Reliability Engineering and System Safety 126 (2014) 87–97

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Nonparametric predictive inference for combined competing risks data

Tahani Coolen-Maturi [a,*], Frank P.A. Coolen [b]

[a] Durham University Business School, Durham University, Durham DH1 3LB, UK
[b] Department of Mathematical Sciences, Durham University, Durham DH1 3LE, UK

ARTICLE INFO

Article history:
Received 4 October 2013
Received in revised form
9 January 2014
Accepted 12 January 2014
Available online 25 January 2014

Keywords:
Imprecise probability
Lower and upper probability
Nonparametric predictive inference
Competing risks
Right-censored data
Combined data

ABSTRACT

The nonparametric predictive inference (NPI) approach for competing risks data has recently been presented, in particular addressing the question due to which of the competing risks the next unit will fail, and also considering the effects of unobserved, re-defined, unknown or removed competing risks. In this paper, we introduce how the NPI approach can be used to deal with situations where units are not all at risk from all competing risks. This may typically occur if one combines information from multiple samples, which can, e.g. be related to further aspects of units that define the samples or groups to which the units belong or to different applications where the circumstances under which the units operate can vary. We study the effect of combining the additional information from these multiple samples, so effectively borrowing information on specific competing risks from other units, on the inferences. Such combination of information can be relevant to competing risks scenarios in a variety of application areas, including engineering and medical studies.

1. Introduction

Nonparametric predictive inference (NPI) is a statistical method based on Hill's assumption $A_{(n)}$ [17], which gives a direct conditional probability for a future observable random quantity, conditional on observed values of related random quantities [1,6]. $A_{(n)}$ does not assume anything else, and can be interpreted as a post-data assumption related to exchangeability [16]. 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 bounds for probabilities via the 'fundamental theorem of probability' [16]. These bounds are lower and upper probabilities in imprecise probability theory [1,23,24]. A suitable, albeit informal, interpretation for lower and upper probabilities, is that a lower probability reflects the evidence in favour of the event of interest while an upper probability, or more accurately the difference between one and an upper probability, reflects the evidence against the event of interest. Short introductions to NPI, imprecise probability and its use in reliability have recently been presented [7,10,11].

In reliability and survival analysis, data on event times are often affected by right-censoring, where for a specific unit or individual it is only known that the event of interest has not yet taken place at a specific time. Coolen and Yan [12] presented a generalization of $A_{(n)}$, called rc-$A_{(n)}$, which is suitable for right-censored data. In comparison to $A_{(n)}$, rc-$A_{(n)}$ uses the additional assumption that, at the moment of censoring, the residual lifetime of a right-censored unit is exchangeable with the residual lifetimes of all other units that have not yet failed or been censored. The assumption rc-$A_{(n)}$ underlies the inferences in this paper, for more details we refer to [12,25].

Coolen et al. [9] introduced NPI for some reliability applications, including lower and upper survival functions for the next future observation, illustrated with an application with competing risks data. They illustrated the lower and upper marginal survival functions, so each restricted to a single failure mode. Competing risks have been the topic of many research papers over recent decades. As examples of applications, Jiang [18] applied a discrete competing risk model to bus motor failure data, Bunea et al. [3] used competing risk methods to analyse military systems data, and Bocchetti et al. [2] applied such methods for the study of reliability of marine diesel engines. Sarhan et al. [22] discussed and illustrated likelihood and classical statistical approaches to competing risks data, Coolen et al. [4] presented a Bayesian competing risk approach to reliability for heat exchangers based on expert judgements. Maturi et al. [21] presented NPI for competing risks data, in particular addressing the question due to which of the competing risks the next unit will fail. Related to this approach, Coolen-Maturi and Coolen [14] considered the

* Corresponding author.
E-mail addresses: tahani.maturi@durham.ac.uk (T. Coolen-Maturi), frank.coolen@durham.ac.uk (F.P.A. Coolen).

0951-8320/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ress.2014.01.007

effects of unobserved, re-defined, unknown or removed competing risks. Recently, Coolen-Maturi [13] introduced NPI to compare two groups under (observed or unobserved) competing risks.

In NPI for competing risks [21], it is assumed that there are K failure modes and a unit fails due to the first occurrence of a failure mode, which is identified with certainty. We should point out that, in this paper, we will use the terms 'failure mode' and 'competing risk' interchangeably with the same meaning. Let X_{n+1} denote the failure time of a future unit, based on n observations, and let the corresponding notation for the failure time including indication of the actual failure mode k be $X_{k,n+1}$. It is important to emphasize that $X_{k,n+1}$ is interpreted as the random failure time of a future unit which is *only* at risk from failure mode k. Different failure modes are assumed to occur independently. The competing risk data per failure mode consist of a number of observed times of failures caused by the specific failure mode considered, and right-censoring times caused by other failure modes or other reasons for right-censoring. Hence rc-$A_{(n)}$ can be applied per failure mode k for inference on $X_{k,n+1}$.

Suppose that, in the available data, u_k failures are caused by failure mode k, at times $x_{k,1} < x_{k,2} < \cdots < x_{k,u_k}$, and let $n-u_k$ be the number of the right-censored observations, $c_{k,1} < c_{k,2} < \cdots < c_{k,n-u_k}$, corresponding to failure mode k; these may be failure times due to other (independent) failure modes, or observations that are right-censored for other reasons, where it is assumed throughout that such censoring processes are independent of $X_{k,n+1}$. For notational convenience, let $x_{k,0} = 0$ and $x_{k,u_k+1} = \infty$. Suppose further that there are s_{k,i_k} right-censored observations in the interval (x_{k,i_k}, x_{k,i_k+1}), denoted by $c^{i_k}_{k,1} < c^{i_k}_{k,2} < \cdots < c^{i_k}_{k,s_{k,i_k}}$, so $\sum_{i_k=0}^{u_k} s_{k,i_k} = n-u_k$. It should be emphasized that we do not assume that each unit considered must actually fail, if a unit does not fail then there will be a right-censored observation recorded for this unit for each failure mode, as we assume that the unit will then be withdrawn from the study, or the study ends, at some point. The random quantity representing the failure time of the next unit, with all K failure modes considered, is $X_{n+1} = \min_{1 \le k \le K} X_{k,n+1}$.

For $i_k = 0, 1, \ldots, u_k$, let $t^{i_k}_{k,i^*_k} = c^{i_k}_{k,i^*_k}$ (i.e. censoring time) for $i^*_k = 1, 2, \ldots, s_{k,i_k}$ and $t^{i_k}_{k,i^*_k} = x_{k,i_k}$ (i.e. failure time or time 0) for $i^*_k = 0$. For notational convenience, let $t^{i_k}_{k,s_{k,i_k}+1} = t^{i_k+1}_{k,0} = x_{k,i_k+1}$ for $i_k = 0, 1, \ldots, u_k - 1$. Let $\tilde{n}_{c_{k,r}}$ and $\tilde{n}_{t^{i_k}_{k,i^*_k}}$ be the number of units in the risk set just prior to time $c_{k,r}$ and $t^{i_k}_{k,i^*_k}$, respectively, with the definition $\tilde{n}_0 = n+1$ for ease of notation. The risk set at a certain time contains all units that have not failed or been right-censored before that time, and hence are indeed still at risk. Fig. 1 represents the data and notation considering failure mode k.

The NPI lower and upper survival functions for the failure time of the next unit due to failure mode k, so if the unit were only at risk from this failure mode, are denoted by $\underline{S}_{X_{k,n+1}}(t)$ and $\overline{S}_{X_{k,n+1}}(t)$, respectively, and are as follows [21,9]. For $t \in (t^{i_k}_{k,a_k}, t^{i_k}_{k,a_k+1}]$ with $i_k = 0, 1, \ldots, u_k$ and $a_k = 0, 1, \ldots, s_{k,i_k}$,

$$\underline{S}_{X_{k,n+1}}(t) = \frac{1}{n+1}\tilde{n}_{t^{i_k}_{k,a_k+1}} \prod_{\{r: c_{k,r} < t^{i_k}_{k,a_k+1}\}} \frac{\tilde{n}_{c_{k,r}}+1}{\tilde{n}_{c_{k,r}}} \tag{1}$$

and for $t \in [x_{k,i_k}, x_{k,i_k+1})$ with $i_k = 0, 1, \ldots, u_k$,

$$\overline{S}_{X_{k,n+1}}(t) = \frac{1}{n+1}\tilde{n}_{x_{k,i_k}} \prod_{\{r: c_{k,r} < x_{k,i_k}\}} \frac{\tilde{n}_{c_{k,r}}+1}{\tilde{n}_{c_{k,r}}} \tag{2}$$

While predictive inference, as considered in this approach, is different to estimation, as it explicitly considers a single future unit instead of estimating characteristics of a population distribution, it is interesting to mention that these NPI lower and upper survival functions bound the well-known Kaplan–Meier estimator [19], which is the nonparametric maximum likelihood estimator for the population survival function in the case of lifetime data with right-censored observations, for more details we refer to [12,15].

If all the units are censored with regard to failure mode k (e.g. if all units failed due to other failure modes, where in this case $i_k = u_k = 0$), then the lower and upper survival functions in (1) and (2) are equal to [14]

$$\underline{S}_{X_{n+1}}(t) = \frac{\tilde{n}_{t^{i_k}_{k,a_k+1}}}{\tilde{n}_{t^{i_k}_{k,a_k+1}}+1} \quad \text{and} \quad \overline{S}_{X_{n+1}}(t) = 1 \tag{3}$$

If the next unit considered is at risk from K independent failure modes, so with its failure time given by $X_{n+1} = \min_{1 \le k \le K} X_{k,n+1}$, then the NPI lower and upper survival functions for its failure time are denoted by $\underline{S}_{X_{n+1}}(t)$ and $\overline{S}_{X_{n+1}}(t)$, respectively, and are equal to

$$\underline{S}_{X_{n+1}}(t) = \prod_{k=1}^{K} \underline{S}_{X_{k,n+1}}(t) \quad \text{and} \quad \overline{S}_{X_{n+1}}(t) = \prod_{k=1}^{K} \overline{S}_{X_{k,n+1}}(t) \tag{4}$$

In Section 2 the main results of this paper are presented, considering combination of information from different groups for several scenarios. This is an important contribution to the literature on competing risks from the perspective of NPI, as in practice one may often have data from a variety of competing risks scenarios which are closely related in the sense that several competing risks occur in all scenarios but there is no full exchangeability (which would allow grouping of all data without further complications) due to some competing risks not applying in all scenarios. Such situations occur frequently in practice. In engineering contexts, the same systems may function in different locations under slightly different circumstances, with several failure modes occurring everywhere but some failure modes specific to one or a few locations. In medical contexts, some diseases may affect both males and females, while other diseases may be gender-specific. This paper presents a general theory of NPI for such circumstances. Section 3 presents an extensive example to illustrate the results, followed by some concluding remarks in Section 4.

2. NPI for combined competing risks data

We now present a generalization of the NPI approach to competing risks, by considering the important situation of different groups of units, such that units from the same group are at risk from the same set of competing risks, but these sets differ for the different groups. Of course, it is typically assumed that there is at least some overlap between the sets of competing risks for different groups. In this case, the information in data from different groups about a specific failure mode, that applied to these groups, can be used to enhance inferences for a unit at risk

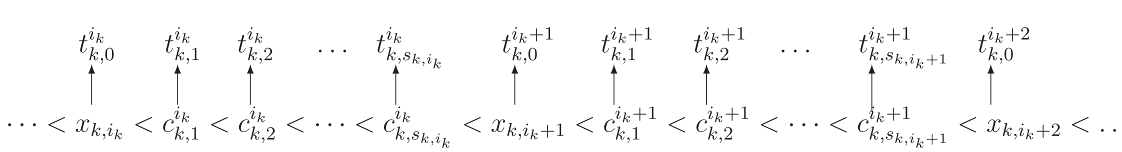

Fig. 1. NPI data representation, considering failure mode k.