

Predictive inference for system reliability after common-cause component failures

Frank P.A. Coolen^{1a}, Tahani Coolen-Maturi²

¹*Department of Mathematical Sciences, Durham University, UK*

²*Durham University Business School, Durham University, UK*

Abstract

This paper presents nonparametric predictive inference for system reliability following common-cause failures of components. It is assumed that a single failure event may lead to simultaneous failure of multiple components. Data consist of frequencies of such events involving particular numbers of components. These data are used to predict the number of components that will fail at the next failure event. The effect of failure of one or more components on the system reliability is taken into account through the system's survival signature. The predictive performance of the approach, in which uncertainty is quantified using lower and upper probabilities, is analysed with the use of ROC curves. While this approach is presented for a basic scenario of a system consisting of only a single type of components and without consideration of failure behaviour over time, it provides many opportunities for more general modelling and inference, these are briefly discussed together with the related research challenges.

Keywords: Common-cause failures, lower and upper probabilities, nonparametric predictive inference, ROC curves, survival signature, system reliability

1. INTRODUCTION

A major consideration for reliability of systems is the possible occurrence of common-cause failures, where multiple components fail simultaneously due to the same underlying cause. This paper considers the reliability of a system

^aCorresponding author: frank.coolen@durham.ac.uk

following a future failure event, with possible common-cause failures of multiple components. In particular, the aim is to develop predictive inference for the reliability of a system with multiple components based on previous failure event data for this system, or for other systems that are fully exchangeable with this system. These data are assumed to consist of the numbers of failing components in past failure events, in each of which at least one component failed. These failure events did not necessarily involve failure of the whole system, but each event involves failure of one or more components. Aspects of ageing are not taken into account, nor any other aspects that are explicitly related to time, usage or other processes. Such aspects are likely to be important in some practical applications, developing methodology to deal with these provides interesting research challenges.

Common-cause failures are important in many applications, as systems with built in redundancy are at increased risk if multiple components may fail simultaneously. Basic concepts of modelling common-cause failures are reviewed by Rasmuson and Kelly [1] and by Mosleh et al [2]. The alpha-factor model, proposed by Mosleh et al [3] is commonly used, and enables straightforward analysis in the framework of Bayesian statistics due to the availability of conjugate prior distributions [4]. A robust Bayesian approach to this model has recently been proposed by Troffaes et al [5], using a set of conjugate priors instead of a single prior distribution, following a standard approach for generalized Bayesian methods in theory of imprecise probabilities [6]. In this paper an alternative statistical method within the theory of imprecise probability is used, namely nonparametric predictive inference (NPI) [7]. For the problem considered here, it makes little difference which specific method for statistical inference is used to predict the number of components failing simultaneously at the next common-cause failure event. Due to its explicitly predictive nature it is attractive to use the NPI approach. The main contribution of the paper is the link from inference on this number of failing components to lower and upper probabilities for the event that the system will still function after the next common-cause failure event. This link is quite straightforward, as shown in this paper, with the use of the system survival signature [8], a recently introduced concept which is closely related to the popular system signature [9] but, in contrast to the latter, is also conceptually straightforward for systems with multiple types of components. The question of the system's reliability following a common-cause failure event is of practical interest, as following such an event one may have more time to consider appropriate maintenance or replacement actions if the

system is still functioning than in case the system has failed. Such further actions are not addressed in this paper, but combined with aspects of failure processes they provide interesting topics for future research.

As mentioned, the main contribution of this paper is to model the link between failure of components and failure of the system, or indeed functioning of the system, for the common-cause failure scenario, through the recently introduced concept of survival signatures [8]. For a system with m components, the state vector $\underline{x} = (x_1, x_2, \dots, x_m) \in \{0, 1\}^m$ is defined such that $x_i = 1$ if the i th component functions and $x_i = 0$ if not, using an arbitrary but fixed labelling of the components. The structure function $\phi : \{0, 1\}^m \rightarrow \{0, 1\}$, defined for all possible \underline{x} , takes the value 1 if the system functions and 0 if the system does not function for the state vector \underline{x} . The structure function can be generalized by defining it as the probability that the system functions for the state vector \underline{x} , which might be relevant if one is not certain about its functioning [10], this is not used in this paper but provides interesting challenges for research and opportunities for applications.

In this paper, attention is restricted to coherent systems, which means that $\phi(\underline{x})$ is not decreasing in any of the components of \underline{x} , so system functioning can never be improved by worse performance of one or more of its components. It is further assumed that $\phi(\underline{0}) = 0$ and $\phi(\underline{1}) = 1$, so the system fails if all its components fail and it functions if all its components function. These assumptions could easily be relaxed but they simplify presentation in this paper and are reasonable for most systems of practical interest. More importantly, attention is restricted in this paper to systems consisting of exchangeable components, which could be called components of a single type. Most practical systems, in particular also networks, consist of components of multiple types. This restriction is merely for sake of simplicity of presentation, the method introduced here can quite straightforwardly be generalized to systems with multiple types of components.

The survival signature for a system with m exchangeable components, denoted by $\Phi(l)$, for $l = 1, \dots, m$, is the probability that the system functions given that *precisely* l of its components function [8]. For coherent systems, $\Phi(l)$ is an increasing function of l , and the second assumption above leads to $\Phi(0) = 0$ and $\Phi(m) = 1$. There are $\binom{m}{l}$ state vectors \underline{x} with precisely l components $x_i = 1$, so with $\sum_{i=1}^m x_i = l$; let S_l denote the set of these state vectors. Due to the exchangeability assumption for the m components, or more precisely the assumed exchangeability of the random quantities representing functioning of the m components, all these state vectors are equally

likely to occur, hence

$$\Phi(l) = \binom{m}{l}^{-1} \sum_{\underline{x} \in S_l} \phi(\underline{x}) \quad (1)$$

Coolen and Coolen-Maturi [8] called $\Phi(l)$ the survival signature because, by its definition, it is closely related to survival of the system, and it is close in nature to the system signature [9]. The survival signature can straightforwardly be generalized to systems with multiple types of components, in contrast to the system signatures for which this is practically impossible [8].

This paper is organised as follows. Section 2 presents the use of non-parametric predictive inference to predict the number of failing components at a future common-cause failure event. In Section 3 these inferences are combined with the survival signature to lead to lower and upper probabilities for the event that the system will still function after the next failure event. As with any newly proposed procedure, it is important to evaluate the performance of the presented method, this is non-trivial for methods using lower and upper probabilities. In Section 4 a novel way to evaluate the performance of such an imprecise probability method is introduced, which is fully in line with the predictive nature of the inferences and makes use of ROC curves. Section 5 concludes the paper with a discussion of the method and related research challenges.

2. PREDICTING THE NUMBER OF FAILING COMPONENTS

For a system with m exchangeable components, which are assumed throughout this paper to all function before the failure event of interest, a common-cause failure event can lead to simultaneous failure of any number $f \in \{1, \dots, m\}$ of components. The alpha-factor model [2, 3] introduces parameters α_j , for $j = 1, \dots, m$, representing the probability that precisely j of the m components in the system fail simultaneously, given that a failure event occurs, hence $\sum_{j=1}^m \alpha_j = 1$. It is possible to select values α_j based on background knowledge of the system, but in this paper we aim at learning about these probabilities from available failure data whilst attempting to add only rather minimal further assumptions. Recently, two Bayesian solutions to the same problem have been presented, one using a non-informative Dirichlet prior distribution [4] and the other using the imprecise Dirichlet model (IDM) [5]. Both of these let the numbers 1 to m of possible simultaneous failures be represented by m categories, with an assumed multinomial

distribution for the numbers of observations in these categories. This is a standard statistical approach, with the IDM the most widely applied imprecise probability model in statistics [6, 11] with also interesting applications in reliability, see for example [12, 13]. There are, however, two issues with this model. Even while the (set of) prior distribution(s) is chosen in order to have little influence on the inferences, they do affect these noticeably in case interest is in categories which have rarely or even never been observed. Furthermore, any ordering of the categories is not taken into account. The latter aspect is particularly relevant if one is interested in events involving unions of categories, for example that the number of simultaneously failing components is at least two.

This paper presents a nonparametric predictive inference (NPI) alternative to the combination of the alpha-factor model with (imprecise) Dirichlet prior(s), also using lower and upper probabilities to quantify uncertainty. NPI for multinomial data has been presented as an alternative to the IDM [14], while NPI for real-valued data, including right-censored observations, has also been applied successfully [15, 16]. However, in this paper attention is restricted to NPI for ordinal data [17] as this explicitly takes the natural ordering of the m categories, namely the numbers $1, \dots, m$, into account. Assume that data are available on n previous common-cause failure events, which are assumed to be exchangeable with the next event, implying that they also all involved m components, and all components, in the observed systems and for the prediction, are exchangeable. Let $n_j \geq 0$ denote the number of the observed common-cause events in which precisely j components failed simultaneously, so $\sum_{j=1}^m n_j = n$. It is important to emphasize that no information or assumption is used about which specific components failed, just the numbers of components failing at the failure events are of interest for the inferences considered in this paper.

It is important to comment briefly on the assumption of exchangeability of all components, both in the observed systems and for prediction. For simplicity, one can consider this assumption as being in line with assuming all components have the same (unknown) probability of failure in case of a common-cause failure event. Hence, one could not, for example, apply the method presented in this paper in a scenario where components that have been in a system for a longer time undergo some wear, in the sense of their probability of failure increasing. If one would deem new components in a system, which replace failed components, to have a lower probability of failure upon a future common-cause failure event than other components in

the system, then one would need to distinguish the components into different types. This can also be modelled through the general concept of the survival signature, but for simplicity of presentation this paper restricts attention to components of a single type. There are, however, scenarios where it seems quite reasonable to assume exchangeability of the components in case of a common-cause failure event, for example if the cause of the failures is due to shocks and the components' capacity to deal with such shocks is independent of their age.

NPI for ordinal data [17] provides lower and upper probabilities for events involving the number Y_{n+1} of failing components at the next, so $n+1$ -st, failure event. These NPI lower and upper probabilities are not available in analytic form for all possible events $Y_{n+1} \in C$ with C any subset of $\{1, \dots, m\}$, generally they are calculated via counting of orderings in an assumed underlying latent variable representation, with a real-valued latent variable assumed to fall into the different categories, ordered on the real-line, determining the observation category. Such an underlying model assumption is rather common in statistics as it leads to quite simple models, but in addition it may be quite realistic in common-cause failure scenarios, for example one might consider there to be an underlying variable corresponding to strength of the cause of the failure event (e.g. strength of wind or earth quake) that would not be directly observable but could play a crucial role in determining the number of simultaneously failing components. There are some special events for which closed-form expressions are available [17] and these include the events required for the theory presented in Section 3 in this paper. This special case involves events of the form $Y_{n+1} \in \{s, \dots, t\}$, with $1 \leq s \leq t \leq m$, so the event that the number of components which fail simultaneously at the next failure event is any value from s to t . This includes events involving just a single number ($s = t$), events that the number of failing components is at least s (with $t = m$), and at most t (with $s = 1$). The event with $s = 1$ and $t = m$ is trivially true, so both the NPI lower and upper probabilities for this event are equal to 1 and this event is excluded from further consideration. Let $n_{s,t} = \sum_{j=s}^t n_j$. Using the notation $(x)^+ = \max(x, 0)$, the NPI lower and

upper probabilities for such events are [17]

$$\underline{P}(Y_{n+1} \in \{s, \dots, t\}) = \begin{cases} \frac{(n_{s,t} - 1)^+}{n + 1} & \text{if } 1 < s \leq t < m \\ \frac{n_{s,t}}{n + 1} & \text{if } s = 1 \text{ or } t = m \end{cases} \quad (2)$$

$$\overline{P}(Y_{n+1} \in \{s, \dots, t\}) = \frac{n_{s,t} + 1}{n + 1} \text{ for } 1 \leq s \leq t \leq m \quad (3)$$

The method presented in this paper uses the lower and upper cumulative distribution functions (CDF), which immediately follow from (2) and (3), with for $j \in \{1, \dots, m - 1\}$

$$\underline{F}(j) = \underline{P}(Y_{n+1} \leq j) = \frac{n_{1,j}}{n + 1} \quad (4)$$

$$\overline{F}(j) = \overline{P}(Y_{n+1} \leq j) = \frac{n_{1,j} + 1}{n + 1} \quad (5)$$

and $\underline{F}(m) = \overline{F}(m) = 1$. Note that Y_{n+1} cannot be equal to 0, as it is assumed that at least one component fails at a failure event, hence $\underline{F}(0) = \overline{F}(0) = 0$.

It is important to emphasize that these NPI lower and upper probabilities are frequentist statistics inferences [6, 7], and can be interpreted as bounds for a confidence level in the occurrence of the stated event of interest, based on the available data. The confidence interpretation results from the exchangeability assumption and can be interpreted as a long-run frequency where, for varying data sets but with exchangeability of observed data and the future observation for each run, the proportion of times that the event will indeed occur will be between the NPI lower and upper probabilities.

3. SYSTEM RELIABILITY AFTER FAILURE EVENT

The survival signature $\Phi(l)$ for a system consisting of m exchangeable components represents the probability that the system functions if precisely l of its components function. So the functioning of the system after the next common-cause failure event can be predicted by combining the survival signature and NPI for ordinal data, as presented in Section 2. The NPI lower probability for the event S that the system still functions after the

next common-cause failure event is

$$\underline{P}(S) = \sum_{j=1}^m \Phi(m-j) [\underline{F}(j) - \underline{F}(j-1)] \quad (6)$$

$$= \sum_{j=1}^{m-1} \frac{n_j}{n+1} \Phi(m-j) \quad (7)$$

and the corresponding NPI upper probability is

$$\overline{P}(S) = \sum_{j=1}^m \Phi(m-j) [\overline{F}(j) - \overline{F}(j-1)] \quad (8)$$

$$= \frac{n_1+1}{n+1} \Phi(m-1) + \sum_{j=2}^{m-1} \frac{n_j}{n+1} \Phi(m-j) \quad (9)$$

Here the logical assumption $\Phi(0) = 0$ has been used, so the system does not function if none of its components function. Furthermore, the assumption that the system is coherent has been used, hence the system survival signature $\Phi(l)$ is non-decreasing in l . These results imply that

$$\overline{P}(S) = \underline{P}(S) + \frac{\Phi(m-1)}{n+1} \quad (10)$$

These NPI lower and upper probabilities are derived by minimising and maximising, respectively, the probability for event S corresponding to every possible precise probability distribution within the bounds specified by the NPI results presented in Section 2. For non-coherent systems the approach can also be used, but the optimisation results are not available in general forms. For a series system, $\Phi(m) = 1$ and $\Phi(l) = 0$ for all $l < m$, so $\underline{P}(S) = \overline{P}(S) = 0$, which is trivial as a series system cannot function if at least one of its components has failed. For a parallel system, $\Phi(l) = 1$ for all $l \geq 1$ so $\underline{P}(S) = \frac{n_{1,m-1}}{n+1} = \frac{n-n_m}{n+1}$ and $\overline{P}(S) = \frac{n_{1,m-1}+1}{n+1} = \frac{n-n_m+1}{n+1}$. For a k -out-of- m system with $k \in \{2, \dots, m-1\}$, which functions if and only if at least k of its m components function, $\Phi(l) = 1$ for $l \geq k$ and $\Phi(l) = 0$ for $l < k$, hence $\underline{P}(S) = \frac{n_{1,m-k}}{n+1}$ and $\overline{P}(S) = \frac{n_{1,m-k}+1}{n+1}$.

From Equation (10) the imprecision $\Delta(S)$ for event S , which is the difference between the corresponding NPI upper and lower probabilities for this event, follows immediately,

$$\Delta(S) = \overline{P}(S) - \underline{P}(S) = \frac{\Phi(m-1)}{n+1} \quad (11)$$

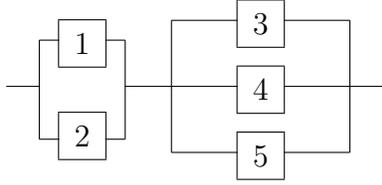


Figure 1: System for Example 1

This imprecision is obviously decreasing as function of n , which is fully in line with intuition. Furthermore, straightforward analysis can be used to derive relations between the NPI lower (or upper) probabilities for the event S for a specific system based on different data sets in which all categories have been observed in the same proportion. Consider a data set A with in total n^a observed failure events, of which n_j^a involved precisely j failing components, and data set B with in total $n^b = wn^a$ observed failure events, of which $n_j^b = wn_j^a$ involved precisely j failing components, and $w > 1$ such that the resulting n_j^b are all integers. While there is not much insight to be gained from such relations, it follows easily that, for all such $w > 1$, the NPI lower and upper probabilities are logically nested, that is

$$\underline{P}^a(S) \leq \underline{P}^b(S) \leq \hat{P}(S) \leq \overline{P}^b(S) \leq \overline{P}^a(S) \quad (12)$$

where $\hat{P}(S) = \sum_{j=1}^{m-1} \frac{n_j}{n} \Phi(m-j)$, the empirical estimator for the probability that the system functions, as based on the empirical proportions $\frac{n_j}{n}$ of the observations for each number of simultaneously failing components. The NPI approach always provides lower and upper probabilities that bound the corresponding empirical estimator in this logical way [7, 18]. All inequalities in (12) tend to be strict except in some trivial cases with either values equal to 0 or 1.

Example 1.

The system in Figure 1 has the survival signature with the following values: $\Phi(1) = 0$, $\Phi(2) = 0.6$, $\Phi(3) = 0.9$ and $\Phi(4) = 1$, in addition to the trivial values $\Phi(0) = 0$ and $\Phi(5) = 1$. Suppose that $n = 10$ failure events for components of such a system have been observed, with $n_1 = 4$ events involving a

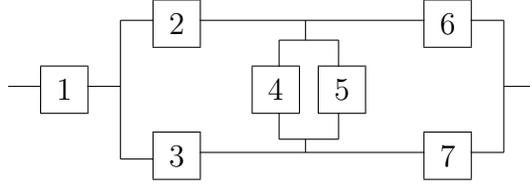


Figure 2: System for Example 2

single failing component, $n_2 = 3$ events involving simultaneous failure of two components, and $n_3 = 2$, $n_4 = 0$ and $n_5 = 1$ the number of events involving the simultaneous failure of three, four and five components, respectively. Interest is in the event S that the system, with all its components currently functioning, will still function after the next failure event, which is assumed to be exchangeable with these 10 previous events with regard to the number of simultaneously failing components. The NPI lower probability for this event, given by Equation (7), is $\underline{P}(S) = 0.718$ and the corresponding NPI upper probability, given by Equation (9), is $\overline{P}(S) = 0.809$. The imprecision in this case is $\Delta(S) = \frac{\Phi(4)}{11} = 0.091$.

As a second data set for the same system consider $n = 40$ failure events with $n_1 = 25$, $n_2 = 10$, $n_3 = 4$, $n_4 = 1$ and $n_5 = 0$. This leads to NPI lower probability $\underline{P}(S) = 0.888$ and $\overline{P}(S) = 0.912$. The imprecision in this case is $\Delta(S) = \frac{\Phi(4)}{41} = 0.024$.

Example 2.

The system in Figure 2 was used by Al-nefaiee and Coolen [19] to illustrate specific aspects of nonparametric predictive inference for system failure time using the system signature. The system signature and system survival signature are directly related for systems with a single type of components, as discussed by Coolen and Coolen-Maturi [8]. The survival signature for this system is given by $\Phi(1) = \Phi(2) = 0$, $\Phi(3) = 0.057$, $\Phi(4) = 0.343$, $\Phi(5) = 0.619$, $\Phi(6) = 0.857$ and the trivial values $\Phi(0) = 0$ and $\Phi(7) = 1$.

Suppose first that available data consist of $n = 100$ observed failure events, with $n_1 = 70$, $n_2 = 20$, $n_3 = 10$ and $n_j = 0$ for $j \in \{4, \dots, 7\}$. This leads to NPI lower and upper probabilities $\underline{P}(S) = 0.7506$ and $\overline{P}(S) =$

0.7591 for system functioning after the next failure event, with imprecision $\Delta(S) = \bar{P}(S) - \underline{P}(S) = \frac{\Phi(6)}{101} = 0.0085$. If there are only $n = 10$ observations, with numbers of failing components occurring in the same ratios, so $n_1 = 7, n_2 = 2, n_3 = 1$ and $n_j = 0$ for $j \in \{4, \dots, 7\}$, then the corresponding NPI lower and upper probabilities are $\underline{P}(S) = 0.6892$ and $\bar{P}(S) = 0.7671$, with imprecision $\Delta(S) = \frac{\Phi(6)}{11} = 0.0779$. Together with $\hat{P}(S) = 0.7581$ these values illustrate the inequalities (12).

Consider further the same data with $n = 100$ except now with $n_3 = 9$ and $n_7 = 1$, so for one observed failure event actually all 7 components failed instead of just three. This leads to $\underline{P}(S) = 0.7472$ and $\bar{P}(S) = 0.7557$, so both slightly smaller than for the earlier case with the nearly identical 100 observations, while the imprecision remains equal to $\Delta(S) = 0.0085$.

4. PERFORMANCE EVALUATION

The method for prediction of system functioning after a future common-cause failure event, as presented in Section 3, is quite straightforward and leads, for each individual application, to lower and upper probabilities for the event that the system will still function after such an event. Due to the use of NPI, these lower and upper probabilities have powerful properties in the frequentist statistics framework [7], and the imprecision, so the difference between the corresponding upper and lower probabilities, reflects the amount of information in the available data. It is important to investigate the behaviour of this method further, in particular to get insight into the quality of the predictions. Due to the use of lower and upper probabilities this is not straightforward, because the additional dimension of uncertainty, reflecting the indeterminacy through imprecision, may not have a clear benefit per se. Indeed, the imprecision is a consequence of limited information from data and limited assumptions. Zaffalon et al [20] have proposed a method to measure the quality of an imprecise prediction by a single number. Their metric consists of an objective component and an aspect related to the decision maker's risk aversion to variability of predictions. Such an approach has advantages, as in contrast to intervals, single numbers can be uniquely ordered, which is helpful to compare different methods. However, an alternative method is proposed here to illustrate and investigate the performance of the presented predictive method. This novel method for performance evaluation does not rely on subjective input and reflects imprecision, which is

important, for example, to study the effect of the number of available data on the inferences.

This new method for investigating the performance of imprecise predictive inferences is presented and applied to investigate how the inferential method presented in Section 3 is influenced by the number of available data. The new method uses the theory of Receiver Operating Characteristic (ROC) curves, a well established method for comparison of the performance of diagnostic methods [21, 22]. While main developments and applications of ROC methods have mostly been reported in medical statistics, they have been increasingly applied in recent years, including to reliability problems. For example, they have been used for the comparison of different statistical models for failures in water supply networks [23] and steel coil manufacturing [24]. Only basic nonparametric methods for ROC analysis are used in this paper, with necessary adaptation to deal with imprecise probabilities as inputs. It is, however, worth mentioning that NPI provides an interesting alternative to the established ROC methods, again with a fully predictive nature [25], its application to scenarios as considered in this paper is left as a topic for future research.

The method presented in Section 3 has as input the numbers n_j , for $j = 1, \dots, m$, of events in which j components failed simultaneously, with data of in total $n = \sum_{j=1}^m n_j$ events. The output of the method consists of the corresponding NPI lower and upper probabilities, $\underline{P}(S)$ and $\overline{P}(S)$, for the event S that the system still functions after the next failure event. At the moment these lower and upper probabilities are calculated, the actual future functioning status of the system will of course be unknown, hence these lower and upper probabilities can be regarded as a diagnostic tool, with higher values more indicative of system functioning. Insight into the performance of this method can be gained through simulation studies, which are described next. It is important here to simulate in such a manner that the quality of the prediction can be investigated. This can be achieved by simulating several cases involving different mixtures of distributions, and then to see if the method picks up the difference well in its predictions. Furthermore, the simulations will illustrate the effect on the inferences of the amount of data available, where typically imprecision should tend to decrease with increasing numbers of available data.

The simulation study is performed as follows, describing one run in detail. Two probability distributions are assumed over the integers $\{1, 2, \dots, m\}$,

from each $n+1$ values are sampled with replacement. Of these $n+1$ values for each distribution, one is drawn by simple random sampling, this will serve as a future observation and be denoted by n_f . The remaining n values sampled from each distribution serve as data, with n_j the number of times the value j is included in this simulated data set. Interestingly, it is not important how these $n+1$ values are simulated as the NPI approach only requires the future observation of interest to be exchangeable with the observed data, which is ensured by the random selection of one of these $n+1$ values to serve as the future observation. The NPI lower and upper probabilities, $\underline{P}(S)$ and $\overline{P}(S)$, are calculated on the basis of this simulated data set $\{n_1, \dots, n_m\}$. The simulated future observation n_f functions as the number out of m components failing at the next failure event, and is used to simulate an actual future observation of the functioning status of the system by drawing a value for a Bernoulli distributed random quantity S_f which takes the value $S_f = 1$ with probability $\Phi(m - n_f)$ and $S_f = 0$ else. Hence, S_f simulates the status of the system after the next failure event, with $S_f = 1$ indicating functioning and $S_f = 0$ failure of the system.

The resulting output from one simulation run is a triple $(\underline{P}(S), \overline{P}(S), S_f)$ for each of the two distributions used for the simulation. The full simulation study consists of a predetermined number K of such independent runs, leading to $2K$ such triples. It is important here that two (or more) probability distributions have been used for the simulations, if only a single distribution were used all values S_f would have been exchangeable and it would not allow investigation of success of the method as there would not be any reason why higher values of \underline{P} and \overline{P} would tend to correspond to value $S_f = 1$ than for lower values of \underline{P} and \overline{P} . Informally, the performance of the predictive method is good if values $S_f = 1$ occur mostly in triples with higher values of $\underline{P}(S)$ and $\overline{P}(S)$ while $S_f = 0$ occurs mostly in triples with lower values of $\underline{P}(S)$ and $\overline{P}(S)$. Hence, these lower and upper probabilities can be considered as diagnostic tools to indicate the future system status, enabling the method's performance to be evaluated using ROC analysis based on the output of these simulations. In particular, the area under the ROC curve (AUC) is an established tool to measure the diagnostic quality of a classification method, where AUC of about 0.5 indicates that the method does not perform better than random allocation while AUC close to 1 indicates excellent performance.

In order to investigate the method's performance, the data resulting from the K simulation runs, so in total $2K$ triples $(\underline{P}(S), \overline{P}(S), S_f)$ as in each

run two probability distributions were used, are divided into two groups, depending on the value of S_f . So $S_f = 1$ for one group of simulation runs with the system functioning after the next failure event, let K_1 denote the number of simulation runs for which the output is in this group. Similarly, $S_f = 0$ for the group of simulation runs with the system failing at the next failure event, the number of simulation runs in this group is denoted by K_0 , of course $K_0 + K_1 = 2K$. Application of ROC analysis to indicate the performance of this method would now be relatively straightforward, e.g. using standard nonparametric methods for such an analysis in order not to make further modelling assumptions, apart from the fact that the indicators consist of intervals $[\underline{P}(S), \overline{P}(S)]$ for each run, while normally for ROC analysis the information on which to base the diagnosis is assumed to be a single value. To deal with this, optimal bounds for the ROC analysis for all possible combinations of probabilities within the intervals $[\underline{P}(S), \overline{P}(S)]$ can be derived, as described next. These bounds reflect the imprecision in the NPI method.

Some further notation is introduced for this ROC analysis. For simulation run $j \in \{1, \dots, K_1\}$ with $S_f = 1$, the NPI lower and upper probabilities for the event that the system will function at the next observation are denoted by $\underline{P}_j^1(S)$ and $\overline{P}_j^1(S)$, respectively. For simulation run $l \in \{1, \dots, K_0\}$ with $S_f = 0$, the corresponding notation is $\underline{P}_l^0(S)$ and $\overline{P}_l^0(S)$. For any set of probabilities $p_j^1 \in [\underline{P}_j^1(S), \overline{P}_j^1(S)]$, $j = 1, \dots, K_1$, and $p_l^0 \in [\underline{P}_l^0(S), \overline{P}_l^0(S)]$, $l = 1, \dots, K_0$, there is a single corresponding nonparametric estimate of the ROC curve, say $\text{ROC}(p^1, p^0)$ with $p^1 = (p_1^1, \dots, p_{K_1}^1)$ and $p^0 = (p_1^0, \dots, p_{K_0}^0)$. It is easy to see that these ROC curve estimates depend monotonically on p^1 and p^0 . If an element of p^1 is increased then the ROC curve can only change by moving upwards, with the area under the ROC curve (denoted by AUC) increasing. Similarly, if an element of p^0 is increased then the ROC curve can only change by moving downwards, with the AUC decreasing. This implies that the set of all $\text{ROC}(p^1, p^0)$ with $p_j^1 \in [\underline{P}_j^1(S), \overline{P}_j^1(S)]$, $j = 1, \dots, K_1$, and $p_l^0 \in [\underline{P}_l^0(S), \overline{P}_l^0(S)]$, $l = 1, \dots, K_0$, is bounded from below by $\underline{\text{ROC}} = \text{ROC}(\underline{P}^1(S), \overline{P}^0(S))$, with $\underline{P}^1(S) = (\underline{P}_1^1(S), \dots, \underline{P}_{K_1}^1(S))$ and $\overline{P}^0(S) = (\overline{P}_1^0(S), \dots, \overline{P}_{K_0}^0(S))$, and this set is bounded from above by $\overline{\text{ROC}} = \text{ROC}(\overline{P}^1(S), \underline{P}^0(S))$, with $\overline{P}^1(S) = (\overline{P}_1^1(S), \dots, \overline{P}_{K_1}^1(S))$ and $\underline{P}^0(S) = (\underline{P}_1^0(S), \dots, \underline{P}_{K_0}^0(S))$. Of course, these bounds are optimal because they are actually obtained for the combinations of lower and upper proba-

| Failing components: | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---------------------|-----|-----|-----|-----|-----|-----|-----|
| A | 0.6 | 0.2 | 0.1 | 0.1 | 0 | 0 | 0 |
| B_1 | 0.4 | 0.2 | 0.1 | 0.1 | 0 | 0 | 0.2 |
| B_2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0 | 0 |
| B_3 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.2 | 0.3 |
| B_4 | 0 | 0 | 0 | 0.1 | 0.1 | 0.2 | 0.6 |

Table 1: Probability distributions used in 4 cases for simulation

| Case | 1 | 2 | 3 | 4 |
|-----------|------------------|------------------|------------------|------------------|
| $n = 10$ | (0.3653, 0.7291) | (0.5225, 0.7755) | (0.6470, 0.8459) | (0.8041, 0.9416) |
| $n = 100$ | (0.5361, 0.6121) | (0.6274, 0.6987) | (0.7228, 0.7893) | (0.8528, 0.9041) |
| $n = 500$ | (0.5622, 0.5969) | (0.6491, 0.6827) | (0.7277, 0.7587) | (0.8592, 0.8832) |

Table 2: Results (\underline{AUC} , \overline{AUC}) of simulation study

bilities as indicated. Optimal bounds for the AUC values for all considered ROC curves follow immediately, with the lower bound \underline{AUC} the area under ROC and the upper bound \overline{AUC} the area under \overline{ROC} .

The method is now implemented for the system in Example 2 in Section 3. Four cases have been investigated, with two probability distributions over the integers $\{1, \dots, 7\}$ used in each case. These cases are specified in Table 1. One distribution was kept constant, it is indicated as distribution A . The second distribution used differs per case, and is indicated by B_i for Case $i \in \{1, 2, 3, 4\}$.

Note that the probability distributions B_3 and particularly B_4 , used in Cases 3 and 4, are substantially different from probability distribution A . Therefore, in Cases 3 and 4 the method should be able to perform well in prediction, as higher values of $\underline{P}(S)$ and \overline{P} will be more likely to result from the simulations using probability distribution B_3 or B_4 than those using A , and hence the corresponding predicted status of the system will also more likely be a failure. Table 2 presents the values \underline{AUC} and \overline{AUC} for this investigation, all based on $K = 1,000$ simulation runs (hence with 2,000 simulated triples in each full simulation), and with different sizes n of the simulated data sets.

The results in Table 2 indicate that the method performs indeed in line

with expectations. The quality of prediction increases if the two probability distributions included in one simulation case differ more substantially, and the imprecision $\overline{AUC} - \underline{AUC}$ decreases for larger values of n . For $n = 10$, the value of \underline{AUC} for Case 1 is less than 0.5, which might be interpreted as indicating that a classification performance better than random allocation is not guaranteed in this case, but it is certainly possible given that the corresponding \overline{AUC} is greater than 0.5. This is actually not surprising as the two distributions used in Case 1 do not differ very much and the number of observations used in each simulation is small. These simulations were also carried out with larger values of K , but the results were close to those reported in Table 2.

This performance evaluation method has been presented here in detail as it appears not to have been reported in the literature before, and it has the clear advantage of reflecting, through the imprecision, the influence of the amount of available data on the performance of the method.

5. DISCUSSION

The nonparametric predictive inference (NPI) approach to statistics [6, 7] has strong frequentist properties and explicitly focuses on prediction, hence on the next event of interest. In this it is different from main methods of frequentist statistics, where emphasis tends to be on estimation of population characteristics. In practice, the latter setting is quite indirect if one is indeed interested in the next event. One could follow estimation inferences by prediction, but such a two-stage approach has many possible problems and, as the NPI approach shows, is not necessary. One could argue about the nature of reliability of a system, whether this is a property of the system (which perhaps would fit better with the classical frequentist approach to statistics) or whether it is a future event, which is the point of view we have taken here. Beyond this, NPI lower and upper probabilities are considered as quantifications of the uncertainty, based on available data and quite weak modelling assumptions, for the event of interest, namely functioning of the system after the next common-cause failure event. Of course, if one would make different modelling assumptions, one would get different uncertainty quantifications, there is no claim in the NPI approach that these are intrinsic properties of the system. This explicit view of reliability related to a future event for which prediction is an adequate inference, also led to the need to judge the performance of the method differently than would have been appropriate if

one had aimed at estimating a system property (or statistically, a ‘population property’), hence the new approach for such a possible performance evaluation method in Section 4.

The basic approach presented in this paper can be generalized in several ways. These are all relevant for further development of the approach towards practical implementation for substantial real world systems and networks, and all require some additional research. A detailed study in which the prediction method presented here is compared with a variety of alternative methods, including the (imprecise) Bayesian approach to the α -factor model will be of interest. In particular, it will also be important to include a variety of measures to evaluate the performances of the methods. As this latter issue of suitable methods for performance evaluation of imprecise probabilistic methods has only quite recently received attention, and is currently an active area of research, such a detailed study of performance of different prediction methods is left as an important topic for future research. It is felt that the performance evaluation method presented in Section 4 has important advantages, investigating its properties in comparison to alternative evaluation methods is therefore also important, ideally in a variety of classification problems.

If a failure event occurs, involving one or more components but not leading to system failure, it may be of interest to consider the reliability of the system in case of a future further failure event. This could provide important insights into whether or not replacement of (some of) the failed components is required. Assuming that inspection after a failure event reveals with certainty which of the system’s components have failed, such inspection leads to a changed structure of the system, with only its functioning components taken into account, for which the survival signature can be computed. This will involve less computation than the derivation of the original survival signature for the full system. Hence, implementation of the approach presented in this paper appears to be straightforward, yet one has to be careful with regard to the use of the available data and whether or not it is reasonable to assume that the number of components failing at the next failure event is independent of the number of components that have failed and also which components have failed. Data on common-cause failure events, as assumed in this paper, cannot automatically be used in case the system has a reduced number of functioning components, in practice it is unlikely that one would have a substantial data set related to the actual reduced system. With some additional assumptions one could use the original data set in an adapted form,

for example by changing each value by taking into account the probability that one or more of the failed components for the data would be in the set of components that have now failed in the system, but this requires some more research. This topic has close links to similar considerations for the system signature [26], yet the statistical aspects mentioned here appear not yet to have been considered in detail in the literature.

Coolen and Coolen-Maturi [8] introduced the survival signature as an alternative to the system signature [9] particularly for its generalization to systems with multiple types of components, which of course is the case for most real world systems and networks. The prediction of system failure time using NPI and the survival signature has been presented by Coolen et al [27]. As shown in this paper, the survival signature provides an attractive and quite straightforward method for inference in case of common-cause failures, which are also important to study for such more general systems with multiple types of components. The approach presented in Section 3 can, quite straightforwardly be generalized to systems with multiple types of components, where care is required due to the fact that an observed failure event in the data would not necessarily imply that at least one component of each type had failed. A more important consideration, however, would be whether or not one can assume independence of numbers of failures of components of different types, at the next failure event. If one is happy to assume this, then indeed the generalization is pretty straightforward, but in practice it may be important to take dependencies between such numbers of failing components into account. If one has historical data for the system, with detailed information about the numbers of components of all types failing simultaneously, then such inference is possible, but NPI for general multivariate ordinal data has not yet been developed. There are several heuristic approaches that could be applied, but the NPI approach is particularly attractive due to its foundational properties as a frequentist statistics procedure, hence further research is required to enable appropriate generalization.

Many reliability problems involve networks, for example in energy provision. Recently, Aslett *et al* [28] have presented Bayesian statistical inference for system reliability using the survival signature, including an introduction of survival signatures for networks, which typically have at least two types of components (the nodes and the links between them). Further research is required to upscale the approach to large real-world networks, but for such networks common-cause failures are important to be taken into account, hence the method presented in this paper will be useful for such applications.

Recently, Coolen and Coolen-Maturi [10] discussed how further aspects of uncertainty and indeterminacy about the system's functioning can be dealt with by generalizing the system structure function to a (possibly imprecise) probability instead of a binary function. In particular, this enables scenarios with multiple tasks to be dealt with, where it is not certain what kind of task the system has to deal with next, and where one can even consider as yet unobserved or even unknown tasks. This generalization can also be combined with the method presented in this paper, which provides a further interesting research topic.

In this paper attention has been restricted to the next failure event. This can be generalized in several ways. One can consider multiple future failure events, where NPI would take the dependence between such future events explicitly into account. NPI for multiple future events for ordinal data has not yet been presented but is relatively straightforward to develop due to the assumed underlying real-valued latent variable structure [17] and the fact that NPI has been presented for multiple future real-valued observations [29]. One can also consider reliability of the system at a certain future moment in time, assuming failure events to happen according to a stochastic process. This is also left as an interesting topic for future research, where particularly any dependence between the process and the numbers of failing components at a failure event would lead to challenging research questions and, in order to implement NPI, substantial amounts of data being required. Related to such a process view over time, it is also of interest to include aspects of wear-out (or burn-in) over time.

The approach presented here can also be used as a basis for a range of asset management decisions, for example one can explore the system reliability if, after a failure event, a few but not all failed components are replaced. Such considerations would probably involve aspects of costs, due to the explicit use of observable random quantities in the NPI approach the formulation of cost functions is quite straightforward. There is no need to rely on concepts such as average costs per unit of time over an infinite time horizon, as often used in traditional Operational Research methods for inspection and maintenance planning based on renewal reward theory, which tend to be made more for mathematical convenience than for their real importance [30]. This provides also a range of interesting challenges for future research.

Acknowledgements

We thank two anonymous reviewers whose comments and suggestions led to improved presentation.

References

- [1] Rasmuson D.M., Kelly D.L. (2008). Common-cause failure analysis in event assessment. *Journal of Risk and Reliability* **222** 521–532.
- [2] Mosleh A., Fleming K.N., Parry G.W., Paula H.M., Worledge D.H., Rasmuson D.M. (1988). Procedures for treating common cause failures in safety and reliability studies: procedural framework and examples. Technical report NUREG/CR-4780 EPRI NP-5613 (Vol. 1). PLG Inc., Newport Beach, CA (USA), January 1988.
- [3] Mosleh A., Fleming K.N., Parry G.W., Paula H.M., Worledge D.H., Rasmuson D.M. (1989). Procedures for treating common cause failures in safety and reliability studies: analytical background and techniques. Technical report NUREG/CR-4780 EPRI NP-5613 (Vol. 2). PLG Inc., Newport Beach, CA (USA), January 1989.
- [4] Kelly D., Atwood C. (2011). Finding a minimally informative Dirichlet prior distribution using least squares. *Reliability Engineering and System Safety* **96** 398–402.
- [5] Troffaes M.C.M., Walter G., Kelly D. (2014). A robust Bayesian approach to modeling epistemic uncertainty in common-cause failure models. *Reliability Engineering and System Safety* **125**, 13–21.
- [6] Augustin T., Coolen F.P.A., de Cooman G., Troffaes M.C.M. (2014). *Introduction to Imprecise Probabilities*. Wiley, Chichester.
- [7] Coolen F.P.A. (2011). Nonparametric predictive inference. In: *International Encyclopedia of Statistical Science*, Lovric M. (Ed.). Springer, Berlin, pp. 968–970.
- [8] Coolen, F.P.A., Coolen-Maturi T. (2012). On generalizing the signature to systems with multiple types of components. In: *Complex Systems and Dependability*, Zamojski W., Mazurkiewicz J., Sugier J., Walkowiak T., Kacprzyk J. (Eds.). Springer, Berlin, pp. 115–130.

- [9] Samaniego F.J. (2007). *System Signatures and their Applications in Engineering Reliability*. Springer, New York.
- [10] Coolen F.P.A., Coolen-Maturi, T. (2015). Modelling uncertain aspects of system dependability with survival signatures. In: *Dependability Problems of Complex Information Systems*, Zamojski W., Sugier J. (Eds). Springer, Berlin, pp. 19–34.
- [11] Walley P. (1996). Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B* **58** 3–34.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] Coolen F.P.A., Yan K.J. (2004). Nonparametric predictive inference with right-censored data. *Journal of Statistical Planning and Inference* **126** 25–54.
- [16] Janurova K., Bris R. (2014). A nonparametric approach to medical survival data: Uncertainty in the context of risk in mortality analysis. *Reliability Engineering and System Safety* **125** 145–152.
- [17] Coolen F.P.A., Coolen-Schrijner P., Coolen-Maturi T., Elkhafifi F.F. (2013). Nonparametric predictive inference for ordinal data. *Communications in Statistics - Theory and Methods* **42** 3478–3496.
- [18] Augustin T., Coolen F.P.A. (2004). Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference* **124** 251–272.

- [19] Al-nefaiee A.H., Coolen F.P.A. (2013). Nonparametric predictive inference for system failure time based on bounds for the signature. *Journal of Risk and Reliability* **227** 513–522.
- [20] Zaffalon M., Corani G., Maua D. (2012). Evaluating credal classifiers by utility-discounted predictive accuracy. *International Journal of Approximate Reasoning* **53** 1282–1301.
- [21] Krzanowski W.J., Hand D.J. (2009). *ROC Curves for Continuous Data*. Chapman & Hall, Boca Raton.
- [22] Pepe M.S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press, Oxford.
- [23] Debon A., Carrion A., Cabrera E., Solano H. (2010). Comparing risk of failure models in water supply networks using ROC curves. *Reliability Engineering and System Safety* **95** 43–48.
- [24] Debon A., Garcia-Diaz J.C. (2012). Fault diagnosis and comparison risk for the steel coil manufacturing process using statistical models for binary data. *Reliability Engineering and System Safety* **100** 102–114.
- [25] Coolen-Maturi T., Coolen-Schrijner P., Coolen F.P.A. (2012). Nonparametric predictive inference for diagnostic accuracy. *Journal of Statistical Planning and Inference* **142** 1141–1150.
- [26] Samaniego F.J., Balakrishnan N., Navarro J. (2009). Dynamic signatures and their use in comparing the reliability of new and used systems. *Naval Research Logistics* **56** 577–591.
- [27] Coolen F.P.A., Coolen-Maturi T., Al-nefaiee A.H. Nonparametric predictive inference for system reliability using the survival signature. *Journal of Risk and Reliability*, to appear.
- [28] Aslett L.J.M., Coolen F.P.A., Wilson S.P. Bayesian inference for reliability of systems and networks using the survival signature. *Risk Analysis*, to appear.
- [29] Arts G.R.J., Coolen F.P.A., van der Laan P. (2004). Nonparametric predictive inference in statistical process control. *Quality Technology and Quantitative Management* **1** 201–216.

- [30] Coolen-Schrijner P., Shaw S.C., Coolen F.P.A. (2009). Opportunity-based age replacement with a one-cycle criterion. *Journal of the Operational Research Society* **60** 1428–1438.