Iranian Journal of Management Studies (IJMS) *http://ijms.ut.ac.ir/*
Vol. 11, No. 2, Spring 2018 *Print ISSN: 2008-7055*
pp. 307-322 *Online ISSN: 2345-3745*
DOI: 10.22059/ijms.2018.236086.672721

Optimal Non-Parametric Prediction Intervals for Order Statistics with Random Sample Size

Elham Basiri[1*], Aylin Pakzad [2]

1. Department of Statistics, Kosar University of Bojnord, Bojnord, Iran
2. Department of Industrial Engineering, Kosar University of Bojnord, Bojnord, Iran

(Received: June 20, 2017; Revised: February 20, 2018; Accepted: March 10, 2018)

Abstract

In many experiments, such as biology and quality control problems, sample size cannot always be considered as a constant value. Therefore, the problem of predicting future data when the sample size is an integer-valued random variable can be an important issue. This paper describes the prediction problem of future order statistics based on upper and lower records. Two different cases for the size of the future sample is considered as fixed and random cases. To do this, we first derive a general formula for the coverage probability of the prediction interval for each case. For the case that the sample size is a random variable, we consider two different distributions for the sample size, such as binomial and Poisson distributions and we study further details. The numerical computations are also given in this paper. Another purpose of this paper is to determine the optimal prediction interval for each case. Finally, the application of the proposed prediction interval is illustrated by analyzing the data in a real-world case study.

Keywords

Prediction interval, random sample size, complete beta function.

* **Corresponding Author,** Email: elham_basiri2000@yahoo.com

Introduction

Let $X=(X_1,\cdots,X_n)$ be a sample of independent and identically distributed (iid) random variables from a distribution with cumulative distribution (CDF) F and probability density function (PDF) f. If $X_{1:n}\le\cdots\le X_{n:n}$ are the order statistics (OSs) from this sample, then the marginal density function of $X_{i:n}, i=1,\cdots,n,$ is given by

$$f_{X_{i:n}}(x)=\frac{1}{\beta(i,n-i+1)}\,f(x)(F(x))^{i-1}\left(\overline{F}(x)\right)^{n-i}, \qquad x\in D_F, \tag{1}$$

where $\overline{F}(x)=1-F(x)$ is the survival function of X-sample, $\beta(.,.)$ denotes the complete beta function.

We refer the reader to David and Nagaraja (2003) and Arnold et al. (2008) and the references therein for more details on the theory and applications of order statistics.

In a sequence of iid random variables as $\{Y_i; i\ge 1\}$ with CDF F and PDF f, an observation Y_j is called an upper (or lower) record value if $Y_j>Y_i$ (or $Y_j<Y_i$) for every $i<j$. Let the first upper and lower record be denoted by $R_1^L=R_1^U=Y_1$, and the r-th upper and lower record be taken as R_r^U and R_r^L (for $r\ge 2$), respectively. The survival function of R_r^U is (see, for example, . ,.Arnord et al, 1998):

$$\overline{F}_{R_r^U}(u)=(\overline{F}(u))\sum_{l=0}^{r-1}\frac{\{-\log(\overline{F}(u))\}^l}{l!}, \qquad u\in D_F. \tag{2}$$

By replacing $\overline{F}$ by F in Equation (2), the survival function of R_r^L can be derived. The theory of records can be used in various topics, including sports fields, meteorology, geophysics, seismology. Interested readers may refer to Arnord etal. (1998) for more details about records.

One of the basic concepts in statistics is the conjecture of the value of an unobserved random variable based on the information obtained from observed events, which is known as the prediction of that

random variable. In many issues, such as time series, regression, quality control, random processes and survival analysis, the prediction problem is used. Also, prediction in various sciences such as meteorology, geology, biology, sociology, geography and economics can be studied. The problem of predicting future data has been studied by many researchers. See, for example, Lawless (1977), Ahsanullah (1980), Hsieh (1997), Raqab and Balakrishnan (2008), Ahmadi and MirMostafaee (2009), Ahmadi et al. (2010), Asgharzadeh and Fallah (2010), Ahmadi and Balakrishnan (2010, 2011), and also Basiri et al. (2016).

In all the articles mentioned above, the size of the sample was a fixed value. In many practical applications, the number of components which are put on the life testing itself, is frequently a random variable. One of the main reasons for this is that in many biological, agricultural and some quality control problems, it may not be possible that the sample size is fixed because some of the observations get lost during the experiment. Assuming the sample size is a random variable, predicting future ordered data has been studied by several authors. See, for example, Soliman (2000), Abd Ellah and Sultan (2005), Sultan and Abd Ellah (2006) and Al-Hussaini and Al-Awadhi (2010). Recently, Basiri and Ahmadi (2015) considered two sample prediction problems for generalized order statistics when the size is random. In this paper, we investigate nonparametric predicting future order statistics based on observed records, assuming the size of the future sample as a random variable.

The rest of this paper is set as follows: In Section 2, a general formula for the coverage probability of prediction intervals for future order statistics based on observed upper and lower records is derived. Two cases for the sample size are assumed, fixed and random. Also, two most used distributions for random sample size are considered. Section 3 is concerned with finding optimal prediction intervals for future random order statistics based on upper and lower records in the both cases of random and fixed samples. In Section 4, a real example is expressed to evaluate the methods outlined in this paper. Finally, a conclusion of the paper is presented in Section 5.

Prediction of Order Statistics

In this section, let $X_{i:N}$ be the i-th order statistic from a future X -sample. Also, let R_j^U and R_j^L, $1 \le j$, the bejupper and lower th-records, .respectivelyBy assuming $N = n_0$ as a fixed value, Ahmadi and Balakrishnan (2010) showed that $(R_r^U, R_s^U), 1 \le r < s,$ is a prediction interval for $X_{i:N}$, when N is a fixed value, with the coverage probability given by

$$\alpha(r,s;i,N) = \phi_1(s;i,N) - \phi_1(r;i,N), \quad (3)$$

Where

$$\phi_1(j;i,N) = \sum_{t=0}^{i-1} \frac{\binom{i-1}{t}(-1)^t}{\beta(i, n_0 - i + 1)(n_0 - i + t + 1)} \left\{ 1 - \frac{1}{(n_0 - i + t + 2)^j} \right\}. \quad (4)$$

Also, based on lower records, they provided a prediction interval as $(R_s^L, R_r^L), 1 \le r < s,$, for $X_{i:N}$ which its coverage probability is

$$\beta(r,s;i,N) = \phi_2(r;i,N) - \phi_2(s;i,N), \quad (5)$$

for

$$\phi_2(j;i,N) = \sum_{t=0}^{n_0 - i} \sum_{k=j}^{\infty} \frac{\binom{n_0 - i}{t}(-1)^t}{\beta(i, n_0 - i + 1)(i + t + 1)^{k+1}}. \quad (6)$$

By considering upper and lower records jointly, Ahmadi and Balakrishnan (2010) showed that $(R_r^L, R_s^U), r, s \ge 1,$ is a prediction interval for $X_{i:N}$, with coverage probability given by

$$\gamma(r,s;i,N) = \phi_1(s;i,N) - \phi_2(r;i,N), \quad (7)$$

When $\phi_1(s;i,N)$ and $\phi_2(r;i,N)$ are defined as in Equations (4) and (6), respectively.

Now, let N be a positive integer-valued random variable. Construction of prediction intervals for future order statistics with random sample size based on upper records is stated in the following theorem.

Theorem 1: Let $X_{i:N}$ be the i-th order statistic from a future X -

sample, when N is a random variable.

Independently, let R_i^U and R_i^L be the i-th observed upper and lower records with the same parent distribution, respectively. Then, $(R_r^U, R_s^U), 1 \le r < s,$ is a prediction interval for $X_{i:N}$ with the coverage probability given by Equation (3) where

$$\phi_1(j;i,N) = \frac{1}{P(N \ge i)} \sum_{n=i}^{\infty} \sum_{t=0}^{i-1} \frac{P(N=n)\binom{i-1}{t}(-1)^t}{\beta(i, n-i+1)(n-i+t+1)} \left\{1 - \frac{1}{(n-i+t+2)^j}\right\}. \quad (8)$$

Proof. First, according to Equation (1) and the results obtained by Raghunandanan and Patil)1972(, the marginal density function of $X_{i:N}$ can be written as

$$f_{X_{i:N}}(x) = \frac{1}{P(N \ge i)} \sum_{n=i}^{\infty} f_{X_{i:n}}(x) P(N=n). \quad (9)$$

By using Equations (2) and (9), we obtain

$$P(R_s^U > X_{i:N}) = \int_{D_F} \overline{F}_{R_s^U}(x) f_{X_{i:N}}(x) dx$$

$$= \frac{1}{P(N \ge i)} \sum_{n=i}^{\infty} \sum_{l=0}^{s-1} \frac{P(N=n)}{\beta(i, n-i+1)}$$

$$\int_{D_F} \frac{(F(x))^{i-1}(\overline{F}(x)))^{n-i+1}}{l!} (-\log(\overline{F}(x)))^l f(x) dx.$$

By setting $y = F(x)$, we find

$$P(R_s^U > X_{i:N}) = \frac{1}{P(N \geq i)} \sum_{n=i}^{\infty} \sum_{l=0}^{s-1} \frac{P(N=n)}{\beta(i, n-i+1)} \int_0^1 \frac{y^{i-1}(1-y)^{n-i+1}}{l!} (-\log(1-y))^l dy$$

$$= \frac{1}{P(N \geq i)} \sum_{n=i}^{\infty} \sum_{l=0}^{s-1} \sum_{t=0}^{i-1} \frac{P(N=n)\binom{i-1}{t}(-1)^t}{\beta(i, n-i+1)} \int_0^{\infty} \frac{e^{-(n-i+t+2)z}}{l!} z^l dz$$

$$= \frac{1}{P(N \geq i)} \sum_{n=i}^{\infty} \sum_{l=0}^{s-1} \sum_{t=0}^{i-1} \frac{P(N=n)\binom{i-1}{t}(-1)^t}{\beta(i, n-i+1)(n-i+t+2)^{l+1}}$$

$$= \frac{1}{P(N \geq i)} \sum_{n=i}^{\infty} \sum_{t=0}^{i-1} \frac{P(N=n)\binom{i-1}{t}(-1)^t}{\beta(i, n-i+1)(n-i+t+1)} \left\{1 - \frac{1}{(n-i+t+2)^s}\right\}$$

$$= \phi_1(s; i, N),$$

Where the second equality is obtained by taking $z = -\log(1-y)$. The required result can be obtained by using the relation $\alpha(r, s; i, N) = P(R_r^U < X_{i:N} < R_s^U) = \phi_1(s; i, N) - \phi_1(r; i, N)$.

Remark2 .Let R_r^L and R_s^L, $1 \leq r < s$, the bert- and s-thlower records, ,respectively. Then $(R_s^L, R_r^L), 1 \leq r < s,$ is a two-sided prediction interval for $X_{i:N}$, and its coverage probability is free of F and is given by Equation (5), where

$$\phi_2(j; i, N) = \frac{1}{P(N \geq i)} \sum_{n=i}^{\infty} \sum_{t=0}^{n-i} \sum_{k=j}^{\infty} \frac{P(N=n)\binom{n-i}{t}(-1)^t}{\beta(i, n-i+1)(i+t+1)^{k+1}}. \qquad (10)$$

Remark3 .Let R_r^L and R_s^U, $1 \leq r, s,$ the berth- lower and s-th upperrecords , ,respectively. Then $(R_r^L, R_s^U), 1 \leq r < s,$ is a two-sided prediction interval for $X_{i:N}$, with coverage probability given in Equation (7), where $\phi_1(s; i, N)$ and $\phi_2(r; i, N)$ are defined as in Equations (8) and (10), respectively.

We now consider some of the most widely used discrete probability distributions for the sample size, N, and we study more details.

Binomial Distribution

Let N be a binomial random variable with parameters M and p, written $B(M, p)$. Then, Relation (8) can be written as:

$$\phi_1(j;i,N)=\frac{1}{\sum_{l=i}^{M}\binom{M}{l}p^l(1-p)^{M-l}}\sum_{n=i}^{M}\sum_{t=0}^{i-1}\frac{\binom{M}{n}p^n(1-p)^{M-n}\binom{i-1}{t}(-1)^t}{\beta(i,n-i+1)(n-i+t+1)}\left\{1-\frac{1}{(n-i+t+2)^j}\right\}.$$

Also, Relation (10) can be re-expressed as:

$$\phi_2(j;i,N)=\frac{1}{\sum_{l=i}^{M}\binom{M}{l}p^l(1-p)^{M-l}}\sum_{n=i}^{M}\sum_{t=0}^{n-i}\sum_{k=j}^{M}\frac{\binom{M}{n}p^n(1-p)^{M-n}\binom{n-i}{t}(-1)^t}{\beta(i,n-i+1)(i+t+1)^{k+1}}.$$

Poisson Distribution

When N has a Poisson distribution with parameter λ, written $P(\lambda)$, then Relation (8) can be changed to:

$$\phi_1(j;i,N)=\frac{1}{\sum_{l=i}^{\infty}\frac{\lambda^l}{l!}}\sum_{n=i}^{\infty}\sum_{t=0}^{i-1}\frac{\lambda^n\binom{i-1}{t}(-1)^t}{n!\,\beta(i,n-i+1)(n-i+t+1)}\left\{1-\frac{1}{(n-i+t+2)^j}\right\}.$$

Also, Relation (10) can be re-expressed as:

$$\phi_2(j;i,N)=\frac{1}{\sum_{l=i}^{\infty}\frac{\lambda^l}{l!}}\sum_{n=i}^{\infty}\sum_{t=0}^{n-i}\sum_{k=j}^{\infty}\frac{\frac{\lambda^n}{n!}\binom{n-i}{t}(-1)^t}{\beta(i,n-i+1)(i+t+1)^{k+1}}.$$

The values of $\alpha(r,s;i,N)$, $\beta(r,s;i,N)$ and $\gamma(r,s;i,N)$ for different choices of i, r, s and different cases for N, are calculated and reported in Table 1. The mathematical package Maple 18 has been used to obtain the numerical computations. From Table 1, we find that the prediction coefficients $\alpha(r,s;i,N)$ and $\beta(r,s;i,N)$ are increasing functions of s but decreasing functions of r, when all other factors are fixed, as we expected. Also, It can be observed that the

prediction coefficient $\gamma(r,s;i,N)$ is an increasing function of s and r, when other values are considered fixed. For predicting lower order statistics considering lower records leads to better results while upper order statistics can be predicted better by considering upper records. Middle order statistics can be predicted better when we consider upper and lower records jointly. Moreover, $\alpha(r,s;i,N)$ is increasing in i while $\beta(r,s;i,N)$ is decreasing in i, when other parameters are fixed.

Table 1. Values of $\alpha(r,s;i,N)$, $\beta(r,s;i,N)$ and $\gamma(r,s;i,N)$ for different distributions of N and some selected values of r, s and i

Distribution of N	i	r \ s	$\alpha(r,s;i,N)$ 8	9	10	$\beta(r,s;i,N)$ 8	9	10	$\gamma(r,s;i,N)$ 1	2	3
N=10		1	0.091	0.091	0.091	0.875	0.891	0.900	0.000	0.184	0.394
	1	2	0.008	0.008	0.008	0.692	0.708	0.716	0.083	0.266	0.477
		3	0.001	0.001	0.001	0.481	0.497	0.506	0.090	0.274	0.485
		1	0.455	0.455	0.455	0.545	0.545	0.545	0.000	0.335	0.480
	5	2	0.144	0.144	0.144	0.211	0.211	0.211	0.311	0.646	0.790
		3	0.037	0.037	0.037	0.066	0.066	0.066	0.418	0.752	0.897
		1	0.875	0.891	0.900	0.091	0.091	0.091	0.000	0.083	0.090
	10	2	0.692	0.708	0.716	0.008	0.008	0.008	0.184	0.266	0.274
		3	0.481	0.497	0.506	0.001	0.001	0.001	0.394	0.477	0.485
	1	1	0.344	0.344	0.344	0.651	0.656	0.658	0.000	0.260	0.420
		2	0.130	0.130	0.130	0.381	0.386	0.388	0.210	0.480	0.640
		3	0.054	0.054	0.054	0.221	0.226	0.228	0.280	0.560	0.720
B(10, 0.2)		1	0.790	0.797	0.801	0.193	0.193	0.193	0.000	0.160	0.190
	5	2	0.535	0.542	0.546	0.036	0.036	0.036	0.260	0.410	0.440
		3	0.324	0.331	0.335	0.007	0.007	0.007	0.470	0.620	0.650
		1	0.956	0.971	0.980	0.091	0.091	0.091	0.000	0.000	0.010
	10	2	0.723	0.738	0.747	0.008	0.008	0.008	0.150	0.240	0.240
		3	0.484	0.499	0.508	0.001	0.001	0.001	0.390	0.480	0.480
	1	1	0.181	0.181	0.181	0.830	0.840	0.844	0.000	0.230	0.440
		2	0.036	0.036	0.036	0.570	0.580	0.584	0.110	0.370	0.590
B(10, 0.5)		3	0.009	0.009	0.009	0.355	0.365	0.369	0.140	0.400	0.620
		1	0.726	0.730	0.732	0.283	0.283	0.283	0.000	0.190	0.250
	5	2	0.433	0.437	0.439	0.072	0.072	0.072	0.280	0.490	0.550
		3	0.234	0.238	0.240	0.012	0.012	0.012	0.480	0.690	0.750
		1	0.900	0.916	0.924	0.091	0.091	0.091	0.000	0.050	0.060
	10	2	0.685	0.701	0.709	0.008	0.008	0.008	0.180	0.270	0.280
		3	0.474	0.490	0.498	0.001	0.001	0.001	0.400	0.480	0.490
	1	1	0.129	0.129	0.129	0.883	0.896	0.903	0.000	0.200	0.430
		2	0.017	0.017	0.017	0.643	0.656	0.663	0.070	0.310	0.540
		3	0.002	0.002	0.002	0.418	0.431	0.438	0.090	0.330	0.550
B(10, 0.8)	5	1	0.637	0.638	0.638	0.220	0.220	0.220	0.000	0.260	0.330
		2	0.311	0.312	0.312	0.105	0.105	0.105	0.470	0.580	0.660
		3	0.138	0.139	0.139	0.028	0.028	0.028	0.640	0.760	0.830
	10	1	0.766	0.781	0.790	0.091	0.091	0.091	0.000	0.190	0.200
		2	0.654	0.669	0.678	0.008	0.008	0.008	0.220	0.300	0.310

Distribution of N	i	r \ s	$\alpha(r,s;i,N)$ 8	9	10	$\beta(r,s;i,N)$ 8	9	10	$\gamma(r,s;i,N)$ 1	2	3
		3	0.462	0.477	0.486	0.001	0.001	0.001	0.410	0.500	0.500
	1	1	0.343	0.343	0.344	0.651	0.656	0.658	0.000	0.260	0.430
		2	0.130	0.130	0.131	0.386	0.391	0.393	0.210	0.470	0.650
		3	0.054	0.054	0.055	0.214	0.219	0.221	0.280	0.550	0.720
P(2)	5	1	0.783	0.789	0.793	0.215	0.215	0.215	0.000	0.160	0.200
		2	0.509	0.515	0.519	0.049	0.049	0.049	0.260	0.430	0.470
		3	0.298	0.304	0.308	0.009	0.009	0.009	0.470	0.640	0.680
	10	1	0.874	0.890	0.898	0.091	0.091	0.091	0.000	0.080	0.090
		2	0.692	0.708	0.716	0.008	0.008	0.008	0.180	0.260	0.270
		3	0.485	0.501	0.509	0.001	0.001	0.001	0.390	0.470	0.480
	1	1	0.195	0.195	0.195	0.772	0.781	0.785	0.000	0.240	0.450
		2	0.046	0.046	0.046	0.549	0.558	0.562	0.160	0.390	0.600
		3	0.011	0.011	0.011	0.339	0.348	0.352	0.200	0.420	0.630
P(5)	5	1	0.681	0.684	0.685	0.197	0.197	0.197	0.000	0.220	0.290
		2	0.390	0.393	0.394	0.080	0.080	0.080	0.400	0.520	0.580
		3	0.202	0.205	0.206	0.011	0.011	0.011	0.590	0.700	0.770
	10	1	0.876	0.891	0.900	0.091	0.091	0.091	0.000	0.080	0.090
		2	0.693	0.708	0.717	0.008	0.008	0.008	0.180	0.270	0.270
		3	0.486	0.501	0.510	0.001	0.001	0.001	0.390	0.470	0.480
	1	1	0.135	0.135	0.135	0.866	0.877	0.883	0.000	0.200	0.430
		2	0.021	0.021	0.021	0.636	0.647	0.653	0.090	0.320	0.540
		3	0.003	0.003	0.003	0.411	0.422	0.428	0.100	0.340	0.560
P(8)	5	1	0.594	0.596	0.596	0.519	0.519	0.519	0.000	0.270	0.370
		2	0.299	0.301	0.301	0.135	0.135	0.135	0.180	0.560	0.660
		3	0.129	0.131	0.131	0.035	0.035	0.035	0.350	0.730	0.830
	10	1	0.876	0.891	0.900	0.091	0.091	0.091	0.000	0.080	0.090
		2	0.693	0.708	0.717	0.008	0.008	0.008	0.180	0.270	0.270
		3	0.486	0.501	0.510	0.001	0.001	0.001	0.390	0.470	0.480

Optimal Prediction Intervals for Random Order Statistics

Obviously, for given α_0, a prediction interval, as $(R_r^U, R_s^U), 1 \le r < s,$ exists if and only if $\alpha(1, M_0; i, N) \ge \alpha_0$, where M_0 is the number of observed upper records. From Equations (3) and (8), this condition is equivalent to:

$$\frac{1}{P(N \ge i)} \sum_{n=i}^{\infty} \sum_{t=0}^{i-1} \frac{P(N=n)\binom{i-1}{t}(-1)^t}{\beta(i, n-i+1)(n-i+t+1)} \left\{ \frac{1}{(n-i+t+2)^1} - \frac{1}{(n-i+t+2)^{M_0}} \right\} \ge \alpha_0. \quad (11)$$

Under Condition (11), if the values i and the coverage level α_0 as well as the distribution of N are all given. Since various indices can be used for constructing the prediction intervals for a pre-fixed level, choosing

the best values for these indices is an important issue. It seems reasonable to choose these indices so that the prediction interval has the shortest length among all the prediction intervals in the same coverage level, which is called the optimal prediction interval. Therefore, in order to determine the optimal prediction interval, we must minimize the mean length of the prediction interval, $E\left(R_s^U - R_r^U\right)$ as an optimization criterion. Since the average length of the interval depends on the parent distribution, we minimize the difference between the predicted distance indices $s-r$ (see, for example, Balakrishnan et al., 2013) as an equivalent approach. Towards this end, first, we take:

$$A(l;i,N)=\frac{1}{P(N\geq i)}\sum_{n=i}^{\infty}\sum_{t=0}^{i-1}\frac{P(N=n)\binom{i-1}{t}(-1)^t}{\beta(i,n-i+1)(n-i+t+2)^{l+1}}.$$

Then, from Equations (3) and (8), we get $\alpha(r,s;i,N)=\sum_{l=r}^{s-1}A(l;i,N)$.

The algorithm below describes the determination of the optimal indices for constructing prediction intervals, which are represented by the symbol (r_{opt},s_{opt}).

Algorithm 1: Suppose the distribution of N is known and the values i and α_0 are all given. Moreover, let Condition (11) hold. In addition, let M_0 be the number of observed upper records, then, the procedure depicted in Figure 1 gives an optimal prediction interval for $X_{i:N}$.

Similar procedures can be considered for finding optimal indices of lower records.

By using Algorithm 1, the optimal indices (r_{opt},s_{opt}) are specified and are presented in Table 2 for $\alpha_0=0.80$ and some selected choices of i when N has one of the three mentioned discrete probability distributions as in Section 2. The results have been obtained by using Maple 18. The optimization command has been utilized for doing Algorithm 1. In Table 2, dash (-) shows that there is no prediction interval at that level.

From Table 2 it can be observed that optimal prediction intervals obtained for different distributions are quite similar for most cases. So, the optimal prediction intervals are relatively stable. Optimal prediction intervals for upper order statistics can be derived by using upper records while for lower order statistics based on lower records. Using upper and lower records jointly can be appropriate for constructing optimal prediction intervals for both upper and lower order statistics.

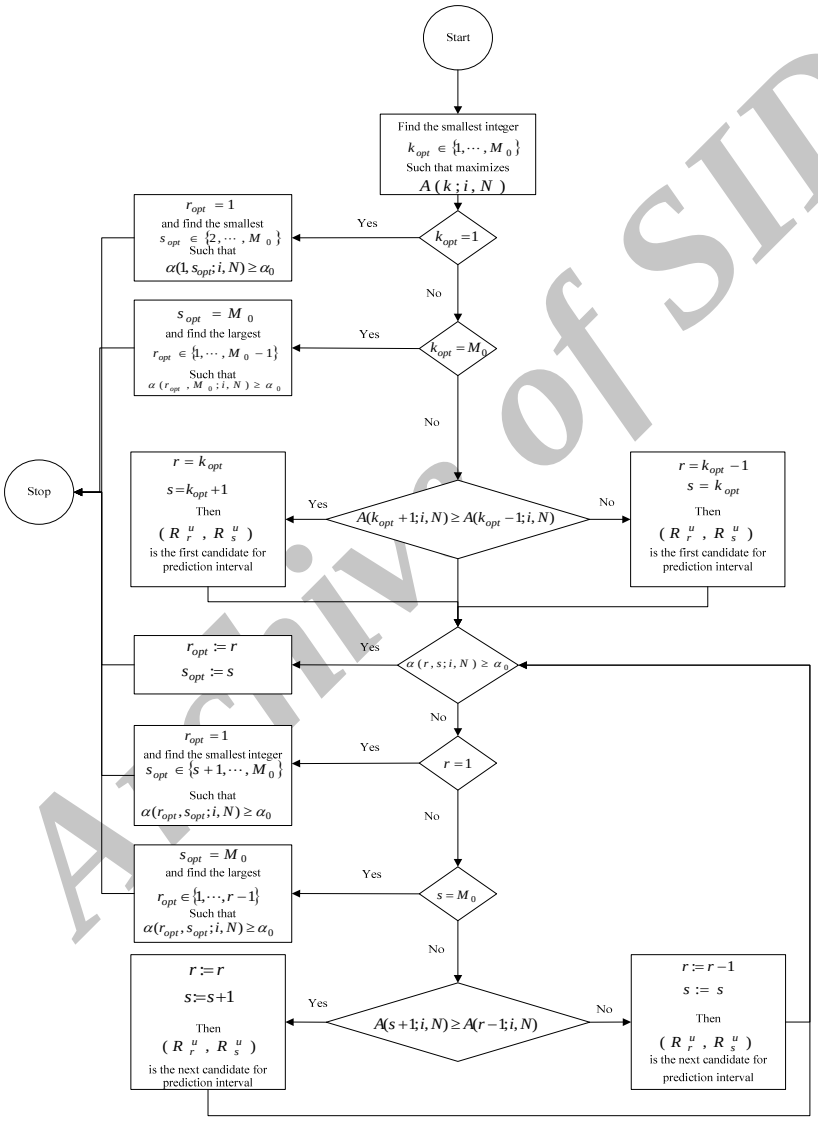

Figure 1. The procedure for finding optimal indices for prediction intervals

Table 2. Values of (r_{opt}, s_{opt}) for $\alpha_0 = 0.80$ and some selected values of i and different distributions for N

Distribution of N	$(R^U_{r_{opt}}, R^U_{s_{opt}})$			$(R^L_{s_{opt}}, R^L_{r_{opt}})$			$(R^L_{r_{opt}}, R^U_{s_{opt}})$		
	i=1	i=5	i=10	i=1	i=5	i=10	i=1	i=5	i=10
N=10	-	-	(1, 7)	(1,7)	-	-	(6,2),(7,1)	(4,2),(3,3)	(1,7),(2,6)
B(10, 0.2)	-	-	(1, 7)	(1,9)	-	-	(5,2),(4,3)	(3,4),(2,5)	(1,6)
B(10, 0.5)	-	-	(1, 7)	(1,6)	-	-	(5,2)	(2,4)	(1,7),(2,6)
B(10, 0.8)	-	-	(1, 7)	(1,7)	-	-	(5,2)	(3,3),(2,4)	(1,6)
P(2)	-	-	(1, 6)	(1,8)	-	-	(5,2),(4,3)	(3,4),(2,5)	(1,6)
P(5)	-	-	(1, 8)	(1,7)	-	-	(5,2)	(2,4)	(1,6)
P(8)	-	-	(1, 7)	(1,6)	-	-	(5,2)	(3,3),(2,4)	(1,6)

Real Data Example

In order to examine the methods outlined in this paper, we consider Sajedi Refining and Packing Raisins Factory as a real case study. The factory was established in 1992, in Quchan, Razavi Khorasan Province. Along with the advancement of technology in this industry, the company tries to improve the quality of products as much as possible. Estimating the number of defective items produced by a machine is an important problem in statistical process control. In a lot of economic sampling plans, the solution of the optimization problem is strongly dependent on the estimation of the fraction of defective items. The biased estimation may lead to improper choice in optimization and, consequently, huge economic penalties (Dasgupta & Mandal, 2008). Estimating the number of defective items has been studied by a few researchers. In this paper, we consider the real data set representing the number of damaged raisins in every 100gr samples of raisins recorded in 2015. By counting the number of inspected raisins in every 100gr samples, we could acquire the fraction of defective raisins is a continuous random variable and takes a value between zero and one. Upper and lower records extracted from these data are reported in Table 3. Here, using the results obtained in Section 3 and Table 2, the optimal prediction intervals are determined and are presented in Table 4. The three mentioned distributions mentioned in Section 2 are considered for the size of the future sample, when $\alpha_0 = 0.80$.

Table 3. The upper and lower records extracted from the fraction of defective raisins

j	1	2	3	4	5	6	7	8	9
R_j^U	0.028	0.031	0.035	0.039	0.041	0.042	0.048	0.056	
R_j^L	0.028	0.026	0.017	0.013	0.010	0.008	0.006	0.003	0.000

Table 4. 80% optimal prediction intervals for some selected values of i and different distributions for N

	$(R_{r_{opt}}^U, R_{s_{opt}}^U)$			$(R_{s_{opt}}^L, R_{r_{opt}}^L)$			$(R_{r_{opt}}^L, R_{s_{opt}}^U)$		
Distribution of N	i=1	i=5	i=10	i=1	i=5	i=10	i=1	i=5	i=10
N=10	-	-	(0.028,0.048)	(0.006,0.028)	-	-	(0.008,0.031), (0.006,0.028)	(0.013,0.031), (0.017,0.035)	(0.028,0.048), (0.026,0.042)
B(10, 0.2)	-	-	(0.028,0.048)	(0.000,0.028)	-	-	(0.010,0.031), (0.013,0.035)	(0.017,0.039), (0.026,0.041)	(0.028,0.042)
B(10, 0.5)	-	-	(0.028,0.048)	(0.008,0.028)	-	-	(0.010,0.031)	(0.026,0.039)	(0.028,0.048), (0.026,0.042)
B(10, 0.8)	-	-	(0.028,0.048)	(0.006,0.028)	-	-	(0.010,0.031)	(0.017,0.035), (0.026,0.039)	(0.028,0.042)
P(2)	-	-	(0.028,0.042)	(0.003,0.028)	-	-	(0.010,0.031), (0.013,0.035)	(0.017,0.039), (0.026,0.041)	(0.028,0.042)
P(5)	-	-	(0.028,0.056)	(0.006,0.028)	-	-	(0.010,0.031)	(0.026,0.039)	(0.028,0.042)
P(8)	-	-	(0.028,0.048)	(0.008,0.028)	-	-	(0.010,0.031)	(0.017,0.035), (0.026,0.039)	(0.028,0.042)

Concluding Remarks

In many experiments, such as biology and quality control, sample size cannot always be considered constant. Therefore, the problem of predicting future data when the sample size is a random variable can be an important issue. In this paper, we first consider the prediction of future order statistics while the sample size is a random variable. Three different distributions, such as degenerate, binomial and Poisson distributions were considered for the size of the future sample. Then, taking into account the optimization criterion for the shortest interval length, which can be obtained by minimizing the mean length of the prediction interval, we determined the optimal prediction intervals in each case, and then compared the results. Finally, it can be concluded that the results are similar for different distributions of future sample size. In other words, the prediction intervals are not very affected by the distribution of the sample size and are almost stable.

Acknowledgements

The author would like to thank the referees and the associate editor for their useful comments and constructive suggestions on the original version of this manuscript, which improved the presentation of the paper considerably.

References

Abd Ellah, A. H., & Sultan, K. S. (2005). Exact Bayesian prediction of exponential lifetime based on fixed and random sample sizes. *Quality Technology & Quantitative Management*, *2*(2), 161-175.

Ahmadi, J., & Balakrishnan, N. (2010). Prediction of order statistics and record values from two independent sequences. *Statistics*, *44*(4), 417-430.

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

Ahmadi, J., & MirMostafaee, S. M. T. K. (2009). Prediction intervals for future records and order statistics coming from two parameter exponential distribution. *Statistics & Probability Letters*, *79*(7), 977-983.

Ahmadi, J., MirMostafaee, S. M. T. K., & Balakrishnan, N. (2010). Nonparametric prediction intervals for future record intervals based on order statistics. *Statistics & probability letters*, *80*(21), 1663-1672.

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.

Al-Hussaini, E. K., & Al-Awadhi, F. (2010). Bayes two-sample prediction of generalized order statistics with fixed and random sample size. *Journal of Statistical Computation and Simulation*, *80*(1), 13-28.

Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (1998). *Records*. New York, NY: John Wiley & Sons.

Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (2008). *A first course in order statistics*. Society for Industrial and Applied Mathematics.

Asgharzadeh, A., & Fallah, A. (2010). Estimation and prediction for exponentiated family of distributions based on records. *Communications in Statistics—Theory and Methods*, *40*(1), 68-83.

Balakrishnan, N., Beutner, E. & Cramer, E. (2013). Computational aspects of statistical intervals based on two Type-II censored samples. *Computational Statistics, 28*, 893-917.

Basiri, E., & Ahmadi, J. (2015). Prediction intervals for generalized-order statistics with random sample size. *Journal of Statistical Computation and Simulation*, *85*(9), 1725-1741.

Basiri, E., Ahmadi, J., & Raqab, M. Z. (2016). Comparison among non-parametric prediction intervals of order statistics. *Communications in Statistics-Theory and Methods*, *45*(9), 2699-2713.

Dasgupta, T., & Mandal, A. (2008). Estimation of process parameters to determine the optimum diagnosis interval for control of defective items. *Technometrics, 50*(2), 167-181.

David, H. A. & Nagaraja, H. N. (2003). *Order statistics* (3rd ed.). Hoboken, New Jersey: John Wiley & Sons.

Hsieh, H. K. (1997). Prediction intervals for Weibull order statistics. *Statistica Sinica*, 1039-1051.

Lawless, J. F. (1977). Prediction intervals for the two parameter exponential distribution. *Technometrics*, *19*(4), 469-472.

Raghunandanan, K., & Patil, S. A. (1972). On order statistics for random sample size. *Statistica Neerlandica*, *26*(4), 121-126.

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

Soliman, A. A. (2000). Bayes prediction in a Pareto lifetime model with random sample size. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *49*(1), 51-62.

Sultan, K. S., & Abd Ellah, A. H. (2006). Exact prediction intervals for exponential lifetime based on random sample size. *International Journal of Computer Mathematics*, *83*(12), 867-878.