DE GRUYTER Demonstratio Mathematica 2024; 57: 20230126

Research Article

Asamh Saleh M. Al Luhayb*

Nonparametric methods of statistical inference for double-censored data with applications

https://doi.org/10.1515/dema-2023-0126
received December 8, 2022; accepted November 6, 2023

Abstract: This article introduces new nonparametric statistical methods for prediction in case of data containing right-censored observations and left-censored observations simultaneously. The methods can be considered as new versions 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} based on each version, and these bounds are compared through two examples. Two interesting features are provided based on the proposed methods. 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

MSC 2020: 62Gxx

1 Introduction

This article introduces new nonparametric statistical methods for prediction using the past data that contain right-censored observations and left-censored observations simultaneously, where these kind of data are called double-censored data in the literature. The methods are proposed to learn about one future observation based on the original sample with few mathematical assumptions. For real-valued data, Hill [1,2] 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 [3,4] generalized the $A_{(n)}$ assumption, and the generalizations are considered for survival analysis. In this article, regarding methods are developed by using the information of censoring observations, and they are referred to by generalization type A and B of $A_{(n)}$ assumption.

This article 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 generalizations of $A_{(n)}$ assumption for double-censored data along with their corresponding justifications. In Section 5, we present bounds for probabilities and survival functions based on the generalizations of $A_{(n)}$ assumption proposed for double-censored data. In Section 6, the new proposed generalizations are used with two applications. In this article, 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.

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

2 $A_{(n)}$ assumption for real-valued univariate data

In 1968, Hill [1] 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 $\frac{1}{n+1}$ is assigned to each created interval. Let $X_1, X_2, \ldots, X_n$ be continuous and exchangeable real-valued random quantities, and let $x_1, x_2, \ldots, x_n$ be the corresponding observed data points. Furthermore, let $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$ 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

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

where $i = 0, 1, 2, \ldots, n$.

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 [5], 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 [6,7]. Augustin and Coolen [8] 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 [9], Coolen and Coolen-Schrijner [10], Coolen et al. [11], Coolen and Van der Laan [12].

Based on data including n event observations, the $A_{(n)}$ assumption provides a partially specified predictive probability distribution for one future observation X_{n+1} via the probabilities assigned to the intervals $(x_{(i)}, x_{(i+1)})$, where $i = 1, 2, \ldots, n$, with $x_{(0)} = 0$ and $x_{(n+1)} = \infty$. These probabilities can lead to derive lower and upper probabilities for any event of interest in terms of X_{n+1} [8]. 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 $(x_{(i)}, x_{(i+1)})$, which are completely within the set A. The upper probability for the event $X_{n+1} \in A$, referred to by $\overline{P}(X_{n+1} \in A)$, is derived by summing all the probabilities for X_{n+1} on intervals $(x_{(i)}, x_{(i+1)})$, which have a nonempty intersection with the set A.

The $A_{(n)}$ assumption has been discussed for multiple statistical inferences in the literature, see, e.g., Hill [2] and De Finetti [5] for more detailed presentation, and the assumption has been contributed with multiple important related works, see, e.g., Dempster [13] and Lane and Sudderth [14] for detailed information. For right-censored data, Mark Berliner and Hill [3] and Coolen and Yan [4] proposed two versions of $A_{(n)}$, and the proposed assumptions will be presented in the next section.

3 Generalizations of $A_{(n)}$ assumption for right-censored univariate data

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 Mark Berliner and Hill [3]. 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 $X_1, X_2, \ldots, X_n$ 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 $0 < t_{(1)} < t_{(2)} < \ldots < t_{(u)}$, where $0 \le u \le n$, and $0 < c_{(1)} < c_{(2)} < \ldots < c_{(v)}$, 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 $(t_{(u)}, t_{(u+1)})$ by the following formula:

$$P_{(u+1)} = (1-\lambda_{(0)})\times\ldots\times(1-\lambda_{(u)})\times\lambda_{(u+1)}, \tag{2}$$

where $\lambda_{(u)} = \frac{1}{n+1-u-C_{(u)}}$, $C_{(u)} = \#\{\text{censors} < t_{(u+1)}\}$ 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 [4] 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, $rc\text{-}A_{(n)}$, provides a partial probability distribution for one future observation, and it is specified via M-function values [4]. The random quantities $X_1, X_2, \ldots, X_n$ 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 $0 < t_{(1)} < t_{(2)} < \ldots < t_{(u)}$, where $0 \le u \le n$, and $0 < c_{(1)} < c_{(2)} < \ldots < c_{(v)}$, where $v = n - u$, respectively. Let the interval $I_i = (t_{(i)}, t_{(i+1)})$, for $0 \le i \le u$, they ordered the right-censoring times in each interval I_i by $c_1^i < c_2^i < \ldots < c_{l_i}^i$, where l_i is the number of censors in I_i. Finally, they formed the open intervals as $(t_{(i)}, t_{(i+1)})$ and $(c_k^i, t_{(i+1)})$, where $1 \le k \le l_i$.

Definition. (M-function). The M-function value is a probability partially specified to each interval $(t_{(i)}, t_{(i+1)})$ or $(c_k^i, t_{(i+1)})$, 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 [4].

Definition. ($rc\text{-}A_{(n)}$). The $rc\text{-}A_{(n)}$ 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 [4]:

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

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

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 $t_{(u+1)} = +\infty$. $M_{X_{n+1}}(t_{(i)}, t_{(i+1)})$ and $M_{X_{n+1}}(c_k^i, t_{(i+1)})$ mean that the random quantity X_{n+1} falls in $(t_{(i)}, t_{(i+1)})$ and $(c_k^i, t_{(i+1)})$, respectively, with a probability M-function value.

From studying the $rc\text{-}A_{(n)}$ 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 [4] 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). Second, the M-function values are limited between 0 and 1, and the values have to sum up to one over all intervals created [4].

4 Generalizations of $A_{(n)}$ assumption for double-censored univariate data

This section provides two generalizations of $A_{(n)}$ assumption along with their justifications. For the justifications, we will include a detailed information of the exact nature of the noninformative censoring assumption implicit in the generalizations, and the justifications 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 postdata and predata 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 postdata assumption will be for right-censored data, and the further predata 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 justifications of the generalizations of $A_{(n)}$, 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. ($\tilde{A}_{(n)}$). 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:

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

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 $(t_{(i)}, t_{(i+1)})$ 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 $(t_{(i)}, t_{(i+1)})$, then the unobserved data point t_{rc} corresponding to that right-censored observation could occur in $(t_{(i)}, t_{(i+1)})$. This leads to assign a probability $\frac{1}{n+1}$ to the interval $(t_{(i)}, t_{rc})$ and another probability $\frac{1}{n+1}$ to the interval $(t_{rc}, t_{(i+1)})$. For the case of having a left-censored observation in $(t_{(i)}, t_{(i+1)})$, then the unobserved data point t_{lc} corresponding to that left-censored observation could occur in $(t_{(i)}, t_{(i+1)})$. This leads to assign a probability $\frac{1}{n+1}$ to the interval $(t_{(i)}, t_{lc})$ and another probability $\frac{1}{n+1}$ to the interval $(t_{lc}, t_{(i+1)})$. 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 subintervals. Therefore, the probability for X_{n+1} to be fallen in $(t_{(i)}, t_{(i+1)})$ 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 $(t_{(i)}, t_{(i+1)})$ would be split into 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 $(rc_{(j)}, \infty)$, and 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 $(t_{(i)}, t_{(i+1)})$, and this leads to split the interval

$(t_{(i)}, t_{(i+1)})$ 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 $(0, lc_{(w)})$, and hence, $P(X_{n+1} \in (0, lc_{(w)})) = \frac{1}{n+1}$.

4.1 Generalization type A

To link the first stage to the second stage of the justification of the generalization type A, each probability assigned to interval $(rc_{(j)}, \infty)$ will be uniformly distributed to the event intervals $(t_{(i)}, t_{(i+1)})$ occurred after $rc_{(j)}$ and the interval $(t_{(i)}, t_{(i+1)})$ including the right-censored observation $rc_{(j)}$. For the left-censored observations, each probability assigned to interval $(0, lc_{(w)})$ will be uniformly distributed to the event intervals $(t_{(i)}, t_{(i+1)})$ occurred before $lc_{(w)}$ and the interval $(t_{(i)}, t_{(i+1)})$ including the left-censored observation $lc_{(w)}$. These assumptions lead to have the following function:

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

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

4.2 Generalization type B

To link the first stage to the second stage of the justification of the generalization type B, each probability assigned to interval $(rc_{(j)}, \infty)$ will be uniformly distributed to the event intervals $(t_{(i)}, t_{(i+1)})$ 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 $(t_{(i)}, t_{(i+1)})$ occurred before $lc_{(w)}$ and the interval $(t_{lc_{(w)}}, lc_{(w)})$, where $t_{lc_{(w)}}$ is the first event time less than $lc_{(w)}$. These assumptions lead to have the following functions:

$$\begin{aligned} P(X_{n+1} \in (t_{(i)}, t_{(i+1)})) = \frac{1}{n+1} &+ \sum_{j=1}^{v} \frac{I(rc_{(j)} < t_{(i)})}{(n+1)(\#\{t_{(\cdot)} > rc_{(j)}\}+1)} \\ &+ \sum_{w=1}^{k} \frac{I(lc_{(w)} > t_{(i+1)})}{(n+1)(\#\{t_{(\cdot)} < lc_{(w)}\}+1)} \\ P(X_{n+1} \in (rc_{(j)}, t_{rc_{(j)}})) &= \frac{1}{(n+1)(\#\{t_{(\cdot)} > rc_{(j)}\}+1)} \\ P(X_{n+1} \in (t_{lc_{(w)}}, lc_{(w)})) &= \frac{1}{(n+1)(\#\{t_{(\cdot)} < lc_{(w)}\}+1)}, \end{aligned} \tag{7}$$

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

From the generalizations A and B of $A_{(n)}$ assumption, four interesting notes should be pointed out. First, the generalization type A divides the sample space into $u + 1$ intervals based on the event observations only, while the generalization type B divides the sample space into $n + 1$ intervals based on the censored and event observations. Second, the intervals created by the generalization A will be never overlapping, contrary to the intervals created by the generalization B. The intervals created by generalization type B are overlapped due to the censored observations, which are used in the intervals' creation. Third, for the case of data not including censored observations, both generalizations will return to the original $A_{(n)}$ assumption proposed for real-valued data. They will be the same. The forth note is that the probabilities, created by either generalization type A or generalization type B, are limited between 0 and 1, and the probabilities have to sum up to one over all intervals created.

5 Imprecise probabilities based on the generalizations of $A_{(n)}$ assumption

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 generalizations of $A_{(n)}$ 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 [8–12] and based on the right-censoring $A_{(n)}$ assumption for right-censored data [15–19]. For many events of interest regarding to X_{n+1}, the probability values proposed by equations (6) and (7) allow bounds for probabilities to be derived analogously as for $A_{(n)}$-based inference [8], 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 [6,7].

Based on data including u event observations, v right-censored observations, and $k = n - (u + v)$ left-censored observations, each generalization of $A_{(n)}$ assumption provides a partially specified predictive probability distribution for one future observation X_{n+1} via the probabilities assigned to the intervals $(t_{(i)}, t_{(i+1)})$, $(rc_{(j)}, t_{rc_{(j)}})$ and $(t_{lc_{(w)}}, lc_{(w)})$, 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 [8]. 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 $(t_{(i)}, t_{(i+1)})$, $(rc_{(j)}, t_{rc_{(j)}})$ and $(t_{lc_{(w)}}, lc_{(w)})$, which are completely within the set A. The upper probability for the event $X_{n+1} \in A$, referred to by $\overline{P}(X_{n+1} \in A)$, is derived by summing all the probabilities for X_{n+1} on intervals $(t_{(i)}, t_{(i+1)})$, $(rc_{(j)}, t_{rc_{(j)}})$ and $(t_{lc_{(w)}}, lc_{(w)})$, which have a nonempty intersection with the set A.

6 Examples

In this section, we apply the generalizations of $A_{(n)}$ assumption to two data sets, including double-censored observations. In each example, two bounds are derived to predict the survival function for next future observation based on each suggested generalization.

6.1 Simulated example

Before applying the generalizations, it is worth to describe how to create a double-censored data set. The simulated data are formed by generating seven observations from Uniform distribution, with parameters $a = 0$ and $b = 60$, and this is repeated three times. The first generated data are considered for event data, and the second and third generated samples are considered for right- and left-censored data. Then, we use equations (8) and (9) to define the double-censored data set. The data are presented in Table 1, which includes

Table 1: The double-censored data for simulated example

x_o	13	17	22	29	35	46	52
d_o	+	+					−

two right-censored observations with sign "+," one left-censored observation with sign "−," and four noncensored observations with no sign.

$$x_o = \max[\min(t_o, rc_o), lc_o], \quad \text{for } o = 1, 2, \ldots, 7 \tag{8}$$

$$d_o = \begin{cases} 1 & \text{if} \quad x_o = t_o \quad \text{(uncensored)} \\ 2 & \text{if} \quad x_o = rc_o \quad \text{(right-censored)} \\ 3 & \text{if} \quad x_o = lc_o \quad \text{(left-censored)}, \end{cases} \tag{9}$$

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

Each generalization of $A_{(n)}$ assumption leads to a partially specified probability distribution, via the probabilities calculated by equations (6) and (7), for survival time of a future observation X_{n+1}, which we refer to by a random quantity X_8 in the simulated example. The probabilities related to the intervals $(t_{(i)}, t_{(i+1)})$, $(rc_{(j)}, t_{rc_{(j)}})$, and $(t_{lc_{(w)}}, lc_{(w)})$ are presented in Tables 2 and 3, and these assignment probabilities can be used to derive bounds for the survival function of X_8, so bounds for $S_{X_8}(t) = P(X_8 > t)$, for $t \geq 0$. We call the maximum lower bound "the lower survival function," denoted by $\underline{S}_{X_8}(t) = \underline{P}(X_8 > t)$, and the minimum upper bound "the upper survival function," denoted by $\overline{S}_{X_8}(t) = \overline{P}(X_8 > t)$. 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 generalizations of $A_{(n)}$ 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 $\overline{S}_{X_8}(t) = \overline{P}(X_8 > t)$, 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}, $\underline{S}_{X_8}(t) = \underline{P}(X_8 > t)$, 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.

Tables 4 and 5 present the survival function values for X_8, specified on intervals created by the simulated data based on both generalizations of $A_{(n)}$ assumption, 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 Tables 4 and 5, 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

Table 2: The generalization A of $A_{(n)}$-based probabilities for X_8 on the intervals $(t_{(i)}, t_{(i+1)})$

Interval	Probability
(0, 22)	0.200
(22, 29)	0.200
(29, 35)	0.200
(35, 46)	0.200
(46, ∞)	0.200

Table 3: The generalization B of $A_{(n)}$-based probabilities for X_8 on the intervals $(t_{(i)}, t_{(i+1)})$, $(rc_{(j)}, t_{rc_{(j)}})$, and $(t_{lc_{(w)}}, lc_{(w)})$

Interval	Probability
(0, 22)	0.150
(13, 22)	0.025
(17, 22)	0.025
(22, 29)	0.200
(29, 35)	0.200
(35, 46)	0.200
(46, 52)	0.025
(46, ∞)	0.175

Table 4: The lower and upper survival functions for X_8 based on the generalization type A of $A_{(n)}$ assumption

$t \in (\cdot, \cdot)$	$\underline{S}_{X_8}(t)$	$\overline{S}_{X_8}(t)$
(0, 22)	0.8	1
(22, 29)	0.6	0.8
(29, 35)	0.4	0.6
(35, 46)	0.2	0.4
(46, ∞)	0	0.2

Table 5: The lower and upper survival functions for X_8 based on the generalization type B of $A_{(n)}$ assumption

$t \in (\cdot, \cdot)$	$\underline{S}_{X_8}(t)$	$\overline{S}_{X_8}(t)$
(0, 13)	0.850	1
(13, 17)	0.825	1
(17, 22)	0.800	0.975
(22, 29)	0.600	0.800
(29, 35)	0.400	0.600
(35, 46)	0.200	0.400
(46, 52)	0.175	0.200
(52, ∞)	0	0.175

or left censored, the difference between corresponding bounds of the survival function at time t becomes large.

In Figure 1, we compared the generalization type A of $A_{(n)}$ assumption to the generalization type B of $A_{(n)}$ assumption. Due to the fact that the generalization type A partitions the sample support into $u + 1$ intervals and all intervals are based on the event observations only, we have less steps in the corresponding lower and upper survival functions, which are indicated by red and blue step lines respectively, in comparison to the lower and upper survival functions found based on the generalization type B, which are indicated by orange and green step lines, respectively. From Figure 1, it is obvious that the lower survival function based on the generalization type A is less than (or equal) the lower survival function based on the generalization type B, and the upper survival function based on the generalization type A is larger than (or equal) the upper survival function based on the generalization type B. After the largest observation, the lower survival functions for X_8 are both equal to zero. However, the upper survival functions always get a positive constant, unless we restrict the data range for X_8 by choosing a finite upper limit. Before the first event time $t_1 = 22$, the upper survival functions are both equal to one, but the lower survival functions are less than one, and they decrease at the censored observations, which is not the case with the upper survival functions. We believe that the generalizations of $A_{(n)}$-based lower and upper survival functions for X_8 are better suited for such graphical

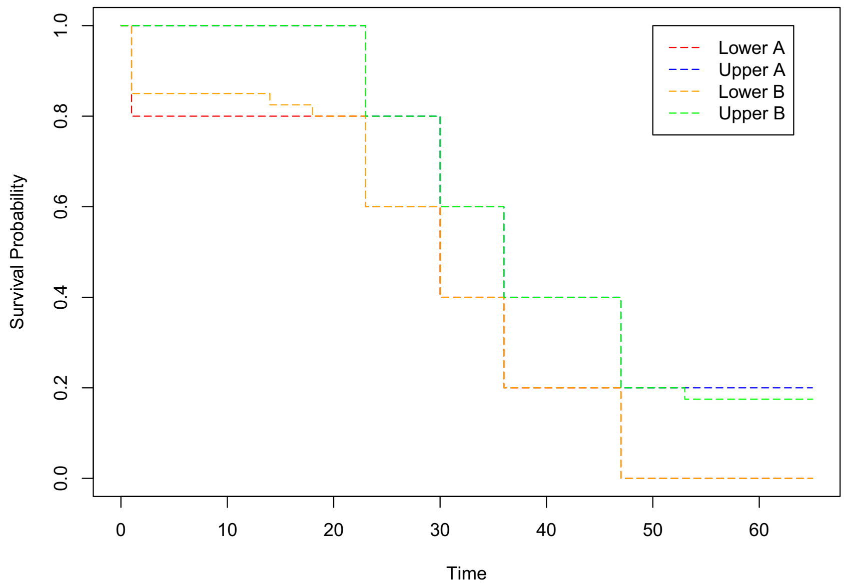

Figure 1: The lower and upper survival functions for X_8 based on the generalizations of $A_{(n)}$ assumption.

presentations, 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 many observations are event, not censored, these proposed methods become nearly identical, and this is obvious in the following example.

6.2 AIDS example

The data set is from a cohort of drug users recruited in a detoxification program in Badalona (Spain) [20]. On the basis of the data, we may predict the lower and upper survival functions for the elapsed time from starting IV-drugs to AIDS diagnosis for the next observation. The data of 232 patients infected with HIV are presented as follows; 136 left AIDS-free (right censored), 14 died with AIDS without prior diagnoses (left censored) and 82 had AIDS while in the program (noncensored).

Figure 2 presents the lower and upper survival functions based on the generalizations of $A_{(n)}$ assumption. It is rational that they are nearly identical and the differences between the lower and upper survival functions become small due to the large sample size with many not censored observations.

7 Treating such cases of ties

For simplicity, we have assumed that there are no ties in the data sets through this article, 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 [3,21]. 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

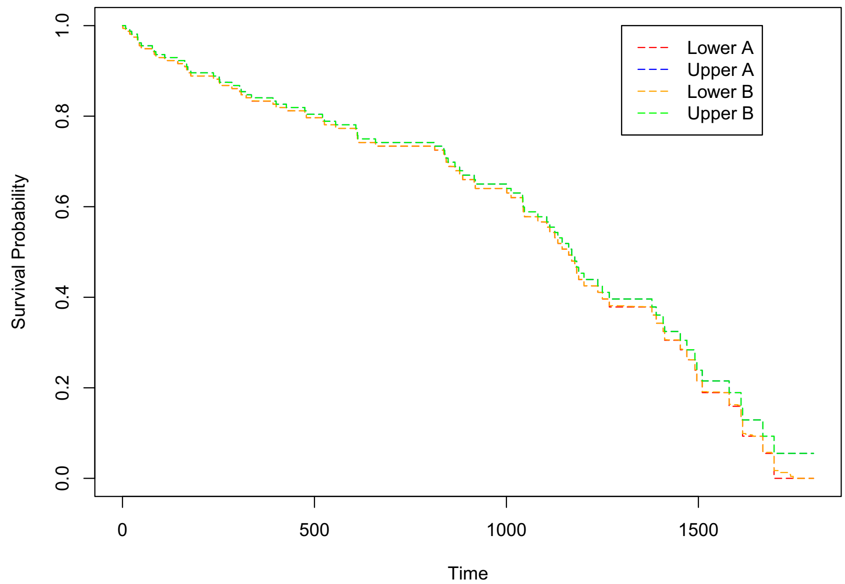

Figure 2: The lower and upper survival functions for X_{1001} based on the generalizations of $A_{(n)}$ assumption.

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 [22].

8 Concluding remarks

The nonparametric predictive inferences with double-censored data for one future observation X_{n+1} introduced in this article 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 [9–12], and other references support the use of rc-$A_{(n)}$-based nonparametric predictive inference for right-censored data [4,23–25]. For bivariate data, Coolen-Maturi et al. [26], Muhammad [27], and Muhammad et al. [28] introduced the use of bivariate-data-$A_{(n)}$-based nonparametric predictive inference.

In this article, we emphasize to keep the generalizations of $A_{(n)}$ assumption with few mathematical assumptions and the inferences based on the proposed assumptions are suited for the situations, which we have vague knowledge about. Of course, the generalizations of $A_{(n)}$-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 generalizations of $A_{(n)}$-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 generalizations of $A_{(n)}$-based inferences are related to robust statistical methods.

Another generalization of $A_{(n)}$ assumption can be proposed for data including double-censored 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. The study by Wu [29] can be a good source to start this future work.

BinHimd [30], BinHimd and Coolen [31], and Coolen and BinHimd [32] 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 [33], and Al Luhayb et al. [34,35] used the rc-$A_{(n)}$ and bivariate-data-$A_{(n)}$ assumptions to provide smoothed bootstrap methods; for more information, it is good to see [36,37]. The generalizations of $A_{(n)}$ 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.

Funding information: Not applicable.

Author contributions: Not applicable.

Conflict of interest: The author states that there is no conflict of interest.

Ethical approval: Not applicable.

Data availability statement: Not applicable.

References

[1] B. M. Hill, *Posterior distribution of percentiles: Bayes' theorem for sampling from a population*, J. Amer. Stat. Assoc. **63** (1968), 677–691.
[2] B. M. Hill, *De finetti's theorem, induction, and $A_{(n)}$ or bayesian nonparametric predictive inference (with discussion)*, Bayesian Stat. **3** (1988), 211–241; J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, (Eds), Oxford University Press.
[3] L. Mark Berliner and B. M. Hill, *Bayesian nonparametric survival analysis*, J. Amer. Stat. Assoc. **83** (1988), 772–779.
[4] F. P. A. Coolen and K. J. Yan, *Nonparametric predictive inference with right-censored data*, J. Stat. Planning Inference **126** (2004), 25–54.
[5] B. De Finetti, *Theory of Probability*, vol. 2, Wiley, London.
[6] P. Walley, *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, London, 1991.
[7] K. Weichselberger, *Elementare grundbegriffe einer allgemeineren wahrscheinlichkeitsrechnung i: Intervallwahrscheinlichkeit als umfassendes konzept*, Physica-Verlag, Heidelberg, 2001.
[8] T. Augustin and F. P. A. Coolen, *Nonparametric predictive inference and interval probability*, J. Stat. Planning Inference **124** (2004), 251–272.
[9] F. P. A. Coolen, *Low structure imprecise predictive inference for bayes' problem*, Stat. Probability Lett. **36** (1998), 349–357.
[10] F. P. A. Coolen and P. Coolen-Schrijner, *Condition monitoring: a new perspective*, J. Operat. Res. Soc. **51** (2000), no. 3, 311–319.
[11] F. P. A. Coolen, P. Coolen-Schrijner, and K. J. Yan, *Nonparametric predictive inference in reliability*, Reliabil. Eng. Syst. Safety **78** (2002), no. 2, 185–193.
[12] F. P. A. Coolen and P. Van der Laan, *Imprecise predictive selection based on low structure assumptions*, J. Stat. Planning Inference **98** (2001), no. 1–2, 259–277.
[13] A. P. Dempster, *On direct probabilities*, J. R. Stat. Soc. Ser. B (Methodological) **25** (1963), no. 1, 100–110.
[14] D. A. Lane and W. D. Sudderth, *Coherent predictive inference*, Sankhyā Indian J. Stat. Ser A **46** (1984), 166–185.
[15] A. A. H. Ahmadini and F. P. A. Coolen, *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), IEEE, 2018, pp. 244–249.
[16] N. A. Ashleik, *Learning for Contingency Tables and Survival Data Using Imprecise Probabilities*, Ph.D. thesis, University of Saskatchewan, 2018.
[17] F. Coolen, A. A. H. Ahmadini, and T. Coolen-Maturi, *Imprecise inference based on the log-rank test for accelerated life testing*, Metrika **84** (2021), no. 6, 913–925.

[18] H. Siddiqa, S. Ali, and I. Shah, *Most recent changepoint detection in censored panel data*, Comput. Stat. **36** (2021), no. 1, 515–540.
[19] K. Fah Tee, K. Pesinis, and T. Coolen-Maturi, *Competing risks survival analysis of ruptured gas pipelines: A nonparametric predictive approach*, Int. J. Pressure Vessels Piping **175** (2019), 103919.
[20] O. Julià and G. Gómez, *Simultaneous marginal survival estimators when doubly censored data is present*, Lifetime Data Anal. **17** (2011), no. 3, 347–372.
[21] E. L. Kaplan and P. Meier, *Nonparametric estimation from incomplete observations*, J. Amer. Stat. Assoc. **53** (1958), 457–481.
[22] W. Q. Meeker and L. A. Escobar, *Statistical Methods for Reliability Data*, Wiley, New York, 1998.
[23] F. P. A. Coolen and K. Jian Yan, *Nonparametric predictive comparison of two groups of lifetime data*, ISIPTA **3** (2003), 148–161.
[24] T. Coolen-Maturi, P. Coolen-Schrijner, and F. P. A. Coolen, *Nonparametric predictive multiple comparisons of lifetime data*, Commun. Stat.-Theory Methods. **41** (2012), no. 22, 4164–4181.
[25] T. A. Maturi, *Nonparametric Predictive Inference for Multiple Comparisons*, Ph.D. thesis, Durham University, UK, 2010, https://npi-statistics.com/pdfs/theses/TM10.pdf.
[26] T. Coolen-Maturi, F. P. A. Coolen, and N. Muhammad, *Predictive inference for bivariate data: Combining nonparametric predictive inference for marginals with an estimated copula*, J. Stat. Theory Practice **10** (2016), 515–538.
[27] N. Muhammad, *Predictive Inference with Copulas for Bivariate Data*, Ph.D. thesis, Durham University, UK, 2016, https://npi-statistics.com/pdfs/theses/NM16.pdf.
[28] N. Muhammad, F. P. A. Coolen, and T. Coolen-Maturi, *Predictive inference for bivariate data with nonparametric copula*, in: The American Institute of Physics (aip) Conference Proceedings, vol. 1750, 2016, pp. 0600041–0600048, DOI: https://aip.scitation.org/doi/abs/10.1063/1.4954609.
[29] S.-F. Wu, *Bayesian interval estimation for the two-parameter exponential distribution based on the right type ii censored sample*, Symmetry **14** (2022), no. 2, 352.
[30] S. BinHimd, *Nonparametric Predictive Methods for Bootstrap and Test Reproducibility*, Ph.D. thesis, Durham University, UK, 2014, https://npi-statistics.com/pdfs/theses/SB14.pdf.
[31] S. BinHimd and F. P. A. Coolen, *On boostrapping using nonparametric predictive inference*, in: The 1st ISM International Statistical Conference (ISM 2012), Malaysia, 2012, pp. 465–468.
[32] F. P. A. Coolen and S. BinHimd, *Nonparametric predictive inference bootstrap with application to reproducibility of the two-sample Kolmogorov-Smirnov test*, J. Stat. Theory Practice **14** (2020), 1–13.
[33] A. S. M. Al Luhayb, *Smoothed bootstrap methods for right-censored data and bivariate data*, Ph.D. thesis, Durham University, 2021.
[34] A. S. M. Al Luhayb, F. P. A. Coolen, and T. Coolen-Maturi, *Generalizing banks smoothed bootstrap method for right-censored data*, in: 29th European Safety and Reliability Conference (ESREL 2019), Hannover, Germany, 2019, pp. 894–901.
[35] A. S. M. Al Luhayb, F. P. A. Coolen, and T. Coolen-Maturi, *Smoothed bootstrap for survival function inference*, in: Proceedings of the International Conference on Information and Digital Technologies (IDT 2019), Zilina, Slovakia, 2019, pp. 297–304.
[36] A. S. M. Al Luhayb, F. P. A. Coolen, and T. Coolen-Maturi, *Smoothed bootstrap for right-censored data*, Commun. Stat.-Theory Methods **52** (2023), 1–25.
[37] A. S. M. Al Luhayb, T. Coolen-Maturi, and F. P. A. Coolen, *Smoothed bootstrap methods for bivariate data*, J. Stat. Theory Practice **17** (2023), no. 3, 37.