Nonparametric Predictive Multiple Comparisons with Censored Data and Competing Risks

Tahani A. Maturi
Dept of Mathematical Sciences,
Durham University, UK
tahani.maturi@durham.ac.uk

Pauline Coolen-Schrijner
Dept of Mathematical Sciences,
Durham University, UK

Frank P.A. Coolen
Dept of Mathematical Sciences,
Durham University, UK
frank.coolen@durham.ac.uk

Abstract

This paper provides an overview of nonparametric predictive inference for comparison of multiple groups of data including right-censored observations. Different right-censoring schemes discussed are early termination of an experiment, progressive censoring and competing risks. Theoretical results are briefly stated, detailed justifications are presented elsewhere. The methods are illustrated and discussed via examples with data from the literature.

Keywords. Competing risks, early termination, nonparametric predictive inference, precedence testing, progressive censoring, right-censored data.

1 Introduction

This paper presents a brief overview of recent results on nonparametric predictive inference (NPI) for multiple comparisons in situations with right-censored observations. Such data typically occur in reliability or survival analysis, due to several reasons. For example, when interest is in a specific failure mode for a technical unit, it may fail due to a different failure cause. If multiple failure modes are of interest, and failure will be due to only a single failure mode, then this situation is known as "competing risks", where an observed failure time is actually a right-censoring time with regard to all failure modes that did not cause the failure. Another reason for right-censoring may be removal of units from a lifetime experiment, normally to save time or reduce cost, but this also occurs if, at some point, one wishes to study units which have not yet failed in an experiment in more detail. If right-censoring is due to an experiment being terminated before all units have failed, multiple comparisons of different groups of units based on such data is known as "precedence testing". If non-failing units are removed from the experiment at several possible stages it is known as "progressive censoring". Recently, we have developed NPI for multiple comparisons for precedence testing, progressive censoring, and competing risks, and these results are briefly presented here and illustrated and discussed via examples. Detailed justifications of the results are presented elsewhere. It should be emphasized that, throughout the paper, unspecified reasons for right-censoring are assumed to be based on processes that are independent of the residual lifetimes of the censored units.

NPI is a statistical method that aims at using relatively few modelling assumptions, it uses lower and upper probabilities to quantify uncertainty. Some basic applications of NPI in reliability were summarized by Coolen, *et al* [12], recently a variety of further applications in this area have been presented, including probabilistic safety assessment if zero failures have been observed [7], prediction of not-yet occurred failure modes [8], comparison of success-failure data [17], and system reliability with optimal redundancy allocation [18]. NPI has also been developed for replacement problems, with specific attention to age replacement of technical units [19, 21]. Imprecise probabilistic methods are attractive in reliability, as their flexibility for dealing with limited information is a particular advantage for dealing with practical aspects of many reliability situations. Utkin and Coolen [35] present an extensive overview of the literature, for a concise overview see [13].

In Section 2 of this paper NPI is briefly introduced, followed in Section 3 by explanation of the way in which NPI deals with right-censored data. Recent developments of NPI for multiple comparisons with the different right-censoring schemes discussed above are presented in Sections 4, 5 and 6, and illustrated and discussed in examples in Section 7. The same notation is used for different quantities in Sections 4-6, but in the general NPI approach to multiple comparisons they all relate to similar concepts, just the interpretations in the specific applications are different per section.

2 Nonparametric predictive inference

Nonparametric predictive inference (NPI) is based on Hill's assumption $A_{(n)}$ [25], which implies direct (lower and upper) probabilities for a future observable random quantity, based on observed values of n related random quantities [6]. Suppose that $X_1, \ldots, X_n, X_{n+1}$ are positive, continuous and exchangeable random quantities representing lifetimes. Let the ordered observed values of $X_1, \ldots, X_n$ be denoted by $x_1 < x_2 < \ldots < x_n < \infty$, and let $x_0 = 0$ and $x_{n+1} = \infty$ for ease of notation, note that the latter is not considered to be an observation for X_{n+1}. We assume that no ties occur, our results can be generalised to allow ties [26]. For positive X_{n+1}, representing a future observation, based on n observations, $A_{(n)}$ assigns $P(X_{n+1} \in (x_i, x_{i+1})) = 1/(n+1)$ for $i = 0, 1, \ldots, n$. $A_{(n)}$ does not assume anything else, and is a post-data assumption related to exchangeability [22]. Hill [24] discusses $A_{(n)}$ in detail, and he also provided a Bayesian justification for $A_{(n)}$ under finite additivity [26]. Inferences based on $A_{(n)}$ can be considered suitable if there is hardly any knowledge about the random quantity of interest, other than the n observations, or if one does not want to use such information. $A_{(n)}$ is not sufficient to derive precise probabilities for many events of interest, but it provides bounds for probabilities via the 'fundamental theorem of probability' [22], which are lower and upper probabilities in interval probability theory [36, 37]. NPI has strong consistency properties within the theory of interval probability [1], attractive frequentist properties, and compares favourably to objective Bayesian methods [6, 24].

3 NPI for right-censored data

Coolen and Yan [16] presented rc-$A_{(n)}$ as a generalization of $A_{(n)}$ for right-censored data, using the additional assumption that, at a moment of censoring, the residual lifetime of a right-censored unit is exchangeable with the residual lifetimes of all other units that have not yet failed or been censored.

Suppose that there are n observations consisting of u event times, $x_1 < x_2 < \ldots < x_u$, and $v(= n - u)$ right-censored observations, $c_1 < c_2 < \ldots < c_v$. Let $x_0 = 0$ and $x_{u+1} = \infty$, and suppose that there are s_i right-censored observations in the interval (x_i, x_{i+1}) at times $c_1^i < c_2^i < \ldots < c_{s_i}^i$, where $\sum_{i=0}^{u} s_i = v$. These data can also be denoted by pairs (t_i, δ_i) for $i = 1, \ldots, n$, where $t_i = x_i$ (so a failure time, or time of other actual event of interest) if $\delta_i = 1$ and $t_i = c_i$ (a right-censored observation) if $\delta_i = 0$. For ease of notation, let $(t_0, \delta_0) = (0, 1)$ and $x_{n+1} = \infty$. The assumption rc-$A_{(n)}$ partially specifies the probability distribution for X_{n+1} by the following M-functions [16], for $i = 1, \ldots, n$:

$$M_{X_{n+1}}(t_i, x_{i+1}) = \frac{1}{n+1} (\tilde{n}_{t_i})^{\delta_i - 1} \prod_{\{r: c_r < t_i\}} \frac{\tilde{n}_{c_r} + 1}{\tilde{n}_{c_r}} \quad (1)$$

where $\tilde{n}_{c_r}$ and $\tilde{n}_{t_i}$ are the numbers of units in the risk set (i.e. that have not yet failed or been censored) just prior to time c_r and t_i, respectively. These M-functions are basic probability assignments in the sense of Shafer [33], and lead to the following precise probabilities for X_{n+1} to be between two consecutive observed failure times x_i and x_{i+1},

$$P(X_{n+1} \in (x_i, x_{i+1})) = \frac{1}{n+1} \prod_{\{r: c_r < x_{i+1}\}} \frac{\tilde{n}_{c_r} + 1}{\tilde{n}_{c_r}} \quad (2)$$

Coolen and Yan [15] developed NPI for comparison of two groups of lifetime data including right-censored observations. By applying the appropriate rc-$A_{(n)}$ assumption for each group, their method is based on comparing the next observation from each group, say X_{n_x+1} and Y_{n_y+1}. The NPI lower and upper probabilities for the event that $X_{n_x+1} < Y_{n_y+1}$ are

$$\underline{P} = \sum_{i=0}^{u_x} \sum_{j=0}^{n_y} 1(x_{i+1} < t_{y,j}) P^X(x_i, x_{i+1}) M^Y(t_{y,j}, y_{j+1})$$

$$\overline{P} = \sum_{i=0}^{n_x} \sum_{j=0}^{u_y} 1(t_{x,i} < y_{j+1}) P^Y(y_j, y_{j+1}) M^X(t_{x,i}, x_{i+1})$$

where $M^X(t_{x,i}, x_{i+1})$, $M^Y(t_{y,j}, y_{j+1})$, $P^X(x_i, x_{i+1})$ and $P^Y(y_j, y_{j+1})$ are as given by (1) and (2), and $1(A)$ is the indicator function that equals 1 if A is true and 0 else. Coolen and Yan [15] did not consider situations with more than two groups, nor the effect of early termination of the lifetime experiment or the specific features of progressive censoring and competing risks. NPI for multiple comparisons for real-valued data without right-censored observations was presented in [14], and NPI multiple comparisons for Bernoulli data in [11].

4 Early termination of experiment

In some circumstances, mostly in order to save costs or time, an experiment to compare lifetimes of units in different groups may be terminated before all units have failed. We assume that all units are placed simultaneously on a lifetime experiment which is terminated at a certain specified time, which may also be the moment a specified number of failures have occurred. The situation where for all units failing before the moment of termination of the experiment the lifetimes are observed, is also known as *precedence testing* in the literature [3]. Coolen-Schrijner *et al* [20] presented NPI for comparison of two groups of lifetime data with early termination of the experiment, say at time T_0, and they illustrated the effect of varying T_0. The resulting data set contains, for each of the two groups in the experiment, failure times prior to T_0 and right-censored observations at T_0 for all units that do not fail before T_0. Maturi *et al* [28] extend this to more than two groups, with a variety of inferential goals for the multiple comparisons in line with different goals as presented in the statistical selection literature [4]. Maturi *et al* [30] present further generalized results, which also generalize the results by Coolen and Yan [15], by developing NPI for comparison of multiple groups of lifetime data including right-censored observations, and with possible early termination of the experiment.

Consider an experiment to compare lifetimes of units from $k \geq 2$ groups, which are assumed to be fully independent, with the experiment starting on all units at time 0. The experiment can be terminated before all units have failed, say at time T_0. This T_0 can be fixed or random, but it is essential that it is assumed not to hold any information on residual time-to-failure for units that have not yet failed. We also allow non-informative right-censoring to occur for some units before the experiment is stopped. For group j, $j = 1, \ldots, k$, n_j units are in the experiment, of which u_j units fail before (or at) T_0, with ordered failure times $0 < x_{j,1} < x_{j,2} < \ldots < x_{j,u_j} \leq T_0$, and with right-censoring times $c_{j,1} < c_{j,2} < \ldots < c_{j,v_j} < T_0$. Let $x_{j,0} = 0$ and $x_{j,u_j+1} = \infty$ $(j = 1, \ldots, k)$, and let s_{j,i_j} be the number of right-censored observations in the interval (x_{j,i_j}, x_{j,i_j+1}), with $x_{j,i_j} < c^{i_j}_{j,1} < c^{i_j}_{j,2} < \ldots < c^{i_j}_{j,s_{j,i_j}} < x_{j,i_j+1}$ and $\sum_{i_j=0}^{u_j} s_{j,i_j} = v_j$, so $n_j-(u_j+v_j)$ units from group j are right-censored at T_0.

For NPI with data containing right-censored observations, and with early termination of the experiment at time T_0, the assumption rc-$A_{(n_j)}$ implies that the following M-function values apply for a nonnegative random quantity X_{j,n_j+1}, on the basis of data consisting of u_j failure times and $(n_j - u_j)$ right-censored observations:

$$M^j_{i_j} = M_{X_{j,n_j+1}}(x_{j,i_j}, x_{j,i_j+1}) = \frac{1}{n_j+1} \prod_{\{r: c_r < x_{j,i_j}\}} \frac{\tilde{n}_{j,c_r}+1}{\tilde{n}_{j,c_r}}$$

$$M^j_{i_j,a_j} = M_{X_{j,n_j+1}}(c^{i_j}_{j,a_j}, x_{j,i_j+1}) = \frac{(\tilde{n}_{j,c^{i_j}_{j,a_j}})^{-1}}{n_j+1} \prod_{\{r: c_r < c^{i_j}_{j,a_j}\}} \frac{\tilde{n}_{j,c_r}+1}{\tilde{n}_{j,c_r}}$$

$$M^j_{T_0} = M_{X_{j,n_j+1}}(T_0, \infty) = \frac{n_j-(u_j+v_j)}{n_j+1} \prod_{\{r: c_r < T_0\}} \frac{\tilde{n}_{j,c_r}+1}{\tilde{n}_{j,c_r}}$$

where $i_j = 0, \ldots, u_j$, $a_j = 1, \ldots, s_{j,i_j}$, and $\tilde{n}_{j,c_r}$ and $\tilde{n}_{j,c^{i_j}_{j,a_j}}$ are the number of units from group j in the risk set just prior to time c_r and $c^{i_j}_{j,a_j}$, respectively. Also

$$P^j_{i_j} = P(X_{j,n_j+1} \in (x_{j,i_j}, x_{j,i_j+1})) = \frac{1}{n_j+1} \prod_{\{r: c_r < x_{j,i_j+1}\}} \frac{\tilde{n}_{j,c_r}+1}{\tilde{n}_{j,c_r}}$$

$$P^j_{T_0} = P(X_{j,n_j+1} \in (T_0, \infty)) = M_{X_{j,n_j+1}}(T_0, \infty) = M^j_{T_0}$$

The NPI lower and upper probabilities for the event that the next observed lifetime from group l is the maximum of all next observed lifetimes for the k groups in the experiment, i.e. $X_{l,n_l+1} = \max_{1\leq j\leq k} X_{j,n_j+1}$, are

$$\begin{aligned}
\underline{P}^{(l)} &= \sum_{i_l=0}^{u_l} \left\{ \prod_{\substack{j=1\\ j\neq l}}^{k} \left[\sum_{i_j=0}^{u_j} 1(x_{j,i_j+1} < x_{l,i_l}) P^j_{i_j} \right] M^l_{i_l} \right. \\
&+ \left. \sum_{a_l=1}^{s_{l,i_l}} \prod_{\substack{j=1\\ j\neq l}}^{k} \left[\sum_{i_j=0}^{u_j} 1(x_{j,i_j+1} < c^{i_l}_{l,a_l}) P^j_{i_j} \right] M^l_{i_l,a_l} \right\} \\
&+ M^l_{T_0} \prod_{\substack{j=1\\ j\neq l}}^{k} \sum_{i_j=0}^{u_j} 1(x_{j,i_j+1} < T_0) P^j_{i_j} \qquad (3)
\end{aligned}$$

$$\begin{aligned}
\overline{P}^{(l)} &= \sum_{i_l=0}^{u_l} P^l_{i_l} \prod_{\substack{j=1\\ j\neq l}}^{k} \left\{ \sum_{i_j=0}^{u_j} 1\,(x_{j,i_j} < x_{l,i_l+1})\, M^j_{i_j} \right. \\
&+ \sum_{i_j=0}^{u_j} \sum_{a_j=1}^{s_{j,i_j}} 1(c^{i_j}_{j,a_j} < x_{l,i_l+1}) M^j_{i_j,a_j} \\
&+ \left. 1\,(T_0 < x_{l,i_l+1})\, M^j_{T_0} \right\} + P^l_{T_0} \qquad (4)
\end{aligned}$$

If the experiment is not terminated before the event times of all units have been observed, so for each unit either the failure time or a right-censoring time not due to the experiment ending, then the terms including T_0 in formulae (3) and (4) disappear, and we get a generalization of the results by Coolen and Yan [15], who only considered NPI for comparison of two groups of lifetime data including right-censored observations. Another special case occurs if there are no right-censored observations before T_0. In this case our method generalizes the results by Coolen-Schrijner *et al* [20], who considered NPI for comparison of two groups with early termination of the experiment, but without earlier right-censoring.

At any value of T_0, we can state that the data provide a strong indication that group l is the best if

$\underline{P}^{(l)} > \overline{P}^{(j)}$ for all $j \neq l$. It might seem attractive to state that, if $\underline{P}^{(l)} > \underline{P}^{(j)}$ and $\overline{P}^{(l)} > \overline{P}^{(j)}$ for all $j \neq l$, there would be a weak indication that group l is the best. The difference between the upper and lower probabilities reflects the amount of information available, it decreases if more relevant information becomes available. A typical feature of NPI for these methods with the experiment terminated at T_0 is that, if T_0 is increased, the upper (lower) probability never increases (decreases), while its value can only change at observed event times.

5 Progressive censoring

Maturi *et al* [29] considered the comparison of two groups, say X and Y, in which progressive censoring schemes are applied for one or both groups. They allow several such censoring schemes, known in the literature as progressive Type-I censoring, progressive Type-II censoring and Type-II progressively hybrid censoring scheme [2]. The main characteristic of progressive censoring is that, at several stages some units are randomly removed from the experiment. For NPI for a progressive Type-II censoring scheme with $R = (R_1, R_2, \ldots, R_r)$, where R_i is the number of units that are removed from the experiment at the ith failure, the assumption rc-$A_{(n)}$ implies that the probability distribution for a nonnegative random quantity X_{n+1} on the basis of data including r real and $n-r$ progressively censored observations, is partially specified by the following M-function values, for $i = 0, 1, \ldots, r$,

$$M^X(x_i, x_{i+1}) = \frac{1}{n+1} \prod_{k=1}^{i-1} \frac{n-k-\sum_{l=1}^{k-1} R_l + 1}{n-k-\sum_{l=1}^{k} R_l + 1} \quad (5)$$

$$M^X(x_i^+, x_{i+1}) = \frac{R_i}{n-i-\sum_{l=1}^{i} R_l + 1} \, M^X(x_i, x_{i+1}) \quad (6)$$

where x_i^+ represents the lower bound for the interval that contains the set of censored units at x_i, $x_0 = 0$ and $x_{r+1} = \infty$. The corresponding NPI probabilities for X_{n+1} to be in (x_i, x_{i+1}) are

$$P^X(x_i, x_{i+1}) = \frac{1}{n+1} \prod_{k=1}^{i} \frac{n-k-\sum_{l=1}^{k-1} R_l + 1}{n-k-\sum_{l=1}^{k} R_l + 1} \quad (7)$$

Suppose that we have two independent groups, X and Y, for which n_x and n_y units, respectively, are placed on a lifetime experiment. Both groups are progressively Type-II censored with the schemes $R^x = (R_1^x, R_2^x, \ldots, R_{r_x}^x)$ and $R^y = (R_1^y, R_2^y, \ldots, R_{r_y}^y)$. Given the data, R^x, R^y, and the assumptions rc-$A_{(n_x)}$ and rc-$A_{(n_y)}$, the NPI lower and the upper probabilities that the next observation from group Y is greater than the next observation from group X, are

$$\underline{P} = \sum_{j=0}^{r_y} \sum_{i=0}^{r_x} 1(x_{i+1} < y_j) P^X(x_i, x_{i+1}) P^Y(y_j, y_{j+1}) \quad (8)$$

$$\overline{P} = \sum_{j=0}^{r_y} \sum_{i=0}^{r_x} 1(x_i < y_{j+1}) P^X(x_i, x_{i+1}) P^Y(y_j, y_{j+1}) \quad (9)$$

We refer to [29] for NPI comparisons in case of progressive Type-I and Type-II progressively hybrid censoring. It should be emphasized that, in classical frequentist methods for such comparisons [2], via hypothesis tests of assumed equality of underlying lifetime distributions, the details of the exact applied censoring scheme are relevant, as they influence the counter-factuals, outcomes of the experiment that were possible but did not occur. In NPI such counterfactuals play no role, as the comparison is directly based on random quantities representing lifetimes of one future unit per group. The different censoring schemes affect the M-function values, but the corresponding derivations of the lower and upper probabilities of interest is similar in all cases.

6 Competing risks

In competing risks, a unit is subject to failure from one of k distinct failure modes. Throughout we assume that these failure modes are independent. Tsiatis [34] showed that competing risks data as considered here do not hold information about dependence of failure modes. We assume that the unit fails due to the first occurrence of a failure caused by one of the possible failure modes, at which moment it is withdrawn from further use. We suppose that such failure observations are obtained for n units, and that failure modes causing failures are known with certainty. As is common in study of failure data under competing risks, we consider for each unit k random quantities, say T_i for $i = 1, \ldots, k$, where T_i represents the unit's time to failure under the condition that failure occurs due to failure mode i. We assume that these T_i are independent continuous random quantities, and the failure time of the unit is $T = \min(T_1, \ldots, T_k)$. Therefore, for each unit considered we can have one failure time and we will know, with certainty, the failure mode that caused the failure. Hence, for the T_i corresponding to the other failure modes, which did not cause the failure of the unit, the unit's observed failure time is a right-censoring time.

For the NPI approach, let the failure time of a future item be denoted by X_{n+1}, and let the corresponding notation for the failure time including indication of the actual failure mode, say failure mode j, be $X_{j,n+1}$ (so X_{n+1} corresponds to an observation T for unit $n+1$, and $X_{j,n+1}$ to T_j, according to the notation in the previous paragraph). As we assume independence between the different failure modes, our competing risk data per failure mode consist of (possibly) a number of observed failure times for failures caused by the specific failure mode considered, and right-censoring times for failures caused by other failure modes. Hence we can apply rc-$A_{(n)}$ per failure mode j, for inference on $X_{j,n+1}$. Let the number of failures caused by failure mode j be u_j and let $\upsilon_j(=n-u_j)$ be the number of the right-censored observations corresponding to failure mode j. It should be emphasized that we do not assume that each unit considered must actually fail, if a unit does not fail then there will be a right-censored observation recorded for this unit for each failure mode, as we assume that the unit will then be withdrawn from the study, or the study ends, at some point. The random quantity representing the failure time of the next unit, with all k failure modes considered, is $X_{n+1} = \min_{1\leq j\leq k} X_{j,n+1}$.

For failure mode j, $j=1,\ldots,k$, we have as data n pairs $(t_{j,i_j}, \delta_{j,i_j})$, for $i_j = 1,\ldots,n$, where $\delta_{j,i_j} = 1$ if a failure at time $t_{j,i_j}(=x_{j,i_j})$ was caused by failure mode j and where $\delta_{j,i_j} = 0$ denotes that the event at the corresponding time $t_{j,i_j}(=c_{j,i_j})$ is, for as far as this specific failure mode j is concerned, a right-censored observation.

We can specify the NPI M-functions for $X_{j,n+1}$ ($j=1,\ldots,k$), similar to (1), as

$$M^j_{t_{j,i_j}} = M^j(t_{j,i_j}, x_{j,i_j+1}) = \frac{(\tilde{n}_{t_{j,i_j}})^{\delta_{j,i_j}-1}}{(n+1)} \prod_{\{r:c_r<t_{j,i_j}\}} \frac{\tilde{n}_{c_r}+1}{\tilde{n}_{c_r}} \quad (10)$$

with $\tilde{n}_{c_r}$ and $\tilde{n}_{t_{j,i_j}}$ the numbers of units in the risk set just prior to times c_r and t_{j,i_j}, respectively. The corresponding NPI probabilities, similar to (2), are

$$P^j = P^j(x_{j,i_j}, x_{j,i_j+1}) = \frac{1}{n+1} \prod_{\{r:c_r<x_{j,i_j+1}\}} \frac{\tilde{n}_{c_r}+1}{\tilde{n}_{c_r}} \quad (11)$$

where x_{j,i_j} and x_{j,i_j+1} are two consecutive observed failure times caused by failure mode j (and $x_{j,0}=0$, $x_{j,n+1}=\infty$).

The event of interest is that a single future unit, which we call the 'next unit', undergoing the same test or process as the n units for which failure data are available, fails due to a specific failure mode, say mode l. The NPI lower and upper probabilities for the event $X_{l,n+1} = \min_{1\leq j\leq k} X_{j,n+1}$, for $l=1,\ldots,k$, are

$$\underline{P}^{(l)} = \sum_{\substack{i_j=0\\ j\neq l}}^{n} \left[\sum_{i_l=0}^{u_l} 1(x_{l,i_l+1} < \min_{\substack{1\leq j\leq k\\ j\neq l}} \{t_{j,i_j}\}) P^l \right] \prod_{\substack{j=1\\ j\neq l}}^{k} M^j_{t_{j,i_j}} \quad (12)$$

$$\overline{P}^{(l)} = \sum_{\substack{i_j=0\\ j\neq l}}^{u_j} \left[\sum_{i_l=0}^{n} 1(t_{l,i_l} < \min_{\substack{1\leq j\leq k\\ j\neq l}} \{x_{j,i_j+1}\}) M^l_{t_{l,i_l}} \right] \prod_{\substack{j=1\\ j\neq l}}^{k} P^j \quad (13)$$

where the first summation signs denote the sums over all i_j from 0 to n or u_j for $j=1,\ldots,k$ but not including $j=l$. The derivation of these NPI lower and upper probabilities is given in [31].

We briefly consider the special case of the general competing risks problem in which there are only two failure modes (so $k=2$), 1 and 2, also denoted by FM1 and FM2, and with all n units considered actually failing due to one of these two failure modes. Therefore, any unit which fails due to FM1 leads to a right-censored observation for FM2, and vice versa. In this case, the number of failures due to FM1 (FM2) is equal to the number of right-censored observations for FM2 (FM1), so $\upsilon_1 = u_2$ and $\upsilon_2 = u_1$. The NPI lower and upper probabilities for the event that the next unit will fail due to FM1 are

$$\underline{P}^{(1)} = \sum_{i_2=0}^{n} \left\{ \sum_{i_1=0}^{u_1} 1(x_{1,i_1+1} < t_{2,i_2}) P^1 \right\} M^2_{t_{2,i_2}} \quad (14)$$

$$\overline{P}^{(1)} = \sum_{i_2=0}^{u_2} \left\{ \sum_{i_1=0}^{n} 1(t_{1,i_1} < x_{2,i_2+1}) M^1_{t_{1,i_1}} \right\} P^2 \quad (15)$$

This special case enables us to illustrate some interesting features of the NPI approach in this setting. We consider two specific scenarios in detail:

(A) all failures due to FM2 come first, followed by all failures due to FM1, meaning that the u_2 failure times of failures due to FM2 are all smaller than the u_1 failure times of failures due to FM1. In this case, the NPI lower and upper probabilities for the event that the next unit will fail due to FM1 are

$$\underline{P}^{(1),\mathrm{A}} = \frac{1}{u_1+1} \sum_{i_2=1}^{\upsilon_2} i_2 \; M^2(c_{2,i_2}, \infty)$$

$$\overline{P}^{(1),\mathrm{A}} = \frac{1}{n+1} \left[\upsilon_2 + 1 + \frac{u_2}{n+1} + \sum_{i_1=1}^{\upsilon_1-1} (\upsilon_1 - i_1) \; M^1(c_{1,i_1}, x_{1,1}) \right]$$

(B) all failures due to FM1 come first, followed by all failures due to FM2, in which case the NPI lower and upper probabilities for the event that the next unit will fail due to FM1 are

$$\underline{P}^{(1),\mathrm{B}} = \frac{1}{n+1}\left[\frac{u_1\, u_2}{u_2+1} + \sum_{i_1=1}^{v_2} i_1\, M^2(c_{2,i_2}, x_{2,1})\right]$$

$$\overline{P}^{(1),\mathrm{B}} = \frac{1}{u_2+1}\left[1+\frac{u_2(u_1+1)}{n+1}+\sum_{i_1=1}^{v_1-1}(v_1-i_1)M^1(c_{1,i_1},\infty)\right]$$

These NPI lower and upper probabilities follow straightforwardly from the general expressions given before. The main reason for highlighting these two special cases is an interesting observation in our study of NPI for competing risks data, namely that case (A) always seems to give the minimal NPI lower and upper probabilities, when all possible orderings of u_2 failures due to FM2 and u_1 failures due to FM1 are considered, while case (B) always seems to give the maximal NPI lower and upper probabilities. For now, we propose this property as a conjecture, which we strongly believe to hold and hope to prove generally in the near future.

7 Examples

7.1 Example I: Early termination

Desu and Raghavarao [23] present recorded times (months) until promotion at a large company, for 19 employees in $k = 3$ departments. The data are: Dept 1: $15, 20+, 36, 45, 58, 60$ ($n_1 = 6$); Dept 2: $12, 25+, 28, 30+, 30+, 36, 40, 45, 48$ ($n_2 = 9$); Dept 3: $30+, 40, 48, 50$ ($n_3 = 4$), where " + " indicates that the employee left the company at that length of service before getting promotion, this is considered to be a right-censored observation (one could argue about whether or not this right-censoring process is independent of the promotion process, but as we only use this data set for illustration, and have no further circumstantial information, we do not address this in more detail). We consider at which department the data suggest that one needs to work the longest to get a promotion. This data set contains tied observations, in NPI these are dealt with by assuming that they differ by a very small amount, in such a way that the lower (or upper) probability of interest is smallest (largest) over all possible ways to break the ties.

To illustrate NPI for multiple comparisons with early termination, as summarized in Section 4, assume that all these employees started working at this company at the same time, and that one considers the data after T_0 months, so all larger observations in the data above are treated as being right-censored at T_0. For several values of T_0, the lower and upper probabilities for the event that one has to work the longest in department l, $\underline{P}^{(l)}$ and $\overline{P}^{(l)}$, for $l = 1, 2, 3$, are presented in Table 1. There is no value of T_0 for which the corresponding data would strongly indicate that one of the departments leads to longest time to promotion, according to the formulation of such indications as explained in Section 4. For several T_0, for example $T_0 = 17$, both the lower and upper probabilities for department 3 are greater than the lower and upper probabilities, respectively, for department 1 and for department 2. As discussed in Section 4, one could argue that this provides a weak indication that department 3 leads to the longest times until promotion. However, the large imprecision in these lower and upper probabilities indicates that the evidence for such a claim is weak, so care must be taken when formulating any conclusion along these lines. For larger values of T_0, department 3 has most imprecision remaining, which reflects that there are only few observations for this department.

T_0	$\underline{P}^{(1)}$	$\overline{P}^{(1)}$	$\underline{P}^{(2)}$	$\overline{P}^{(2)}$	$\underline{P}^{(3)}$	$\overline{P}^{(3)}$
11	0	1	0	1	0	1
14	0	1	0	0.903	0	1
17	0	0.863	0	0.903	0.011	1
27	0	0.863	0	0.903	0.011	1
33	0	0.863	0	0.797	0.024	1
38	0	0.714	0	0.659	0.089	1
42	0.068	0.714	0.025	0.540	0.114	0.833
47	0.081	0.615	0.032	0.434	0.197	0.833
49	0.167	0.615	0.032	0.354	0.216	0.748
52	0.239	0.615	0.032	0.354	0.216	0.662
59	0.239	0.615	0.032	0.354	0.216	0.662
61	0.239	0.615	0.032	0.354	0.216	0.662

Table 1: Lower and upper probabilities, Example I

7.2 Example II: Progressive censoring

In this example, we illustrate the above presented NPI approach for comparison of two groups of lifetime data under several progressive censoring schemes. We use a subset of Nelson's data [32] on breakdown times (in minutes) of an insulating fluid that is subject to high voltage stress. The data are given below, 10 units per group involved in the experiment, so $n_x = n_y = 10$.
X : 0.49, 0.64, 0.82, 0.93, 1.08, 1.99, 2.06, 2.15, 2.57, 4.75
Y : 1.34, 1.49, 1.56, 2.10, 2.12, 3.83, 3.97, 5.13, 7.21, 8.71
We present the NPI lower and upper probabilities that group Y is better than group X, by comparing single next future observations from both groups, X_{11} and Y_{11}. The appropriate assumptions rc-$A_{(n)}$ are again made per group, and it is assumed that the groups are fully independent.

Suppose that progressive Type-II censoring is applied to group Y, with three units withdrawn from the experiment at the first observed breakdown time

for group Y (at $y_1 = 1.34$), and two units for this group withdrawn at the last observed breakdown time, $y_5 = 5.13$, so with $R^y = (3, 0, 0, 0, 2)$. It is also assumed that all breakdown times for the units from group X are observed. Assume that, with y^c denoting a right-censored observation at time y, the data actually observed in this case are
$X : 0.49, 0.64, 0.82, 0.93, 1.08, 1.99, 2.06, 2.15, 2.57, 4.75$
$Y : 1.34, 1.34^c, 1.34^c, 1.34^c, 1.49, 1.56, 2.12, 5.13, 5.13^c, 5.13^c$
The NPI lower and upper probabilities are $\underline{P}(Y_{11} > X_{11}) = 0.6139$ and $\overline{P}(Y_{11} > X_{11}) = 0.8052$.

Now suppose that the progressive Type-II censoring scheme is applied to both groups X and Y, with $R^x = (2, 1, 0, 1, 0, 0)$ and $R^y = (1, 2, 0, 3)$ and resulting in the following data, $X : 0.49, 0.49^c, 0.49^c, 0.64, 0.64^c, 0.93, 1.99, 1.99^c, 2.06, 4.75$
$Y : 1.34, 1.34^c, 1.49, 1.49^c, 1.49^c, 2.10, 2.12, 2.12^c, 2.12^c, 2.12^c$
These data lead to NPI lower and upper probabilities $\underline{P}(Y_{11} > X_{11}) = 0.5148$ and $\overline{P}(Y_{11} > X_{11}) = 0.8506$.

Precedence testing can be considered as a special case of progressive censoring. Suppose that the experiment is terminated as soon as the fifth breakdown from group Y is observed, i.e. at time $y_5 = 2.12$. Then the breakdown times of five units from group Y are right-censored at that time, together with three units from group X, resulting in data
$X : 0.49, 0.64, 0.82, 0.93, 1.08, 1.99, 2.06, 2.12^c, 2.12^c, 2.12^c$
$Y : 1.34, 1.49, 1.56, 2.10, 2.12, 2.12^c, 2.12^c, 2.12^c, 2.12^c, 2.12^c$
For these data, NPI gives $\underline{P}(Y_{11} > X_{11}) = 0.5289$ and $\overline{P}(Y_{11} > X_{11}) = 0.8264$. Coolen-Schrijner *et al* [20] present several results for NPI precedence testing, including the attractive fact that, if one increases the end-time of the experiment, such an NPI lower (upper) probability for comparison of two groups never decreases (increases).

7.3 Example III: Competing risks

In this example, a well-known data set from the literature [27] is used to illustrate some aspects of the NPI method for dealing with competing risks. The data contain information about 36 units of a new model of a small electrical appliance which were tested, and where the lifetime observation per unit consists of the number of completed cycles of use until the unit failed. These data are presented in Table 2, which also includes the specific failure mode (FM) that caused the unit to fail. In the study, there were 18 different ways in which an appliance could fail, so 18 failure modes, but to illustrate the NPI method we will first reduce this to two failure modes, thereafter we consider grouping into three failure modes. Three units in the test did not fail before the end of the experiment, so for these units we have right-censored observations (2565, 6367 and 13403) for all failure modes considered, indicated by '-' for the failure mode in Table 2.

# cycles	FM	# cycles	FM	# cycles	FM
11	1	1990	9	3034	9
35	15	2223	9	3034	9
49	15	2327	6	3059	6
170	6	2400	9	3112	9
329	6	2451	5	3214	9
381	6	2471	9	3478	9
708	6	2551	9	3504	9
958	10	2565	-	4329	9
1062	5	2568	9	6367	-
1167	9	2702	10	6976	9
1594	2	2761	6	7846	9
1925	9	2831	2	13403	-

Table 2: Failure data for electrical appliance test

The two most frequently occurring failure modes in these data are FM9, which caused 17 units to fail, and FM6 which caused 7 failures. We consider how likely it is that the next unit, say unit 37, would fail due to FM9, assuming it would undergo the same test and its number of completed cycles would be exchangeable with these numbers for the 36 units reported. Let us first group all failure modes other than FM9 together, and consider these jointly as a failure mode, so we consider the NPI approach with 2 failure modes, FM9 and, say, 'other failure mode' (OFM). There are still three units that do not fail and for which we only have right-censored observations (RC). The data corresponding to this definition of failure modes are presented in Table 3.

FM9	1167	1925	1990	2223	2400	2471
	2551	2568	3034	3034	3112	3214
	3478	3504	4329	6976	7846	
OFM	11	35	49	170	329	381
	708	958	1062	1594	2327	2451
	2702	2761	2831	3059		
RC	2565	6367	13403			

Table 3: Failure data for FM9, OFM and RC

In this case there are tied observations, as two units have failed due to FM9 after 3034 completed cycles. To deal with this, we assume a small difference between these values, such that their ordering does not change with regard to observations of units in other groups, so, we assume that one of these two units actually failed after 3035 completed cycles. If such a tie would occur among different groups, then one can break it similarly in two ways, different for upper and lower probabilities in such a way that these are maximal and minimal, respectively, over the possible ways of breaking such ties, without changing the order of these observations with respect to all other observations. For competing risks data, a failure time observation caused by one failure mode is simultaneously a right-censored observation for all other failure modes. This situation is dealt with in the NPI approach, as is common in many statistical approaches, by assuming that the right-censoring time is just beyond the failure time. For the three right-censored observations for units that were not observed to fail, we also have tied observations for the two failure modes considered (FM9 and OFM), so for both these right-censoring times coincide. We deal with this again by assuming that for one of the failure modes this event occurred fractionally later than for the other, and then we calculate the lower and upper probabilities for the event of interest by considering the maximum and minimum of the upper and lower probabilities, respectively, corresponding to the different possible orderings of these 'un-tied' right-censoring times.

The NPI lower and upper probabilities for the event that unit 37 will fail due to FM9 are

$$\underline{P}(X_{37}^{FM9} < X_{37}^{OFM}) = 0.4358,$$

$$\overline{P}(X_{37}^{FM9} < X_{37}^{OFM}) = 0.5804$$

while the corresponding NPI lower and upper probabilities for unit 37 to fail due to OFM are

$$\underline{P}(X_{37}^{OFM} < X_{37}^{FM9}) = 0.4196,$$

$$\overline{P}(X_{37}^{OFM} < X_{37}^{FM9}) = 0.5642$$

These lower and upper probabilities satisfy the conjugacy property as, implicit in our method, it is assumed that the experiment on unit 37 would actually continue until it fails, and this is assumed to happen with certainty. NPI can be generalized to take the possibility of 'non-failure' of the next unit by a certain time into account, but we have not developed this further. On the basis of these NPI lower and upper probabilities, one could interpret the data as containing weak evidence that the event that unit 37 will fail due to FM9 is (a bit) more likely than for it to fail due to another failure mode, with all the other failure modes grouped together as done in this case.

Let us now group the failure modes differently, by considering FM9 and FM6 separately, causing 17 and 7 units to fail, respectively. We group all the other failure modes together into OFM. The data used here are given in Table 4. The NPI lower and upper probabilities for the event that unit 37 will fail due to FM9, due to FM6 or due to OFM, are

$$\underline{P}\left(X_{37}^{FM9} < \min\left\{X_{37}^{FM6}, X_{37}^{OFM}\right\}\right) = 0.3915\ ,$$

FM9	1167	1925	1990	2223	2400	2471
	2551	2568	3034	3034	3112	3214
	3478	3504	4329	6976	7846	
FM6	170	329	381	708	2327	2761
	3059					
OFM	11	35	49	958	1062	1594
	2451	2702	2831			
RC	2565	6367	13403			

Table 4: Failure data for FM9, FM6, OFM and RC

$$\overline{P}\left(X_{37}^{FM9} < \min\left\{X_{37}^{FM6}, X_{37}^{OFM}\right\}\right) = 0.5804$$

$$\underline{P}\left(X_{37}^{FM6} < \min\left\{X_{37}^{FM9}, X_{37}^{OFM}\right\}\right) = 0.1749\ ,$$

$$\overline{P}\left(X_{37}^{FM6} < \min\left\{X_{37}^{FM9}, X_{37}^{OFM}\right\}\right) = 0.3279$$

$$\underline{P}\left(X_{37}^{OFM} < \min\left\{X_{37}^{FM6}, X_{37}^{FM9}\right\}\right) = 0.2265\ ,$$

$$\overline{P}\left(X_{37}^{OFM} < \min\left\{X_{37}^{FM6}, X_{37}^{FM9}\right\}\right) = 0.3808$$

Since

$$\underline{P}\left(X_{37}^{FM9} < \min\left\{X_{37}^{FM6}, X_{37}^{OFM}\right\}\right) >$$
$$\overline{P}\left(X_{37}^{FM6} < \min\left\{X_{37}^{FM9}, X_{37}^{OFM}\right\}\right)$$

one could interpret the data as providing strong evidence that unit 37 is more likely to fail due to FM9 than due to FM6, in this setting with all other failure modes grouped into OFM. If one adopts a subjective interpretation of lower and upper probabilities in terms of prices for desirable gambles, in line with Walley [36], then these lower and upper probabilities would imply that, for any price between 0.3279 and 0.3915, one would be willing both to buy the gamble which pays 1 if unit 37 fails due to FM9 and to sell the gamble which pays 1 if unit 37 fails due to FM6. If one has a quick look at the data, one may be surprised that FM6 is not the more likely one to lead to failure, as it has caused relatively many early failures. However, it only caused failure of 7 out of the 36 units tested, the comparisons would be different if the data were not competing risks data on the same units but failure times for independent groups without the important aspect of a failure due to one failure mode providing a right-censored observation for all other failure modes. Similarly, strong evidence that unit 37 is more likely to fail due to FM9 than due to OFM can be claimed because

$$\underline{P}\left(X_{37}^{FM9} < \min\left\{X_{37}^{FM6}, X_{37}^{OFM}\right\}\right) >$$
$$\overline{P}\left(X_{37}^{OFM} < \min\left\{X_{37}^{FM6}, X_{37}^{FM9}\right\}\right)$$

Comparison of these two cases illustrates some features that are different in statistics using lower and upper probabilities when compared to methods using precise probabilities. The lower and upper probabilities for unit 37 to fail due to FM9 are $[0.4358, 0.5804]$ in the first case, with all other failure modes grouped together, and $[0.3915, 0.5804]$ in the second case, with FM6 also taken separately. In the latter case, there is more imprecision in these upper and lower probabilities, while data are represented in more detail. This increase in imprecision, actually the fact that these upper and lower probabilities are nested with more imprecision if data are represented in more detail, is in line with a fundamental principle of NPI proposed and discussed by Coolen and Augustin [9, 10] in the context of multinomial data. This leads to the conjecture that, for such competing risks data, if more failure modes are treated separately instead of being grouped together, then lower and upper probabilities for an event that the next unit's failure is caused by a specific failure mode are nested, with imprecision increasing with the number of failure modes used. We hope to prove this conjecture in the near future.

The two NPI upper probabilities for the event that unit 37 will fail due to FM9, for the cases with all other failure modes grouped together (first case) and with FM6 separated (second case), are both equal to 0.5804. This is a consequence of the fact that this upper probability is realized with the extreme assignments of probability masses in the intervals created by the data in accordance to the lower survival function for FM9 and the upper survival function for the other failure modes. With all failure modes assumed to be independent, the upper survival function for the other failure modes combined is actually the same, whether or not FM6 is considered separately, this was discussed by Coolen *et al* [12], who presented individual NPI lower and upper survival functions and also considered the data used in this example, but they did not develop the NPI method for multiple comparisons that underlies the NPI method for competing risks presented here.

Acknowledgments

This work forms part of the research for my PhD degree, with Pauline Coolen-Schrijner as main supervisor. To my great sadness, Pauline passed away on 23 April 2008 due to her illness. She was a very supportive and nice person, who helped me very much during my study and she has a strong influence on my future career. I am grateful to you Pauline. *Tahani*

We are grateful to two anonymous referees for suggesting improvements to this paper.

References

[1] T. Augustin and F.P.A. Coolen. Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference*, 124: 251–272, 2004.

[2] N. Balakrishnan and R. Aggarwala. *Progressive Censoring : Theory, Methods, and Applications*. Birkha, 2000.

[3] N. Balakrishnan and H.K.T. Ng. *Precedence-Type Tests and Applications*. Wiley, 2006.

[4] R.E. Bechhofer, T.J. Santner and D. Goldsman. *Design and Analysis of Experiments for Statistical Selection, Screening and Multiple Comparisons*. Wiley, 1995.

[5] F.P.A. Coolen. Comparing two populations based on low stochastic structure assumptions. *Statistics and Probability Letters*, 29: 297–305, 1996.

[6] F.P.A. Coolen. On nonparametric predictive inference and objective Bayesianism. *Journal of Logic, Language and Information*, 15: 21–47, 2006.

[7] F.P.A. Coolen. On probabilistic safety assessment in case of zero failures. *Journal of Risk and Reliability*, 220: 105–114, 2006.

[8] F.P.A. Coolen. Nonparametric prediction of unobserved failure modes. *Journal of Risk and Reliability*, 221: 207–216, 2007.

[9] F.P.A. Coolen and T. Augustin. Learning from multinomial data: a nonparametric predictive alternative to the imprecise dirichlet model. In *ISIPTA'05: Proceedings of the Fourth International Symposium on Imprecise Probability Theory and Applications*, Cozman *et al* (eds), pp 125–134, 2005.

[10] F.P.A. Coolen and T. Augustin. A nonparametric predictive alternative to the imprecise dirichlet model: the case of a known number of categories. *International Journal of Approximate Reasoning*, 50: 217–230, 2009.

[11] F. P. A. Coolen and P. Coolen-Schrijner. Nonparametric predictive comparison of proportions. *Journal of Statistical Planning and Inference*, 137: 23–33, 2007.

[12] F.P.A. Coolen, P. Coolen-Schrijner and K.J. Yan. Nonparametric predictive inference in reliability. *Reliability Engineering and System Safety*, 78: 185–193, 2002.

[13] F.P.A. Coolen and L.V. Utkin. Imprecise reliability. In *Wiley Encyclopedia of Quantitative Risk Analysis and Assessment*, Melnick and Everitt (eds), pp 875–881. Wiley, 2008.

[14] F.P.A. Coolen and P. van der Laan. Imprecise predictive selection based on low structure assumptions. *Journal of Statistical Planning and Inference*, 98: 259–277, 2001.

[15] F.P.A. Coolen and K.J. Yan. Nonparametric predictive comparison of two groups of lifetime data. In *ISIPTA'03: Proceedings of the Third International Symposium on Imprecise Probabilities and their Applications*, Bernard *et al* (eds), pp 148–161, 2003.

[16] F.P.A. Coolen and K.J. Yan. Nonparametric predictive inference with right-censored data. *Journal of Statistical Planning and Inference*, 126: 25–54, 2004.

[17] P. Coolen-Schrijner and F.P.A. Coolen. Nonparametric predictive comparison of success-failure data in reliability. *Journal of Risk and Reliability*, 221: 319–327, 2007.

[18] P. Coolen-Schrijner, F.P.A. Coolen and I.M. MacPhee. Nonparametric predictive inference for systems reliability with redundancy allocation. *Journal of Risk and Reliability*, 222: 463–476, 2008.

[19] P. Coolen-Schrijner, F.P.A. Coolen and S.C. Shaw. Nonparametric adaptive opportunity-based age replacement strategies. *Journal of the Operational Research Society*, 57: 63–81, 2006.

[20] P. Coolen-Schrijner, T.A. Maturi and F.P.A. Coolen. Nonparametric predictive precedence testing for two groups. *Journal of Statistical Theory and Practice,*, to appear.

[21] P. Coolen-Schrijner, S.C. Shaw and F.P.A. Coolen. Opportunity-based age replacement with a one-cycle criterion. *Journal of the Operational Research Society*, to appear.

[22] B. De Finetti. *Theory of Probability.* Wiley, 1974.

[23] M.M. Desu and D. Raghavarao. *Nonparametric Statistical Methods for Complete and Censored Data.* Chapman and Hall, 2004.

[24] B.M. Hill. De Finetti's theorem, induction, and A_n, or Bayesian nonparametric predictive inference (with discussion). In *Bayesian Statistics 3*, Lindley *et al* (eds), pp 211–241, 1988.

[25] B.M. Hill. Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association*, 63: 677–691, 1968.

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

[27] J.F. Lawless. *Statistical Models and Methods for Lifetime Data.* Wiley, 2003.

[28] T. A. Maturi, P. Coolen-Schrijner, and F. P. A. Coolen. Nonparametric predictive selection with early experiment termination. In preparation.

[29] T.A. Maturi, P. Coolen-Schrijner and F.P.A. Coolen. Nonparametric predictive comparison of lifetime data under progressive censoring. In submission.

[30] T.A. Maturi, P. Coolen-Schrijner and F.P.A. Coolen. Nonparametric predictive comparison of lifetime data with early termination of experiments. In submission.

[31] T.A. Maturi, P. Coolen-Schrijner and F.P.A. Coolen. On nonparametric predictive inference for competing risks. In *Proceedings of the 18th Advances in Risk and Reliability Technology Symposium*, Lisa Bartless (ed), Loughborough University, pp 196–211, 2009.

[32] W. Nelson. *Applied Life Data Analysis.* Wiley, 2004.

[33] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, 1976.

[34] A. Tsiatis. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences of the United States of America*, 72: 20–22, 1975.

[35] L.V. Utkin and F.P.A. Coolen. Imprecise reliability: an introductory overview. In *Computational Intelligence in Reliability Engineering, Volume 2: New Metaheuristics, Neural and Fuzzy Techniques in Reliability*, Levitin (ed), pp 261–306. Springer, 2007.

[36] P. Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, 1991.

[37] K. Weichselberger. *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I. Intervalwahrscheinlichkeit als umfassendes Konzept.* Physika, 2001.