Nonparametric predictive inference with right-censored data

Ke-Jian Yan

A Thesis presented for the degree of
Doctor of Philosophy

Statistics and Probability Group
Department of Mathematical Sciences
University of Durham
England

July 2002

To my parents, and my wife.

Nonparametric predictive inference with right-censored data

Ke-Jian Yan

Submitted for the degree of Doctor of Philosophy
July 2002

Abstract

This thesis considers nonparametric predictive inference for lifetime data that include right-censored observations.

The assumption $A_{(n)}$ proposed by Hill in 1968 provides a partially specified predictive distribution for a future observation given past observations. But it does not allow right-censored data among the observations. Although Berliner and Hill in 1988 presented a related nonparametric method for dealing with right-censored data based on $A_{(n)}$, they replaced 'exact censoring information' (ECI) by 'partial censoring information' (PCI), enabling inference on the basis of $A_{(n)}$. We address if ECI can be used via a generalization of $A_{(n)}$.

We solve this problem by presenting a new assumption 'right-censoring $A_{(n)}$' (rc-$A_{(n)}$), which generalizes $A_{(n)}$. The assumption rc-$A_{(n)}$ presents a partially specified predictive distribution for a future observation, given the past observations including right-censored data, and allows the use of ECI. Based on rc-$A_{(n)}$, we derive nonparametric predictive inferences (NPI) for a future observation, which can also be applied to a variety of predictive problems formulated in terms of the future observation.

As applications of NPI, we discuss grouped data and comparison of two groups of lifetime data, which are problems occurring frequently in reliability and survival analysis.

Declaration

This thesis is the result of research carried out by the author between October 1998 and June 2002 in the Department of Mathematical Sciences at the University of Durham, under the supervision of Dr. Frank Coolen. No part of this thesis has been submitted elsewhere for any other degree or qualification.

Chapters 1 and 2 contain necessary background material and no claim of originality is made. The remaining work is believed to be original. Chapters 3, 4 and 5 are based on joint work with my supervisor, Dr. Frank Coolen, and these can be found in [19, 20, 21], which have been submitted to 'Journal of Statistical Planning and Inference', 'Reliability Engineering and System Safety' and 'Statistics and Probability Letters', respectively. A summary of results in Chapter 3 can be also found in [22], which was presented at the Third International Conference on Mathematical Methods in Reliability, Trondheim (Norway), June 2002.

Acknowledgements

It is a pleasure to thank my supervisor, Dr. Frank Coolen, for the careful advice and guidance of my research. He checked my work very carefully, even my spelling and grammar. He was so generous with his time on supervising me. His instruction will continue to benefit my research work in the future.

I would like to thank the Department of Mathematical Sciences of the University of Durham for funding my research studies, and for providing me with good working conditions. In particular I would like to thank the Statistics Group for their pleasant and supportive help.

I would like to thank the Graduate Society of the University of Durham for support by providing an accommodation grant.

Finally, I also thank my family for their support during all these years.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

Statistical analysis of lifetime data is a topic of considerable interest in areas such as medicine and engineering. The field has developed rapidly in the past half century, and many statistical methods for lifetime data have been presented. By studying these methods, we can find that most earlier methods mainly involved parametric models. An advantage of parametric models is that they are often specified by only a few parameters. However, it is often difficult to derive such parametric models, which considerably affects the use of parametric models. A classical method presented by Kaplan and Meier [46] proposed a nonparametric method for lifetime data. After that, nonparametric methods have been widely applied in statistical analysis for lifetime data.

We know that, in the study of lifetime data, incomplete observations due to censoring often occur. As the most common form of incomplete observation, right-censored data are considered in most nonparametric methods, such as the Kaplan-Meier estimator of the survival function. Other nonparametric methods introduced in Chapter 2, such as the 'standard life table estimator' [48] for grouped data, and Mantel's test [50] for comparison of two groups of lifetime data, present how to deal with right-censored observations in different problem situations. These nonparametric methods have a common character on dealing with right-censored observations, that is they do not take all censoring times precisely into account.

All methods mentioned above are based on estimation of the survival functions, and they are not intended for prediction of a future observation, or other predictive inferences. Estimation is often important, but prediction also plays a key role in real decision-making processes [1, 10, 37]. Talking of prediction, we may consider Bayesian prediction. Conventional Bayesian methods yield a predictive posterior distribution, using a prior distribution for a parameter. But in this procedure, the selection of a statistical model and a prior distribution may be difficult. Particularly, if there is no appropriate model, Bayesian prediction becomes difficult. Hill [39] proposed the assumption $A_{(n)}$ for prediction in the case of extremely vague a prior knowledge about characteristics of the underlying source of observations, sometimes it is also called low structure Bayesian prediction [38]. Based on the assumption $A_{(n)}$, Berliner and Hill [6] presented a nonparametric predictive method based on lifetime data including right-censored observations. A disadvantage of their method is that they use so-called 'partial censoring information' instead of the exact censoring information, which makes a slight change to the censored data.

This thesis presents a new way to deal with censoring information in our nonparametric predictive methods. By generalizing Hill's $A_{(n)}$, a new assumption is presented, called 'right-censoring $A_{(n)}$'. Based on this new assumption, we obtain predictive inferences which take all censoring times precisely into account. At the same time, we can also extend these inferences to other predictive problems, such as grouped data and comparison of two groups of lifetime data.

1.2 Lifetime data

In statistical analysis, data may, for example, arise from the following situations: (1) The survival times of patients in a clinical trial; (2) The lifetimes of machine components in industrial reliability; (3) The duration of periods of unemployment in economics; (4) The lengths of tracks on a photographic plate in particle physics; (5) The number of years until death of people who have bought life assurance policies. The data in such situations are often referred to as 'lifetime data' even though the observations may not refer to lifetimes in the strictest sense. Mathematically we can think of a 'lifetime' as a one-dimensional non-negative random variable. Let T denote the lifetime random variable, then $T \in [0, \infty)$. Lifetime data are often encountered in both medicine and engineering applications such as those in cases (1) and (2) respectively. The study of lifetime data in engineering applications is normally referred to as reliability analysis, whilst in medicine we often talk about survival analysis. Cases (3) and (4) illustrate that lifetime data are encountered in a wide range of disciplines such as economics and science. Case (5) is an important consideration when setting premiums for life assurance policies.

In order to determine lifetime data precisely, three basic elements are needed: *(i) A starting point for measuring time (the time origin); (ii) A finishing point for measuring time (ending event of interest); (iii) A scale for measuring time.* The time origin can be viewed as the starting point of the measuring process. The individuals in the study may have different time origins. For example, most clinical trials have staggered entry, and each patient's lifetime is measured from their date of entry into the trial (first time), rather than from the date of the first entry into the trial. For the end point, first there must be a defined event related to particular time points. For example, in medical work, this event could be death from a specific cause (e.g. lung cancer) or the recurrence of a disease after treatment. The scale for measuring time is often real (clock) time, but could also be the operating time of a system, the mileage of a car, or some measure of cumulative load encountered.

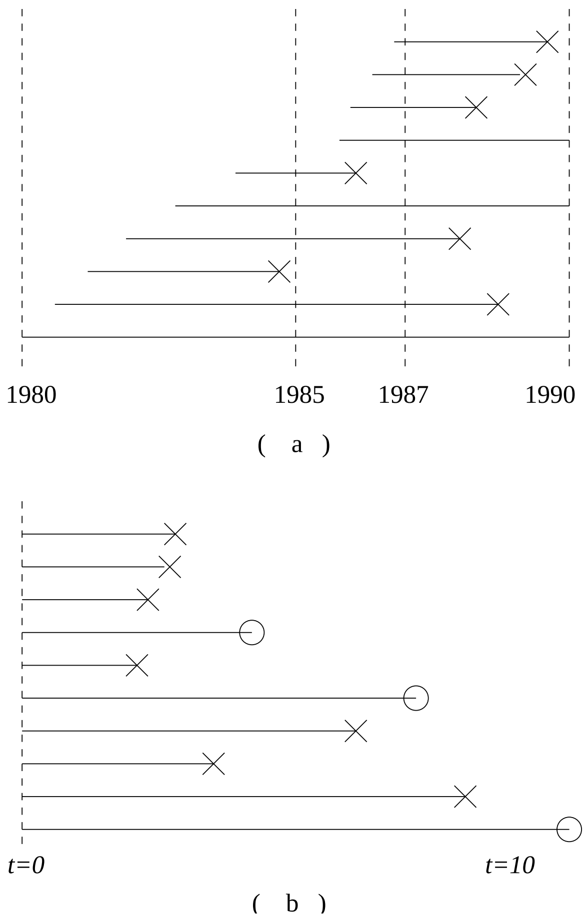

Figure 1.1: (a) Real time; (b) time T from entry ($\times$, death; $\bigcirc$, censoring).

Figure 1.1 (a) gives the real times for ten individuals with staggered entry and follow-up until 1990, and using death as point event. Figure 1.1 (b) illustrates the lifetimes for these ten individuals respectively. It should be noticed that seven of them are dead before 1990, and three of them are still alive at 1990. So we can obtain the exact lifetimes for those who are dead before 1990. For those who are still alive at 1990, we only know that their lifetimes exceed certain times. Such observations refer to a special feature of lifetime data, these are known as right-censored observations and will be discussed in the next section.

1.3 Censoring

We review censoring, closely following Lawless [48]. Censoring arises in various ways. Formally, an observation is said to be *right-censored* at c if it is only known that the lifetime is greater than c. For example, when a patient has been given a certain treatment, a right-censoring time might arise if the patient is still alive at the end of the time period set aside for observation. Similarly, an observation is said to be *left-censored* at c if it is known only that the observation is less than c; this situation might arise if a patient were put on test, but only checked for reaction every month. If at the first check after one month, the patient is found to have died, then we only know that his lifetime was less than one month. In this example, if the patient was found to have died between the second and third checks (that is, the patient was alive at the second check, but had died by the third check) then we would know that the patient had a lifetime between two and three months. This is an example of *interval censoring*. Obviously, right-censoring and left-censoring are two special types of interval censoring. As an incomplete observation in the study of lifetime data, right-censoring is the most common form. In this thesis, we present a nonparametric predictive method based on lifetime data, including right-censored observations. Throughout the thesis, except Section 3.7 (where left-censored observations are considered), we will refer to all lifetime data as ‘event time’, if it is a time at which the event of interest actually occurred, or ‘right-censoring time’.

On analysing censored data, there are some important assumptions about the nature of the censoring and its relationship to the event process. Following Meeker and Escobar [52], we describe these assumptions. First, a censoring time can be random, but it is often a predetermined value due to practical reasons. For example, in a life test experiment of n patients, a decision is made to terminate a study at a date on which not all patients' lifetimes will be known, then right-censored observations for such an experiment will occur. In order for standard censored data analysis methods to be valid, it is necessary that the censoring time of an observation depends only on the history of the observed event process. Using future events to stop observation could cause bias. The second assumption is that censoring is non-informative. For right-censoring, this means that such an event is only known not yet to have taken place at the corresponding right-censoring time, and no further information with regard to the corresponding event time is available.

As censored data are often encountered in collection of lifetime data, undoubtedly, we must be able to deal with it in statistical analysis. In Chapter 2, we will review some nonparametric methods, and discuss how they deal with right-censored data.

1.4 Outline of the thesis

This thesis considers nonparametric predictive inference for lifetime data including right-censored observations, based on the new assumption 'right-censoring $A_{(n)}$'. In Chapter 2 we briefly review some nonparametric methods presented for lifetime data and discuss how the right-censored data are dealt with in these methods. Hill's assumption $A_{(n)}$ is also reviewed in this chapter. In Chapter 3, we generalize Hill's $A_{(n)}$, and present the assumption right-censoring $A_{(n)}$ (rc-$A_{(n)}$). The assumption rc-$A_{(n)}$ and corresponding nonparametric predictive inference (NPI) are the main topic of this chapter, and indeed of this thesis. They present a new way for dealing with right-censored data in the nonparametric situation. In Chapter 4, we apply rc-$A_{(n)}$ and NPI to grouped data with right-censored observations. We also compare our method with alternative nonparametric methods. In Chapter 5, we apply rc-$A_{(n)}$ and NPI to predictive comparison of two groups of lifetime data including right-censored observations, and compare our approach with an alternative nonparametric method. Finally, we summarize our main results, along with some concluding remarks, in Chapter 6.

Chapter 2

Nonparametric inference and right-censored data

2.1 Introduction

Nonparametric methods are widely used in statistics. In practice, they are often attractive as they allow more flexibility than the use of parametric models. As Hill [40] remarked: 'In fact, nonparametric analyses represent the great majority of statistical situations, whilst parametric models are appropriate only in quite limited cases'.

In this chapter, we briefly review some nonparametric methods for lifetime data, and discuss how these methods deal with right-censored data. In Section 2.2, we introduce the classical nonparametric method by Kaplan and Meier [46]. As a nonparametric estimator of a population survival function, the Kaplan-Meier method [46] presents a tool for analyzing censored data. For nonparametric predictive analysis, the assumption $A_{(n)}$ has been proposed by Hill [39]. In Section 2.3 we present Hill's $A_{(n)}$ and briefly discuss possible inferences based on this assumption. Although $A_{(n)}$ does not apply to censored data, it provides an important tool for nonparametric predictive analysis. Later we will use this assumption to present our nonparametric predictive inference with right-censored data. Based on the assumption $A_{(n)}$, Berliner and Hill [6] present a nonparametric method for predictive analysis in case of right-censored data, which is described in Section 2.4. Section 2.5 discusses 'grouped data', focusing on two methods, the standard life table estimator [48] and the method by Coolen [13], based on Walley's [58] imprecise Dirichlet model. These two methods are also used to compare with our nonparametric method presented in

Chapter 4. Section 2.6 reviews comparison of two groups of lifetime data with right-censored observations, and Mantel's test [50] is considered in this section. Later in Chapter 5, Mantel's test is used to compare with our nonparametric method. Finally, in Section 2.7 we briefly add a few concluding remarks.

2.2 Kaplan-Meier estimator

In this section, we discuss the nonparametric estimator of the survival function for data including right-censored observations, presented by Kaplan and Meier [46], which is also known as the 'Product-Limit' (PL) estimator. This method is widely used, and presented in about all textbooks on survival analysis, e.g. [23, 45, 48, 53, 55].

Before we introduce the Kaplan-Meier method, we first give an important concept used in the method. If the events of interest are the deaths of individuals, the *risk set* at time t is the set of individuals known to be alive (i.e. alive and uncensored) at time t, and denoted as n_t. In this thesis, at an event or censoring time t, n_t does not include the individual corresponding to the observation, so n_t is always equal to the number of event and censoring times greater than t. In addition, we use $\tilde{n}_t$ to denote the number of individuals known to be alive just prior to t.

Suppose that there are observations on n individuals, and there are k $(k \leq n)$ distinct observed event times $t_1 < t_2 < \cdots < t_k$, where it is possible to have multiple events at t_j, let d_j be the number of events at t_j. In addition to the lifetimes $t_1, \ldots, t_k$, assume that there are $n - \sum_{j=1}^{k} d_j$ right-censored observations for individuals whose event times are not observed. Let there be l different right-censoring times, $c_1 < \ldots < c_l$. The PL estimator of the survival function, on the basis of these observed data, is

$$\hat{S}(t) = \prod_{j:t_j \leq t} \frac{\tilde{n}_{t_j} - d_j}{\tilde{n}_{t_j}}, \tag{2.1}$$

where $\tilde{n}_{t_j}$ is the number of individuals at risk just prior to t_j.

The PL estimator is a step function, which is constant on $[t_j, t_{j+1})$, for $j = 0, 1, \ldots, k-1$ with $t_0 = 0$, and decreases at event time t_j by a factor $(\tilde{n}_{t_j} - d_j)/\tilde{n}_{t_j}$. If the largest observation is at event time t_k, then the PL estimator is zero on $[t_k, \infty)$. If the largest observation is a right-censoring at c_l, then the PL estimator is a positive constant on $[t_k, c_l)$, but for interval $[c_l, \infty)$ it is often left undefined.

On the interval $[0, t_1)$, the PL estimator is equal to one. In the PL estimator, every drop of value happens at an event time, there is no change at censoring times. So we can say that censoring times do not have any direct effects on the PL estimate, their only effect is on the size of the later steps.

The PL estimator provides a nonparametric estimate of the survival function corresponding to the lifetime distribution for a population, and it is the Maximum Likelihood Estimator (MLE) [46], as such generalizing the empirical survival function in case of no censorings. It should be noted that, for the PL estimator to be the nonparametric MLE, the implicit assumption is made that attention is restricted to the class of all probability distribution functions [46]. The discrete model that underlines this estimator is described in detail by Lawless [48, Section 2.3].

Now we illustrate the PL estimator via an example.

Example 1

The data for this example are from the Dukes' C colorectal cancer patients of McIllmurray and Turkie [51]. The data are on survival of 24 patients with Dukes' C colorectal cancer randomly assigned to receive control treatment. These survival times are being measured in months, and given in Table 2.1, together with the PL estimator $\hat{S}(t)$. Figure 2.1 is a plot of the PL estimator.

t_j	$\tilde{n}_{t_j}$	d_j	$\hat{S}(t)$
0	24	0	1
6	23	4	0.8261
8	19	2	0.7391
12	17	2	0.6522
20	10	1	0.5780
24	8	1	0.5136
30	4	1	0.3852
42	1	1	0

Table 2.1: Dukes' C colorectal cancer survival data.

The Kaplan-Meier method is regularly used to graphically present data including right-censored observations. Many nonparametric methods for inference based on lifetime data, for example, the standard life table estimator [48] for grouped data, are related to this method.

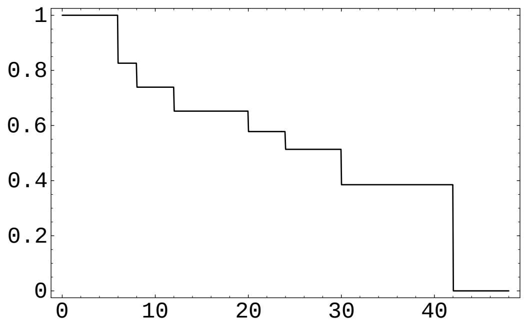

Figure 2.1: PL estimator of survival function (Example 1).

2.3 Assumption $A_{(n)}$ and imprecise probability

In this section, we discuss Hill's [39] assumption $A_{(n)}$, together with predictive inference based on this assumption. Our nonparametric methods which are presented in this thesis, are based on this assumption, and generalize it for the case of data including right-censored observations.

2.3.1 Overview of $A_{(n)}$

The assumption $A_{(n)}$ was proposed by Hill [39, 40], for prediction in the case of extremely vague a prior knowledge about characteristics of the underlying source of the observation. Let t_i, for $i = 1, \ldots, n$, be data values obtained by sampling from a population, and let $t_{(i)}$ be their ordered values (in increasing order of magnitude). Let T_i be the corresponding pre-data random quantities, so that the data consist of the observed values, $T_i = t_i,\ i = 1, \ldots, n$. Following Hill [42], $A_{(n)}$ is defined as follows.

1. The observable random quantities $T_1, \ldots, T_n$ are exchangeable. (In the original definition of $A_{(n)}$ [39], exchangeability was not included allowing more general situations.)

2. Ties have probability 0. (Generalization to include possible ties is straightforward, see Hill [40], but leads to more awkward notation.)

3. Given data t_i, $i = 1, \ldots, n$, the probability that the next observation falls in the open interval $I_j = (t_{(j)}, t_{(j+1)})$ is $1/(n+1)$, for all $j = 0, \ldots, n$, where we define $t_{(0)} = -\infty$ (or, for example, $t_{(0)} = 0$ when dealing with non-negative random quantities) and $t_{(n+1)} = \infty$.

It is clear that $A_{(n)}$ is a post-data assumption related to finite exchangeability [30], see Hill [40] for a detailed presentation and discussion of $A_{(n)}$, and an overview of related work, including important contributions by Dempster [31] and Lane and Sudderth [47]. Hill [42] presents a class of parametric models, called 'splitting processes', with a member which results exactly in $A_{(n)}$ as posterior predictive assuming finite additivity, hence providing a nonparametric Bayesian justification for $A_{(n)}$.

A natural interpretation of $A_{(n)}$ is in terms of ranks, namely the rank of the next observation amongst all observations will be equal to any possible value with probability $1/(n+1)$. Prior to the data $\{t_1, \ldots, t_n\}$, this is just an implication of exchangeability, so $A_{(n)}$ can be considered as a 'post-data version of exchangeability', the data carry information on location, but no information whatsoever on the rank of the future observation, which indeed corresponds to absence of prior knowledge.

De Finetti's representation theorem [30] uses a similar setting to justify a Bayesian framework for learning about an underlying parameter, and a probability distribution for that parameter, but he relies on the assumption that indeed there is an infinite sequence of random quantities involved, whereas our interest is mostly in inference on a single future observation. Even more, the Bayesian approach, as justified by De Finetti's [30] important results, explicitly needs a specified prior distribution, and together with the conditional independence of future observations (conditional on an unknown parameter) this adds quite a bit more structure to the data.

2.3.2 $A_{(n)}$ and imprecise probability

The assumption $A_{(n)}$ is not sufficient to derive precise probabilities for many possible events of interest. However, it does provide bounds for probabilities, by what is essentially an application of De Finetti's 'fundamental theorem of probability' [30] or Walley's 'natural extension' [57]. In this situation, some related predictive inferences, based on the assumption $A_{(n)}$, can be expressed using imprecise probability. In this subsection we review the related concepts and properties of imprecise probability, and describe $A_{(n)}$-based imprecise probabilities which are bounds for the predictive survival function in case of no censoring.

(I) Imprecise probability

The idea to use imprecise probabilities dates back at least to the middle of the nineteenth century [8]. Since then, the use of imprecise probabilities has been suggested in many areas of statistics. Recently, there has been increasing activity in this area by researchers from widely varying backgrounds, resulting in a series of conferences [27, 28], special issues of journals [7, 24, 26] and a webpage [29].

Extending De Finetti's theory [30] to imprecise probability, or more generally imprecise previsions, Walley [57] provides a rigorous generalization of the concept of probability, based on a behavioural interpretation of subjective imprecise probability as bets with possibly differing maximum buying price $\overline{P}$ and minimum selling price $\underline{P}$. Augustin and Coolen [4] propose an expression for imprecise probability. According to such an expression, the imprecise probability, for an event of interest A, can be expressed by two optimal bounds,

$$\underline{P}(A) = \inf P(A),$$

$$\overline{P}(A) = \sup P(A).$$

An important consequence for these two bounds is that $\underline{P}(A)$ and $\overline{P}(A)$ are conjugate,

$$\underline{P}(A) = 1 - \overline{P}(A^c), \tag{2.2}$$

where A^c is the complementary event to A. The conjugacy property can often be used to simplify the calculation of imprecise probabilities for events of interest and their complementary events (we will use this in Chapter 5).

Here we mainly introduced the related concepts and conjugacy property of imprecise probability. They will be referred to throughout this thesis. For a complete introduction and overview of imprecise probability, we refer to Walley [57].

(II) Imprecise survival functions

Now we illustrate how imprecise probabilities are derived for $T_{n+1} > t$, giving imprecise survival functions based on the assumption $A_{(n)}$ [17].

The survival function represents the probability for an individual of surviving past a certain moment of time. The survival function for an individual with random positive lifetime T is defined as $S_T(t) = P(T > t)$. Assuming observed event times for n individuals, ordered as $t_{(1)} < t_{(2)} < \ldots < t_{(n)}$, and denoting $t_{(0)} = 0$ and $t_{(n+1)} = \infty$, the assumption $A_{(n)}$ gives direct predictive probabilities for the lifetime T_{n+1} of a further individual, at $t_{(j)}$ this leads to predictive survival function for T_{n+1} equal to

$$S_{T_{n+1}}(t_{(j)}) = \frac{n+1-j}{n+1}, \quad \text{for} \quad j = 0, \ldots, n.$$

Without further assumptions it is not possible to give a precise value for this survival function at times other than previously observed event times, as $A_{(n)}$ assigns probability mass $1/(n+1)$ to the open intervals between observed event times, and to the intervals $[0, t_{(1)})$ and $(t_{(n)}, \infty)$, but does not put any further restrictions on the distribution of the probability mass within each such interval. Therefore, the only inference we can derive at, without additional assumptions, consists of lower and upper bounds for the survival function, where we aim at deriving the maximum lower bound, denoted by $\underline{S}$, and the minimum upper bound, denoted by $\overline{S}$, which are consistent with the probability assessment according to $A_{(n)}$. To derive $\underline{S}(t)$, one can shift the probability mass in the interval in which t lies to the left end-point of that interval (i.e. to the infimum value of the open interval), leading to

$$\underline{S}_{T_{n+1}}(t) \;=\; S_{T_{n+1}}(t_{(j+1)}) \;=\; \frac{n-j}{n+1} \quad \text{for} \quad t \in (t_{(j)}, t_{(j+1)}), \ \text{with} \ j = 0, \ldots, n.$$

This $\underline{S}_{T_{n+1}}(t)$ is the optimal lower bound of $S_{T_{n+1}}(t)$ based on $A_{(n)}$, without additional assumptions. We call this $\underline{S}$ the *lower survival function* for T_{n+1}. Similarly, one derives the optimal upper bound of $S_{T_{n+1}}(t)$, called *upper survival function* $\overline{S}$, by shifting the probability mass per interval to the right end-point (the supremum of the open interval), leading to

$$\overline{S}_{T_{n+1}}(t) \;=\; S_{T_{n+1}}(t_{(j)}) \;=\; \frac{n+1-j}{n+1} \quad \text{for} \quad t \in (t_{(j)}, t_{(j+1)}), \ \text{with} \ j = 0, \ldots, n.$$

Notice that, for any $t > t_{(n)}$, we have $\underline{S}_{T_{n+1}}(t) = 0$ and $\overline{S}_{T_{n+1}}(t) = 1/(n+1)$, while for any value t in $(0, t_{(1)})$ we have $\underline{S}_{T_{n+1}}(t) = n/(n+1)$ and $\overline{S}_{T_{n+1}}(t) = 1$.

Example 2 illustrates the lower and upper survival functions based on the assumption $A_{(n)}$.

Example 2

The following data are the ordered numbers of millions of revolutions to failure for each of 23 ball bearings [25, Section 2.9].

17.88	28.92	33.00	41.52	42.12	45.60	48.40	51.84
51.96	54.12	55.56	67.80	68.64	68.64	68.88	84.12
93.12	98.64	105.12	105.84	127.92	128.04	173.40	

Based on these 23 observations, the assumption $A_{(23)}$ provides predictive probabilities for T_{24} as described above, leading to lower and upper survival functions for T_{24} as given in Figure 2.2. It should be remarked that there are two tied observations, at 68.64. Although $A_{(n)}$ is presented assuming no ties in the data, it can be seen that we now get a predictive point probability $P(T_{24} = 68.64) = 1/24$. We can think of these tied observations as being not really identical, with the tie being caused by rounding, with the probability for the very small interval between such two observations still equal to 1/24. One may, of course, doubt the correctness of such a predictive point probability for an apparently continuous random quantity. However, this point probability is actually caused by two identical values already observed, so it merely indicates that this value could well occur again.

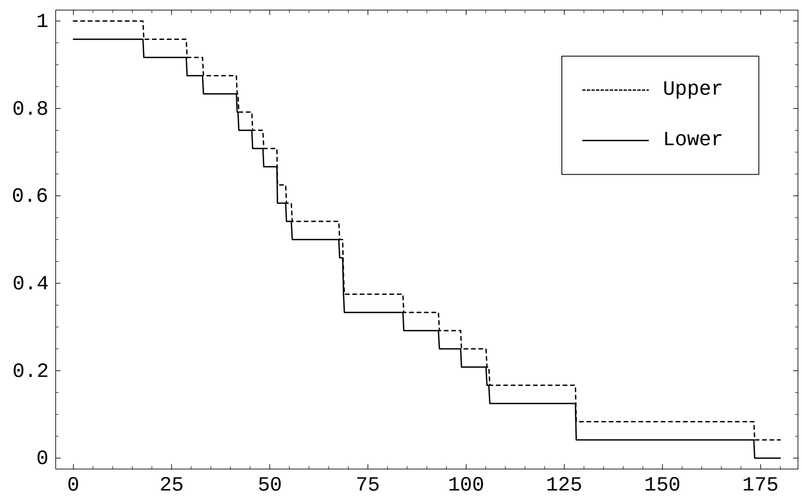

Figure 2.2: Ball bearings example: survival functions for T_{24}.

(III) Inference based on $A_{(n)}$

The position of $A_{(n)}$-based inference in the theory of imprecise probability has been studied in detail by Augustin and Coolen [4]. These inferences have a predictive and nonparametric nature, which is referred to as nonparametric predictive inference [4]. Several examples of $A_{(n)}$-based nonparametric predictive inference have been presented, e.g. [3, 15, 16, 18].

Inferences based on $A_{(n)}$ seem suitable if there is hardly any knowledge about the random quantities of interest, other than the first n observations, or, which may be more realistic, if one explicitly does not want to use such information. This may occur, for example, if one wants to study the (often hidden) effects of additional structural assumptions underlying statistical models or methods. Inferences based on such restricted knowledge have also been called 'low structure inferences' [38] and 'black-box inferences' [47]. In addition, $A_{(n)}$-based inferences are entirely flexible, valid for few data, although high imprecision may be the fair price of only little information, and valid for many data as its asymptotics are closely related to those of the empirical distribution function.

2.4 Berliner-Hill method

By using the assumption $A_{(n)}$, Berliner and Hill [6] presented a nonparametric predictive method on the basis of data including right-censored observations.

Let $T_1, \ldots, T_n$ be observable random quantities, assume that they are exchangeable, and that ties have probability 0. Suppose we have observations from these n random quantities, consisting of u event times and $v = n - u$ right-censored observations. Let $t_{(1)} < t_{(2)} < \ldots < t_{(u)}$ denote the order statistics for observed event times, and $c_{(1)} < c_{(2)} < \ldots < c_{(v)}$ denote the order statistics for the right-censoring times. For convenience, let the random quantities $T_1, \ldots, T_u$ correspond to the event times $t_{(1)}, \ldots, t_{(u)}$, and the random quantities T_{u+j}, for $j = 1, \ldots, v$, correspond to the v censored observations $c_{(1)}, \ldots, c_{(v)}$. So the data consist of the survival times $T_i = t_{(i)}$, for $i = 1, \ldots, u$, and censoring times $T_{u+j} \geq c_{(j)}$, for $j = 1, \ldots, v$. Let T_{n+1} denote the next observation.

The censoring information provided by the right-censoring times, is called *exact censoring information* (ECI), and denoted as

$$\text{ECI} = \{T_{u+j} \geq c_{(j)} : j = 1, \ldots, v\}.$$

A further concept, called *partial censoring information* (PCI), is used in the Berliner-Hill method. For each censored observation $c_{(j)}$, $j = 1, \ldots, v$, let $\tilde{t}_j$ be the largest observed event time (or 0) smaller than $c_{(j)}$. Then PCI is defined as

$$\text{PCI} = \{T_{u+j} \geq \tilde{t}_j : j = 1, \ldots, v\}.$$

Following the Berliner-Hill method [6], predictive probabilities for the next observation can be derived as below.

Assuming $A_{(n)}$, let l_i denote the number of censored observations in interval $(t_{(i)}, t_{(i+1)})$, and $\tilde{l}_i = \sum_{k=0}^{i} l_k$ for $i = 0, 1, \ldots, u$, then the Berliner-Hill method specifies the following predictive probabilities

$$\begin{aligned} P(T_{n+1} \in (0, t_{(1)}) \mid PCI) &= \lambda_0, \\ P(T_{n+1} \in (t_{(i)}, t_{(i+1)}) \mid PCI) &= (1-\lambda_0) \times \cdots \times (1-\lambda_{i-1}) \times \lambda_i, \\ &\quad \text{for } i = 1, \ldots, u, \end{aligned}$$

where

$$\lambda_i = \frac{1}{n-(i-1)-\tilde{l}_i}, \quad \text{for } i = 0, 1, \ldots, u.$$

Berliner and Hill [40] use PCI instead of ECI in their method, which allows them to deal with the censoring information by computation of the appropriate conditional probabilities for T_{n+1}, conditioned on the observed event times and PCI for the random quantities corresponding to the censoring times, and using $A_{(n)}$ without further assumptions.

Berliner and Hill [40] also give upper and lower bounds of predictive probabilities for the next observation T_{n+1}. The upper bound is obtained by moving the censored observations in an interval $(t_{(i)}, t_{(i+1)})$ just to the right of its left end-point, which is identical to replacing ECI by PCI. The lower bound is obtained by moving the censored observations in $(t_{(i)}, t_{(i+1)})$ just to the right of its right end-point. Although indeed this provides bounds for predictive probabilities for the next observation T_{n+1}, it adds some information to the data, which is not justified by these data.

Berliner and Hill [40] present a survival function based on PCI by assuming that the probability mass is uniform per interval, which leads to a continuous survival function. Let us denote this 'uniform Berliner-Hill survival function' by $S^{BH}_{T_{n+1}}(t)$, then

$$S^{BH}_{T_{n+1}}(t) = S^{BH}_{T_{n+1}}(t_{(i)}) - \frac{t - t_{(i)}}{t_{(i+1)} - t_{(i)}} P(T_{n+1} \in (t_{(i)}, t_{(i+1)}) \mid PCI),$$

for $t \in (t_{(i)}, t_{(i+1)})$, where $S^{BH}_{T_{n+1}}(t_{(i)}) = 1 - \sum_{j=1}^{i} P(T_{n+1} \in (t_{(j-1)}, t_{(j)}) \mid PCI)$ with $S^{BH}_{T_{n+1}}(t_{(0)}) = 1$, for $t_{(0)} = 0$. Obviously, the uniform Berliner-Hill survival function beyond the largest event time is influenced by the choice of an upper bound for the random quantity T_{n+1}. Without such an upper bound, a uniform Berliner-Hill survival function cannot be defined on this interval.

Example 3

Suppose that we have three event times, 2, 3, 12, and five right-censoring times, 9, 10, 10.5, 11, 11.5. We assume $A_{(8)}$. Table 2.2 gives the predictive probabilities for T_9 according to the uniform Berliner-Hill method as outlined above.

$(t_{(i)}, t_{(i+1)})$	$P(T_9 \in (t_{(i)}, t_{(i+1)}) \mid PCI)$	$S^{BH}_{T_{n+1}}(t)$
$(0, 2)$	0.111	$1 - 0.056t$
$(2, 3)$	0.111	$1.111 - 0.111t$
$(3, 12)$	0.389	$0.907 - 0.043t$
$(12, \infty)$	0.389	——

Table 2.2: The predictive probabilities and survival function, according to the uniform Berliner-Hill method.

Consider $t = 8$. There are two event times less than 8, and 5 censoring times and one event time greater than 8. This means that 6 out of 8 individuals will be at risk at the time 8. So, intuitively, the predictive survival function for T_9 should be larger than the result derived in the example, $S_{T_9}(8) = 0.563$. The uniform Berliner-Hill method uses PCI, that the random quantities, corresponding to censoring times, exceed $t_2 = 3$, instead of exceeding $c_1 = 9,\ c_2 = 10,\ c_3 = 10.5,\ c_4 = 11,\ c_5 = 11.5$, respectively, and assumes that the probability mass in each open interval between event times is uniformly distributed. It should be noticed that because there is not a finite upper bound for the observations, the uniform Berliner-Hill survival function is not defined on the interval $(12, \infty)$ based on such a uniform assumption.

2.5 Nonparametric methods for grouped data with right-censoring

2.5.1 Introduction

In reliability and survival analysis, data are frequently recorded in groups, with the time-axis partitioned into a finite number of intervals, and the data only consisting of numbers of event times and numbers of censoring times per interval. A well-known example of such data is the use of so-called 'life tables' [48]. In reliability contexts, such data may typically appear on lifetimes of non-critical components in systems, where e.g. once a month the components are inspected, showing if they have failed or not. In such situations, right-censoring could be due to component failures caused by competing risks, which are not the main failure modes under consideration, or by components being replaced due to a predetermined preventive replacement policy. Grouping data is one of the most widely used methods of portraying lifetime data. Although grouped data have been used for a long time, the elaboration of their statistical properties has been a much more recent development because of the problems that censoring introduces [48].

Suppose the time-axis is divided into $k+1$ intervals, $I_z = [a_z, a_{z+1})$, $z = 0, 1, \ldots, k$, with $a_0 = 0$, and $a_{k+1} = \infty$. For each member of a random sample of n individuals from the population, suppose that one observes either an event time or a right-censoring time. However, the data are grouped, so only the numbers of event times and censoring times in intervals I_z are known, and not the exact event times and censoring times. Let e_z be the number of event times in I_z, and c_z the number of right-censoring times in I_z. Let $e = \sum_{z=0}^{k} e_z$ and $c = \sum_{z=0}^{k} c_z$, so $e + c = n$.

Based on grouped data, nonparametric methods are presented. Lawless [48] describes the so-called standard life table estimator based on grouped data including right-censored observations, which is the nonparametric maximum likelihood estimator. However, in some sense, this method is arbitrary in adjustment of censoring mechanism, by effectively assuming that censorings took place at the middle of the interval. Other nonparametric methods, such as presented by Elveback [33] and Chiang [11], are derived on more formal grounds than the standard life table estimator, but there is still quite some arbitrariness in the adjustment to censoring. When there are relatively few censored data, or time intervals are not wide, there is not much difference between estimators such as Elveback's, Chiang's and the standard life table estimator. Coolen [14] adapted Walley's [58] imprecise Dirichlet model for grouped data including right-censored observations. In this method, censorings are assumed to take place at right-hand points of the time intervals. In this section, we mainly introduce two methods, the standard life table estimator [48], and the method by Coolen [13]. We will compare our new method for such data with these two methods in Chapter 4.

2.5.2 The standard life table estimator

We review the standard life table estimator, closely following Lawless [48]. Let the underlying survival function for grouped data be $S(t)$. For $I_z = [a_z, a_{z+1})$, we define:

$$p_z = P(T \notin I_z \mid T \geq a_z) \quad \text{and} \quad q_z = P(T \in I_z \mid T \geq a_z).$$

Obviously, $q_z = 1 - p_z$, and $p_z = S(a_{z+1})/S(a_z)$, so the survival function at a_{z+1} is

$$S(a_{z+1}) = p_0 p_1 \cdots p_z, \quad \text{for } z = 0, 1, \ldots, k. \tag{2.3}$$

Let n_{a_z} be the number of individuals at risk at a_z. The idea of the standard life table estimator is to employ (2.3) in obtaining an estimate of $S(a_{z+1})$, via the estimates of q_z and p_z. The usual procedure is as follows.

If there are no censored observations in I_z, then an estimate of q_z is $\hat{q}_z = e_z/n_{a_z}$. However, if there are censored observations in I_z, e_z/n_{a_z} might be expected to underestimate q_z. Therefore, an adjustment is required due to the censored observations. The standard life table estimator uses the following estimate of q_z in the situation that there are censored observations in I_z, that is

$$\hat{q}_z = \frac{e_z}{n_{a_z} - c_z/2} = \frac{e_z}{n'_{a_z}}.$$

The denominator $n'_{a_z} = n_{a_z} - c_z/2$ can be thought of as an effective number of individuals at risk over I_z. Once estimates $\hat{q}_z$ and $\hat{p}_z = 1 - \hat{q}_z$ have been calculated, we can estimate $S(a_{z+1})$ by $\hat{S}(a_{z+1}) = \hat{p}_0 \hat{p}_1 \cdots \hat{p}_z$ with $z = 0, 1, \ldots, k$.

It should be remarked that the adjustment for dealing with censored observations is quite arbitrary in the standard life table estimator. In some situations other estimates of q_z may be preferable. For example, if all censored observations in I_z are close to a_{z+1}, the estimate $\hat{q}_z = e_z/n_{a_z}$ might be more appropriate, whereas if all censored observations in I_z are close to a_z, $\hat{q}_z = e_z/(n_{a_z} - c_z)$ might be more appropriate. Clearly, any adjustments for dealing with censored observations effectively adds some additional information to grouped data, which is not justified by these data. Our method presented for grouped data in Chapter 4 does not need to add such assumption for censored observations within I_z.

Example 4

Table 2.3 gives the standard life table estimator of the survival function for grouped lifetime data, given by Lawless [48].

The example illustrates that standard life table estimator gives a survival function estimate at points a_z, for $z = 0, 1, \ldots, k$, with $\hat{S}(0) = 1$. The estimator of the

$I_z = [a_z, a_{z+1})$	n_{a_z}	e_z	c_z	n'_{a_z}	$\hat{q}_z$	$\hat{p}_z$	$\hat{S}(a_{z+1})$
$[0, 1)$	356	60	0	356	0.1685	0.8315	0.8315
$[1, 2)$	296	48	0	296	0.1622	0.8378	0.6966
$[2, 3)$	248	30	0	248	0.1210	0.8790	0.6123
$[3, 4)$	218	28	35	200.5	0.1397	0.8603	0.5268
$[4, 5)$	155	19	49	130.5	0.1456	0.8544	0.4501
$[5, 6)$	87	12	41	66.5	0.1804	0.8196	0.3689
$[6, \infty)$	34	34	0	34	1	0	0

Table 2.3: The standard life table estimate.

survival function at ∞ is assumed to be equal to 0.

2.5.3 The imprecise Dirichlet model for grouped data

We review the imprecise Dirichlet model for grouped data, closely following Coolen [13]. In this method, censorings are dealt with assuming that they are at times a_z, i.e. the censoring times within interval I_z are assumed to take place at the right-end point a_{z+1} of this interval.

Walley [58] introduced an imprecise Dirichlet model related to multinomial data. Let the multinomial model have parameter vector $\theta = (\theta_0, \theta_1, \cdots, \theta_k)$, with $\sum_{z=0}^{k} \theta_z = 1$ and all $\theta_z \geq 0$, and

$$P(T \in I_z \mid \theta) = \theta_z, \quad \text{for } z = 0, 1, \ldots, k.$$

In a Bayesian framework, a Dirichlet distribution is a conjugate prior for this parameter θ. A Dirichlet prior distribution is specified by the density function

$$\pi(\theta \mid s, t) \propto \prod_{z=0}^{k} \theta_z^{st_z - 1},$$

with $t_z > 0$ for $z = 0, 1, \ldots, k$ and $\sum_{z=0}^{k} t_z = 1$, and s is a parameter with $s > 0$. This prior distribution is uniquely determined by (s, t). Combining this prior distribution with the likelihood, based on e event times and c censoring times, leads to a posterior distribution as $f(\theta|e, c, s, t)$. This posterior distribution is a generalised Dirichlet distribution, as introduced by Connor and Mosimann [12], and analyzed in detail by Lochner [49]. For statistical inference about θ_z in the Bayesian framework, the expected value $E(\theta_z|e, c, s, t)$ according to the posterior distribution can be obtained, see van Noortwijk *et al.* [56]. Clearly, $E(\theta_z|e, c, s, t)$ is a set of expected values for

θ_z, since it is determined by (s,t). Based on such a set of expected values for θ_z, Coolen [14] derives the optimal lower and upper bounds for the expected value of θ_z,

$$\underline{E}(\theta_z|e,c,s) = \inf_t \Big\{ E(\theta_z|e,c,s,t) \mid t_i > 0,\ \sum_{i=0}^{k} t_i = 1 \Big\}$$

$$\overline{E}(\theta_z|e,c,s) = \sup_t \Big\{ E(\theta_z|e,c,s,t) \mid t_i > 0,\ \sum_{i=0}^{k} t_i = 1 \Big\},$$

as

$$\underline{E}(\theta_0|e,c,s) = \frac{e_0}{e+c+s}$$

$$\underline{E}(\theta_z|e,c,s) = \frac{e_z}{\sum_{j=z}^{k}(e_j+c_j)+s} \times \prod_{h=0}^{z-1} \Big\{ \frac{\sum_{j=h}^{k-1}(e_{j+1}+c_j)+s}{\sum_{j=h}^{k}(e_j+c_j)+s} \Big\}, \quad z = 1,\dots,k-1$$

$$\underline{E}(\theta_k|e,c,s) = \frac{e_k+c_{k-1}}{e_k+e_{k-1}+c_{k-1}+s} \times \prod_{h=0}^{k-2} \Big\{ \frac{\sum_{j=h}^{k-1}(e_{j+1}+c_j)+s}{\sum_{j=h}^{k}(e_j+c_j)+s} \Big\}$$

and

$$\overline{E}(\theta_0|e,c,s) = \frac{e_0+s}{e+c+s}$$

$$\overline{E}(\theta_z|e,c,s) = \frac{e_z+s}{\sum_{j=z}^{k}(e_j+c_j)+s} \times \prod_{h=0}^{z-1} \Big\{ \frac{\sum_{j=h}^{k-1}(e_{j+1}+c_j)+s}{\sum_{j=h}^{k}(e_j+c_j)+s} \Big\}, \quad z = 1,\dots,k.$$

The choice of s is discussed in detail by Walley [58], who shows that, when attempting to model a lack of prior information, the choices $s = 1$ or $s = 2$ are reasonably cautious. The choice $s = 0$ would reduce the imprecise Dirichlet model to a precise Dirichlet model.

Example 5

The example is from Coolen [14]. Suppose that the partition of the time-axis consists of $k = 5$ intervals with $a_1 = 2$, $a_2 = 4$, $a_3 = 6$ and $a_4 = 8$. The number of event times in every interval are 1,0,0,2,0, and the number of censoring times 1,3,5,3,0. Table 2.4 gives the optimal lower and upper bounds of $E(\theta_z|e,c,s,t)$, for such grouped data.

The bounds for the expected value of θ_z in the imprecise Dirichlet model are Bayesian imprecise predictive probabilities for a future observation.

	[0,2)	[2,4)	[4,6)	[6,8)	[8,∞)
$\overline{E}(\theta_z\|e,c,s=2)$	0.1765	0.1255	0.1569	0.5378	0.6723
$\overline{E}(\theta_z\|e,c,s=1)$	0.1250	0.0670	0.0852	0.4688	0.6250
$\underline{E}(\theta_z\|e,c,s=1)$	0.0625	0	0	0.3125	0.4688
$\underline{E}(\theta_z\|e,c,s=2)$	0.0588	0	0	0.2689	0.4034

Table 2.4: Bounds of $E(\theta_z|e,c,s,t)$ for Example 5.

2.6 Nonparametric methods for comparison of two groups of lifetime data

2.6.1 Introduction

Comparison of two groups of lifetime data including right-censored observations is often required, for example in medical applications. For such comparison, an often used method is via parametric models for lifetime data, such as exponential distributions, and then testing the equality of parameters. Alternative nonparametric method is also often used, such as Mantel's test [50], Gehan-Breslow's test [35, 36], and Breslow's test [9]. These nonparametric methods compare the unknown survival functions from two groups of lifetime data, by testing a null hypothesis of equal survival functions.

Coolen [13] presented a nonparametric method for comparison of two different groups via predictive inferences for a future observation, based on $A_{(n)}$, but this did not allow censored data. In Chapter 5, we will generalize the method by Coolen [13], allowing right-censored data. In this section we briefly discuss Mantel's test [50], which we will compare with our method in Chapter 5.

2.6.2 Mantel's test

We review Mantel's test [50] for comparison of two groups of lifetime data, closely following Hollander and Wolfe [43].

Suppose that there are n_a and n_b observations in groups A and B, respectively. Let S_a denote the underlying survival function of group A, and S_b the underlying survival function of group B. Now combine the lifetime data from the two groups together and let $t_1 < t_2 < \cdots < t_m$ be the distinct event times of these two groups. Let $\tilde{n}_{a,t_k}$ ($\tilde{n}_{b,t_k}$) be the number of individuals from group A (B) who were at risk just before time t_k, and let $\tilde{n}_{t_k} = \tilde{n}_{a,t_k} + \tilde{n}_{b,t_k}$, for $1 \leq k \leq m$. Let $d_{a,k}$ ($d_{b,k}$) be the number of event times from group A (B) at t_k, and let $d_k = d_{a,k} + d_{b,k}$, for $1 \leq k \leq m$. Under null hypothesis $H_0 : S_a = S_b$, the statistic

$$M_c = \frac{\sum_{k=1}^{m}(d_{a,k} - E_{a,k})}{\sqrt{\sum_{k=1}^{m} V_{a,k}}},$$

where

$$E_{a,k} = \frac{d_k \tilde{n}_{a,t_k}}{\tilde{n}_{t_k}}$$

and

$$V_{a,k} = \frac{d_k(\tilde{n}_{t_k} - d_k)\tilde{n}_{a,t_k}\tilde{n}_{b,t_k}}{\tilde{n}_{t_k}^2(\tilde{n}_{t_k} - 1)}$$

has approximately a $N(0,1)$ distribution, if n_a and n_b are not too small and there are not too many censorings. The comparison of S_a and S_b is given by testing statistic, which is described as below,

1. One-side test of H_0 against alternatives for which survival times for group B tend to be longer than those for group A. To test at the approximate α-level of significance, if $M_c \geq z_\alpha$ (the critical value of α significance level), then reject H_0, otherwise do not reject;

2. One-side test of H_0 against alternatives for which survival times for group A tend to be longer than those for group B. To test at the approximate α-level of significance, if $M_c \leq -z_\alpha$ then reject H_0, otherwise do not reject;

3. Two-side test of H_0 against alternatives for which survival times for group B have a different distribution than that for group A. To test at the approximate α-level of significance, if $|M_c| \geq z_{\frac{\alpha}{2}}$, then reject H_0, otherwise do not reject.

This is Mantel's [50] test. We illustrate this method via an example by Hollander and Wolfe [43, Section 11.7].

Example 6

The data in Table 2.5 are from a clinical trial on Hodgkin's disease, a cancer of the lymph system. We will also consider these data in Chapter 5. The following two treatments were considered, (A) radiation treatment of the affected node and, (B) radiation treatment of the affected node plus all nodes in the trunk of the body. The data represent the relapse-free survival times in days. If a relapse had not occurred before the end of the data analysis, then the observation for that patient is right-censored.

Treatment A		Treatment B	
86	822	173	> 1726
107	836	498	> 1763
141	> 1309	615	> 1807
296	1375	950	> 1879
312	> 1378	> 1190	> 1889
330	> 1446	> 1242	> 1897
346	> 1540	1408	> 1968
364	> 1645	> 1493	> 1972
401	> 1818	> 1572	> 2022
419	> 1910	> 1576	> 2070
505	> 1953	> 1585	> 2177
570	> 2052	> 1684	
688		> 1699	

Table 2.5: Relapse-free survival times for Hodgkin's disease patients ($> t$ indicates right-censoring at t).

The test statistic in Mantel's test is $M_c = 3.25$, which gives an approximate one-sided P-value of 0.0006. Thus there is strong evidence that total nodal radiation is more effective than radiation of affected nodes in preventing or delaying the recurrence of early stage Hodgkin's disease.

In Mantel's test, the censoring times within interval (t_k, t_{k+1}) do not have any direct effects on the calculation of $E_{a,k}$, their effects are on the calculation of $E_{a,k+1}$. We can say that Mantel's test is similar to Kaplan-Meier estimator in dealing with censored observations.

2.7 Remarks

Statistical inference related to informative censoring is an important topic, both theoretically and related to application. However, as remarked by Coolen [14], it seems that only quite complicated model assumptions or a direct subjective approach are suitable to deal with the kind of information on lifetimes that may arise.

A successful alternative statistical approach for lifetime data has been developed based on theory of counting processes and martingales, an excellent overview is given by Andersen, *et al* [2]. The novel methods presented in this thesis are not directly related to counting processes and martingales, comparison with such methods is an interesting topic for future research.

From the above discussion, either common nonparametric methods or classical statistical methods, they are all not capable of taking censoring times precisely into account, when dealing with the censoring information resulting from such non-informative censoring mechanism. In the following chapters, we present a novel nonparametric predictive method for dealing with right-censored data, based on a non-informative censoring assumption.

Chapter 3

Right-censoring $A_{(n)}$ and nonparametric predictive inference

3.1 Introduction

In this chapter a new nonparametric predictive method is presented based on data including right-censored observations. Basically, the method is an attempt to learn about a future observation from past observations, including right-censored data, while adding only few additional structural assumptions. The method is based on Hill's assumption $A_{(n)}$ [39, 40]. However, the presence of right-censored data requires further attention, which is the main topic of this chapter. In Section 3.2, some further assumptions related to $A_{(n)}$ are introduced and justified. Based on these assumptions, Section 3.3 presents a new assumption, which is called 'right-censoring $A_{(n)}$' (rc-$A_{(n)}$). Section 3.4 presents 'nonparametric predictive inference' (NPI) for a new observation, based on rc-$A_{(n)}$. In Section 3.5 this new inferential method is compared to the established methods by Berliner and Hill [6] and Kaplan and Meier [46]. Throughout the first five sections of this chapter we assume that there are no ties present in the data, to keep notation relatively straightforward. However, in Section 3.6 the possibilities for the treatment of ties are discussed. Section 3.7 considers an application of this new inferential method for data sets including left-censored observations, using our method for right-censored data and a monotone only decreasing transformation of the data.

3.2 Preparing for right-censoring $A_{(n)}$

3.2.1 Introduction

In Section 2.3 the assumption $A_{(n)}$ [39, 40, 42] was discussed, which provides a partially specified predictive distribution for a future observation given past observations, consisting of exact event times. For lifetime data there are often right-censored data among the observations. Although Berliner and Hill [6] present a related nonparametric method for dealing with right-censored data based on $A_{(n)}$, they replace 'exact censoring information' (ECI), i.e. the exact observed censoring times, by 'partial censoring information' (PCI), in effect shifting each exact censoring time back to the nearest smaller observed event time, enabling inference on the basis of $A_{(n)}$ alone. The question addressed in this thesis is if ECI can be used, via a generalization of $A_{(n)}$.

In this chapter, the assumption $A_{(n)}$ is extended via a generalization called 'right-censoring $A_{(n)}$' (rc-$A_{(n)}$), which presents a partially specified predictive distribution for a future observation, given the past observations including right-censored data, and indeed allows the use of ECI. As the preparation for this generalization of $A_{(n)}$, this section presents some assumptions related to $A_{(n)}$. In Subsection 3.2.2 an assumption denoted by $\tilde{A}_{(n)}$ is presented, related to $A_{(n)}$, which assumes that the predictive distribution for a future observation consists of probability masses defined on two kinds of open interval, one is formed by consecutive event times and the other is formed by a censoring time and infinity. Dealing with the probability masses on intervals formed by censoring times and infinity requires further attention. Subsection 3.2.3 presents a further assumption called 'shifted-$\tilde{A}_{(n)}$', which presents a partially specified predictive distribution for the random quantity related to a right-censored observation in case of a 'non-informative censoring' assumption, related to exchangeability of a right-censored observation with all other random quantities in the risk set at the censoring time. In Subsection 3.2.4 an assumption called 'right-censoring $\tilde{A}_{(n)}$' (rc-$\tilde{A}_{(n)}$) is presented, based on $\tilde{A}_{(n)}$ and shifted-$\tilde{A}_{(n)}$, for dealing with the probability mass on an interval formed by a censoring time and infinity.

Throughout this section (and indeed this entire chapter, except Section 3.6) we assume that there are no ties of any kind in the data set, so also no ties among censored observations, nor censoring times coinciding with observed event times. Suppose that the data available, and on which to base predictive probabilities for a future observation, are as follows.

Data notation:
Assume that information is available on n exchangeable nonnegative real-valued random quantities, $T_1, T_2, \ldots, T_n$. For n observations, consisting of u event times $t_{(1)}, t_{(2)}, \ldots, t_{(u)}$ and v right-censoring times $c_{(1)}, c_{(2)}, \ldots, c_{(v)}$, assume that $0 < t_{(1)} < t_{(2)} < \ldots < t_{(u)}$ and $0 < c_{(1)} < c_{(2)} < \ldots < c_{(v)}$ are the ordered data. In addition, the notation $t_{(0)} = 0$ and $t_{(u+1)} = \infty$ is used, unless explicitly stated otherwise. Let $I_i = (t_{(i)}, t_{(i+1)})$, for $i = 0, 1, \ldots, u$, and let $c_1^i < c_2^i < \ldots < c_{l_i}^i$ denote the ordered censoring times within I_i, where l_i is the number of censoring times in I_i, of course these l_i are nonnegative and sum up to v.

In Chapter 2 the assumption $A_{(n)}$ [39, 40, 42] was discussed. $A_{(n)}$ provides a partially specified predictive distribution for a future observation given past observations, and describes this predictive distribution via probability masses in open intervals between the observed event times. These probability masses are restricted to those intervals, but there are no further specifications or restrictions on the spread of the probability mass within such an interval. The generalization of $A_{(n)}$, presented in the next section, will specify predictive probabilities in a similar way. Therefore, a notation for probability mass is introduced, which is called M-function.

Definition 1 (M-function)
A partial specification of a probability distribution for a real-valued random quantity T can be provided via probability masses assigned to intervals, without any further restriction on the spread of the probability mass within each interval. A probability mass assigned, in such a way, to an interval (a, b) is denoted by $M_T(a, b)$, and referred to as M-function value for T on (a, b).

The intervals in this definition can be of any nature, but in this thesis, when assuming no ties in the data, M-function values are only used on open intervals, which is the reason of presenting the definition with intervals denoted as (a, b). Clearly, all M-function values for T specified on all intervals should sum up to one, and each M-function value should be in $[0, 1]$. This definition does not require all probability mass for T in (a, b) to be specified in a single M-function value, but clearly one can find a minimal representation by using only a single M-function value for all probability mass specified in such a way to an interval. Predictive probabilities according to $A_{(n)}$, for the next observation T_{n+1}, can be specified by $M_{T_{n+1}}(t_{(j)}, t_{(j+1)}) = 1/(n+1)$ for $j = 0, 1, \ldots, n$.

Generally, once a partial specification of a probability distribution for a random quantity is available in terms of M-function values, optimal bounds (i.e. minimum upper and maximum lower bounds) for probabilities of events involving this random quantity can be derived, where the bounds often correspond logically to limits occurring when the probability masses within the intervals are moved towards the boundaries of the intervals.

3.2.2 Effect of right-censored data on $A_{(n)}$

(I) The assumption $\tilde{A}_{(n)}$

The assumption $A_{(n)}$ provides a partially specified probability distribution for a future observation T_{n+1}, in terms of M-function values, based on n observed event times. But there might be right-censored observations among the data. In this situation, what is the effect of right-censored data on such a partially specified probability distribution for T_{n+1}? This leads to the definition of a generalization of $A_{(n)}$, which is denoted by $\tilde{A}_{(n)}$.

Definition 2 ($\tilde{A}_{(n)}$)
The assumption $\tilde{A}_{(n)}$ is that the probability distribution for a nonnegative random quantity T_{n+1}, on the basis of data including u event times, $t_{(1)} < t_{(2)} < \ldots < t_{(u)}$, and $v = n - u$ right-censoring times, $c_{(1)} < c_{(2)} < \ldots < c_{(v)}$, is partially specified by the following M-function values:

(i) $\tilde{M}_{T_{n+1}}(t_{(i)}, t_{(i+1)}) = 1/(n+1), \quad \text{for} \quad i = 0, 1, \ldots, u;$

(ii) $\tilde{M}_{T_{n+1}}(c_{(j)}, \infty) = 1/(n+1), \quad \text{for} \quad j = 1, \ldots, v.$

We use the notation $\tilde{M}$ to emphasize that these M-function values are based on the assumption $\tilde{A}_{(n)}$ to distinguish notation from M-function values based on rc-$A_{(n)}$ later via this thesis.

The assumption $\tilde{A}_{(n)}$ partially specifies the predictive probabilities for a future observation via M-function values, in the case of right-censoring times among the observations, without any further assumptions. These probability masses are defined on two kinds of open intervals, one is formed by consecutive event times, and the other is formed by a censoring time and infinity. It is straightforward to see that, if the data do not include any right-censored observations, so $v = 0$ and $u = n$, then $\tilde{A}_{(n)}$ is identical to $A_{(n)}$. Let us consider an example to explain the assumption $\tilde{A}_{(n)}$. We use this as a didactic example throughout this chapter.

Example 7

Suppose that we have data consisting of four event times, 3, 6, 8, 9, and two right-censoring times, 4 and 7. Let T_7 denote a corresponding random quantity for a future observation.

According to $\tilde{A}_{(6)}$, the M-function values for T_7 are

$$\tilde{M}_{T_7}(0,3) = \tilde{M}_{T_7}(3,6) = \tilde{M}_{T_7}(6,8) = \tilde{M}_{T_7}(8,9) = \tilde{M}_{T_7}(9,\infty) = 1/7$$

and

$$\tilde{M}_{T_7}(4,\infty) = \tilde{M}_{T_7}(7,\infty) = 1/7.$$

(II) Justification of $\tilde{A}_{(n)}$

The justification of $\tilde{A}_{(n)}$, in relation to $A_{(n)}$, is as follows. The intervals created by the observed events times, $(t_{(i)}, t_{(i+1)})$, for $i = 0, 1, \ldots, u$, are each assigned a probability mass of $1/(n+1)$, by $A_{(n)}$. Considering one such interval, the total mass in it could actually be more than $1/(n+1)$ due to the presence of one or more right-censoring times within this interval. However, any additional probability mass due to such right-censoring times would not necessarily be restricted to lie within this interval, without additional assumptions, indeed leading to M-function values on intervals from a right-censoring time to infinity. If there is a right-censoring time in interval $(t_{(i)}, t_{(i+1)})$, then the unobserved event time corresponding to this right-censoring time perhaps could also have fallen in this same interval, in which case the M-function value $1/(n+1)$ assigned to $(t_{(i)}, t_{(i+1)})$ would actually be assigned to a smaller interval, namely from $t_{(i)}$ to that unobserved event time. However, without any further assumptions on where such an observed event time would fall within the interval $(t_{(i)}, t_{(i+1)})$, this probability mass $1/(n+1)$ cannot justifiably be restricted to a sub-interval, hence $\tilde{M}_{T_{n+1}}(t_{(i)}, t_{(i+1)}) = 1/(n+1)$.

A right-censoring at time $c_{(j)}$ only means that the corresponding event time would exceed $c_{(j)}$. If this event time were actually observed, denote it by t_c, then its effect under the assumption $A_{(n)}$ would be to split one of the intervals $(t_{(i)}, t_{(i+1)})$, created by the observed event times, and add probability mass $1/(n+1)$ to $(t_c, t_{(i+1)})$, while restricting the mass $1/(n+1)$ already assigned to this interval $(t_{(i)}, t_{(i+1)})$ to $(t_{(i)}, t_c)$, which is discussed above. Hence, this right-censoring time can be considered as carrying with it a probability mass $1/(n+1)$, which it would assign to interval $(t_c, t_{(i+1)})$ if t_c were actually observed. But because it is only known that t_c would exceed $c_{(j)}$, the only statement about this probability mass $1/(n+1)$ for T_{n+1} that can be justified, without further assumptions, is that it will fall in $(c_{(j)}, \infty)$, hence $\tilde{M}_{T_{n+1}}(c_{(j)}, \infty) = 1/(n+1)$.

3.2.3 Shifted-$\tilde{A}_{(n)}$ for right-censored random quantities

The assumption $\tilde{A}_{(n)}$ partially specifies a predictive probability distribution for a future observation T_{n+1} via M-function values. Proceeding with these M-function values would enable inference via bounds on probabilities concerning T_{n+1}. However, the probability masses assigned to intervals $(c_{(j)}, \infty)$ would cause wide bounds on probabilities. So it would be useful if we can split the probability masses assigned to intervals $(c_{(j)}, \infty)$ further into masses on sub-intervals. From the justification of $\tilde{A}_{(n)}$, it can be found that $\tilde{M}_{T_{n+1}}(c_{(j)}, \infty)$ is a consequence of assuming no further information for the random quantity, corresponding to $c_{(j)}$, on where it would actually be in $(c_{(j)}, \infty)$. This indicates that the random quantity, corresponding to $c_{(j)}$, is a key quantity for $\tilde{M}_{T_{n+1}}(c_{(j)}, \infty)$. We propose an additional assumption for the random quantity corresponding to $c_{(j)}$, say T_j^c, for which we only observed $T_j^c > c_{(j)}$. An assumption for such right-censored random quantities is presented, which is related to $\tilde{A}_{(n)}$, and called 'shifted-$\tilde{A}_{(n)}$'.

(I) The assumption shifted-$\tilde{A}_{(n)}$

Suppose again that there are n observations consisting of u event times, $t_{(1)} < t_{(2)} < \ldots < t_{(u)}$, and v right-censored observations, $c_{(1)} < c_{(2)} \ldots < c_{(v)}$. Let $t_{(0)} = 0$ and $t_{(u+1)} = \infty$. Let T_j^c, $j = 1, \ldots, v$, denote the random quantity corresponding to the right-censoring at $c_{(j)}$.

Definition 3 (shifted-$\tilde{A}_{(n)}$)
For given $c_{(j)}$ $(1 \leq j \leq v)$, assume $c_{(j)} \in (t_{(k)}, t_{(k+1)})$ $(0 \leq k \leq u)$, and let $n_{c_{(j)}}$ be the number of these n random quantities in the risk set at time $c_{(j)}$. Then shifted-$\tilde{A}_{(n)}$ provides a partial specification of the probability distribution of T_j^c, conditional on $T_j^c > c_{(j)}$, by the following definitions:

(i) $M_{T_j^c}(t_{(h)}, t_{(h+1)}) = 1/(n_{c_{(j)}} + 1), \quad$ for $\; h = k+1, \ldots, u$;

(ii) $M_{T_j^c}(c_{(j)}, t_{(k+1)}) = 1/(n_{c_{(j)}} + 1)$;

(iii) $M_{T_j^c}(c_{(l)}, \infty) = 1/(n_{c_{(j)}} + 1), \quad$ for $\; l = j+1, \ldots, v$.

The M-function values for T_j^c on all other intervals are zero. The sum of these M-function values is equal to one, so indeed they partially specify a probability distribution for T_j^c over $(c_{(j)}, \infty)$. For convenience, shifted-$\tilde{A}_{(n)}$ will be also denoted by $\tilde{A}_{(n_{c_{(j)}};\ c_{(j)})}$, or, both if it is clear which random quantity T_j^c it relates to, and to indicate this assumption more generally, by $\tilde{A}_{(n_c;\ c)}$.

The assumption $\tilde{A}_{(n_c;\ c)}$ provides a partially specified probability distribution for the random quantity corresponding to the right-censoring at $c_{(j)}$. Clearly, $\tilde{A}_{(n)}$ is the special case that could be denoted as $\tilde{A}_{(n;0)}$, although of course it is not assumed that there is a right-censoring at time 0. Let us consider the didactic example again.

Example 7 (continued)
There are two right-censoring times, 4 and 7, among the observations. Let T_1^c and T_2^c denote the random quantities corresponding to the right-censorings at times 4 and 7, respectively. According to the assumption $\tilde{A}_{(4;\ 4)}$, the probability distribution for T_1^c is partially specified as

$$M_{T_1^c}(4, 6) = M_{T_1^c}(6, 8) = M_{T_1^c}(8, 9) = M_{T_1^c}(9, \infty) = \frac{1}{5}$$

and

$$M_{T_1^c}(7, \infty) = \frac{1}{5}.$$

The assumption $\tilde{A}_{(2;\ 7)}$ for T_2^c gives

$$M_{T_2^c}(7, 8) = M_{T_2^c}(8, 9) = M_{T_2^c}(9, \infty) = \frac{1}{3}.$$

(II) Justification of shifted-$\tilde{A}_{(n)}$

The justification of shifted-$\tilde{A}_{(n)}$, in relation to $A_{(n)}$ and $\tilde{A}_{(n)}$, is based on the natural interpretation of these as post-data assumptions related to finite exchangeability, as discussed in Section 2.3, and the following properties of exchangeable random quantities.

Property 1 If random quantities $T_1, T_2, \ldots, T_r$ are exchangeable, then the random quantities in any subset of $T_1, T_2, \ldots, T_r$ are exchangeable.

Property 2 If random quantities $T_1, T_2, \ldots, T_s$ are exchangeable, then they are also exchangeable when all are conditioned on exceeding a given value c.

Property 1 is straightforward, Property 2 follows from the definition of conditional probability. These two properties can be naturally combined to say that the random quantities in a subset of exchangeable random quantities, and conditioned on all random quantities in the subset exceeding c, are again exchangeable, as long as the subset is determined by the criterion of exceeding c only. In other words, the selection of the subset is not based on any further possible variables or information. We believe that exchangeability of all the random quantities known to be in the risk set just prior to c is a natural assumption for dealing with a random quantity that is right-censored at time c, and indeed implies an assumption of 'non-informative censoring'. The assumption $\tilde{A}_{(n_c;\ c)}$ is a post-data equivalence of the discussed combination of Property 1 and Property 2, assuming such non-informative censoring. Related to Property 1, it is worth mentioning that Hill [39] showed that $A_{(n)}$ implies $A_{(m)}$ for all $m \leq n$.

Using the possible interpretation in terms of ranks, $\tilde{A}_{(n_c;\ c)}$ implies that a random quantity that is right-censored at time c, has equal probability to have any of the ranks $1, 2, \ldots, n_c + 1$ when restricted to the $n_c + 1$ random quantities in the risk set just prior to c, also when the value of the other random quantities are known (this is the same argument as discussed in Section 2.3, that observations carry information on location but not on ranks of the related non-observed random quantities).

3.2.4 Right-censoring $\tilde{A}_{(n)}$ for M-functions on $(c_{(j)}, \infty)$

The assumption $\tilde{A}_{(n)}$, presented in Subsection 3.2.2, gives M-function values for T_{n+1} defined on all the intervals $(t_{(i)}, t_{(i+1)})$ and $(c_{(j)}, \infty)$ created by the observations. The assumption shifted-$\tilde{A}_{(n)}$, presented in Subsection 3.2.3, gives M-function values for the random quantities corresponding to the censored observations. The main task now is to link the random quantities corresponding to censored observations to the random quantity of interest in the nonparametric predictive inference, namely T_{n+1}. This leads to splitting of the total M-function value for T_{n+1} assigned to $(c_{(j)}, \infty)$. Therefore, an additional assumption is presented, which we call 'right-censoring $\tilde{A}_{(n)}$', also denoted as 'rc-$\tilde{A}_{(n)}$'.

(I) The assumption right-censoring $\tilde{A}_{(n)}$

Suppose that there are n observations consisting of u event times, $t_{(1)} < t_{(2)} < \ldots < t_{(u)}$, and v $(= n-u)$ right-censored observations, $c_{(1)} < c_{(2)} < \ldots < c_{(v)}$. Let $t_{(0)} = 0$ and $t_{(u+1)} = \infty$. Assume that $k \in \{0, 1, \ldots, u\}$ is such that $c_{(j)} \in (t_{(k)}, t_{(k+1)})$. Let $M_{T_{n+1}}(c_{(j)}, \infty) = p_{(c_{(j)})}$ be the M-function value for T_{n+1} due to censoring at $c_{(j)}$ for $j = 1, \ldots, v$. Let $M^{c_{(j)}}_{T_{n+1}}(a, b)$ denote the M-function value as a separate entity from the total M-function value of $M_{T_{n+1}}(c_{(j)}, \infty)$, where (a, b) is a sub-interval of $(c_{(j)}, \infty)$. Then right-censoring $\tilde{A}_{(n)}$ (rc-$\tilde{A}_{(n)}$) is described as follows:

Definition 4 (rc-$\tilde{A}_{(n)}$)
For given $M_{T_{n+1}}(c_{(j)}, \infty) = p_{(c_{(j)})}$, let $n_{c_{(j)}}$ be the number of random quantities in the risk set at $c_{(j)}$. Then rc-$\tilde{A}_{(n)}$ splits the probability mass of $M_{T_{n+1}}(c_{(j)}, \infty)$ as

(i) $M^{c_{(j)}}_{T_{n+1}}(t_{(h)}, t_{(h+1)}) = p_{(c_{(j)})}/(n_{c_{(j)}} + 1)$; for $h = k+1, \ldots, u$;

(ii) $M^{c_{(j)}}_{T_{n+1}}(c_{(j)}, t_{(k+1)}) = p_{(c_{(j)})}/(n_{c_{(j)}} + 1)$;

(iii) $M^{c_{(j)}}_{T_{n+1}}(c_{(l)}, \infty) = p_{(c_{(j)})}/(n_{c_{(j)}} + 1)$; for $l = j+1, \ldots, v$.

The $M^{c_{(j)}}$-function values for T_{n+1} on all other intervals are equal to zero. For convenience, the assumption right-censoring $\tilde{A}_{(n)}$ (rc-$\tilde{A}_{(n)}$) will also be denoted by $\tilde{A}_{(n;\ n_{c_{(j)}};\ c_{(j)};\ p_{(c_{(j)})})}$, or, both if it is clear which random quantity it relates to, and to indicate this assumption more generally, by $\tilde{A}_{(n;\ n_c;\ c;\ p_{(c)})}$. As a special case of $\tilde{A}_{(n;\ n_c;\ c;\ p_{(c)})}$, $\tilde{A}_{(n;\ n;\ 0;\ 1)}$ is identical to $\tilde{A}_{(n)}$.

(II) Justification of $\tilde{A}_{(n;\ n_c;\ c;\ p_{(c)})}$

The probability mass $p_{(c_{(j)})} = M_{T_{n+1}}(c_{(j)}, \infty)$ is the mass for T_{n+1} assigned to $(c_{(j)}, \infty)$ as a consequence of the censoring of random quantity T^c_j at $c_{(j)}$. The assumption $\tilde{A}_{(n_{c_{(j)}};\ c_{(j)})}$ for T^c_j allows further conditioning on sub-intervals of $(c_{(j)}, \infty)$ in which T^c_j would actually occur as event time, according to $\tilde{A}_{(n_{c_{(j)}};\ c_{(j)})}$ as defined in Subsection 3.2.3, which provides a partially specified probability distribution for T^c_j.

Let (a, b) be a sub-interval of $(c_{(j)}, \infty)$, with a any observation greater than or equal to $c_{(j)}$, and corresponding b either the smallest observed event time greater than a or infinity. If T_j^c would actually take on a value, as event time, in interval (a, b), say $T_j^c = d \in (a, b)$, then all the mass $p_{(c_{(j)})}$ for T_{n+1}, due to this censoring, would be assigned to the interval (d, b), according to $\tilde{A}_{(n)}$. However, without any additional assumptions, nothing more is known about the exact location of d, so the mass for sub-intervals of (a, b) cannot be specified, and, by similar argument, all other M-function values for T_{n+1} assigned to intervals ending at b cannot be assigned to shorter intervals, ending at d, without further assumptions on the location of the unknown d. Therefore, by what is essentially an application of the theorem of total probability, using the assumption $\tilde{A}_{(n_{c_{(j)}};\ c_{(j)})}$ for T_j^c, we get

$$\begin{aligned} M_{T_{n+1}}^{c_{(j)}}(a,b) &= p_{(c_{(j)})} \times P(T_j^c \in (a,b) | T_j^c > c_{(j)}) + 0 \times P(T_j^c \notin (a,b) | T_j^c > c_{(j)}) \\ &= p_{(c_{(j)})} \times M_{T_j^c}(a,b) \\ &= p_{(c_{(j)})} \times \frac{1}{n_{c_{(j)}} + 1}. \end{aligned}$$

(III) Deriving $p_{(c)}$

Now we consider how to derive $p_{(c_{(j)})}$ for the M-function assigned to $(c_{(j)}, \infty)$. The assumption $\tilde{A}_{(n)}$ gave M-function values for T_{n+1} for all intervals $(t_{(i)}, t_{(i+1)})$ and $(c_{(j)}, \infty)$, with $\tilde{M}_{T_{n+1}}(c_{(j)}, \infty)$ equal for all $j = 1, \ldots, v$. However, according to the assumption rc-$\tilde{A}_{(n)}$, the M-function value for T_{n+1} on $(c_{(j)}, \infty)$ also depends on contributions from previous censorings to this interval. So the value $p_{(c_{(j)})}$ is the sum of all probability masses for T_{n+1} assigned to $(c_{(j)}, \infty)$, both due to the assumptions $\tilde{A}_{(n)}$ and rc-$\tilde{A}_{(n)}$. The following theorem provides the value of $p_{(c_{(j)})}$.

Theorem 1
Given the data as described in this section, and for any right-censoring time $c_{(j)}$ $(1 \le j \le v)$, assume $\tilde{A}_{(n)}$ and $\tilde{A}_{(n;\ n_{c_{(r)}};\ c_{(r)};\ p_{(c_{(r)})})}$ for $r = 1, \ldots, j-1$. Then

$$p_{(c_{(j)})} = M_{T_{n+1}}(c_{(j)}, \infty) = \frac{1}{n+1} \prod_{\{r: r<j\}} \frac{\tilde{n}_{c_{(r)}} + 1}{\tilde{n}_{c_{(r)}}},$$

where $j = 1, \ldots, v$, and $\tilde{n}_{c_{(r)}}$ is the number of individuals in the risk set just prior to time $c_{(r)}$. The product terms are defined as one if the product is taken over an empty set.

Proof
Starting with the assumption $\tilde{A}_{(n)}$, a part of the probability mass for T_{n+1} was

assigned to interval $(c_{(j)}, \infty)$, for $j = 1, \ldots, v$, after which this mass is further divided on the basis of the assumptions rc-$\tilde{A}_{(n)}$, giving the masses $M^{c_{(j)}}_{T_{n+1}}(a, b)$.

First consider what happens due to the first censoring at time $c_{(1)}$. Clearly,

$$p_{(c_{(1)})} = M_{T_{n+1}}(c_{(1)}, \infty) = \tilde{M}_{T_{n+1}}(c_{(1)}, \infty) = \frac{1}{n+1},$$

as there have been no previous censorings that affect this M-function value, where the second equality is based on the link to $\tilde{A}_{(n)}$.

Secondly, consider the effect of the second censoring time $c_{(2)}$, taking into account the previous censoring at $c_{(1)}$. Clearly,

$$M_{T_{n+1}}(c_{(2)}, \infty) = \tilde{M}_{T_{n+1}}(c_{(2)}, \infty) + M^{c_{(1)}}_{T_{n+1}}(c_{(2)}, \infty) = \frac{1}{n+1} + M^{c_{(1)}}_{T_{n+1}}(c_{(2)}, \infty),$$

where, assuming $\tilde{A}_{(n;\ n_{c_{(1)}};\ c_{(1)};\ p_{(c_{(1)})})}$ and using that $\tilde{n}_{c_{(1)}} = n_{c_{(1)}} + 1$,

$$\begin{aligned} M^{c_{(1)}}_{T_{n+1}}(c_{(2)}, \infty) &= \frac{p_{(c_{(1)})}}{n_{c_{(1)}} + 1} \\ &= \frac{1}{n+1} \times \frac{1}{\tilde{n}_{c_{(1)}}}. \end{aligned}$$

Hence,

$$p_{(c_{(2)})} = M_{T_{n+1}}(c_{(2)}, \infty) = \frac{1}{n+1} \times \frac{\tilde{n}_{c_{(1)}} + 1}{\tilde{n}_{c_{(1)}}}.$$

Now the proof can be completed via induction. Assume, for $j = 2, \ldots, v$, that

$$p_{(c_{(j-1)})} = M_{T_{n+1}}(c_{(j-1)}, \infty) = \frac{1}{n+1} \prod_{\{r:r<j-1\}} \frac{\tilde{n}_{c_{(r)}} + 1}{\tilde{n}_{c_{(r)}}}.$$

The M-function value for T_{n+1} on the interval $(c_{(j)}, \infty)$ is equal to the originally assigned mass $1/(n+1)$, resulting from $\tilde{A}_{(n)}$, and the contributions to this interval resulting from each of the previous censorings, so

$$M_{T_{n+1}}(c_{(j)}, \infty) = \frac{1}{n+1} + \sum_{l=1}^{j-1} M^{c_{(l)}}_{T_{n+1}}(c_{(j)}, \infty).$$

And, similarly, the corresponding value for the interval $(c_{(j-1)}, \infty)$ is

$$M_{T_{n+1}}(c_{(j-1)}, \infty) = \frac{1}{n+1} + \sum_{l=1}^{j-2} M^{c_{(l)}}_{T_{n+1}}(c_{(j-1)}, \infty).$$

From the assumptions $\tilde{A}_{(n;\ n_{c_{(r)}};\ c_{(r)};\ p_{c_{(r)}})}$, for $r = 1, \ldots, j-2$, we have

$$M^{c_{(l)}}_{T_{n+1}}(c_{(j-1)}, \infty) = M^{c_{(l)}}_{T_{n+1}}(c_{(j)}, \infty) \quad \text{for all } \ l = 1, \ldots, j-2,$$

so

$$\begin{aligned} M_{T_{n+1}}(c_{(j)},\infty) &= M_{T_{n+1}}(c_{(j-1)},\infty) + M_{T_{n+1}}^{c_{(j-1)}}(c_{(j)},\infty) \\ &= p_{(c_{(j-1)})} + p_{(c_{(j-1)})} \times \frac{1}{\tilde{n}_{c_{(j-1)}}} \\ &= p_{(c_{(j-1)})} \times \frac{\tilde{n}_{c_{(j-1)}}+1}{\tilde{n}_{c_{(j-1)}}} \\ &= \Big[\frac{1}{n+1} \prod_{\{r:r<j-1\}} \frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}}\Big] \times \frac{\tilde{n}_{c_{(j-1)}}+1}{\tilde{n}_{c_{(j-1)}}} \\ &= \frac{1}{n+1} \prod_{\{r:r<j\}} \frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}}, \end{aligned}$$

which completes the proof. □

We illustrate $\tilde{A}_{(n;\ n_{c_{(r)}};\ c_{(r)};\ p_{(c_{(r)})})}$ again via the didactic example.

Example 7 (continued)
The assumption $\tilde{A}_{(6)}$ assigned a probability mass of 1/7 to intervals $(4,\infty)$ and $(7,\infty)$. Clearly, $p_{(4)} = M_{T_7}(4,\infty) = \tilde{M}_{T_7}(4,\infty) = 1/7$. The assumption $\tilde{A}_{(6;\ 4;\ 4;\ 1/7)}$ implies that $p_{(4)} = M_{T_7}(4,\infty)$ can be split into

$$M_{T_7}^4(4,6) = M_{T_7}^4(6,8) = M_{T_7}^4(8,9) = M_{T_7}^4(9,\infty) = \frac{1}{7} \times \frac{1}{5} = \frac{1}{35}$$

and

$$M_{T_7}^4(7,\infty) = \frac{1}{7} \times \frac{1}{5} = \frac{1}{35}.$$

By Theorem 1,

$$p_{(7)} = M_{T_7}(7,\infty) = \frac{1}{7} \times (1 + \frac{1}{4+1}) = \frac{6}{35}.$$

Clearly, the value of $p_{(7)} = M_{T_7}(7,\infty)$ is obtained by both considering the effect of the censoring time 7 and taking into account the previous censoring at 4. Further use of $\tilde{A}_{(6;\ 2;\ 7;\ 6/35)}$ splits $p_{(7)} = M_{T_7}(7,\infty) = 6/35$ into

$$M_{T_7}^7(7,8) = M_{T_7}^7(8,9) = M_{T_7}^7(9,\infty) = \frac{6}{35} \times \frac{1}{3} = \frac{2}{35}.$$

Now, the M-function values assigned to intervals formed by a censoring time and infinity are all split into M-function values on sub-intervals.

3.3 Right-censoring $A_{(n)}$

In this section a new assumption 'right-censoring $A_{(n)}$', also denoted as 'rc-$A_{(n)}$', is introduced, based on the combined assumptions in Section 3.2. The assumption $\tilde{A}_{(n)}$ presented in Subsection 3.2.2, related to $A_{(n)}$, considered the effect of right-censored data, giving a partially specified probability distribution for a future observation by assigning predictive probability to intervals $(t_{(i)}, t_{(i+1)})$ and $(c_{(j)}, \infty)$ for $i = 0, 1, \ldots, u$ and $j = 1, 2, \ldots, v$, expressed via M-function values. The assumption shifted-$\tilde{A}_{(n)}$ presented in Subsection 3.2.3 provides a partially specified probability distribution for the random quantity corresponding to the right-censored observation. The assumption $\tilde{A}_{(n;\ n_c;\ c;\ p_c)}$ presented in Subsection 3.2.4, assumed that the M-function value for $(c_{(j)}, \infty)$ is equal to $p_{c_{(j)}}$, and divides this into corresponding M-function values for sub-intervals, and $p_{c_{(j)}}$ is derived by taking into account the contribution from all censorings that occurred before $c_{(j)}$. Therefore, the M-function values for T_{n+1} are finally all assigned to intervals $(t_{(i)}, t_{(i+1)})$ or $(c^i_k, t_{(i+1)})$ for $i = 0, 1, \ldots, u$ and $k = 1, 2, \ldots, l_i$, by combining all M-function values for T_{n+1} that are defined on the same interval as a consequence of the initial assumption $\tilde{A}_{(n)}$ and $\tilde{A}_{(n;\ n_c;\ c;\ p_c)}$. This assumption is called 'right-censoring $A_{(n)}$', also denoted as 'rc-$A_{(n)}$'.

3.3.1 The assumption right-censoring $A_{(n)}$

Definition 5 (rc-$A_{(n)}$)
Suppose that there are n observations consisting of u event times, $t_{(1)} < t_{(2)} < \ldots < t_{(u)}$, and v $(= n - u)$ right-censored observations, $c_{(1)} < c_{(2)} \ldots < c_{(v)}$. Let $t_{(0)} = 0$ and $t_{(u+1)} = \infty$. Then right-censoring $A_{(n)}$ (rc-$A_{(n)}$) partially specifies the probability distribution for the next observation T_{n+1} by the following M-function values,

$$
\begin{aligned}
M_{T_{n+1}}(t_{(i)}, t_{(i+1)}) &= \frac{1}{n+1} \prod_{\{r:\, c_{(r)} < t_{(i)}\}} \frac{\tilde{n}_{c_{(r)}} + 1}{\tilde{n}_{c_{(r)}}}, \\
M_{T_{n+1}}(c^i_k, t_{(i+1)}) &= \frac{1}{(n+1)\tilde{n}_{c^i_k}} \prod_{\{r:\, c_{(r)} < c^i_k\}} \frac{\tilde{n}_{c_{(r)}} + 1}{\tilde{n}_{c_{(r)}}},
\end{aligned}
$$

where $i = 0, 1, \ldots, u$ and $k = 1, \ldots, l_i$, and $\tilde{n}_{c_{(r)}}$ is the number of individuals in the risk set just prior to time $c_{(r)}$. The product terms are defined as one if the product is taken over an empty set.

Clearly, With a total of n observations, the probability distribution for T_{n+1} according to rc-$A_{(n)}$ is partially specified by $n+1$ M-function values, one related to each of the n observations, namely specified on the open interval from the observation till the next observed event time (or infinity), and one on the interval $(0, t_{(1)})$. These $n+1$ M-function values for T_{n+1} sum up to one, and if there are no right-censored observations, so the data consists of n actually observed event times, then rc-$A_{(n)}$ is identical to Hill's $A_{(n)}$.

The calculation of the M-function values according to rc-$A_{(n)}$ is illustrated by the didactic example.

Example 7 (continued)
The M-function values for the next observation T_7, according to rc-$A_{(6)}$, can be derived as Table 3.1.

Interval	M-function
(0, 3)	5/35
(3, 6)	5/35
(4, 6)	1/35
(6, 8)	6/35
(7, 8)	2/35
(8, 9)	8/35
$(9, \infty)$	8/35

Table 3.1: M-function values for T_7 (Example 7).

3.3.2 Justification of right-censoring $A_{(n)}$

The justification of the M-function values for T_{n+1} in the definition of rc-$A_{(n)}$ can now be completed, following the assumptions $\tilde{A}_{(n)}$ and $\tilde{A}_{(n;\ n_c;\ c;\ p_{(c)})}$.

First, consider $M_{T_{n+1}}(c_k^i, t_{(i+1)})$. According to the assumption $\tilde{A}_{(n;\ n_c;\ c;\ p_{(c)})}$, and using Theorem 1,

$$p_{(c_k^i)} = M_{T_{n+1}}(c_k^i, \infty) = \frac{1}{n+1} \prod_{\{r:\, c_{(r)} < c_k^i\}} \frac{\tilde{n}_{c_{(r)}} + 1}{\tilde{n}_{c_{(r)}}},$$

and

$$M_{T_{n+1}}^{c_k^i}(c_k^i, t_{(i+1)}) = \frac{p_{(c_k^i)}}{\tilde{n}_{c_k^i}},$$

which is the probability mass assigned to the interval $(c_k^i, t_{(i+1)})$ as a result of the censoring at c_k^i, which has taken into account the effect of all previous censorings.

It is important to emphasize here that $p_{(c^i_k)}$ includes all the values $M^{c_{(r)}}_{T_{n+1}}(c^i_k, \infty)$ for all censoring times $c_{(r)} < c^i_k$. Therefore, this is the only probability mass that is, in such a way, assigned and restricted to this interval, hence

$$\begin{aligned} M_{T_{n+1}}(c^i_k, t_{(i+1)}) &= M^{c^i_k}_{T_{n+1}}(c^i_k, t_{(i+1)}) \\ &= \frac{1}{(n+1)\tilde{n}_{c^i_k}} \prod_{\{r:\, c_{(r)} < c^i_k\}} \frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}}, \end{aligned}$$

as stated in the definition of rc-$A_{(n)}$.

Secondly, consider $M_{T_{n+1}}(t_{(i)}, t_{(i+1)})$. This probability mass consists of the initial value $1/(n+1)$, assigned to this interval by the assumption $\tilde{A}_{(n)}$, and contributions related to each of the censorings that occurred before $t_{(i)}$ according to the corresponding $\tilde{A}_{(n;\ n_c;\ c;\ p_{(c)})}$ assumptions, so

$$M_{T_{n+1}}(t_{(i)}, t_{(i+1)}) = \frac{1}{n+1} + \sum_{\{r:\, c_{(r)} < t_{(i)}\}} M^{c_{(r)}}_{T_{n+1}}(t_{(i)}, t_{(i+1)}).$$

Derivation of this M-function value is simplified by use of the following equalities, justified in rc-$\tilde{A}_{(n)}$ and Theorem 1,

$$M_{T_{n+1}}(c^{i-1}_{l_{i-1}}, \infty) = \frac{1}{n+1} + \sum_{\{r:\, c_{(r)} < c^{i-1}_{l_{i-1}}\}} M^{c_{(r)}}_{T_{n+1}}(c^{i-1}_{l_{i-1}}, \infty),$$

and

$$M^{c_{(r)}}_{T_{n+1}}(t_{(i)}, t_{(i+1)}) = M^{c_{(r)}}_{T_{n+1}}(c^{i-1}_{l_{i-1}}, \infty), \quad \text{for all } r \text{ with } c_{(r)} < c^{i-1}_{l_{i-1}}.$$

This leads to

$$\begin{aligned} M_{T_{n+1}}(t_{(i)}, t_{(i+1)}) &= M_{T_{n+1}}(c^{i-1}_{l_{i-1}}, \infty) + M^{c^{i-1}_{l_{i-1}}}_{T_{n+1}}(t_{(i)}, t_{(i+1)}) \\ &= M_{T_{n+1}}(c^{i-1}_{l_{i-1}}, \infty) + M_{T_{n+1}}(c^{i-1}_{l_{i-1}}, \infty) \times \frac{1}{\tilde{n}_{c^{i-1}_{l_{i-1}}}} \\ &= M_{T_{n+1}}(c^{i-1}_{l_{i-1}}, \infty) \times \frac{\tilde{n}_{c^{i-1}_{l_{i-1}}}+1}{\tilde{n}_{c^{i-1}_{l_{i-1}}}} \\ &= \left[\frac{1}{n+1} \prod_{\{r:\, c_{(r)} < c^{i-1}_{l_{i-1}}\}} \frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}}\right] \times \frac{\tilde{n}_{c^{i-1}_{l_{i-1}}}+1}{\tilde{n}_{c^{i-1}_{l_{i-1}}}} \\ &= \frac{1}{n+1} \prod_{\{r:\, c_{(r)} < t_{(i)}\}} \frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}}, \end{aligned}$$

as stated in the definition of rc-$A_{(n)}$.

3.3.3 Discussion of rc-$A_{(n)}$

The assumption rc-$A_{(n)}$ provides a partially specified predictive probability distribution for a future observation, based on data including right-censored observations, and expressed via M-function values. The justification of rc-$A_{(n)}$ is based on the relation between exchangeability and Hill's $A_{(n)}$, and the fact that exchangeability still holds for real-valued random quantities when they are all conditioned on exceeding a value t.

The assumed exchangeability of the random quantities $T_1, \ldots, T_n$ (before the data are available) is not strictly required, but, as in Hill's presentation of $A_{(n)}$, it is natural in relation to the post-data assumption rc-$A_{(n)}$, as situations where one would not be willing to assume exchangeability before the data, yet would be willing to assume rc-$A_{(n)}$ once data are available, will be rare, and the assumed exchangeability also simplifies the discussion. A hypothetical example, in which one may not wish to assume exchangeability before the data, yet could still be happy to use inferences based on rc-$A_{(n)}$ after n observations, could occur if one would expect data to be a time-series, but when studying actual observations strongly doubt this prior belief and hence would wish to proceed without taking such prior beliefs further into account.

As discussed, the post-data assumption rc-$A_{(n)}$ includes an assumption on the censoring mechanism, or, more particularly, the information available about this mechanism both from background knowledge and from the data. Theoretically, this non-informative censoring assumption is a post-data equivalence of exchangeability of all random quantities known to be at risk just prior to a censoring time, implying that each of those at risk had equal chance to be the next one censored. This means that after the data have become available, one does not have any information suggesting that the items that have been censored were actually selected on the grounds of some criterion dependent on their random event time.

For simplicity of presentation, rc-$A_{(n)}$ has been discussed assuming that there are no ties of any nature present in the data, so no two observations (event or censoring) happen at the same time. Generalizations to include possible ties are discussed in Section 3.6.

The partially specified probability distribution for T_{n+1}, via the M-function values as given by rc-$A_{(n)}$, is not explicitly considered as a predictive posterior distribution within the Bayesian framework, a point of view taken by Hill when he presented $A_{(n)}$ [39, 40], and indeed formally justified by the splitting process he presented [42] as a parametric model for $A_{(n)}$ in the Bayesian context. This raises the interesting question whether rc-$A_{(n)}$ can be justified from similar perspective, which has not been analyzed in detail yet. In the Bayesian framework, one would derive a conditional distribution for $v+1$ random quantities, namely the v random quantities corresponding to the right-censored observations, each of course conditioned on surviving its right-censoring time, and T_{n+1}. Justification of rc-$A_{(n)}$ should then be based on consideration of the marginal posterior for T_{n+1}. However, it seems that this requires further assumptions and careful definition of the splitting process [42] to deal with the right-censored observations, probably along lines similar to the assumption shifted-$\tilde{A}_{(n)}$ as discussed in Section 3.2.

The assumption rc-$A_{(n)}$ partially specifies the probability distribution for the next observation by the M-function values. These M-function values can lead to inferences, which will be discussed in the next section, in terms of bounds of the survival function for T_{n+1}, which are closely linked to the method presented by Berliner and Hill [6], which in itself is no justification, yet does serve to support the use of such nonparametric predictive inferences. However, inferences considered based on $A_{(n)}$ and rc-$A_{(n)}$ will be most naturally interpreted in a frequentist context, with a valuable additional Bayesian justification for $A_{(n)}$ provided by Hill [42].

3.4 Inference based on rc-$A_{(n)}$

This section presents inference based on the partially specified probability distribution for T_{n+1}, via the M-function values as given by rc-$A_{(n)}$. Obviously, such inferences will have a predictive nature, directly in terms of T_{n+1}, which can be interpreted as 'the next observation', along similar lines as nonparametric predictive inferences based on $A_{(n)}$ for data without censored observations [3, 4, 15, 16, 18]. For many events of interest, in terms of T_{n+1}, the M-function values only allow bounds for probabilities to be derived, where the maximum lower bound is called the 'lower probability', and the minimum upper bound is called the 'upper probability', following terminology from theory of imprecise probabilities [57]. In general, these lower and upper probabilities are derived analogously as described by Augustin and Coolen [4] for inferences based on $A_{(n)}$. We use the item 'nonparametric predictive inference' (NPI) for inferences based on rc-$A_{(n)}$.

In this section, the data are again assumed to be u event times and v $(= n - u)$ right-censoring times, as described in Section 3.2, and throughout we assume rc-$A_{(n)}$.

3.4.1 Probabilities for $T_{n+1} \in (t_{(i)}, t_{(i+1)})$

Berliner and Hill [6] present a nonparametric predictive method for the next observation T_{n+1} on the basis of data including right-censored observations, based on the assumption $A_{(n)}$, as described in Section 2.4, which leads to predictive probabilities for T_{n+1} on the intervals $(t_{(i)}, t_{(i+1)})$, with $t_{(0)} = 0$ and $t_{(u+1)} = \infty$, so $P(T_{n+1} \in (t_{(i)}, t_{(i+1)}))$ for $i = 0, 1, \ldots, u$. The M-functions as specified by the assumption rc-$A_{(n)}$, can also lead straightforwardly to such probabilities for events $T_{n+1} \in (t_{(i)}, t_{(i+1)})$. The fact that these precise probabilities can be derived is caused by the intervals on which M-function values based on rc-$A_{(n)}$ are specified, as each of these is fully contained in one single interval $(t_{(i)}, t_{(i+1)})$.

Theorem 2

If we have data according to the description in Section 3.2.1, then the probabilities for events $T_{n+1} \in (t_{(i)}, t_{(i+1)})$, based on the assumption rc-$A_{(n)}$, are

$$P(T_{n+1} \in (t_{(i)}, t_{(i+1)})) = \frac{1}{n+1} \prod_{\{r:\, c_{(r)} < t_{(i+1)}\}} \frac{\tilde{n}_{c_{(r)}} + 1}{\tilde{n}_{c_{(r)}}}, \quad i = 0, 1, \ldots, u,$$

where $\tilde{n}_{c_{(r)}}$ is the number of individuals in the risk set just prior to time $c_{(r)}$, and the product term is defined as one if the product is taken over an empty set.

Proof

(i) For $l_i = 0$, i.e. no censored observations in the interval $(t_{(i)}, t_{(i+1)})$, this is straightforward from the definition of rc-$A_{(n)}$, since then $P(T_{n+1} \in (t_{(i)}, t_{(i+1)})) = M_{T_{n+1}}(t_{(i)}, t_{(i+1)})$.

(ii) Now consider the situation where there is at least one censored observation in $(t_{(i)}, t_{(i+1)})$, so $l_i \geq 1$. This probability is the sum of all probability masses assigned to the interval $(t_{(i)}, t_{(i+1)})$ and its sub-intervals via the M-function values for T_{n+1}, so

$$\begin{aligned}
& P(T_{n+1} \in (t_{(i)}, t_{(i+1)})) \\
= \; & M_{T_{n+1}}(t_{(i)}, t_{(i+1)}) + \sum_{k=1}^{l_i} M_{T_{n+1}}(c_k^i, t_{(i+1)}) \\
= \; & \frac{1}{n+1} \prod_{\{r:c_{(r)}<t_{(i)}\}} \frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}} + \sum_{k=1}^{l_i} \Big[\frac{1}{(n+1)\tilde{n}_{c_k^i}} \prod_{\{r:c_{(r)}<c_k^i\}} \frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}} \Big] \\
= \; & \Big[\frac{1}{n+1} \prod_{\{r:c_{(r)}<t_{(i)}\}} \frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}} \Big] \times \Big[1 + \sum_{k=1}^{l_i} \frac{1}{\tilde{n}_{c_k^i}} \prod_{\{r:t_{(i)}<c_{(r)}<c_k^i\}} \frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}} \Big] \\
= \; & \Big[\frac{1}{n+1} \prod_{\{r:c_{(r)}<t_{(i)}\}} \frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}} \Big] \times \Big[\prod_{\{r:t_{(i)}<c_{(r)}<t_{(i+1)}\}} \frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}} \Big] \\
= \; & \frac{1}{n+1} \prod_{\{r:c_{(r)}<t_{(i+1)}\}} \frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}},
\end{aligned}$$

where the fourth equality is based on the following lemma. □

Lemma 1

The following two equalities hold, for all $i = 0, 1, \ldots, u$,

$$(i) \qquad \sum_{k=2}^{l_i} \frac{1}{\tilde{n}_{c_k^i}\tilde{n}_{c_{k-1}^i}} = \frac{1}{\tilde{n}_{c_{l_i}^i}} - \frac{1}{\tilde{n}_{c_1^i}}, \text{ for } l_i \geq 2.$$

$$(ii) \qquad 1 + \sum_{k=1}^{l_i} \Big[\frac{1}{\tilde{n}_{c_k^i}} \prod_{\{r:t_{(i)}<c_{(r)}<c_k^i\}} \frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}} \Big] = \prod_{\{r:t_{(i)}<c_{(r)}<t_{(i+1)}\}} \frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}},$$

for $l_i \geq 1$.

The product term is defined as one if the product is taken over an empty set, except if stated otherwise.

Proof

The essential property used to prove both these equalities is

$$\tilde{n}_{c_k^i} = \tilde{n}_{c_{k+1}^i} + 1, \quad \text{for } k = 1, \ldots, l_i - 1. \tag{3.1}$$

(i) For $l_i = 2$, according (3.1),

$$\frac{1}{\tilde{n}_{c_2^i}\tilde{n}_{c_1^i}} = \frac{\tilde{n}_{c_1^i} - \tilde{n}_{c_2^i}}{\tilde{n}_{c_2^i}\tilde{n}_{c_1^i}} = \frac{1}{\tilde{n}_{c_2^i}} - \frac{1}{\tilde{n}_{c_1^i}}$$

For $l_i \geq 3$, the proof is via induction. Suppose that, for $s \in \{2, \ldots, l_i - 1\}$,

$$\sum_{k=2}^{s} \frac{1}{\tilde{n}_{c_k^i}\tilde{n}_{c_{k-1}^i}} = \frac{1}{\tilde{n}_{c_s^i}} - \frac{1}{\tilde{n}_{c_1^i}},$$

then

$$\begin{aligned}\sum_{k=2}^{s+1}\frac{1}{\tilde{n}_{c_k^i}\tilde{n}_{c_{k-1}^i}} &= \sum_{k=2}^{s}\frac{1}{\tilde{n}_{c_k^i}\tilde{n}_{c_{k-1}^i}}+\frac{1}{\tilde{n}_{c_{s+1}^i}\tilde{n}_{c_s^i}} = \frac{1}{\tilde{n}_{c_s^i}}-\frac{1}{\tilde{n}_{c_1^i}}+\frac{1}{\tilde{n}_{c_{s+1}^i}\tilde{n}_{c_s^i}}\\ &= \frac{1}{\tilde{n}_{c_s^i}}-\frac{1}{\tilde{n}_{c_1^i}}+\frac{1}{\tilde{n}_{c_{s+1}^i}}-\frac{1}{\tilde{n}_{c_s^i}} = \frac{1}{\tilde{n}_{c_{s+1}^i}}-\frac{1}{\tilde{n}_{c_1^i}},\end{aligned}$$

which completes the proof.

(ii) If $l_i = 1$, then both sides of the equality are easily seen to be equal to

$$1+\frac{1}{\tilde{n}_{c_1^i}},$$

as the product term is defined as one if the product is taken over an empty set. Now let $l_i \geq 2$, then

$$\prod_{\{r:t_{(i)}<c_{(r)}<t_{(i+1)}\}}\frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}}=\prod_{k=1}^{l_i}\frac{\tilde{n}_{c_k^i}+1}{\tilde{n}_{c_k^i}}=\frac{\tilde{n}_{c_1^i}+1}{\tilde{n}_{c_{l_i}^i}},$$

and similarly, for $k = 2, \ldots, l_i$,

$$\prod_{\{r:t_{(i)}<c_{(r)}<c_k^i\}}\frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}}=\prod_{s=1}^{k-1}\frac{\tilde{n}_{c_s^i}+1}{\tilde{n}_{c_s^i}}=\frac{\tilde{n}_{c_1^i}+1}{\tilde{n}_{c_{k-1}^i}}.$$

This leads to

$$\begin{aligned}& 1+\sum_{k=1}^{l_i}\left[\frac{1}{\tilde{n}_{c_k^i}}\prod_{\{r:t_{(i)}<c_{(r)}<c_k^i\}}\frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}}\right]\\ =\;& 1+\frac{1}{\tilde{n}_{c_1^i}}+\sum_{k=2}^{l_i}\left[\frac{1}{\tilde{n}_{c_k^i}}\times\frac{\tilde{n}_{c_1^i}+1}{\tilde{n}_{c_{k-1}^i}}\right]\\ =\;& 1+\frac{1}{\tilde{n}_{c_1^i}}+\left(\tilde{n}_{c_1^i}+1\right)\times\sum_{k=2}^{l_i}\left[\frac{1}{\tilde{n}_{c_k^i}\tilde{n}_{c_{k-1}^i}}\right]\\ =\;& \frac{\tilde{n}_{c_1^i}+1}{\tilde{n}_{c_1^i}}+\left(\tilde{n}_{c_1^i}+1\right)\times\left[\frac{1}{\tilde{n}_{c_{l_i}^i}}-\frac{1}{\tilde{n}_{c_1^i}}\right]\\ =\;& \frac{\tilde{n}_{c_1^i}+1}{\tilde{n}_{c_{l_i}^i}}\\ =\;& \prod_{\{r:t_{(i)}<c_{(r)}<t_{(i+1)}\}}\frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}},\end{aligned}$$

where part (i) of this lemma has been used for the third equality, and which completes the proof of this lemma. $\square$

The following corollary gives an alternative formula which might simplify calculation of the probabilities in Theorem 2.

Corollary 1

Using the same setting as in Theorem 2, the probabilities for events $T_{n+1} \in (t_{(i)}, t_{(i+1)})$ can be also derived by

$$P(T_{n+1} \in (t_{(i)}, t_{(i+1)})) = P(T_{n+1} \in (t_{(i-1)}, t_{(i)})) \times \left[\frac{\tilde{n}_{c_1^i} + 1}{\tilde{n}_{c_{l_i}^i}}\right],$$

where $i = 1, \ldots, u$, and c_1^i and $c_{l_i}^i$ are the smallest and largest censoring times, respectively, in interval I_i.

Proof

Using Theorem 2 the probabilities of events $T_{n+1} \in (t_{(i)}, t_{(i+1)})$ are

$$P(T_{n+1} \in (t_{(i)}, t_{(i+1)})) = \frac{1}{n+1} \prod_{\{r: c_{(r)} < t_{(i+1)}\}} \frac{\tilde{n}_{c_{(r)}} + 1}{\tilde{n}_{c_{(r)}}}.$$

Adjusting the right-hand side of this formula gives

$$\begin{aligned} P(T_{n+1} \in (t_{(i)}, t_{(i+1)})) &= \left[\frac{1}{n+1} \prod_{\{r: c_{(r)} < t_{(i)}\}} \frac{\tilde{n}_{c_{(r)}} + 1}{\tilde{n}_{c_{(r)}}}\right] \left[\prod_{j=1}^{l_i} \frac{\tilde{n}_{c_j^i} + 1}{\tilde{n}_{c_j^i}}\right] \\ &= P(T_{n+1} \in (t_{(i-1)}, t_{(i)})) \prod_{j=1}^{l_i} \frac{\tilde{n}_{c_j^i} + 1}{\tilde{n}_{c_j^i}} \\ &= P(T_{n+1} \in (t_{(i-1)}, t_{(i)})) \times \left[\frac{\tilde{n}_{c_1^i} + 1}{\tilde{n}_{c_{l_i}^i}}\right], \end{aligned}$$

where the last equation follows from $\tilde{n}_{c_j^i} = \tilde{n}_{c_{j+1}^i} + 1$, for $j = 1, \ldots, l_i - 1$. $\square$

The following corollary gives the relationship between $P(T_{n+1} \in (t_{(i)}, t_{(i+1)}))$ and the M-function value on $(t_{(i+1)}, t_{(i+2)})$, which can easily be derived from the expression for $P(T_{n+1} \in (t_{(i)}, t_{(i+1)}))$ in Theorem 2, and the M-function values as specified by rc-$A_{(n)}$.

Corollary 2

Using the same setting as in Theorem 2, we have

$$P(T_{n+1} \in (t_{(i)}, t_{(i+1)})) = M_{T_{n+1}}(t_{(i+1)}, t_{(i+2)}) \text{ for } i = 0, 1, \ldots, u-1. \tag{3.2}$$

Example 7 (continued)

By using Theorem 2 or Corollary 1, the probabilities for T_7 in open intervals between consecutive event times can be derived, see Table 3.2. Comparison of these probabilities with the corresponding M-function values in Table 3.1 illustrates that indeed these probabilities are the sums of all M-function values per interval, and that $P(T_7 \in (t_{(i)}, t_{(i+1)})) = M_{T_7}(t_{(i+1)}, t_{(i+2)})$ for $i = 0, 1, \ldots, 3$.

$(t_{(i)}, t_{(i+1)})$	$P(T_7 \in (t_{(i)}, t_{(i+1)}))$
(0, 3)	5/35
(3, 6)	6/35
(6, 8)	8/35
(8, 9)	8/35
$(9, \infty)$	8/35

Table 3.2: The probabilities for $T_7 \in (t_{(i)}, t_{(i+1)})$ (Example 7).

3.4.2 Imprecise probabilities based on rc-$A_{(n)}$

The assumption rc-$A_{(n)}$ provides a partially specified probability distribution for the observable random quantity T_{n+1}, via M-function values on the intervals $(t_{(i)}, t_{(i+1)})$ and $(c_k^i, t_{(i+1)})$, for $i = 0, 1, \ldots, u$ and $k = 1, \ldots, l_i$, with $t_{(0)} = 0$ and $t_{(u+1)} = \infty$. These M-function values can be used to derive lower and upper probabilities for events of interest in terms of T_{n+1}, via the same principle used by Augustin and Coolen [4] for $A_{(n)}$-based inference, where no additional assumptions on the distribution of the probability masses per interval are made.

For a sub-interval (a, b) of $(0, \infty)$, it is clear that precise values for probabilities of events $T_{n+1} \in (a, b)$ can be derived when a and b are two event times (or 0 or ∞). However, for other choices of (a, b), precise values for the probability of an event $T_{n+1} \in (a, b)$ cannot be derived. Instead *lower and upper probabilities* for such events $T_{n+1} \in (a, b)$ are introduced. The lower probability is derived by summing only the probability masses that necessarily must be in (a, b), so it is a sum of M-function values for T_{n+1} on intervals which are completely within (a, b). The upper probability for this event is derived by summing all the probability masses that can be in (a, b), so it is a sum of the M-function values for T_{n+1} on intervals that have a non-empty intersection with (a, b). The justification of these lower and upper probabilities is as the optimal lower and upper bounds, denoted by $\underline{P}(T_{n+1} \in (a, b))$ and $\overline{P}(T_{n+1} \in (a, b))$, which represent the maximum lower bound and the minimum upper bound for $P(T_{n+1} \in (a, b))$, respectively. Using the notation presented in $A_{(n)}$-based inference by Augustin and Coolen [4], these two lower and upper probabilities can be expressed as

$$\underline{P}(T_{n+1} \in (a,b)) = \sum_{(x,y)\subset(a,b)} M_{T_{n+1}}(x,y), \tag{3.3}$$

$$\overline{P}(T_{n+1} \in (a,b)) = \sum_{(x,y)\cap(a,b)\neq\emptyset} M_{T_{n+1}}(x,y). \tag{3.4}$$

3.4.3 Lower and upper survival functions

In survival analysis one is often interested in the survival function, which represents the probability that an individuals lifetime T exceeds a certain time t, denoted as $S_T(t) = P(T > t)$. On the basis of data including right-censored observations, as described in Section 3.2, the assumption rc-$A_{(n)}$ can be used to derive bounds for the survival function of T_{n+1}, i.e. optimal lower and upper bounds for $S_{T_{n+1}}(t) = P(T_{n+1} > t)$, for $t \geq 0$. In this subsection, these bounds will be discussed, with the maximum lower bound called the *lower survival function*, denoted by $\underline{S}_{T_{n+1}}(t) = \underline{P}(T_{n+1} > t)$, and the minimum upper bound called the *upper survival function*, denoted by $\overline{S}_{T_{n+1}}(t) = \overline{P}(T_{n+1} > t)$. These imprecise probabilities can be derived as described in Subsection 3.4.2, as the particular case of imprecise probabilities for events $T_{n+1} \in (t, \infty)$. According to the discussion of imprecise probabilities in Subsection 3.4.2, it is clear that the lower and upper survival functions are equal at observed event times $t_{(i)}$, and at such times their value can be derived using the precise probabilities presented in Subsection 3.4.1,

$$\underline{S}_{T_{n+1}}(t_{(i)}) = \overline{S}_{T_{n+1}}(t_{(i)}) = \sum_{j=i}^{u} P(T_{n+1} \in (t_{(j)}, t_{(j+1)})), \quad \text{for} \quad i = 0, 1, \ldots, u,$$

where $t_{(0)} = 0$, at which value both the lower and upper survival function are equal to one, assuming that there are no observed events at $t = 0$.

(I) The upper survival function

The upper survival function for T_{n+1} is easiest to derive, due to the fact that the M-function values in rc-$A_{(n)}$ are all defined on intervals with an observed event time (or infinity) as right end-point. To derive $\overline{S}_{T_{n+1}}(t)$, for $t > 0$, as much probability mass as possible, as described by the M-function values, must be put in the interval (t, ∞). At $t \in (t_{(i)}, t_{(i+1)})$, the upper survival function is therefore the sum of all the M-function values defined on intervals starting at $t_{(i)}$ or greater values. Here we use that the probability masses, specified by the M-function values, on intervals starting at right-censoring times $c_k^i \in (t_{(i)}, t_{(i+1)})$, $k = 1, \ldots, l_i$, are all defined on intervals $(c_k^i, t_{(i+1)})$, with no further restriction on these probability masses, which therefore can always be put in the subinterval $(t, t_{(i+1)})$ of $(t_{(i)}, t_{(i+1)})$. This implies that, for all $i = 0, 1, \ldots, u$,

$$\overline{S}_{T_{n+1}}(t) = \overline{S}_{T_{n+1}}(t_{(i)}), \quad \text{for all} \quad t \in [t_{(i)}, t_{(i+1)}).$$

Expressing this upper survival function via M-function values gives

$$\overline{S}_{T_{n+1}}(t) = \sum_{j=i}^{u} \left[M_{T_{n+1}}(t_{(j)}, t_{(j+1)}) + \sum_{k=1}^{l_j} M_{T_{n+1}}(c_k^j, t_{(j+1)}) \right]. \tag{3.5}$$

An alternative expression, using the probabilities for T_{n+1} in the intervals between two consecutive event times (or zero and infinity) as described in Subsection 3.4.1, is

$$\overline{S}_{T_{n+1}}(t) = \sum_{j=i}^{u} P(T_{n+1} \in (t_{(j)}, t_{(j+1)})). \tag{3.6}$$

Obviously, the upper survival function $\overline{S}_{T_{n+1}}(t)$ is a step-function, which is constant between observed event times, and decreases at event time $t_{(i)}$ by the value $P(T_{n+1} \in (t_{(i-1)}, t_{(i)}))$. This implies that, on the interval $[0, t_{(1)})$, the upper survival function is equal to one, while on the interval $[t_{(u)}, \infty)$ it is a positive constant, which is clearly a consequence of the fact that no further assumptions are added to the data, so also no assumptions on the tail of the distribution. The upper survival function does not change value at observed censoring times, but censorings do affect the amount with which this function decreases at later event times.

(II) The lower survival function

At $t \in (t_{(i)}, t_{(i+1)})$, the lower survival function for T_{n+1} at $t > 0$ is derived by only summing the M-function values for intervals which are entirely within (t, ∞). This leads to

$$\begin{aligned} \underline{S}_{T_{n+1}}(t) &= \sum_{j=i+1}^{u} P(T_{n+1} \in (t_{(j)}, t_{(j+1)})) + \sum_{\{k: c_k^i \geq t\}} M_{T_{n+1}}(c_k^i, t_{(i+1)}), \\ & \quad \text{for } t \in (t_{(i)}, t_{(i+1)}). \end{aligned} \tag{3.7}$$

The sum of the M-function values in formula (3.7) is only over intervals starting at observed censoring times $c_k^i \in [t, t_{(i+1)})$, the effects of later censorings are of course included in the values $P(T_{n+1} \in (t_{(j)}, t_{(j+1)}))$, for $j = i+1, \ldots, u$. This lower survival function can be derived by calculating the relevant probabilities and M-function values, as presented in Subsection 3.4.1 and Section 3.3, respectively.

The lower survival function for T_{n+1} is also a step-function, decreasing at right-censoring times as well as at event times. A great advantage of this is in the graphical presentation of these functions, as the effects of all observations on the lower and upper survival functions can be seen clearly. The lower survival function is zero beyond the largest observation, both if this largest observation is an event time or a censoring time. Calculation of the lower survival function can be slightly simplified if the upper survival function is available, via (for $i = 0, 1, \ldots, u$)

$$\begin{aligned} \underline{S}_{T_{n+1}}(t) &= \overline{S}_{T_{n+1}}(t) - P(T_{n+1} \in (t_{(i)}, t_{(i+1)})) + \sum_{\{k: c_k^i \geq t\}} M_{T_{n+1}}(c_k^i, t_{(i+1)}) \\ &= S_{T_{n+1}}(t_{(i+1)}) + \sum_{\{k: c_k^i \geq t\}} M_{T_{n+1}}(c_k^i, t_{(i+1)}), \\ &\quad \text{for } \ t \in (t_{(i)}, t_{(i+1)}), \end{aligned} \tag{3.8}$$

We use the didactic example to illustrate the lower and upper survival functions for T_{n+1}. Subsection 3.4.4 presents a larger example with data from the literature.

Example 7 (continued)
Table 3.3 gives the lower and upper survival functions for T_7, specified on intervals created by the data.

$t \in (.,\ .)$	$\underline{S}(t)$	$\overline{S}(t)$
(0, 3)	30/35	1
(3, 4)	25/35	30/35
(4, 6)	24/35	30/35
(6, 7)	18/35	24/35
(7, 8)	16/35	24/35
(8, 9)	8/35	16/35
$(9, \infty)$	0	8/35

Table 3.3: The lower and upper survival functions (Example 7).

The values of the lower and upper survival functions at observations are easily derived from Table 3.3, using 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. An effect of right-censored data is that the difference between corresponding upper and lower survival functions increases at an observed censoring time, so at times 4 and 7 in this example. The lower and upper survival function are also shown in Figure 3.1.

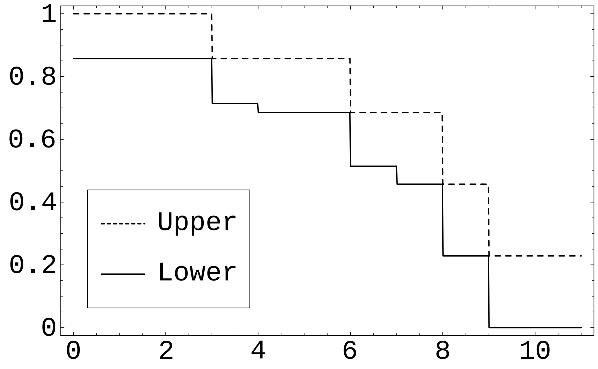

Figure 3.1: Upper and lower survival functions (Example 7).

3.4.4 Example

In this subsection nonparametric predictive inference (NPI) is illustrated, using data from the literature.

Example 8
The data for this example are given in Table 3.4, and were also used by Parmar and Machin [54, Section 4.2] to illustrate nonparametric methods for survival data. This is a subset of data obtained from 183 patients entered into a randomised Phase III trial conducted by the Medical Research Council Working Party on Advanced Carcinoma of the Cervix. The data are on survival of 30 patients with cervical cancer, recruited to a randomised trial aimed at analysing the effect of addition of a radiosensitiser to radiotherapy (New therapy - 'treatment B'), via comparison to the use of radiotherapy alone (Control - 'treatment A'). Of these 30 patients, 16 received Treatment A and 14 treatment B. The data are in days since start of the study, the event of interest is death of the patient caused by this cancer. Further variables recorded for patients in the original study are not taken into account. Of course, the inference of main interest in such a study is comparison of the two treatments. But here attention is restricted to predictive inference per treatment, and we only use this subset of all the data to illustrate NPI, so in effect considering the two treatments separately. Methods for comparison of two treatments on the basis of

such data, using NPI based on rc-$A_{(n)}$, will be presented in Chapter 5 where also these data are discussed again.

Control - A	New therapy - B
90	272
142	362
150	373
269	> 383
291	> 519
> 468	> 563
680	> 650
837	827
> 890	> 919
1037	> 978
> 1090	> 1100
> 1113	1307
1153	> 1360
1297	> 1476
1429	
> 1577	

Table 3.4: Cervical cancer survival data ($> t$ indicates right-censoring at t).

The assumptions rc-$A_{(16)}$ for treatment A, and rc-$A_{(14)}$ for treatment B, lead to partially specified probability distributions, via M-function values, for survival time of a future patient undergoing such treatment A or B, which are denoted by random quantities $T_{A,17}$ and $T_{B,15}$. These M-function values lead straightforwardly to precise probabilities for $T_{A,17}$ and $T_{B,15}$ on intervals between two consecutive event times (or zero, infinity), as described in Subsection 3.4.1. Tables 3.5 and 3.6 give the M-function values for $T_{A,17}$ and $T_{B,15}$, and probabilities $P(T_{A,17} \in (t_{A,(i)}, t_{A,(i+1)}))$, for $i = 0, 1, \ldots, 11$, and $P(T_{B,15} \in (t_{B,(j)}, t_{B,(j+1)}))$, for $j = 0, 1, \ldots, 5$.

Tables 3.5 and 3.6, show again that the probability of $T_{n+1} \in (t_{(i)}, t_{(i+1)})$ is the sum of all M-function values for T_{n+1} assigned to the interval $(t_{(i)}, t_{(i+1)})$, and its sub-intervals based on rc-$A_{(n)}$. Another relationship, described in Corollary 2 in Subsection 3.4.2, is also illustrated, namely

$$\begin{aligned} P(T_{A,17} \in (t_{A,(i)}, t_{A,(i+1)})) &= M_{T_{A,17}}(t_{A,(i+1)}, t_{A,(i+2)}) \quad \text{for } i = 0, 1 \ldots, 15, \\ P(T_{B,15} \in (t_{B,(j)}, t_{B,(j+1)})) &= M_{T_{B,15}}(t_{B,(j+1)}, t_{B,(j+2)}) \quad \text{for } j = 0, 1 \ldots, 13, \end{aligned}$$

with $t_{A,(0)} = 0,\ t_{A,(17)} = \infty$ and $t_{B,(0)} = 0,\ t_{B,(15)} = \infty$.

The assumption rc-$A_{(n)}$ can also be used to derive the upper and lower survival functions for T_{n+1}, as described in Subsection 3.4.3. Table 3.7 gives the upper and lower survival functions for $T_{A,17}$ and $T_{B,15}$, using notation $\underline{S}^A$ for $\underline{S}_{T_{A,17}}(t)$, and so

Treatment A; $T_{A,17}$		
Interval $I_{A,i}$	M-function	$P(T_{A,17} \in I_{A,i})$
$(0, 90)$	0.059	0.059
$(90, 142)$	0.059	0.059
$(142, 150)$	0.059	0.059
$(150, 269)$	0.059	0.059
$(269, 291)$	0.059	0.059
$(291, 680)$	0.059	0.064
$(468, 680)$	0.005	
$(680, 837)$	0.064	0.064
$(837, 1037)$	0.064	0.072
$(890, 1037)$	0.008	
$(1037, 1153)$	0.072	0.101
$(1090, 1153)$	0.012	
$(1113, 1153)$	0.017	
$(1153, 1297)$	0.101	0.101
$(1297, 1429)$	0.101	0.101
$(1429, \infty)$	0.101	0.202
$(1577, \infty)$	0.101	

Table 3.5: M-function values for $T_{A,17}$ and $P(T_{A,17} \in I_{A,i})$ (Example 8).

Treatment B; $T_{B,15}$		
Interval $I_{B,j}$	M-function	$P(T_{B,15} \in I_{B,j})$
$(0, 272)$	0.067	0.067
$(272, 362)$	0.067	0.067
$(362, 373)$	0.067	0.067
$(373, 827)$	0.067	0.100
$(383, 827)$	0.006	
$(519, 827)$	0.007	
$(563, 827)$	0.009	
$(650, 827)$	0.011	
$(827, 1307)$	0.100	0.175
$(919, 1307)$	0.017	
$(978, 1307)$	0.023	
$(1100, 1307)$	0.035	
$(1307, \infty)$	0.175	0.525
$(1360, \infty)$	0.088	
$(1476, \infty)$	0.262	

Table 3.6: M-function values for $T_{B,15}$ and $P(T_{B,15} \in I_{B,j})$ (Example 8).

Control; $T_{A,17}$			Treatment; $T_{B,15}$		
$t \in (.,.)$	$\underline{S}^A(t)$	$\overline{S}^A(t)$	$t \in (.,.)$	$\underline{S}^B(t)$	$\overline{S}^B(t)$
$(0, 90)$	0.941	1	$(0, 272)$	0.933	1
$(90, 142)$	0.882	0.941	$(272, 362)$	0.867	0.933
$(142, 150)$	0.824	0.882	$(362, 373)$	0.800	0.867
$(150, 269)$	0.765	0.824	$(373, 383)$	0.733	0.800
$(269, 291)$	0.706	0.765	$(383, 519)$	0.727	0.800
$(291, 468)$	0.647	0.706	$(519, 563)$	0.720	0.800
$(468, 680)$	0.642	0.706	$(563, 650)$	0.711	0.800
$(680, 837)$	0.578	0.642	$(650, 827)$	0.700	0.800
$(837, 890)$	0.513	0.578	$(827, 919)$	0.600	0.700
$(890, 1037)$	0.505	0.578	$(919, 978)$	0.583	0.700
$(1037, 1090)$	0.433	0.505	$(978, 1100)$	0.560	0.700
$(1090, 1113)$	0.421	0.505	$(1100, 1307)$	0.525	0.700
$(1113, 1153)$	0.404	0.505	$(1307, 1360)$	0.350	0.525
$(1153, 1297)$	0.303	0.404	$(1360, 1476)$	0.263	0.525
$(1297, 1429)$	0.202	0.303	$(1476, \infty)$	0	0.525
$(1429, 1577)$	0.101	0.202			
$(1577, \infty)$	0	0.202			

Table 3.7: Lower and upper survival functions for cervical cancer example.

on, based on the assumptions rc-$A_{(16)}$ and rc-$A_{(14)}$. The lower and upper survival functions, per treatment, are indeed equal at observed event times, and the upper survival function is constant between observed event times, while the lower survival function decreases at each observation. The difference between corresponding upper and lower survival functions increases at observed censoring times. If there are more censorings, this difference becomes larger, which is illustrated by the fact that the difference for group B becomes larger than for group A, for large t, in this example. The upper survival functions are equal to one between zero and the first event time, while the lower survival functions are equal to zero beyond the largest observation. Figures 3.2 presents plots of these lower and upper survival functions for treatment A and treatment B.

These lower and upper survival functions suggest that survival tends to be better under the treatment B than the treatment A. We will return to this example in Chapter 5 to investigate this suggestion further.

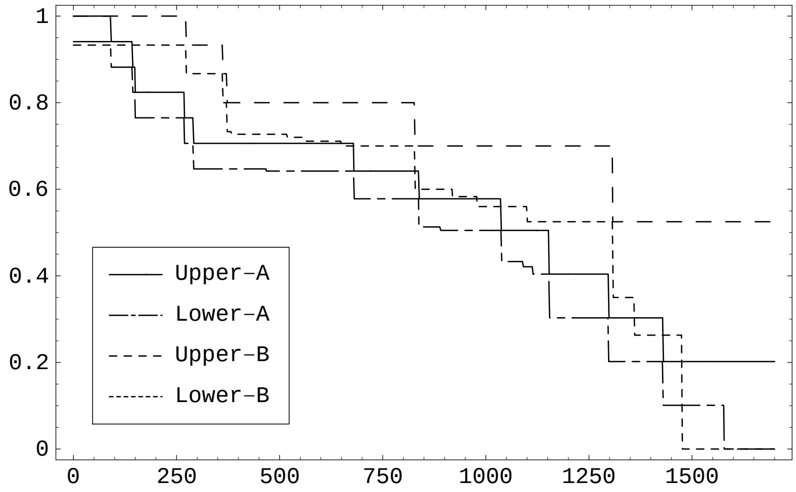

Figure 3.2: Upper and lower survival functions (Example 8).

3.5 Comparison with alternative nonparametric methods

Section 3.4 presented a nonparametric predictive method for inference based on the assumption rc-$A_{(n)}$ in the case of data including right-censored observations. As introduced in Chapter 2, there are many nonparametric methods which are used in statistical analysis. In this section, we compare this new inferential method with two alternative nonparametric methods, namely the Berliner-Hill method and the Kaplan-Meier method, as described in Sections 2.2 and 2.4.

3.5.1 Comparison with the Berliner-Hill method

The Berliner-Hill method derives predictive probabilities for T_{n+1}, which we denote as P^{BH}, to be in intervals between consecutive event times, and these probabilities turn out to be identical to the rc-$A_{(n)}$-based $P(T_{n+1} \in (t_{(i)}, t_{(i+1)}))$.

Lemma 2

For n given observations consisting of u event times, $t_{(1)} < t_{(2)} < \ldots < t_{(u)}$, and v right-censored observations, $c_{(1)} < c_{(2)} < \ldots < c_{(v)}$, let $t_{(0)} = 0$, and $t_{(u+1)} = \infty$, then we have

$$P(T_{n+1} \in (t_{(i)}, t_{(i+1)})) = P^{BH}(T_{n+1} \in (t_{(i)}, t_{(i+1)})), \quad \text{for} \;\; i = 0, 1, \ldots, u.$$

Proof

Let $\tilde{l}_i = \sum_{k=0}^{i} l_k$ and $\lambda_i = 1/[n - (i-1) - \tilde{l}_i]$, for $i = 0, 1, \ldots, u$. First consider the interval $(0, t_1)$. Based on the assumption $A_{(n)}$, and using Theorem 2,

$$P(T_{n+1} \in (0, t_{(1)})) = \frac{1}{n - \tilde{l}_0 + 1} = P^{BH}(T_{n+1} \in (0, t_{(1)})).$$

For the interval (t_1, t_2), using Corollary 1 from Section 3.4, and $P(T_{n+1} \in (0, t_{(1)}))$ from above, we get

$$\begin{aligned}
P(T_{n+1} \in (t_1, t_2)) &= P(T_{n+1} \in (0, t_{(1)})) \times \left[\frac{\tilde{n}_{c_1^1} + 1}{\tilde{n}_{c_{l_1}^1}}\right] \\
&= \frac{1}{n - \tilde{l}_0 + 1} \times \frac{n - \tilde{l}_0}{n - \tilde{l}_1} \\
&= \frac{n - \tilde{l}_0}{n - \tilde{l}_0 + 1} \times \frac{1}{n - \tilde{l}_1} \\
&= (1 - \lambda_0) \times \lambda_1 \\
&= P^{BH}(T_{n+1} \in (t_{(1)}, t_{(2)})).
\end{aligned}$$

Now induction can be used to complete the proof. Assume, for $j \in \{2, \ldots, u\}$, that

$$P(T_{n+1} \in (t_{(j-1)}, t_{(j)})) = P^{BH}(T_{n+1} \in (t_{(j-1)}, t_{(j)})).$$

Corollary 1 from Section 3.4 gives

$$P(T_{n+1} \in (t_{(j)}, t_{(j+1)})) = P(T_{n+1} \in (t_{(j-1)}, t_{(j)})) \times \frac{n - \tilde{l}_{j-1} - j + 1}{n - \tilde{l}_j - j + 1}.$$

According to the Berliner-Hill method,

$$P^{BH}(T_{n+1} \in (t_{(j-1)}, t_{(j)})) = (1 - \lambda_0) \times (1 - \lambda_1) \times \cdots (1 - \lambda_{j-2}) \times \lambda_{j-1},$$

with the induction assumption, this gives

$$\begin{aligned} P(T_{n+1} \in (t_{(j)}, t_{(j+1)})) &= (1-\lambda_0) \times (1-\lambda_1) \times \cdots (1-\lambda_{j-2}) \times \lambda_{j-1} \times \\ &\quad \frac{n - \tilde{l}_{j-1} - j + 1}{n - \tilde{l}_j - j + 1} \\ &= (1-\lambda_0) \times (1-\lambda_1) \times \cdots (1-\lambda_{i-2}) \times \\ &\quad \frac{n - \tilde{l}_{j-1} - j + 1}{n - \tilde{l}_{j-1} - j + 2} \times \frac{1}{n - \tilde{l}_j - j + 1} \\ &= (1-\lambda_0) \times (1-\lambda_1) \times \cdots (1-\lambda_{j-1}) \times \lambda_j \\ &= P^{BH}(T_{n+1} \in (t_{(j-1)}, t_{(j)})), \end{aligned}$$

which completes the proof of Lemma 2. □

This similarity between these two methods is intuitively logical, as Berliner and Hill assume that the censorings, under PCI, still happened in the same intervals created by the consecutive lifetimes, hence the number of censorings in each such interval remains the same. When considering our method, it is clear that the total probability assigned to such intervals $(t_{(i)}, t_{(i+1)})$ does not depend on the exact location of the censoring times within these intervals.

The most important difference between these two methods shows in the survival functions. Berliner and Hill [6] suggest two methods for specifying the survival function for T_{n+1} in more detail. One is to distribute the probability mass uniformly per interval $(t_{(i)}, t_{(i+1)})$, which is attractive as it leads to a continuous and precisely specified survival function. Let us denote this 'uniform Berliner-Hill survival function' by $S^{BH}_{T_{n+1}}(t)$. For $t \in (t_{(i)}, t_{(i+1)})$, $i = 0, 1, \ldots, u$, assuming t_{u+1} is a finite upper bound for T_{n+1} (else S^{BH} cannot be defined on the final interval), then

$$\begin{aligned} S^{BH}_{T_{n+1}}(t) &= \overline{S}_{T_{n+1}}(t) - \int_{t_{(i)}}^{t} \frac{P(T_{n+1} \in (t_{(i)}, t_{(i+1)}))}{t_{(i+1)} - t_{(i)}} dt \\ &= \overline{S}_{T_{n+1}}(t) - \frac{t - t_{(i)}}{t_{(i+1)} - t_{(i)}} P(T_{n+1} \in (t_{(i)}, t_{(i+1)})). \end{aligned} \quad (3.9)$$

From (3.9), it is clear that $S^{BH}_{T_{n+1}}(t) \leq \overline{S}_{T_{n+1}}(t)$. Let us now compare the uniform Berliner-Hill survival function with our lower survival function. Let us consider $t \in (t_{(i)}, c^i_1)$, from (3.8), we have

$$\underline{S}_{T_{n+1}}(t) = \overline{S}_{T_{n+1}}(t) - M_{T_{n+1}}(t_{(i)}, t_{(i+1)}).$$

Using (3.9), the uniform Berliner-Hill survival function can be expressed further, using Corollary 1, as

$$S^{BH}_{T_{n+1}}(t) = \overline{S}_{T_{n+1}}(t) - \frac{t - t_{(i)}}{t_{(i+1)} - t_{(i)}} \times P(T_{n+1} \in (t_{(i-1)}, t_{(i)})) \times \frac{\tilde{n}_{c^i_1} + 1}{\tilde{n}_{c^i_{l_i}}}.$$

To get $S^{BH}_{T_{n+1}}(t) > \underline{S}_{T_{n+1}}(t)$, for all $t \in (t_{(i)}, c^i_1)$, we must have,

$$\frac{t - t_{(i)}}{t_{(i+1)} - t_{(i)}} < \frac{\tilde{n}_{c^i_{l_i}}}{\tilde{n}_{c^i_1} + 1}. \tag{3.10}$$

The right-hand side of (3.10) is less than one. If c^i_1 is close to $t_{(i+1)}$, it is possible that a $t \in (t_{(i)}, c^i_1)$ can be found such that the left-hand side of (3.10) exceeds the right-hand side, and so $S^{BH}_{T_{n+1}}(t) < \underline{S}_{T_{n+1}}(t)$, at that value t, so $S^{BH}_{T_{n+1}}(t)$ can be smaller than $\underline{S}_{T_{n+1}}(t)$.

Example 3 (continued)
Consider Example 3 from Chapter 2 again. Table 3.8 gives the lower and upper survival function for T_9, assuming rc-$A_{(8)}$.

$(t_{(i)}, t_{(i+1)})$	$\underline{S}_{T_9}(t)$	$\overline{S}_{T_9}(t)$
(0,2)	0.889	1
(2,3)	0.778	0.889
(3,9)	0.667	0.778
(9,10)	0.648	0.778
(10,10.5)	0.622	0.778
(10.5,11)	0.583	0.778
(11,11.5)	0.519	0.778
(11.5,12)	0.389	0.778
(12, $+\infty$)	0	0.389

Table 3.8: Lower and upper survival functions for Example 3.

Comparison of the uniform Berliner-Hill survival function and the lower survival function based on T_9, leads to $S^{BH}_{T_9}(t) < \underline{S}_{T_9}(t)$, for some t values, for example for $t \in (9, 11.5)$. Figure 3.3 presents these lower and upper survival functions, together with the uniform Berliner-Hill survival function (BH). It should be noticed that the uniform Berliner-Hill survival function is undefined beyond the largest event time due to no finite upper bound for the observations.

3.5.2 Comparison with the Kaplan-Meier method

Hill [41] compared the Berliner-Hill method with the Kaplan-Meier method, which, although its explicit inferential aim is quite different, namely estimation of the un-

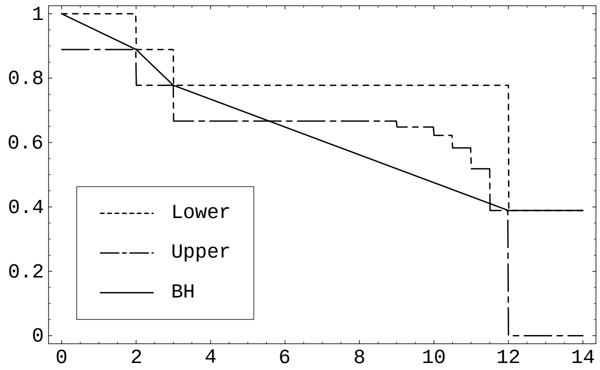

Figure 3.3: Lower and upper survival function, and the uniform BH survival function (Example 3).

derlying population survival function instead of prediction for one future individual, turns out to be pretty similar. Hill concludes that the Berliner-Hill method always gives more mass to the upper tail of the distribution than the Kaplan-Meier method does, and that such underestimation of the tail of the survival distribution is the primary practical defect of the Kaplan-Meier method. Clearly, comparison of our inferential method with the Kaplan-Meier method would lead to the identical conclusion, as the probabilities between observed event times is the same for our inferential method and the Berliner-Hill method.

Our lower survival function for T_{n+1} becomes zero after the largest observation, which is also the case for the Kaplan-Meier estimator if this observation is an event time. Our upper survival function always remains positive, unless one restricts the range of possible values for T_{n+1} by choosing a finite upper bound. Many authors have suggested minor variations to the Kaplan-Meier estimator with regard to what happens in the upper tail [2]. Comparison of the first interval is also of interest. The Kaplan-Meier method gives zero mass to the interval $(0, t_{(1)})$, even if there are censored observations in this interval. This coincides with our upper survival function, which is equal to one in this interval, yet the corresponding lower survival function is less than one, and decreases at each censoring time. The Kaplan-Meier survival function estimator, which is the nonparametric maximum likelihood estimator [46], is also regularly used to graphically present data including right-censored observations. We suggest that our lower and upper survival functions for T_{n+1} are better suited for such graphical presentation, as they indeed give a complete picture of the data, including censoring times, and can directly be interpreted from a predictive perspective. Of course, when there are many observations, most of which are event times, then all these methods are similar.

3.5.3 Example

We illustrate the comparisons in Subsections 3.5.1 and 3.5.2 by continuing with Example 8.

Example 8 (continued)
Figures 3.4 and 3.5 give the lower and upper survival functions based on rc-$A_{(n)}$, together with the uniform Berliner-Hill survival function and the Kaplan-Meier estimate, for treatments A and B, respectively. In these figures, it has been assumed that 1700 is a known upper bound for these observations and random quantity.

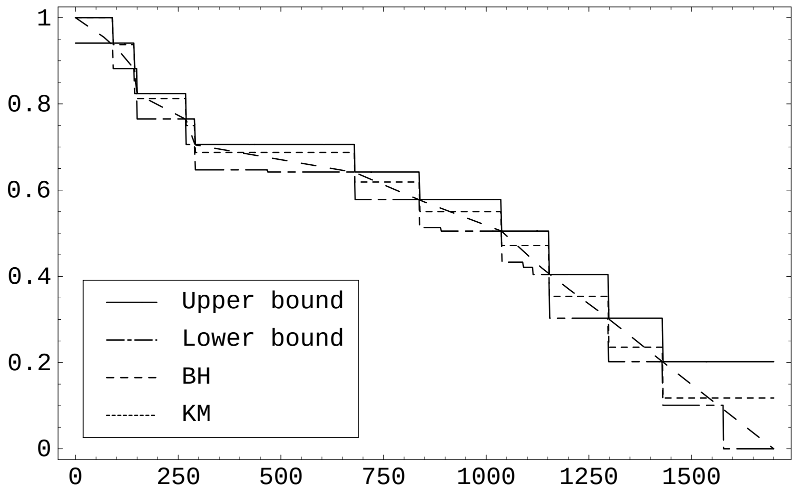

Figure 3.4: Survival functions for $T_{A,17}$ (Example 8).

From Figures 3.4 and 3.5, we see that the Kaplan-Meier estimate puts quite a lot of mass beyond the largest event time. Although our inferential method plotted this estimate as constant after the largest observed event time, it could also have left it undefined after the largest observation, which is a censoring time for both treatments. The uniform Berliner-Hill survival function beyond the largest event time, in both figures, is clearly influenced by the choice (just for the presentation) to set 1700 as an upper bound for the observations, we could also have left this uniform Berliner-Hill survival function undefined in the upper tail in both figures. However, this conveniently chosen upper bound again illustrates that the uniform Berliner-Hill survival function can actually become smaller than our lower survival function, which happens in Figure 3.4 in a very small interval just prior to the censoring time 1577.

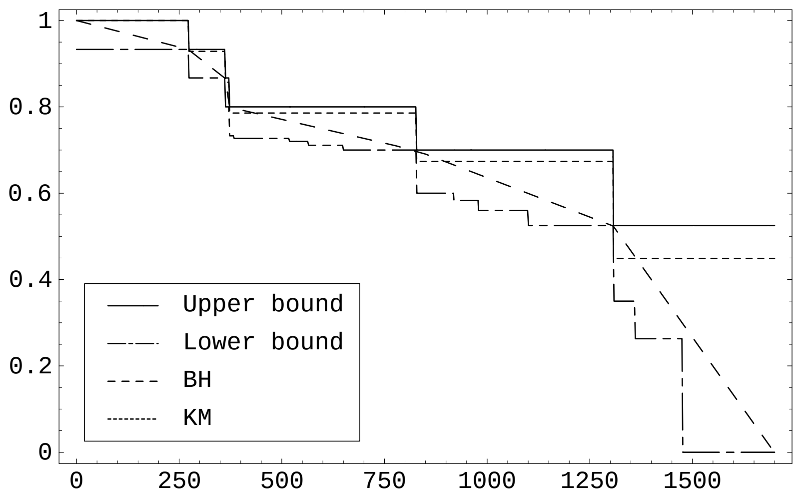

Figure 3.5: Survival functions for $T_{B,15}$ (Example 8).

At observed event times, our upper and lower survival functions are equal, and indeed also coincide with the uniform Berliner-Hill survival function. The inferential method based on rc-$A_{(n)}$, via the lower survival function, is the only one of these three that clearly indicates where censorings take place.

3.6 Treatment of ties

The assumption rc-$A_{(n)}$ and related nonparametric predictive inference, as presented in the previous sections, did not allow ties of any kind in the data set. However, tied observations frequently occur in practice, for example due to the data-collection methods, or the discrete nature of measurements or data representation (as Meeker and Escobar [52] discuss in detail). As there are two kinds of observed data, event times and censoring times, three kinds of ties may occur: *(i) Tied event times; (ii) Tied censoring times; (iii) Ties among event times and censoring times.* In this section, the discussion will focus on the possible solutions for dealing with such ties, in order to derive related predictive inferences based on rc-$A_{(n)}$.

A possible method for dealing with ties is suggested by Kaplan and Meier [46] and Berliner and Hill [6]. They suggest to break the ties by assuming that the censoring occurs just after the corresponding observed event times. The method presented in this chapter can be generalized to allow all three kinds of ties. For each case, the resulting probabilities and lower and upper survival functions can be regarded as limits of those appearing when all ties are broken by adding very small values $\epsilon > 0$ to tied observations, to get different values but keeping to the initial order as far as the non-tied observations are concerned, and then letting the ϵ's decrease to zero. For example, if data are measured in days, a tie consisting of two events and two censorings at time 8 could be broken by assuming that the events happened at times 8 and 8.001, and the censorings at 8.002 and 8.003. Then the methods presented in this chapter can be directly applied, and this would lead to inferences which are practically identical to those based on the exact observations (else, one could make the added ϵ's even smaller). This procedure is the same as was suggested by Berliner and Hill [6]. There is one interesting consequence of tied event times, say at time t_t, namely that our method, like $A_{(n)}$-based inference in general, then gives a positive predictive probability for $T_{n+1} = t_t$. However, this seems quite natural because the fact that already more than one event happened at time t_t supports the idea that future events can also happen at this time. We illustrate the method for dealing with ties in an example.

Example 9

The data for this example are given in Table 3.9, and were also used by Berliner and Hill [6] to illustrate their method in case of ties among observations. The data were given by Freireich, *et al* [34]. The data are on survival of 42 patients with acute leukemia, recruited to a randomized trial aimed at assessing the ability of 6-mercaptopurine (Treatment B) to maintain remission, via comparison to the use of a placebo (Control A). Of these 42 patients, 21 received Treatment B and 21 Control A. The data are in weeks since the start of the study, the event of interest is

Control-A		Treatment-B	
1	8	6	> 17
1	8	6	> 19
2	11	6	> 20
2	11	> 6	22
3	12	7	23
4	12	> 9	> 25
4	15	10	> 32
5	17	> 10	> 32
5	22	> 11	> 34
8	23	13	> 35
8		16	

Table 3.9: Acute leukemia survival data ($> t$ indicates right-censoring at t).

Interval	M-function	Interval		$P(T_{B,22} \in I_i)$	$\underline{S}_{T_{B,22}}(t)$	$\overline{S}_{T_{B,22}}(t)$
(0,6)	0.0455	(0,6)		0.0455	0.9545	1
6	0.0910	6		0.0910		
(6,7)	0.0480	(6,7)		0.0480	0.8157	0.8636
(7,10)	0.0480	(7,10)	(7,9)	0.0510	0.7677	0.8157
(9,10)	0.0030		(9,10)		0.7647	
(10,13)	0.0546	(10,13)	(10,11)	0.0588	0.7101	0.7647
(11,13)	0.0042		(11,13)		0.7059	
(13,16)	0.0588	(13,16)		0.0588	0.6470	0.7059
(16,22)	0.0588	(16,22)	(16,17)	0.0809	0.5882	0.6470
(17,22)	0.0059		(17,19)		0.5823	
(19,22)	0.0072		(19,20)		0.5751	
(20,22)	0.0090		(20,22)		0.5662	
(22,23)	0.0809	(22,23)		0.0809	0.4853	0.5662
$(23,\infty)$	0.0809	$(23,\infty)$	(23,25)	0.4853	0.4044	0.4853
$(25,\infty)$	0.0162		(25,32)		0.3882	
$(32,\infty)$	0.0647		(32,34)		0.3235	
$(34,\infty)$	0.0809		(34,35)		0.2426	
$(35,\infty)$	0.2426		$(35,\infty)$		0	

Table 3.10: Predictive probabilities for $T_{B,22}$ (Example 9).

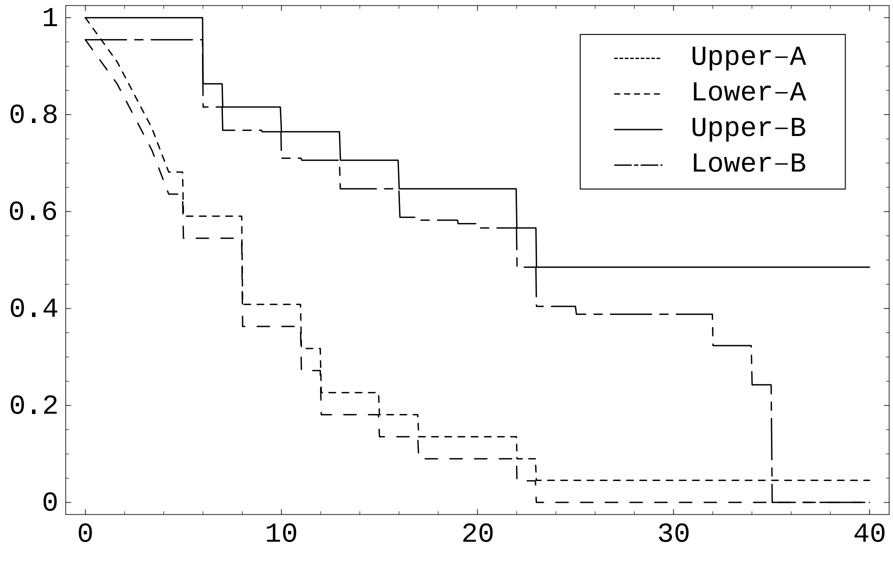

Figure 3.6: The upper and lower survival functions with ties (Example 9).

remission time of a patient. Of course, the inference of main interest in such a study is comparison of the two treatments, but here attention is restricted to predictive inference per treatment.

We are interested in the remission time of a future patient undergoing Control A

Interval	$P(T_{A,22} \in I_i)$	$\underline{S}_{T_{A,22}}(t)$	$\overline{S}_{T_{A,22}}(t)$
(0,1)	0.0455	0.9545	1
1	0.0455		
(1,2)	0.0455	0.8635	0.9090
2	0.0455		
(2,3)	0.0455	0.7725	0.8180
(3,4)	0.0455	0.7270	0.7725
4	0.0455		
(4,5)	0.0455	0.6360	0.6815
5	0.0455		
(5,8)	0.0455	0.5450	0.5905
8	0.1365		
(8,11)	0.0455	0.3630	0.4085
11	0.0455		
(11,12)	0.0455	0.2720	0.3175
12	0.0455		
(12,15)	0.0455	0.1810	0.2265
(15,17)	0.0455	0.1355	0.1810
(17,22)	0.0455	0.0900	0.1355
(22,23)	0.0455	0.0455	0.0900
$(23, \infty)$	0.0455	0	0.0455

Table 3.11: Predictive probabilities for $T_{A,22}$ (Example 9).

or Treatment B, denoted by random quantities $T_{A,22}$ and $T_{B,22}$, respectively. Tables 3.10 and 3.11 present NPI for Control A and Treatment B, respectively, by using the method discussed above for dealing with ties. For example, the probabilities for $T_{B,22}$ and $T_{A,22}$ are positive at times of ties events. Figure 3.6 gives plots of these lower and upper survival functions, suggesting that survival tends to be better under the treatment than the control.

3.7 Use of rc-$A_{(n)}$ for left-censored data

In this section, nonparametric predictive inference (NPI) based on rc-$A_{(n)}$ is discussed in the case of left-censored observations. As introduced in Section 1.3, left-censored data can also arise in survival analysis, although it is less common than right-censored data, and it means that there are observations of the form 'event has happened before time t', but nothing else is known. Hill [41] shows how the Berliner-Hill method for right-censored data can be used in such a case, by applying a monotonically decreasing transformation of the time, namely $w = 1/t$ for $t > 0$, and then using the $A_{(n)}$-based Berliner-Hill method for the transformed data, which is possible as originally left-censored observations have now become right-censored, and by transforming back one gets predictive inferences for T_{n+1} at the correct time-scale. This procedure can also be used to apply our NPI for data consisting of event time and left-censored observations. We use an example to illustrate the method.

Example 10

Suppose a set of observed data consists of four event times, 2,5,15,18, and three left-censoring times 8,10,16. How can we derive at predictive inference for T_8 based on the assumption rc-$A_{(n)}$.

For convenience, denote the original data as $t_{(1)} = 2,\ t_{(2)} = 5,\ t_{(3)} = 15,\ t_{(4)} = 18$ and $c_{(1)} = 8,\ c_{(2)} = 10,\ c_{(3)} = 16$, then a new data set can be obtained by applying the monotonically decreasing transformation as described above, leading to $t'_{(1)} = 1/18,\ t'_{(2)} = 1/15,\ t'_{(3)} = 1/5,\ t'_{(4)} = 1/2$, and $c'_{(1)} = 1/16,\ c'_{(2)} = 1/10,\ c'_{(3)} = 1/8$. Clearly, $c'_{(1)}$, $c'_{(2)}$ and $c'_{(3)}$ are right-censored observations after this transformation. In this situation, predictive inference for $T'_8 = 1/T_8$ can be derived based on this new data set and the assumption rc-$A_{(8)}$, as illustrated in Table 3.12.

Interval	M-function	Interval		$P(T'_8 \in (t'_{(i)}, t'_{(i+1)}))$	$\underline{S}_{T'_8}(t)$	$\overline{S}_{T'_8}(t)$
$(0, \frac{1}{18})$	0.1250	$(0, \frac{1}{18})$		0.1250	0.8750	1
$(\frac{1}{18}, \frac{1}{15})$	0.1250	$(\frac{1}{18}, \frac{1}{15})$	$(\frac{1}{18}, \frac{1}{16})$	0.1458	0.7500	0.8750
$(\frac{1}{16}, \frac{1}{15})$	0.0208		$(\frac{1}{16}, \frac{1}{15})$		0.7292	
$(\frac{1}{15}, \frac{1}{5})$	0.1458	$(\frac{1}{15}, \frac{1}{5})$	$(\frac{1}{15}, \frac{1}{10})$	0.2431	0.5834	0.7292
$(\frac{1}{10}, \frac{1}{5})$	0.0365		$(\frac{1}{10}, \frac{1}{8})$		0.5469	
$(\frac{1}{8}, \frac{1}{5})$	0.0608		$(\frac{1}{8}, \frac{1}{5})$		0.4861	
$(\frac{1}{5}, \frac{1}{2})$	0.2431	$(\frac{1}{5}, \frac{1}{2})$		0.2431	0.2431	0.4861
$(\frac{1}{2}, \infty)$	0.2431	$(\frac{1}{2}, \infty)$		0.2431	0	0.2431

Table 3.12: Predictive probabilities for T'_8 (Example 10).

Because of the monotonically decreasing transformation, we derive the M-function values and probabilities for T_8 by

$$M_{T_{n+1}}(t_{(i)}, t_{(i+1)}) = M_{T'_{n+1}}(1/t_{(i+1)}, 1/t_{(i)})$$

$$M_{T_{n+1}}(t_{(i)}, c^i_r) = M_{T'_{n+1}}(1/c^i_r, 1/t_{(i)})$$

$$P(T_{n+1} \in (t_{(i)}, t_{(i+1)})) = P(T'_{n+1} \in (1/t_{(i+1)}, 1/t_{(i)})),$$

where $i = 0, 1, \ldots, 4$, with $t_{(0)} = 0$ and $t_{(5)} = \infty$, and c^i_r is a left-censoring time in $(t_{(i)}, t_{(i+1)})$, for $r = 1, \ldots, l_i$. Lower and upper survival functions for T_8 are derived by

$$\underline{S}_{T_{n+1}}(t) = 1 - \overline{S}_{T'_{n+1}}(t') \quad \text{and} \quad \overline{S}_{T_{n+1}}(t) = 1 - \underline{S}_{T'_{n+1}}(t'),$$

for $t' \in (1/t_{(i+1)}, 1/t_{(i)})$.

Table 3.13 gives the related values for T_8 using the results in Table 3.12. Figure 3.7 is a plot of the lower and upper survival functions for T_8.

Interval	M-function	Interval		$P(T_8 \in (t_{(i)},\ t_{(i+1)}))$	$\underline{S}_{T_8}(t)$	$\overline{S}_{T_8}(t)$
(0,2)	0.2431	(0,2)		0.2431	0.7569	1
(2,5)	0.2431	(2,5)		0.2431	0.5138	0.7569
(5,8)	0.0608	(5,15)	(5,8)	0.2431	0.2707	0.5138
(5,10)	0.0365		(8,10)			0.4530
(5,15)	0.1458		(10,15)			0.4165
(15,16)	0.0208	(15,18)	(15,16)	0.1458	0.1249	0.2707
(15,18)	0.1250		(16,18)			0.2499
(18,∞)	0.1250	(18,∞)		0.1250	0	0.1249

Table 3.13: Predictive probabilities for T_8 (Example 10).

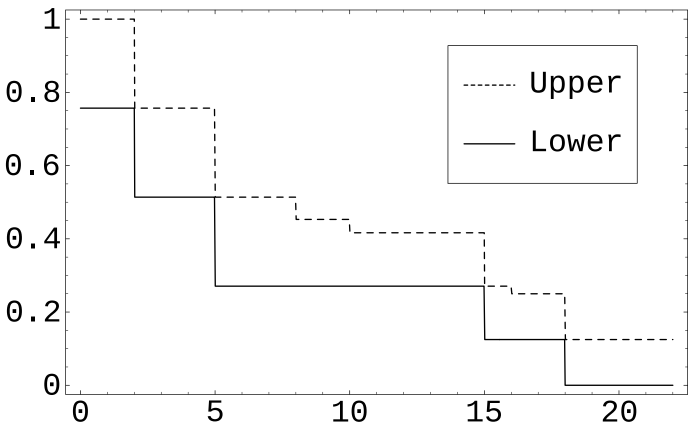

Figure 3.7: The upper and lower survival functions for T_8 (Example 10).

In this case, the lower survival function for the future observation T_{n+1}, based on the lifetime data consisting of event times and left-censoring times, is a step-function which is constant between observed event times. The lower survival function is zero beyond the largest observation, both if this largest observation is an event time or a left-censoring time. The upper survival function for T_{n+1} is also a step-function, which now decreases at each observation, so at left-censoring times as well as at event times. The upper survival function is equal to one on the interval between zero and the smallest observation. It should be addressed that the difference between upper and lower survival functions for the future observation T_{n+1}, based on the lifetime data consisting of event times and left-censoring times, is larger for smaller time t.

The method presented in this section can be used for data sets including only left-censoring times. Dealing with data including both right- and left-censored observations, or other forms of censoring, from a similar perspective, remains a topic for future research.

3.8 Concluding remarks

The assumption rc-$A_{(n)}$ provides a partially specified predictive probability distribution for a future observation, based on data including right-censored observations. The rc-$A_{(n)}$-based inference is an explicit attempt to keep structural assumptions minimal, and is therefore naturally suited if one only has extremely vague knowledge about the situation which is being modeled, other than that provided by the data set.

The assumption rc-$A_{(n)}$ is a generalization of $A_{(n)}$. An advantage of rc-$A_{(n)}$ is that predictive probabilities for a future observation can be derived using the exact censoring information, and it leads to explicit changes in the lower survival function at censoring times, which is not the case in the Berliner-Hill method [6] and the Kaplan-Meier method [46]. The inferential method based on rc-$A_{(n)}$ can be generalized to allow ties of any nature in the data. The assumption rc-$A_{(n)}$ can also be used to derive nonparametric predictive inferences for other problems formulated in terms of T_{n+1}, such as grouped data and comparison of groups of lifetime data, which will be presented in Chapters 4 and 5.

Chapter 4

Nonparametric predictive inference for grouped data

4.1 Introduction

This chapter applies nonparametric predictive inference (NPI), as presented in Chapter 3, to grouped data including right-censored observations. As stated in Section 2.5, such data frequently occur in situations concerning reliability and survival analysis. The statistical method presented for grouped data in this chapter is based on quite minimal modelling assumptions, and is directly in terms of a random quantity representing a future observation. We will assume that either a well-specified event happens, at a particular time, to each item for which we have an observation, or that a time is reported at which such an event has not yet occurred, where such right-censoring is assumed to be non-informative, as discussed in Section 1.3 and, with regard to the assumption rc-$A_{(n)}$, in Section 3.2. We restrict attention to non-negative random quantities, so to random quantities and observations on the time-axis $[0, \infty)$. However, the method presented is more widely applicable, as only a finite partition of (part of) the real line is required.

Section 4.2 discusses, for grouped lifetime data, the influence of different possible configurations of event times and censoring times, and presents the main principle which enables derivation of optimal bounds for predictive probabilities. Sections 4.3 presents lower and upper probabilities per interval, and Section 4.4 presents lower and upper predictive survival functions. In Section 4.5 the method is illustrated, and briefly compared with some alternative methods, via an example. Finally, in Section 4.6 we add some concluding remarks about the method and results presented in this chapter.

4.2 Grouped lifetime data and configurations

We consider NPI for grouped lifetime data, where event and censoring times are not actually observed, but only the numbers of each per interval are available, for a finite number of intervals forming a partition of the time-axis. We assume that indeed no further information is available on the times of events and censorings within the intervals, that the intervals have been determined independently of the data, and that censoring was non-informative, so it occurs due to a mechanism that is independent of the remaining lifetimes.

We use the same notations for grouped data introduced in Section 2.5. At the same time, we also add some new notations in this chapter. Although the exact observation times are not given for grouped data, the orderings of event and censoring times within the intervals (we will call this 'configurations') is important, therefore, let the e_z ordered event times in I_z be denoted by $t_1^z < t_2^z < \ldots < t_{e_z}^z$, and the c_z ordered censoring times in I_z by $c_1^z < c_2^z < \ldots < c_{c_z}^z$. Let $n_z = e_z + c_z$ be the total number of observations in I_z, so $n = \sum_{z=0}^{k} n_z$ be the total number of observations. Let $I_i^z = [t_i^z, t_{i+1}^z)$ denote the sub-intervals of I_z, for $i = 1, \ldots, e_z - 1$, with $I_0^z = [a_z, t_1^z)$ and $I_{e_z}^z = [t_{e_z}^z, a_{z+1})$. For convenience, we assume no ties in this notation, but ties could be included without changing the main results in Sections 4.3 and 4.4 by dealing with them as described in Section 3.6.

To develop NPI based on such data, the obvious problem is that the exact observation times, as needed for the M-function values for T_{n+1} in the definition of rc-$A_{(n)}$, are not available. These M-function values depend on the order of the event and censoring times, which is called 'the configuration' of the observations. For each configuration, we would have M-function values which partially specify the probability distribution for T_{n+1}, and we could derive corresponding optimal bounds for probabilities of the form $T_{n+1} \in B$, for $B \subset [0, \infty)$, as presented in Subsection 3.4.2. To derive such bounds for situations where B is equal to an interval I_s (or a union of several such intervals), it turns out that the exact location of the event and censoring times within the intervals is not relevant, only their configuration. As there is a finite number of possible configurations, one could calculate the M-function values according to all possible configurations, calculate optimal bounds for $T_{n+1} \in B$ for each configuration, and take the minimum (maximum) of the lower (upper) bounds per configuration as the lower (upper) probability for this event, that is the maximum lower (minimum upper) bound that can be justified without adding any further assumptions about the configurations. Although this is a natural way to proceed, the amount of computation involved would be enormous, making this approach not feasible for all but the smallest data sets. Luckily, however, we can derive these optimal bounds far quicker, as we can derive the configurations that actually lead to such bounds, therefore deleting the need to calculate M-functions for any other configurations. Such 'optimal configurations' will be presented in the next two sections, considering events $T_{n+1} \in B$ with $B = I_z$ and $B = [a_z, \infty)$. The corresponding optimal configurations are all based on the following principle.

Theorem 3 (Optimal Configuration Principle, OCP)
If a pair of neighbouring observations consists of an event time t_j and a censoring time c_l, then, for any given set of further data consisting of $u-1$ event times and $v-1$ censoring times, the particular order of t_j and c_l influences the M-function values for T_{n+1}, according to rc-$A_{(n)}$, as follows:

Let $C(tc)$ be the configuration for which $t_j < c_l$, and let $C(ct)$ be the configuration for which $c_l < t_j$. Let $M^{tc}_{T_{n+1}}$ denote the M-function values corresponding to $C(tc)$, and $M^{ct}_{T_{n+1}}$ those corresponding to $C(ct)$, based on rc-$A_{(n)}$. Then:

1. M-function values on intervals before t_j and c_l do not depend on the order of t_j and c_l. So, for intervals (t_i, t_{i+1}) and (c^i_k, t_{i+1}), with $t_{i+1} < \min(t_j, c_l)$, we have that

$$\begin{aligned} M^{tc}_{T_{n+1}}(t_i, t_{i+1}) &= M^{ct}_{T_{n+1}}(t_i, t_{i+1}), \\ M^{tc}_{T_{n+1}}(c^i_k, t_{i+1}) &= M^{ct}_{T_{n+1}}(c^i_k, t_{i+1}). \end{aligned}$$

2. M-function values on intervals after t_j and c_l are larger for $C(tc)$ than for $C(ct)$. So, for intervals (t_i, t_{i+1}), with $t_i > \max(t_j, c_l)$, and (c^i_k, t_{i+1}), with $c^i_k > \max(t_j, c_l)$, we have that

$$\begin{aligned} M^{tc}_{T_{n+1}}(t_i, t_{i+1}) &> M^{ct}_{T_{n+1}}(t_i, t_{i+1}), \\ M^{tc}_{T_{n+1}}(c^i_k, t_{i+1}) &> M^{ct}_{T_{n+1}}(c^i_k, t_{i+1}). \end{aligned}$$

Proof
The relations between the different M-function values in OCP follow directly from the definition of rc-$A_{(n)}$, as for intervals before t_j and c_l, all the factors in the products in the definitions of the M-function values are not affected by the order of t_j and c_l, while for intervals after t_j and c_l, all these factors but one remain unchanged,

and the one changing factor is larger for $C(tc)$ than for $C(ct)$, since $\tilde{n}_{c_l}$ is one smaller for $C(tc)$ than for $C(ct)$. □

It turns out that OCP is sufficient for the inferences considered in the next two sections, as the optimal configurations of the event and censoring times per interval, leading to the lower and upper probabilities presented in Sections 4.3 and 4.4, will be justified straightforwardly by OCP. If one would wish to consider inferences on events $T_{n+1} \in B$, with B not consisting of one or more of the intervals I_z, then further aspects of the configurations may need to be considered. For OCP we did not consider the tied observation $t_j = c_l$, as the way that NPI would deal with such a tie introduced, as presented in Section 3.6, would not affect the optimal configurations, which is the reason why we do not have to focus on possible ties within grouped data.

4.3 Predictive probabilities per interval

In this section, we consider the probability that a future observation falls in one of the predetermined intervals I_z, based on grouped data consisting of n observations, using the notation introduced in Section 4.2 and the assumption rc-$A_{(n)}$. Clearly, we cannot derive precise probabilities for such events, both due to the nature of the M-functions in rc-$A_{(n)}$ and the way that the data are presented, without knowledge of the exact configurations within the intervals I_z, $z = 0, 1 \ldots, k$. Instead, we derive optimal bounds for these probabilities, i.e. the maximum lower bounds and minimum upper bounds that can be justified on the basis of the grouped data and rc-$A_{(n)}$. These bounds are lower and upper probabilities [57], denoted by $\underline{P}(T_{n+1} \in I_z)$ and $\overline{P}(T_{n+1} \in I_z)$, respectively. The lower probability $\underline{P}(T_{n+1} \in I_z)$ is determined by only summing up the probability masses that necessarily must be in I_z, according to the M-function values and the particular optimal configurations for all relevant intervals for which this probability mass is minimal. The upper probability $\overline{P}(T_{n+1} \in I_z)$ is determined by summing up all probability masses that could be in I_z, with the configurations for which this total probability mass is maximal. The principle OCP, derived in Section 4.2, is sufficient to derive the configurations corresponding to these lower and upper probabilities.

We first consider the lower probabilities $\underline{P}(T_{n+1} \in I_z)$, for $z = 0, 1, \ldots, k$. For given configurations for each of the intervals, this lower probability is derived by only summing up the M-function values, according to rc-$A_{(n)}$, on intervals that are fully within $I_z = [a_z, a_{z+1})$, so summing up the values $M_{T_{n+1}}(t_i^z, t_{i+1}^z)$ and the $M_{T_{n+1}}(c_k^i, t_{i+1})$ with $(c_k^i, t_{i+1}) \subset (t_i^z, t_{i+1}^z)$, for $i = 1, \ldots, e_z - 1$.

Let us now consider the configurations for all intervals I_t, $t = 0, 1, \ldots, k$, that lead to minimum probability for T_{n+1} to be in a particular I_z, so the configurations for which the above M-function values are minimal, we denote these configurations as $\underline{C}_{I_z}(I_t)$, and they are specified in Theorem 4.

Theorem 4 (Configurations for $\underline{P}(T_{n+1} \in I_z)$)
Consider NPI, based on rc-$A_{(n)}$, for grouped data. For $z = 1, \ldots, k-1$, the following configurations lead to minimum probability mass for T_{n+1} in I_z:

$$\underline{C}_{I_z}(I_t) = \{c_{c_t}^t < t_1^t\}, \quad \text{for } t < z,$$

where $\{c_{c_t}^t < t_1^t\}$ is used to denote that all censorings in I_t are assumed to take place before the event times. The optimal configuration of the interval I_z itself is:

$$\underline{C}_{I_z}(I_z) = \{t_{e_z}^z < c_1^z\},$$

so all events are assumed to happen prior to the censorings in this interval. Finally, the configurations in intervals beyond I_z do not influence the M-function values within I_z, so we do not need to specify $\underline{C}_{I_z}(I_t)$ for $t > z$.

For I_0, the only configuration that affects the M-function values within this interval is that of I_0 itself, for which the optimal configuration is

$$\underline{C}_{I_0}(I_0) = \{t_{e_0}^0 < c_1^0\},$$

and $\underline{C}_{I_0}(I_t)$ does not need to be specified for $t > 0$.

For $I_k = [a_k, \infty)$, the configurations of all intervals I_t with $t < k$ are relevant, but the actual configuration within I_k plays no role, as this configuration would effectively only serve to move probability mass within this interval. Hence,

$$\underline{C}_{I_k}(I_t) = \{c_{c_t}^t < t_1^t\}, \quad \text{for } t < k,$$

and $\underline{C}_{I_k}(I_k)$ does not need to be specified.

Proof
These optimal configurations follow by (possibly repeated) application of OCP, presented in Section 4.2. For example, when considering I_z with $z = 1, \ldots, k$, OCP implies that, for all intervals I_t with $t < z$, any pair of neighbouring event time t_j and censoring time c_l within such an interval should be configured like $C(ct)$, in

Theorem 3, so the censoring should be assumed to occur before the event time, in order to minimize the M-function values that together make up the lower probability for $T_{n+1} \in I_z$. For any configuration, OCP can be applied repeatedly, every time moving a censoring time c_l, that is immediately to the right of an event time t_j, to the left of t_j. This can be continued until all censoring times in each interval I_t, with $t < z$, are assumed to occur before the event times in the same interval, leading to $\{c^t_{c_t} < t^t_1\}$ for $t < z$.

The optimal configuration in the interval I_z, for $z = 0, 1, \ldots, k-1$, follows from a simple argument, namely that the configuration $\underline{C}_{I_z}(I_z)$ must be the configuration that leads to maximum M-function values within the later intervals, I_t with $t > z$, as the sum of the M-function values in $I_z, I_{z+1}, \ldots, I_k$ is constant once the optimal configurations in the earlier intervals I_t with $t < z$ are determined. Hence, with a similar argument as before, (possibly repeated use of) OCP leads to

$$\underline{C}_{I_z}(I_z) = \{t^z_{e_z} < c^z_1\}$$

so all event times are assumed to take place before the censorings in I_z.

An optimal configuration not yet covered is $\underline{C}_{I_k}(I_k)$, but clearly different configurations in I_k will affect the particular M-function values on intervals within I_k, but the sum of these M-function values remains constant as there are no intervals to the right of I_k that can take over some of the probability mass that is in I_k. Similarly, changes in configurations within intervals I_t, with $t > z$, have no relevance for the M-function values within I_z, so do not need to be considered when determining $\underline{P}(T_{n+1} \in I_z)$. $\square$

With the optimal configurations for $\underline{P}(T_{n+1} \in I_z)$, as given in Theorem 4, and the definition of M-function values in rc-$A_{(n)}$, these lower probabilities can now be determined by summing only M-function values on intervals that are completely within I_z. It is important to remark that here, as in the rest of this chapter, we assume that no single observation coincides with one of the values a_z that create the grouped data partition. (This would slightly complicate matters, but as we assumed that the a_z were determined independently of the data, and indeed probably before the data became available, this seems a reasonable assumption, even more since we do not actually know the precise observation times.)

Theorem 5 ($\underline{P}(T_{n+1} \in I_z)$)
For grouped data, using the notation introduced in Section 4.2, the lower probabilities for the events $T_{n+1} \in I_z$, with $z = 0, 1, \ldots, k$, according to the assumption

rc-$A_{(n)}$, are

$$\begin{aligned}
\underline{P}(T_{n+1} \in I_0) &= \frac{e_0}{n+1}, \\
\underline{P}(T_{n+1} \in I_z) &= 0, \text{ for } z = 1, \ldots, k-1, \text{ if } e_z \in \{0, 1\}, \\
\underline{P}(T_{n+1} \in I_z) &= \frac{e_z - 1}{n+1} \times \prod_{t=0}^{z-1} \Big(1 + \frac{c_t}{n - \sum_{w=0}^{t-1} n_w - c_t + 1}\Big), \\
& \quad \text{for } z = 1, \ldots, k-1, \text{ if } e_z \geq 2, \\
\underline{P}(T_{n+1} \in I_k) &= \frac{n_k}{n+1} \times \prod_{t=0}^{k-1} \Big(1 + \frac{c_t}{n - \sum_{w=0}^{t-1} n_w - c_t + 1}\Big).
\end{aligned}$$

Proof

(i) First we consider the lower probability for $T_{n+1} \in I_0$. When $e_0 = 0$, which means no event times in I_0, the lower probability for $T_{n+1} \in I_0$ is clearly equal to zero. Now let us consider the case of $e_0 > 0$. According to Theorem 4, $\underline{C}_{I_0}(I_0)$ is given as

$$t_1^0 < \ldots < t_{e_0}^0 < c_1^0 < \ldots < c_{c_0}^0.$$

Based on this configuration, the lower probability for $T_{n+1} \in I_0$ is

$$\begin{aligned}
\underline{P}(T_{n+1} \in I_0) &= P(T_{n+1} \in I_0^0) + P(T_{n+1} \in I_1^0) + \ldots + \\
& \quad P(T_{n+1} \in I_{e_0-1}^0) + \underline{P}(T_{n+1} \in I_{e_0}^0),
\end{aligned}$$

where $\underline{P}(T_{n+1} \in I_{e_0}^0)$ is the lower bound of the probability for $T_{n+1} \in I_{e_0}^0$. From NPI based on rc-$A_{(n)}$, $\underline{P}(T_{n+1} \in I_{e_0}^0)$ can actually be zero. Obviously, for this configuration $P(T_{n+1} \in I_i^0)$ is equal to $1/(n+1)$ for all $i = 0, 1, \ldots, e_0 - 1$, so the lower probability for $T_{n+1} \in I_0$ is equal to $e_0/(n+1)$.

(ii) Now let us consider the lower probability for $T_{n+1} \in I_z$, $z = 1, \ldots, k-1$. First we consider $e_z \in \{0, 1\}$. Based on the optimal configurations (Theorem 4), it is clear that no probability mass for T_{n+1} is necessarily within I_z. So the lower probability for $T_{n+1} \in I_z$ is equal to zero according to NPI based on rc-$A_{(n)}$.

Next we consider the case $e_z \geq 2$. The lower probability for $T_{n+1} \in I_z$ is derived by summing only the probability masses that necessarily must be in I_z. Therefore, from the optimal configurations (Theorem 4), the lower probability for $T_{n+1} \in I_z$ is

$$\begin{aligned}
\underline{P}(T_{n+1} \in I_z) &= \underline{P}(T_{n+1} \in I_0^z) + P(T_{n+1} \in I_1^z) + \ldots + \\
& \quad P(T_{n+1} \in I_{e_z-1}^z) + \underline{P}(T_{n+1} \in I_{e_z}^z),
\end{aligned}$$

where $\underline{P}(T_{n+1} \in I_0^z)$ and $\underline{P}(T_{n+1} \in I_{e_z}^z)$ are lower bounds of the probabilities for $T_{n+1} \in I_0^z$ and $T_{n+1} \in I_{e_z}^z$. From NPI based on rc-$A_{(n)}$, these two lower bounds can both be zero. So, the lower probability for $T_{n+1} \in I_z$, is equal to

$$\underline{P}(T_{n+1} \in I_z) = P(T_{n+1} \in I_1^z) \ldots + P(T_{n+1} \in I_{e_z-1}^z).$$

For the optimal configuration of $\underline{P}(T_{n+1} \in I_z)$, clearly, there are no censoring times in $I_1^z, \ldots, I_{e_z-1}^z$, so

$$P(T_{n+1} \in I_1^z) = P(T_{n+1} \in I_2^z) = \ldots = P(T_{n+1} \in I_{e_z-1}^z),$$

and these probability values are equal to $P(T_{n+1} \in (t_{e_{z-2}}^{z-2}, t_1^{z-1}))$ (Corollary 1). For the optimal configurations, we can derive $P(T_{n+1} \in (t_{e_{z-2}}^{z-2}, t_1^{z-1}))$. To complete the proof we now show that the probability for $T_{n+1} \in (t_{e_{z-2}}^{z-2}, t_1^{z-1})$, corresponding to such configurations, is

$$P(T_{n+1} \in (t_{e_{z-2}}^{z-2}, t_1^{z-1})) = \frac{1}{n+1} \times \prod_{t=0}^{z-1} \Big(1 + \frac{c_t}{n - \sum_{w=0}^{t-1} n_w - c_t + 1}\Big). \tag{4.1}$$

For the interval $(0, t_1^0)$, under $\underline{C}_{I_z}(I_0)$, $\{c_{c_0}^0 < t_1^0\}$, using Theorem 2 of Section 3.4, we have

$$\begin{aligned} P(T_{n+1} \in (0,\ t_1^0)) &= \frac{1}{n+1} \prod_{\{r:\, c_{(r)} < t_1^0\}} \frac{\tilde{n}_{c_{(r)}} + 1}{\tilde{n}_{c_{(r)}}} \\ &= \frac{1}{n+1} \times \frac{n+1}{n - c_0 + 1} \\ &= \frac{1}{n+1} \times \Big(1 + \frac{c_0}{n - c_0 + 1}\Big), \end{aligned}$$

as stated in (4.1). Next we use an induction step to complete the proof. Assume, for $h \in \{2, \ldots, z-1\}$, that

$$P(T_{n+1} \in (t_{e_{h-2}}^{h-2}, t_1^{h-1})) = \frac{1}{n+1} \times \prod_{t=0}^{h-1} \Big(1 + \frac{c_t}{n - \sum_{w=0}^{t-1} n_w - c_t + 1}\Big).$$

For $\underline{C}_{I_z}(I_{h-1})$ and $\underline{C}_{I_z}(I_h)$ using Corollary 1 of Section 3.4, we get

$$\begin{aligned} P(T_{n+1} \in (t_{e_{h-1}}^{h-1}, t_1^h)) &= P(T_{n+1} \in (t_{e_{h-1}-1}^{h-1},\ t_{e_{h-1}}^{h-1})) \times \Big(\frac{\tilde{n}_{c_1^h} + 1}{\tilde{n}_{c_{c_h}^h}}\Big) \\ &= P(T_{n+1} \in (t_{e_{h-1}-1}^{h-1}, t_{e_{h-1}}^{h-1})) \times \\ &\quad \Big(1 + \frac{c_h}{n - \sum_{w=0}^{h-1} n_w - c_h + 1}\Big). \end{aligned}$$

For $\underline{C}_{I_z}(I_{h-1})$, it is known that there are no censoring times in $I_1^{h-1}, \ldots, I_{e_{h-1}-1}^{h-1}$, so we have

$$P(T_{n+1} \in (t_{e_{h-1}-1}^{h-1}, t_{e_{h-1}}^{h-1})) = P(T_{n+1} \in (t_{e_{h-2}}^{h-2}, t_1^{h-1})).$$

Continuing the calculation above,

$$
\begin{aligned}
P(T_{n+1} \in (t^{h-1}_{e_{h-1}}, t^h_1)) &= P(T_{n+1} \in (t^{h-2}_{e_{h-2}}, t^{h-1}_1)) \times \\
&\quad \Big(1 + \frac{c_h}{n - \sum_{w=0}^{h-1} n_w - c_h + 1}\Big) \\
&= \frac{1}{n+1} \times \prod_{t=0}^{h-1} \Big(1 + \frac{c_t}{n - \sum_{k=0}^{t-1} n_w - c_t + 1}\Big) \\
&\quad \times \Big(1 + \frac{c_h}{n - \sum_{w=0}^{h-1} n_w - c_h + 1}\Big) \\
&= \frac{1}{n+1} \times \prod_{t=0}^{h} \Big(1 + \frac{c_j}{n - \sum_{w=0}^{t-1} n_w - c_t + 1}\Big),
\end{aligned}
$$

which completes the proof for the probability of $P(T_{n+1} \in (t^{z-2}_{e_{z-2}}, t^{z-1}_1))$, given as (4.1), so the lower probability for $T_{n+1} \in I_z$, for $z = 1, \ldots, k-1$, if $e_z \geq 2$, is as stated in the corollary.

(iii) Finally we consider the lower probability for $T_{n+1} \in I_k$, which is

$$
\underline{P}(T_{n+1} \in I_k) = \underline{P}(T_{n+1} \in I^k_0) + P(T_{n+1} \in I^k_1) + \ldots + P(T_{n+1} \in I^k_{e_k}).
$$

As presented in Theorem 4, under $\underline{C}_{I_k}(I_k)$, the order of event times and censoring times is irrelevant. But for calculation of $\underline{P}(T_{n+1} \in I_k)$, we need to assume an order of these lifetimes in I_k. For convenience to calculate, here we assume

$$
t^k_1 < c^k_1 < \cdots < c^k_{c_k} < t^k_2.
$$

For such an order of lifetimes in I_k, we know that $\underline{P}(T_{n+1} \in I^k_0) = 0$, and $P(T_{n+1} \in I^k_1) = \cdots = P(T_{n+1} \in I^k_{e_k})$. Similar to the analysis in (i) and (ii), $\underline{P}(T_{n+1} \in I_k)$ can be derived via

$$
\underline{P}(T_{n+1} \in I_k) = e_k P(T_{n+1} \in I^k_1).
$$

Under the optimal configuration for $\underline{P}(T_{n+1} \in I_k)$, we can use formula (4.1) to get

$$
P(T_{n+1} \in (t^{k-2}_{e_{k-2}}, t^{k-1}_1)) = \frac{1}{n+1} \times \prod_{t=0}^{k-1} \Big(1 + \frac{c_t}{n - \sum_{w=0}^{t-1} n_w - c_t + 1}\Big).
$$

By Corollary 1 of Section 3.4, we have

$$
\begin{aligned}
P(T_{n+1} \in I^k_1) &= P(T_{n+1} \in (t^{k-2}_{e_{k-2}},\ t^{k-1}_1)) \times \frac{e_k + c_k}{e_k} \\
&= \frac{e_k + c_k}{(n+1)e_k} \times \prod_{t=0}^{k-1} \Big(1 + \frac{c_t}{n - \sum_{w=0}^{t-1} n_w - c_t + 1}\Big).
\end{aligned}
$$

Hence, by using $\underline{P}(T_{n+1} \in I_k) = e_k P(T_{n+1} \in I^k_1)$, the lower probability for $T_{n+1} \in I_k$ is as stated in the theorem. □

These optimal configurations can of course not hold simultaneously, if indeed there are censored observations in the relevant intervals, which immediately confirms the super-additivity of such lower probabilities [57], i.e.

$$\underline{P}(T_{n+1} \in I_z \cup I_t) \geq \underline{P}(T_{n+1} \in I_z) + \underline{P}(T_{n+1} \in I_t), \text{ for } z \neq t,$$

where the inequality is strict if there are censored observations in intervals which affect the relevant M-function values as described in Theorem 4.

The upper probabilities $\overline{P}(T_{n+1} \in I_z)$, for $z = 0, 1, \ldots, k$, are derived in a similar way as the lower probabilities above, but now all the probability masses, as specified by the M-function values, that could actually be within I_z are included. The optimal configurations are specified in Theorem 6, using $\overline{C}_{I_z}(I_t)$ to denote the configuration in I_t that leads to maximum probability for T_{n+1} to be in I_z.

Theorem 6 (Configurations for $\overline{P}(T_{n+1} \in I_z)$)
Consider NPI, based on rc-$A_{(n)}$, for grouped data. For $z = 1, \ldots, k-1$, the following configurations lead to maximum probability mass for T_{n+1} in I_z:

$$\overline{C}_{I_z}(I_t) = \{t^t_{e_t} < c^t_1\}, \quad \text{for } t < z.$$

The optimal configuration of the interval I_z itself is:

$$\overline{C}_{I_z}(I_z) = \{c^z_{c_z} < t^z_1\}.$$

The configurations in intervals beyond I_z do not influence the M-function values within I_z, so we do not need to specify $\overline{C}_{I_z}(I_t)$ for $t > z$.

For I_0, the optimal configuration is

$$\overline{C}_{I_0}(I_0) = \{c^0_{c_0} < t^0_1\},$$

and $\overline{C}_{I_0}(I_t)$ does not need to be specified for $t > 0$.

For I_k, the optimal configurations are:

$$\overline{C}_{I_k}(I_t) = \{t^t_{e_t} < c^t_1\}, \quad \text{for } t < k,$$

and $\overline{C}_{I_k}(I_k)$ does not need to be specified.

Proof
These optimal configurations follow again from OCP, along the same lines as the proof of Theorem 4, but of course now using OCP to derive the maximum probability mass in I_z, leading to configurations where, when compared to those in Theorem

6, the order of all events and all censorings per interval is turned around. $\square$

The optimal configurations for $\overline{P}(T_{n+1} \in I_z)$, as given in Theorem 6, and the definition of M-function values in rc-$A_{(n)}$, lead to the upper probabilities in Theorem 7, where all M-function values on intervals that have non-empty intersection with I_z are summed up to give the upper probability for T_{n+1} to be in I_z.

Theorem 7 ($\overline{P}(T_{n+1} \in I_z)$)
For grouped data, using the notation introduced in Section 4.2, the upper probabilities for the events $T_{n+1} \in I_z$, with $z = 0, 1, \ldots, k$, according to the assumption rc-$A_{(n)}$, are

$$\begin{aligned}\overline{P}(T_{n+1} \in I_z) &= \frac{e_z + 1}{n+1} \times \Big[\prod_{t=0}^{z-2}\Big(1 + \frac{c_t}{n - \sum_{w=0}^{t} n_w - c_t + 1}\Big)\Big] \\ &\quad \times \Big(1 + \frac{c_{z-1} + c_z}{n - \sum_{w=0}^{z-1} n_w - c_z + 1}\Big),\end{aligned}$$

where the product term is defined as one if the product is over an empty set, and $c_{-1} = 0$.

Proof
For interval $I_z = [a_z, a_{z+1})$, $z = 0, 1, \ldots, k$, the upper probability for $T_{n+1} \in I_z$ is derived by summing all probability masses which can be in I_z. There are e_z event times in I_z, leading to $e_z + 1$ sub-intervals, I_0^z, $I_1^z, \ldots, I_{e_z}^z$, created by these event times, that have non-empty intersection with I_z. Therefore, under the optimal configuration, the probability for $T_{n+1} \in I_z$ is

$$\begin{aligned}\overline{P}(T_{n+1} \in I_z) &= \overline{P}(T_{n+1} \in I_0^z) + P(T_{n+1} \in I_1^z) + \ldots + \\ &\quad P(T_{n+1} \in I_{e_z-1}^z) + \overline{P}(T_{n+1} \in I_{e_z}^z),\end{aligned}$$

where $\overline{P}(T_{n+1} \in I_0^z)$ and $\overline{P}(T_{n+1} \in I_{e_z}^z)$ are upper bounds of the probabilities for $T_{n+1} \in I_0^z$ and $T_{n+1} \in I_{e_z}^z$, respectively. It is clear that there are no censoring times in $I_1^z, \ldots, I_{e_z-1}^z$ under the optimal configuration for $\overline{P}(T_{n+1} \in I_z)$, so

$$P(T_{n+1} \in I_1^z) = P(T_{n+1} \in I_2^z) = \ldots = P(T_{n+1} \in I_{e_z-1}^z).$$

Under $\overline{C}_{I_z}(I_{z-1})$ and $\overline{C}_{I_z}(I_z)$, using Corollary 1 of Section 3.4, we get

$$P(T_{n+1} \in I_1^z) = P(T_{n+1} \in I_1^{z-1}) \times \Big(1 + \frac{c_{z-1} + c_z}{n - \sum_{w=0}^{z-1} n_w - c_z + 1}\Big).$$

Now we prove that the probability for $T_{n+1} \in I_1^{z-1}$ is

$$P(T_{n+1} \in I_1^{z-1}) = \frac{1}{n+1} \times \prod_{t=0}^{z-2}\Big(1 + \frac{c_t}{n - \sum_{w=0}^{t} n_w + 1}\Big). \tag{4.2}$$

When $z = 1$, under $\overline{C}_{I_z}(I_0)$, $P(T_{n+1} \in I_1^0) = 1/(n+1)$, as expressed in (4.2). Next, we use induction to complete the proof. Assume that, for $r \in \{2, \dots, z-2\}$, that

$$P(T_{n+1} \in I_1^r) = \frac{1}{n+1} \times \prod_{t=0}^{r-1} \Big(1 + \frac{c_t}{n - \sum_{w=0}^{t} n_w + 1}\Big),$$

then using Corollary 1 of Section 3.4,

$$\begin{aligned} P(T_{n+1} \in I_1^{r+1}) &= P(T_{n+1} \in I_1^r) \times \Big(1 + \frac{c_r}{n - \sum_{w=0}^{r} n_w + 1}\Big) \\ &= \frac{1}{n+1} \times \prod_{t=0}^{r-1} \Big(1 + \frac{c_t}{n - \sum_{w=0}^{t} n_w + 1}\Big) \\ &\quad \times \Big(1 + \frac{c_r}{n - \sum_{w=0}^{r} n_w + 1}\Big) \\ &= \frac{1}{n+1} \times \prod_{t=0}^{r} \Big(1 + \frac{c_t}{n - \sum_{w=0}^{t} n_w + 1}\Big), \end{aligned}$$

which completes the proof for the probability of $P(T_{n+1} \in I_1^{z-1})$. Under $\overline{C}_{I_z}(I_z)$, $\overline{P}(T_{n+1} \in I_0^z)$ and $\overline{P}(T_{n+1} \in I_{e_z}^z)$ are equal to $P(T \in I_1^z)$. Then, from (4.2), the upper probabilities for $T_{n+1} \in I_z$ follow as stated in the theorem. □

As for the lower probabilities above, these optimal configurations cannot occur simultaneously if there are censored data in the relevant intervals. This confirms sub-additivity of such upper probabilities [57], i.e.

$$\overline{P}(T_{n+1} \in I_z \cup I_t) \leq \overline{P}(T_{n+1} \in I_z) + \overline{P}(T_{n+1} \in I_t) \quad \text{for } z \neq t,$$

with strict inequality if there are censored observations in intervals which affect the relevant M-function values according to Theorem 6.

Example 11

The grouped lifetime data set in this example is given by Coolen [14] as an illustrative example. Now it is used to illustrate the lower and upper probabilities for T_{n+1} in intervals as presented in this section. Table 4.1 gives the data and the related lower and upper probabilities.

The example illustrates that, if there are more censored data in an interval I_z, the difference between lower and upper probabilities increases.

Interval	e_z	c_z	$\underline{P}(T_{59} \in I_z)$	$\overline{P}(T_{59} \in I_z)$
I_0	6	0	0.1017	0.1186
I_1	20	1	0.3220	0.3628
I_2	17	2	0.2764	0.3356
I_3	12	0	0.2211	0.2622

Table 4.1: The lower and upper probabilities for $T_{59} \in I_z$ (Example 11).

4.4 Predictive survival functions

In this section, we derive lower and upper probabilities for the events that T_{n+1} is greater than t, for $t \in I_z$, which define lower and upper survival functions for T_{n+1} for grouped data, denoted by $\underline{S}_{T_{n+1}}(t) = \underline{P}(T_{n+1} > t)$ and $\overline{S}_{T_{n+1}}(t) = \overline{P}(T_{n+1} > t)$, respectively. The super-additivity and sub-additivity of lower and upper probabilities, respectively, mentioned in Section 4.3, prevent us from using the results from Section 4.3 directly. Instead, we must again find the optimal configurations, which luckily turn out to be rather straightforward in this case. We present these configurations together with the upper and lower survival functions based on these configurations, in Theorems 8 and 9. At the same time, some further, but straightforward, notation for the optimal configurations is introduced in these theorems.

Theorem 8 ($\overline{S}_{T_{n+1}}(t)$)
Consider NPI, based on rc-$A_{(n)}$, for grouped data. Using the notation introduced in Section 4.2, the upper survival function $\overline{S}_{T_{n+1}}(t) = \overline{P}(T_{n+1} > t)$, at $t \in I_z$, is obtained for the configurations

$$\overline{C}^S_{I_z}(I_v) = \{t^v_{e_v} < c^v_1\}, \quad \text{for } v < z.$$

For I_z, the optimal configuration is:

$$\overline{C}^S_{I_z}(I_z) = \{t < \min(t^z_1, c^z_1)\}.$$

The configurations in intervals beyond I_z do not influence $\overline{S}_{T_{n+1}}(t)$, for $t \in I_z$, so we do not need to specify $\overline{C}^S_{I_z}(I_v)$ for $v > z$.

The corresponding upper predictive survival function can be derived as follows. Let $\overline{S}_z$ denote the value of the upper survival function for $t \in I_z = [a_z, a_{z+1})$, with $z = 0, 1, \ldots, k$, then,

$$\overline{S}_0 = 1,$$

and, for $z \geq 1$,

$$\overline{S}_z = \overline{S}_{z-1} - e_{z-1} p^{z-1},$$

with

$$p^0 = \frac{1}{n+1},$$

and, for $z \geq 2$,

$$p^{z-1} = p^{z-2} \times \frac{n - \sum_{w=0}^{z-3} n_w - e_{z-2} + 1}{n - \sum_{w=0}^{z-2} n_w + 1}.$$

Proof

The proof of the optimal configurations follows again from OCP, but the optimal configuration of the interval I_z takes a bit more consideration now, as we are actually interested in the upper survival function at time $t \in I_z$. To get maximum probability mass to the right of t, all observations in I_z are assumed to be greater than t, in which case the actual ordering does not influence the upper survival function value at t.

Next, we determine the value of the upper survival function for $t \in I_z = [a_z, a_{z+1})$. First, consider $t \in I_0$. According to the optimal configuration, it is known that all probability masses defined via M-function values on intervals which are (partially) in I_0, can be assumed to be in (t, ∞), so all probability masses for T_{n+1} can be in (t, ∞), which implies that $\overline{S}_{T_{n+1}}(t)$ is equal to one for $t \in I_0$, so $\overline{S}_0 = 1$.

Now we consider the upper survival function for $\overline{S}_{T_{n+1}}(t)$ for $t \in I_z$, $z = 1, \ldots, k$. All probability masses defined via M-function values on intervals which have non-empty intersection with I_z can be assumed to be in (t, ∞), so this upper survival function is the sum of all probability masses which could be in $[a_z, \infty)$. For $t \in I_z$, we have

$$\overline{S}_z = \overline{S}_{z-1} - \left[\overline{P}(T_{n+1} \in I_0^{z-1}) + P(T_{n+1} \in I_1^{z-1}) + \ldots + P(T_{n+1} \in I_{e_{z-1}-1}^{z-1})\right],$$

as $\overline{S}_z$ is equal to $\overline{S}_{z-1}$ minus the probability masses that necessarily must be in I_{z-1}. Clearly, $\overline{P}(T_{n+1} \in I_0^{z-1})$ is equal to $P(T_{n+1} \in (t_{e(z-2)}^{z-2},\ t_1^{z-1}))$ for $z = 2, \ldots, k$, and $\overline{P}(T_{n+1} \in I_0^0) = P(T_{n+1} \in I_0^0)$. For the configuration $\overline{C}_{I_z}^S(I_{z-1})$, $\overline{P}(T_{n+1} \in I_0^{z-1})$ and $P(T_{n+1} \in I_i^{z-1})$, for $i = 1, \ldots, e_{z-1} - 1$ are equal. So

$$\overline{S}_z = \overline{S}_{z-1} - e_{z-1} P(T_{n+1} \in I_1^{z-1}).$$

Using Corollary 1 of Section 3.4,

$$P(T_{n+1} \in I_1^{z-1}) = P(T_{n+1} \in I_1^{z-2}) \times \frac{n - \sum_{w=0}^{z-3} n_w - e_{z-2} + 1}{n - \sum_{w=0}^{z-2} n_w + 1}.$$

To simplify the notation, denote $P(T_{n+1} \in I_1^h)$ as p^h for $h = 0, 1, \ldots, z-1$, then

$$\overline{S}_z = \overline{S}_{z-1} - e_{z-1} p^{z-1}$$

with $\overline{S}_0 = 1$, and

$$p^{z-1} = p^{z-2} \times \frac{n - \sum_{w=0}^{z-3} n_w - e_{z-2} + 1}{n - \sum_{w=0}^{z-2} n_w + 1}, \text{ for } z \geq 2,$$

with $p^0 = 1/(n+1)$. $\square$

Obviously, we use the fact that no information is provided by grouped data except the number of event times and censoring times. Because the event and censoring times in $I_z = [a_z, a_{z+1})$ could be anywhere in this interval, without additional assumptions, this implies that

$$\overline{S}_{T_{n+1}}(t) = \overline{S}_{T_{n+1}}(a_z), \quad \text{for all } t \in I_z = [a_z, \ a_{z+1}).$$

Therefore, the upper survival function $\overline{S}_{T_{n+1}}(t)$ is a step-function, which only decreases at the a_z. The upper survival function is constant on I_z, so it is continuous from the right at points a_z.

Next we consider the lower survival function, $\underline{S}_{T_{n+1}}(t)$, which is also a step-function, but we will show that it is continuous from the left at a_z, so $\underline{S}_{T_{n+1}}(t)$ is constant on $(a_z, a_{z+1}]$, which is denoted as $\tilde{I}_z = (a_z, a_{z+1}]$. The optimal configuration and formula to calculate $\underline{S}_{T_{n+1}}(t)$ is given in Theorem 9.

Theorem 9 ($\underline{S}_{T_{n+1}}(t)$)
Consider NPI, based on rc-$A_{(n)}$, for grouped data. Using the notation introduced in Section 4.2, the lower survival function $\underline{S}_{T_{n+1}}(t) = \underline{P}(T_{n+1} > t)$, at $t \in \tilde{I}_z = (a_z, a_{z+1}]$, is obtained for the configuration

$$\underline{C}^S_{\tilde{I}_z}(I_v) = \{c^v_{c_v} < t^v_1\}, \quad \text{for } v < z.$$

The optimal configuration of the interval I_z takes a bit more consideration now, as we are actually interested in the lower survival function at time $t \in \tilde{I}_z$. To get minimum probability mass to the right of t, all observations in I_z are assumed to be less than t, and again ordered such that the censorings are assumed to happen before the event times:

$$\underline{C}^S_{\tilde{I}_z}(I_z) = \{c^z_{c_z} < t^z_1 \text{ and } t^z_{e_z} < t\}.$$

The configurations in intervals beyond I_z do not influence $\overline{S}_{T_{n+1}}(t)$, for $t \in \tilde{I}_z$, so we do not need to specify $\underline{C}^S_{\tilde{I}_z}(I_v)$ for $v > z$.

Based on the assumption rc-$A_{(n)}$ and such grouped data, the lower predictive survival function can be derived as follows. Let $\underline{S}_{\tilde{z}}$ denote the value of the lower survival function at $t \in \tilde{I}_z = (a_z, a_{z+1}]$, for $z = 0, 1, \ldots, k$, then,

$$\underline{S}_{\tilde{0}} = \frac{n - n_0}{n - c_0 + 1},$$

and, for $z \geq 1$,

$$\underline{S}_{\tilde{z}} = \underline{S}_{\tilde{z-1}} + q^{z-1} - (e_z + 1)q^z,$$

with

$$q^0 = \frac{1}{n - c_0 + 1},$$

and, for $z \geq 1$,

$$q^z = q^{z-1} \times \frac{n - \sum_{w=0}^{z-1} n_w + 1}{n - \sum_{w=0}^{z-1} n_w - c_z + 1}.$$

For completeness, it seems reasonable to define $\underline{S}_{\tilde{0}} = 1$, assuming that no events or censorings actually happened at time 0.

Proof

The proof of the optimal configuration follows again straightforwardly from OCP. Next, we determine the value of the lower survival function for $t \in \tilde{I}_z = (a_z, a_{z+1}]$. For convenience, we denote $\tilde{I}^z_i = (t^z_i, t^z_{i+1}]$ which are sub-intervals of $\tilde{I}_z$, for $i = 1, \ldots, e_z - 1$, and $\tilde{I}^z_0 = (a_z, t^z_1]$ and $\tilde{I}^z_{e_z} = (t^z_{e_z}, a_{z+1}]$.

First, we consider the lower survival function $\underline{S}_{T_{n+1}}(t)$ for $t \in \tilde{I}_0$. According to the optimal configuration, all probability masses defined via M-function values on intervals which have non-empty intersection with $\tilde{I}_0$ can be assumed to be in $(0, t)$, so the total probability mass for T_{n+1} that necessarily must be in (t, ∞), corresponding to $\underline{C}^S_{\tilde{I}_0}(I_0)$, is

$$\begin{aligned}
\underline{S}_{\tilde{0}} &= 1 - P(T_{n+1} \in I^0_0) \times (e_0 + 1) \\
&= 1 - \frac{e_0 + 1}{n - c_0 + 1} \\
&= \frac{n - n_0}{n - c_0 + 1}.
\end{aligned}$$

Now we consider $\underline{S}_{T_{n+1}}(t)$ for $t \in \tilde{I}_z$, $z = 1, \ldots, k$. This survival function is the sum of probability masses which necessarily must be in (t, ∞). According to the optimal configuration,

$$\begin{aligned}
\underline{S}_{\tilde{z}} = \; & \underline{S}_{\tilde{z-1}} - \underline{P}(T_{n+1} \in \tilde{I}^z_0) - P(T_{n+1} \in \tilde{I}^z_1) - \cdots - \\
& P(T_{n+1} \in \tilde{I}^z_{e_z-1}) - \overline{P}(T_{n+1} \in \tilde{I}^z_{e_z}),
\end{aligned} \tag{4.3}$$

as $\underline{S}_{\tilde{z}}$ is equal to $\underline{S}_{\widetilde{z-1}}$ minus the probability masses that is possibly in I_z. Equation (3.3) implies that

$$\underline{P}(T_{n+1} \in \tilde{I}_0^z) = P(T_{n+1} \in (t_{e_{z-1}}^{z-1}, t_1^z)) - M_{T_{n+1}}(t_{e_{z-1}}^{z-1}, t_1^z),$$

and

$$\overline{P}(T_{n+1} \in \tilde{I}_{e_z}^z) = M_{T_{n+1}}(t_{e_z}^z, t_1^{z+1}).$$

According to the optimal configuration and Corollary 2, we have

$$M_{T_{n+1}}(t_{e_{z-1}}^{z-1}, t_1^z) = P(T_{n+1} \in \tilde{I}_1^{z-1}) = \ldots = P(T_{n+1} \in \tilde{I}_{e_{z-1}-1}^{z-1})$$

$$M_{T_{n+1}}(t_{e_z}^z, t_1^{z+1}) = P(T_{n+1} \in \tilde{I}_1^z) = \ldots = P(T_{n+1} \in \tilde{I}_{e_z-1}^z).$$

So (4.3) can be expressed as

$$\underline{S}_{\tilde{z}} = \underline{S}_{\widetilde{z-1}} + P(T_{n+1} \in \tilde{I}_1^{z-1}) - (e_z + 1)P(T_{n+1} \in \tilde{I}_1^z).$$

By using the notation $q^h = P(T_{n+1} \in \tilde{I}_1^h)$, for $h = 1, \ldots, u$, we have

$$\underline{S}_{\tilde{z}} = \underline{S}_{\widetilde{z-1}} + q^{z-1} - (e_z + 1)q^z.$$

Using Corollary 1 of Section 3.4,

$$q^z = q^{z-1} \times (1 + \frac{c_z}{n - \sum_{w=0}^{z-1} n_w - c_z + 1}),$$

and for q^0, using Theorem 2 of Section 3.4, we have

$$q^0 = P(T_{n+1} \in \tilde{I}_1^0) = \frac{1}{n - c_0 + 1}.$$

This completes the proof. □

Example 11 (continued)

The grouped lifetime data given in Example 11 are now used to illustrate the lower and upper survival function for T_{n+1}. Table 4.2 gives the corresponding values of $\underline{S}_{T_{59}}(t)$ and $\overline{S}_{T_{59}}(t)$, for $t \in \tilde{I}_z$ and $t \in I_z$ respectively, for $z = 0, 1, \ldots, 3$.

These lower and upper survival functions illustrate some of the issues by using NPI to grouped data. For example, although there are less event times in the interval I_2, the difference of upper and lower survival function values in this I_2 is still larger. This is because M-functions for intervals within I_2 depend on censoring in I_1.

Interval	$\underline{S}_{T_{59}}(t)$	Interval	$\overline{S}_{T_{59}}(t)$
$\tilde{I}_0$	0.8814	I_0	1
$\tilde{I}_1$	0.5355	I_1	0.8984
$\tilde{I}_2$	0.2211	I_2	0.5593
$\tilde{I}_3$	0	I_3	0.2622

Table 4.2: Lower and upper survival function for grouped data (Example 11).

4.5 Comparison with alternative nonparametric methods

In this section, we compare the results presented in Sections 4.3 and 4.4 with alternative nonparametric methods for grouped data, which were introduced in Section 2.5.

First, we compare the lower and upper survival functions with the so-called standard life table estimator, which we denote by $\hat{S}(t)$. The standard life table estimator provides an estimate of the underlying population lifetime distribution function, so its aim is not directly prediction as it is in our method. It deals with censorings per interval by effectively assuming that censorings took place at the middle of the interval, enabling a precise estimate of the survival function corresponding to the underlying lifetime distribution at the time points a_i which create the intervals. This estimator does not consider other time points. We do not discuss further variations, which can often be interpreted as following from slightly different assumptions for the exact censoring times within an interval [48].

Secondly, we compare our inferences with the imprecise Dirichlet model (Coolen [14]), which gives bounds for estimated values of parameters θ_z in a Bayesian multinomial model, where the categories are the same intervals as used in this chapter, and the parameters have the standard interpretation in terms of proportion of a large population to fall in each of these intervals, so again the inferential goal differs from our method, although the corresponding Bayesian predictive probability would suggest the probability for a future observation to fall in a particular interval to be equal to the expected value of the corresponding parameter, which allows comparison with our method. However, censorings are dealt with differently in the imprecise Dirichlet model, as they are only assumed to take place at times a_z. In the example below, the c_z censorings within interval I_z are, for the imprecise Dirichlet model, assumed to take place at the right-end point a_{z+1} of this interval, so after the e_z event times in this interval, which explains why the corresponding bounds for the expected parameter values, $\underline{E}(\theta_z)$ and $\overline{E}(\theta_z)$, are relatively close to our upper probability $\overline{P}(T_{n+1} \in I_z)$, which is based on similar configurations for the intervals left of I_z, as shown in Section 4.3.

For completeness, we should remark that the imprecise Dirichlet model [14] requires specification of one parameter (denoted by s in [14, 58]), which effectively controls the size of the set of corresponding Bayesian prior distributions. For larger values the bounds for the expected values can fall outside our corresponding lower and upper probabilities as based on rc-$A_{(n)}$.

It should be remarked that these methods all give numerical values which are quite similar. The underlying reason is that all can be explained, with some small variations according to goal of inference and the assumptions added for dealing with the censorings, via Efron's [32] redistribution of probability mass process, which in addition also underlies well-known nonparametric methods for inference based on right-censored data, such as the product-limit estimator by Kaplan and Meier [46]. Next, the comparison is illustrated using data from the literature.

Example 12

The data set used in this example is given by Berkson and Gage [5], to describe the survival experience of a group of patients who underwent operations in connection with a type of malignant disease. The data were also used by Lawless [48] to illustrate the standard life table estimator. Effectively, the time-axis (in years) is partitioned into 11 intervals I_z, with a total number of 374 observations, consisting of e_z event times and c_z censoring times per interval, as given in Table 4.3.

Table 4.4 presents the lower and upper predictive survival functions, as derived in Section 4.4, together with the corresponding survival function estimates according to the standard life table method [48], at points a_z, denoted by $\hat{S}(a_z)$. We can find that the value of this standard life table estimator at a_z is always between our lower and upper survival functions for T_{n+1} at a_z.

The standard life table method does not define the survival function estimate at other time points, the values of the lower and upper survival functions based on rc-$A_{(n)}$ at other t are given in Table 4.5. As discussed, $\underline{S}_{T_{n+1}}(t) = \underline{S}_{T_{n+1}}(a_{z+1})$ for $t \in \tilde{I}_z = (a_z, a_{z+1}]$, $\overline{S}_{T_{n+1}}(t) = \overline{S}_{T_{n+1}}(a_z)$ for $t \in I_z = [a_z, a_{z+1})$. In the final interval, beyond $a_{10} = 10$, the lower survival function is equal to 0, whereas the upper survival function remains equal to 0.236 without any further assumptions added.

I_z	e_z	c_z
$[0,1)$	90	0
$[1,2)$	76	0
$[2,3)$	51	0
$[3,4)$	25	12
$[4,5)$	20	5
$[5,6)$	7	9
$[6,7)$	4	9
$[7,8)$	1	3
$[8,9)$	3	5
$[9,10)$	2	5
$[10,\infty)$	47	0

Table 4.3: Event and censoring times per interval.

a_z	$\underline{S}_{T_{375}}(a_z)$	$\overline{S}_{T_{375}}(a_z)$	$\hat{S}(a_z)$
0	1	1	1
1	0.757	0.760	0.759
2	0.555	0.557	0.556
3	0.419	0.421	0.420
4	0.346	0.355	0.350
5	0.286	0.296	0.291
6	0.262	0.275	0.268
7	0.247	0.261	0.254
8	0.243	0.257	0.250
9	0.230	0.245	0.237
10	0.220	0.236	0.228

Table 4.4: Lower and upper predictive survival functions, and standard life table estimator.

(a_z, a_{z+1})	$\underline{S}_{T_{375}}(t)$	$\overline{S}_{T_{375}}(t)$
$(0,1)$	0.757	1
$(1,2)$	0.555	0.760
$(2,3)$	0.419	0.557
$(3,4)$	0.346	0.421
$(4,5)$	0.286	0.355
$(5,6)$	0.262	0.296
$(6,7)$	0.247	0.275
$(7,8)$	0.243	0.261
$(8,9)$	0.230	0.257
$(9,10)$	0.220	0.245
$(10,\infty)$	0	0.236

Table 4.5: Lower and upper predictive survival function for (a_z, a_{z+1}).

Table 4.6 gives the lower and upper probabilities for $T_{n+1} \in I_z$, together with the lower and upper expected values for the parameters θ_z ($s = 1, 2$) in the imprecise Dirichlet model [14].

	rc-$A_{(n)}$		imprecise Dirichlet model			
			s=1		s=2	
I_z	$\underline{P}(T_{375} \in I_z)$	$\overline{P}(T_{375} \in I_z)$	$\underline{E}(\theta_z)$	$\overline{E}(\theta_z)$	$\underline{E}(\theta_z)$	$\overline{E}(\theta_z)$
$[0, 1)$	0.2400	0.2427	0.2400	0.2427	0.2394	0.2447
$[1, 2)$	0.2000	0.2053	0.2027	0.2053	0.2021	0.2027
$[2, 3)$	0.1333	0.1387	0.1360	0.1387	0.1356	0.1410
$[3, 4)$	0.0640	0.0750	0.0667	0.0693	0.0665	0.0718
$[4, 5)$	0.0548	0.0642	0.0586	0.0616	0.0584	0.0643
$[5, 6)$	0.0181	0.0272	0.0216	0.0247	0.0215	0.0276
$[6, 7)$	0.0100	0.0193	0.0137	0.0172	0.0137	0.0205
$[7, 8)$	0	0.0081	0.0039	0.0078	0.0039	0.0116
$[8, 9)$	0.0078	0.0177	0.0122	0.0163	0.0121	0.0202
$[9, 10)$	0.0043	0.0147	0.0089	0.0133	0.0088	0.0176
$[10, \infty)$	0.2200	0.2357	0.2313	0.2357	0.2292	0.2380

Table 4.6: Lower and upper predictive probabilities for $T_{n+1} \in I_z$, and lower and upper expected values for θ_z in imprecise Dirichlet model.

Table 4.6 illustrate some of issues addressed in discussion above. For example, the corresponding bounds for the expected parameter values, $\underline{E}(\theta_z)$ and $\overline{E}(\theta_z)$, are relatively close to the upper probability $\overline{P}(T_{n+1} \in I_z)$. In this example, we see that, for $s = 1$, $\overline{P}(T_{n+1} \in I_z) \geq \overline{E}(\theta_z)$ holds for all z, but this inequality does not hold generally for $s = 2$. It is difficult to prove this inequality for $s = 1$, because the calculation vectors of these two methods are really not the same when applied. It should be noticed that, for $s = 1$, $\overline{P}(T_{n+1} \in I_z)$ and $\overline{E}(\theta_z)$ are equal when there are no censorings in I_t, for $t \leq z$. There is only one event time in the interval $[7, 8)$, which leads to $\underline{P}(T_{375} \in [7, 8))$ is equal to zero. These upper and lower predictive probabilities illustrate the issues of using NPI to grouped data, i.e., the difference of upper and lower probabilities in I_z is larger following more censorings happening before I_z.

4.6 Concluding remarks

The inferences for T_{n+1} based on grouped data including right-censored observations presented in this chapter, are directly in terms of lower and upper probabilities for $T_{n+1} \in I_z$, and the lower and upper survival functions of T_{n+1}. Related methods such as those given in Section 2.5 require the use of additional assumptions about the event times and censoring times. In our method, the optimal configurations for lower and upper probabilities are not the same for different events, leading, for example, again to sub-additivity of lower probabilities.

It is important to consider how these lower and upper probabilities can be used and interpreted. Although they could be considered from a Bayesian perspective, the most logical interpretation has a more classical frequentist nature, in the sense that, if the method were applied very often, the lower and upper probabilities would be bounds on the frequencies with which the relevant events would occur. It is clear that, if there are only few data, these bounds may become wide, hence may not lead to strong enough inferences, which would indicate that either more data or additional assumptions are necessary. Finally, we should remark that, if the exact event and censoring times were actually known, then the corresponding lower and upper survival functions (as presented in Chapter 3) would of course be within the bounds for grouped data as presented in Section 4.4.

Chapter 5

Nonparametric predictive comparison of two groups of lifetime data

5.1 Introduction

This chapter applies right-censoring $A_{(n)}$, and corresponding nonparametric predictive inference (NPI), as presented in Chapter 3, to the problem of comparing two groups of data, or, if one prefers to use such terminology, two underlying populations, where the data include right-censored observations. The comparison is based on the lower and upper probabilities for the event that the next observation from one population is greater than the next observation from the other population. This generalizes the results presented by Coolen [13], who did not allow censoring. Throughout this chapter, we again assume that censoring is non-informative, as discussed in Section 3.2. We also assume that the two populations compared are independent, in the sense that any information about the random quantities from one population does not influence our inferences on random quantities from the other population.

Section 5.2 presents the main result on predictive comparison of two groups of lifetime data. In Section 5.3, the method is illustrated, and briefly compared with an alternative nonparametric method, as discussed in Section 2.6, via two examples. In Section 5.4, we briefly discuss how the method can be adapted for dealing with tied observations, and we add some concluding remarks about the method.

5.2 Predictive comparison of two groups of lifetime data

For comparison of two groups of lifetime data, we use the notation as introduced in Chapter 3, but consistently add an index, a or b, corresponding to the groups, which are called A and B. For example, for group A we have n_a observations, consisting of the event times $0 < t_{a,1} < \ldots < t_{a,u_a}$ and right-censoring times $0 < c_{a,1} < \ldots < c_{a,v_a}$, and the right-censoring times in the interval $(t_{a,i}, t_{a,i+1})$ are denoted by $c^i_{a,1} < \ldots < c^i_{a,l_{a,i}}$. We refer to these as A-event times and A-censoring times. Let T_{a,n_a+1} denote a future observation from population A, etc. Throughout we assume that there are no ties at all among the observations. Dealing with ties is straightforward, but would make notation more awkward, brief comments on this are given in the final section. We assume that information on one group does not have any effect on probabilities of random quantities corresponding to the other group, so that T_{a,n_a+1} and T_{b,n_b+1} are independent and that data from group A (B) does not influence the probabilities for T_{b,n_b+1} (T_{a,n_a+1}). This is informally summarized by stating that the groups are independent.

Some additional notation is required, effectively counting the number of observed B-event times to the left of observations from A:

$$\begin{aligned} s_b(t_{a,i}) &= \#\{t_{b,j} \,|\, t_{b,j} < t_{a,i},\ j = 1, \ldots, u_b\}, \quad \text{for } \ i = 1, \ldots, u_a, \\ s_b(c^i_{a,k}) &= \#\{t_{b,j} \,|\, t_{b,j} < c^i_{a,k},\ j = 1, \ldots, u_b\}, \quad \text{for } \ i = 1, \ldots, u_a \ \text{ and} \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad k = 1, \ldots, l_{a,i}, \end{aligned}$$

and, similarly, the number of right-censoring times from group B in the interval $(t_{b,s_b(t_{a,i})}, t_{a,i})$ is denoted by,

$$s^c_b(t_{a,i}) = \#\{c_{b,j} \,|\, c_{b,j} \in (t_{b,s_b(t_{a,i})}, t_{a,i}),\ j = 1, \ldots, u_b\}, \quad \text{for } \ i = 1, \ldots, u_a + 1,$$

where, $t_{b,s_b(t_{a,i})}$ is the largest event time from group B smaller than $t_{a,i}$.

The main results of this chapter, namely lower and upper probabilities for events $T_{a,n_a+1} > T_{b,n_b+1}$, based on the assumptions rc-$A_{(n_a)}$ and rc-$A_{(n_b)}$, are presented as two theorems. The proofs of these theorems are simplified via a lemma, which is presented first, and which justifies the use of a variation of the theorem of total probability with conditioning on nested intervals, with probability distributions partially specified via M-function values.

Lemma 3

For $s \geq 2$, let $J_l = (j_l, r)$, with $j_1 < j_2 < \ldots < j_s < r$, so we have nested intervals

$J_1 \supset J_2 \supset \ldots \supset J_s$ with the same right end-point r (which may be infinity). We consider two independent real-valued random quantities, say X and Y. Let the probability distribution for X be partially specified via M-function values, with all probability mass $P(X \in J_1)$ described by the s M-function values $M_X(J_l)$, so $\sum_{l=1}^{s} M_X(J_l) = P(X \in J_1)$. Then, without additional assumptions, we have

$$\sum_{l=1}^{s} P(Y < j_l) M_X(J_l) \leq P(Y < X, X \in J_1) \leq P(Y < r) P(X \in J_1),$$

and these bounds are optimal, so they are the maximum lower and minimal upper bounds that generally hold.

Proof

For any number s of nested intervals, the proof follows the same principle, so for ease of notation we present it for $s = 3$. We use the theorem of total probability to condition further on the partition $\{J_3, J_2 \setminus J_3, J_1 \setminus J_2\}$ of J_1 for the random quantity X. The probability distribution of X on J_1 is partially specified via M-function values for X defined on J_1, J_2, J_3. Let $M_X^l(J)$ denote the (unknown) part of the M-function value $M_X(J_l)$ that is actually in $J \subset J_l$, so we have

$$\begin{aligned}
P(X \in J_3) &= M_X^3(J_3) + M_X^2(J_3) + M_X^1(J_3), \\
P(X \in J_2 \setminus J_3) &= M_X^2(J_2 \setminus J_3) + M_X^1(J_2 \setminus J_3), \\
P(X \in J_1 \setminus J_2) &= M_X^1(J_1 \setminus J_2), \\
M_X(J_1) &= M_X^1(J_1 \setminus J_2) + M_X^1(J_2 \setminus J_3) + M_X^1(J_3), \\
M_X(J_2) &= M_X^2(J_2 \setminus J_3) + M_X^2(J_3), \\
M_X(J_3) &= M_X^3(J_3).
\end{aligned}$$

These M-function values are not further specified, but we can now use the theorem of total probability, and then derive bounds by solving the constrained optimization problems. The lower bound follows from (with $J_4 = \emptyset$ for ease of notation)

$$\begin{aligned}
P(Y < X, X \in J_1) &= \sum_{l=1}^{3} P(Y < X, X \in J_l \setminus J_{l+1}) \\
&= \sum_{l=1}^{3} P(Y < X \mid X \in J_l \setminus J_{l+1}) P(X \in J_l \setminus J_{l+1}) \\
&= P(Y < X \mid X \in J_1 \setminus J_2) M_X^1(J_1 \setminus J_2) + \\
&\quad P(Y < X \mid X \in J_2 \setminus J_3)[M_X^2(J_2 \setminus J_3) + M_X^1(J_2 \setminus J_3)] + \\
&\quad P(Y < X \mid X \in J_3)[M_X^3(J_3) + M_X^2(J_3) + M_X^1(J_3)].
\end{aligned}$$

With the constraints on these M-function values as given above, the lower bound is achieved by effectively putting the probability masses for X at the infimum of each interval on which they are defined by the M-function values, so setting

$$M_X^1(J_2 \setminus J_3) = M_X^1(J_3) = M_X^2(J_3) = 0,$$

and taking the lower bounds for the conditional probabilities for $Y < X$, given $X \in I$, for the relevant I above, by replacing $X \in I$ by $X = \inf(I)$, leading to the terms $Y < j_l$ in the lower bound. The upper bound can be derived simultaneously, but is rather trivial as these nested intervals have the same right end-point. The fact that these bounds are optimal, without additional assumptions, follows easily from this construction. $\square$

Bounds for the probability of $T_{a,n_a+1} > T_{b,n_b+1}$, based on the assumptions rc-$A_{(n_a)}$ and rc-$A_{(n_b)}$, are presented in the following theorems. As these bounds are optimal, without any additional assumptions, they are lower and upper probabilities, which we denote by $\underline{P}(T_{a,n_a+1} > T_{b,n_b+1})$ and $\overline{P}(T_{a,n_a+1} > T_{b,n_b+1})$, respectively.

Theorem 10 $\left((\underline{P}(T_{a,n_a+1} > T_{b,n_b+1})\right)$
Assume that data are available from two independent groups, A and B, using the notation above. Based on the assumptions rc-$A_{(n_a)}$ and rc-$A_{(n_b)}$, the lower probability for the event $T_{a,n_a+1} > T_{n_b+1}$ can be derived as

$$\begin{aligned}
&\underline{P}(T_{a,n_a+1} > T_{b,n_b+1}) \\
= &\sum_{i=0}^{u_a} \left\{ \left[\sum_{j=0}^{s_b(t_{a,i})-1} P(T_{b,n_b+1} \in (t_{b,j}, t_{b,j+1})) \right] M_{T_{a,n_a+1}}(t_{a,i}, t_{a,i+1}) \right. \\
&\left. + \sum_{k=1}^{l_{a,i}} \left(\left[\sum_{j=0}^{s_b(c^i_{a,k})-1} P(T_{b,n_b+1} \in (t_{b,j}, t_{b,j+1})) \right] M_{T_{a,n_a+1}}(c^i_{a,k}, t_{a,i+1}) \right) \right\}.
\end{aligned}$$

Proof
The assumption rc-$A_{(n_a)}$ leads straightforwardly to precise probabilities for events $T_{a,n_a+1} \in I_{a,i}$, for $i = 0, 1, \ldots, u_a$. We have

$$P(T_{a,n_a+1} > T_{b,n_b+1}) = \sum_{i=0}^{u_a} P(T_{b,n_b+1} < T_{a,n_a+1}, T_{a,n_a+1} \in (t_{a,i}, t_{a,i+1})), \qquad (5.1)$$

and applying Lemma 3 for each of the terms within this sum, we get

$$
\begin{aligned}
&P(T_{a,n_a+1} > T_{b,n_b+1}) \\
\geq\ &\sum_{i=0}^{u_a} \Big[P(T_{b,n_b+1} < t_{a,i}) M_{T_{a,n_a+1}}(t_{a,i}, t_{a,i+1}) \\
&\quad + \sum_{k=1}^{l_{a,i}} P(T_{b,n_b+1} < c^i_{a,k}) M_{T_{a,n_a+1}}(c^i_{a,k}, t_{a,i+1}) \Big].
\end{aligned} \tag{5.2}
$$

According to the definition of $s_b(t_{a,i})$ and $s_b(c^i_{a,k})$, $t_{b,s_b(t_{a,i})}$ is the largest B-event time smaller than A-event time $t_{a,i}$, and $t_{b,s_b(c^i_{a,k})}$ is the largest B-event time smaller than A-censoring time $c^i_{a,k}$, so (5.2) is equivalent to

$$
\begin{aligned}
&P(T_{a,n_a+1} > T_{b,n_b+1}) \\
\geq\ &\sum_{i=0}^{u_a} \Bigg\{ \Bigg[\sum_{j=0}^{s_b(t_{a,i})-1} P(T_{b,n_b+1} \in (t_{b,j}, t_{b,j+1})) + P(T_{b,n_b+1} \in (t_{b,s_b(t_{a,i})}, t_{a,i})) \Bigg] \times \\
&M_{T_{a,n_a+1}}(t_{a,i}, t_{a,i+1}) + \\
&\sum_{k=1}^{l_{a,i}} \Bigg(\Bigg[\sum_{j=0}^{s_b(c^i_{a,k})-1} P(T_{b,n_b+1} \in (t_{b,j}, t_{b,j+1})) + P(T_{b,n_b+1} \in (t_{b,s_b(c^i_{a,k})}, c^i_{a,k})) \Bigg] \times \\
&M_{T_{a,n_a+1}}(c^i_{a,k}, t_{a,i+1}) \Bigg) \Bigg\}.
\end{aligned} \tag{5.3}
$$

In (5.3), $P(T_{b,n_b+1} \in (t_{b,s_b(t_{a,i})}, t_{a,i}))$ and $P(T_{b,n_b+1} \in (t_{b,s_b(c^i_{a,k})}, c^i_{a,k}))$ do not have precise values, because $t_{a,i}$ and $c^i_{a,k}$ are not B-event times, and the interval $(t_{b,s_b(t_{a,i})}, t_{a,i})$ is just a sub-interval of $(t_{b,s_b(t_{a,i})}, t_{b,s_b(t_{a,i})+1})$, and $(t_{b,s_b(c^i_{a,k})}, c^i_{a,k})$ is a sub-interval of $(t_{b,s_b(c^i_{a,k})}, t_{b,s_b(c^i_{a,k})+1})$, which are formed by two consecutive B-event times. According to the assumption rc-$A_{(n_b)}$, the maximum lower bounds, without additional assumptions, for $P(T_{b,n_b+1} \in (t_{b,s_b(t_{a,i})}, t_{a,i}))$ and for $P(T_{b,n_b+1} \in (t_{b,s_b(c^i_{a,k})}, c^i_{a,k}))$ are zero. Therefore

$$
\begin{aligned}
&P(T_{a,n_a+1} > T_{n_b+1}) \\
\geq\ &\sum_{i=0}^{u_a} \Bigg\{ \Bigg[\sum_{j=0}^{s_b(t_{a,i})-1} P(T_{b,n_b+1} \in (t_{b,j}, t_{b,j+1})) \Bigg] M_{T_{a,n_a+1}}(t_{a,i}, t_{a,i+1}) \\
&\quad + \sum_{k=1}^{l_{a,i}} \Bigg(\Bigg[\sum_{j=0}^{s_b(c^i_{a,k})-1} P(T_{b,n_b+1} \in (t_{b,j}, t_{b,j+1})) \Bigg] M_{T_{a,n_a+1}}(c^i_{a,k}, t_{a,i+1}) \Bigg) \Bigg\}.
\end{aligned} \tag{5.4}
$$

We cannot improve the lower bound in (5.4) without further assumptions. Hence, (5.4) provides the maximum lower bound for $P(T_{a,n_a+1} > T_{n_b+1})$ based on the assumptions rc-$A_{(n_a)}$ and rc-$A_{(n_b)}$. $\square$

Theorem 11 $\left(\overline{P}(T_{a,n_a+1} > T_{b,n_b+1})\right)$
For the same setting as in Theorem 10, the upper probability for $T_{a,n_a+1} > T_{n_b+1}$ can be derived as

$$
\begin{aligned}
&\overline{P}(T_{a,n_a+1} > T_{b,n_b+1}) \\
= &\sum_{i=0}^{u_a} \Bigg\{ \Bigg[\sum_{j=0}^{s_b(t_{a,i+1})-1} P(T_{b,n_b+1} \in (t_{b,j}, t_{b,j+1})) \\
&\quad + P(T_{b,n_b+1} \in (t_{b,s_b(t_{a,i+1})-1}, t_{b,s_b(t_{a,i+1})})) \\
&\quad + \sum_{l=1}^{s_b^c(t_{a,i+1})} M_{t_{b,n_b+1}}(c_{b,l}^{s_b^c(t_{a,i+1})}, t_{b,s_b(t_{a,i+1})+1}) \Bigg] P(T_{a,n_a+1} \in (t_{a,i}, t_{a,i+1})) \Bigg\}.
\end{aligned}
$$

Proof
Using the upper bound in Lemma 3 for each of the terms in the sum in (5.1), we get

$$
\begin{aligned}
&P(T_{a,n_a+1} > T_{b,n_b+1}) \\
\leq &\sum_{i=0}^{u_a} \Big[P(T_{b,n_b+1} < t_{a,i+1}) P(T_{a,n_a+1} \in (t_{a,i}, t_{a,i+1})) \Big]. \qquad (5.5)
\end{aligned}
$$

According to the definition of $s_b(t_{a,i})$, $t_{b,s_b(t_{a,i+1})}$ is the largest B-event time smaller than A-event time t_{a,i_a+1}, so from (5.5),

$$
\begin{aligned}
&P(T_{a,n_a+1} > T_{b,n_b+1}) \\
\leq &\sum_{i=0}^{u_a} \Bigg[\sum_{j=0}^{s_b(t_{a,i+1})-1} P(T_{b,n_b+1} \in (t_{b,j}, t_{b,j+1})) \\
&\quad + P(T_{b,n_b+1} \in (t_{b,s_b(t_{a,i+1})}, t_{a,i+1})) \Bigg] P(T_{a,n_a+1} \in (t_{a,i}, t_{a,i+1})) \qquad (5.6)
\end{aligned}
$$

In (5.6), only $P(T_{b,n_b+1} \in (t_{b,s_b(t_{a,i+1})}, t_{a,i+1}))$, for $i = 0, 1, \ldots, u_a$, is not determined precisely, because $t_{a,i+1}$ is not a B-event time, and the interval $(t_{b,s_b(t_{a,i+1})}, t_{a,i+1})$ is just a subinterval of $(t_{b,s_b(t_{a,i+1})}, t_{b,s_b(t_{a,i+1})+1})$ which is formed by two consecutive B-event times. According to rc-$A_{(n_b)}$, the upper probability for $P(T_{b,n_b+1} \in (t_{b,s_b(t_{a,i+1})}, t_{a,i+1}))$ can be derived as presented in Subsection 3.4.2. According to the definition of $s_b^c(t_{a,i+1})$, and equation (3.4), we have

$$
\begin{aligned}
&\overline{P}(T_{b,n_b+1} \in (t_{b,s_b(t_{a,i+1})}, t_{a,i+1})) \\
= &M_{T_{b,n_b+1}}(t_{b,s_b(t_{a,i+1})}, t_{b,s_b(t_{a,i+1})+1}) + \sum_{l=1}^{s_b^c(t_{a,i+1})} M_{T_{b,n_b+1}}(c_{b,l}^{s_b(t_{a,i+1})}, t_{b,s_b(t_{a,i+1})+1}).
\end{aligned}
$$

Hence, the minimum upper bound for $P(T_{a,n_a+1} > T_{b,n_b+1})$ can be derived from (5.6), together with equation (3.2) leading to the upper probability for $T_{a,n_a+1} > T_{b,n_b+1}$

$$\begin{aligned}
&\overline{P}(T_{a,n_a+1} > T_{b,n_b+1}) \\
= &\sum_{i=0}^{u_a} \left\{ \left[\sum_{j=0}^{s_b(t_{a,i+1})-1} P(T_{b,n_b+1} \in (t_{b,j}, t_{b,j+1})) \right.\right. \\
&\quad + P(T_{b,n_b+1} \in (t_{b,s_b(t_{a,i+1})-1}, t_{b,s_b(t_{a,i+1})})) \\
&\quad \left.\left. + \sum_{l=1}^{s_b^c(t_{a,i+1})} M_{t_{b,n_b+1}}(c_{b,l}^{s_b^c(t_{a,i+1})}, t_{b,s_b(t_{a,i+1})+1}) \right] P(T_{a,n_a+1} \in (t_{a,i}, t_{a,i+1})) \right\}.
\end{aligned}$$

$\square$

Theorems 10 and 11 present lower and upper probabilities for the event $T_{a,n_a+1} > T_{b,n_b+1}$. Although these imprecise probabilities are not available in a nice closed form, calculation is relatively easy as the individual terms are all product forms following from the definition of rc-$A_{(n)}$. If the data do not include any right-censorings, these lower and upper probabilities are identical to those presented by Coolen [13]. Although these formula become complex, the underlying idea for these optimal bounds is straightforward. The lower probability for $T_{a,n_a+1} > T_{b,n_b+1}$, based on the rc-$A_{(n)}$ assumptions per group, puts the probability masses, as specified by the M-function values for T_{a,n_a+1}, at the infimums of the intervals on which corresponding M-function values are specified, and for T_{b,n_b+1} at the supremums of the intervals, so at this bound the probability masses are effectively least supportive for this event given the partial specifications via M-function values. Of course, the upper probability just relates to these probability masses being put at the other end-points per interval.

We have presented the lower and upper probabilities for $T_{a,n_a+1} > T_{b,n_b+1}$. Similar results are available for the complementary event $T_{b,n_b+1} > T_{a,n_a+1}$, which can be derived by interchanging the indices for the groups above. However, it is not necessary to calculate lower and upper probabilities for both these events, because the well-known conjugacy property for imprecise probabilities, $\underline{P}(E) = 1 - \overline{P}(E^c)$, holds here [57], where E^c is the complementary event to E. Informally, this holds because our bounds are optimal, and correspond to the same assessments based on the rc-$A_{(n)}$ assumptions per group. One could opt to only compute either the lower or upper probabilities for both these events, requiring only a single algorithm, and using this relation to derive the other imprecise probabilities of interest. Implicit in our results is the assumption that the probability of $T_{a,n_a+1} = T_{b,n_b+1}$ is zero, which is reasonable for our method as long as there are no ties among the event times of different groups (it would particularly become a problem if an event time had been observed twice or more in each group, we discuss ties briefly in the final section), and which is a consequence of our method of comparison, where effectively we always put probability masses at end-points of different intervals.

5.3 Examples

In this section, our nonparametric predictive method for comparison of two groups of data is illustrated, using data from the literature.

Example 13
The data were used to illustrate NPI in Subsection 3.4.4 and concern the survival times of patients with cervical cancer, when given either treatment A (Control) or B (New therapy). We now focus on the comparison between these two treatments by using the method presented in Section 5.2.

The data from treatment A consist of 11 event times and 5 censoring times, i.e. $u_a = 11$, $v_a = 5$ and $n_a = 16$, and from treatment B we have 5 event times and 9 censoring times, i.e. $u_b = 5$, $v_b = 9$ and $n_b = 14$. Let $T_{a,17}$ and $T_{b,15}$ be two random quantities representing the next observations for treatment A and B, respectively. Then the method presented in Section 5.2 leads to

$$\overline{P}(T_{a,17} > T_{b,15}) = 0.473 \quad \text{and} \quad \underline{P}(T_{a,17} > T_{b,15}) = 0.226,$$

By the conjugacy property (2.2) for imprecise probabilities, this implies

$$\overline{P}(T_{b,15} > T_{a,17}) = 0.774 \quad \text{and} \quad \underline{P}(T_{b,15} > T_{a,17}) = 0.527.$$

These values indicate that the data provide fairly strong evidence for $T_{b,15} > T_{a,17}$, suggesting that treatment B has more ability than treatment A to prolong the patient's survival time. So, if a patient with cervical cancer was offered a choice between treatment A or B, she might be willing to choose B.

To compare our method to an alternative nonparametric method for inference for such data, we use Mantel's test [50] (Section 2.6), which gives an approximate one-sided p-value of 0.1020. Because Mantel's test can be used for comparison of unknown survival functions from two treatments, such a p-value may not be regarded as strong enough evidence against the null-hypothesis, namely there is some evidence that the survival times for treatment B is greater than A.

Treatment A		Treatment B	
86	822	173	> 1726
107	836	498	> 1763
141	> 1309	615	> 1807
296	1375	950	> 1879
312	> 1378	> 1190	> 1889
330	> 1446	> 1242	> 1897
346	> 1540	1408	> 1968
364	> 1645	> 1493	> 1972
401	> 1818	> 1572	> 2022
419	> 1910	> 1576	> 2070
505	> 1953	> 1585	> 2177
570	> 2052	> 1684	
688		> 1699	

Table 5.1: Relapse-free Survival Times for Hodgkin's disease patients ($> t$ indicates right-censoring at t).

Example 14

The data in Table 5.1 were used to illustrate Mantel's test in Section 2.6. We now use these data to illustrate our method presented in Section 5.2, leading to the lower and upper probabilities

$$\overline{P}(T_{b,25} > T_{a,26}) = 0.893 \quad \text{and} \quad \underline{P}(T_{b,25} > T_{a,26}) = 0.557.$$

These values indicate that the data provide pretty strong evidence for event $T_{b,25} > T_{a,26}$. In Section 2.6, by applying Mantel's test to these data, we derived an approximate one-sided p-value of 0.0006. So Mantel's test also suggest strongly that the survival functions corresponding to these two treatments are not equal, in other words, the nodal radiation is more effective than radiation of affected nodes in preventing or delaying the recurrence of early stage Hodgkin's disease.

It should be remarked that it might happen that Mantel's test would reject a null hypothesis when we would still have $\underline{P}(T_{a,n_a+1} > T_{b,n_b+1}) < 0.5 < \overline{P}(T_{a,n_a+1} > T_{b,n_b+1})$, because Mantel's test is based on the comparison of two unknown survival functions from groups A and B, while our method compares two future observations from these two groups, respectively. For example, suppose that all observations from group A are about equal to 3 years, and half the observations from group B are about equal to 2 years and the other half about 4 years, then with lots of observations for each group, Mantel's test will reject that two unknown survival functions of A and B are equal, but our method would give $\underline{P}(T_{a,n_a+1} > T_{b,n_b+1}) < 0.5 < \overline{P}(T_{a,n_a+1} > T_{b,n_b+1})$.

5.4 Concluding remarks

This chapter presents a novel method to compare two groups of lifetime data including right-censored observations. The method uses NPI as presented in Chapter 3, via comparison of the future observations from both groups A and B, and leads to upper and lower probabilities for $T_{a,n_a+1} > T_{b,n_b+1}$. This comparison is predictive, in which it is different with Mantel's test [50], which tests a hypothesis of equal survival functions. This nonparametric method generalizes the method by Coolen [13], which does not allow right-censored data.

It should be remarked that, although our method is presented by assuming that there are no ties in the data set, the method can also be adapted for dealing with tied observations. As there are two groups of lifetime data, ties may occur within each group as discussed in Section 3.6, and between the groups. For the first situation, our method can be adapted as in Section 3.6. For the second situation, if there are ties between an event time and a censoring time, or between censoring times, it does not affect our nonparametric predictive method. If there are ties between event times in both groups, one should break them into all possible orderings among the groups, calculate lower (upper) probabilities for each such ordering, and then take the minimum (maximum) of all these lower (upper) probabilities as the actual lower (upper) probability to be used for the comparison.

Coolen and van der Laan [18] presented NPI comparison of more than two groups, not allowing censored data. It will be relatively straightforward to generalize their results along the lines of this chapter, which is left as a topic for future research.

Chapter 6

Summary and concluding remark

This thesis presents nonparametric predictive analysis for lifetime data including right-censored observations. Although many nonparametric methods were proposed for such data, they are all not capable of dealing with exact censoring information (ECI).

The assumption $A_{(n)}$ [39] is a sound basis for prediction in case of vague prior knowledge of a probability distribution for observed random quantities. It provides a partially specified predictive distribution for a future observation given past observations, consisting of exact event times. However, $A_{(n)}$ does not allow right-censoring data in observations. Although Berliner and Hill [6] address the same problem in their method, they use partial censoring information. The question we address is how $A_{(n)}$ can be generalized to deal with ECI.

We generalized $A_{(n)}$, and presented the assumption right-censoring $A_{(n)}$ (rc-$A_{(n)}$), which is related to exchangeability of a right-censored observation with other random quantities in the risk set at the censoring time. The assumption rc-$A_{(n)}$ provides a partially specified predictive probability distribution for a future observation. Inference based on rc-$A_{(n)}$ uses ECI.

Although rc-$A_{(n)}$ is not sufficient to derive precise probabilities for many events of interest, it does provide bounds for probabilities. The derived lower and upper survival functions for T_{n+1} are well suited for graphical presentation giving a complete picture of the data, including right-censoring times. The lower survival function for T_{n+1} shows explicit changes at censoring times, which is not the case in the Berliner-Hill method [6] and Kaplan-Meier methods [46].

The assumption rc-$A_{(n)}$ and related inference can be used for other problems formulated in terms of T_{n+1}. The nonparametric predictive inference for grouped data is one such application. Grouped data with right-censored observations frequently occur in reliability and survival analysis. Our nonparametric method presented for grouped data is based on quite minimal modelling assumptions, and is directly in terms of T_{n+1}. We derived the optimal bounds for predictive probabilities by using the optimal configuration principle (OCP). An advantage of our nonparametric method is that it does not require the use of additional assumptions about the event times and censoring times, on which many other related methods for grouped data depend.

The nonparametric predictive comparison of two groups of lifetime data presents another application of rc-$A_{(n)}$. The comparison of two groups of lifetime data is also an elementary problem in statistics, which arises often in medical research. Our method presents the comparison of a future observation from each of the two groups, and leads to upper and lower probabilities for the event that the next observation from one population is greater than the next observation from the other population, which is different with other nonparametric methods proposed for such comparison. Our method takes all censoring times precisely into account.

The nonparametric predictive inference presented in the thesis is based on data including right-censored observations. Basically, these inferences are an attempt to keep structural assumptions minimal, and they are therefore suited if there is only extremely vague knowledge about the situation which is being modelled, other than that provided by the data set. Alternatively, such inferences can also be used as a basis for studying the influence of additional modelling assumptions on ultimate inferences, or related decisions, when we wish to use methods with more structure explicitly taken into account. The imprecise nature of these inferences may also lead to situations where optimal decisions are not derived. For example, when comparing two medical treatments predictively, on the basis of data including many censored observations, the range of survival functions between the lower and upper survival functions per treatment may well include pairs that would lead to preference of either treatment. In such cases, our method makes clear that strong inferences may not be possible based only on the data, so further modelling assumptions or more data are required. From this perspective, our inferences are related to robust statistical methods [4, 44].

In this thesis, we have only considered predictive inference for a single future observation. Extension of such inference to multiple future observations is of interest, particularly as random quantities representing future observations are mutually dependent [15, 39]. It would also be interesting to compare our lower and upper survival functions with other methods for lifetime data as presented in the literature, for example estimators based on counting processes [2]. More generally, detailed development and analysis, along the lines of Augustin and Coolen [4], of interval probabilities resulting from our partially specified probability distributions via M-function values, might provide an important contribution to the theory of interval probability.

Bibliography

[1] Aitchison, J. and Dunsmore, I.R. (1975). **Statistical Prediction Analysis**. Cambridge University Press, Cambridge, U.K.

[2] Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993). **Statistical Models Based on Counting Processes**. Springer-Verlag, New York.

[3] Arts, G.R.J., Coolen, F.P.A. and van der Laan, P. (2002). *Nonparametric predictive inference in statistical process control*. In submission.

[4] Augustin, T. and Coolen, F.P.A. (2002). *Nonparametric predictive inference and interval probability*. In submission.

[5] Berkson, J. and Gage, R.P. (1950). *Calculation of survival rates for cancer*. Proc. Staff Meet. Mayo Clin., **25**, 270-286.

[6] Berliner, L.M. and Hill, B.M. (1988). *Bayesian nonparametric survival analysis (with discussion)*. Journal of the American Statistical Association, **83**, 772-784.

[7] Bernard, J.M.(ed.) (2002). *Special issue on imprecise probabilities and their applications*. Journal of Statistical Planning and Inference, **105**, issue 1.

[8] Boole, G. (1854). **An Investigation of the Laws of Thought on which are founded the Mathematical Theories of Logic and Probabilities**. Walton and Maberley, London. (Reprinted: Dover, New York, 1951.)

[9] Breslow, N. (1970). *A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship*. Biometrika, **57**, 579-594.

[10] Chen, W.C., Hill, B.M., Greenhouse, J.B. and Fayos, J.V. (1985). *Bayesian Analysis of Survival Curves for Cancer Patients Following Treatment (with discussion).* In: **Bayesian Statistics 2** (J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith, eds.), Elsevier Science Publishers, 299-328.

[11] Chiang, C.L. (1960). *A stochastic study of the life table and its applications: I. Probability distributions of the biometric functions.* Biometrics, **16**, 618-635.

[12] Connor, R.J. and Mosimann, J.E. (1969). *Concepts of independence for proportions with a generalization of the Dirichlet distribution.* Journal of the American Statistical Association, **64**, 194-206.

[13] Coolen, F.P.A. (1996). *Comparing two populations based on low stochastic structure assumptions.* Statistics & Probability Letters, **29**, 296-305.

[14] Coolen, F.P.A. (1997). *An imprecise Dirichlet model for Bayesian analysis of failure data including right-censored observations.* Reliability Engineering and System Safety, **56**, 61-68.

[15] Coolen, F.P.A. (1998). *Low structure imprecise predictive inference for Bayes' problem.* Statistics & Probability Letters, **36**, 349-357.

[16] Coolen, F.P.A. and Coolen-Schrijner, P. (2000). *Condition monitoring: a new perspective.* Journal of the Operational Research Society, **51**, 311-319.

[17] Coolen, F.P.A., Coolen-Schrijner, P. and Yan, K.J. (2002). *Nonparametric predictive inference in reliability.* To appear in Reliability Engineering and System Safety.

[18] 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.

[19] Coolen, F.P.A. and Yan, K.J. (2002). *Nonparametric predictive inference with right-censored data.* In submission.

[20] Coolen, F.P.A. and Yan, K.J. (2002). *Nonparametric predictive inference for grouped lifetime data.* In submission.

[21] Coolen, F.P.A. and Yan, K.J. (2002). *Nonparametric predictive comparison of two groups of lifetime data.* In submission.

[22] Coolen, F.P.A. and Yan, K.J. (2002). *The use of right-censored data in nonparametric predictive inference.* Proceedings of the Third International Conference on Mathematical Methods in Reliability, Trondheim (Norway), June 2002, 155-158.

[23] Cox, D.R. and Oakes, D. (1984). **Analysis of Survival Data**. Chapman and Hall, London.

[24] Cozman, F. and Moral, S. (eds.) (2000). *Special volume on imprecise probabilities.* International Journal of Approximate Reasoning, **24**, 121-299.

[25] Crowder, M.J., Kimber, A.C., Smith, R.L. and Sweeting, T.J. (1991). **Statistical Analysis of Reliability Data**. Chapman and Hall, London.

[26] de Cooman, G. (ed.) (2000). *Special issue on imprecise probabilities.* Risk, Decision and Policy, **5**, 107-181.

[27] de Cooman, G., Cozman, F.G., Moral, S. and Walley, P. (eds.) (1999). **ISIPTA'99: Proceedings of the First International Symposium on Imprecise Probabilities and their Applications**. University of Ghent, Belgium.

[28] de Cooman, G., Fine, T.L., Moral, S. and Seidenfeld, T. (eds.) (2001). **ISIPTA'01: Proceedings of the Second International Symposium on Imprecise Probabilities and their Applications**. Cornell University, Ithaca, New York.

[29] de Cooman, G. and Walley, P. (eds.) (2001). **The Imprecise Probability Project**. www.sipta.org.

[30] De Finetti, B. (1974). **Theory of Probability (2 volumes)**. Wiley, London.

[31] Dempster, A.P. (1963). *On Direct Probabilities.* Journal of the Royal Statistical Society, B, **25**, 100-114.

[32] Efron, B. (1967). *The two sample problem with censored data.* In: **Proceedings of the Fifth Berkeley Symposium on Mathematical**

Statistics and Probability 4, University of California Press, 837-853.

[33] Elveback, L. (1958). *Estimation of survivorship in chronic disease: The 'actuarial' method.* Journal of the American Statistical Association, **53**, 420-440.

[34] Freireich, J., Gehan, E., Frei, E., et al (1963). *The efffect of 6-mercaptoburine on the doration of steroid-induced remissions in acute leukemia: a model for evaluation of other potentially useful therapy.* Blood, **21**, 699–716.

[35] Gehan, E.A. (1965a). *A generalized Wilcoxon test for comparing arbitrarily singly-censored samples.* Biometrika, **52**, 203-223.

[36] Gehan, E.A. (1965b). *A generalized two-sample Wilcoxon test for doubly censored data.* Biometrika, **52**, 650-653.

[37] Geisser, S. (1982). *Aspects of the predictive and estimative approaches in the determination of probabilities (with discussion).* Biometrics, **38**, 75-93.

[38] Geisser, S. (1993). **Predictive Inference: an Introduction**. Chapman and Hall, London.

[39] Hill, B.M. (1968). *Posterior distribution of percentiles: Bayes theorem for sampling from a finite population.* Journal of the American Statistical Association, **63**, 677-691.

[40] Hill, B.M. (1988). *De Finetti's theorem, induction, and An, or Bayesian nonparametric predictive inference (with discussion).* In: **Bayesian Statistics 3** (J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds.), Oxford University Press, 211-241.

[41] Hill, B.M. (1992). *Bayesian nonparametric survival analysis: a comparison of the Kaplan-Meier and Berliner-Hill estimators.* Survival Analysis: State of the Art (J.P. Klein and P.K. Goel, eds.), NATO ASI series, Kluwer Academic Publishers, 25-46.

[42] Hill, B.M. (1993). *Parametric models for $A_{(n)}$: splitting processes and mixtures.* Journal of the Royal Statistical Society, B, **55**, 423-433.

[43] Hollander, M. and Wolfe, D.A. (1999). **Nonparametric Statistical Methods**. Wiley, New York.

[44] Huber, P.J. (1981). **Robust Statistics**. Wiley, New York.

[45] Johnson, N. and Elandt-Johnson, R. (1980). **Survival Models and Data Analysis**. Wiley, New York.

[46] Kaplan, E.L. and Meier, P. (1958). *Nonparametric estimation from incomplete observations.* Journal of the American Statistical Association, **53**, 457-481.

[47] Lane, D.A. and Sudderth, W.D. (1984). *Coherent predictive inference.* Sankhyā: The Indian Journal of Statistics, A, **46**, 166-185.

[48] Lawless, J.F. (1982). **Statistical Models and Methods for Lifetime Data**. Wiley, New York.

[49] Lochner, R.H. (1975). *A generalized Dirichlet distribution in Bayesian life testing.* Journal of the Royal Statistical Society, B, **37**, 103-113.

[50] Mantel, N. (1966). *Evaluation of survival data and two new rank order statistics arising in its consideration.* Cancer Chemotherapy Reports, **50**, 113-170.

[51] McIllmurray, M.B. and Turkie W (1987). *Controlled trial of γ linolenic acid in Dukes' C colorectal cancer.* British Medical Journal, **294**, 1260; **295**, 475.

[52] Meeker, W.Q. and Escobar, L.A. (1998). **Statistical Methods for Reliability Data**. Wiley, New York.

[53] Miller, R.G., Jr. (1981). **Survival Analysis**. Wiley, New York.

[54] Parmar, M.K.B. and Machin, D. (1995). **Survival Analysis: a Practical Approach**. Wiley, Chichester.

[55] Prentice, R. and Kalbfleisch, J. (1980). **The Statistical Analysis of Failure Time Data**. Wiley, New York.

[56] van Noorwijk, J.M., Dekker, R., Cooke, R.M. and Mazzuchi, T.A. (1992). *Expert judgement in maintenance optimization.* IEEE Transactions on Reliability, **41**, 427-432.

[57] Walley, P. (1991). **Statistical Reasoning with Imprecise Probabilities**. Chapman and Hall, London.

[58] Walley, P. (1996). *Inferences from multinomial data: learning about a bag of marbles (with discussion)*. Journal of the Royal Statistical Society, B, **58**, 3-57.