Malaysian Journal of Mathematical Sciences 10(3): 329–343 (2016)

MALAYSIAN JOURNAL OF MATHEMATICAL SCIENCES

Journal homepage: http://einspem.upm.edu.my/journal

Prediction Bounds for Order Statistics and Progressive Censored Values from Two Independent Samples

Abbas Rasouli * and S. Rajaei Salmasi

Department of Statistics, University of Zanjan, Iran

E-mail: rasouli@znu.ac.ir
**Corresponding author*

ABSTRACT

The aim of this paper is to construction prediction bounds for future progressive Type-II random censored sample from an independent Y-sample by observations of n ordered X-sample and reverse problem. We show that these prediction bounds have closed form and distribution of them are free. Finally, theoretical results are illustrated by simulation samples and two real data sets.

Keywords: Order statistics, Progressive censored data, Prediction bounds, Lifetime experiment.

1. Introduction

Suppose $Y_1, \ldots, Y_n$ are random sample from a continuous cumulative distribution function (cdf) $F(x)$ and probability density function (pdf) $f(x)$. Let us introduce the corresponding order statistics (OS) from the Y_i's with $Y_{1:n} \leq Y_{2:n} \leq \ldots \leq Y_{n:n}$. In life-testing and reliability experiments we may encounter with units that are lost or removed from experimentation before failure. In such case we say that our random sample is censored. Censoring can be used as a way to limit the time, cost, or a combination of both. The progressive Type-II censored (PTC) scheme is one of the most important censoring designs in lifetime studies that is more flexible and has follow description: Suppose n experimental items of the same kind are placed on the test; at the ith failure time, we observe $X^{\mathbf{R}}_{(i,m):n}$ and R_i items from the remaining items are removed from the experiment; $i = 1, 2, \ldots, m$. Here, $m + \sum_{i=1}^{m} R_i = n$ and $\mathbf{R} = (R_1, \ldots, R_m)$ is PTC scheme. For more detail see Cohen, 1963), Balakrishnan (2007), Balakrishnan and Sandhu (1995, 1996), Balakrishnan and Aggarwala (2000), Salehi et al. (2015).

Prediction or prediction problem has special position in statistics and it has two known types in lifetime studies: one sample prediction and two sample prediction. In the one sample prediction; we would predict the future random variables from the same sequence which is dependent on the observed variables. In the two sample prediction; the observed sample and the future sample are stochastically independent. Both, one and two sample prediction problems have discussed by several authors; see Ahsanullah (1980), Kaminsky and Nelson (1998), Ahmadi and Balakrishnan (2005, 2011) and Raqab and Balakrishnan (2008).

The aim of this article is construction prediction bounds for future PTC sample based on observed OS and inverse, prediction bounds for OS based on future PTC data. In next section, we try to make prediction bounds for OS from a future Y-sample by PTC observations from X-sample. In section 3, we present prediction bounds for a PTC order statistics of a future sample by a sample of observed OS. All these prediction bounds will proved that are free distributed. In sections 4 and 5 we present one sided prediction bounds for a future sample by an observed sample, also two numerical examples to illustration the proposed procedures.

2. Prediction of future order statistics based on progressive Type-II censored sample

In this section, two-sided prediction bounds, for one of OS from a future sample, are obtained. These prediction bounds are constructed by an observed PTC sample. It is shown that $\alpha_1(p, q, n, k) = P(M_p \leq Y_{k:n} \leq M_q)$; (the coverage probabilities of these bounds) does not depend on the parent distribution F; where M_r, is the rth PTC order statistic from X-sample and $Y_{k:n}$ is kth ordinary order statistic from Y-sample. As mention before, let

$\mathbf{R} = (R_1, \ldots, R_m)$ be PTC scheme and X is an absolutely continuous random variable. The marginal distribution of $X^{\mathbf{R}}_{(s,m):n}$; $(1 \le s \le m \le n)$, is given by

$$f^{\mathbf{R}}_{(s,m):n}(x_s) = k' \sum_{i=0}^{s-1} d_{i,s-1}(R_1 + 1, \ldots, R_{s-1} + 1) f(x_s)(1 - F(x_s)^{R'_i - 1}, \quad (1)$$

where $x_s \in (L_f, U_f)$ and $d_{i,s-1}(b_1, \ldots, b_r) = \frac{(-1)^i}{\{\prod_{j=1}^{i} \sum_{k=r-i+1}^{r-i+j} b_k\}\{\prod_{j=1}^{r-i} \sum_{k=j}^{r-i} b_k\}}$, with $\prod_{j=1}^{0} h_j = 1, \sum_{j=i}^{i-1} h_j = 0$; and $k' = n(n - R_1 - 1) \ldots (n - R_1 - \ldots - R_{s-1} - s + 1)$, $R'_i = (1 + R^*_s) + \sum_{j=s-i}^{s-1}(1 + R_j)$, $R^*_s + s = n - R_{s-1} - \ldots - R_1$; see for more detail Balakrishnan et. al. (2002).

Theorem 1. Suppose M_r is the rth PTC order statistic from the X-sample of i.i.d continuous random variables with the cdf $F(x)$ and the pdf $f(x)$. let $Y_{i:n} \le Y_{2:n} \le \ldots \le Y_{n:n}$ are the OS from a future random sample of size n with the same distribution. Then $\alpha_1(p, q, n, k) = P(M_p \le Y_{k:n} \le M_q)$, $1 \le p \le q \le n$, is distribution free coverage probability of two-sided prediction bound for $Y_{k:n}$; $1 \le k \le n$, and is given by

$$\begin{aligned} \alpha_1(p, q, n, k) &= n_k^{k'} \sum_{l=0}^{k-1} k_l \left\{ \sum_{w=0}^{p-1} \frac{d_1}{R'_w} \left[(n - k + l + 1)^{-1} - (n + R'_w - k + l + 1)^{-1} \right] \right. \\ &\quad \left. - \sum_{w=0}^{q-1} \frac{d_2}{R'_w} \left[(n - k + l + 1)^{-1} - (n + R'_w - k + l + 1)^{-1} \right] \right\}, \end{aligned}$$

where $n_k^{k'} = k' \binom{n}{k} k$, $k_l = (-1)^l \binom{k-1}{l}$ also d_1 and d_2 are respectively $d_{w,p-1}(R_1 + 1, \ldots, R_{p-1} + 1)$ and $d_{w,q-1}(R_1 + 1, \ldots, R_{q-1} + 1)$.

Proof. Suppose y is given, by continuity assumption for M_m; $(1 \le m \le n)$, we get

$$P(M_p \le y \le M_q) = P(y \ge M_p) - P(M_q \le y). \quad (2)$$

By Eq. (1); Eq. (2) can be obtained as

$$\begin{aligned} & k' \sum_{w=0}^{p-1} d_1 \int_{-\infty}^{y} f(t)(1 - F(t))^{R'_w - 1} dt - k' \sum_{w=0}^{q-1} d_2 \int_{-\infty}^{y} f(t)(1 - F(t))^{R'_w - 1} dt \\ = \quad & k' \left[\sum_{w=0}^{p-1} d_1 \frac{1 - (1 - F(y))^{R'_w}}{R'_w} - \sum_{w=0}^{q-1} d_2 \frac{1 - (1 - F(y))^{R'_w}}{R'_w} \right]. \quad (3) \end{aligned}$$

Now, using the total probability role, we will have

$$\begin{aligned} P(M_p \leq Y_{k:n} \leq M_q) &= \int_{-\infty}^{+\infty} P(M_p \leq Y_{k:n} \leq M_q \mid Y_{k:n} = y) dF_{Y_{k:n}}(y) \\ &= \int_{-\infty}^{+\infty} P(M_p \leq y \leq M_q) dF_{Y_{k:n}}(y), \end{aligned} \tag{4}$$

where $F_{Y_{k:n}}(y)$ is cdf of $Y_{k:n}$; $(1 \leq k \leq n)$.

With substituting the pdf of $Y_{k:n}$ and expression (3) in Eq. (4), we will have

$$\begin{aligned} P(M_p \leq Y_{k:n} \leq M_q) &= n_k^{k'} \int_{-\infty}^{+\infty} \left[\sum_{w=0}^{p-1} d_1 \frac{1-(1-F(y))^{R'_w}}{R'_w} \right. \\ &\quad \left. - \sum_{w=0}^{q-1} d_2 \frac{1-(1-F(y))^{R'_w}}{R'_w} \right] (F(y))^{k-1}(1-F(y))^{n-k} f(y) dy \\ &= n_k^{k'} \int_0^1 [\left[\sum_{w=0}^{p-1} \frac{d_1}{R'_w} (1-y^{R'_w}) \right. \\ &\quad \left. - \sum_{w=0}^{q-1} \frac{d_2}{R'_w} (1-y^{R'_w}) \right] (1-y)^{k-1} y^{n-k} dy \\ &= n_k^{k'} \sum_{l=0}^{k-1} k_l \left\{ \sum_{w=0}^{p-1} \frac{d_1}{R'_w} \left[(n-k+l+1)^{-1} \right. \right. \\ &\quad \left. - (n+R'_w-k+l+1)^{-1} \right] \\ &\quad \left. - \sum_{w=0}^{q-1} \frac{d_2}{R'_w} \left[(n-k+l+1)^{-1} - (n+R'_w-k+l+1)^{-1} \right] \right\}. \end{aligned}$$

Suppose the desired confidence level α_0 and n, k are known. With appropriate the parameters p and q we can hope that $\alpha_1(p,q,n,k)$ exceeds α_0. Of course, The confidence coefficient may not equal to α_0, because $\alpha_1(p,q,n,k)$ is a step function but may be chosen to be slightly larger α_0. We know obviously that the choice of p and q is not unique and we would like to construct a prediction bounds as short as possible among all prediction bounds with the same level. So, for a given confidence level α_0 and specified k and n, the two-sided prediction bounds exist if and only if

$$p(M_p \leq Y_{k:n} \leq M_q) \geqslant \alpha_0.$$

In Tables 1 and 3 we have obtained the exact values of $\alpha_1(p, q; n, k)$ for $n = 12$ and 19 and two PTC schemes and some selected amounts of p, q and k. Tables 2 and 4 are related to simulation values of $\alpha_1(p, q; n, k)$ based on 10000 time repudiation of computing coverage probabilities $p(M_p \leq Y_{k:n} \leq M_q)$ and the parent exponential distribution with mean 1.

Table 1: The amounts of $\alpha_1(p, q, n, k)$ for $n = 12$ and some amounts of p, q and k, with censoring scheme $(5, 0, 0, 0, 0, 0, 0)$.

			q			
n	k	p	4	5	6	7
12	3	1	0.7730	0.8543	0.8833	0.8903
		2	0.4511	0.5325	0.5611	0.5685
		3	0.1724	0.2537	0.2827	0.2897
12	5	1	0.6372	0.8275	0.9333	0.9733
		2	0.4896	0.6799	0.7852	0.8257
		3	0.2504	0.4408	0.5466	0.5866
12	7	1	0.3768	0.6224	0.8338	0.9583
		2	0.3316	0.5772	0.7886	0.9131
		3	0.2074	0.4531	0.6644	0.7889
12	9	1	0.1506	0.3419	0.6045	0.8592
		2	0.1419	0.3332	0.5958	0.8505
		3	0.1028	0.2941	0.5567	0.8115
12	11	1	0.0294	0.1012	0.2738	0.5954
		2	0.0287	0.1005	0.2731	0.5946
		3	0.0232	0.0950	0.2676	0.5892

Table 2: Simulation amounts of $\alpha_1(p,q,n,k)$ for $n = 12$ and some amounts of p, q and k, with censoring scheme $(5,0,0,0,0,0,0)$.

			q			
n	k	p	4	5	6	7
12	3	1	0.7750	0.8543	0.8821	0.8915
		2	0.4530	0.5351	0.5628	0.5697
		3	0.1739	0.2535	0.2858	0.2888
12	5	1	0.6360	0.8280	0.9334	0.9736
		2	0.4892	0.6796	0.7847	0.8247
		3	0.2499	0.4403	0.5469	0.5868
12	7	1	0.3775	0.6221	0.8330	0.9579
		2	0.3315	0.5769	0.7880	0.9134
		3	0.2084	0.4538	0.6630	0.7884
12	9	1	0.1509	0.3412	0.6051	0.8588
		2	0.1418	0.3338	0.5962	0.8515
		3	0.1041	0.2949	0.5581	0.8120
12	11	1	0.0299	0.1016	0.2721	0.5977
		2	0.0292	0.1000	0.2738	0.5954
		3	0.0226	0.0926	0.2679	0.5866

Table 3: The amounts of $\alpha_1(p,q,n,k)$ for $n = 19$ and some amounts of p, q and k, with censoring scheme $(0,0,0,2,3,5,0,0,0)$.

			q				
n	k	p	5	6	7	8	9
19	2	1	0.6731	0.7243	0.7513	0.7561	0.7567
		2	0.4163	0.4675	0.4945	0.4993	0.4999
		3	0.2183	0.2694	0.2964	0.3013	0.3018
		4	0.0862	0.1374	0.1644	0.1692	0.1698
19	6	1	0.3675	0.5707	0.8474	0.9584	0.9866
		2	0.3325	0.5357	0.8123	0.9234	0.9616
		3	0.2613	0.4645	0.7412	0.8522	0.8804
		4	0.1538	0.3570	0.6337	0.7447	0.7729
19	10	1	0.0521	0.1414	0.4857	0.7958	0.9589
		2	0.0508	0.1401	0.4844	0.7945	0.9575
		3	0.0460	0.1352	0.4795	0.7896	0.9527
		4	0.0332	0.1225	0.4668	0.7769	0.9400
19	14	1	0.0018	0.0103	0.1447	0.4423	0.7977
		2	0.0018	0.0103	0.1447	0.4423	0.7977
		3	0.0017	0.0102	0.1446	0.4422	0.7976
		4	0.0014	0.0099	0.1443	0.4419	0.7973

Table 4: Simulation amounts of $\alpha_1(p,q,n,k)$ for $n=19$ and some amounts of p, q and k, with censoring scheme $(0,0,0,2,3,5,0,0,0)$.

			q				
n	k	p	5	6	7	8	9
19	2	1	0.6725	0.7234	0.7522	0.7545	0.7558
		2	0.4156	0.4671	0.4934	0.4989	0.4994
		3	0.2168	0.2675	0.2958	0.3004	0.3017
		4	0.0861	0.1236	0.1374	0.1659	0.1700
19	6	1	0.3644	0.5708	0.8464	0.9583	0.9867
		2	0.3306	0.5342	0.8130	0.9228	0.9509
		3	0.2615	0.4626	0.7412	0.8513	0.8801
		4	0.1534	0.3590	0.6338	0.7440	0.7747
19	10	1	0.0522	0.1413	0.4836	0.7953	0.9563
		2	0.0500	0.1411	0.4824	0.7940	0.9586
		3	0.0449	0.1355	0.4791	0.7875	0.9530
		4	0.0330	0.1230	0.4625	0.7758	0.9394
19	14	1	0.0018	0.0111	0.1451	0.4447	0.7997
		2	0.0017	0.0105	0.1443	0.4436	0.7976
		3	0.0017	0.0105	0.1410	0.4419	0.7971
		4	0.0014	0.0088	0.1396	0.4408	0.7937

Let's k and n be fixed. From Table 1, we see that the prediction coefficient $\alpha_1(p,q,n,k)$ is decreasing in p and increasing in q. Based on two samples with sizes $n=12$; first sample ordinary ordered and second sample ordered progressively with PTC scheme $(5,0,0,0,0,0)$. We consider when k increase, (other two parameters p and q are fixed) coverage probabilities decrease; (compare $\alpha_1(1,4;12,3)=0.7730$ with $\alpha_1(1,4;12,11)=0.0294$) or when q increase (other two parameters p and k are fixed) coverage probabilities increase; (compare $\alpha_1(2,4;12,5)$ with $\alpha_1(2,7;12,5)$ or other values in this Table). It is natural because (for example in last comparison) prediction intervals become long and follow it coverage probabilities increase.

Table 3 provides $\alpha_1(p,q,n,k)$ for $n=19$ and selected values of k, p and q with different PTC schemes. From this Table, similar to Table 1, we consider that for fixed k and n, the prediction coefficient $\alpha_1(p,q,n,k)$ is decreasing in p and increasing in q.

3. Prediction of future progressive Type-II censored based on order sample

Our aim here is to construction $100(1-\alpha)\%$ two sided prediction bounds $(Y_{p:n}, Y_{q:n})$; $1 \leq p \leq q \leq n$, for M_l; the lth PTC order statistic from a future X-sample such that $P(Y_{p:n} \leq M_l \leq Y_{q:n}) = 1-\alpha$. We will prove that the coverage probability of prediction bounds $(Y_{p:n}, Y_{q:n})$; $1 \leq p \leq q \leq n$, i.e. $P(Y_{p:n} \leq M_l \leq Y_{q:n})$ that here is denoted by $\alpha_2(p, q; n, l)$ dose not depend on the parent distribution F.

Theorem 2. By the assumptions of Theorem 1, then $(Y_{p:n}, Y_{q:n})$, $1 \leq p \leq q \leq n$, is a two-sided (distribution free) prediction bounds for the lth PTC order statistic M_l; $(1 \leq l \leq n)$, and we see that

$$\alpha_2(p, q; l, n) = k' \sum_{i=0}^{l-1} \sum_{t=p}^{q-1} \sum_{w=0}^{t} d_3 \binom{n}{t} \binom{t}{w} \frac{(-1)^w}{n-t+R_i'+w}, \tag{5}$$

where $d_3 = d_{i,l-1}(R_1+1, \ldots, R_{l-1}+1)$.

Proof. For a given m, by continuity assumption on the ordinary order statistic $Y_{k:n}$; $(k \geq 1)$, we can write

$$P(Y_{p:n} \leq m \leq Y_{q:n}) = P(Y_{p:n} \leq m) - P(Y_{q:n} \leq m). \tag{6}$$

Now, by the binomial expansion for $F_{Y_{r:n}}(y)$; $(1 \leq r \leq n)$, as

$$F_{r:n}(x) = \sum_{i=r}^{n} \binom{n}{i} (F(x))^i (1-F(x))^{n-x}, \tag{7}$$

and substitution in Eq. (6), we get

$$P(Y_{p:n} \leq m \leq Y_{q:n}) = \sum_{t=p}^{q-1} \binom{n}{t} (F(m))^t (1-F(m))^{n-t}.$$

Let $f_{M_l}(m)$ is pdf M_l; $(l \geq 1)$. For $P(Y_{p:n} \leq M_l \leq Y_{q:n})$, by conditioning on $M_l = m$, we have

$$\begin{aligned} P(Y_{p:n} \leq M_l \leq Y_{q:n}) &= \int_{-\infty}^{+\infty} P(Y_{p:n} \leq M_l \leq Y_{q:n} \mid M_l = m) f_{M_l}(m) dm \\ &= \int_{-\infty}^{+\infty} P(Y_{p:n} \leq m \leq Y_{q:n}) f_{M_l}(m) dm \\ &= k' \sum_{t=p}^{q-1} \binom{n}{t} \sum_{i=0}^{l-1} d_3 \int_{-\infty}^{+\infty} F(m)^t (1-F(m))^{n-t+R_i'-1} f(m) dm \\ &= k' \sum_{t=p}^{q-1} \binom{n}{t} \sum_{i=0}^{l-1} d_3 \int_0^1 (1-m)^t m^{n-t+R_i'-1} dm \\ &= k' \sum_{i=0}^{l-1} \sum_{t=p}^{q-1} \sum_{w=0}^{t} d_3 \binom{n}{t} \binom{t}{w} \frac{(-1)^w}{n-t+R_i'+w}, \end{aligned}$$

which is the coverage probability $\alpha_2(p, q, n, l)$, presented in Eq. (5).

Corollary 1: By assumptions of Theorem 2,

(I) : $Y_{p:n}$, $p \leq n$, is a lower prediction bound, and

(II) : $Y_{q:n}$, $q \leq n$, is an upper prediction bound for M_l; $(l \geq 1)$.

Here, in Tables 5, 7 we have presented the exact values of $\alpha_2(p, q; n, l)$ for $n = 10$ and 17 and two PTC schemes and some selected values of p, q and k. Also; Tables 6 and 8 are related to simulation amounts of $\alpha_2(p, q; n, l)$ based on 10000 time repetition of computing $p(Y_{p:n} \leq M_l \leq Y_{q:n})$ and parent distribution exponential with mean 1.

Table 5: The amounts of $\alpha_2(p, q, n, l)$ for $n = 10$ and some amounts of p, q and k, with censoring scheme $(0, 0, 0, 0, 0, 4)$.

			q			
n	l	p	4	6	8	10
10	2	1	0.6114	0.7345	0.7604	0.7630
		2	0.3482	0.4713	0.4972	0.4999
		3	0.1393	0.2623	0.2882	0.2909
10	4	1	0.4566	0.7717	0.9217	0.9551
		2	0.3482	0.6633	0.8134	0.8467
		3	0.1857	0.5008	0.6508	0.6842
10	6	1	0.1795	0.4945	0.8196	0.9783
		2	0.1562	0.4713	0.7964	0.9551
		3	0.1000	0.4150	0.7401	0.8988

From Table 5 it can be see that when n and l are fixed, the coverage probabilities $\alpha_2(p, q, n, l)$ are decreasing in p and increasing in q.

Table 6: Simulation amounts of $\alpha_2(p, q, n, l)$ for $n = 10$ and some amounts of p, q and k, with censoring scheme. $(0, 0, 0, 0, 0, 4)$.

			q			
n	l	p	4	6	8	10
10	2	1	0.6226	0.7226	0.7701	0.7731
		2	0.3858	0.4795	0.4986	0.4995
		3	0.1444	0.2706	0.2948	0.2953
10	4	1	0.4660	0.7819	0.9247	0.9597
		2	0.3512	0.6734	0.8222	0.8497
		3	0.1933	0.5022	0.6559	0.6838
10	6	1	0.1836	0.4946	0.8262	0.9800
		2	0.1611	0.4736	0.7941	0.9587
		3	0.1076	0.4219	0.7431	0.9043

Table7: The amounts of $\alpha_2(p, q, n, l)$ for $n = 17$ and some amounts of p, q and k, with censoring scheme $(0, 1, 1, 0, 3, 0, 1, 2, 0)$.

			q				
n	l	p	5	8	11	14	17
17	1	2	0.2201	0.2410	0.2423	0.2424	0.2424
		3	0.0913	0.1122	0.1135	0.1136	0.1136
		4	0.0290	0.0499	0.1051	0.0513	0.0513
17	3	2	0.6499	0.6769	0.7060	0.7082	0.7082
		3	0.3026	0.4801	0.5092	0.5114	0.5114
		4	0.1307	0.3082	0.3373	0.3394	0.3395
17	5	2	0.3758	0.7497	0.9001	0.9264	0.9279
		3	0.2789	0.6529	0.8033	0.8296	0.8311
		4	0.1478	0.5218	0.6722	0.6985	0.6999
17	7	2	0.1145	0.4284	0.7774	0.9572	0.9913
		3	0.0964	0.4104	0.7594	0.9391	0.9733
		4	0.0600	0.3739	0.7229	0.9027	0.9369

From Table 7 it can be see that when n and l are fixed, the coverage probabilities $\alpha_2(p, q, n, l)$ are decreasing in p and increasing in q.

Table 8: Simulation amounts of $\alpha_2(p, q, n, l)$ for $n = 17$ and some amounts of p, q and k, with censoring scheme $(0, 1, 1, 0, 3, 0, 1, 2, 0)$.

			q				
n	l	p	5	8	11	14	17
17	1	2	0.2262	0.2483	0.2491	0.2517	0.2550
		3	0.0947	0.1171	0.1183	0.1158	0.1187
		4	0.0291	0.0486	0.0535	0.0543	0.0549
17	3	2	0.4984	0.6860	0.7126	0.7145	0.7147
		3	0.3088	0.4878	0.5169	0.5185	0.5193
		4	0.1376	0.3170	0.3429	0.3482	0.3527
17	5	2	0.3830	0.7530	0.9039	0.9330	0.9339
		3	0.2826	0.6626	0.8046	0.8307	0.8343
		4	0.1544	0.5242	0.6757	0.6967	0.6985
17	7	2	0.1170	0.4342	0.7846	0.9606	0.9924
		3	0.0924	0.4147	0.7637	0.9425	0.9911
		4	0.0628	0.3812	0.7346	0.9038	0.9372

4. One-sided prediction for future OS

Let us an order statistics to be fixed. In this section we try to construct one-sided prediction bounds for this order statistic by a future sample. Construction of prediction bounds are based on observed PTC order statistics. It is proved that coverage probabilities of prediction bounds, i.e. $\alpha_3(p,n,k) = P(Y_{k:n} \geq M_p)$, are free of the parent distribution F.

Theorem 3. Suppose $X_1, X_2, \cdots, X_n$ are i.i.d continuous random variables with cdf $F(x)$ and pdf $f(x)$. Let kth order statistic from a future random sample of size n denoted by $Y_{k:n}$. If M_p is the pth PTC order statistics from the X-sample, then (M_p, ∞) is one-sided prediction bound for $Y_{k:n}$; and probability of this event dose not depend on the F and is calculated by

$$\alpha_3(p,n,k) = n_k^{k'} \sum_{i=0}^{p-1} \sum_{w=0}^{k-1} d_4 \frac{k_w}{R_i^{\prime}} \left[(n-k+w+1)^{-1} - \left(n + R_i^{\prime} - k + w + 1\right)^{-1} \right],$$

where $d_4 = d_{i,p-1}(R_1+1, \ldots, R_{p-1}+1)$.

Proof. By continuity assumption for random statistics M_p; $(p \geq 1)$, for given y, we can write $P(M_p \leq y) = F_{M_p}(y)$.

Now based on Eq. (1), we have

$$P(M_p \leq y) = k' \sum_{i=0}^{p-1} d_4 \frac{1-(1-F(y))^{R_i'}}{R_i'}. \tag{8}$$

Using the total probability role, we can write

$$\begin{aligned} P(M_p \quad \leq \quad Y_{k:n}) &= \int_{-\infty}^{+\infty} P(Y_{k:n} \geq M_p \mid Y_{k:n} = y) dF_{Y_{k:n}}(y) \\ &= n_k^{k'} \sum_{i=0}^{p-1} \frac{d_4}{R_i'} \int_{-\infty}^{+\infty} \left(1 - \frac{1}{(1-F(y))^{-R_i'}}\right) F(y)^{k-1} (1-F(y))^{n-k} f(y) dy \\ &= n_k^{k'} \sum_{i=0}^{p-1} \frac{d_4}{R_i'} \int_0^1 (1 - y^{R_i'+l-1})(1-y)^{k-1} y^{n-k} dy \\ &= n_k^{k'} \sum_{i=0}^{p-1} \sum_{w=0}^{k-1} \frac{k_w d_4}{R_i'} \left[(n-k+w+1)^{-1} - (n+R_i'-k+w+1)^{-1} \right]. \end{aligned}$$

5. Illustrative examples

In this section, Two data sets are used to verify correctness of results of previous sections. First set of data is related rainfall recorded at LACC between years 1900 until 2000. Second set is reported by (Lawless, 1982, p. 288) and is related to ball bearing (number of million revolutions before failing for each ball bearing). Observed order values of annual rainfall are presented in Table 9. In the first example, prediction bounds for the future PTC order observations are obtained with prediction probability of at least $\alpha_0 = 0.89$. These intervals are reported in Table 10. In the second example, under a PTC scheme, the progressively censored sample related to the ball bearings data is presented in Table 11. Using this data set, two and one sided prediction bounds for future order statistics $Y_{k:n}$ are calculated and results are reported in Tables 12 and 13 respectively.

Table 9: Ordered annual rainfall at $LACC$ during 1900-2000.

r	1	2	3	4	5	6	7	8	9	10
$Year$	1960	1958	1923	1971	1975	1974	1989	1986	1969	1963
$Y_{r:100}$	4.85	5.58	6.67	7.17	7.21	7.22	7.35	7.66	7.74	7.93
r	20	30	50	70	80	90	95	98	99	100
$Year$	1980	1941	1928	1926	1921	1937	1992	1982	1940	1977
$Y_{r:100}$	8.96	11.10	12.66	18.03	19.66	23.43	27.36	31.28	32.76	33.44

Table 10: Future progressive type-II censored Prediction with censoring scheme $(0, 0, 0, 0, 0, 4)$ based on observed order statistics from Table 9.

(n, l)	(p, q)	(Y_p, Y_q)	α
$(10, 6)$	$(1, 10)$	(4.85, 7.93)	0.9783
$(10, 6)$	$(2, 10)$	(5.58, 7.93)	0.9551
$(10, 6)$	$(3, 10)$	(6.67, 7.93)	0.8988
$(10, 4)$	$(1, 10)$	(4.85, 7.93)	0.9551
$(10, 4)$	$(1, 8)$	(4.85, 7.66)	0.9217

Table 11: Progressive Type-II censored for the ball bearing data

r	1	2	3	4	5	6	7	8	9	10
r_i	0	0	0	3	0	0	0	3	0	0
M_i	17.88	28.92	33.00	41.52	51.84	51.96	55.12	55.56	68.88	84.12
r	11	12								
r_i	0	5								
M_i	93.12	98.64								

Table 12: Prediction bounds for OS based on observed progressive censored in Table 11.

(n,k)	(p,q)	(M_p, M_q)	α
$(23,5)$	$(2,5)$	(28.92, 51.84)	0.4306
$(23,5)$	$(2,6)$	(28.92, 51.96)	0.5847
$(23,5)$	$(2,7)$	(28.92, 55.12)	0.7049
$(23,7)$	$(3,4)$	(33.00, 41.52)	0.0745
$(23,7)$	$(2,4)$	(28.92, 41.52)	0.1183
$(23,9)$	$(3,4)$	(33.00, 41.52)	0.0278
$(23,9)$	$(2,4)$	(28.92, 41.52)	0.0405
$(23,9)$	$(1,7)$	(17.88, 55.12)	0.3252
$(23,9)$	$(2,7)$	(28.92, 55.12)	0.3210

Table 13: One sided prediction bounds for OS based on observed progressive censored in Table 11.

(n,k)	p	M_p	α
$(23,2)$	3	33.00	0.3038
$(23,4)$	3	33.00	0.6674
$(23,6)$	5	51.84	0.6145
$(23,6)$	7	55.12	0.3124
$(23,12)$	1	17.84	0.9999
$(23,14)$	7	55.12	0.9036

6. Conclusions

In this paper, we have derived prediction bounds for the OS from the future PTC data and vice versa; by the same parent distribution assumption for two independent random samples. These intervals have exact expressions and are distribution free.

References

Ahmadi, J. and Balakrishnan, N. (2005). Distribution-free confidence intervals for quantile intervals based on current records. *Statistics & probability letters*, 75(3):190–202.

Ahmadi, J. and Balakrishnan, N. (2011). Distribution-free prediction intervals for order statistics based on record coverage. *Journal of the Korean Statistical Society*, 40(2):181–192.

Ahsanullah, M. (1980). Linear prediction of record values for the two parameter exponential distribution. *Annals of the Institute of Statistical Mathematics*, 32(1):363–368.

Balakrishnan, N. (2007). Progressive censoring methodology: an appraisal. *Test*, 16(2):211–259.

Balakrishnan, N. and Aggarwala, R. (2000). *Progressive censoring: theory, methods, and applications.* Springer Science & Business Media.

Balakrishnan, N. and Sandhu, R. (1995). A simple simulational algorithm for generating progressive type-ii censored samples. *The American Statistician*, 49(2):229–230.

Balakrishnan, N. and Sandhu, R. (1996). Best linear unbiased and maximum likelihood estimation for exponential distributions under general progressive type-ii censored samples. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 1–9.

Cohen, A. Truncated and censored samples: Theory and applications. 1991. *Marcel Dekker Inc*, 3:379–423.

Cohen, A. C. (1963). Progressively censored samples in life testing. *Technometrics*, 5(3):327–339.

Kaminsky, K. S. and Nelson, P. I. (1998). 15 prediction of order statistics. *Handbook of statistics*, 17:431–450.

Lawless, J. F. (1982). *Statistical models and methods for lifetime data.* John Wiley & Sons.

Raqab, M. Z. and Balakrishnan, N. (2008). Prediction intervals for future records. *Statistics & Probability Letters*, 78(13):1955–1963.

Salehi, M., Ahmadi, J., and Balakrishnan, N. (2015). Prediction of order statistics and record values based on ordered ranked set sampling. *Journal of Statistical Computation and Simulation*, 85(1):77–88.