Adaptive Age Replacement Strategies based on Nonparametric Predictive Inference

P. Coolen-Schrijner, F.P.A. Coolen

Department of Mathematical Sciences

University of Durham

Durham, DH1 3LE, England

Abstract

We consider an age replacement problem using nonparametric predictive inference (NPI) for the lifetime of a future unit. Based on n observed failure times, NPI provides lower and upper bounds for the survival function for a future lifetime X_{n+1}, which are lower and upper survival functions in the theory of interval probability, and which lead to upper and lower cost functions, respectively, for age replacement based on the renewal reward theorem. Optimal age replacement times for X_{n+1} follow by minimising these cost functions.

Although the renewal reward theorem implicitly assumes that the corresponding optimal strategy will be used for a long period, we study the effect on this strategy when the observed value for X_{n+1}, which is either an observed failure time or a right-censored observation, becomes available. This is possible due to the fully adaptive nature of our nonparametric approach, and the next optimal strategy will be for X_{n+2}. We compare the optimal strategies for X_{n+1} and X_{n+2} both analytically and via simulation studies.

Our NPI-based approach is fully adaptive to the data, to which it adds only few structural assumptions. We discuss the possible use of this approach, and indeed the wider importance of the conclusions of this study to situations where one wishes to combine the statistical aspects of estimating a lifetime distribution with the more traditional Operational Research approach of determining optimal replacement strategies for lifetime distributions which are assumed to be known.

Keywords: $A_{(n)}$; Adaptive age replacement strategies; Nonparametric predictive inference; Renewal reward theorem.

Introduction

Age replacement strategies, where a unit is replaced upon failure or on reaching a predetermined age, whichever occurs first, provide simple and intuitively attractive replacement guidelines for technical units. Within theory of stochastic processes, the optimal preventive replacement age, in the sense of leading to minimal expected costs per unit of time when the strategy is used for a sequence of similar units over a long period of time, is derived by application of the renewal reward theorem, see e.g. Barlow and Proschan [1]. In practice, this procedure is also used even though one realises that the resulting optimal strategy may only be used for a few such cycles, for example because the unit would normally undergo some technical updates within reasonable period of time, or one wishes to keep the option open to change the policy in light of new information that may occur during the process. Attention to age replacement has predominantly been based on a classical Operational Research perspective, where the probability distribution for the lifetime of the unit is assumed to be known. Recently, age replacement has been considered from Bayesian perspective [2], allowing the assumed parametric lifetime distribution to be updated, within the Bayesian framework, when new data from the process become available. Such procedures can be called 'adaptive', in the sense that the optimal preventive replacement time may change over time. To counter the fact that such changing optimal preventive replacement times are in conflict with the assumption of long-term use of the same strategy, on which the renewal reward theorem is based, Mazzuchi and Soyer [2] replaced the renewal reward criterion by minimisation of expected costs per unit of time over a single cycle. The use of Bayesian statistics for such replacement problems has been advocated by several authors, not only because of the opportunity to update the prior distribution for the parameters of the lifetime distribution, within the assumed parametric family of distributions, but also because of the subjective probabilistic underpinnings for Bayesian statistics. To deal with scarce information, as may regularly be the case in practice, one could attempt to entirely base replacement decisions on expert judgements, via elicitation of a lifetime distribution for the unit [3, 4].

As an alternative to these approaches, we combine the OR-based decision making aspects for age replacement with nonparametric predictive inference for the lifetime distribution, making rather minimal structural assumptions for this distribution, which enables study of the way that resulting optimal replacement strategies adapt to available data. Indeed, this approach implicitly assumes the presence of failure data, which we assume throughout the paper and briefly discuss in the concluding section. We still base the optimisation criterion for the preventive replacement time on the renewal reward theorem. While we do not suppose that the same policy will be used in the long run, and thus cannot appeal to the renewal reward theorem to justify the use of the long run cost per unit time as the decision criterion, there is no doubt that, intuitively, the expected cost of a cycle over the expected cycle length is a reasonable decision criterion. Using it has the advantage that it allows study of the manner in which optimal replacement times according to this criterion adapt to data, and indeed to new process data once the procedure is in place, when no distributional assumptions are added to the data. This work further develops the work in Coolen-Schrijner and Coolen [5], where the basic results for such age replacement based on nonparametric predictive inference (NPI) were presented. There, on the basis of n failure times, NPI was used to derive probabilities for X_{n+1}, the random lifetime of the next unit to be used in the process, and for which the age replacement strategy is going to be adopted. In this paper, we study how the optimal age replacement strategy changes depending on the observation of X_{n+1} under the applied optimal age replacement rule, so we focus on X_{n+2}, using the information of the first n failure times and the information from unit $n+1$, which can be either a failure before, or a right-censored observation at, the optimal age replacement time. We derive analytical results for the optimal age replacement time for this unit $n+2$, and analyse it further via simulations.

Coolen and Newby [6] first used NPI for replacement problems, focussing on corrective replacement. This work was followed by Coolen and Coolen-Schrijner [7], who used NPI for preventive replacement decisions in case of continuous condition monitoring, with deterioration of a unit modelled via a known number of states for its condition. Recently, next to a variety of statistical applications of NPI, its use has also been suggested for problems in reliability [8] and queueing [9]. Further progress on adaptive replacement strategies based on NPI is now possible due to recent development of NPI for data sets including right-censored observations [10], which is also essential for the work in this paper.

The outline of this paper is as follows. First, we briefly provide the necessary details of NPI, referring to the literature for justifications and further discussions. Then, we briefly summarize the main results from Coolen-Schrijner and Coolen [5], which we also use in this paper. In the main part of this paper, we present theoretical results on the optimal age replacement time for X_{n+2}, focussing on the way it adapts with regard to information on X_{n+1} as coming from the process, and further insights are provided by simulation studies. Finally, we briefly comment on the main conclusions of this work, and its relevance beyond NPI when one wishes to combine statistical methods for learning from data directly with Operational Research methods for decision making, and we point out some topics for future research. The proofs of the analytical results are presented in an appendix.

Nonparametric predictive inference

Nonparametric predictive inference (NPI) is based on Hill's assumption $A_{(n)}$, which is suitable for probabilistic predictions in case one wishes to add very little extra information to observed data. Denoting n ordered observed lifetimes by $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$, $A_{(n)}$ [11, 12] specifies direct probabilities [13] for a future lifetime, X_{n+1}, by

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

for $j = 0, \ldots, n$, where, for ease of notation, $x_{(0)} = 0$ and $x_{(n+1)} = \infty$, or $x_{(n+1)} = r$ if we can safely assume a finite upper bound r for the support of X_{n+1} (note that the latter is not an observed value of X_{n+1}). $A_{(n)}$ is a post-data assumption related to exchangeability [14]. For a further discussion of $A_{(n)}$ and an overview of related work, see Hill [15]. $A_{(n)}$-based inference is predictive and nonparametric, and suitable if there is hardly any knowledge about the random quantities of interest, other than the n observations, or if one explicitly does not want to use such information, e.g. if one wants to study the effects of assumptions underlying statistical models. Clearly, $A_{(n)}$ does not provide precise probabilities for all possible events of interest, as it only partially defines a probability distribution for X_{n+1}. However, it does provide optimal bounds for all probabilities of interest involving X_{n+1}, by application of De Finetti's 'fundamental theorem of probability' [14]. Such bounds are lower and upper probabilities within the theory of interval probability [16, 17, 18], Augustin and Coolen [12] prove strong consistency properties of NPI within this theory, which justifies its use from statistical perspective.

To enable study of adaptive behaviour based on process data in this paper, we also need a generalized version of $A_{(n)}$ which allows right-censored data. The general theory for this has recently been developed by Coolen and Yan [10], to which paper we refer for technical details and justification. The generalization is called 'right-censoring $A_{(n)}$', denoted by rc-$A_{(n)}$, for the purpose of this paper we only need two special cases. Suppose that we have the n observed failure times, denoted as before. Now assume that we also have an observation of X_{n+1}, but a right-censoring at a previously observed $x_{(k)}$, for a $k \in \{1, \ldots, n\}$, so the only information about unit $n+1$ is $X_{n+1} > x_{(k)}$. Then the assumption rc-$A_{(n+1)}$ provides the following partial specification of a probability distribution for X_{n+2}, the random lifetime of the next unit to be used:

$$P(X_{n+2} \in (x_{(j)}, x_{(j+1)})) = \frac{1}{n+2}, \quad \text{for } 0 \leq j < k,$$
$$P(X_{n+2} \in (x_{(j)}, x_{(j+1)})) = \frac{n+2-k}{(n+2)(n+1-k)}, \quad \text{for } k \leq j \leq n.$$

As this is the form of rc-$A_{(n)}$ most used in this paper, we have presented it explicitly. We do, however, prove some results on situations where several further lifetimes are all right-censored at the same value $x_{(k)}$. For that special situation, we use the following generalization of the above specification. Suppose that units $n+1$ to $n+m$, with $m \geq 1$, have all been right-censored at the same previously observed $x_{(k)}$. Then the assumption rc-$A_{(n+m)}$ provides the following partial specification of a probability distribution for X_{n+m+1}:

$$P(X_{n+m+1} \in (x_{(j)}, x_{(j+1)})) = \frac{1}{n+m+1}, \quad \text{for } 0 \leq j < k,$$
$$P(X_{n+m+1} \in (x_{(j)}, x_{(j+1)})) = \frac{n+m+1-k}{(n+m+1)(n+1-k)}, \quad \text{for } k \leq j \leq n.$$

In this paper, these specifications will be used to provide lower and upper predictive survival functions when we study the adaptive nature of our optimal age replacement strategies from NPI perspective. We also study the case where the observation for unit $n+1$ is actually a failure time, which occurs if that unit fails before the predetermined age replacement time. For such situations, a partially specified probability distribution for X_{n+2} follows directly from the assumption $A_{(n+1)}$ with all $n+1$ observed failure times.

The general form of rc-$A_{(n)}$ as presented by Coolen and Yan [10] allows generalization of our method for any data set consisting of failure times and right-censored observations. However, general analytical results are harder to achieve, and difference between lower and upper cost functions can become large in case of many censored observations. Other types of censoring are difficult to deal with given the current state of development of the statistical theory of NPI. Coolen and Yan [10] briefly indicate how NPI can deal with left-censored data, but that method can only be used if there are no right-censored data available.

There is an intuitive link between NPI and nonparametric likelihood estimation, for example one can consider rc-$A_{(n)}$, used explicitly for probabilistic prediction for a future observation, as a predictive alternative to the well-known Product-Limit estimator by Kaplan and Meier [19], this is explained in detail by Coolen and Yan [10].

We should briefly mention here that, although the above forms of $A_{(n)}$ and rc-$A_{(n)}$ appear not to allow ties between the n observed failure times, such situations can be dealt with rather straightforwardly [10], namely by regarding tied observations as if they are very close but different, and then letting the difference decrease to zero. In practice this is of little relevance, as long as the data are recorded in sufficient detail to make tied observations very unlikely.

Optimal age replacement for X_{n+1}

In this section, we formulate the basic age replacement problem studied in this paper, and summarize the key results from Coolen-Schrijner and Coolen [5] which form the basis for the study reported in this paper.

We consider an age replacement problem in which an item is replaced upon failure ('corrective replacement') at cost $c_2 > 0$, or upon reaching the age T ('preventive replacement') at cost

$c_1 > 0$, whichever occurs first. We restrict attention to $c_1 < c_2$, a logical requirement to make preventive replacement possibly worthwhile. In the classical setting, a unit's lifetime is represented by a random quantity, say X, assumed to belong to a population of independent and identically distributed random quantities. For this case, we denote the survival function for X by $S(x) = P(X > x)$.

The aim is to determine the optimal preventive replacement age T. One possible method is to base the criterion function on renewal reward theory (see e.g. Barlow and Proschan [1]), in which case the optimal replacement age follows from minimising the long-run average costs per unit time, which, by applying the renewal reward theorem, are equal to the expected costs per cycle divided by the expected length of a cycle, where a cycle is the time between two consecutive replacements. This leads to long-run average costs per unit time, $C(T)$, given by

$$C(T) = \frac{c_2 - S(T)(c_2 - c_1)}{\int_0^T S(x)dx}. \tag{1}$$

Coolen-Schrijner and Coolen [5] do not assume a known survival function $S(x)$, and not even restrict to a parametric family of underlying distributions [2], but use NPI to derive lower and upper survival functions for X_{n+1}, on the basis of n observed failure times, and study the optimal replacement ages according to these lower and upper survival functions. An attractive feature of this approach is that the replacement problem is directly formulated in terms of the lifetime random quantity of the next unit. In addition, using a nonparametric statistical approach reduces to almost minimal the influence of modelling assumptions, which are often hidden, for example via assumed parametric families of distributions. In this paper, we will build on the previous work [5] by studying how the optimal replacement ages adapt when the process moves on to the next stage, that is when information from unit $n+1$ becomes available, assuming that the optimal NPI replacement strategy had indeed been used for this unit. In order to present the results of this study in the next section, we first summarize the key results from the previous paper [5].

The assumption $A_{(n)}$ assigns probability masses for X_{n+1}, the lifetime of the next unit to be used, to the open intervals $(x_{(j)}, x_{(j+1)})$, $j = 0, \ldots, n$, created by n observed failure times, but it does not put any further restrictions on the distributions of the probability masses within each such interval. This immediately leads to precise values for the corresponding survival function of X_{n+1} at the points $x_{(j)}$, namely

$$S_{X_{n+1}}(x_{(j)}) = \frac{n+1-j}{n+1} \text{ for } j = 0, \ldots, n+1,$$

but it does not provide precise values for the survival function at other times without adding further assumptions. However, optimal lower and upper bounds for this survival function, consistent with the probability assessment according to $A_{(n)}$, are easily derived. The maximum lower bound, $\underline{S}(x)$, is obtained by shifting the probability mass in the interval in which x lies to the left end-point of the interval, while the minimum upper bound, $\overline{S}(x)$, is obtained by shifting the probability mass in the interval in which x lies to the right end-point of the interval. These bounds for the survival function of X_{n+1} are

$$\underline{S}_{X_{n+1}}(x) = S_{X_{n+1}}(x_{(j+1)}) = \frac{n-j}{n+1} \text{ for } x \in (x_{(j)}, x_{(j+1)}),\ j = 0, \dots, n, \tag{2}$$

$$\overline{S}_{X_{n+1}}(x) = S_{X_{n+1}}(x_{(j)}) = \frac{n+1-j}{n+1} \text{ for } x \in (x_{(j)}, x_{(j+1)}),\ j = 0, \dots, n. \tag{3}$$

We call $\underline{S}(\cdot)$ and $\overline{S}(\cdot)$ lower and upper survival functions, respectively [8], they indeed provide lower and upper probabilities within the theory of interval probability [16, 17, 18].

The cost function $C(T)$, as given by (1), is decreasing as function of $S(\cdot)$, in the sense that $C(T)$ decreases if $S(x)$ increases for $x \in (0, T]$. This implies that the above lower and upper $A_{(n)}$-based survival functions straightforwardly lead to bounds for this cost function, corresponding to all such cost functions for possible survival functions between $\underline{S}(\cdot)$ and $\overline{S}(\cdot)$. The maximum lower bound for $C(T)$, which we call the lower cost function for X_{n+1} and denote by $\underline{C}_{X_{n+1}}(T)$, is derived for the upper survival function $\overline{S}_{X_{n+1}}(x)$ for $x \in (0, T]$, whereas the minimum upper bound, called the upper cost function for X_{n+1} and denoted by $\overline{C}_{X_{n+1}}(T)$, is derived for the lower survival function $\underline{S}_{X_{n+1}}(x)$ for $x \in (0, T]$. These lower and upper cost functions are [5]

$$\underline{C}_{X_{n+1}}(x_{(j)}) = \frac{jc_2 + (n+1-j)c_1}{(n+1-j)x_{(j)} + \sum_{l=1}^{j} x_{(l)}}, \quad j = 1, \dots, n+1, \tag{4}$$

$$\underline{C}_{X_{n+1}}(T) = \frac{jc_2 + (n+1-j)c_1}{(n+1-j)T + \sum_{l=1}^{j} x_{(l)}}, \quad \text{for } T \in (x_{(j)}, x_{(j+1)}),\ j = 0, \dots, n, \tag{5}$$

$$\overline{C}_{X_{n+1}}(x_{(j)}) = \frac{jc_2 + (n+1-j)c_1}{(n+1-j)x_{(j)} + \sum_{l=1}^{j-1} x_{(l)}}, \quad j = 1, \dots, n+1, \tag{6}$$

$$\overline{C}_{X_{n+1}}(T) = \frac{(j+1)c_2 + (n-j)c_1}{(n-j)T + \sum_{l=1}^{j} x_{(l)}}, \quad \text{for } T \in (x_{(j)}, x_{(j+1)}),\ j = 0, \dots, n. \tag{7}$$

Because $\underline{S}_{X_{n+1}}(T) = 0$ for $T > x_{(n)}$, $\overline{C}_{X_{n+1}}(\cdot)$ is constant beyond $x_{(n)}$. Moreover, $\overline{C}_{X_{n+1}}(T) > \overline{C}_{X_{n+1}}(x_{(n)})$ for $T > x_{(n)}$. Hence, when determining the optimal age replacement time in the sense of minimising the upper cost function, we do not have to consider replacement times larger than $x_{(n)}$.

We should emphasize that there is no simple relation between the optimum replacement times corresponding to the lower and upper cost functions, for the same data set, nor indeed are there general relations when also considering optimum replacement times corresponding to survival functions between the lower and upper survival functions.

In case we do not replace preventively, we have

$$\overline{C}_{X_{n+1}}(x_{(n+1)}) = \frac{c_2}{\frac{1}{n+1}\sum_{l=1}^{n} x_{(l)}} \quad \text{and} \quad \underline{C}_{X_{n+1}}(x_{(n+1)}) = \frac{c_2}{\frac{1}{n+1}\sum_{l=1}^{n+1} x_{(l)}},$$

where the denominator of the lower cost function is only finite if we assume a known upper bound $x_{(n+1)} = r$ for the support of X_{n+1}. This is a minor complication when determining the optimal age replacement time in the sense of minimising the lower cost function for X_{n+1}, to prevent complications we mostly restrict attention to $T \in (0, x_{(n)}]$, but return briefly to this issue later.

Throughout this paper, notation like $x_{(j)}^-$ is to be interpreted as 'just before $x_{(j)}$', such that the adherent probability mass [14] to the left of $x_{(j)}$ is considered to be to the right of $x_{(j)}^-$ in the extreme situation related to the location of the probability masses corresponding to $\overline{S}_{X_{n+1}}(\cdot)$.

In [5] it was shown that $\overline{C}_{X_{n+1}}(\cdot)$ and $\underline{C}_{X_{n+1}}(\cdot)$ are both discontinuous at the observed failure times $x_{(j)}$, but in between these failure times they are continuous and strictly decreasing. It followed that the global minimum of $\overline{C}_{X_{n+1}}(\cdot)$ is assumed in one of the points $x_{(j)}$, $j = 1, \ldots, n$, and the minimum of $\underline{C}_{X_{n+1}}(\cdot)$ on $(0, x_{(n)}]$ is assumed in one of the $x_{(j)}^-$, $j = 1, \ldots, n$. The values of $\overline{C}_{X_{n+1}}(x_{(j)})$ are given in (6), while the values of $\underline{C}_{X_{n+1}}(x_{(j)}^-)$ are given by

$$\underline{C}_{X_{n+1}}(x_{(j)}^-) = \frac{(j-1)c_2 + (n+2-j)c_1}{(n+2-j)x_{(j)} + \sum_{l=1}^{j-1} x_{(l)}}. \tag{8}$$

If we assume a known upper bound r for the support of X_{n+1}, $\underline{C}_{X_{n+1}}(\cdot)$ is also strictly decreasing on $(x_{(n)}, r)$, so we must also consider $\underline{C}_{X_{n+1}}(r^-)$ for finding the global minimum on $(0, r]$. If this minimum is attained in r^-, then it is better not to replace preventively. Alternatively, we can calculate a critical value r^* such that if you think that an upper bound for the support of X_{n+1} is smaller than r^*, then the optimal age replacement time for $\underline{C}_{X_{n+1}}(\cdot)$ over the interval $(0, r]$ is the optimal T that was derived when attention was restricted to $T \in (0, x_{(n)}]$. But if you think that an upper bound for the support of X_{n+1} is larger than r^*, then it is better not to replace unit $n+1$ preventively. Before we present the new results on the adaptive nature of our NPI-based age replacement procedure in the next section, we illustrate NPI-based age replacement for unit $n+1$ in a small example, which we will also use later in this paper.

Example 1 Suppose we have 8 lifetimes: 1, 2, 5, 7, 8, 9, 12, 20. Each preventive replacement costs $c_1 = 1$, while each corrective replacement costs $c_2 = 10$. We would like to find the optimal age replacement times T_9^* for X_9 in the sense of minimising the upper and lower cost functions, using the renewal argument, $A_{(8)}$ and these data. The upper cost function needs only to be computed for $T = x_{(j)}$:

	$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$
	1	2	5	7	8	9	12	20
$\overline{C}_{X_9}(x_{(i)})$	9/4	27/15	12/11	45/43	54/47	63/50	9/7	81/64

so the optimal age replacement time T_9^* for X_9 in the sense of minimising the upper cost function is 7, with corresponding upper costs 45/43. The lower cost function needs only to be computed for $T = x_{(j)}^-$:

	$x_{(1)}^-$	$x_{(2)}^-$	$x_{(3)}^-$	$x_{(4)}^-$	$x_{(5)}^-$	$x_{(6)}^-$	$x_{(7)}^-$	$x_{(8)}^-$
	1^-	2^-	5^-	7^-	8^-	9^-	12^-	20^-
$\underline{C}_{X_9}(x_{(i)}^-)$	1	18/17	27/38	18/25	9/11	54/59	63/68	6/7

so the optimal age replacement time T_9^* for X_9 over the interval $(0, 20]$ in the sense of minimising the lower cost function is 5^-, with corresponding lower costs 27/38. The critical value of r^* is 50. Figure 1 is a plot of these upper and lower cost functions.

Coolen-Schrijner and Coolen [5] performed simulation studies to see how well this procedure performed. These showed that there was no clear relation between the optima resulting from minimising the lower and upper cost functions for the same data set, as these were frequently the same, but also differed with no fixed order. A further conclusion was that already for fairly small data sets ($n = 10$) the NPI-based optima were reasonably close to theoretical optima as related to chosen distributions for simulation. Although there was quite some variation in the NPI-based optima over different data sets, this approach often performed better than when using the classical approach with wrongly assumed lifetime distributions. Of course, the larger the data set, the better the NPI-based method performs. This was reason to also study the way that this NPI-based age replacement method adapts to the process information on unit $n+1$, assuming that the optimal NPI-based replacement age according to the criterion chosen, i.e. minimal upper or lower cost function, has been applied for this unit. The analytical results of this study are presented in the next section, thereafter further insights are provided via a simulation study.

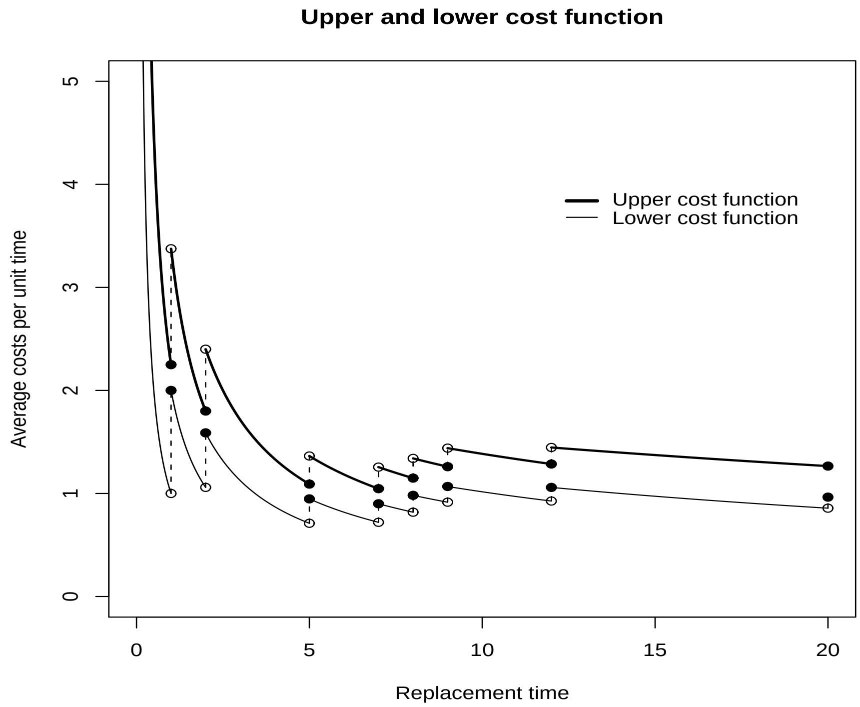

Figure 1: Lower and upper cost functions for Example 1

Adaptive age replacement strategy

As minimisation of the upper and lower cost functions with regard to age replacement of unit $n+1$, using NPI and renewal reward theory, may lead to two different optimal replacement times, we denote these optima by $T^*_{up,n+1}$ and $T^{-*}_{low,n+1}$, respectively. Early in this section, we present some general theory for which it is not relevant to make the distinction, in which case we use T^*_{n+1} as generic notation for an optimal age replacement time that was used for unit $n+1$. We now consider the effect of using such an optimal policy, and the resulting information about the lifetime X_{n+1} of unit $n+1$, on the optimal NPI-based age replacement policy for unit $n+2$, with random lifetime X_{n+2}. Although our cost functions are based on the renewal reward theorem, which implicitly assumes that an age replacement strategy, once determined, is used for a long period (i.e. many cycles), we do think it is interesting to study how optimal replacement times would actually adapt to new data from the process, because the assumption of a constant replacement time seems more defendable from classical OR perspective, where a known lifetime distribution is assumed, than when one attempts to use the process information to infer such a lifetime distribution. If, however, our study would reveal that the optimal strategy is unlikely to change much on the basis of the new observation, this would suggest that the use of this criterion is not unreasonable even when one wishes to use an adaptive method like ours.

Under the optimal age replacement strategy, the observation for X_{n+1} is either a failure time less than T^*_{n+1}, in which case the unit was replaced correctively, or a right-censored observation at T^*_{n+1}, in which case the unit was replaced preventively. We present our study for these two cases separately, where the first is reasonably straightforward but the second case requires a lot more attention and leads to particularly interesting results.

I. Unit $n+1$ fails before T^*_{n+1}.

If unit $n+1$ fails before T^*_{n+1}, we can directly apply the results presented in the previous section [5] on the $n+1$ failure times, with $A_{(n)}$ replaced by $A_{(n+1)}$, to obtain the optimal age replacement time for unit $n+2$. Here, it seems not only natural to use the same cost function, upper or lower, to determine the consecutive optimum replacement times, but this also allows us to study the effect of such a new observation on the NPI-based optimum replacement times. In the next section we study this effect via simulations, we first briefly illustrate this procedure via an example.

Example 1 (continued) Suppose that, using the optimal replacement time for unit $n+1$ according to the upper cost function, which was equal to 7, unit 9 fails at time 4. This adds an observed failure time to the data set, and theory presented in the previous section now implies that the optimum replacement time for unit 10 is found by taking the minimal value of this cost function at the observed failure times, these values are given below.

	$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$	$x_{(9)}$
	1	2	4	5	7	8	9	12	20
$\overline{C}_{X_{10}}(x_{(i)})$	19/9	28/17	37/31	46/37	55/47	64/51	73/54	82/60	91/68

Hence, the optimal age replacement time $T^*_{up,10}$ for X_{10}, is still equal to 7, so has not changed in this case. Simulations will later show that it is not always the case that this optimum does not change, and that it can change in both directions if a failure time has been observed.

II. Unit $n+1$ is preventively replaced at T^*_{n+1}.

The case where unit $n+1$ is preventively replaced at T^*_{n+1} is of more interest, first of all because in practical situations where age replacement with a finite replacement time is cost effective, one tends to have relatively many preventive replacements, so this situation tends to occur more frequently than corrective replacements. Secondly, the mathematical study of the effect of such information on the optimal replacement age for unit $n+2$ leads to some interesting results, and uses recently developed statistical methods for NPI with right-censored data, as reviewed earlier in this paper.

Throughout this section, let us assume that unit $n+1$ was preventively replaced at $T^*_{n+1} = x_{(k)}$, for some $k \in \{1, \ldots, n\}$, the optimum NPI-based age replacement time for that unit according to the cost function used (lower or upper). As a technical detail, we should point out that the lower cost function would actually have had its optimum in an $x^-_{(k)}$, but in practice there would be no noticeable difference between a replacement at $x^-_{(k)}$ or at $x_{(k)}$, hence we assume throughout that the replacement actually took place at an $x_{(k)}$, also when the lower cost function had been used.

When considering age replacement of the unit $n+2$ in this process, the relevant data set now consists of the n original failure times, $x_{(1)} < \ldots < x_{(n)}$, together with a right censored observation at $x_{(k)}$. For NPI based on such data, the assumption rc-$A_{(n+1)}$ [10] can be applied, leading to the probabilities for X_{n+2} as presented before in the introductory section on NPI. These predictive probabilities lead to lower and upper survival functions for the random lifetime X_{n+2}, which coincide at the observed failure times $x_{(j)}$, $j = 1, \ldots, n$,

$$\underline{S}_{X_{n+2}}(x_{(j)}) = \overline{S}_{X_{n+2}}(x_{(j)}) = \frac{n+2-j}{n+2} \text{ for } 0 \le j \le k, \tag{9}$$

$$\underline{S}_{X_{n+2}}(x_{(j)}) = \overline{S}_{X_{n+2}}(x_{(j)}) = \frac{(n+2-k)(n+1-j)}{(n+2)(n+1-k)} \text{ for } k < j \le n+1. \tag{10}$$

At all other times these functions are again the optimal bounds that are consistent with the rc-$A_{(n+1)}$-based probability specification, and are

$$\underline{S}_{X_{n+2}}(x) = \underline{S}_{X_{n+2}}(x_{(j+1)}) = \frac{n+1-j}{n+2} \text{ for } x \in (x_{(j)}, x_{(j+1)}),\ 0 \le j < k, \tag{11}$$

$$\underline{S}_{X_{n+2}}(x) = \underline{S}_{X_{n+2}}(x_{(j+1)}) = \frac{(n+2-k)(n-j)}{(n+2)(n+1-k)} \text{ for } x \in (x_{(j)}, x_{(j+1)}),\ k \le j \le n, \tag{12}$$

$$\overline{S}_{X_{n+2}}(x) = \overline{S}_{X_{n+2}}(x_{(j)}) = \frac{n+2-j}{n+2} \text{ for } x \in (x_{(j)}, x_{(j+1)}),\ 0 \le j \le k, \tag{13}$$

$$\overline{S}_{X_{n+2}}(x) = \overline{S}_{X_{n+2}}(x_{(j)}) = \frac{(n+2-k)(n+1-j)}{(n+2)(n+1-k)} \text{ for } x \in (x_{(j)}, x_{(j+1)}),\ k < j \le n. \tag{14}$$

These upper and lower survival functions are constant between event times. The upper survival function decreases at $x_{(j)}$ by the value $P(X_{n+2} \in (x_{(j-1)}, x_{(j)}))$, while the lower survival function decreases at $x_{(j)}$ by the value $P(X_{n+2} \in (x_{(j)}, x_{(j+1)}))$. The effect of a right-censoring at $x_{(k)}$ is increased difference between upper and lower survival function beyond $x_{(k)}$.

The upper cost function for X_{n+2} is given in the following lemma, and is derived by substituting the lower survival function for X_{n+2} into the cost function (1), the detailed proof is given in the appendix.

Lemma 1 The upper cost function for X_{n+2} is given by

$$\overline{C}_{X_{n+2}}(x_{(j)}) = \frac{jc_2 + (n+2-j)c_1}{(n+2-j)x_{(j)} + \sum_{l=1}^{j-1} x_{(l)}} \quad \text{for } 0 \leq j \leq k, \tag{15}$$

$$\overline{C}_{X_{n+2}}(x_{(j)}) = \frac{(nj+2j-jk-k)c_2 + (n+2-k)(n+1-j)c_1}{(n+1-k)\sum_{l=1}^{k-1} x_{(l)} + (n+2-k)\sum_{l=k}^{j-1} x_{(l)} + (n+2-k)(n+1-j)x_{(j)}} \tag{16}$$
$$\text{for } k < j \leq n+1,$$

$$\overline{C}_{X_{n+2}}(T) = \frac{(j+1)c_2 + (n+1-j)c_1}{(n+1-j)T + \sum_{l=1}^{j} x_{(l)}} \quad \text{for } T \in (x_{(j)}, x_{(j+1)}),\ 0 \leq j < k, \tag{17}$$

$$\overline{C}_{X_{n+2}}(T) = \frac{(n+2-2k+nj+2j-kj)c_2 + (n+2-k)(n-j)c_1}{(n+1-k)\sum_{l=1}^{k-1} x_{(l)} + (n+2-k)\sum_{l=k}^{j} x_{(l)} + (n+2-k)(n-j)T} \tag{18}$$
$$\text{for } T \in (x_{(j)}, x_{(j+1)}),\ k \leq j \leq n.$$

The lower cost function for X_{n+2} is given in the following lemma, and is obtained by substituting the upper survival function in the cost function (1). We omit the proof of the next lemma as it is similar to the proof of Lemma 1.

Lemma 2 The lower cost function for X_{n+2} is given by

$$\underline{C}_{X_{n+2}}(x_{(j)}) = \frac{jc_2 + (n+2-j)c_1}{(n+2-j)x_{(j)} + \sum_{l=1}^{j} x_{(l)}} \quad \text{for } 0 \leq j \leq k, \tag{19}$$

$$\underline{C}_{X_{n+2}}(x_{(j)}) = \frac{(nj+2j-jk-k)c_2 + (n+2-k)(n+1-j)c_1}{(n+1-k)\sum_{l=1}^{k} x_{(l)} + (n+2-k)\sum_{l=k+1}^{j-1} x_{(l)} + (n+2-k)(n+2-j)x_{(j)}} \tag{20}$$
$$\text{for } k < j \leq n+1,$$

$$\underline{C}_{X_{n+2}}(T) = \frac{jc_2 + (n+2-j)c_1}{(n+2-j)T + \sum_{l=1}^{j} x_{(l)}} \quad \text{for } T \in (x_{(j)}, x_{(j+1)}),\ 0 \leq j \leq k, \tag{21}$$

$$\underline{C}_{X_{n+2}}(T) = \frac{(nj+2j-kj-k)c_2 + (n+2-k)(n+1-j)c_1}{(n+1-k)\sum_{l=1}^{k} x_{(l)} + (n+2-k)\sum_{l=k+1}^{j} x_{(l)} + (n+2-k)(n+1-j)T} \tag{22}$$
$$\text{for } T \in (x_{(j)}, x_{(j+1)}),\ k < j \leq n.$$

Analogous to the situation for X_{n+1}, as discussed in the previous section, the lower survival function for X_{n+2} has no probability mass beyond the largest observation, that is, $\underline{S}_{X_{n+2}}(T) = 0$ for $T > x_{(n)}$, and consequently the upper cost function for X_{n+2} is constant beyond $x_{(n)}$. From

Lemma 1 we get that $\overline{C}_{X_{n+2}}(T) > \overline{C}_{X_{n+2}}(x_{(n)})$ for $T > x_{(n)}$, so to determine the corresponding optimal age replacement time for unit $n+2$, which we denote by $T^*_{up,n+2}$, we can restrict attention to the interval $(0, x_{(n)}]$. As was the case with the lower cost function for unit $n+1$, we must either restrict attention to $T \in (0, x_{(n)}]$ when considering $\underline{C}_{X_{n+2}}(T)$, or assume a finite upper bound r for the support of X_{n+2}. We deal with this in the same way as for X_{n+1}, that is, we determine the optimal age replacement time for the lower cost function for X_{n+2} over the interval $(0, x_{(n)}]$, which we denote by $T^{-*}_{low,n+2}$. If $\underline{C}_{X_{n+2}}(T^{-*}_{low,n+2})$ is greater than $\underline{C}_{X_{n+2}}(r^-)$, then it is better not to replace at all, otherwise the optimal age replacement time for unit $n+2$, according to this criterion, is $T^{-*}_{low,n+2}$. Alternatively, we can again calculate a critical value r^*.

The following lemma (proof in the appendix) presents some mathematical properties of these upper and lower cost functions for age replacement of unit $n+2$, which then immediately imply Theorem 1, which again makes optimisation of these cost functions computationally straightforward.

Lemma 3

a. $\overline{C}_{X_{n+2}}(\cdot)$ is continuous and strictly decreasing in $T \in (x_{(j)}, x_{(j+1)})$ for $j = 0, \ldots, n-1$. Moreover, $\overline{C}_{X_{n+2}}(\cdot)$ is continuous from the left in $x_{(j)}$ for $j = 1, \ldots, n$ and every $x_{(j)}$ is a local minimum.

b. $\underline{C}_{X_{n+2}}(\cdot)$ is continuous and strictly decreasing in $T \in (x_{(j)}, x_{(j+1)})$ for $j = 0, \ldots, n$. Moreover, $\underline{C}_{X_{n+2}}(\cdot)$ is continuous from the right in $x_{(j)}$ for $j = 1, \ldots, n$ and every $x^-_{(j)}$ is a local minimum.

Theorem 1

a. The minimum of $\overline{C}_{X_{n+2}}(\cdot)$ is assumed in one of the points $x_{(j)}$, $j = 1, \ldots, n$.

b. The minimum of $\underline{C}_{X_{n+2}}(\cdot)$ over $(0, x_{(n)}]$ is assumed in one of the points $x^-_{(j)}$, $j = 1, \ldots, n$.

The values of $\overline{C}_{X_{n+2}}(x_{(j)})$ are given in (15) and (16) while the values of $\underline{C}_{X_{n+2}}(x^-_{(j)})$ are given by

$$\underline{C}_{X_{n+2}}(x^-_{(j)}) = \frac{(j-1)c_2 + (n+3-j)c_1}{(n+3-j)x_{(j)} + \sum_{l=1}^{j-1} x_{(l)}} \quad \text{for } 1 \leq j \leq k+1, \tag{23}$$

$$\underline{C}_{X_{n+2}}(x^-_{(j)}) = \frac{(nj - n + 2j - 2 - kj)c_2 + (n+2-k)(n+2-j)c_1}{(n+1-k)\sum_{l=1}^{k} x_{(l)} + (n+2-k)\sum_{l=k+1}^{j-1} x_{(l)} + (n+2-k)(n+2-j)x_{(j)}} \tag{24}$$

$$\text{for } k+2 \leq j \leq n.$$

We illustrate this procedure via an example.

Example 2 Suppose we have 10 lifetimes $1, 2, \ldots, 9, 10$. A preventive replacement costs $c_1 = 1$, while corrective replacement costs $c_2 = 10$. Applying the results from the previous section, the optimal age replacement time for unit 11, in the sense of minimising the lower cost function, is equal to $T^{-*}_{low,11} = 2^-$ with corresponding lower costs $\underline{C}_{X_{n+1}}(x^-_{(2)}) = 0.952$. Now suppose unit 11 is still functioning at time 2^-, so it is preventively replaced at time 2, leading to a right-censored observation. Theorem 1b can now be applied, and leads to optimal age replacement time $T^{-*}_{low,12} = 3^-$ for X_{12}, with corresponding lower costs $\underline{C}_{X_{n+2}}(x^-_{(3)}) = 0.909$.

In this example, the right-censored observation for unit 11 leads to an increase of the NPI-based optimal age replacement time, where minimisation of the lower cost function is used as criterion. It may be intuitively attractive that a preventive replacement leads to an increase of the optimal age replacement time. The following two theorems imply that, in this setting, a preventive replacement cannot lead to a smaller optimal age replacement time, which holds both for the lower and for the upper cost function. For the upper cost function, a preventive replacement also does not lead to a larger optimal replacement time, so the minimum of the upper cost function remains unchanged in case of a new observation in the form of a unit preventively replaced at its optimal replacement time according to the same criterion. These two theorems are actually formulated as stronger results, namely considering any number $m \geq 1$ units which are all preventively replaced at the optimal $x_{(k)}$ or $x^-_{(k)}$, following the n units that provide the observed failure times $x_{(1)} < \ldots < x_{(n)}$. For this setting, the next unit considered is unit $n+m+1$, and the required assumption rc-$A_{(n+m)}$ has been discussed in the introductory section on NPI, where also the relevant corresponding probabilities were specified. The proofs of Theorems 2 and 3 are given in the Appendix.

Theorem 2 If $T^{-*}_{low,n+1} = x^-_{(k)}$ minimises $\underline{C}_{X_{n+1}}(\cdot)$, and units $n+1$ to $n+m$, for $m \geq 1$, are all preventively replaced at $x^-_{(k)}$, then $\underline{C}_{X_{n+m+1}}(\cdot)$ is minimised at $T^{-*}_{low,n+m+1}$ with $T^{-*}_{low,n+m+1} \geq T^{-*}_{low,n+1}$.

Theorem 3 If $T^{*}_{up,n+1} = x_{(k)}$ minimises $\overline{C}_{X_{n+1}}(\cdot)$, and units $n+1$ to $n+m$, for $m \geq 1$, are all preventively replaced at $x_{(k)}$, then $\overline{C}_{X_{n+m+1}}(\cdot)$ is minimised at $T^{*}_{up,n+m+1}$ with $T^{*}_{up,n+m+1} = T^{*}_{up,n+1}$.

These two theorems provide general relations between optimal age replacement times which are in accordance with intuition, except for the fact that we have $T^{*}_{up,n+m+1} = T^{*}_{up,n+1}$ in

Theorem 3, instead of just an inequality. As the upper cost function is attained for the lower survival function, this might be explained from the fact that in the case of $m \geq 1$ right-censored observations at $x_{(k)}$, the lower survival function for unit $n+m+1$ has a point mass immediately after $x_{(k)}$ that has increased as a result of each such right-censoring, and hence may form a barrier for moving the optimal replacement time to the right. This point mass is not placed at the same position for the upper survival function for unit $n+m+1$, explaining the inequality in Theorem 2 for the corresponding result.

A simulation study of adaptive age replacement strategies

In this section we present results from simulation studies to illustrate our method and discuss several of its features. All simulations are performed with the statistical package R [20]. The lifetimes are simulated from a known distribution, enabling us to compare the optimal replacement times corresponding to our lower and upper cost functions with the theoretical optimal replacement time, which is the result of minimising (1) for the distribution used in the simulation. We have restricted attention to Weibull distributions with scale parameter 1, but differing shape parameters α (denoted by $W(\alpha, 1)$). Without loss of generality, we use $c_1 = 1$ in all simulations, as only the cost ratio c_2/c_1 is relevant for the location of the minimum of $C(\cdot)$.

Table 1 gives the theoretical optimal replacement times T^* and the corresponding minimal costs $C(T^*)$ for lifetime distributions $W(2,1)$, $W(3,1)$ and $W(1.2,1)$. We have also included the limiting values of these cost functions for $T \to \infty$, denoted by $C(\infty)$, which relate to no preventive replacement being carried out. As these Weibull distributions have shape parameter greater than 1, they all model wearout, so indeed finite replacement strategies may be optimal. We also include $\Lambda(\infty) = (C(\infty) - C(T^*))/C(T^*)$, which we will use for comparison of our method with the theoretical results in case these distributions are known. The values of $\Lambda(\infty)$ indicate the loss, relative to the optimal costs, if no preventive replacements were carried out. For such age replacement, effectiveness of preventive replacement depends largely on the variance of the underlying lifetime distribution, where increasing variance reduces the cost savings that can be achieved by preventive replacements. This is illustrated here by the large value of $\Lambda(\infty)$ for $W(3,1)$, which has the smallest variance of these three distributions, and the small value for $W(1.2,1)$, which has the largest variance. This latter value indicates that the cost function corresponding to $W(1.2,1)$ is very flat beyond the optimal replacement time.

	$W(2,1)$		$W(3,1)$		$W(1.2,1)$	
	$c_2 = 10$	$c_2 = 50$	$c_2 = 10$	$c_2 = 50$	$c_2 = 10$	$c_2 = 50$
T^*	0.3365	0.1431	0.3825	0.2170	0.6861	0.1522
$C(T^*)$	6.0561	14.0239	3.9494	6.9215	10.0161	40.3527
$C(\infty)$	11.2838	56.4190	11.1985	55.9923	10.6309	53.1544
$\Lambda(\infty)$	0.863	3.023	1.835	7.090	0.0614	0.317

Table 1: Summary of theoretical results for age replacement with known lifetime distributions

As before, let $T^{-*}_{low,n+1} = \text{argmin}\, \underline{C}_{X_{n+1}}(T)$, which denotes that $T^{-*}_{low,n+1}$ is the optimum replacement time corresponding to minimisation of $\underline{C}_{X_{n+1}}(T)$, and let $T^{-*}_{low,n+2} = \text{argmin}\, \underline{C}_{X_{n+2}}(T)$, $T^{*}_{up,n+1} = \text{argmin}\, \overline{C}_{X_{n+1}}(T)$ and $T^{*}_{up,n+2} = \text{argmin}\, \overline{C}_{X_{n+2}}(T)$, and for comparisons we will use $\Lambda_{low,n+1} = (C(T^{-*}_{low,n+1}) - C(T^*))/C(T^*)$, $\Lambda_{low,n+2} = (C(T^{-*}_{low,n+2}) - C(T^*))/C(T^*)$, $\Lambda_{up,n+1} = (C(T^{*}_{up,n+1}) - C(T^*))/C(T^*)$ and $\Lambda_{up,n+2} = (C(T^{*}_{up,n+2}) - C(T^*))/C(T^*)$. These Λ's indicate how good our optimum replacement times are compared to the theoretical optimum, judged by comparing the loss in long-run average costs per unit of time that would be incurred by using our optimum instead of the theoretical optimum, as fraction of the long-run average costs per unit of time in the theoretical optimum. Effectiveness of our method can be studied by comparing these Λ's to the corresponding $\Lambda(\infty)$-values in Table 1. Finally, n denotes the number of initially observed lifetimes, and throughout we use $A_{(n)}$ for our inference leading to $T^{-*}_{low,n+1}$ and $T^{*}_{up,n+1}$, and $A_{(n+1)}$ or rc-$A_{(n+1)}$ leading to $T^{-*}_{low,n+2}$ and $T^{*}_{up,n+2}$, depending on whether the observation involving unit $n+1$ is a failure time or a preventive replacement. In each case we have simulated 10000 times. Tables 2, 3 and 4 present the simulation results for the cases that the lifetimes are simulated from a $W(2,1)$, $W(3,1)$ and $W(1.2,1)$ distribution, respectively.

¿From the tables we see that the means of $T^{-*}_{low,n+1}$, $T^{-*}_{low,n+2}$, $T^{*}_{up,n+1}$ and $T^{*}_{up,n+2}$ are all larger than the corresponding theoretical T^*'s. However, as the distributions of the values of $T^{-*}_{low,n+1}$, $T^{-*}_{low,n+2}$, $T^{*}_{up,n+1}$ and $T^{*}_{up,n+2}$, from these simulations, are all skewed to the right, the medians may be better indications of performance of our method. We also see that for all three distributions the mean, median and standard deviation of $T^{-*}_{low,n+1}$ and $T^{*}_{up,n+1}$ are larger than the mean, median and standard deviation of $T^{-*}_{low,n+2}$ and $T^{*}_{up,n+2}$, respectively, except for the standard deviation of $W(1.2,1)$ with $n = 100$ and $c_2 = 10$. The differences between these corresponding values are the smallest for $W(3,1)$ and the largest for $W(1.2,1)$, which agrees with the fact that the variance of $W(3,1)$ and $W(1.2,1)$ are 0.0098 and 0.5639, respectively (the variance of $W(2,1)$ equals 0.1138).

	$T^{-*}_{low,n+1}$	$\Lambda_{low,n+1}$	$T^{-*}_{low,n+2}$	$\Lambda_{low,n+2}$	$T^{*}_{up,n+1}$	$\Lambda_{up,n+1}$	$T^{*}_{up,n+2}$	$\Lambda_{up,n+2}$
CASE 1-1	$W(2,1), c_2 = 10, n = 10$							
mean	0.4369	0.0966	0.4242	0.0905	0.5456	0.1370	0.5166	0.1200
median	0.3967	0.0472	0.3880	0.0439	0.5022	0.0808	0.4769	0.0669
sd	0.2011	0.1236	0.1921	0.1189	0.2313	0.1535	0.2175	0.1409
CASE 1-2	$W(2,1), c_2 = 10, n = 100$							
mean	0.3553	0.0313	0.3529	0.0310	0.3694	0.0315	0.3673	0.0312
median	0.3453	0.0152	0.3433	0.0151	0.3588	0.0157	0.3564	0.0155
sd	0.0926	0.0414	0.0917	0.0409	0.0948	0.0412	0.0941	0.0409
CASE 1-3	$W(2,1), c_2 = 50, n = 100$							
mean	0.1652	0.0720	0.1647	0.0717	0.1944	0.0897	0.1938	0.0889
median	0.1557	0.0358	0.1551	0.0354	0.1850	0.0460	0.1845	0.0457
sd	0.0614	0.0935	0.0612	0.0935	0.0665	0.1131	0.0662	0.1119

Table 2: Simulation results for $W(2,1)$

The tables show further that, in most cases sd($T^{-*}_{low,n+1}$) is the smallest for $W(3,1)$, and the largest for $W(1.2,1)$, except for CASEs 1-3 and 2-3, and the same ordering (with a few exceptions) also tends to hold for sd($T^{-*}_{low,n+2}$), sd($T^{*}_{up,n+1}$) and sd($T^{*}_{up,n+2}$). For individual cases (not shown) $T^{-*}_{low,n+1}$ and $T^{*}_{up,n+1}$ ($T^{-*}_{low,n+2}$ and $T^{*}_{up,n+2}$, respectively) are often at the same $x_{(j)}$, or else $T^{*}_{up,n+1}$ tends to be larger than $T^{-*}_{low,n+1}$, but the reverse also occurs. For $W(2,1)$ and $W(3,1)$ we see that the mean, median and standard deviation of $T^{-*}_{low,n+1}$ ($T^{-*}_{low,n+2}$) are smaller than the mean, median and standard deviation, respectively, of $T^{*}_{up,n+1}$ ($T^{*}_{up,n+2}$), but this does not hold in general for $W(1.2,1)$.

If we increase the number of observed lifetimes (n) from 10 to 100, the means and medians of $T^{-*}_{low,n+1}$, $T^{-*}_{low,n+2}$, $T^{*}_{up,n+1}$ and $T^{*}_{up,n+2}$ all get closer to the theoretical T^*'s. Increasing the cost of corrective replacement (c_2) from 10 to 50 tends to lead to earlier replacement. The differences between the means and the medians of $T^{-*}_{low,n+1}$ and $T^{-*}_{low,n+2}$, as well as those for $T^{*}_{up,n+1}$ and $T^{*}_{up,n+2}$, become smaller as n or c_2 increases.

For $W(2,1)$ and $W(3,1)$ we see that the mean, median and standard deviation of $\Lambda_{low,n+1}$ is larger than the mean, median and standard deviation of $\Lambda_{low,n+2}$, but for $W(1.2,1)$ the opposite holds. The mean, median and standard deviation of $\Lambda_{up,n+1}$ are also larger than those of $\Lambda_{up,n+2}$ for $W(2,1)$ and $W(3,1)$, but for $W(1.2,1)$ there does not seem to be a clear trend.

	$T^{-*}_{low,n+1}$	$\Lambda_{low,n+1}$	$T^{-*}_{low,n+2}$	$\Lambda_{low,n+2}$	$T^{*}_{up,n+1}$	$\Lambda_{up,n+1}$	$T^{*}_{up,n+2}$	$\Lambda_{up,n+2}$
CASE 2-1	$W(3,1), c_2 = 10, n = 10$							
mean	0.4821	0.1437	0.4727	0.1326	0.5471	0.2031	0.5303	0.1795
median	0.4688	0.0725	0.4609	0.0675	0.5329	0.1209	0.5164	0.1021
sd	0.1473	0.1798	0.1421	0.1663	0.1509	0.2279	0.1449	0.2078
CASE 2-2	$W(3,1), c_2 = 10, n = 100$							
mean	0.3983	0.0317	0.3972	0.0313	0.4096	0.0334	0.4083	0.0331
median	0.3936	0.0152	0.3928	0.0150	0.4054	0.0159	0.4041	0.0157
sd	0.0705	0.0427	0.0700	0.0419	0.0712	0.0454	0.0710	0.0448
CASE 2-3	$W(3,1), c_2 = 50, n = 100$							
mean	0.2501	0.0935	0.2494	0.0926	0.2762	0.1262	0.2753	0.1242
median	0.2448	0.0446	0.2441	0.0439	0.2704	0.0631	0.2696	0.0619
sd	0.0657	0.1233	0.0655	0.1224	0.0673	0.1613	0.0670	0.1589

Table 3: Simulation results for $W(3,1)$

If we increase the number of observed lifetimes from 10 to 100 then the means, medians and standard deviations of $\Lambda_{low,n+1}$, $\Lambda_{low,n+2}$, $\Lambda_{up,n+1}$ and $\Lambda_{up,n+2}$ all become smaller. Moreover, the differences between the means and the medians of the $\Lambda_{low,n+1}$ and $\Lambda_{low,n+2}$, as well as those for $\Lambda_{up,n+1}$ and $\Lambda_{up,n+2}$, become smaller for all three distributions.

If we increase the cost of corrective replacement from 10 to 50, then the means, medians and standard deviations of $\Lambda_{low,n+1}$, $\Lambda_{low,n+2}$, $\Lambda_{up,n+1}$ and $\Lambda_{up,n+2}$ all become larger. However, if we compare the relative values of these Λ's, as fractions of the corresponding $\Lambda(\infty)$-values, then they tend to be pretty similar on increasing the cost from 10 to 50 for both $W(2,1)$ and $W(3,1)$, yet for $W(1.2,1)$ the effects are again different. These simulations show that, for this latter case, our method has much more variation due both to the more varying data simulated, and the fact that the corresponding theoretical cost function is rather flat, which together lead to much more variation in the optimal replacement times suggested by our method, where in particular very early preventive replacement is bad in terms of costs.

In our simulation study we have also recorded the number of times that the optimal age replacement time is decreasing or increasing after the information on unit $n+1$, i.e. either an observed failure time or a right-censoring time, becomes available, see Table 5. Here the number

	$T^{-*}_{low,n+1}$	$\Lambda_{low,n+1}$	$T^{-*}_{low,n+2}$	$\Lambda_{low,n+2}$	$T^{*}_{up,n+1}$	$\Lambda_{up,n+1}$	$T^{*}_{up,n+2}$	$\Lambda_{up,n+2}$
CASE 3-1	$W(1.2,1), c_2 = 10, n = 10$							
mean	0.9583	0.0714	0.9237	0.0748	1.1129	0.0340	1.0424	0.0367
median	0.5294	0.0384	0.5041	0.0392	0.7782	0.0194	0.7196	0.0196
sd	1.0405	0.1043	1.0192	0.1089	0.9699	0.0493	0.9331	0.0551
CASE 3-2	$W(1.2,1), c_2 = 10, n = 100$							
mean	0.8969	0.0311	0.8912	0.0316	0.8132	0.0256	0.8022	0.0259
median	0.6050	0.0143	0.5988	0.0145	0.6345	0.0118	0.6264	0.0117
sd	0.9578	0.0454	0.9634	0.0461	0.6380	0.0385	0.6259	0.0392
CASE 3-3	$W(1.2,1), c_2 = 50, n = 100$							
mean	0.1763	0.0733	0.1745	0.0737	0.2352	0.0467	0.2324	0.0464
median	0.1168	0.0339	0.1160	0.0342	0.1681	0.0240	0.1665	0.0237
sd	0.2328	0.0998	0.2301	0.1005	0.2321	0.0583	0.2259	0.0581

Table 4: Simulation results for $W(1.2, 1)$

between brackets in the second row is the number of times that an increase corresponds with a situation where we had a right-censored observation. For the other rows, Theorems 2 and 3 imply that these cases cannot occur in case of a right-censored observation. See Table 6 for the total number of right-censored observed lifetimes in the simulations.

We see that in the vast majority of the simulations, $T^{-*}_{low,n+1} = T^{-*}_{low,n+2}$ and $T^{*}_{up,n+1} = T^{*}_{up,n+2}$, as the number of times that $T^{-*}_{low,n+1} < T^{-*}_{low,n+2}$ plus the number of times that $T^{-*}_{low,n+1} > T^{-*}_{low,n+2}$ equals 10000 minus the number of times that $T^{-*}_{low,n+1} = T^{-*}_{low,n+2}$, and of course the same is true for T^{*}_{up}.

¿From Table 6 we see that there are more right-censored observations for the lower cost function than for the upper cost function. This is because most of the time we have that $T^{*}_{up,n+1} \geq T^{-*}_{low,n+1}$, which implies that a right-censored observation for the lower cost function need not be a right-censored observation for the upper cost function. Also, the smaller the variance of the distribution used for the simulations, the more right-censorings (preventive replacements) occur. This can be explained from the fact that a larger variance, and hence more spread of the lifetimes, will give a higher chance of small values which imply corrective replacement, making age replacement strategies less successful than for smaller variance.

Table 5 shows that in some simulations with an observed failure time for unit $n+1$, hence a corrective replacement of this unit, we have $T^{*}_{up,n+2} > T^{*}_{up,n+1}$, so for the next unit it would be

	CASE 1-1	CASE 1-2	CASE 1-3	CASE 2-1	CASE 2-2	CASE 2-3
$T^{-*}_{low,n+2} < T^{-*}_{low,n+1}$	1241	319	162	1088	215	146
$T^{-*}_{low,n+2} > T^{-*}_{low,n+1}$	335(238)	119(38)	77(47)	221(196)	74(48)	35(28)
$T^{*}_{up,n+2} < T^{*}_{up,n+1}$	1578	323	177	1223	228	151
$T^{*}_{up,n+2} > T^{*}_{up,n+1}$	184	86	49	51	34	11
	CASE 3-1	CASE 3-2	CASE 3-3			
$T^{-*}_{low,n+2} < T^{-*}_{low,n+1}$	1600	518	315			
$T^{-*}_{low,n+2} > T^{-*}_{low,n+1}$	787(117)	277(20)	174(24)			
$T^{*}_{up,n+2} < T^{*}_{up,n+1}$	1723	548	364			
$T^{*}_{up,n+2} > T^{*}_{up,n+1}$	495	275	186			

Table 5: Number of times optimal age replacement time is increasing or decreasing

	CASE 1-1	CASE 1-2	CASE 1-3	CASE 2-1	CASE 2-2	CASE 2-3
Lower cost function	8114	8748	9699	8683	9365	9808
Upper cost function	7241	8644	9600	8253	9306	9759
	CASE 3-1	CASE 3-2	CASE 3-3			
Lower cost function	5376	5322	8895			
Upper cost function	4445	5255	8476			

Table 6: Number of right-censored observed lifetimes in 10000 simulations

optimal to replace preventively later. Such a situation is illustrated in more detail in Example 3 below. This may be counter intuitive, but one way to get a feeling for this result might be by considering that early replacements tend to be better if the new unit will be quite reliable early on. The new observation makes us doubt such early reliability more, hence it may be better to leave a unit in operation longer. This is combined, however, with higher expected costs, as also illustrated in Example 3. This same effect also occurs for the lower cost function.

Example 3 The 10 ordered observed lifetimes: 0.2571, 0.2885, 0.4716, 0.5038, 0.7454, 0.7799, 1.0055, 1.2370, 1.4912 and 1.5504, lead to $T^*_{up,11} = 0.2571$, with corresponding upper costs of 7.7780. Suppose that the observed lifetime for X_{11} is 0.1555, then $T^*_{up,12} = 0.4716$, with corresponding upper costs of 10.7280. Hence, the optimal replacement time has increased after observing an early failure time, and the corresponding expected costs have also increased.

To summarize the results of this simulation study, we note that our method tends to work reasonably well for data sets as small as $n = 10$, and particularly well for $n = 100$, and, that the information on unit $n + 1$ has more effect for $n = 10$ than for $n = 100$. Our method is, however, less successful if data are drawn from distributions with larger variance. Of course, these conclusions agree fully with intuition, where the latter point made agrees with the general fact that age replacement is less effective if the variance of the lifetimes is larger, yet this effect is stronger in our method as it not only has to deal with a relatively flat theoretical cost function, but also with more variation in the data sets. Further important insights are provided in the cases where unit $n+1$ is correctively replaced after an early failure, as our study indicates that this quite frequently leads to a larger replacement time for the next unit.

Conclusions

In this paper, we have studied the adaptive nature of our NPI-based age replacement strategies, in particular how optimal replacement times adapt to information from the first unit to which our optimal strategies are actually applied. If this unit $n + 1$ is preventively replaced, then we derived theoretical results on the change in optimal replacement time, most importantly that this would never become smaller, while in case of corrective replacement our simulation study showed that the optimal replacement time could actually move in any direction.

Our NPI-based method requires available failure data, which one may criticize from practical perspective. However, our study via simulations based on assumed theoretical distributions highlighted some features which may have been surprising, and provided some more valuable insights. One should be careful that, if one wishes to infer underlying distributions from process data and combine this with the standard age replacement approach with an assumed known distribution, then the outcomes may vary substantially depending on the data. However, in many cases it would still seem to give reasonably good replacement times, which is a pleasing conclusion. We should point out that one may reasonably doubt the use of NPI, as one would expect underlying distributions to be somewhat smooth. However, our conclusions are useful beyond NPI, in the sense that the larger and more flexible an assumed parametric family of probability distributions would be, the closer a best-fitting member of such a family to given process data will be to the empirical distribution, and hence, from predictive perspective, the closer the behaviour of corresponding optimal age replacement times would be to our NPI-based optimal replacement times. In particular, this implies that one can expect that, on the basis of about 10 observed failures, one can reasonably well determine good age replacement times in many situations, and that early corrective replacements of a unit may lead to movement of the optimal replacement time in both directions. Finally, if available failure data indicate large variation, we should be more sceptical about effective age replacement than if there seems to be less variation.

Although this study has given valuable insights into adaptiveness of age replacement strategies with regard to available data, it could be argued that the cost function should not be based on the renewal reward theorem if indeed one wishes to allow changing replacement times per cycle. In future research, we will study the effect of using a one-cycle cost function instead, and we will compare that with the approach and conclusions from this paper. It will also be interesting to compare our method with alternative adaptive methods, where parametrical lifetime distributions are assumed, and where data can be used to estimate the relevant parameters [2]. As we explicitly aimed at studying the effect the data have on optimal age replacement strategies without adding further modelling assumptions for the lifetime distributions, and due to considerations with regard to the length of this paper, we have not included such comparisons here.

Appendix

Proof of Lemma 1

Proof. For the situation that $0 \leq j \leq k$ and $T = x_{(j)}$, we have

$$\begin{aligned}\int_0^T \underline{S}_{X_{n+2}}(x)\,dx &= \sum_{l=0}^{j-1}\int_{x_{(l)}}^{x_{(l+1)}} \underline{S}_{X_{n+2}}(x)\,dx = \sum_{l=0}^{j-1}\int_{x_{(l)}}^{x_{(l+1)}} \underline{S}_{X_{n+2}}(x_{(l+1)})\,dx = \sum_{l=0}^{j-1}\int_{x_{(l)}}^{x_{(l+1)}} \frac{(n+1-l)}{n+2}\,dx \\ &= \sum_{l=0}^{j-1}\frac{n+1-l}{n+2}(x_{(l+1)}-x_{(l)}) = \frac{1}{n+2}\left\{\sum_{l=1}^{j}(n+2-l)x_{(l)} - \sum_{l=0}^{j-1}(n+1-l)x_{(l)}\right\} \\ &= \frac{n+2-j}{n+2}x_{(j)} + \frac{1}{n+2}\sum_{l=1}^{j-1}x_{(l)}\end{aligned}$$

Substituting this into the cost function (1) yields (15). The formulas (16), (17) and (18) can be derived in the same way using

$$\int_0^T \underline{S}_{X_{n+2}}(x)\,dx = \sum_{l=0}^{k-1}\int_{x_{(l)}}^{x_{(l+1)}} \underline{S}_{X_{n+2}}(x_{(l+1)})\,dx + \sum_{l=k}^{j-1}\int_{x_{(l)}}^{x_{(l+1)}} \underline{S}_{X_{n+2}}(x_{(l+1)})\,dx, \text{ for } T = x_{(j)},\ k < j \leq n+1,$$

$$\int_0^T \underline{S}_{X_{n+2}}(x)\,dx = \sum_{l=0}^{j-1}\int_{x_{(l)}}^{x_{(l+1)}} \underline{S}_{X_{n+2}}(x_{(l+1)})\,dx + \int_{x_{(j)}}^{T} \underline{S}_{X_{n+2}}(x_{(j+1)})\,dx, \text{ for } T \in (x_{(j)}, x_{(j+1)}),\ 0 \leq j < k,$$

$$\int_0^T \underline{S}_{X_{n+2}}(x)\,dx = \sum_{l=0}^{k-1}\int_{x_{(l)}}^{x_{(l+1)}} \underline{S}_{X_{n+2}}(x_{(l+1)})\,dx + \sum_{l=k}^{j-1}\int_{x_{(l)}}^{x_{(l+1)}} \underline{S}_{X_{n+2}}(x_{(l+1)})\,dx + \int_{x_{(j)}}^{T} \underline{S}_{X_{n+2}}(x_{(j+1)})\,dx,$$
$$\text{for } T \in (x_{(j)}, x_{(j+1)}),\ k \leq j \leq n,$$

respectively. □

Proof of Lemma 3

Proof. $\overline{C}_{X_{n+2}}(\cdot)$ and $\underline{C}_{X_{n+2}}(\cdot)$ are continuous as $\underline{S}_{X_{n+2}}(\cdot)$ and $\overline{S}_{X_{n+2}}(\cdot)$, respectively, are continuous in $T \in (x_{(j)}, x_{(j+1)})$, for $j = 0, \ldots, n$. From (17) and (18) it follows immediately that, for $j = 0, \ldots, n-1$, $\overline{C}_{X_{n+2}}(\cdot)$ is strictly decreasing in $T \in (x_{(j)}, x_{(j+1)})$ ($\overline{C}_{X_{n+2}}(\cdot)$ is constant for $T \in (x_{(n)}, x_{(n+1)})$). From (21) and (22) it follows that, for $j = 0, \ldots, n$, $\underline{C}_{X_{n+2}}(\cdot)$ is strictly decreasing in $T \in (x_{(j)}, x_{(j+1)})$. $\overline{C}_{X_{n+2}}(\cdot)$ is continuous from the left in $x_{(j)}$ for $j = 1, \ldots, n$ as, for $1 \leq j \leq k-1$, for $j = k$ and for $k+1 \leq j \leq n$,

$$\lim_{\epsilon \downarrow 0} \overline{C}_{X_{n+2}}(x_{(j)} - \epsilon) = \overline{C}_{X_{n+2}}(x_{(j)}) < \lim_{\epsilon \downarrow 0} \overline{C}_{X_{n+2}}(x_{(j)} + \epsilon).$$

$\underline{C}_{X_{n+2}}(\cdot)$ is continuous from the right in $x_{(j)}$ for $j = 1, \ldots, n$ as, for $1 \leq j \leq k$, for $j = k+1$ and for $k+2 \leq j \leq n$,

$$\lim_{\epsilon \downarrow 0} \underline{C}_{X_{n+2}}(x_{(j)} + \epsilon) = \underline{C}_{X_{n+2}}(x_{(j)}) > \lim_{\epsilon \downarrow 0} \underline{C}_{X_{n+2}}(x_{(j)} - \epsilon).$$

□

Proof of Theorem 2

Proof. If, for $T_1 < T_2$, the long-run average cost per unit time $C(T)$ (see (1)) satisfies $C(T_1) > C(T_2)$, then this is equivalent to

$$\frac{S(T_1) \int_0^{T_2} S(x)dx - S(T_2) \int_0^{T_1} S(x)dx}{\int_{T_1}^{T_2} S(x)dx} < \frac{c_2}{c_2 - c_1}. \tag{A.1}$$

Take T_1 to be equal to $x_{(j)}^-$, T_2 to be equal to $x_{(k)}^-$ and substitute $\overline{S}_{X_{n+1}}(\cdot)$ for $S(\cdot)$ (as $\overline{S}(\cdot)$ yields $\underline{C}(\cdot)$), then we know that, if $x_{(j)}^- < x_{(k)}^-$, (A.1) holds for X_{n+1}. We want to prove that this implies that (A.1) also holds for X_{n+m+1}, $m \geq 1$. In this setting, condition (A.1) for X_{n+1} equals

$$\frac{\overline{S}_{X_{n+1}}(x_{(j)}^-) \int_0^{x_{(k)}^-} \overline{S}_{X_{n+1}}(x)dx - \overline{S}_{X_{n+1}}(x_{(k)}^-) \int_0^{x_{(j)}^-} \overline{S}_{X_{n+1}}(x)dx}{\int_{x_{(j)}^-}^{x_{(k)}^-} \overline{S}_{X_{n+1}}(x)dx} < \frac{c_2}{c_2 - c_1} \tag{A.2}$$

where

$$\overline{S}_{X_{n+1}}(x^-_{(j)}) = \frac{n+2-j}{n+1}, \quad \overline{S}_{X_{n+1}}(x^-_{(k)}) = \frac{n+2-k}{n+1}, \tag{A.3}$$

$$\begin{aligned}
\int_0^{x^-_{(k)}} \overline{S}_{X_{n+1}}(x)dx &= \sum_{l=0}^{k-2} \int_{x_{(l)}}^{x_{(l+1)}} \overline{S}_{X_{n+1}}(x)dx + \lim_{\epsilon\downarrow 0} \int_{x_{(k-1)}}^{x_{(k)}-\epsilon} \overline{S}_{X_{n+1}}(x)dx \\
&= \sum_{l=0}^{k-2} \frac{n+1-l}{n+1}(x_{(l+1)} - x_{(l)}) + \lim_{\epsilon\downarrow 0} \left(\frac{n+2-k}{n+1}\right)(x_{(k)} - \epsilon - x_{(k-1)}) \\
&= \sum_{l=0}^{k-1} \frac{n+1-l}{n+1}(x_{(l+1)} - x_{(l)}) = \frac{1}{n+1}\left\{\sum_{l=1}^{k-1} x_{(l)} + (n+2-k)x_{(k)}\right\}, \qquad \text{(A.4)}
\end{aligned}$$

$$\int_0^{x^-_{(j)}} \overline{S}_{X_{n+1}}(x)dx = \sum_{l=0}^{j-1} \frac{n+1-l}{n+1}(x_{(l+1)} - x_{(l)}) = \frac{1}{n+1}\left\{\sum_{l=1}^{j-1} x_{(l)} + (n+2-j)x_{(j)}\right\}, \tag{A.5}$$

$$\begin{aligned}
\int_{x^-_{(j)}}^{x^-_{(k)}} \overline{S}_{X_{n+1}}(x)dx &= \sum_{l=j}^{k-1} \frac{n+1-l}{n+1}(x_{(l+1)} - x_{(l)}) \\
&= \frac{1}{n+1}\left\{\sum_{l=j+1}^{k-1} x_{(l)} + (n+2-k)x_{(k)} - (n+1-j)x_{(j)}\right\}. \qquad \text{(A.6)}
\end{aligned}$$

Here, (A.5) and (A.6) can be derived in a similar way as (A.4). The expressions to the right of the second equality sign in (A.4), (A.5) and (A.6) are only mentioned as they provide an easy way to calculate the corresponding integrals.

Let $q \equiv \frac{c_2}{c_2 - c_1} (> 1)$ and substitute (A.3)-(A.6) into (A.2), then condition (A.2) becomes equivalent to

$$\frac{\sum_{l=0}^{j-1}(n+1-l)(x_{(l+1)} - x_{(l)})}{\sum_{l=j}^{k-1}(n+1-l)(x_{(l+1)} - x_{(l)})} < \frac{q(n+1)-(n+2-j)}{k-j}. \tag{A.7}$$

So we know that if $x^-_{(j)} < x^-_{(k)}$ then (A.7) holds for X_{n+1}. Now, we want to prove that this implies that (A.1) holds for X_{n+m+1}, $m \geq 1$, that is, we want to prove that

$$\frac{\overline{S}_{X_{n+m+1}}(x^-_{(j)}) \int_0^{x^-_{(k)}} \overline{S}_{X_{n+m+1}}(x)dx - \overline{S}_{X_{n+m+1}}(x^-_{(k)}) \int_0^{x^-_{(j)}} \overline{S}_{X_{n+m+1}}(x)dx}{\int_{x^-_{(j)}}^{x^-_{(k)}} \overline{S}_{X_{n+m+1}}(x)dx} < q. \tag{A.8}$$

Here

$$\overline{S}_{X_{n+m+1}}(x_{(j)}^{-}) = \frac{n+m+2-j}{n+m+1}, \quad \overline{S}_{X_{n+m+1}}(x_{(k)}^{-}) = \frac{n+m+2-k}{n+m+1}, \tag{A.9}$$

$$\begin{aligned} \int_0^{x_{(k)}^{-}} \overline{S}_{X_{n+m+1}}(x)dx &= \sum_{l=0}^{k-1} \frac{n+m+1-l}{n+m+1}(x_{(l+1)} - x_{(l)}) \\ &= \frac{1}{n+m+1}\left\{\sum_{l=0}^{k-1} x_{(l)} + (n+m+2-k)x_{(k)}\right\}, \end{aligned} \tag{A.10}$$

$$\begin{aligned} \int_0^{x_{(j)}^{-}} \overline{S}_{X_{n+m+1}}(x)dx &= \sum_{l=0}^{j-1} \frac{n+m+1-l}{n+m+1}(x_{(l+1)} - x_{(l)}) \\ &= \frac{1}{n+m+1}\left\{\sum_{l=0}^{j-1} x_{(l)} + (n+m+2-j)x_{(j)}\right\}, \end{aligned} \tag{A.11}$$

$$\begin{aligned} \int_{x_{(j)}^{-}}^{x_{(k)}^{-}} \overline{S}_{X_{n+m+1}}(x)dx &= \sum_{l=j}^{k-1} \frac{n+m+1-l}{n+m+1}(x_{(l+1)} - x_{(l)}) \\ &= \frac{1}{n+m+1}\left\{\sum_{l=j}^{k-1} x_{(l)} + (n+m+2-k)x_{(k)} - (n+m+2-j)x_{(j)}\right\}. \end{aligned} \tag{A.12}$$

Substituting (A.9)-(A.12) into (A.8) yields that we have to prove the following statement:

$$\frac{\sum_{l=0}^{j-1}(n+m+1-l)(x_{(l+1)} - x_{(l)})}{\sum_{l=j}^{k-1}(n+m+1-l)(x_{(l+1)} - x_{(l)})} < \frac{q(n+m+1)-(n+m+2-j)}{k-j}. \tag{A.13}$$

As $q > 1$ we have that $q(n+m+1)-(n+m+2-j) > q(n+1)-(n+2-j)$ so that the RHS of (A.7) is less than the RHS of (A.13). Hence, it is sufficient to prove that the LHS of (A.7) is greater than the LHS of (A.13) for (A.13) to hold. So we have to prove that

$$\frac{\sum_{l=0}^{j-1}(n+1-l)(x_{(l+1)} - x_{(l)})}{\sum_{l=j}^{k-1}(n+1-l)(x_{(l+1)} - x_{(l)})} > \frac{\sum_{l=0}^{j-1}(n+m+1-l)(x_{(l+1)} - x_{(l)})}{\sum_{l=j}^{k-1}(n+m+1-l)(x_{(l+1)} - x_{(l)})}. \tag{A.14}$$

But (A.14) is equivalent to

$$\sum_{l=0}^{j-1}\sum_{s=j}^{k-1}(x_{(l+1)} - x_{(l)})(x_{(s+1)} - x_{(s)})\left[(n+1-l)(n+m+1-s)-(n+1-s)(n+m+1-l)\right] > 0 \tag{A.15}$$

and (A.15) is true as $(n+1-l)(n+m+1-s)-(n+1-s)(n+m+1-l) = m(s-l) > 0$. $\square$

Proof of Theorem 3

Proof. The proof consists of two parts. In the first part we prove that $T^*_{up,n+m+1} \nless T^*_{up,n+1}$ and in the second part we prove that $T^*_{up,n+m+1} \ngtr T^*_{up,n+1}$.

Part 1: $T^*_{up,n+m+1} \nless T^*_{up,n+1}$

¿From the proof of Theorem 2 we know that for $T_1 < T_2$, the long-run average cost per unit time $C(T)$ satisfies (A.1). Now taking T_1 to be equal to $x_{(j)}$, T_2 to be equal to $x_{(k)}$ and substituting $\underline{S}_{X_{n+1}}(\cdot)$ for $S(\cdot)$ (as $\underline{S}(\cdot)$ yields $\overline{C}(\cdot)$), we know that, if $x_{(j)} < x_{(k)}$, (A.1) holds also for X_{n+1}. We want to prove that this implies that (A.1) also holds for X_{n+m+1}.

Condition (A.1) for X_{n+1} equals

$$\frac{\underline{S}_{X_{n+1}}(x_{(j)}) \int\limits_0^{x_{(k)}} \underline{S}_{X_{n+1}}(x)dx - \underline{S}_{X_{n+1}}(x_{(k)}) \int\limits_0^{x_{(j)}} \underline{S}_{X_{n+1}}(x)dx}{\int\limits_{x_{(j)}}^{x_{(k)}} \underline{S}_{X_{n+1}}(x)dx} < \frac{c_2}{c_2 - c_1} \tag{A.16}$$

where

$$\underline{S}_{X_{n+1}}(x_{(j)}) = \frac{n+1-j}{n+1}, \quad \underline{S}_{X_{n+1}}(x_{(k)}) = \frac{n+1-k}{n+1}, \tag{A.17}$$

$$\int\limits_0^{x_{(k)}} \underline{S}_{X_{n+1}}(x)dx = \sum_{l=0}^{k-1} \frac{n-l}{n+1}(x_{(l+1)} - x_{(l)}) = \frac{1}{n+1}\left\{\sum_{l=0}^{k-1} x_{(l)} + (n+1-k)x_{(k)}\right\}, \tag{A.18}$$

$$\int\limits_0^{x_{(j)}} \underline{S}_{X_{n+1}}(x)dx = \sum_{l=0}^{j-1} \frac{n-l}{n+1}(x_{(l+1)} - x_{(l)}) = \frac{1}{n+1}\left\{\sum_{l=0}^{j-1} x_{(l)} + (n+1-j)x_{(j)}\right\}, \tag{A.19}$$

$$\begin{aligned}\int\limits_{x_{(j)}}^{x_{(k)}} \underline{S}_{X_{n+1}}(x)dx &= \sum_{l=j}^{k-1} \frac{n-l}{n+1}(x_{(l+1)} - x_{(l)}) \\ &= \frac{1}{n+1}\left\{\sum_{l=j}^{k-1} x_{(l)} + (n+1-k)x_{(k)} - (n+1-j)x_{(j)}\right\}.\end{aligned} \tag{A.20}$$

Let again $q \equiv \frac{c_2}{c_2 - c_1} (> 1)$ and substitute (A.17)-(A.20) into (A.16), then condition (A.16) becomes equivalent to

$$\frac{\sum\limits_{l=0}^{j-1}(n-l)(x_{(l+1)} - x_{(l)})}{\sum\limits_{l=j}^{k-1}(n-l)(x_{(l+1)} - x_{(l)})} < \frac{q(n+1) - (n+1-j)}{k-j}. \tag{A.21}$$

So we know that if $x_{(j)} < x_{(k)}$ then (A.21) holds for X_{n+1}. Now, we want to prove that this implies that (A.1) holds for X_{n+m+1}, $m \geq 1$, that is, we want to prove that

$$\frac{\underline{S}_{X_{n+m+1}}(x_{(j)}) \int_0^{x_{(k)}} \underline{S}_{X_{n+m+1}}(x)dx - \underline{S}_{X_{n+m+1}}(x_{(k)}) \int_0^{x_{(j)}} \underline{S}_{X_{n+m+1}}(x)dx}{\int_{x_{(j)}}^{x_{(k)}} \underline{S}_{X_{n+m+1}}(x)dx} < q. \tag{A.22}$$

Here

$$\underline{S}_{X_{n+m+1}}(x_{(j)}) = \frac{n+m+1-j}{n+m+1}, \quad \underline{S}_{X_{n+m+1}}(x_{(k)}) = \frac{n+m+1-k}{n+m+1}, \tag{A.23}$$

$$\int_0^{x_{(k)}} \underline{S}_{X_{n+m+1}}(x)dx = \sum_{l=0}^{k-1} \frac{n+m-l}{n+m+1}(x_{(l+1)} - x_{(l)}), \tag{A.24}$$

$$\int_0^{x_{(j)}} \underline{S}_{X_{n+m+1}}(x)dx = \sum_{l=0}^{j-1} \frac{n+m-l}{n+m+1}(x_{(l+1)} - x_{(l)}), \tag{A.25}$$

$$\int_{x_{(j)}}^{x_{(k)}} \underline{S}_{X_{n+m+1}}(x)dx = \sum_{l=j}^{k-1} \frac{n+m-l}{n+m+1}(x_{(l+1)} - x_{(l)}). \tag{A.26}$$

Substituting (A.23)-(A.26) into (A.22) yields that we have to prove the following statement:

$$\frac{\sum_{l=0}^{j-1}(n+m-l)(x_{(l+1)} - x_{(l)})}{\sum_{l=j}^{k-1}(n+m-l)(x_{(l+1)} - x_{(l)})} < \frac{q(n+m+1) - (n+m+1-j)}{k-j}. \tag{A.27}$$

As again $q(n+m+1) - (n+m+1-j) > q(n+1) - (n+2-j)$, so that the RHS of (A.21) is less than the RHS of (A.27), we are ready if the LHS of (A.21) is greater than the LHS of (A.27). So we have to prove that

$$\frac{\sum_{l=0}^{j-1}(n-l)(x_{(l+1)} - x_{(l)})}{\sum_{l=j}^{k-1}(n-l)(x_{(l+1)} - x_{(l)})} > \frac{\sum_{l=0}^{j-1}(n+m-l)(x_{(l+1)} - x_{(l)})}{\sum_{l=j}^{k-1}(n+m-l)(x_{(l+1)} - x_{(l)})}. \tag{A.28}$$

But (A.28) is equivalent to

$$\sum_{l=0}^{j-1}\sum_{s=j}^{k-1}(x_{(l+1)} - x_{(l)})(x_{(s+1)} - x_{(s)})\left[(n-l)(n+m-s) - (n-s)(n+m-l)\right] > 0 \tag{A.29}$$

and (A.29) is true as $(n-l)(n+m-s) - (n-s)(n+m-l) = m(s-l) > 0$.

Part 2: $T^*_{up,n+m+1} \ngtr T^*_{up,n+1}$

First we prove that condition (A.32) must hold as $x_{(k)}$ yields minimum of $\overline{C}_{X_{n+1}}(\cdot)$. Choose $x_{(k)}$, $x_{(j)}$ with $j > k$. As $x_{(k)}$ gives minimal $\overline{C}_{X_{n+1}}(\cdot)$, we know that $\overline{C}_{X_{n+1}}(x_{(j)}) > \overline{C}_{X_{n+1}}(x_{(k)})$ for $k < j$, which is equivalent to

$$\frac{\underline{S}_{X_{n+1}}(x_{(k)}) \int_0^{x_{(j)}} \underline{S}_{X_{n+1}}(x)dx - \underline{S}_{X_{n+1}}(x_{(j)}) \int_0^{x_{(k)}} \underline{S}_{X_{n+1}}(x)dx}{\int_{x_{(k)}}^{x_{(j)}} \underline{S}_{X_{n+1}}(x)dx} > \frac{c_2}{c_2 - c_1} \tag{A.30}$$

where $\underline{S}_{X_{n+1}}(x_{(j)})$, $\underline{S}_{X_{n+1}}(x_{(k)})$, $\int_0^{x_{(k)}} \underline{S}_{X_{n+1}}(x)dx$ and $\int_0^{x_{(j)}} \underline{S}_{X_{n+1}}(x)dx$ are given in (A.17)-(A.19) and

$$\int_{x_{(k)}}^{x_{(j)}} \underline{S}_{X_{n+1}}(x)dx = \sum_{l=k}^{j-1} \left(\frac{n-l}{n+1}\right)(x_{(l+1)} - x_{(l)}). \tag{A.31}$$

Substituting (A.17)-(A.19) and (A.31) into (A.30) gives

$$\frac{\sum_{l=0}^{k-1}(n-l)(x_{(l+1)} - x_{(l)})}{\sum_{l=k}^{j-1}(n-l)(x_{(l+1)} - x_{(l)})} > \frac{q(n+1) - (n+1-k)}{j-k}. \tag{A.32}$$

So we know that if $x_{(k)} < x_{(j)}$ then (A.32) holds for X_{n+1}. We now prove that if there exists a $j > k$ such that $x_{(j)}$ gives minimal $\overline{C}_{X_{n+m+1}}(\cdot)$, then condition (A.38) must hold. As $x_{(j)}$ gives minimal $\overline{C}_{X_{n+m+1}}(\cdot)$ we have $\overline{C}_{X_{n+m+1}}(x_{(k)}) > \overline{C}_{X_{n+m+1}}(x_{(j)})$, which is equivalent to

$$\frac{\underline{S}_{X_{n+m+1}}(x_{(k)}) \int_0^{x_{(j)}} \underline{S}_{X_{n+m+1}}(x)dx - \underline{S}_{X_{n+m+1}}(x_{(j)}) \int_0^{x_{(k)}} \underline{S}_{X_{n+m+1}}(x)dx}{\int_{x_{(k)}}^{x_{(j)}} \underline{S}_{X_{n+m+1}}(x)dx} < \frac{c_2}{c_2 - c_1} \tag{A.33}$$

where

$$\underline{S}_{X_{n+m+1}}(x_{(k)}) = \frac{n+m+1-k}{n+m+1}, \quad \underline{S}_{X_{n+m+1}}(x_{(j)}) = \frac{n+m+1-k}{n+m+1} \cdot \frac{n+1-j}{n+1-k}, \tag{A.34}$$

$$\int_0^{x_{(k)}} \underline{S}_{X_{n+m+1}}(x)dx = \sum_{l=0}^{k-1} \frac{n+m-l}{n+m+1}(x_{(l+1)} - x_{(l)}), \tag{A.35}$$

$$\int_0^{x_{(j)}} \underline{S}_{X_{n+m+1}}(x)dx = \sum_{l=0}^{k-1} \frac{n+m-l}{n+m+1}(x_{(l+1)} - x_{(l)}) + \sum_{l=k}^{j-1} \frac{n+m+1-k}{n+m+1} \cdot \frac{n-l}{n+1-k}(x_{(l+1)} - x_{(l)}), \tag{A.36}$$

and

$$\int_{x_{(k)}}^{x_{(j)}} \underline{S}_{X_{n+m+1}}(x)dx = \sum_{l=k}^{j-1} \frac{n+m+1-k}{n+m+1} \cdot \frac{n-l}{n+1-k} (x_{(l+1)} - x_{(l)}). \tag{A.37}$$

Substituting (A.34)-(A.37) into (A.33) yields

$$\frac{\sum_{l=0}^{k-1} (n+m-l)(x_{(l+1)} - x_{(l)})}{\sum_{l=k}^{j-1} (n-l)(x_{(l+1)} - x_{(l)})} < \frac{q(n+m+1) - (n+m+1-k)}{j-k}. \tag{A.38}$$

We will now prove that condition (A.32) and (A.38) can never occur simultaneously, that is, we will prove that, in this setting, it is never possible that $T^*_{up,n+m+1} = x_{(k+i)}$ for $i = 1, 2, \ldots, n-k$, given that $T^*_{up,n+1} = x_{(k)}$. Substituting $j = k+i$ into (A.32) yields

$$\frac{i\sum_{l=0}^{k} x_{(l)} + i(n-k)x_{(k)}}{\sum_{l=k}^{k+i-1} (n-l)(x_{(l+1)} - x_{(l)})} > q(n+1) - (n+1-k) \tag{A.39}$$

whereas substituting $j = k+i$ into (A.38) yields

$$\frac{i\sum_{l=0}^{k} x_{(l)} + i(n+m-k)x_{(k)}}{\sum_{l=k}^{k+i-1} (n-l)(x_{(l+1)} - x_{(l)})} < q(n+m+1) - (n+m+1-k). \tag{A.40}$$

Combining (A.39) and (A.40) gives the following inequality

$$\begin{aligned} q(n+m+1) - (n+m+1-k) &> \frac{i\sum_{l=0}^{k} x_{(l)} + i(n+m-k)x_{(k)}}{i\sum_{l=0}^{k} x_{(l)} + i(n-k)x_{(k)}} \cdot \frac{i\sum_{l=0}^{k} x_{(l)} + i(n-k)x_{(k)}}{\sum_{l=k}^{k+i-1} (n-l)(x_{(l+1)} - x_{(l)})} \\ &> \frac{i\sum_{l=0}^{k} x_{(l)} + i(n+m-k)x_{(k)}}{i\sum_{l=0}^{k} x_{(l)} + i(n-k)x_{(k)}} [q(n+1) - (n+1-k)] \end{aligned}$$

which turns out to be equal to

$$q < \frac{-\sum_{l=0}^{k-1} x_{(l)}}{kx_{(k)} - \sum_{l=0}^{k-1} x_{(l)}} \leq 0. \tag{A.41}$$

But this gives a contradiction as $q \equiv \dfrac{c_2}{c_2 - c_1} > 1$. Hence, $T^*_{up,n+m+1} \not> T^*_{up,n+1}$.

Combining part 1 and part 2 of the proof yields that $T^*_{up,n+m+1} = T^*_{up,n+1},\ m \geq 1$.

□

Acknowledgement

This research is supported by the UK Engineering and Physical Research Sciences Council, grant GR/R92530/01. We thank the referees for detailed comments and suggestions.

References

[1] Barlow RE and Proschan F (1965). *Mathematical Theory of Reliability.* Wiley: New York.

[2] Mazzuchi TA and Soyer R (1996). A Bayesian perspective on some replacement strategies. *Reliability Engineering and System Safety* **51**: 295-303.

[3] Apeland S and Scarf PA (2003). A fully subjective approach to capital equipment replacement. *Journal of the Operational Research Society* **54**: 371-378.

[4] Percy DF (2002). Bayesian enhanced stategic decision making for reliability. *European Journal of Operational Research* **139**: 133-145.

[5] Coolen-Schrijner P and Coolen FPA (2004). Nonparametric predictive inference for age replacement with a renewal argument. *Quality and Reliability Engineering International*, to appear.

[6] Coolen FPA and Newby MJ (1997). Guidelines for corrective replacement based on low stochastic structure assumptions. *Quality and Reliability Engineering International* **13**: 177-182.

[7] Coolen FPA and Coolen-Schrijner P (2000). Condition monitoring: a new perspective. *Journal of the Operational Research Society* **51**: 311-319.

[8] Coolen FPA, Coolen-Schrijner P and Yan KJ (2002). Nonparametric predictive inference in reliability. *Reliability Engineering and System Safety* **78**: 185-193.

[9] Coolen FPA and Coolen-Schrijner P (2003). A nonparametric predictive method for queues. *European Journal of Operational Research* **145**: 425-442.

[10] Coolen FPA and Yan KJ (2004). Nonparametric predictive inference with right-censored data. *Journal of Statistical Planning and Inference*, to appear.

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

[12] Augustin T and Coolen FPA (2004). Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference*, to appear.

[13] Dempster AP (1963). On direct probabilities. *Journal of the Royal Statistical Society B* **25**: 100-110.

[14] De Finetti B (1974). *Theory of Probability.* Wiley: London.

[15] Hill BM (1988). De Finett's Theorem, Induction, and $A_{(n)}$ or Bayesian nonparametric predictive inference. In: Bernardo JM et al. (eds). *Bayesian Statistics* **3**, 211-241.

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

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

[18] Weichselberger K (2001). *Elementare Grundbegriffe einer Allgemeineren Wahrscheinlichkeitsrechnung I. Intervallwahrscheinlichkeit als Umfassendes Konzept.* Physika: Heidelberg (in German).

[19] Kaplan EL and Meier P (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**: 457-481.

[20] The Comprehensive R Archive Network (CRAN). http://cran.r=project.org/.