Early Termination of Experiments in Nonparametric Predictive Comparisons

Tahani A. Maturi, Pauline Coolen-Schrijner, and Frank P.A. Coolen

Department of Mathematical Sciences, Durham University,
Durham, DH1 3LE, UK
(e-mail: tahani.maturi@durham.ac.uk, pauline.schrijner@durham.ac.uk,
frank.coolen@durham.ac.uk)

Abstract. Coolen and van der Laan (2001) presented nonparametric predictive inference (NPI) for comparison of groups of units, aimed at selection of the 'best group', where 'best' relates to maximum value of a real-valued random quantity corresponding to each unit, e.g. its failure time. NPI is a statistical approach based on few assumptions in addition to data, made possible by the use of lower and upper probabilities to quantify uncertainty. We extend the results by Coolen and van der Laan by considering experiments that are terminated before all units have failed, and the effect of this on the inferences. It is shown that continuing an experiment, as long as not all units have failed, will never lead to increased imprecision, which will indeed decrease if further failures are observed.

1 Introduction

We consider comparison of failure times of units from different groups, simultaneously placed on a life-testing experiment with each unit failing at most once, and we focus on the effect of early termination of the experiment before all units have failed. In the statistical literature, this scenario occurs in 'precedence testing', an excellent overview is presented by Balakrishnan and Ng (2006). Coolen and van der Laan (2001) proposed a nonparametric predictive approach for comparison of different groups, with emphasis on selection of the best group. Their inferences are in terms of lower and upper probabilities for events that compare the failure times of one further unit from each group. Lower and upper probabilities generalize classical probabilities, where imprecision reflects the amount of information available. A lower (upper) probability for an event A, denoted by $\underline{P}(A)$ ($\overline{P}(A)$), can be interpreted in several ways (Augustin and Coolen (2004)). From subjective point of view, it can be interpreted as supremum buying (infimum selling) price for a gamble on the event A, while more generally it can be interpreted as the maximum lower (minimum upper) bound for the probability of A that follows from the (restricted) assumptions made and the available data. Informally, $\underline{P}(A)$ ($\overline{P}(A)$) can be considered to reflect the evidence in favour of (against) event A. We study the effect of early termination of the experiment on the lower and upper probabilities of Coolen and van der Laan (2001).

2 Main results

Nonparametric predictive inference (NPI) is a statistical method based on Hill's assumption $A_{(n)}$ (Hill (1968)), which gives direct (lower and upper) probabilities for a future observable random quantity, based on observed values of n related random quantities (Augustin and Coolen (2004)). Inferences based on $A_{(n)}$ are predictive and nonparametric, and are 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. NPI provides an attractive alternative to objective Bayesian analysis, as discussed by Coolen (2006), who also discusses related literature and other NPI methods and applications. Coolen and Yan (2004) presented rc-$A_{(n)}$ as a generalization of $A_{(n)}$ for right-censored data. In comparison to $A_{(n)}$, rc-$A_{(n)}$ uses the extra assumption that, at a moment of censoring, the residual time-to-failure of a right-censored unit is exchangeable with the residual time-to-failure of all other units that have not yet failed or been censored.

We consider a life-testing experiment on units of $k \geq 2$ groups. The experiment can be terminated before all units have failed, e.g. based on a chosen stop criterion. In our approach it is irrelevant what, if any, stop criterion is used, as long as it does not hold any information on residual time-to-failure for units that have not yet failed. Let T_0 be the time at which the experiment is terminated, so data for units that have not failed before T_0 are right-censored observations at T_0. For group j, $j = 1, \ldots, k$, n_j units are included in the experiment, of which r_j units failed before (or at) T_0, with ordered failure times $0 < x_{j,1} < x_{j,2} < \ldots < x_{j,r_j} \leq T_0$ (we assume no tied observations, generalization is straightforward). To compare these k groups, we consider a further unit from each group which was not involved in this experiment, with X_{j,n_j+1} the random failure time for the further unit from group j, assumed to be exchangeable with the failure times of the n_j units of the same group included in the experiment. The assumption rc-$A_{(n_j)}$ implies that total probability mass 1 for X_{j,n_j+1} is divided over the intervals created by the observations for this group, without any further restrictions or assumptions for the probability mass within these intervals, as follows: probability mass $\frac{1}{n_j+1}$ is assigned to each of the intervals $(0, x_{j,1})$ and $(x_{j,i-1}, x_{j,i})$ for $i = 2, \ldots, r_j$, and also to the interval (x_{j,r_j}, ∞); the remaining probability mass $\frac{n_j - r_j}{n_j+1}$ is assigned to the interval (T_0, ∞). Following Coolen and van der Laan (2001), we compare these k groups by considering the random failure times of such further units, one for each group. We restrict attention here to the events $X_{l,n_l+1} = \max_{1 \leq j \leq k} X_{j,n_j+1}$, for $l = 1, \ldots, k$. These assignments of probability masses to intervals formed by the observations and T_0 do not lead to precise probabilities for events of interest, but, with the additional assumption that the k groups are fully independent, optimal bounds can be derived by shifting the probability masses to either the right- or left-end points of each interval (Maturi et al (2008)). These optimal bounds are the lower probabilities ($l = 1, \ldots, k$)

$$\begin{aligned}\underline{P}^{(l)} &= \underline{P}(X_{l,n_l+1} = \max_{1\le j\le k} X_{j,n_j+1}) \\ &= \frac{1}{\prod_{j=1}^{k}(n_j+1)} \left\{ \sum_{i_l=1}^{r_l} \prod_{\substack{j=1\\ j\neq l}}^{k} \sum_{i_j=1}^{r_j} 1\{x_{j,i_j} < x_{l,i_l}\} + (n_l - r_l) \prod_{\substack{j=1\\ j\neq l}}^{k} r_j \right\}\end{aligned}$$

and the upper probabilities ($l = 1, \ldots, k$)

$$\begin{aligned}\overline{P}^{(l)} &= \overline{P}(X_{l,n_l+1} = \max_{1\le j\le k} X_{j,n_j+1}) \\ &= \frac{1}{\prod_{j=1}^{k}(n_j+1)} \sum_{i_l=1}^{r_l} \prod_{\substack{j=1\\ j\neq l}}^{k} \left(1 + \sum_{i_j=1}^{r_j} 1\{x_{j,i_j} < x_{l,i_l}\} \right) + \frac{n_l - r_l + 1}{n_l + 1}\end{aligned}$$

T_0 influences these lower and upper probabilities (only) through the r_j. The proofs of these lower and upper probabilities will be presented elsewhere (Maturi et al (2008)), together with similar results for more general events including selection of subsets of groups as was done by Coolen and van der Laan (2001), who only considered complete experiments. In Section 3, we illustrate these lower and upper probabilities via an example. Before that, we state some special cases (Maturi et al (2008)).

If $r_l = 0$, so the experiment is terminated before the first failure of group l is observed, then

$$\underline{P}^{(l)} = \frac{n_l}{\prod_{j=1}^{k}(n_j+1)} \prod_{\substack{j=1\\ j\neq l}}^{k} r_j \quad , \quad \overline{P}^{(l)} = 1$$

This upper probability is equal to one, which can be interpreted as representing that one cannot exclude, on the basis of only the data from this experiment, the possibility that units from group l would never fail. If $r_j = 0$, for all $j \neq l$, so the experiment is terminated before the first failure of each group $j \neq l$ is observed, then

$$\underline{P}^{(l)} = 0 \ , \ \ \overline{P}^{(l)} = \frac{r_l}{\prod_{j=1}^{k}(n_j+1)} + \frac{n_l - r_l + 1}{n_l + 1}$$

This lower probability is zero, representing the possibility that units of all groups other than group l would never fail. If the experiment is terminated before a single unit, of any group, has failed, then $\underline{P}^{(l)} = 0$ and $\overline{P}^{(l)} = 1$ for all groups. These extreme examples illustrate an attractive feature of the use of lower and upper probabilities in quantifying the strength of statistical information, in an intuitive manner that is not possible with precise probabilities.

If T_0 increases, $\underline{P}^{(l)}$ never decreases and $\overline{P}^{(l)}$ never increases, and they can only change if further failures are observed. If a further failure is observed from group l, but no further failures from any other group, then $\underline{P}^{(l)}$ does not change but $\overline{P}^{(l)}$ decreases (except if $r_j = n_j$, for all $j \neq l$, in which case it does not change). If only a further failure is observed from another group j^* (so $j^* \neq l$), then $\overline{P}^{(l)}$ does not change but $\underline{P}^{(l)}$ increases (except if $r_l = n_l$ or if at least one $r_j = 0$ for $j \neq l, j^*$, in which cases it does not change). This is in line with the possible interpretation of lower (upper) probabilities as representing the evidence in favour of (against) the event of interest.

3 Example

To illustrate this NPI method for comparison of k groups, we use data from Coolen and van der Laan (2001), presented in Table 1. We interpret these data as failure times of units from $k = 4$ different groups, with $n_1 = 20$, $n_2 = 18$, $n_3 = 15$ and $n_4 = 3$. In addition to these data, we assume rc-$A_{(n_j)}$ per group (Coolen and Yan (2004)), and that these 4 groups are fully independent. We have computed lower probabilities $\underline{P}^{(l)}$ and upper probabilities $\overline{P}^{(l)}$, for $l = 1, \ldots, 4$, as functions of T_0. For some ranges of values of T_0, these lower and upper probabilities are presented in Table 2. For values of T_0 covering the range from the minimum (4.50) to the maximum (9.16) observations, $\underline{P}^{(l)}$ and $\overline{P}^{(l)}$ are shown, per group l, in Figure 1. This example illustrates the effect of termination of this experiment at time T_0, in which case failure times greater than T_0 are not available, but are right-censored observations at T_0. The special cases discussed in Section 2 are clearly illustrated.

Group	
1	5.01 5.04 5.60 5.78 6.43 6.53 6.96 7.00 7.21 7.58 8.12 8.26 8.27 8.34 8.62 8.66 8.91 8.94 9.05 9.16
2	4.50 4.86 5.10 5.15 5.17 5.34 5.99 6.18 6.72 7.39 7.44 7.46 7.47 7.76 8.38 8.42 8.52 8.81
3	6.84 6.91 7.22 7.24 7.25 7.35 7.55 7.62 7.69 7.98 7.99 8.04 8.08 8.18 8.97
4	4.71 8.20 9.03

Table 1. Failure times for 4 groups

For $T_0 \geq 7.55$, $\underline{P}^{(4)} > \underline{P}^{(j)}$ and $\overline{P}^{(4)} > \overline{P}^{(j)}$ for $j = 1, 2, 3$, this might suggest that group 4 is the most likely group to lead to maximum failure time if one further unit is used from each group. However, for $T_0 \leq 7.24$,

T_0	r_1	r_2	r_3	r_4	$\underline{P}^{(1)}$	$\overline{P}^{(1)}$	$\underline{P}^{(2)}$	$\overline{P}^{(2)}$	$\underline{P}^{(3)}$	$\overline{P}^{(3)}$	$\underline{P}^{(4)}$	$\overline{P}^{(4)}$
$[4.86, 5.01)$	0	2	0	1	0	1	0	0.8949	0	1	0	0.7501
$[7.21, 7.22)$	9	9	3	1	0.0137	0.5815	0.0095	0.5287	0.0455	0.8313	0.0190	0.7501
$[7.22, 7.24)$	9	9	4	1	0.0176	0.5815	0.0127	0.5287	0.0455	0.7766	0.0254	0.7501
$[7.47, 7.55)$	9	13	6	1	0.0357	0.5815	0.0190	0.3401	0.0582	0.6673	0.0550	0.7501
$[7.55, 7.58)$	9	13	7	1	0.0413	0.5815	0.0208	0.3401	0.0582	0.6158	0.0641	0.7501
$[7.99, 8.04)$	10	14	11	1	0.0660	0.5426	0.0292	0.2961	0.0646	0.4157	0.1206	0.7501
$[8.66, 8.81)$	16	17	14	2	0.1707	0.4161	0.0756	0.2175	0.0809	0.2682	0.2336	0.6058
$[8.81, 8.91)$	16	18	14	2	0.1751	0.4161	0.0756	0.1948	0.0822	0.2682	0.2423	0.6058
$[8.91, 8.94)$	17	18	14	2	0.1751	0.4019	0.0756	0.1948	0.0836	0.2682	0.2522	0.6058
$[9.05, 9.16)$	19	18	15	3	0.1991	0.3878	0.0756	0.1948	0.0850	0.2481	0.2747	0.5820
$[9.16, \infty)$	20	18	15	3	0.1991	0.3878	0.0756	0.1948	0.0850	0.2481	0.2747	0.5820

Table 2. Lower and upper probabilities for some T_0

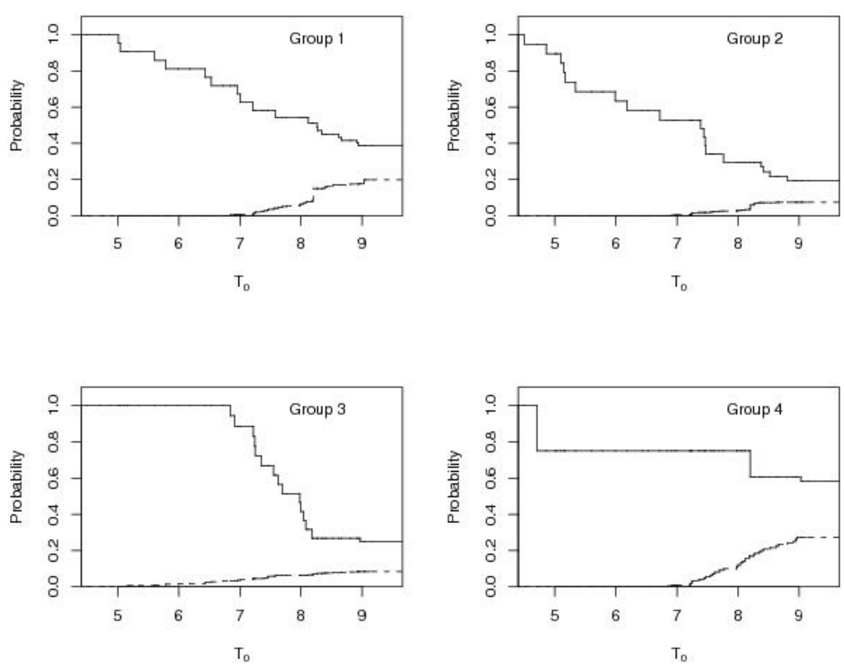

Fig. 1. The best group: lower and upper probabilities

group 3 has the greatest corresponding lower and upper probabilities, so if the experiment were stopped before time 7.24 then there would be an argument to select group 3, if the overall aim were to select a single group. If, for one group l, $\underline{P}^{(l)} > \overline{P}^{(j)}$ for all $j \neq l$, this would be a strong indication that group l is the best, this situation does not occur here. The imprecision $\overline{P}^{(l)} - \underline{P}^{(l)}$, for $l = 1, \ldots, 4$, is a decreasing function of T_0, so there is always a possible benefit from continuing the experiment (until all units have failed). For larger values of T_0, such that most units have failed, group 4 has most imprecision remaining, reflecting that there are only few observations for group 4. In Maturi et al (2008) we also consider these inferences with group 4 removed, which leads to substantial reduction in imprecision for groups 1-3.

4 Concluding remarks

This short paper only gives a little insight into the attractive opportunities provided by the use of lower and upper probabilities to quantify uncertainty. In the application presented, the influence of early termination of an experiment is obvious, as more failure observations lead to less imprecision and hence to better quality input for decisions on selection of one of the groups. A more detailed presentation (Maturi et al (2008)) will include discussion on the use of such lower and upper probabilities to support selection decisions. One possible argument against the use of lower and upper probabilities is that they, apparently, can lead to 'indecision'. We strongly feel that imprecision merely points out the lack of sufficient evidence from data if the method does not clearly provide a full ranking of possible decisions, and hence that, in such situations, one must either get more data (e.g. by continuing the experiment or repeating it with more units) or add further assumptions or information from other sources. Although we present this NPI approach as an attractive method for comparison of groups of units, we would not dismiss other methods, for example classical precedence testing (Balakrishnan and Ng (2006)). We see a real strength in using several methods simultaneously, and carefully studying the corresponding inferences, we discuss this further in Maturi et al (2008).

References

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

Balakrishnan, N. and Ng, H.K.T. (2006): *Precedence-Type Tests and Applications.* John Wiley & Sons.

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

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

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

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

Maturi, T.A., Coolen-Schrijner, P. and Coolen, F.P.A. (2008): Nonparametric Predictive Selection with Early Termination of Experiments. *In preparation.*