independent of the fish weights. If one introduces the random sample size as an extension of a model (mainly for statistical inference), one can usually assume that it is independent of the underlying variables. Other practical applications could be the study of sizes or intensity of brightness for comets and shooting stars penetrating the atmosphere of earth in certain areas.

In this paper, we discuss Prediction Intervals (PIs) for future order statistics, $Y_{1:N}, Y_{2:N}, \ldots, Y_{N:N}$, from a future independent sample (Y-sample) of random size N from a continuous cdf $F(x)$; based on informative sample (X-sample) $X_{1:M}, X_{2:M}, \ldots, X_{M:M}$ of random size M from the same distribution. Throughout this paper, we assume that the rv's N and M are independent of $Y_{1:n}, Y_{2:n}, \ldots, Y_{n:n}$ and the items of the informative sample $X_{1:m}, X_{2:m}, \ldots, X_{m:m}$, for any n and m, respectively. This paper is organized as follows. In Section 2, we present some useful preliminaries. In Section 3, we derive Pls for a future order statistic. In Sections 4 and 5, outer and inner PIs for future order statistic intervals, respectively, are derived. Numerical computations for picking the appropriate order statistics for the establishment of these

PIs are performed and presented in their respective sections. In Section 6, we show how the increments of order statistics of the observed X-sample can be used to construct upper and lower prediction limits for increments of order statistics of the future Y-sample. Finally, in Section 7, a practical example on a biometric data is used to illustrate all the results developed here.

Preliminaries

Let $1 \leq r < s \leq m$. Then, the interval $(X_{r:m}, X_{s:m})$ is termed the *coverage* and $I_{r,s:m} = X_{s:m} - X_{r:m}$, the *increment*. The marginal pdf of the r-th order statistic, $U_{r:n}$ from uniform $U(0, 1)$ [see David and Nagaraja (2003)] is:

$$f_{r:n}(u) = \frac{n!}{(r-1)!(n-r)!} u^{r-1}(1-u)^{n-r}, 0 < u < 1 \tag{1}$$

The joint pdf of $U_{r:n}$ and $U_{s:n}$ is similarly given by:

$$f_{r,s:n}(u,v) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} u^{r-1}(v-u)^{s-r-1}(1-v)^{n-s}, \tag{2}$$
$$0 < u < v < 1.$$

The following well-known lemma [see for example, David and Nagaraja (2003), P.18] will be quite helpful in simplifying some arguments in the proofs of the results established in the subsequent sections.

Lemma 1

With $U_i = F(X_i)$ and $V_j = F(Y_j)$, we have:

$$U_{r:n} \overset{d}{=} F(X_{r:n}), \tag{3}$$

where, $U_{i:n}$ and $V_{j:n}$ are the i-th and j-th order statistics arising from samples of independent and identically distributed (iid) uniform rv's U_i's, V_j's and $\overset{d}{=}$ denotes identical in distribution.

The following identities are quite useful in showing various results in the next sections.

The pdf and cdf of the i-th order statistic from a sample of size n can be rewritten as follows:

$$f_{i:n}(u) = n B_{i-1,n-1}(u), 0 \leq u \leq 1,$$

$$F_{i:n}(u) = I_u(i, n-i+1) = \sum_{k=i}^{n} B_{k,n}(u), 0 \leq u \leq 1, \tag{4}$$

respectively, where:

$$I_u(i, n-i+1) = \frac{1}{B(i, n-i+1)} \int_0^u x^{i-1}(1-x)^{n-i} dx,$$

and

$$B_{j,n}(u)\binom{n}{j}u^j(1-u)^{n-j}, 0 < u < 1,$$

$$j = 0, 1,...,n, \quad n = 0, 1,...,$$

are incomplete beta function and Bernstein polynomials, respectively, with $B(i,n-i+1)$ being the complete beta function. Note that when a and b are positive integers $B(a, b) = (a-1)!(b-1)!/(a + b-1)!$.

PIs for a Future Order Statistic

Let $X_{1:M}, X_{2:M},...,X_{M:M}$ be order statistics from the X-sample with random size M of iid rv's from $F(x)$. For given $0 < \alpha < 1$, suppose we are interested in obtaining $100(1-\alpha)\%$ PI for the k-th order statistic $Y_{k:N}$ from the Y-sample with random size N of the form $(X_{r:M}, X_{s:M})$, $1 \le r < s \le M$, such that $P(X_{r:M} \le Y_{k:N} \le X_{s:M}) = 1-\alpha$. We refer to the interval $(X_{r:M}, X_{s:M})$ as a $100(1-\alpha)\%$ PI for $Y_{k:N}$. In this section, we derive such two-sided PIs for $Y_{k:N}$ with coverage probabilities being free of the parent distribution F.

Theorem 1

Let $\{X_i, i \ge 1\}$ and $\{Y_i, i \ge 1\}$ be two independent sequences of iid rv's from the same cdf $F(x)$ having random sample sizes M and N, respectively. Then, $(X_{r:M}, X_{s:M})$ is a PI, based on the X-sample, for the future order statistic $Y_{k:N}$ from the Y-sample, with the corresponding prediction coefficient, being free of F and is given by:

$$\pi_1(r,s;k) = \frac{1}{P(M \ge s)P(N \ge k)}$$

$$\sum_{m=s}^{\infty}\sum_{n=k}^{\infty}\sum_{i=r}^{s-1}\frac{\binom{k+i-1}{i}\binom{m+n-k-i}{m-i}}{\binom{m+n}{m}}P_M(m)P_N(n),$$

where, $P_M(m)$ and $P_N(n)$ stand for $P(M = m)$ and $P(N = n)$, respectively.

Proof:

We have:

$$P(X_{i:m} \le x) = P(X_{i:m} \le x, X_{j:m} < x) + P(X_{i:m} \le x, X_{j:m} \ge x) \quad (5)$$

$$= P(X_{j:m} < x) + P(X_{i:m} \le x \le X_{j:m}).$$

By the continuity of F, we immediately obtain:

$$P(X_{i:m} \le x \le X_{j:m}) = F_{i:m}(x) - F_{j:m}(x). \quad (6)$$

By the conditioning arguments, (3) and (5), we can write:

$$P(X_{r:m} \le Y_{k:n} \le X_{s:m}) = P(U_{r:m} \le V_{k:n} \le U_{s:m})$$

$$= \int_0^1 P(U_{r:m} \le v \le U_{s:m})f_{k:n}(v)dv$$

$$= \int_0^1 [F_{r:m}(v) - F_{s:m}(v)]f_{k:n}(v)dv \quad (7)$$

$$= \sum_{i=r}^{s-1}\frac{\binom{k+i-1}{i}\binom{m+n-k-i}{m-i}}{\binom{m+n}{m}}.$$

Now, let us adopt the coverage probability based on $(X_{r:M}, X_{s:M})$. On using the facts that $M \ge s$ and $N \ge k$ and applying the conditioning arguments on the rv's M and N (Raghunandanan and Patil, 1972; Buhrman, 1973), we obtain:

$$P(X_{r:M} \le Y_{k:N} \le X_{s:M}) = P(U_{r:M} \le V_{k:N} \le U_{s:M})$$

$$= \frac{1}{P(M \ge s)P(N \ge k)}\sum_{m=s}^{\infty}\sum_{n=k}^{\infty}P(U_{r:m} \le V_{k:n} \le U_{s:m})P_M(m)P_N(n). \quad (8)$$

Therefore the result in Theorem 1 follows readily by (7) and (8).

Remark 1:

From Theorem 1, $(X_{r:M}, X_{s:M})$ can be picked as a $100(1-\alpha)$ PI for $Y_{k:N}$ with a specified coefficient level $1-\alpha$ ($0 < \alpha < 1$) such that $\pi_1(r, s; k) \ge 1-\alpha$. The prediction coefficient $\pi_1(r, s; k)$, which is free of the parent distribution F, can thus be computed simply based on the probability values of the rv's M and N. Letting $r = 0$ and $X_{r:M} = -\infty$, we have one-sided PI of the form $(-\infty, X_{s:M})$ immediately from above formula. Similarly, if s is an integer such that $P(r \le M < s) = 1$ and $X_{s:M} = \infty$, we will get $(X_{r:M}, \infty)$ as one-sided PI for $Y_{k:N}$. In this case, the prediction coefficient is reduced to:

$$\pi_1(r,s;k) = \frac{1}{P(M \ge r)P(N \ge k)}$$

$$\sum_{m=r}^{\infty}\sum_{n=k}^{\infty}\sum_{i=r}^{m}\frac{\binom{k+i-1}{i}\binom{m+n-k-i}{m-i}}{\binom{m+n}{i}}P_M(m)P_N(n).$$

Remark 2

Under the settings and conditions of Theorem 1 with M and N being integers such that $P(M = m) = P(N = n) = 1$, $m \ge s$ and $n \ge k$, then $\pi_1(r, s; k)$ reduces to:

$$\pi_1(r,s;k) = \sum_{i=r}^{s-1} \frac{\binom{k+i-1}{i}\binom{m+n-k-i}{m-i}}{\binom{m+n}{m}}.$$

Corollary 1:

Under the settings and conditions of Theorem 1:

a) *If M, N have binomially distributions with parameters v and p [M, N ~ B(v,p)], then the probabilities coefficients in Theorem 1 can be represented as:*

$$P_M^s(m) = \frac{P_M(m)}{P(M \ge s)} = \frac{\binom{v}{m}p^m(1-p)^{v-m}}{\sum_{m=s}^{v}\binom{v}{m}p^m(1-p)^{v-m}},$$

and

$$P_N^k(n) = \frac{P_N(n)}{P(N \ge k)} = \frac{\binom{v}{n}p^n(1-p)^{v-n}}{\sum_{n=k}^{v}\binom{v}{n}p^n(1-p)^{v-n}}.$$

In this case, the prediction confidence coefficient reduces to:

$$\pi_1(r,s;k) = \sum_{m=s}^{v}\sum_{n=k}^{v}\sum_{i=r}^{s-1} \frac{\binom{k+i-1}{i}\binom{m+n-k-i}{m-i}}{\binom{m+n}{m}} P_M^s(m)P_N^k(n).$$

b) *If M,N have negative binomial distributions with parameters v and p [M,N~NB(v, p)], then the probabilities coefficients in Theorem 1 becomes:*

$$P_M^s(m) = \begin{cases} \binom{m-1}{v-1}p^v(1-p)^{m-v}, & \text{if } s \le v, \\[2ex] \dfrac{\binom{m-1}{v-1}p^v(1-p)^{m-v}}{1-\sum_{m=v}^{s-1}\binom{m-1}{v-1}p^v(1-p)^{m-v,}} & \text{if } s > v, \end{cases}$$

and:

$$P_N^k(n) = \begin{cases} \binom{n-1}{v-1}p^v(1-p)^{n-v}, & \text{if } k \le v, \\[2ex] \dfrac{\binom{n-1}{v-1}p^v(1-p)^{n-v}}{1-\sum_{n=v}^{k-1}\binom{n-1}{v-1}p^v(1-p)^{n-v,}} & \text{if } k > v. \end{cases}$$

Therefore, the prediction confidence coefficient turns out:

$$\pi_1(r,s;k) = \sum_{m=\max(s,v)}^{\infty}\sum_{n=\max(k,v)}^{\infty} \sum_{i=r}^{s-1} \frac{\binom{k+i-1}{i}\binom{m+n-k-i}{m-i}}{\binom{m+n}{m}} P_M^s(m)P_N^k(n).$$

c) *If M, N have Poisson distributions with parameters λ [M, N~ Poisson(λ)], then the probabilities coefficients in Theorem 1 becomes:*

$$P_M^s(m) = \frac{e^{-\lambda}\lambda^m/m!}{1-\sum_{m=0}^{s-1}e^{-\lambda}\lambda^m/m!},$$

and

$$P_N^k(n) = \frac{e^{-\lambda}\lambda^n/n!}{1-\sum_{n=0}^{k-1}e^{-\lambda}\lambda^k/k!}.$$

The prediction confidence coefficient is readily obtained as:

$$\pi_1(r,s;k) = \sum_{m=s}^{\infty}\sum_{n=k}^{\infty}\sum_{i=r}^{s-1} \frac{\binom{k+i-1}{i}\binom{m+n-k-i}{m-i}}{\binom{m+n}{m}} P_M^s(m)P_N^k(n).$$

In the following lemma, we provide $(1 - \alpha)$ 100% PI that contains at least *t* observations of the *Y*-sample $(N \ge t)$.

Lemma 2:

Let us define $J_{t,N}$ as the number of future observations $\ge t$ out of N such that $N \ge t$. Under the condition and setting of Theorem 1, the coverage probability of $J_{t,N}$ based on the two-sided interval of the form $(X_{r:M}, X_{s:M})$ is:

$$P(X_{r:m} < J_{t,N} < X_{s:M})$$

$$= \frac{1}{P(M \ge s)P(N \ge t)}\sum_{m=s}^{\infty}\sum_{n=t}^{\infty}\sum_{i=t}^{n} \frac{\binom{s-r+i-1}{i}\binom{m+n-s+r-i}{n-i}}{\binom{m+n}{m}} P_M(m)P_N(n).$$

Proof:

Given $M = m$, $N = n > t$, we have:

$$P(X_{r:m} < J_{t,n} < X_{s:m}) = P\left(F(X_{s:m}) - F(X_{r:M}) \ge \frac{t}{n}\right)$$

$$= P(U_{s-r:m} \ge V_{t:n}).$$

Using the conditioning argument, the result follows immediately.

19

Corollary 2:

If the experiment conditions have not changed, we may assume $P(M = x) = P(N = x)$ and then get PI, $(X_{r:\hat{m}}, X_{s:\hat{m}}) \approx (X_{r:M}, X_{s:M})$ for $Y_{k:N}$, where $\hat{m}$ is the observed value of the rv M. Under the condition that both samples have equal random sizes (say N), the coverage probability in Theorem 1 is simplified to:

$$\pi_1(r,s;k) = \frac{1}{P(N \geq \max(k,s))}$$

$$\sum_{n=\max(k,s)}^{\infty} \sum_{i=r}^{s-1} \frac{\binom{k+i-1}{i}\binom{2n-k-i}{n-i}}{\binom{2n}{n}} P_N(n).$$

Remark 3:

It is worth mentioning that one can't apply Theorem 1 for any future order statistics except for the lower extremes. Otherwise, the value of k in $\pi_1(r, s, k)$ depends on N, which does not make any sense. However, if the goal is to estimate a specific parent quantile ξ_p of order $p(0 < p < 1)$, one may refer to Al-Mutairi and Raqab (2017). For predicting the sample median when the sample size N is random, we first obtain an integer k such that $P(N \leq k) \geq 1/2$ and $P(N \geq k) \geq 1/2$, then we immediately predict the corresponding value for $Y_{k:N}$. For example, if $N \sim B(20, 0.3)$; then $k = 6$ and $Y_{6:N}$ has to be predicted. Let us consider the case with $P(N = n_0) = 1$, where n_0 is either the observed value of N or just a hypothetical value or even the predicted value of N. Then, we may consider $(X_{r:\hat{m}}, X_{s:\hat{m}}) \approx (X_{r:M}, X_{s:M})$, as the PI for the future median $Y_{k:N}$, where $\hat{m}$ is the observed value of M and $k = \left[\frac{n_0}{2}\right]$, with $[x]$ being the greatest integer less than or equal to x.

The values of $\pi_1(r,s;k)$ are presented in Table 1 for various choices of r, s and k. If the desired level of the PI is specified to be $1-\alpha$, we may choose r and s so that $\pi_1(r, s; k) \geq 1 - \alpha$. Evidently, the distribution of the sample size plays an important role in evaluating the prediction coefficients. It can be easily checked that for fixed r, $\pi_1(r, s; k)$ increases rapidly when s increases for small values of s. For large values of s, $\pi_1(r, s; k)$ changes very slowly as s increases. That is, the computed values of these coefficients are not sensitive to the large values of $s-r$.

Outer PIs for an Order Interval

Suppose $(Y_{k:N}, Y_{l:N})$ is a future interval of order statistics from the Y-sample and that we are interested in obtaining $100(1-\alpha)\%$ PI for it of the form $(X_{r:M},$ $X_{s:M})$ such that $P(X_{r:M} < Y_{k:M} < Y_{l:N} < X_{s:M}) \geq 1-\alpha$. Then, $(X_{r:M}, X_{s:M})$ is termed the *outer PI* for the interval $(Y_{k:N}, Y_{l:N})$. In this section, we describe how such these PIs can be constructed.

Theorem 2:

Under the settings and assumptions of Theorem 1, $(X_{r:M}, X_{s:M})$ is an outer PI, based on the X-sample, for the future interval of order statistics $(Y_{k:N}, Y_{l:N})$ from the Y-sample, with the corresponding prediction coefficient, being free of F, given by:

$$\pi_2(r,s,k,l) = \frac{1}{P(M \geq s)P(N \geq l)}$$

$$\times \sum_{m=s}^{\infty}\sum_{n=l}^{\infty}\sum_{i=r}^{s-1}\sum_{j=0}^{s-i-1} \frac{\binom{k+i-1}{i}\binom{j+l-k-1}{j}\binom{m+n-l-i-j}{m-i-j}}{\binom{m+n}{m}} P_M(m)P_N(n),$$

Proof:

From Lemma 1, $F(X_{i:M}) \stackrel{d}{=} U_{i:M}$ and $F(Y_{j:N}) \stackrel{d}{=} V_{j:N}$. For $1 \leq r < s \leq M$ and $1 \leq k < l \leq N$, let us consider:

$$P(X_{r:M} < Y_{k:N} < Y_{l:N} < X_{s:M}) = P(U_{r:M} < V_{k:N} < V_{l:N} < U_{s:M})$$

$$= \frac{1}{P(M \geq s)P(N \geq l)}\sum_{m=s}^{\infty}\sum_{m=l}^{\infty} P(U_{r:m} < V_{k:n} < V_{l:n} < U_{s:m})P_M(m)P_N(n). \quad (9)$$

$$= \frac{1}{P(M \geq s)P(N \geq l)}\sum_{m=s}^{\infty}\sum_{n=l}^{\infty}\int_0^1\int_{v_k}^1 P(U_{r:m} < v_{k:n} < v_{l:n} < U_{s:m})$$

$$\times f_{k,l:n}(v_k,v_l)\,dv_l\,dv_k\,P_M(m)P_N(n).$$

From Lemma 1 and the fact that $(U_{r:n}, U_{s:m})$ with m and n being fixed, are jointly distributed as the r-th and s-th order statistics from the standard uniform distribution, we first have:

$$P(U_{r:m} < v_k < v_l < U_{s:m}) =$$

$$\frac{m!}{(r-1)!(s-r-1)!(m-s)!}\int_{v_l}^1\int_0^{v_k} w_1^{r-1}(w_2-w_1)^{s-r-1}(1-w_2)^{m-s}\,dw_1\,dw_2$$

$$= \int_{v_l}^1\left\{\int_0^{v_k}\frac{m!}{(r-1)!(s-r-1)!(m-s)!}w_1^{r-1}\left(1-\frac{w_1}{w_2}\right)^{s-r-1}dw_1\right\} \quad (10)$$

$$\times w_2^{s-r-1}(1-w_2)^{m-s}\,dw_2.$$

Since the inside integral is an incomplete beta function, it can be expressed as a sum of binomial probabilities using (4). Then the resulting expression is:

$$P(U_{r:m} < v_k < v_l < U_{s:m}) = s\binom{m}{s}$$

$$\sum_{i=r}^{s-1}\binom{s-1}{i}\int_{v_l}^1 v_k^i(w_2-v_k)^{s-i-1}(1-w_2)^{m-s}\,dw_2 \quad (11)$$

Table 1: Values of $\pi_1(r, s; k)$ for some choices of r, s and k

k	r	p	B(20, p)				NB(6, p)			
			$s = 6$	$s = 10$	$s = 14$	$s = 18$	$s = 6$	$s = 10$	$s = 14$	$s = 18$
6	1	0.30	0.497	0.971	0.992	0.990	0.492	0.915	0.985	0.989
		0.50	0.495	0.957	0.991	0.990	0.495	0.949	0.990	0.990
		0.70	0.492	0.930	0.990	0.990	0.497	0.966	0.991	0.990
		0.90	0.490	0.903	0.987	0.990	0.498	0.975	0.992	0.990
	2	0.30	0.484	0.946	0.961	0.956	0.461	0.882	0.951	0.952
		0.50	0.472	0.931	0.960	0.956	0.473	0.921	0.957	0.955
		0.70	0.462	0.900	0.958	0.956	0.483	0.941	0.960	0.956
		0.90	0.456	0.869	0.953	0.955	0.490	0.952	0.961	0.956
	3	0.30	0.447	0.889	0.895	0.886	0.395	0.814	0.880	0.880
		0.50	0.418	0.871	0.894	0.886	0.421	0.859	0.889	0.883
		0.70	0.397	0.834	0.891	0.885	0.445	0.882	0.893	0.885
		0.90	0.386	0.799	0.883	0.884	0.463	0.895	0.895	0.886
	4	0.30	0.367	0.788	0.789	0.778	0.290	0.708	0.772	0.772
		0.50	0.321	0.768	0.788	0.778	0.327	0.756	0.783	0.776
		0.70	0.292	0.729	0.785	0.778	0.364	0.781	0.787	0.777
		0.90	0.279	0.692	0.775	0.777	0.396	0.794	0.789	0.778
7	1	0.30	0.271	0.941	0.996	0.996	0.347	0.858	0.985	0.995
		0.50	0.316	0.918	0.995	0.996	0.311	0.905	0.992	0.996
		0.70	0.347	0.875	0.993	0.996	0.275	0.933	0.995	0.996
		0.90	0.358	0.840	0.987	0.996	0.244	0.949	0.996	0.996
	2	0.30	0.267	0.932	0.982	0.980	0.332	0.842	0.967	0.976
		0.50	0.307	0.908	0.981	0.980	0.303	0.894	0.977	0.978
		0.70	0.334	0.862	0.979	0.979	0.270	0.923	0.981	0.979
		0.90	0.342	0.823	0.971	0.979	0.241	0.941	0.982	0.979
	3	0.30	0.252	0.905	0.947	0.939	0.297	0.804	0.927	0.933
		0.50	0.283	0.878	0.945	0.939	0.279	0.863	0.940	0.937
		0.70	0.299	0.827	0.942	0.939	0.255	0.896	0.944	0.938
		0.90	0.302	0.783	0.931	0.938	0.231	0.915	0.947	0.939
	4	0.30	0.215	0.849	0.880	0.868	0.230	0.736	0.855	0.859
		0.50	0.230	0.818	0.878	0.868	0.228	0.801	0.871	0.864
		0.70	0.233	0.761	0.873	0.867	0.217	0.838	0.877	0.866
		0.90	0.230	0.712	0.859	0.866	0.203	0.860	0.880	0.868
8	1	0.30	0.131	0.973	0.995	0.998	0.224	0.768	0.974	0.997
		0.50	0.174	0.840	0.994	0.998	0.176	0.824	0.988	0.998
		0.70	0.220	0.786	0.990	0.998	0.137	0.861	0.993	0.998
		0.90	0.242	0.747	0.978	0.998	0.110	0.885	0.995	0.998
	2	0.30	0.130	0.870	0.990	0.991	0.218	0.762	0.966	0.988
		0.50	0.171	0.836	0.988	0.991	0.172	0.820	0.982	0.990
		0.70	0.215	0.780	0.984	0.991	0.136	0.858	0.987	0.991
		0.90	0.235	0.740	0.971	0.991	0.109	0.883	0.990	0.991
	3	0.30	0.124	0.860	0.973	0.970	0.200	0.743	0.945	0.964
		0.50	0.161	0.824	0.971	0.970	0.162	0.806	0.963	0.968
		0.70	0.198	0.764	0.966	0.970	0.130	0.847	0.970	0.969
		0.90	0.214	0.719	0.950	0.969	0.106	0.873	0.73	0.970
	4	0.30	0.109	0.834	0.936	0.927	0.162	0.703	0.901	0.917
		0.50	0.137	0.795	0.933	0.927	0.137	0.775	0.923	0.923
		0.70	0.162	0.727	0.926	0.926	0.113	0.820	0.932	0.925
		0.90	0.171	0.676	0.907	0.925	0.095	0.849	0.936	0.926

By a change of variable $z = (1 - w_2)/(1 - v_k)$, the integral in (11) can be transformed to an incomplete beta function, which upon using the identity in (4), yields:

$$P(U_{r:m} < v_k < v_l < U_{s:m}) = \sum_{i=r}^{s-1} \sum_{j=0}^{s-i-1} \binom{m}{i} \binom{m-i}{j} v_k^i (1 - v_k)^{m-i}$$

$$\times \left(\frac{v_l - v_k}{1 - v_k} \right)^j \left(\frac{1 - v_l}{1 - v_k} \right)^{m-i-j}. \qquad (12)$$

Upon substituting the expression in (12) into Equation (9) and using the change of variable arguments, the result readily follows. ∇

Corollary 3:

Under the settings and the condition that both samples are taken with equal sizes (say, N), then the corresponding prediction coefficient is:

$$\pi_2(r,s;k,l) = \frac{1}{P(N \geq \max(l,s))}$$

$$\sum_{n=\max(k,s)}^{\infty} \sum_{i=r}^{s-1} \sum_{j=0}^{s-i-1} \frac{\binom{k+i-1}{i}\binom{j+l-k-1}{j}\binom{2n-l-i-j}{n-i-j}}{\binom{2n}{n}} P_N(n).$$

The values of $\pi_2(r, s, k, l)$ are presented in Table 2 for various choices of r, s, k and l. If the desired level of the outer PI is specified to be 1-α_0, we may choose appropriate values of r and s so that $\pi_2(r, s, k, l) \geq 1-\alpha_0$. It is naturally expected that the outer PIs of the order intervals are observed to be wider to achieve the specified level 0.90. Therefore, given a specified level 1-α_0, it is more likely to pick up outer PI for $(Y_{k:N}, Y_{l:N})$ based $(X_{r:M}, Y_{s:M})$ when $s{-}r$ gets large. Clearly, the randomness of the sample size is affecting the behavior of $\pi_2(r, s, k, l)$ when comparing with the fixed sample size. It can be seen from Table 2 that $\pi_2(r, s, k, l)$ increases when s moves away from r but it is not sensitive to the large values of s.

Inner PIs for an Order Interval

Suppose we are interested in obtaining a 100(1-α)% PI for the future interval of order statistics $(Y_{k:N}, Y_{l:N})$ from the Y-sample of the form $(X_{r:M}, X_{s:M})$ such that $P(Y_{k:N} < X_{r:M} < X_{s:M} < Y_{l:N}) \geq 1-\alpha$. Then, $(X_{r:M}, X_{s:M})$ is termed the *inner PI* for the interval $(Y_{k:N}, Y_{l:N})$: In this section, we describe how such inner PIs can be constructed.

Theorem 3:

Under the setting and assumption of Theorem 1, $(X_{r:M}, X_{s:M})$ is an inner PI, based on the X-sample, for the future record interval $(Y_{k:N}, Y_{l:N})$ from the Y-sample, with the corresponding prediction coefficient, being free of F, given by:

$$\pi_3(r,s;k,l) = \frac{1}{P(M \geq s)P(N \geq l)}$$

$$\sum_{m=s}^{\infty}\sum_{n=l}^{\infty}\sum_{i=0}^{r-1} \frac{\binom{k+i-1}{i}\binom{m+n-i-k}{m-i}}{\binom{m+n}{m}} P_M(m) P_N(n).$$

$$- \frac{1}{P(M \geq s)P(N \geq l)}$$

$$\sum_{m=s}^{\infty}\sum_{n=l}^{\infty}\sum_{i=0}^{r-1}\sum_{j=0}^{s-i-1} \frac{\binom{k+i-1}{i}\binom{l+j-k-1}{j}\binom{m+n-l-i-j}{m-i-j}}{\binom{m+n}{m}}$$

$$\times P_M(m) P_N(n).$$

Proof:

Using arguments similar to those in the proof of Theorem 2, we have:

$$P(y_k < X_{r:m} < X_{s:m} < y_l) = P(F(y_k) < U_{r:m} < U_{s:m} < F(y_l))$$

$$= \frac{m!}{(r-1)!(s-r-1)!(m-s)!}$$

$$\int_{F(y_k)}^{F(y_l)}\int_{F(y_k)}^{w_2} w_1^{r-1}(w_2-w_1)^{s-r-1}(1-w_2)^{m-s} dw_1\, dw_2$$

$$= \sum_{i=0}^{r-1}\binom{m}{i}[F(y_k)]^i[1-F(y_k)]^{m-i}$$

$$- \sum_{i=0}^{r-1}\sum_{j=0}^{s-r-1}\binom{m}{i}\binom{m-i}{j}[F(y_k)]^i$$

$$[F(y_l)-F(y_k)]^j[1-F(y_l)]^{m-i-j}. \qquad (13)$$

Now, upon using the conditioning argument, we can write:

$$P(V_{k:n} < U_{r:m} < U_{s:m} < V_{l:n})$$

$$= \int_0^1\int_{v_k}^1 P(V_{k:n} : U_{r:m} < U_{s:m} < V_{l:n}|V_{k:n}=v_k, V_{l:n}=v_l) \qquad (14)$$

$$\times f_{k,l:n}(v_k, v_l)dv_l\, dv_k.$$

By using change of variables, the integral in (14) can be transformed into an incomplete beta integral which can be written as a finite sum of binomial probabilities using (4). The required result in Theorem 3 then follows by conditioning arguments on M and N: ▽

Remark 4:

When N and M are degenerate rv's at m and n such that $P(M = m) = P(N = n) = 1$ with $m \geq s$ and $n \geq l$; then $\pi_3(r, s, k, l)$ is turned to be:

$$\pi_3(r,s;k,l) = \sum_{i=0}^{r-1} \frac{\binom{k+i-1}{i}\binom{m+n-i-k}{m-i}}{\binom{m+n}{m}}$$

$$- \sum_{i=0}^{r-1}\sum_{j=0}^{s-i-1} \frac{\binom{k+i-1}{i}\binom{l+j-k-1}{j}\binom{m+n-l-i-j}{m-i-j}}{\binom{m+n}{m}}.$$

Table 3 presents the values of $\pi_3(r, s; k, l)$ for various choices of r, s; k and l. If the desired level of the inner prediction interval is specified to be 1-α_0, we may choose appropriate values of r and s so that $\pi_3(r, s; k, l) \geq 1-\alpha_0$. It is worth mentioning here that it is expected to have shortest inner PIs to attain the specified level 0.90.

Table 3 shows that, for fixed r, $\pi_3(r, s; k, l)$ reaches its maximum quickly for small values of $s-r$. The values of $\pi_3(r, s; k, l)$ are not sensitive to the small values of $s-r$ and change significantly when $s-r$ gets large.

Table 2: Values of $\pi_2(r, s; k, l)$ for some choices of r, s, k and l

(k, l)	r	p	B(20, p)				NB(20, p)			
			s = 6	s = 10	s = 14	s = 18	s = 6	s = 10	s = 14	s = 18
(5, 6)	1	0.30	0.490	0.960	0.979	0.977	0.472	0.835	0.949	0.969
		0.50	0.485	0.946	0.979	0.977	0.373	0.848	0.956	0.971
		0.70	0.480	0.918	0.978	0.977	0.475	0.861	0.963	0.974
		0.90	0.478	0.890	0.974	0.977	0.476	0.874	0.969	0.976
	2	0.30	0.453	0.904	0.917	0.911	0.399	0.761	0.874	0.894
		0.50	0.436	0.888	0.916	0.911	0.403	0.775	0.883	0.899
		0.70	0.422	0.856	0.914	0.911	0.407	0.790	0.892	0.903
		0.90	0.415	0.824	0.908	0.910	0.411	0.806	0.900	0.907
	3	0.30	0.374	0.799	0.807	0.799	0.291	0.645	0.758	0.778
		0.50	0.343	0.782	0.806	0.799	0.296	0.660	0.768	0.783
		0.70	0.322	0.747	0.803	0.799	0.300	0.676	0.778	0.789
		0.90	0.311	0.713	0.796	0.798	0.306	0.693	0.787	0.794
	4	0.30	0.254	0.650	0.660	0.655	0.170	0.507	0.619	0.639
		0.50	0.218	0.634	0.660	0.655	0.174	0.521	0.628	0.643
		0.70	0.196	0.601	0.658	0.654	0.178	0.535	0.636	0.647
		0.90	0.187	0.569	0.651	0.654	0.182	0.551	0.644	0.651
(5, 7)	1	0.30	0.177	0.926	0.978	0.977	0.355	0.760	0.928	0.965
		0.50	0.293	0.902	0.977	0.977	0.352	0.773	0.939	0.969
		0.70	0.332	0.858	0.975	0.977	0.349	0.787	0.949	0.973
		0.90	0.342	0.821	0.969	0.977	0.346	0.802	0.959	0.975
	2	0.30	0.157	0.871	0.916	0.911	0.289	0.685	0.853	0.890
		0.50	0.254	0.845	0.915	0.911	0.289	0.701	0.866	0.897
		0.70	0.282	0.797	0.912	0.911	0.288	0.717	0.879	0.902
		0.90	0.286	0.756	0.903	0.910	0.288	0.734	0.891	0.907
	3	0.30	0.119	0.766	0.806	0.799	0.196	0.571	0.737	0.774
		0.50	0.187	0.740	0.805	0.799	0.197	0.587	0.751	0.781
		0.70	0.201	0.689	0.801	0.799	0.198	0.605	0.764	0.788
		0.90	0.201	0.646	0.791	0.798	0.200	0.623	0.777	0.793
	4	0.30	0.071	0.619	0.660	0.655	0.101	0.437	0.598	0.635
		0.50	0.106	0.593	0.658	0.655	0.103	0.452	0.611	0.641
		0.70	0.110	0.546	0.655	0.654	0.104	0.467	0.623	0.646
		0.90	0.108	0.505	0.646	0.654	0.106	0.484	0.634	0.651
(5, 8)	1	0.30	0.125	0.857	0.976	0.977	0.256	0.670	0.896	0.959
		0.50	0.165	0.823	0.974	0.977	0.250	0.681	0.911	0.965
		0.70	0.207	0.767	0.970	0.977	0.243	0.694	0.926	0.971
		0.90	0.227	0.726	0.957	0.977	0.236	0.708	0.941	0.975
	2	0.30	0.106	0.803	0.914	0.911	0.200	0.598	0.821	0.884
		0.50	0.137	0.767	0.912	0.911	0.196	0.611	0.838	0.893
		0.70	0.169	0.708	0.907	0.911	0.193	0.626	0.855	0.900
		0.90	0.182	0.663	0.891	0.910	0.188	0.642	0.873	0.906
	3	0.30	0.074	0.702	0.804	0.799	0.126	0.488	0.705	0.768
		0.50	0.094	0.666	0.802	0.799	0.125	0.502	0.723	0.777
		0.70	0.112	0.604	0.796	0.799	0.123	0.518	0.741	0.786
		0.90	0.119	0.557	0.779	0.798	0.122	0.535	0.760	0.793
	4	0.30	0.039	0.560	0.657	0.655	0.058	0.362	0.568	0.629
		0.50	0.048	0.526	0.655	0.655	0.058	0.375	0.584	0.637
		0.70	0.055	0.469	0.650	0.654	0.058	0.389	0.601	0.644
		0.90	0.057	0.425	0.635	0.654	0.058	0.405	0.617	0.650

Table 3: Values of $\pi_3(r, s; k, l)$ for some choices of $r, s; k$ and l

(k, l)	r	p	B(20, p)				NB(5, p)			
			s = r+1	s = r+4	s = r+7	r+10	s = r+1	s = r+4	s = r+7	r+10
(2, 18)	5	0.30	0.911	0.911	0.907	0.860	0.907	0.904	0.879	0.769
		0.50	0.911	0.911	0.907	0.858	0.910	0.909	0.897	0.817
		0.70	0.911	0.911	0.916	0.853	0.911	0.910	0.903	0.842
		0.90	0.910	0.910	0.902	0.835	0.911	0.911	0.907	0.856
	6	0.30	0.956	0.955	0.946	0.842	0.952	0.945	0.901	0.731
		0.50	0.956	0.955	0.945	0.839	0.955	0.952	0.928	0.785
		0.70	0.956	0.955	0.944	0.832	0.956	0.954	0.939	0.817
		0.90	0.955	0.954	0.937	0.806	0.956	0.955	0.945	0.837
	7	0.30	0.980	0.978	0.956	0.737	0.975	0.962	0.889	0.638
		0.50	0.979	0.978	0.955	0.734	0.978	0.972	0.927	0.683
		0.70	0.979	0.977	0.952	0.726	0.979	0.976	0.944	0.711
		0.90	0.979	0.976	0.940	0.700	0.979	0.978	0.954	0.732
	8	0.30	0.991	0.991	0.987	0.494	0.986	0.961	0.845	0.492
		0.50	0.991	0.987	0.935	0.494	0.989	0.977	0.894	0.493
		0.70	0.991	0.986	0.930	0.996	0.990	0.984	0.919	0.494
		0.90	0.990	0.983	0.912	0.494	0.991	0.986	0.934	0.494
(3, 18)	5	0.30	0.795	0.751	0.799	0.799	0.794	0.790	0.766	0.664
		0.50	0.795	0.749	0.795	0.749	0.797	0.796	0.785	0.710
		0.70	0.799	0.799	0.794	0.745	0.798	0.798	0.791	0.734
		0.90	0.798	0.798	0.790	0.727	0.799	0.799	0.794	0.748
	6	0.30	0.886	0.885	0.876	0.776	0.879	0.872	0.829	0.666
		0.50	0.886	0.885	0.875	0.773	0.883	0.881	0.857	0.721
		0.70	0.885	0.885	0.873	0.766	0.885	0.884	0.869	0.751
		0.90	0.884	0.883	0.867	0.741	0.885	0.885	0.875	0.771
	7	0.30	0.939	0.938	0.916	0.703	0.932	0.919	0.847	0.602
		0.50	0.939	0.938	0.915	0.700	0.937	0.931	0.886	0.649
		0.70	0.939	0.937	0.911	0.692	0.938	0.935	0.904	0.677
		0.90	0.938	0.935	0.900	0.666	0.939	0.937	0.914	0.698
	8	0.30	0.970	0.966	0.917	0.481	0.962	0.937	0.822	0.475
		0.50	0.970	0.966	0.914	0.481	0.967	0.955	0.872	0.478
		0.70	0.969	0.965	0.909	0.481	0.969	0.962	0.898	0.480
		0.90	0.969	0.961	0.891	0.480	0.970	0.965	0.913	0.481
(4, 18)	5	0.30	0.655	0.654	0.651	0.612	0.651	0.648	0.625	0.533
		0.50	0.655	0.654	0.651	0.610	0.653	0.652	0.641	0.575
		0.70	0.654	0.654	0.650	0.606	0.654	0.654	0.648	0.596
		0.90	0.654	0.653	0.647	0.591	0.655	0.654	0.650	0.609
	6	0.30	0.778	0.778	0.769	0.677	0.771	0.764	0.723	0.572
		0.50	0.778	0.778	0.768	0.674	0.776	0.773	0.751	0.625
		0.70	0.778	0.777	0.766	0.668	0.777	0.776	0.762	0.654
		0.90	0.777	0.776	0.760	0.644	0.778	0.778	0.768	0.672
	7	0.30	0.868	0.866	0.845	0.644	0.858	0.845	0.774	0.543
		0.50	0.868	0.866	0.844	0.640	0.864	0.858	0.814	0.590
		0.70	0.867	0.865	0.840	0.633	0.866	0.863	0.833	0.618
		0.90	0.866	0.863	0.829	0.607	0.867	0.866	0.843	0.638
	8	0.30	0.927	0.923	0.874	0.455	0.915	0.890	0.777	0.442
		0.50	0.926	0.922	0.872	0.455	0.922	0.911	0.829	0.449
		0.70	0.926	0.921	0.867	0.454	0.925	0.918	0.855	0.453
		0.90	0.925	0.917	0.848	0.452	0.926	0.922	0.870	0.455

PIs for Order Statistics Increments

The increments of order statistics $I_{r,s:M} = X_{s:M} - X_{r:M}, M \geq s$ based on X-sample can be used in similar settings to construct PIs for the increments of order statistics $I^*_{k,l:N} = Y_{l:N} - Y_{k:N}, N \geq l$ for Y-sample. For $s = r+1$ and $l = k + 1$, $I_{r,s:M}$ and $I^*_{k,l:N}$ are simply the spacing of two consecutive order statistics. The order statistics increments are usually used to measure the variation in the case of location-scale distributions; see, for example, David and Nagaraja (2003, p.160). In this section, we describe how the record spacings $I_{r,s:M}$ can be used to construct upper and lower prediction limits for the increments of order statistics $I^*_{k,l:N}$. Let us first start with the following:

$$\pi_4(r,s;k,l) = P(X_{s:M} - X_{r:M} \geq Y_{l:N} - Y_{k:N})$$

$$= \frac{1}{P(M \geq s)P(N \geq l)} \qquad (15)$$

$$\sum_{m=s}^{\infty}\sum_{n=l}^{\infty} P(X_{s:m} - X_{r:m} \geq Y_{l:n} - Y_{k:n})P_M(m)P_N(n).$$

We have also:

$$P(X_{s:m} - X_{r:m} \geq Y_{l:n} - Y_{k:n})$$
$$= P(X_{s:m} - X_{r:m} \geq Y_{l:n} - Y_{k:n}, X_{r:m} \leq Y_{k:n})$$
$$+ P(X_{s:m} - X_{r:m} \geq Y_{l:n} - Y_{k:n}, X_{r:m} > Y_{k:n})$$
$$\geq P(X_{s:m} \geq Y_{l:n}, X_{r:m} \leq Y_{k:n})$$
$$\geq \max\{P(X_{s:m} \geq Y_{l:n}) + P(X_{r:m} \leq Y_{k:n}) - 1, 0\} \qquad (16)$$
$$= \max\{P(X_{r:m} \leq Y_{k:n}) - P(X_{s:m} < Y_{l:n}), 0\}$$
$$L_4(r,s;k,l).$$

By conditioning argument and the fact that $P(X_{r:m} \leq u) = \sum_{i=r}^{m} B_{i,m}(u)$, we obtain:

$$P(X_{r:m} \leq Y_{k:n}) = \sum_{i=r}^{m} \frac{\binom{i+k-1}{i}\binom{m+n-i-k}{m-i}}{\binom{m+n}{m}}. \qquad (17)$$

Upon using (17) and substituting (16) into (15), we immediately conclude:

$$\pi_4(r,s;k,l) \geq \max \left\{ \begin{array}{c} \frac{1}{P(M \geq s)\,P(N \geq l)} \\ \sum_{m=s}^{\infty}\sum_{n=l}^{\infty}\sum_{i=r}^{m} \frac{\binom{i+k-1}{i}\binom{m+n-i-k}{m-i}}{\binom{m+n}{m}} \\ P_M(m)P_N(n) \\ -\frac{1}{P(M \geq s)\,P(N \geq l)} \\ \sum_{m=s}^{\infty}\sum_{n=l}^{\infty}\sum_{i=s}^{m} \frac{\binom{i+l-1}{i}\binom{m+n-i-l}{m-i}}{\binom{m+n}{m}} \\ P_M(m)P_N(n), 0 \end{array} \right\} = L_4(r,s;k,l).$$

Proceeding similarly, we also obtain:

$$\pi_5(r,s;k,l) = P(X_{v:M} - X_{u:M} \leq Y_{l:N} - Y_{k:N}) \geq L_5(r,s;k,l)$$

$$\max \left\{ \begin{array}{c} \frac{1}{P(M \geq s)P(N \geq l)} \\ \sum_{m=s}^{\infty}\sum_{n=l}^{\infty}\sum_{i=v}^{m} \frac{\binom{i+k-1}{i}\binom{m+n-i-l}{m-i}}{\binom{m+n}{m}} P_M(m)P_N(n) \\ -\frac{1}{P(M \geq s)P(N \geq l)} \\ \sum_{m=s}^{\infty}\sum_{n=l}^{\infty}\sum_{i=u}^{m} \frac{\binom{i+k-1}{i}\binom{m+n-i-k}{m-i}}{\binom{m+n}{m}} P_M(m)P_N(n), 0 \end{array} \right\}.$$

On the other hand, we clearly have:

$$P(X_{s:M} - X_{r:M} \geq Y_{l:N} - Y_{k:N})$$
$$\geq P(X_{r:M} \leq Y_{k:N} \leq Y_{l:N} \leq X_{s:M})$$
$$= \pi_2(r,s;k,l)$$

and:

$$P(X_{v:M} - X_{u:M} \leq Y_{l:N} - Y_{k:N})$$
$$\geq P(X_{k:N} \leq Y_{u:M} \leq X_{v:M} \leq Y_{l:N})$$
$$= \pi_3(u,v;k,l),$$

where, $\pi_2(r, s; k, l)$ and $\pi_3(u, v; k, l)$) are as defined in Theorems 2 and 3, respectively.

Therefore:

$$\pi_4(r,s;k,l) = P(X_{s:M} - X_{r:M} \geq Y_{l:N} - Y_{k:N})$$
$$\geq \max\{\pi_2(r,s;k,l), L_4(r,s;k,l)\}$$

and:

$$\pi_5(r,s;k,l) = P(X_{s:M} - X_{r:M} \leq Y_{l:N} - Y_{k:N})$$
$$\geq \max\{\pi_3(r,s;k,l), L_5(r,s;k,l)\}.$$

Thus, $X_{s:M} - X_{r:M}$ and $X_{v:M} - X_{u:M}$ are then the upper and lower prediction limits for $Y_{l:N} - Y_{k:N}$ with prediction coefficient being at least $\max\{\pi_2(r, s; k, l); L_4(r, s; k, l)\}$ and $\max\{\pi_3(r, s; k, l), L_5(r, s; k, l)\}$.

Remark 5:

The increments of order statistics $I_{r,s:M}$ represent the width of the outer and inner PIs. Since the expected width of the PI can be considered as an optimality criterion while comparing different intervals, evaluation of $E(I_{r,s:M})$ is of natural interest. In this regard, sharp bounds for $E(I_{r,s:m})$ established by Raqab (2003) would be useful.

Illustrative Data Analysis

In this section, we apply the procedures developed here on a biometric data set analyzed by Prentice (1973) and Lawless (2003). The data represent the survival times of lung cancer patients, measured from the start of chemotherapy treatment for each patient.

The survival times, in weeks, are given below:

0.14 0.29 2.14 2.57 2.71 2.86 6.14 6.29

7.29 12.0 12.86 23.43 28.71 33.0 48.57.

The methods developed in the previous sections will be applied to analyze the survival times data assuming a random sample size N and that 20 patients begin the therapeutic program. From past experience, it is known that 20% of patients had left the therapeutic program for one reason or another. Consequently, a binomial random sample size is a reasonable assumption. That is, $N \sim B(20; 0.70)$. Based on the observed ordered data above, PIs for future order statistics and order intervals from another independent sample of random size under similar conditions were obtained with prediction coefficient of at least $1-\alpha_0 = 0.90$. These intervals are presented in Tables 4 and 5, respectively.

To see how the PIs of future order statistics compare for random and fixed sample sizes, we have also presented in Table 4, the prediction coefficients based on fixed sample size $n = 20$. In Table 4, it is checked that the PIs based on the informative sample are shorter than those for fixed sample size with probability coefficient being at least the specified level $1-\alpha = 0.90$. For example, let us consider the PI of 6-th order statistic from Y-sample for random sample size as well as for fixed sample size. The shortest PI of $Y_{6:N}$ is $(X_{2:M}, X_{10:M})$ = (0.29, 2.0) with probability coverage 0.90 which is the same as the coefficient level 0.90 when the sample size is random. But for fixed sample size, the shortest PI interval of $Y_{6:20}$ is $(X_{2:20}, X_{11:20})$ = (0.29, 12.86) with the coefficient level 0.902. Similar observations can be noticed for the PI of $Y_{8:N}$. Generally, one can conclude that the randomness of the N along its respective distribution represent essential factors in allowing the PIs to become shorter when compared with fixed sample size. The outer and inner PIs of a future order interval for both random sample size N and fixed sample size $N = 20$ are displayed in Tables 5 and 6, respectively. From Tables 5 and 6, it is evident that the outer PIs for random size N are shorter than those for fixed sample size $N = 20$ under the available coefficient level being at least the specified desired level $1-\alpha = 0.90$. On the contrary, these tables shows that the inner PIs of random N is wider than the inner PIs for fixed $N = 20$.

Table 4: PIs for $Y_{k:N}$

	B(20, 0.7)						Fixed sample size				
k	$1-\alpha$	r	s	$(X_{r:M}, X_{s:M})$	$\pi_1(r, s; k)$	k	$1-\alpha$	r	s	$(X_{r:20}, X_{s:20})$	$\pi_1(r, s, k)$
5	0.90	1	9	(0.14,7.29)	0.918	5	0.90	1	10	(0.14,12.0)	0.929
	0.90	2	11	(0.29,12.86)	0.907		0.90	2	12	(0.29,23.43)	0.900
6	0.90	1	10	(0.14,12.0)	0.930	6	0.90	1	11	(0.14,12.86)	0.937
	0.90	2	10	(0.29,12.0)	0.900		0.90	2	11	(0.29,12.86)	0.902
7	0.90	1	11	(0.14,12.86)	0.938	7	0.90	1	11	(0.14,12.86)	0.900
	0.90	2	11	(0.29,12.86)	0.924		0.90	2	12	(0.29,23.43)	0.923
	0.90	3	12	(2.14,23.43)	0.922		0.90	3	13	(2.14,28.71)	0.908
8	0.90	1	11	(0.14,12.86)	0.900	8	0.90	1	12	(0.14,23.43)	0.900
	0.90	2	12	(0.29,23.43)	0.939	8	0.90	2	13	(0.29,28.71)	0.934
	0.90	3	12	(1.14,23.43)	0.922	8	0.90	3	13	(1.14,28.71)	0.911
	0.90	4	13	(2.57,28.71)	0.914	8	0.90	4	15	(2.57,48.57)	0.910

Table 5: Outer and inner PIs for $(Y_{k:N}, Y_{l:N})$ when N is random

	Outer PI						Inner PI				
(k, l)	$1-\alpha$	r	s	$(X_{r:M}, X_{s:M})$	$\pi_1(r, s; k, l)$	(k, l)	$1-\alpha$	r	s	$(X_{r:M}, X_{s:M})$	$\pi_1(r, s; k, l)$
(5,7)	0.90	1	11	(0.14,12.86)	0.920	(2,13)	0.90	5	8	(2.71,6.29)	0.900
	2	13		(0.29,28.71)	0.907		0.90	6	9	(2.86,7.29)	0.912
	-	-	-	-			0.90	8	10	(6.29,12.0)	0.900
(5,8)	0.90	1	12	(0.14,23.43)	0.925	(3,15)	0.90	7	11	(6.14,12.86)	0.905
	0.90	2	14	(0.29,33.0)	0.907		0.90	8	11	(6.29,12.86)	0.935
	-	-	-	-			0.90	9	12	(7.29,23.43)	0.909
(6,9)	0.90	1	13	(0.14,28.71)	0.943	(4,15)	0.90	8	10	(6.29,12.0)	0.915
	0.90	2	13	(0.29,28.71)	0.912		0.90	9	11	(7.29,12.86)	0.929
	-	-	-	-	-		0.90	10	12	(12.0,23.43)	0.907

Table 6: Outer and inner PIs for $(Y_{k:N}, Y_{l:N})$ when N is fixed

	Outer PI						Inner PI				
(k, l)	$1-\alpha$	r	s	$(X_{r:20}, X_{s:20})$	$\pi_1(r, s; k, l)$	(k, l)	$1-\alpha$	r	s	$(X_{r:20}, X_{s:20})$	$\pi_1(r, s; k, l)$
(5,7)	0.90	1	12	(0.14,23.43)	0.921	(2,13)	0.90	5	6	(2.71,2.86)	0.900
	0.90	2	15	(0.29,48.57)	0.904		0.90	6	8	(2.86,6.29)	0.900
	-	-	-	-	-		0.90	7	8	(6.14,6.29)	0.922
(5,8)	0.90	1	13	(0.14,28.71)	0.920	(3,15)	0.90	7	9	(6.14,7.29)	0.910
	0.90	2	15	(0.29,48.57)	0.900		0.90	8	10	(6.29,12.0)	0.915
	-	-	-	-	-		0.90	9	10	(7.29,12.0)	0.932
(6,9)	0.90	1	14	(0.14,33.0)	0.934	(4,15)	0.90	8	9	(6.29,7.29)	0.900
	0.90	2	14	(0.29,33.0)	0.900		0.90	9	10	(7.29,12.0)	0.906

Further, we can also obtain the upper prediction limits for the increments of order statistics from the future sample with coefficient at least $1-\alpha_0 = 0.90$. For example, the upper prediction limit for $Y_{12:N} - Y_{5:N}$ is $X_{9:N} - X_{1:N} = 7.17$ with prediction coefficient 0.902 while the upper prediction limit for $Y_{15:N} - Y_{5:N}$ is $X_{14:N} - X_{2:N} = 32.71$ with prediction coefficient 0.903.

Acknowledgement

Author's Contributions

The authors carried out this work together. Both authors read and approved the final version of the manuscript.

Ethics

The authors declare that there is no conflict interests regarding the publication of this manuscript.

References

Abdel-Hamid, A.H. and E.K. AL-Hussaini, 2014. Bayesian prediction for type-II progressive censored data from the Rayleigh distribution under progressive-stress model. J. Stat. Computation Simulation, 84: 1297-1312.

Ahmadi, J. and N. Balakrishnan, 2011. Distribution-free prediction intervals for order statistics based on record coverage. J. Korean Stat. Society, 40: 181-192.

Arnold, B.C., N. Balakrishnan and H.K. Nagraraja, 2008. *A First Course in Order Statistics*. 1st Edn., SIAM, Philadelphia, ISBN-10: 0898716489, pp: 279.

Asgharzadeh, A. and R. Valiollahi, 2010. Prediction intervals for proportional hazard rate models based on progressively type II censored samples. Commun. Korean Stat. Society, 17: 99-106.

Balakrishnan, N., C.T. Lin and P.S. Chan, 2005. A comparison of two simple prediction intervals for exponential distribution. IEEE Trans. Reliability, 54: 27-33.

Barakat, H.M., M.E. El-Adll and A.E. Aly, 2011. Exact prediction intervals for future exponential lifetime based on ramdom generalized order statistics. Comput. Math. Applic., 61: 1366-1378.

Barakat, H.M., M. El-Adll and A.E. Aly, 2014. Prediction intervals of future observations for a sample of random size from any continuous distribuiton. Math. Comput. Simulation, 97: 1-13.

Barakat, H.M., E.M. Nigm, M.E. El-Adll and M. Yusuf, 2016. Prediction of future generalized order statistics based on exponential distribution with random sample size. Stat. Papers. DOI: 10.1007/s00362-016-0779-2

Barlow, R.E. and F. Proschan, 1981. Statistical Theory of Reliability and Life Testing. 2nd Edn., To begin With, Silver Spring, pp: 290.

Basak, I., P. Basak and N. Balakrishnan, 2006. On some predictors of times to failures of censored items in progressively censored sample. Comput. Stat. Data Anal., 50: 1313-1337.

Al-Mutairi, J.S. and M.Z. Raqab, 2017. Confidence intervals for quantiles based on samples of random sizes, Stat. Papers. DOI: 10.1007/s00362-017-0935-3

Basiri, E., J. Ahmadi and M.Z. Raqab, 2016. Comparison among non-parametric prediction intervals of order statistics. Commun. Stat. Theory Meth., 45: 2699-2713.

Buhrman, J.M., 1973. On order statistics when the sample size has a binomial distribution. Stat. Neerlandica, 27: 125-126.

David, H.A. and H.N. Nagaraja, 2003. *Order Statistics*. 3rd Edn, John Wiley and Sons, Hoboken, New Jersey.

El-Adll, M.E., 2011. Predicting future lifetime based on random number of three parameters Weibull distribution. Math. Comput. Simula., 81: 1842-1854.

El-Adll, M.E. and A.E. Aly, 2016. Prediction intervals of future generalized order statistics from Pareto distribution. J. Applied Stat. Sci., 22: 111-125.

Fligner, M.A. and D.A. Wolfe, 1976. Some applications of sample analogues to the probability integral transformation and a coverage property. Am. Statistician, 30: 78-85.

Fligner, M.A. and D.A. Wolfe, 1979. Nonparametric prediction intervals for a future sample median. J. Am. Stat. Assoc., 74: 453-456.

Gulati, S. and W.J. Padgett, 2003. *Parametric and Nonparametric Inference from Record-Breaking Data*. 1st Edn., Springer-Verlag, New York.

Kaminsky, K.S. and P.I. Nelson, 1998. Prediction Intervals, In: Handbook of Statistics, Balakrishnan, N. and C.R. Rao (Eds.), Amesterdam, North Holland, pp: 431-450.

Krewski, D., 1976. Distribution-free confidence intervals for quantile intervals. J. Am. Stat. Assoc., 71: 420-422.

Lawless, J.F., 2003. Statistical Models and Methods for Lifetime Data. 2nd Edn, John Wiley and Sons, Inc, USA.

Prentice, R.L., 1973. Exponential survival with censoring and explanatory variables. Biometrika, 60: 279-288.

Raghunandanan, K. and S.A. Patil, 1972. On order statistics for random sample size. Stat. Neerlandica, 26: 121-126.

Raqab, M.Z., 2003. Evaluating improvements for spacings of order statistics. Extremes, 6: 259-273.

Raqab, M.Z. and M.T. Madi, 2002. Bayesian prediction of the total time on test using doubly censored Rayleigh data. J. Stat. Computa. Simulation, 72: 781-789.

Sultan, K.S. and A.H.A. Ellah, 2006. Exact prediction intervals for exponential lifetime based on random sample size. Int. J. Comput. Math., 83: 867-878.

Wilks, S.S., 1962. Mathematical Statistics. 1st Edn., John Wiley and Sons, New York.