Frank P.A. Coolen
Durham University, Durham, UK

Nonparametric predictive inference in reliability and risk: recent developments

Keywords
competing risks, imprecise reliability, lower and upper probability, nonparametric predictive inference, system reliability, unobserved or unknown failure modes

Abstract
During the last two decades, statistical methods using lower and upper probabilities have become increasingly popular. One such method is Nonparametric Predictive Inference (NPI), which makes relatively few modelling assumptions. Due to the specific nature of many reliability and risk scenarios, NPI provides attractive new solutions to many problems in these fields. This paper provides an introductory overview to this area, including examples on competing risks, system reliability and prediction of unobserved or even unknown failure modes.

1 Introduction

Traditionally, uncertainty quantification has been done mostly by the use of precise probabilities: For each event A, a single (classical, precise) probability $P(A)$ is used, typically satisfying Kolmogorov's axioms. Whilst this has been very successful in many applications, it has long been recognized to have severe limitations. Classical probability requires a very high level of precision and consistency of information, and thus it is often too restrictive to carefully represent the multi-dimensional nature of uncertainty. Perhaps the most straightforward restriction is that the quality of underlying knowledge cannot be adequately represented using a single probability measure. An increasingly popular and successful generalization is available through the use of *lower and upper probabilities*, denoted by $\underline{P}(A)$ and $\overline{P}(A)$ respectively, with $0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1$. Walley [45] presents a detailed introduction of the theory using the term 'imprecise probability', with a subjective interpretation of lower and upper probability as maximum price one is willing to pay for a gamble that pays 1 if A occurs and 0 else, and minimum price at which one is willing to sell the gamble, respectively. Weichselberger [47, 48], using the term 'interval probability', presents a theory that generalizes Kolmogorov's axioms without explicitly requiring an interpretation. These theories, and further alternatives that have been presented, are very similar for as far as practical applications are concerned, with differences typically in smaller mathematical details. Clearly, the special case with $\underline{P}(A) = \overline{P}(A)$ for all events A provides precise probability, whilst $\underline{P}(A) = 0$ and $\overline{P}(A) = 1$ represents complete lack of knowledge about A, with a flexible continuum in between. A typical consistency property is conjugacy, that is $\underline{P}(A^c) = 1 - \overline{P}(A)$, with A^c the complementary event of A. It is useful to keep the following informal interpretation in mind: $\underline{P}(A)$ can be interpreted as reflecting the evidence in favour of event A, and $1-\overline{P}(A)$ as reflecting the evidence against A hence in favour of A^c. The lower and upper probabilities in this paper typically result from an attempt to base inference on relatively few modelling assumptions, not sufficiently strong to lead to precise probabilities for many events of interest. In this case, the lower and upper probabilities can be considered to be the maximum lower and minimum upper bounds, respectively, for all precise probabilities that are in agreement with the assumptions made and the data-based inferences following from these assumptions. For a recent short introduction to lower and upper probabilities with many references to the literature see [22][1].

It should be emphasized that the term 'imprecise probability', which has become widely accepted following Walley [45], may provide a false impression as lower and upper probabilities enable more accurate quantification of uncertainty than precise probability. In applications, clear advantages over the established theory of precise probability have been demonstrated, some of these will be discussed later in this paper. This justifies the further development of imprecise probability, particularly towards building a complete methodological framework for applications in statistics, decision support, and related fields including reliability and risk. Imprecise probability provides important new methods that promise greater flexibility for uncertainty quantification. Its advantages include the possibility to deal with conflicting evidence, to base inferences on weaker assumptions than needed for precise probabilistic methods, and to allow for simpler and more realistic elicitation of subjective information, as imprecise probability does not require experts to represent their judgements through a full probability distribution, which often does not reflect their beliefs appropriately.

Following Walley [45], many of the imprecise probability-based contributions to statistics follow a generalized Bayesian approach, using a standard precise parametric sampling model with a set of prior distributions. In particular, the use of models from the exponential family is popular in conjunction with classes of conjugate priors. Walley's Imprecise Dirichlet Model (IDM) for inference in case of multinomial data [46] has attracted particular attention, also with applications to lifetime data including right-censored observations [8] and with partial information on dependencies in failure data [42]. In these models, taking new information into account is effectively done by updating all elements of the set of prior distributions as in Bayesian statistics with precise prior distributions, leading to a set of posterior distributions which forms the basis for inferences. From technical perspective this procedure is therefore closely related to robust Bayesian inference [5], but by reporting the indeterminacy resulting from limited information, use and interpretation of the resulting imprecise posterior goes beyond a simple sensitivity and robustness analysis. There are also interesting arguments to vary the set of distributions, going from prior to posterior, when working with lower and upper probabilities. One such an approach, staying close to generalized Bayesian updating but including an extra parameter to reduce the distribution set upon updating, which allows some additional control of the imprecision, was presented in [7] and also included in an introduction of imprecise Bayesian reliability analysis [21].

[1] See also *www.sipta.org*

2 Imprecise reliability

Reliability analysis is an important application area of statistics and probability theory in engineering, with several specific features which often complicate application of standard methods. Such features include data censoring, for example due to maintenance activities, lack of knowledge and information on dependence between random quantities, for example if failures occur due to competing risks, and required use of expert judgements, for example when new or upgraded versions of units are used. While mathematical approaches for dealing with such issues have been presented within the framework of statistics using precise probabilities, the use of imprecise probability provides exciting new ways for dealing with such challenges in reliability. One of the first approaches that generalized probability in reliability was fuzzy reliability theory [6], but this suffers from vagueness about axioms and rules for combination of information, and lack of clear interpretation of the results. During the last two decades, imprecise probability has received increasing attention and interesting applications have been reported, particularly also in the reliability and risk literature. It is widely accepted that, by generalizing precise probability theory in a mathematically sound manner, with clear axioms and interpretations, this theory provides a better approach to generalized uncertainty quantification then its current alternatives.

An extensive introduction to imprecise reliability, together with a discussion of many applications, has been presented in [43], while [23] provides a concise recent overview. As an example of an imprecise reliability application, Fetz and Tonon [33] consider bounds for the probability of failure of a series system when no information is available about dependence between failure probabilities of different modes. They consider several models, including random sets and p-boxes, and they provide a detailed list of references to the literature on such topics. They also discuss some computational methods, which is an important aspect of application of imprecise reliability to medium or larger size practical problems. One of the possible ways in which output from imprecise probability methods can be useful is in the study of sensivity of model outputs with respect to variations in input parameters. An interesting recent study [40] presents such an approach to a large-scale modelling problem to assess reliability in an aerospace engineering application, comparing the use of classical probabilities and a variety of imprecise probability methods.

Imprecise probabilistic approaches to statistics are of great value to reliability problems. An interesting recent development is combination of imprecise Bayesian methods for some paramaters with a generalized maximum likelihood approach for other parameters in an inferential problem, where the former can for example be used to explicitly deal with incomplete expert judgements while the latter can be appropriate on aspects of the problem for which suitable data are available. This has been explored for software reliability growth models, using the maximum likelihood approach for the temporal growth aspect together with imprecise Bayesian methods for the parameters modelling the stationary aspects of the model [44].

3 Nonparametric predictive inference

Nonparametric predictive inference (NPI) is a statistical method based on Hill's assumption $A_{(n)}$ [34], which gives a direct conditional probability for a future observable random quantity, conditional on observed values of related random quantities [3, 11]. Suppose that $X_1, \ldots, X_n, X_{n+1}$ are continuous and exchangeable random quantities. Let the ordered observed values of $X_1, \ldots, X_n$ be denoted by $x_{(1)} < x_{(2)} < \ldots < x_{(n)} < \infty$, and let $x_{(0)} = -\infty$ and $x_{(n+1)} = \infty$ for ease of notation, to be replaced by other known bounds of support, such as $x_{(0)} = 0$, if the random quantities represent lifetimes. For a future observation X_{n+1}, based on n observations, $A_{(n)}$ [34] is, for $j = 1, \ldots, n+1$

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

$A_{(n)}$ does not assume anything else and is a post-data assumption related to finite exchangeability. Hill [35] discusses $A_{(n)}$ in detail. Inferences based on $A_{(n)}$ are predictive and nonparametric, and 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, e.g. to study effects of additional assumptions underlying other statistical methods. $A_{(n)}$ is not sufficient to derive precise probabilities for many events of interest, but it provides optimal bounds for probabilities for all events of interest involving X_{n+1}. These bounds are lower and upper probabilities in the theories of imprecise probability [45] and interval probability [48], and as such they have strong consistency properties [3]. NPI is a framework of statistical theory and methods that use these $A_{(n)}$-based lower and upper probabilities. Several variations of $A_{(n)}$ are used for different inferences. For example, NPI has been presented for Bernoulli data, multinomial data and lifetime data with right-censored observations. NPI enables inferences for $m \geq 1$ future observations, with their interdependence explicitly taken into account, and based on sequential assumptions $A_{(n)}, \ldots, A_{(n+m-1)}$. NPI provides a solution to some explicit goals formulated for objective (Bayesian) inference, which cannot be obtained when using precise probabilities [11]. NPI is exactly calibrated [37], which is a strong consistency property in frequentist statistics, and it never leads to results that are in conflict with inferences based on empirical probabilities.

NPI for Bernoulli random quantities [9] is based on a latent variable representation of Bernoulli data as real-valued outcomes of an experiment in which there is a completely unknown threshold value, such that outcomes to one side of the threshold are successes and to the other side failures. The use of $A_{(n)}$ together with lower and upper probabilities enables inference without a prior distribution on the unobservable threshold value, as is needed in Bayesian statistics where this threshold value is typically represented by a parameter. Suppose that there is a sequence of $n+m$ exchangeable Bernoulli trials, each with 'success' and 'failure' as possible outcomes, and data consisting of s successes in n trials. Let Y_1^n denote the random number of successes in trials 1 to n, then a sufficient representation of the data for NPI is $Y_1^n = s$, due to the assumed exchangeability of all trials. Let Y_{n+1}^{n+m} denote the random number of successes in trials $n+1$ to $n+m$. Let $R_t = \{r_1, \ldots, r_t\}$, with $1 \leq t \leq m+1$ and $0 \leq r_1 < r_2 < \ldots < r_t \leq m$, and, for ease of notation, define $\binom{s+r_0}{s} = 0$. Then the NPI upper probability for the event $Y_{n+1}^{n+m} \in R_t$, given data $Y_1^n = s$, for $s \in \{0, \ldots, n\}$, is

$$\overline{P}(Y_{n+1}^{n+m} \in R_t | Y_1^n = s) = \binom{n+m}{n}^{-1} \times$$

$$\sum_{j=1}^{t} \left[\binom{s+r_j}{s} - \binom{s+r_{j-1}}{s} \right] \binom{n-s+m-r_j}{n-s}$$

The corresponding NPI lower probability is derived via the conjugacy property $\underline{P}(Y_{n+1}^{n+m} \in R_t|Y_1^n = s) = 1 - \overline{P}(Y_{n+1}^{n+m} \in R_t^c|Y_1^n = s)$ where $R_t^c = \{0, 1, \ldots, m\} \backslash R_t$.

For multinomial data, a latent variable representation via segments of a probability wheel has been presented, together with a corresponding adaptation of $A_{(n)}$ [16]. For data including right-censored observations, as often occur in lifetime data analysis, NPI is based on a variation of $A_{(n)}$ which effectively uses a similar exchangeability assumption for the future lifetime of a right-censored unit at its moment of censoring [25]. This method provides an attractive predictive alternative to the well-known Kaplan-Meier estimatoe (KME) for such data.

Many applications of NPI have been presented in the literature, for a concise overview see [14][2]. These include solutions to problems in statistics, operational research, reliability and risk. In statistics, NPI provides frequentist solutions to problems which do not depend on counterfactuals, which play a role in hypothesis testing and are often criticized by opponents of frequentist statistics. An important advantage of the use of lower and upper probabilities is that one does not need to add assumptions to data which one feels are not justified. A nice example occurs in precedence testing, where experiments to compare different groups may be terminated early in order to save costs or time [32]. In such cases, the NPI lower and upper probabilities are the sharpest bounds corresponding to all possible orderings of the not-fully observed data. NPI has been applied for comparisons of multiple groups of proportions data [18], where the number m of future observations per group plays an interesting role in the inferences and where particularly interesting situations occur in case of high-reliability applications with few, if any, failures per group [28]. Effectively, if m increases the inferences tend to become more imprecise, while imprecision tends to decrease if the number of observations in the data set increases. Several recent applications of NPI in the areas of reliability and risk are illustrated next, an early discussion with some example applications of NPI in reliability was presented in [19].

4 NPI in reliability and risk

Several aspects of data as typically occur in reliability and risk applications lead to interesting inferences when lower and upper probabilities are used, where NPI particularly shows such aspects due to the limited influence of modelling assumptions. These aspects include data sets consisting of relatively few observations and often including right-censored observations, success-failure data which actually contain few or even zero failures, and the wish to draw inferences on failure modes that have not yet been observed or have not even been specified. Some of these topics are discussed in this section, mostly via illustrative examples. For details on the theoretical results and their justifications, together with further examples and more discussion of properties, we refer to the original papers in which these approaches have been presented.

4.1 System reliability

NPI for Bernoulli data has been implemented for system reliability, with test information on components of several types that are exchangeable with those in the system [1, 31, 38]. This is illustrated in the following example, see [1] for further discussion and presentation of general formulae and their derivation.

Example 1

Consider a system consisting of 2 subsystems in series configuration with components of 2 types, A and B. Subsystem 1 is a 2-out-of-4 system (so this functions if at least two of its components function), with 2 type A components and 2 type B components; subsystem 2 is a 1-out-of-4 system, also with 2 components of each type. Assume that 2 components of type A were tested, both of which functioned successfully, and also 2 components of type B were tested of which only 1 functioned successfully. The NPI lower probability for the event that this system functions successfully is equal to 0.664. Suppose that, to increase the system's reliability by increasing redundancy, extra components can be added to the system, keeping the requirements that at least 2 components function in subsystem 1 and at least 1 in subsystem 2, but adding extra components to subsystems. It is assumed that there are no cost considerations, only the number of extra components that can be added is restricted, and these extra components can be of any type and added to any of the two subsystems. Table 1 presents the optimal allocation of 1 to 11 extra components ('Extra' in the first column; m_a^1 denotes the total number of components of type A in subsystem 1, *et cetera*), in the sense of maximum resulting NPI lower probability for the event that the system functions (denoted by $\underline{P}$).

If one extra component is allowed, it is optimal to add a component of type A to subsystem

[2] see also *www.npi-statistics.com*

Extra	$(m_a^1, m_b^1, m_a^2, m_b^2)$	$\underline{P}$
1	(3,2,2,2)	0.775
2	(4,2,2,2)	0.827
3	(5,2,2,2)	0.856
4	(5,2,3,2)	0.882
5	(6,2,3,2)	0.901
6	(6,2,4,2)	0.914
7	(7,2,4,2)	0.972
8	(8,2,4,2)	0.936
9	(8,2,5,2)	0.944
10	(9,2,5,2)	0.950
11	(9,3,5,2)	0.956

Table 1: Example system reliability

1. This is fully as expected, since type A components seem to be more reliable than type B components based on the test results, and subsystem 1 has less redundancy in the original system than subsystem 2. With two further extra components allowed, both would be chosen and assigned similarly. However, if four extra components are allowed, they would still all be of type A but now one of these would be added to subsystem 2. For up to 10 extra components, it is optimal to choose them all to be of type A and added to the subsystems as presented in Table 1. However, if an eleventh extra component is allowed, a component of type B would be added to subsystem 1. This illustrates an important aspect of NPI for system reliability, namely that it takes explicitly into account that the reliabilities of components of one type in the system are statistically dependent, as a result of the limited information from the test data. Effectively, if one has the system with already 10 extra components added in the optimal manner, it has become quite a reliable system. If, however, this system does not function, it implies that the components of type A are far less reliable than had been expected following the test results. Hence, at this point it is better to add a component of type B than a further one of type A. This illustrates that diversity in redundancy allocation can result directly from maximisation of reliability, and is due to the limited knowledge about the reliability of the components of different types.

4.2 System survival time

Recently, NPI has been presented for survival time of coherent systems consisting of exchangeable components [2], introducing the combination of system signatures and lower and upper probabilities. System signatures are a powerful tool for qualifying reliability of such systems and can be used to quantify aspects of system reliability such as its failure time distribution [41]. Consider a system consisting of m exchangeable components, it could be said that such components are all 'of the same type'. Let the random failure time of the system be T_S, and let $T_{j:m}$ be the j-th order statistic of the m random component failure times for $j = 1, \ldots, m$, with $T_{1:m} \leq T_{2:m} \leq \ldots \leq T_{m:m}$. The system's signature is defined to be the m-vector q with j-th component

$$q_j = P(T_S = T_{j:m})$$

so q_j is the probability that the system failure occurs at the moment of the j-th component failure. Assume that $\sum_{j=1}^{m} q_j = 1$, so the system functions if all components function, has failed if all components have failed, and system failure can only occur at times of component failures. The survival function of the system failure time can be derived by

$$P(T_S > t) = \sum_{j=1}^{m} q_j P(T_{j:m} > t) \tag{1}$$

The NPI lower and upper survival functions for systems with exchangeable components are derived by generalizing expression (1) to lower and upper probabilities [2]. Suppose that in a test of n components, exchangeable with those in the system considered, the observed failure times were $t_1 < t_2 < \ldots < t_n$. For ease of notation, define $t_0 = 0$ and $t_{n+1} = \infty$. These n observations partition the non-negative real-line into $n+1$ intervals $I_i = (t_{i-1}, t_i)$ for $i = 1, \ldots, n+1$. Consider reliability of a system with m components, so interest is in the m failure times of those components, say $T_1, \ldots, T_m$. The test data and $T_1, \ldots, T_m$ are linked via repeated use of the assumption $A_{(n)}$ [3, 11, 20]. Let $S_j = \#\{T_l \in I_i,\ l = 1, \ldots, m\}$, then

$$P(\bigcap_{j=1}^{n+1} \{S_j = s_j\}) = \binom{n+m}{n}^{-1}$$

for all $(s_1, \ldots, s_{n+1})$ with s_j non-negative integers and $\sum_{j=1}^{n+1} s_j = m$. For any event involving the m future observations, the number of such orderings for which this event holds can be counted. The NPI lower probability for the event of interest is derived by counting all orderings for which this event has to hold, while the corresponding upper probability is derived by counting all orderings for which this event can hold [20]. The order statistics of the m future observations $T_1, \ldots, T_m$ are the $T_{1:m} \leq T_{2:m} \leq \ldots \leq T_{m:m}$ as introduced before. The following probabilities for $T_{j:m}$, for $j = 1, \ldots, m$, are derived by counting the relevant orderings, and hold for $i = 1, \ldots, n+1$,

$$P(T_{j:m} \in I_i) =$$

$$\binom{i+j-2}{i-1}\binom{n-i+1+m-j}{n-i+1}\binom{n+m}{n}^{-1} \quad (2)$$

For this event $T_{j:m} \in I_i$ NPI provides a precise probability, as each of the $\binom{n+m}{n}$ equally likely orderings of n test observations and m future observations has the j-th ordered future observation in precisely one interval I_i. The probabilities (2) straightforwardly lead to the following NPI lower and upper survival functions for $T_{j:m}$, these are the sharpest bounds for the probability of the event $T_{j:m} > t$ that can be justified without further assumptions. The NPI lower survival function for $T_{j:m}$ is, for $t \in (t_{i-1}, t_i]$,

$$\underline{S}_{T_{j:m}}(t) = \underline{P}(T_{j:m} > t) = \sum_{l=i+1}^{n+1} P(T_{j:m} \in I_l)$$

and the corresponding NPI upper survival function is, for $t \in [t_{i-1}, t_i)$,

$$\overline{S}_{T_{j:m}}(t) = \overline{P}(T_{j:m} > t) = \sum_{l=i}^{n+1} P(T_{j:m} \in I_l)$$

At observed failure times t_i there is no imprecision in the NPI lower and upper survival functions, that is $\underline{S}_{T_{j:m}}(t_i) = \overline{S}_{T_{j:m}}(t_i)$ for $i = 1, \ldots, n$, while $\underline{S}_{T_{j:m}}(0) = \overline{S}_{T_{j:m}}(0) = 1$. Beyond the largest observed component failure time in the test, the NPI lower survival function is equal to zero but the NPI upper survival function remains positive,

$$\underline{S}_{T_{j:m}}(t) = 0$$

$$\overline{S}_{T_{j:m}}(t) = \prod_{l=j}^{m} \frac{l}{n+l} > 0 \quad \text{for } t > t_n$$

This reflects that there is no evidence in favour of such components, and hence the system, surviving past time t_n (this is reflected by the lower survival function being equal to zero), but the evidence against this is limited as there are only n observations thus far (this is reflected by the upper survival function being a positive decreasing function of n).

Combining NPI and system signatures, the NPI lower and upper survival functions for the failure time T_S of a coherent system consisting of m exchangeable components, with the system structure represented by signature q, are

$$\underline{S}_{T_S}(t) = \sum_{j=1}^{m} q_j \underline{P}(T_{j:m} > t) \quad (3)$$

$$\overline{S}_{T_S}(t) = \sum_{j=1}^{m} q_j \overline{P}(T_{j:m} > t) \quad (4)$$

It should be emphasized that this generalization of (1) is non-trivial, as the right-hand sides of (3) and (4) both involve m optima involving related random quantities and the equalities in (3) and (4) only hold if all the optima can be attained simultaneously; this is the case in NPI [2] but not generally in imprecise probability.

Example 2.
There are six coherent systems with $m = 4$ exchangeable components. Suppose that $n = 4$ components exchangeable with those in such a system were tested, leading to ordered failure times $t_1 < t_2 < t_3 < t_4$, which create the partition $I_1, \ldots, I_5$ of the positive real-line. Table 4.2 presents the NPI lower and upper survival functions for T_S, from (3) and (4), for each of these six systems, with the signatures as given.

q	$(1,0,0,0)$		$(0,0,0,1)$		$(0,\frac{1}{3},\frac{2}{3},0)$	
i	$\underline{S}_{T_S}$	$\overline{S}_{T_S}$	$\underline{S}_{T_S}$	$\overline{S}_{T_S}$	$\underline{S}_{T_S}$	$\overline{S}_{T_S}$
1	0.50	1	0.99	1	0.88	1
2	0.21	0.50	0.93	0.99	0.67	0.88
3	0.07	0.21	0.79	0.93	0.41	0.67
4	0.01	0.07	0.50	0.79	0.17	0.41
5	0	0.01	0	0.50	0	0.17
q	$(\frac{1}{4},\frac{1}{4},\frac{1}{2},0)$		$(0,\frac{2}{3},\frac{1}{3},0)$		$(0,\frac{1}{2},\frac{1}{4},\frac{1}{4})$	
i	$\underline{S}_{T_S}$	$\overline{S}_{T_S}$	$\underline{S}_{T_S}$	$\overline{S}_{T_S}$	$\underline{S}_{T_S}$	$\overline{S}_{T_S}$
1	0.79	1	0.83	1	0.87	1
2	0.56	0.79	0.59	0.83	0.67	0.87
3	0.33	0.56	0.33	0.59	0.44	0.67
4	0.13	0.33	0.12	0.33	0.21	0.44
5	0	0.13	0	0.12	0	0.21

Table 2: $\underline{S}_{T_S}(t)$ and $\overline{S}_{T_S}(t)$ for $t \in I_i$

These results illustrate the NPI lower and upper survival functions, and can be used to compare the reliabilities of these six systems, with of course the parallel system with signature $q = (0,0,0,1)$ being the most reliable and the series system with signature $q = (1,0,0,0)$ the least reliable. The NPI approach combined with signatures can also be used to compare failure times of different systems, as presented in [2].

4.3 Competing risks

In reliability and survival analysis, data often contain right-censored observations, for example due to failures caused by different failure modes, also called 'competing risks'. The NPI approach for competing risks has been presented recently [26, 39], it requires the introduction of the NPI lower and upper survival function in case of data containing right-censored observations [25, 39]. These are based on the assumption rc-$A_{(n)}$ [25], which effectively applies $A_{(n)}$ to the unobserved residual time beyond the moment of right-censoring for a censored unit. Suppose that there are n observations consisting of u failure times, $x_1 < x_2 < \ldots < x_u$, and $n-u$ right-censored observations, $c_1 < c_2 < \ldots < c_{n-u}$. Let $x_0 = 0$ and $x_{u+1} = \infty$. Suppose further that there are s_i right-censored observations in the interval (x_i, x_{i+1}), denoted by $c_1^i < c_2^i < \ldots < c_{s_i}^i$, so $\sum_{i=0}^{u} s_i = n-u$. Let $t_0^i = x_i$ for $i = 0, u$, $t_a^i = c_a^i$ for $a = 1, \ldots, s_i$, and $t_{s_i+1}^i = t_0^{i+1} = x_{i+1}$ for $i = 0, 1, \ldots, u-1$. The NPI lower and upper survival functions for the failure time X_{n+1} of the next unit are denoted by $\underline{S}(t)$ and $\overline{S}(t)$, respectively, and are as follows [19, 25, 39]. For $t \in [t_a^i, t_{a+1}^i)$ with $i = 0, 1, \ldots, u$ and $a = 0, 1, \ldots, s_i$,

$$\underline{S}(t) = \frac{1}{n+1}\, \tilde{n}_{t_a^i} \prod_{\{r: c_r < t_a^i\}} \frac{\tilde{n}_{c_r} + 1}{\tilde{n}_{c_r}}$$

and for $t \in [x_i, x_{i+1})$ with $i = 0, 1, \ldots, u$,

$$\overline{S}(t) = \frac{1}{n+1}\, \tilde{n}_{x_i} \prod_{\{r: c_r < x_i\}} \frac{\tilde{n}_{c_r} + 1}{\tilde{n}_{c_r}}$$

It should be mentioned that these NPI lower and upper survival functions always bound the Kaplan-Meier estimate, which is the well-known nonparametric maximum likelihood estimate based on such data. The NPI upper survival function only decreases at observed failure times but the NPI lower survival function also decreases at right-censoring times, by a smaller step than at observed failure times. This is in line with intuition, as the lower survival function reflects the evidence in favour of surviving, which is reduced by an observed failure but also by a right-censoring, as beyond such a time the evidence on survival comes from fewer units. The upper survival function reflects the evidence against surviving, which is clearly stronger from the moment of an observed failure time, but is not affected by a right-censoring as this holds no evidence in favour of the failure event. Beyond the largest observation, both if this is a failure time or a right-censored time, the lower survival function is equal to zero, reflecting that there is no evindence in the data that supports surviving past later times, but the upper survival function remains positive which reflects that larger observations cannot be excluded on the basis of the data alone.

In a situation with competing risks [26, 39], which are assumed to be independent, NPI lower and upper survival functions are derived conditional on each failure mode, where failures due to other failure modes are right-censored observations. The assumed independence of the competing risks allow the overall NPI lower and upper survival functions for the units considered, affected by all competing risks, to be calculated by taking the products of all the corresponding NPI lower and upper survival functions conditioned on the failure modes. A question particularly considered in [39] is which failure mode will cause the next unit to fail, which is illustrated in the following example [26]. This also shows an important advantage of the NPI approach, namely it can include unobserved and even (any number of) unknown failure modes in the analysis.

Example 3.
The data in Table 3 are a subset of a well-known data set from the literature [36] that was also used in [26, 39]. Twelve units of a new model of a small electrical appliance were tested, the lifetime observation per unit consists of the number of completed cycles of use until the unit failed. There were 18 possible failure modes identified, the specific failure mode (FM) that caused the unit to fail is given in the table.

# cycles	FM	# cycles	FM	# cycles	FM
381	6	1594	2	2471	9
708	6	1925	9	2702	10
958	10	2327	6	2831	2
1167	9	2451	5	3112	9

Table 3: Failure data for electrical appliances

The two most frequently occurring failure modes in these data are FM9 and FM6, which caused 4 and 3 units to fail, respectively. The event of interest is that the next unit, say unit 13, will fail due to a specific failure mode, assuming it would undergo the same test and its number of completed cycles would be exchangeable with these numbers for the 12 units reported.

Six cases are presented, (A) to (F), each involving a different grouping of the failure modes in 2, 3 or 4 re-defined failure modes, the corresponding NPI lower and upper probabilities are presented in Table 4. The term OFM is introduced for 'other failure modes' grouped into one single re-defined failure mode. For example in case (A), OFM is considered to be a single new failure mode containing all originally defined failure modes except failure mode 9. For each case the sums of the lower probabilities and of the upper probabilities are also given, these illustrate that they are constant for a fixed number of re-defined failure modes. For further discussion of properties of these inferences we refer to [26], where also the corresponding NPI lower and upper survival functions are investigated and illustrated.

	$\underline{P}$	$\overline{P}$		$\underline{P}$	$\overline{P}$
(A)			(B)		
9	0.237	0.481	6	0.208	0.452
OFM	0.519	0.764	OFM	0.548	0.793
	0.755	1.245		0.755	1.245
(C)			(D)		
9	0.190	0.481	9	0.190	0.481
6	0.187	0.452	3	0	0.245
OFM	0.255	0.556	OFM	0.442	0.764
	0.632	1.489		0.632	1.489
(E)			(F)		
9	0.157	0.481	3	0	0.245
6	0.168	0.452	4	0	0.245
3	0	0.245	7	0	0.245
OFM	0.212	0.556	OFM	0.536	1
	0.536	1.734		0.536	1.734

Table 4: Results Example 3

In case (A) all failure modes except FM9 are together re-defined as one failure mode (OFM). Suppose that there may be an additional unknown failure mode that may cause the units to fail, but this has not happened for the tested units. To illustrate the effect of such a possible further failure mode, consider case (D) where FM3 has been defined although it had not led to a failure for the tested units. Including such an unobserved failure mode has led to decrease of the lower probabilities for the event that unit 13 will fail due to FM9 or due to OFM, which follows by comparison of cases (A) and (D) and can also be seen in cases (C) and (E), while the corresponding upper probabilities have remained the same. This reflects that the unknown failure mode could possibly lead to the next failure, which would make the other failure modes less likely causes for it, but it cannot be excluded that this unknown failure mode may not have any effect at all, as reflected by those unchanged upper probabilities. Case (F) shows that multiple unobserved failure modes can be considered, each of these has lower probability 0 and the same positive upper probability for the event that it causes unit 13 to fail. Any number of such unobserved failure modes could be included, they would all have the same lower and upper probabilities. This uses the sub-additivity of upper probabilities to great advantage [45, 47, 48]. The upper probability for OFM in Case (F) is one, reflecting that all units observed failed due to one of the failure modes in OFM.

4.4 **Unobserved failure modes**

While NPI for competing risks [26, 39] enables inference on unobserved or even unknown failure modes, it does not quantify how likely such failure modes are to occur. The NPI approach has also been presented for multinomial data [15, 16], which explicitly enables inference on the event that the next observation will fall into either a defined but not yet observed category or even into an undefined category, which can be interpreted as an 'unknown unknown' event. This approach was presented explicitly for prediction of unobserved failure modes in [13]. Typically, if outcome categories have not occurred yet, the NPI lower probability of the next observation falling in such a category is zero, but the corresponding NPI upper probability is positive and depends on whether or not the category is explicitly defined, on the total number of categories or whether this number is unknown, and on the number of categories observed so far. Such NPI upper probabilities can be used to support cautious decisions, which are often deemed attractive in reliability and risk analysis. The following example [13] illustrates this approach.

Example 4.
Suppose that a database contains detailed information on failures experienced during warranty periods of a particular product. Currently 200 failures have been recorded, with 5 different failure modes specified. The producer is interested in the event that the next recorded failure of such a product during its warranty period is caused by another failure mode than these 5 already recorded. First assume that there is no clear assumption or knowledge about the number of possible failure modes. Suppose that interest is in the event that the next reported failure is caused by any as yet unseen failure mode, represented by $X_{201} \in UN$, then the NPI-based upper probabil-

ity for this event is equal to 5/200. If, however, the producer has actually specified two further possible failure modes, which have not yet been recorded so far, and interest is in the event that the next failure mode will be one of these two, then this method gives a different answer. In this case, let these two failure modes be denoted by DN_1 and DN_2, then the NPI-based upper probability for the event $X_{201} \in DN_1 \cup DN_2$ is equal to 2/200. The NPI-based lower probabilities for both these events are 0, reflecting that there is no actual evidence in the data for these events to occur.

Now suppose that these 200 failures were instead caused by 25 different failure modes, not including the defined DN_1 and DN_2. Then the upper probability for $X_{201} \in UN$ changes to 25/200, but the upper probability for $X_{201} \in DN_1 \cup DN_2$ remains 2/200. It is in line with intuition that the changed data affect this first upper probability, as the fact that more failure modes have been recorded suggests that there may be more different failure modes that can cause this product to fail (no knowledge is assumed on the number of possible failure modes). For the second event considered here, the reasoning is somewhat different, as effectively interest is in two specific, as yet unseen, failure modes, and there is no actual difference in the data available that is naturally suggesting that either of these two failure modes has become more likely, as in both cases there is actually no real evidence that they can lead to failures during the warranty period.

Suppose now that the producer is absolutely certain that at most 40 different failure modes exist for this product. This would not affect the above upper probabilities in case so far only 5 failure modes had been observed, but in the case of 25 observed failure modes, it reduces the upper probability for the event $X_{201} \in UN$ to 15/200, reflecting that there are only 15 possible failure modes not yet recorded. Assuming that DN_1 and DN_2 are among those 15 failure modes, the upper probability for $X_{201} \in DN_1 \cup DN_2$ remains 2/200. Clearly, if there had been 40 well defined failure modes, of which 25 had already caused failures and with no other failure modes possible, then the 15 which had not yet been recorded could be denoted by DN_i for $i = 1, \dots, 15$, in which case

$$\overline{P}(X_{201} \in UN) = \overline{P}(X_{201} \in \bigcup_{i=1}^{15} DN_i) = \frac{15}{200}$$

4.5 Further applications

NPI provides an attractive framework for decision support in a wide range of problems where the focus is naturally on a future observation. There are several further examples of applications of NPI in reliability and risk that are of interest in themselves, for the further applications that can be considered using similar approaches, and for the related research opportunities. Powerful methods for replacement decisions of technical units have been presented which are fully adaptive to process data and in simulation studies have shown to perform very well [27, 28, 30]. As NPI is particularly attractive in situations with zero failures, its applications for reliability demonstration [17] and probabilistic safety assessment [12] provide interesting alternatives to classical approaches. It is worth noting as well that NPI can also deal with grouped data [24] as often occur in real-world applications, for example if lifetime and failure data are collected only on a monthly basis for non-critical hardware.

5 Concluding remarks

Imprecise probability and its applications in statistical inference and decision support offer a wide range of research challenges. On foundations, key aspects such as updating have not yet been fully explored, and different approaches have different conditioning rules. The relation between imprecision and information requires further study, and many of the most frequently used statistical methods (such as complex regression models) have not yet been fully generalized to deal with imprecise probability. In cases where generalizations are easily found, it may be unclear which of many possible approaches is most suitable. Of course, early developments of new theoretic frameworks tend to include illustrative applications to mostly text-book style problems. The next stage required towards widely applicable methods involves upscaling, where in particular computational aspects provide many challenges. Even methods such as simulation, mostly straightforward with precise probabilities, become non-trivial with lower and upper probabilities.

For applications which require the use of subjective information, elicitation of expert judgements is less demanding when lower and upper probabilities are used, but while practical aspects of elicitation have been widely studied this has, thus far, only included very few studies involving lower and upper probabilities. In decision making, algorithms to find optimal solutions need to be developed further and implemented for large-scale applications. As many problems have a sequential nature, ways of representing sequential solutions efficiently also need to be developed, the more so as classical techniques such as backward induction and dynamic programming often cannot be extended directly. All relevant aspects of theory and application of imprecise probabilities will be addressed in an introductory book that is under development [4].

Imprecise reliability is a relatively new area of research, with methods presented that are inspired by practical problems but that are not yet suitable for applications of substantial size. The main challenges are in upscaling the methods to become useful for practical problems, together with aspects of implementation which include consideration of elicitation, model choice and computation.

The models for imprecise reliability that have been presented so far are still pretty basic. Generally, imprecise approaches can be found that generalize the established methods in varying ways, so in addition to developing new methods one must find ways to decide on how useful they are, which requires careful consideration of fundamental aspects of uncertainty and information. Hybrid methods, which combine imprecise models where useful to model indeterminacy with precise models where possible due to sufficient data or information, provide exciting opportunities for research, with issues that must be addressed including interpretation of results and choice of models and methods.

Development of NPI is gathering momentum, inferential problems for which NPI solutions have recently been presented or are being developed include aspects of medical diagnosis with the use of ROC curves, robust classification, quality control and acceptance sampling. Main research challenges for NPI include its generalization for multidimensional data, which is similarly challenging for NPI as for general nonparametric methods due to the lack of a natural ordering of the data. NPI theory and methods that enable information from covariates to be taken into account also provide interesting and challenging research opportunities. A research monograph introducing NPI theory, methods and applications is currently in development, further information is available from www.npi-statistics.com.

Acknowledgements

This paper is substantially based on articles [22, 23] jointly written with Thomas Augustin (Munich), Matthias Troffaes (Durham) and Lev Utkin (St Petersburg). It is further based on the article [14] and on journal publications with co-authors as refered to in the main text. Their contributions to these papers, and more generally to the research projects reported on in this paper, is greatly valued and has been essential to achieve the results presented in this brief overview.

References

[1] Aboalkhair, A.M., Coolen, F.P.A., MacPhee, I.M. (2011). Nonparametric predictive inference for system reliability with subsystems consisting of multiple component types. *Proceedings 19th Advances in Reliability and Risk Technology Symposium*, Stratford-upon-Avon (UK), April 2011, to appear.

[2] Al-Nefaiee, A., Coolen, F.P.A. (2011). Nonparametric predictive inference for failure times of systems consisting of exchangeable components. *Proceedings 19th Advances in Reliability and Risk Technology Symposium*, Stratford-upon-Avon (UK), April 2011, to appear.

[3] Augustin, T., Coolen, F.P.A. (2004). Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference* **124**, 251-272.

[4] Augustin, T, Coolen, F.P.A., de Cooman, G., Troffaes, M.C.M. (2011). *An Introduction to Imprecise Probabilities.* Wiley, to appear (expected early 2012).

[5] Berger, J.O. (1990). Robust Bayesian analysis: sensitivity to the prior. *Journal of Statistical Planning and Inference* **25**, 303-328.

[6] Cai, K.Y. (1996). *Introduction to Fuzzy Reliability.* Kluwer Academic Publishers.

[7] Coolen, F.P.A. (1993). Imprecise conjugate prior densities for the one-parameter exponential family of distributions. *Statistics & Probability Letters* **16**, 337-342

[8] 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.

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

[10] Coolen, F.P.A. (2004). On the use of imprecise probabilities in reliability. *Quality and Reliability Engineering International* **20**, 193-202.

[11] Coolen, F.P.A. (2006). On nonparametric predictive inference and objective Bayesianism. *Journal of Logic, Language and Information* **15**, 21-47.

[12] Coolen, F.P.A. (2006). On probabilistic safety assessment in case of zero failures. *Journal of Risk and Reliability* **220**, 105-114.

[13] Coolen, F.P.A. (2007). Nonparametric prediction of unobserved failure modes. *Journal of Risk and Reliability* **221**, 207-216.

[14] Coolen, F.P.A. (2011). Nonparametric predictive inference. In: *International Encyclopedia of Statistical Science*, M. Lovric (Ed.). Springer, pp. 968-970.

[15] Coolen, F.P.A., Augustin, T. (2005). Learning from multinomial data: a nonparametric predictive alternative to the Imprecise Dirichlet Model. In: *Proceedings ISIPTA'05* (www.sipta.org/isipta05), pp. 125-134.

[16] Coolen, F.P.A., Augustin, T. (2009). 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.

[17] Coolen, F.P.A., Coolen-Schrijner, P. (2005). Nonparametric predictive reliability demonstration for failure-free periods. *IMA Journal of Management Mathematics* **16**, 1-11.

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

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

[20] Coolen, F.P.A., Maturi T. (2010). Nonparametric predictive inference for order statistics of future observations. In: *Combining Soft Computing and Statistical Methods in Data Analysis*, C. Borgelt, *et al* (Eds.). Springer, pp. 97-104.

[21] Coolen, F.P.A., Newby, M.J. (1994). Bayesian reliability analysis with imprecise prior probabilities. *Reliability Engineering and System Safety* **43**, 75-85.

[22] Coolen, F.P.A., Troffaes, M.C., Augustin, T. (2011). Imprecise probability. In: *International Encyclopedia of Statistical Science*, M. Lovric (Ed.). Springer, 645-648.

[23] Coolen, F.P.A., Utkin, L.V. (2011). Imprecise reliability. In: *International Encyclopedia of Statistical Science*, M. Lovric (Ed.). Springer, 649-650.

[24] Coolen, F.P.A., Yan, K.J. (2003). Nonparametric predictive inference for grouped lifetime data[3]. *Reliability Engineering and System Safety* **80**, 243-252.

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

[26] Coolen-Maturi, T., Coolen, F.P.A. (2011). Unobserved, re-defined, unknown or removed failure modes in competing risks. *Journal of Risk and Reliability*, to appear.

[27] Coolen-Schrijner, P., Coolen, F.P.A. (2004). Adaptive age replacement based on nonparametric predictive inference. *Journal of the Operational Research Society* **55**, 1281-1297.

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

[29] Coolen-Schrijner, P., Coolen, F.P.A. (2007). Nonparametric adaptive age replacement with a one-cycle criterion. *Reliability Engineering and System Safety* **92**, 74-84.

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

[31] Coolen-Schrijner, P., Coolen, F.P.A., MacPhee, I.M. (2008). Nonparametric

[3]This version contains many text setting errors due to publisher's mistakes at the final stage of the publishing process. It is better to use the material as presented in the PhD thesis by K.J. Yan, available from *www.npi-statistics.com*

predictive inference for systems reliability with redundancy allocation. *Journal of Risk and Reliability* **222**, 463-476.

[32] Coolen-Schrijner P., Maturi, T.A., Coolen, F.P.A. (2009). Nonparametric predictive precedence testing for two groups. *Journal of Statistical Theory and Practice* **3**, 273-287.

[33] Fetz, T., Tonon, F. (2008). Probability bounds for series systems with variables constrained by sets of probability measures. *International Journal of Reliability and Safety* **2**, 309-339.

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

[35] Hill, B.M. (1988). De Finetti's theorem, induction, and $A_{(n)}$ or Bayesian nonparametric predictive inference (with discussion). In: *Bayesian Statistics 3*, J.M. Bernardo, *et al.* (Eds.). Oxford University Press, pp. 211-241.

[36] Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data* (2nd ed.). Wiley.

[37] Lawless, J.F., Fredette, M. (2005). Frequentist prediction intervals and predictive distributions. *Biometrika* **92**, 529-542.

[38] MacPhee, I.M., Coolen, F.P.A., Aboalkhair, A.M. (2009). Nonparametric predictive system reliability with redundancy allocation following component testing. *Journal of Risk and Reliability* **223**, 181-188.

[39] Maturi, T.A., Coolen-Schrijner, P., Coolen, F.P.A. (2010). Nonparametric predictive inference for competing risks. *Journal of Risk and Reliability* **224**, 11-26.

[40] Oberguggenberger, M., King, J., Schmelzer, B. (2009). Classical and imprecise probability methods for sensitivity analysis in engineering: a case study. *International Journal of Approximate Reasoning* **50**, 680-693.

[41] Samaniego, F. (2007). *System Signatures and their Applications in Engineering Reliability.* Springer.

[42] Troffaes, M.C.M., Coolen, F.P.A. (2009). Applying the Imprecise Dirichlet Model in cases with partial observations and dependencies in failure data. *International Journal of Approximate Reasoning* **50**, 257-268.

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

[44] Utkin, L.V., Zatenko, S.I., Coolen, F.P.A. (2009). Combining imprecise Bayesian and maximum likelihood estimation for reliability growth models. In: *Proceedings ISIPTA'09* (www.sipta.org/isipta09), pp. 421-430.

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

[46] 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.

[47] Weichselberger, K. (2000). The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning* **24**, 149-170.

[48] Weichselberger, K. (2001). *Elementare Grundbegriffe einer Allgemeineren Wahrscheinlichkeitsrechnung I. Intervallwahrscheinlichkeit als Umfassendes Konzept.* Physika.