Pak. J. Statist.
2023 Vol. 39(4), 485-500

NONPARAMETRIC STATISTICAL METHOD FOR PREDICTION IN CASE OF DATA INCLUDING DOUBLE-CENSORED OBSERVATIONS

Asamh Saleh M. Al Luhayb
Department of Mathematics, College of Science, Qassim University
P.O. Box 6644, Buraydah 51452, Saudi Arabia.
Email: a.alluhayb@qu.edu.sa

ABSTRACT

This paper introduces a new nonparametric statistical method for prediction in case of data containing right-censored observations and left-censored observations simultaneously. The method can be considered as a new version of Hill's $A_{(n)}$ assumption for double-censored data. Two bounds are derived to predict the survival function for one future observation X_{n+1} and compare those bounds with two nonparametric maximum likelihood estimators of the survival function. Two interesting features are provided based on the proposed method. The first one is the detailed graphical presentation of the effects of right- and left-censoring. The second feature is that the lower and upper survival functions can be derived.

KEYWORDS

Double-censored data, Hill's $A_{(n)}$ assumption, nonparametric predictive inference, prediction, right-censoring- $A_{(n)}$ assumption, statistical model.

1. INTRODUCTION

This paper introduces a new nonparametric statistical method for prediction using the past data that contain right-censored observations and left-censored observations simultaneously, where this kind of data is called double-censored data in the literature. The method is proposed to learn about one future observation based on the original sample with few mathematical assumptions. For real-valued data, Hill (1968) and Hill (1988) presented the $A_{(n)}$ assumption for prediction when there is few knowledge about the underlying distribution, and this assumption provides certain probabilities for one future observation based on the past observations. For right-censored data, Berliner and Hill (1988), Coolen and Yan (2004) generalized the $A_{(n)}$ assumption, and the methods are considered for survival analysis. In this paper, a regarding method is developed by using the information of censoring observations, and it is referred to by 'double-censoring $A_{(n)}$, (dc- $A_{(n)}$).

This paper is organized as follows. In Section 2, a brief overview of the $A_{(n)}$ assumption is given for real-valued data. Section 3 presents the generalizations of $A_{(n)}$ assumption for right-censored data. Section 4 presents the double-censoring $A_{(n)}$

assumption along with the corresponding justification. In Section 5, we present bounds for probabilities and survival functions based on the dc- $A_{(n)}$ assumption. In Section 6, the new proposed method is compared to two nonparametric maximum likelihood estimators, which are based on the expectation-maximization algorithm and the self-consistency algorithm. In this paper, we assume that ties can occur with probability zero to make notation simple, but Section 7 briefly discusses the way that ties can be dealt with. The final section provides some concluding remarks and discusses some future related research.

2. $A_{(\mathrm{n})}$ ASSUMPTION FOR DATA INCLUDING EVENT OBSERVATIONS ONLY

Hill (1968) introduced the $A_{(n)}$ assumption for prediction when there is no prior knowledge about the underlying distribution. The data support is partitioned into $n+1$ intervals using the observed data points and a probability $1/(n+1)$ is assigned to each created interval. Let $X_1, X_2, ..., X_n$ be continuous and exchangeable real-valued random quantities, and let $x_1, x_2, ..., x_n$ be the corresponding observed data points. Furthermore, let $x_{(1)} < x_{(2)} < ... < x_{(n)}$ be the ordered observations and let $x_{(0)} = -\infty$ (or $x_{(0)} = 0$ for positive random quantities) and $x_{(n+1)} = \infty$. For one future observation X_{n+1}, the $A_{(n)}$ assumption is

$$P\left(X_{n+1} \in \left(x_{(i)}, x_{(i+1)}\right)\right) = \frac{1}{n+1} \tag{1}$$

It should be noted that the $A_{(n)}$ assumption is suitable to provide bounds for probabilities, known as imprecise probabilities, by using the theorem of probability proposed by De Finetti (1974), but it is not good to derive precise probabilities for many functions of interest. The bounds created for probabilities can lead to valuable information in the case of uncertainty of event or in the case of indeterminacy caused by restricted information Walley (1991) and Weichselberger (2001). Augustin and Coolen (2004) derived strong consistency properties for nonparametric predictive inference in interval probability theory based on the $A_{(n)}$ assumption, and multiple examples have been presented by Coolen (1998), Coolen and Coolen-Schrijner (2000), Coolen et al. (2002) and Coolen and Van der Laan (2001).

The $A_{(n)}$ assumption has been discussed for multiple statistical inferences in the literature, see e.g. De Finetti (1974) and Hill (1988) for more detailed presentation, and the assumption has been contributed with multiple important related work, see e.g. Dempster (1963), Lane and Sudderth (1984) for detailed information. For right-censored data, Berliner and Hill (1988) and Coolen and Yan (2004) proposed two versions of $A_{(n)}$ and the proposed assumptions will be presented in the next section.

3. $A_{(n)}$ ASSUMPTION FOR DATA INCLUDING RIGHT-CENSORED OBSERVATIONS

The $A_{(n)}$ assumption for real-valued data is introduced as certain predictive probabilities are assigned to open intervals created by the observed data points with no further assumptions or restrictions on the spread of probabilities within the intervals. In 1988, the $A_{(n)}$ assumption is generalized for data containing right-censored observations by Berliner and Hill (1988). They use the same technique and provide a partial probability distribution via certain values for next future observation X_{n+1}, and the probability masses are assigned to open intervals with no more constraints on the spread of the probability mass within each interval. Let $X_1, X_2, ..., X_n$ be exchangeable positive random quantities, and there are u event observations, and v right-censored observations. The event and right-censored observations are ordered as $0 < t_{(1)} < t_{(2)} < ... < t_{(u)}$, where $0 \le u \le n$, and $0 < c_{(1)} < c_{(2)} < ... < c_{(v)}$, where $v = n - u$, respectively. They assign a certain probability for one future observation X_{n+1} to be fallen in any two ordered event observations $\left(t_{(u)}, t_{(u+1)}\right)$ by the following formula

$$P_{(u+1)} = \left(1-\lambda_{(0)}\right) \times ... \times \left(1-\lambda_{(u)}\right) \times \lambda_{(u+1)} \tag{2}$$

where $\lambda_{(u)} = \frac{1}{n+1-u-C_{(u)}}, C(u) = \# \left\{censors < t_{(u+1)}\right\}$ and $P_{(0)} = \lambda_{(0)}$.

This proposed assumption was the first attempt to generalize $A_{(n)}$ assumption for right-censored data, and it was a good start to deal with the case of data including right-censored observations. Coolen and Yan (2004) provided an alternative approach, named by the right-censoring $A_{(n)}$ assumption, which is nicer because it does not neglect the censored observations as Berliner and Hill does to create the intervals partitioned the sample space.

The right-censoring $A_{(n)}$ assumption, rc-$A_{(n)}$, provides a partial probability distribution for one future observation, and it is specified via *M*-function values (Coolen and Yan, 2004). The random quantities $X_1, X_2, ..., X_n$ are assumed to be exchangeable, nonnegative and real-valued, and there are u event observations, and v right-censoring observations. The event and right-censoring observations are ordered as $0 < t_{(1)} < t_{(2)} < ... < t_{(u)}$, where $0 \le u \le n$, and $0 < c_{(1)} < c_{(2)} < ... < c_{(v)}$, where $v = n - u$, respectively. Let the interval $I_i = \left(t_{(i)}, t_{(i+1)}\right)$, for $0 \le i \le u$, they ordered the right-censoring times in each interval I_i by $c_1^i < c_2^i < \cdots < c_{li}^i$ where l_i is the number of censors in I_i. Finally, they formed the open intervals as $\left(t_{(i)}, t_{(i+1)}\right)$ and $\left(c_k^i, t_{(i+1)}\right)$ where $1 \le k \le l_i$.

Definition (*M*-function). The *M*-function value is a probability partially specified to each interval $\left(t_{(i)},\, t_{(i+1)}\right)$ or $\left(c_k^i, t_{(i+1)}\right)$ where $1 \le k \le l_i$ and $0 \le i \le u$, so that the next future observation X_{n+1} falls in an interval with a probability M -function value (Coolen and Yan, 2004).

Definition (*rc*- $A_{(\mathrm{n})}$). The *rc*- $A_{(\mathrm{n})}$ assumption is that the probability distribution for one positive random quantity X_{n+1} based on data including u event observations and v censored observations is partially assigned through M -function values as follows (Coolen and Yan, 2004):

$$M_{X_{n+1}}\left(t_{(i)}, t_{(i+1)}\right) = \frac{1}{n+1} \prod_{\{r: c_{(r)} < t_{(i)}\}} \frac{\tilde{n}_{c_{(r)}} + 1}{\tilde{n}_{c_{(r)}}} \tag{3}$$

$$M_{X_{n+1}}\left(c_k^i, t_{(i+1)}\right) = \frac{1}{(n+1)\tilde{n}_{c_k^i}} \prod_{\{r: c_{(r)} < c_k^i\}} \frac{\tilde{n}_{c_{(r)}} + 1}{\tilde{n}_{c_{(r)}}} \tag{4}$$

where $\tilde{n}_{c_{(r)}}$ is the number of observations not experiencing the event of study just before time $c_{(r)}$ plus one. $t_{(0)} = 0$ and $t_{(u+1)} = +\infty$. $M_{X_{n+1}}\left(t_{(i)}, t_{(i+1)}\right)$ and $M_{X_{n+1}}\left(c_k^i, t_{(i+1)}\right)$ mean that the random quantity X_{n+1} falls in $\left(t_{(i)}, t_{(i+1)}\right)$ and $\left(c_k^i, t_{(i+1)}\right)$, respectively, with a probability *M*-function value.

From studying the *rc*- $A_{(n)}$ assumption, two interesting notes have been experienced, and it is good to list them to have a wider picture for the assumption. First, Coolen and Yan (2004) presented a new version of the $A_{(n)}$ assumption for data containing right-censored observations by using the same technique of $A_{(n)}$ assumption, but the intervals are overlapped due to the censored observations. They provided a partial probability distribution via the *M*-function values for a random quantity X_{n+1}, and the probability masses are assigned to open intervals with no more constraints on the spread of the probability mass within each interval. The probability mass specified to such interval (a, b) is referred to by $M_{X_{n+1}}(a, b)$ and interpreted by *M*-function value for X_{n+1} on (a, b). Secondly, the *M*-function values are limited between 0 and 1, and the values have to sum up to one over all intervals created (Coolen and Yan, 2004).

4. $A_{(n)}$ ASSUMPTION FOR DATA INCLUDING DOUBLE-CENSORED OBSERVATIONS

This section provides the double-censoring $A_{(n)}$ assumption, dc- $A_{(n)}$ along with its justification. For the justification, we will include a detailed information of the exact nature of the non-informative censoring assumption implicit in dc- $A_{(n)}$, and the justification will be provided in two stages. First, the influence of double-censored observations on the $A_{(n)}$ assumption will be presented, with no further assumptions added. In the second stage, further post-data and pre-data assumptions, which are strongly related to $A_{(n)}$, are considered to derive partially specified predictive probability distributions for the random quantities that were double-censored. The further post-data assumption will be for right-censored data and the further pre-data assumption will be for left-censored data.

Based on the $A_{(n)}$ assumption, a partially specified probability distribution for X_{n+1} is provided by using the observed data points. As first stage in our justification of dc-$A_{(n)}$, the influence of double-censored data on the partially specified probability distribution for X_{n+1} is considered without further constraints, and this leads to generalize $A_{(n)}$ assumption for double-censored data. This generalization will be referred to by $\tilde{A}_{(n)}$, and the definition is justified as follows:

Definition $\left(\tilde{A}_{(n)}\right)$. The $\tilde{A}_{(n)}$ assumption is that the probability distribution for one positive random quantity X_{n+1} based on data including u event observations, $t_{(1)} < t_{(2)} < \ldots < t_{(u)}$, v right-censored observations, $rc_{(1)} < rc_{(2)} < \ldots < rc_{(v)}$, and $k = n - (u + v)$ left-censored observations, $lc_{(1)} < lc_{(2)} < \ldots < lc_{(k)}$, is partially specified through certain probabilities as follows

$$
\begin{aligned}
&P\left(X_{n+1} \in \left(t_{(i)}, t_{(i+1)}\right)\right) = \frac{1}{n+1} \\
&P\left(X_{n+1} \in \left(rc_{(j)}, +\infty\right)\right) = \frac{1}{n+1} \\
&P\left(X_{n+1} \in \left(0, lc_{(w)}\right)\right) = \frac{1}{n+1}
\end{aligned}
\tag{5}
$$

where $i = 0,1,2,\ldots,u, j = 1,2,\ldots,v$ and $w = 1,2,\ldots,k$, with $t_{(0)} = 0$ and $t_{(u+1)} = +\infty$.

The justification of $\tilde{A}_{(n)}$ assumption, in relation to the $A_{(n)}$ assumption, is as follows. The intervals in the form of $\left(t_{(i)}, t_{(i+1)}\right)$ are each assigned a minimum probability mass of $\frac{1}{n+1}$, by the $A_{(n)}$ assumption. If we consider one such an interval, then the total mass in it could be more than $\frac{1}{n+1}$ due to the presence of double-censored observations. Any additional probability mass due to such double-censored observations does not have to be restricted to lie within this interval, without any assumptions, leading us to the probabilities on intervals from a right-censored observation to infinity and from 0 to a left-censored observation. For the case of having a right-censored observation in $\left(t_{(i)}, t_{(i+1)}\right)$, then the unobserved data point t_{rc} corresponding to that right-censored observation could occur in $\left(t_{(i)}, t_{(i+1)}\right)$. This leads to assign a probability $\frac{1}{n+1}$ to the interval $\left(t_{(i)}, t_{rc}\right)$ and another probability $\frac{1}{n+1}$ to the interval $\left(t_{rc}, t_{(i+1)}\right)$. For the case of having a left-censored observation in $\left(t_{(i)}, t_{(i+1)}\right)$, then the unobserved data point t_{lc} corresponding to that left-censored observation could occur in $\left(t_{(i)}, t_{(i+1)}\right)$. This leads to assign a probability $\frac{1}{n+1}$ to the interval $\left(t_{(i)}, t_{lc}\right)$ and another probability $\frac{1}{n+1}$ to the interval $\left(t_{lc}, t_{(i+1)}\right)$. However, due to the lack of information about the true times of double-censored observations, we cannot assign a probability $\frac{1}{n+1}$ to the sub-intervals. Therefore, the probability for X_{n+1} to be fallen in $\left(t_{(i)}, t_{(i+1)}\right)$ is set equal to $\frac{1}{n+1}$.

The only information known about a right-censored observation $rc_{(j)}$ is that the corresponding event would occur after time $rc_{(j)}$. This implies that if this time were observed, then one of the intervals in the form of $\left(t_{(i)}, t_{(i+1)}\right)$ would be split to two intervals and each interval has a probability $\frac{1}{n+1}$ based on the $A_{(n)}$ assumption. However, because it is only known that the corresponding event would exceed $rc_{(j)}$, the only statement about this probability mass $\frac{1}{n+1}$ for X_{n+1} that can be justified, with no more constraints, is that it will fall in $(rc_{(j)}, \infty)$, hence $P(X_{n+1} \in (rc_{(j)}, \infty)) = \frac{1}{n+1}$.

For a left-censored time $lc_{(w)}$, we know that the corresponding event occurred before time $lc_{(w)}$. From this information, the event is occurred in one of the intervals $\left(t_{(i)}, t_{(i+1)}\right)$, and this leads to split the interval $\left(t_{(i)}, t_{(i+1)}\right)$ to two intervals and each one has a probability $\frac{1}{n+1}$ based on the $A_{(n)}$ assumption. But, because it is only known that the corresponding event occurred before $lc_{(w)}$, the only statement about this probability mass $\frac{1}{n+1}$ for X_{n+1} that can be justified, with no more constraints, is that it will fall in $(0, lc_{(w)})$, hence $P\left(X_{n+1} \in (0, lc_{(w)})\right) = \frac{1}{n+1}$.

To link the first stage to the second stage of the justification, each probability assigned to interval $(rc_{(j)}, \infty)$ will be uniformly distributed to the event intervals $\left(t_{(i)}, t_{(i+1)}\right)$ occurred after $rc_{(j)}$ and the interval $(rc_{(j)}, t_{rc_{(j)}})$, where $t_{rc_{(j)}}$ is the first event time greater than $rc_{(j)}$. For the left-censored observations, each probability assigned to interval $(0, lc_{(w)})$ will be uniformly distributed to the event intervals $\left(t_{(i)}, t_{(i+1)}\right)$ occurred before $lc_{(w)}$ and the interval $\left(t_{lc_{(w)}}, lc_{(w)}\right)$, where $t_{lc_{(w)}}$ is the first event time less than $lc_{(w)}$. These assumptions lead to have the following functions

$$P\left(X_{n+1} \in \left(t_{(i)}, t_{(i+1)}\right)\right) = \frac{1}{n+1} + \sum_{j=1}^{v} \frac{I(rc_{(j)} < t_{(i)})}{(n+1)\left(\#\{rc_{(j)} < t_{(i)}\} + 1\right)} + \sum_{w=1}^{k} \frac{I\left(lc_{(w)} > t_{(i+1)}\right)}{(n+1)\left(\#\left\{lc_{(w)} > t_{(i+1)}\right\} + 1\right)} \tag{6}$$

$$P\left(X_{n+1} \in \left(rc_{(j)}, t_{rc_{(j)}}\right)\right) = \frac{1}{(n+1)\left(\#\{rc_{(j)} < t_{(i)}\} + 1\right)}$$

$$P\left(X_{n+1} \in \left(t_{lc_{(w)}}, lc_{(w)}\right)\right) = \frac{1}{(n+1)\left(\#\{lc_{(w)} > t_{(i+1)}\} + 1\right)}$$

where $i = 0,1,2,\ldots,u, j = 1,2,\ldots,v$ and $w = 1,2,\ldots,k$, with $t_{(0)} = 0$ and $t_{(u+1)} = \infty$. I(.) is the indicator function, $t_{rc_{(j)}}$ is the first event time greater than $rc_{(j)}$ and $t_{lc_{(w)}}$ is the first event time less than $lc_{(w)}$.

It should be noted that if data does not include censored observations, the double-censoring $A_{(n)}$ assumption will return to the $A_{(n)}$ assumption. They will be identical.

5. IMPRECISE PROBABILITIES BASED ON dc- $A_{(n)}$

The use of the partially specified predictive probability distribution for the next future observation X_{n+1} is introduced in this section, through the probabilities as given by the double-censoring $A_{(n)}$ assumption. Such inferences have a predictive nature in terms of the next future observation X_{n+1} along similar lines as nonparametric predictive inferences based on the $A_{(n)}$ assumption for real-valued data [Augustin and Coolen (2004), Coolen (1998), Coolen and Coolen-Schrijner (2000), Coolen et al. (2002) and Coolen and Van der Laan (2001)] and based on the right-censoring $A_{(n)}$ assumption for right-censored data [Ahmadini and Coolen (2018), Ashleik (2018), Coolen et al. (2021), Siddiqa et al. (2021) and Tee et al. (2019)]. For many events of interest regarding to X_{n+1}, the probability values proposed by equations (6) allow bounds for probabilities to be derived analogously as for $A_{(n)}$-based inference (Augustin and Coolen, 2004), where the maximum lower bound is referred to by the "lower probability", and the minimum upper bound is referred to by the "upper probability". These bounds follow the terminology proposed in the theory of imprecise probabilities [Walley (1991) and Weichselberger (2001)].

Based on data including u event observations, v right-censored observations and $k = n - (u + v)$ left-censored observations, the double-censoring $A_{(n)}$ assumption provides a partially specified predictive probability distribution for one future observation X_{n+1} via the probabilities assigned to the intervals $\left(t_{(i)}, t_{(i+1)}\right)$, $\left(rc_{(j)}, t_{rc_{(j)}},\right)$ and $\left(t_{lc_{(w)}}, lc_{(w)}\right)$, where $i = 1,2,\ldots,u, j = 1,2,\ldots,v$ and $w = 1,2,\ldots,k$, with $t_{(0)} = 0$ and $t_{(u+1)} = \infty$. These probabilities can lead to derive lower and upper probabilities for any event of interest in terms of X_{n+1}, by applying the same technique used for $A_{(n)}$-based inference [Augustin and Coolen, 2004]. If we are interested in the event $X_{n+1} \in A$, with A a set of the nonnegative real values, then the lower probability for this event, referred to by $\underline{P}(X_{n+1} \in A)$, is derived by summing only the probabilities for X_{n+1} on intervals $\left(t_{(i)}, t_{(i+1)}\right)$, $\left(rc_{(j)}, t_{rc_{(j)}},\right)$ and $\left(t_{lc_{(w)}}, lc_{(w)}\right)$, which are completely within the set A. The upper probability for the event $X_{n+1} \in A$, referred to by $\bar{P}(X_{n+1} \in A)$, is derived by summing all the probabilities for X_{n+1} on intervals $\left(t_{(i)}, t_{(i+1)}\right)$, $\left(rc_{(j)}, t_{rc_{(j)}},\right)$ and $\left(t_{lc_{(w)}}, lc_{(w)}\right)$, which have a nonempty intersection with the set A.

Table 1
The Simulated Data Examples

Example	***N***	**Event Distribution**	**Right-Censored Distribution**	**Left-Censored Distribution**
1	20	Weibull ($\alpha = 2, \beta = 1.5$)	Uniform ($\alpha_1 = 0$, $b_1 = 3.5$)	Uniform ($\alpha_2 = 0$, $b_2 = 1$)
2	1000	Weibull ($\alpha = 2, \beta = 1.5$)	Uniform ($\alpha_1 = 2.5$, $b_1 = 3.5$)	Uniform ($\alpha_2 = 0$, $b_2 = 0.5$)

Table 2
The Simulated Double-Censored Data for the first Example

x_0	d_0	x_0	d_0
0.575	2	0.887	2
0.720	2	0.893	3
0.777	1	0.925	3
0.812	3	0.975	3
0.829	3	0.987	2
0.834	3	1.041	1
0.838	1	1.058	1
0.851	1	1.227	1
0.875	1	1.234	1
0.886	2	1.308	1

6. COMPARISON WITH ALTERNATIVE NONPARAMETRIC METHODS

The partially specified predictive probability distribution for X_{n+1}, which can be derived by the double-censoring $A_{(n)}$ assumption, is used in two simulated data examples, which are presented in Table 1, to provide the imprecise probabilities for such events of interest. However before applying the method, it is worth to show how to create the double-censored data. The simulated data is created by generating n observations from Weibull distribution, with shape parameter α and scale parameter β, and generating n observations from Uniform distribution, with parameters α_1 and b_1, and generating n observations from Uniform distribution, with parameters α_2 and b_2, where these three distributions are assumed for event, right-censored and left-censored data, respectively. Then, we use Equations (7) and (8) to define the double-censored data. Due to small data set in the first example, the double-censored observations are presented in Table 2.

$$x_o = max\,[min(t_o, rc_o), lc_o]\ ,\text{for } o = 1, 2, \ldots, n \qquad (7)$$

$$\mathrm{d_o} = \begin{cases} 1 \text{ if } x_o = t_o \text{ (uncensored)} \\ 2 \text{ if } x_o = rc_o \text{ (right} - \text{censored)} \\ 3 \text{ if } x_o = lc_o \text{ (left} - \text{censored)} \end{cases} \qquad (8)$$

where x_o is the time and d_o is the censored indicator.

The dc-$A_{(n)}$ assumption leads to a partially specified probability distribution, via the probabilities calculated by Equations (6), for survival time of a future observation X_{n+1}, which we refer to by a random quantity X_{21} in the first example. The probabilities related to the intervals $\left(t_{(i)}, t_{(i+1)}\right)$, $\left(rc_{(j)}, t_{rc_{(j)}},\right)$ and $\left(t_{lc_{(w)}}, lc_{(w)}\right)$ are presented in Table 3, and these assignment probabilities can be used to derive bounds for the survival function of X_{21}, so bounds for $S_{X_{21}}(t) = P(X_{21} > t)$, for $t \geq 0$. We call the maximum lower bound "the lower survival function", denoted by $\underline{S}_{X_{21}}(t) = \underline{P}(X_{21} > t)$, and the minimum upper bound "the upper survival function", denoted by $\bar{S}_{X_{21}}(t) = \bar{P}(X_{21} > t)$. These imprecise probabilities are derived as described in Section 5.

The upper survival function for X_{n+1} can be easily derived, due to the fact that the probabilities defined based on the dc-$A_{(n)}$ assumption are all specified on the intervals $(t_{(i)}, t_{(i+1)})$, $(rc_{(j)}, t_{rc_{(j)}})$ and $(t_{lc_{(w)}}, lc_{(w)})$. To derive $\bar{S}_{X_{21}}(t) = \bar{P}(X_{21} > t)$, for $t > 0$, we sum up all mass probabilities related to the intervals $(t_{(i)}, t_{(i+1)})$, $(rc_{(j)}, t_{rc_{(j)}})$ and $(t_{lc_{(w)}}, lc_{(w)})$, which have intersections with the set (t, ∞). To compute the lower survival function for X_{n+1}, $\underline{S}_{X_{21}}(t) = \underline{P}(X_{21} > t)$, for $t > 0$, we sum up all mass probabilities related to the intervals $(t_{(i)}, t_{(i+1)})$, $(rc_{(j)}, t_{rc_{(j)}})$ and $(t_{lc_{(w)}}, lc_{(w)})$, which are completely within the set (t, ∞). Both bounds of the survival function are step functions, and they both equal to one at time 0 due to the assumption that there are no observed events at time 0, $t_{(0)} = 0$. In the last interval, $(., \infty)$, the upper survival function bound is a positive constant, but the lower survival function bound is equal to zero.

Table 3
dc- $A_{(n)}$ -based Probabilities for X_{21} on the Intervals
$(t_{(i)}, t_{(i+1)})$, $(rc_{(j)}, t_{rc_{(j)}})$ and $(t_{lc_{(w)}}, lc_{(w)})$

Interval	Probability
(0, 0.777)	0.148
(0.575, 0.777)	0.005
(0.720, 0.777)	0.005
(0.777, 0.812)	0.024
(0.777, 0.829)	0.024
(0.777, 0.834)	0.024
(0.777, 0.838)	0.086
(0.838, 0.851)	0.086
(0.851, 0.875)	0.086
(0.875, 0.893)	0.010
(0.875, 0.925)	0.010
(0.875, 0.975)	0.010
(0.875, 1.041)	0.057
(0.886, 1.041)	0.008
(0.887, 1.041)	0.008
(0.987, 1.041)	0.008
(1.041, 1.058)	0.081
(1.058, 1.227)	0.081
(1.227, 1.234)	0.081
(1.234, 1.308)	0.081
(1.308, ∞)	0.081

Table 4
The Lower and Upper Survival Functions for X_{21}

$t \in (.,.)$	$\underline{S}_{X_{21}}(t)$	$\bar{S}_{X_{21}}(t)$
(0, 0.575]	0.852	1
[0.576, 0.719]	0.848	1
[0.720, 0.776]	0.843	1
[0.777, 0.812]	0.686	0.843
[0.813, 0.828]	0.686	0.819
[0.829, 0.833]	0.686	0.795
[0.834, 0.837]	0.686	0.771
[0.838, 0.850]	0.600	0.686
[0.851, 0.874]	0.514	0.600
[0.875, 0.886]	0.429	0.514
[0.887, 0.892]	0.413	0.514
[0.893, 0.925]	0.413	0.505
[0.926, 0.975]	0.413	0.495
[0.976, 0.987]	0.413	0.486
[0.988, 1.040]	0.405	0.486
[1.041, 1.057]	0.324	0.405
[1.058, 1.226]	0.243	0.324
[1.227, 1.233]	0.162	0.243
[1.234, 1.307]	0.081	0.162
[1.308, ∞)	0	0.081

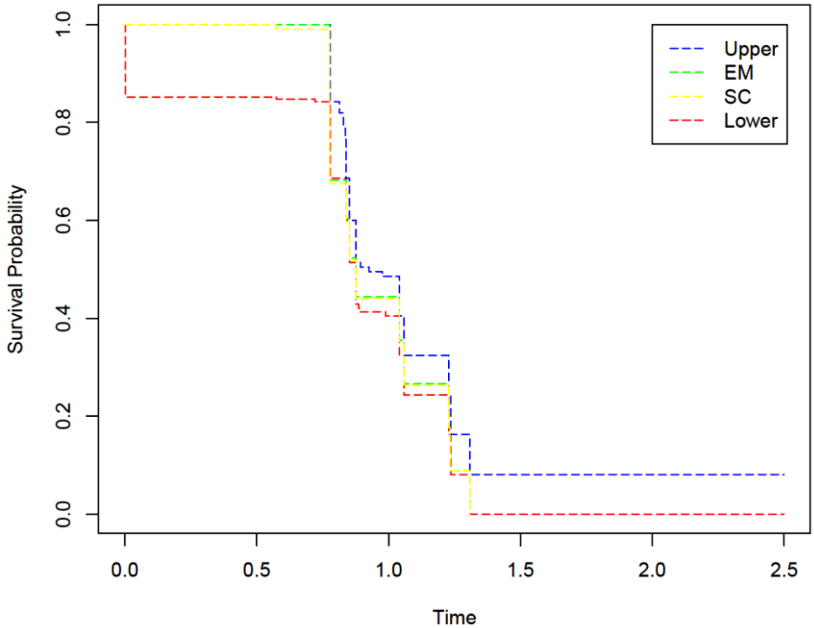

Figure 1: The Lower and Upper Survival Functions for X_{21} along with the Precise Estimates based on the EM and SC Methods

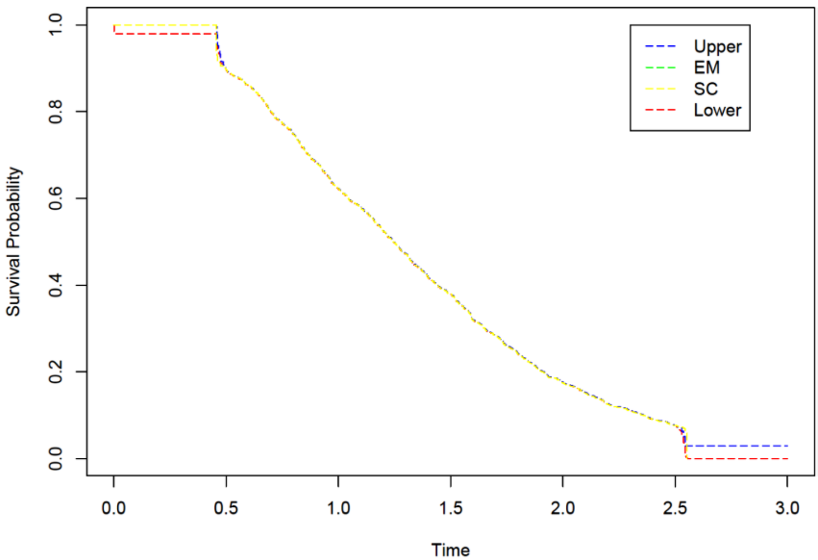

Figure 2: The Lower and Upper Survival Functions for X_{1001} along with the Precise Estimates based on the EM and SC Methods

Table 4 gives the survival function values for X_{21}, specified on intervals created by the simulated data from the first example, and the results are presented in Figure 1. The values of the lower and upper survival functions at observation times can be easily derived from Table 4, with respect to the fact that the lower survival function is continuous from the left at all observations, and the upper survival function is continuous from the right at event times. The right-censored and left-censored observations increase the difference between corresponding upper and lower survival functions at any censored observation. In case of data including many censored observations, regardless whether they are right-censored or left-censored, the difference between corresponding bounds of the survival function at time t becomes large.

In Figure 1, we compared our proposed method with the nonparametric maximum likelihood estimators, which are based on the expectation-maximization algorithm, EM, and the self-consistency algorithm, SC, although their main goal is quite different, namely estimation of the underlying population survival function instead of prediction for one future observation X_{n+1}. We can conclude that the proposed method always gives more probability to the upper bound of the distribution than the precise estimation methods do, and that such underestimation of the tail of the survival distribution is the primary practical deficiency of the precise estimation methods. After the largest observation, the lower survival function for X_{21} and the precise estimates are both zero. However, the upper survival function always gets a positive constant, unless we restrict the data range for X_{21} by choosing a finite upper limit. Before the first event time $t_1 = 0.777$, the upper survival function and the precise estimates are both equal to one, but the lower survival function is less than one, and it decreases at the censored observations, which is not the case with the upper survival function and the precise estimation methods. The nonparametric maximum likelihood estimators have been recently used in the literature to graphically present double-censored data. We believe that the dc-$A_{(n)}$-based lower and upper survival functions for X_{21} are better suited for such graphical presentation, as they indeed give a wider picture of the double-censored data, and from a predictive perspective, they can easily be interpreted. For the case of large data and most observations are event, not censored, these methods become nearly identical; and this is obvious in Figure 2, where the simulated data is from the second example in Table 1 and the sample size is 1000.

7. TREATING SUCH CASES OF TIES

For simplicity, we have assumed that there are no ties in the data sets through this paper, but in real studies, ties could occur in different scenarios. There are seven kinds of ties that could occur: tied event observations, tied right-censoring observations, tied left-censoring observations, ties among event and right-censoring observations, ties among event and left-censoring observations, ties among left-censoring and right-censoring observations, and ties among event and left-censoring observations with right-censoring observations. With the first three situations, we break the tied observations by adding a small value to those ties. With the fourth situation, we assume that the right-censoring times occur after the event observations, where this assumption has been widely used in the literature [Berliner and Hill (1988) and Kaplan and Meier (1958)]. With the fifth situation, we assume that the left-censoring times occur before the event observations. With the sixth situation, we assume that the right-censoring times occur after the left-censoring times. With the last situation, we assume that the right-censoring times occur after the event observations and the left-censoring times occur before the event times.

An interesting result of having tied event time observations, say at time t_t, is that our method, like rc-$A_{(n)}$-based inferences and $A_{(n)}$-based inferences in general, then gives a positive predictive probability for X_{n+1} to occur at t_t. This seems rational due to the fact that more than one event happened at time t_t supports the case that future events can also occur at the time t_t. Of course, such ties have been widely occurred due to data display and the discrete nature of measurement [Meeker and Escobar, 1998].

8. CONCLUDING REMARKS

The nonparametric predictive inferences with double-censored data for one future observation X_{n+1} introduced in this paper can be used in multiple different ways. Imprecise probabilities and survival functions, as presented in Sections 5 and 6, can be used for a variety of inferential predictive problems, where the predictive inferences have been presented in many statistical applications. Some references suggested the use of $A_{(n)}$-based nonparametric predictive inference for real-valued data [Coolen (1998), Coolen et al. (2000), Coolen et al. (2002) and Coolen, and Van der Laan (2001)], and other references support the use of rc-$A_{(n)}$-based nonparametric predictive inference for right-censored data [Coolen and Yan (2003), Coolen and Yan (2004), Coolen-Maturi et al. (2012) and Maturi (2010)]. For bivariate data, Coolen-Maturi et al. (2016)], Muhammad (2016), Muhammad et al. (2016) introduced the use of bivariate-data-$A_{(n)}$-based nonparametric predictive inference.

In this paper, we emphasize to keep the double-censoring $A_{(n)}$ assumption with few mathematical assumptions and the inferences based on the proposed assumption are suited for the situations, which we have vague knowledge about. Of course, the dc-$A_{(n)}$-based inferences may not be able to lead to the optimal decisions, but they can be used as a basis for studying the effect of additional modeling assumptions on final inferences, or related decisions, in the case of wishing to use methods with more structure. For example, when comparing two survival groups predictively, in the case of data including many double-censored observations, the range of survival functions between the lower and upper bounds per group may lead to preference of either group. In such cases, the dc-$A_{(n)}$-based inferences conclude that strong inferences may not be possible based only on the data, so there is a need to add further modeling assumptions or to add more observations. From this situation, the dc-$A_{(n)}$-based inferences are related to robust statistical methods.

Another generalization of $A_{(n)}$ assumption can be proposed for data including double-ensored observations by using the Exponential distribution for the right-censored observations. This distribution has been widely used to model lifetime data including right-censored observation. Wu (2022) can be a good source to start this future work.

BinHimd (2014), BinHimd and Coolen (2012) and Coolen and BinHimd (2020) show that the $A_{(n)}$ assumption can be used to provide a smoothed bootstrap method for real-valued data. For right-censored data and bivariate data, Al-Luhayb (2021), Al-Luhayb et al. (2019a & 2019b) used the rc- $A_{(n)}$ and bivariate-data- $A_{(n)}$ assumptions to provide smoothed bootstrap methods. The dc- $A_{(n)}$ assumption can smooth the bootstrap method for double-censored data, and the bootstrap method can be used for survival analysis and testing. These extensions will be left as future research topics.

ACKNOWLEDGMENTS

The researcher would like to thank the Deanship of Scientific Research, Qassim University for funding the publication of this project. Further thanks go to the reviewers for the valuable comments and suggestions.

REFERENCES

1. Ahmadini, A.A. and Coolen, F.P. (2018). On the use of an imprecise statistical method for accelerated life testing data using the power-law link function. In 2018 *3rd International Conference on System Reliability and Safety (ICSRS)* (244-249). IEEE.
2. Al-Luhayb, A.S.M. (2021). *Smoothed Bootstrap Methods for Right-Censored Data and Bivariate Data*. Ph.D. thesis. Durham University.
3. Al Luhayb, A.S., Coolen, F.P.A. and Coolen-Maturi, T. (2019a). Generalizing Banks smoothed bootstrap method for right-censored data. In *29th European Safety and Reliability Conference* (ESREL 2019), Hannover (Germany). pp. 894-901.
4. Al Luhayb, A.S., Coolen-Maturi, T. and Coolen, F.P. (2019b). Smoothed bootstrap for survival function inference. In *2019 International Conference on Information and Digital Technologies* (IDT) (pp. 296-303). IEEE.
5. Ashleik, N.A. (2018). *Learning for Contingency Tables and Survival Data Using Imprecise Probabilities*. Ph.D. thesis. University of Saskatchewan.
6. Augustin, T. and Coolen, F.P. (2004). Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference*, 124(2), 251-272.
7. Berliner, L.M. and Hill, B.M. (1988). Bayesian nonparametric survival analysis. *Journal of the American Statistical Association*, 83(403), 772-779.
8. BinHimd, S. (2014). *Nonparametric predictive methods for bootstrap and test reproducibility* (Doctoral dissertation, Durham University).
9. Binhimd, S. and Coolen, F.P.A. (2012). On boostrapping using nonparametric predictive inference. In *Proceedings of the 1st ISM International Statistical Conference* (ISM 2012) (pp. 465-468).
10. Coolen, F., Ahmadini, A.A.H. and Coolen-Maturi, T. (2021). Imprecise inference based on the log-rank test for accelerated life testing. *Metrika*, 84, 913-925.
11. Coolen, F.P.A. (1998). Low structure imprecise predictive inference for bayes' problem. *Statistics & Probability Letters*, 36, 349-357.
12. Coolen, F.P.A. and BinHimd, S. (2020). Nonparametric predictive inference bootstrap with application to reproducibility of the two-sample kolmogorov-smirnov test. *Journal of Statistical Theory and Practice*, 14, 1-13.
13. Coolen, F.P.A. and Coolen-Schrijner, P. (2000). Condition monitoring: A new perspective. *Journal of the Operational Research Society*, 51, 311-319.
14. Coolen, F.P.A., Coolen-Schrijner, P. and Yan, K.J. (2002). Nonparametric predictive inference in reliability. *Reliability Engineering & System Safety*, 78, 185-193.
15. Coolen, F.P.A. and Van der Laan, P. (2001). Imprecise predictive selection based on

low structure assumptions. *Journal of Statistical Planning and Inference*, 98, 259-277.
16. Coolen, F.P. and Yan, K.J. (2003). Nonparametric Predictive Comparison of Two Groups of Lifetime Data. In *ISIPTA*, 3, 148-161.
17. Coolen, F.P.A. and Yan, K.J. (2004). Nonparametric predictive inference with right-censored data. *Journal of Statistical Planning and Inference*, 126, 25-54.
18. Coolen-Maturi, T., Coolen, F.P.A. and Muhammad, N. (2016). Predictive inference for bivariate data: Combining nonparametric predictive inference for marginals with an estimated copula. *Journal of Statistical Theory and Practice*, 10, 515-538.
19. Coolen-Maturi, T., Coolen-Schrijner, P. and Coolen, F.P. (2012). Nonparametric predictive multiple comparisons of lifetime data. *Communications in Statistics-Theory and Methods*, 41, 4164-4181.
20. De Finetti, B. (1974). *Theory of probability* (2 volumes). Wiley, London.
21. Dempster, A.P. (1963). On direct probabilities. *Journal of the Royal Statistical Society: Series B (Methodological)*, 25, 100-110.
22. Hill, B.M. (1968). Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association*, 63, 677-691.
23. Hill, B.M. (1988). De finetti's theorem, induction, and $A_{(n)}$ or bayesian nonparametric predictive inference (with discussion). In: *Bayesian Statistics*, 3, 211-241.
24. Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53, 457-481.
25. Lane, D.A. and Sudderth, W.D. (1984). Coherent predictive inference. *Sankhyā: The Indian Journal of Statistics*, Series A, 166-185.
26. Maturi, T. (2010). *Nonparametric predictive inference for multiple comparisons*. Doctoral dissertation, Durham University, UK.
27. Meeker, W.Q. and Escobar, L.A. (1998). *Statistical methods for reliability data*. Wiley, New York.
28. Muhammad, N. (2016). *Predictive Inference with Copulas for Bivariate Data*. Ph.D. thesis. Durham University, UK. https://npi-statistics.com/ pdfs/theses/ NM16.pdf.
29. Muhammad, N., Coolen, F.P.A. and Coolen-Maturi, T. (2016). Predictive inference for bivariate data with nonparametric copula. *The American Institute of Physics (AIP) Conference Proceedings* 1750, 0600041-0600048. https://aip.scitation.org/doi/abs/ 10.1063/1.4954609.
30. Siddiqa, H., Ali, S. and Shah, I. (2021). Most recent changepoint detection in censored panel data. *Computational Statistics*, 36, 515-540.
31. Tee, K.F., Pesinis, K. and Coolen-Maturi, T. (2019). Competing risks survival analysis of ruptured gas pipelines: A nonparametric predictive approach. *International Journal of Pressure Vessels and Piping*, 175, 103919.
32. Walley, P. (1991). *Statistical reasoning with imprecise probabilities*, Chapman and Hall, London.
33. Weichselberger, K. (2001). Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I: Intervallwahrscheinlichkeit als umfassendes Konzept (Vol. 1). Springer-Verlag, Berlin Heidelgert GmbH
34. Wu, S.F. (2022). Bayesian interval estimation for the two-parameter exponential distribution based on the right type ii censored sample. *Symmetry*, 14, 352.