Nonparametric Predictive Inference for Age Replacement with a Renewal Argument

P. Coolen-Schrijner* *Pauline.Schrijner@durham.ac.uk*
F.P.A. Coolen *Frank.Coolen@durham.ac.uk*

Department of Mathematical Sciences
University of Durham, Durham, DH1 3LE, UK

Abstract

We consider an age replacement problem with cost function based on the renewal reward theorem. However, instead of assuming a known probability distribution for the lifetimes, we apply Hill's assumption $A_{(n)}$ for predicting probabilities for the lifetime of a future item. Lower and upper bounds for the survival function of a future item are used, resulting in upper and lower cost functions. Minimising these upper and lower cost functions to obtain the optimal age replacement times is simplified due to the special form of these functions.

To discuss some features of our approach, we first study the consequences of using n equally spaced percentiles from a known distribution instead of n observed data. Secondly, we report on a simulation study where the lifetimes are simulated from known distributions, so that the optimal replacement times corresponding to our approach can be compared with the theoretical optimal replacement times.

1 Introduction

Nonparametric predictive inference (NPI) is a recently developed statistical approach, using Hill's assumption $A_{(n)}$ [8] for prediction in case of vague prior knowledge of a probability distribution for real-valued random quantities. Let $x_{(j)},\ j = 1, \ldots, n$, be ordered data. Let X_j be the corresponding unordered random quantities. Hill [10] defines $A_{(n)}$ as follows:

(i) the observable random quantities $X_1, \ldots, X_n$ are exchangeable;

(ii) ties have probability 0 (generalisation is easy but requires awkward notation);

(iii) given $x_{(j)}, j = 1, \ldots, n$, the probability that the next observation falls in the interval $I_j = (x_{(j)}, x_{(j+1)})$ is $1/(n+1)$, for each $j = 0, \ldots, n$, where $x_{(0)} = -\infty$ (or $x_{(0)} = 0$ if $X_i \geq 0$) and $x_{(n+1)} = \infty$ (or a known upper bound for X_j).

*Corresponding author

For convenience, we drop the brackets in $x_{(j)}$ in the rest of the paper, and just assume $x_1 < x_2 < \ldots < x_n$. $A_{(n)}$ is a post-data assumption related to exchangeability [7], see [9] for discussion of $A_{(n)}$ and an overview of related work. Hill [10] showed that $A_{(n)}$ leads to De Finetti-coherent inference from a Bayesian perspective. Clearly, $A_{(n)}$-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.

The assumption $A_{(n)}$ is not sufficient to derive precise probabilities for many events of interest. However, it does provide bounds for such probabilities [2]. Applications of NPI in reliability and operational research have been presented [3, 4, 5, 6]. In [4] NPI is used to develop predictive probability results for the waiting time for customers in a queue, including guidelines on which queue to join in case of multiple queues. In [3, 6] NPI is applied to replacement problems. In [3] guidelines for preventive replacement decisions are presented, under the assumption that a unit's condition, which is continuously monitored, is in one of k possible states, and transitions occur at random times. If data are available in the form of n observed transition times for each of the k states, NPI is used for prediction of the time in each state, and probabilistic results for residual lifetime of the current unit, when entering a state, are derived via convolutions. In [6] an NPI-based strategy was developed for corrective replacement when observed failure times are available, addressing the question whether or not unit $n+1$ should be installed, given the failure times of the first n units used consecutively in a production process. In [5] a variety of NPI-based results are introduced to reliability, including attention to survival functions.

In this paper, we consider an age replacement problem, as formulated for a stochastic process with cost function following from application of the renewal reward theorem, see e.g. Barlow and Proschan [1]. The classical setting is as follows. Let $X_1, X_2, \ldots$ be independent identically distributed random quantities, representing lifetimes of units. Assume that the distribution of these random quantities is known, and has survival function $S(x) = P(X_j > x)$. The age replacement rule prescribes replacement of an item upon failure ('corrective replacement') or upon reaching the age T ('preventive replacement'), whichever occurs first. For each preventive replacement a fixed cost of $c_1 > 0$ is incurred, while each corrective replacement costs a fixed $c_2 > c_1$. Using the renewal argument [1] the long-run average costs per unit time, $C(T)$, is given by

$$C(T) = \frac{c_2 - S(T)(c_2 - c_1)}{\int_0^T S(x)dx}. \tag{1.1}$$

Without loss of generality, we will choose $c_1 = 1$ in all examples, as only the cost ratio c_2/c_1 is relevant for the location of the minimum of $C(\cdot)$.

For a known probability distribution, this cost function $C(\cdot)$ can be minimised in order to find the optimal replacement time T^*. However, there has been growing attention during the past decade to the estimation of the underlying distribution of the X_j, which is of particular interest in applications with actual failure data. For example, Mazzuchi and Soyer [11] considered a Bayesian approach, assuming that the underlying

distribution is a Weibull with unknown shape and scale parameter, and updating an assumed prior distribution for these parameters when process data become available. They correctly comment on the fact that the principle behind use of the above cost function $C(\cdot)$ is such that it should only be used if the resulting strategy is going to be used for a long period of time in the future, else, a cost function based on considering only a single cycle might be more appropriate.

In this paper, we analyse what inferences we can derive by adding only few mathematical assumptions to failure data, but we do assume that the resulting strategy will be used for a period long enough to justify the use of a cost function of the form above. Therefore, the results presented here provide a method that adapts fully to the available failure data at the time that an optimal strategy is determined, but it is assumed that this strategy will be fixed from then on. In later research we will consider even more adaptive strategies, which will explicitly adapt to each new piece of information from the process, and using optimisation over one cycle. It should be remarked that the cost function above is widely accepted for age replacement, with often not much emphasis on whether or not the resulting strategy will actually be used forever on.

Let us denote by $x_1 < x_2 < \ldots < x_n$ the ordered lifetimes of the first n units, and write $x_0 = 0$ and $x_{n+1} = \infty$. The assumption $A_{(n)}$ defines direct predictive probabilities for the lifetime X_{n+1} of a further item as $P(X_{n+1} \in (x_j, x_{j+1})) = 1/(n+1),\ j = 0, \ldots, n$. This leads to predictive survival function for X_{n+1} equal to [5]

$$S_{X_{n+1}}(x_j) = \frac{n+1-j}{n+1}, \text{ for } j = 0, \ldots, n+1. \tag{1.2}$$

As $A_{(n)}$ assigns probability mass to the open intervals $(x_j, x_{j+1}),\ j = 0, \ldots, n$, but does not put any further restrictions on the distribution of the probability mass within each such interval, it is not possible to give a precise value for the survival function at times other than previously observed event times without further assumptions. However, we can derive lower and upper bounds for the survival function, consistent with the probability assessment according to $A_{(n)}$. The maximum lower bound, $\underline{S}(x)$, can be obtained by shifting the probability mass in the interval in which x lies to the left end-point of the interval, leading to

$$\underline{S}_{X_{n+1}}(x) = S_{X_{n+1}}(x_{j+1}) = \frac{n-j}{n+1} \text{ for } x \in (x_j, x_{j+1}),\ j = 0, \ldots, n. \tag{1.3}$$

Similarly, the minimum upper bound, $\overline{S}(x)$, can be obtained by shifting the probability mass in the interval in which x lies to the right end-point of the interval, leading to

$$\overline{S}_{X_{n+1}}(x) = S_{X_{n+1}}(x_j) = \frac{n+1-j}{n+1} \text{ for } x \in (x_j, x_{j+1}),\ j = 0, \ldots, n. \tag{1.4}$$

$\underline{S}(\cdot)$ and $\overline{S}(\cdot)$ are called the *lower* and *upper survival function* for X_{n+1}, respectively [5]. It should be emphasized that the actual positions of probability masses $1/(n+1)$ per interval, for these upper and lower survival functions, is such that they are very close to the observed x_j (or 0 or ∞), but not actually at these points. This is what De Finetti [7] calls 'adherent probability' to a point, which implies that the probability mass is not actually in the point, but inside each open interval that contains the point. For

our applications this is relevant with regard to the continuity properties of the lower and upper survival functions and the related cost functions, where, at the x_j, $\underline{S}(\cdot)$ is continuous from the left, and $\overline{S}(\cdot)$ continuous from the right.

In the following sections we will study the use of these NPI-based lower and upper survival functions in the age replacement problem with the renewal argument. Sections 2 and 3 present the main results, namely that the lower and upper survival functions for X_{n+1} lead to the optimal upper and lower bounds, respectively, for the cost function for age replacement of the next unit. It is also shown that the minimisation of the corresponding cost functions is fairly straightforward, as effectively only the values at (or just before) observed failure times need to be considered. In Section 4 we discuss our upper and lower cost functions if instead of n data values we would actually use the $100j/(n+1)$-percentiles, for $j = 1, \ldots, n$, of a known probability distribution in our NPI approach. In Section 5 we illustrate our approach via simulations, which enable us to discuss its performance compared to theoretical optima. Finally, some concluding remarks are given in Section 6.

2 Upper Cost function

In this section we use a renewal argument and the lower survival function (1.3) to derive the optimal NPI upper bound for the long-run average costs per cycle for X_{n+1}. We will show that the optimal time to replace is in one of the points $x_j,\ j = 1, \ldots, n$, which significantly simplifies optimisation.

¿From (1.1) it is clear that the optimal NPI upper bound for the cost function for X_{n+1}, denoted by $\overline{C}_{X_{n+1}}(T)$, is obtained by substituting the lower survival function $\underline{S}_{X_{n+1}}(x)$ for $S(x)$, for $x \in (0, T]$.

Lemma 2.1

$$\overline{C}_{X_{n+1}}(x_j) = \frac{jc_2 + (n+1-j)c_1}{(n-j+1)x_j + \sum_{l=1}^{j-1} x_l}, \quad j = 1, \ldots, n+1, \tag{2.1}$$

$$\overline{C}_{X_{n+1}}(T) = \frac{(j+1)c_2 + (n-j)c_1}{(n-j)T + \sum_{l=1}^{j} x_l}, \quad \textit{for } T \in (x_j, x_{j+1}),\ j = 0, \ldots, n. \tag{2.2}$$

Proof. First consider $T = x_j,\ j = 1, \ldots, n+1$. Then

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

Substituting this into (1.1), we obtain the optimal upper bound for the cost function for X_{n+1} in $x_j,\ j = 1, \ldots, n+1$, as given in (2.1).

For $T \in (x_j, x_{j+1}),\ j = 0, \ldots, n$, we get

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

Substituting this into (1.1), we obtain the optimal upper bound for the cost function for X_{n+1} in $T \in (x_j, x_{j+1}),\ j = 0, \ldots, n$, as given in (2.2) □

An NPI-based upper bound for the cost function for X_{n+1} in the case that we do not repair preventively, that is, $T = \infty$, is given by $\overline{C}_{X_{n+1}}(x_{n+1}) = c_2 / \left(\frac{1}{n+1}\sum_{l=1}^{n} x_l\right)$, where the denominator is the lower expectation of X_{n+1} corresponding to $A_{(n)}$ and the data. Here the lower expectation of X_{n+1}, $\underline{E}(X_{n+1})$, is obtained by shifting all probability mass in each interval $(x_j, x_{j+1}),\ j = 0, \ldots, n$, to the left end-point of the interval, so $\underline{E}(X_{n+1}) = \frac{1}{n+1}\sum_{l=1}^{n} x_l$.

The lower survival function has no probability mass beyond the largest observation, so $\forall T > x_n$ we have $\underline{S}_{X_{n+1}}(T) = 0$. As a consequence, the upper cost function is constant for $T > x_n$. From (2.1) and (2.2) it is clear that $\overline{C}_{X_{n+1}}(T) > \overline{C}_{X_{n+1}}(x_n)$ for $T > x_n$, so we do not have to consider replacement times $T > x_n$ when determining the optimal replacement time in the sense of minimising the upper cost function.

Lemma 2.2 *$\overline{C}_{X_{n+1}}(\cdot)$ is a continuous and strictly decreasing function in $T \in (x_j, x_{j+1})$, $j = 0, \ldots, n-1$. Moreover, $\overline{C}_{X_{n+1}}(\cdot)$ is continuous from the left in $x_j,\ j = 1, \ldots, n$, and every x_j is a local minimum.*

Proof. $\overline{C}_{X_{n+1}}(\cdot)$ is a continuous function in $T \in (x_j, x_{j+1}),\ j = 0, \ldots, n-1$, as $\underline{S}_{X_{n+1}}(\cdot)$ is continuous in all such T, and $\overline{C}_{X_{n+1}}(\cdot)$ is clearly strictly decreasing in all $T \in (x_j, x_{j+1}),\ j = 0, \ldots, n-1$. Finally, for $j = 1, \ldots, n$,

$$\lim_{\epsilon \downarrow 0} \overline{C}_{X_{n+1}}(x_j + \epsilon) = \frac{(j+1)c_2 + (n-j)c_1}{(n-j+1)x_j + \sum_{l=1}^{j-1} x_l} > \overline{C}_{X_{n+1}}(x_j)$$

and

$$\lim_{\epsilon \downarrow 0} \overline{C}_{X_{n+1}}(x_j - \epsilon) = \overline{C}_{X_{n+1}}(x_j),$$

so $\overline{C}_{X_{n+1}}(\cdot)$ is continuous from the left in $x_j,\ j = 1, \ldots, n$, and each $x_j,\ j = 1, \ldots, n$, is a local minimum of $\overline{C}_{X_{n+1}}(\cdot)$. □

The following theorem follows immediately from Lemma 2.2.

Theorem 2.1 *The minimum of $\overline{C}_{X_{n+1}}(\cdot)$ is assumed in one of the points $x_j,\ j = 1, \ldots, n$.*

This theorem tells us that, in order to determine the optimum T in the sense of minimising $\overline{C}_{X_{n+1}}(\cdot)$, we only have to consider the points $T = x_j,\ j = 1, \ldots, n$, the previously observed failure times.

Example 2.1 Suppose we have observed 5 lifetimes: $4, 6, 10, 11, 15$. Each preventive replacement costs $c_1 = 1$, while each corrective replacement costs $c_2 = 10$. We would like to find the optimal replacement time for item 6 in the sense of minimising $\overline{C}_{X_6}(\cdot)$, using the renewal argument, the assumption $A_{(5)}$, and the data.

By Theorem 2.1 we only have to calculate $\overline{C}_{X_6}(x_j)$ for $j = 1, \ldots, 5$,

	x_1	x_2	x_3	x_4	x_5
	4	6	10	11	15
$\overline{C}_{X_6}(x_i)$	3/4	6/7	33/40	1	51/46

The optimal replacement time is $T = 4$, with corresponding upper costs $\overline{C}_{X_6}(x_1) = 3/4$. An upper bound for the cost function for X_6 if we do not repair preventively is given by $\overline{C}_{X_6}(\infty) = 30/23$. Figure 1 is a plot of $\overline{C}_{X_6}(\cdot)$, together with $\underline{C}_{X_6}(\cdot)$, which we discuss in the next section.

3 Lower Cost function

In this section we use a renewal argument and the upper survival function (1.4) to derive the optimal NPI lower bound for the long-run average costs per cycle for X_{n+1}. We will show that the optimal time to replace is in one of the points $x_j^-,\ j = 1, \ldots, n$, which are to be interpreted as 'just before x_j', and are such that the adherent probability mass to 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 $\overline{S}_{X_{n+1}}(\cdot)$. So, to determine the optimal value of the control parameter T in the sense of minimising the lower cost function, we again only have to check a finite number of values.

¿From (1.1) it is clear that the optimal NPI lower bound for the cost function for X_{n+1}, denoted by $\underline{C}_{X_{n+1}}$, is obtained by substituting the upper survival function $\overline{S}_{X_{n+1}}(x)$ for $S(x)$, for $x \in (0, T]$.

Lemma 3.1

$$\underline{C}_{X_{n+1}}(x_j) = \frac{jc_2 + (n+1-j)c_1}{(n-j+2)x_j + \sum_{l=1}^{j-1} x_l}, \quad j = 1, \ldots, n+1, \tag{3.1}$$

$$\underline{C}_{X_{n+1}}(T) = \frac{jc_2 + (n+1-j)c_1}{(n-j+1)T + \sum_{l=1}^{j} x_l}, \quad \textit{for } T \in (x_j, x_{j+1}),\ j = 0, \ldots, n. \tag{3.2}$$

Proof. First consider $T = x_j,\ j = 1, \ldots, n+1$. Then

$$\int_0^T \overline{S}_{X_{n+1}}(x)dx = \int_0^{x_j} \overline{S}_{X_{n+1}}(x)dx = \sum_{l=0}^{j-1} \int_{x_l}^{x_{l+1}} \overline{S}_{X_{n+1}}(x)dx = \sum_{l=0}^{j-1} (x_{l+1} - x_l) \left(\frac{n+1-l}{n+1}\right)$$
$$= \frac{1}{n+1}\left((n-j+2)x_j + \sum_{l=1}^{j-1} x_l\right).$$

Substituting this and (1.2) into (1.1), we obtain the lower bound (3.1) for the cost function for X_{n+1} in $x_j,\ j = 1, \ldots, n+1$. For $T \in (x_j, x_{j+1}),\ j = 0, \ldots, n$, we obtain

$$\int_0^T \overline{S}_{X_{n+1}}(x)dx = \sum_{l=0}^{j-1} \int_{x_l}^{x_{l+1}} \overline{S}_{X_{n+1}}(x)dx + \int_{x_j}^{T} \overline{S}_{X_{n+1}}(x)dx = \frac{1}{n+1}\left((n+1-j)T + \sum_{l=1}^{j} x_l\right).$$

Substituting this and (1.4) into (1.1), we obtain the lower bound (3.2) for the cost function for X_{n+1} in $T \in (x_j, x_{j+1}),\ j = 0, \ldots, n$. □

A lower bound for the cost function for X_{n+1} in the case that we do not repair preventively ($T = \infty$), is given by the expected cost c_2 divided by the upper expectation of X_{n+1} according to $A_{(n)}$ and the data. This corresponds to $\underline{C}_{X_{n+1}}(\infty)$ as given in (3.1), with $x_{n+1} = \infty$. The upper expectation $\overline{E}(X_{n+1})$ is obtained by shifting all probability mass in the interval in which x lies to the right end-point of the interval. So, we have $\overline{E}(X_{n+1}) = \frac{1}{n+1}\sum_{j=1}^{n+1} x_j$. But, as long as we do not assume a known finite upper bound for x_{n+1}, and thus use $x_{n+1} = \infty$, we have $\overline{E}(X_{n+1}) = \infty$ and consequently $\underline{C}_{X_{n+1}}(x_{n+1}) = 0$. This is a minor complication that we avoid, in first instance, by restricting attention to the interval $(0, x_n]$. However, we return to this issue after Theorem 3.1.

Lemma 3.2 *$\underline{C}_{X_{n+1}}(\cdot)$ is a continuous and strictly decreasing function in $T \in (x_j, x_{j+1})$, $j = 0, \ldots, n$. Moreover, $\underline{C}_{X_{n+1}}(\cdot)$ is continuous from the right in $x_j,\ j = 1, \ldots, n$, and every $x_j^-,\ j = 1, \ldots, n$, is a local minimum.*

Proof. The proof that $\underline{C}_{X_{n+1}}(\cdot)$ is a continuous and strictly decreasing function in $T \in (x_j, x_{j+1})$, $j = 0, \ldots, n$, is similar to the proof of Lemma 2.2. As, for $j = 0, \ldots, n-1$,

$$\lim_{\epsilon \downarrow 0} \underline{C}_{X_{n+1}}(x_{j+1} - \epsilon) = \frac{jc_2 + (n+1-j)c_1}{(n+1-j)x_{j+1} + \sum_{l=1}^{j} x_l} < \underline{C}_{X_{n+1}}(x_{j+1})$$

and

$$\lim_{\epsilon \downarrow 0} \underline{C}_{X_{n+1}}(x_{j+1} + \epsilon) = \underline{C}_{X_{n+1}}(x_{j+1}),$$

it follows that $\underline{C}_{X_{n+1}}(\cdot)$ is continuous from the right in $x_j,\ j = 1, \ldots, n$, and that each $x_j^-,\ j = 1, \ldots, n$, is a local minimum of $\underline{C}_{X_{n+1}}(\cdot)$. □

The following theorem is an immediate result of Lemma 3.2.

Theorem 3.1 *The minimum of $\underline{C}_{X_{n+1}}(T)$ over $(0, x_n]$ is assumed in one of the points $x_j^-,\ j = 1, \ldots, n$.*

If we assume a known finite right end-point r for the interval I_{n+1}, so that $x_{n+1} = r$ and $I_{n+1} = (x_n, r)$, then $A_{(n)}$ places mass $1/(n+1)$ in this interval and no probability mass to the right of r. This gives $\overline{E}(X_{n+1}) = \frac{1}{n+1}(\sum_{j=1}^{n} x_j + r)$ as the upper expectation for X_{n+1}. Now we have to compare the minimal costs corresponding to the optimal T

obtained by Theorem 3.1 with $\underline{C}_{X_{n+1}}(r^-)$. If $\underline{C}_{X_{n+1}}(r^-)$ is smaller than the minimal costs corresponding to the optimal T obtained by Theorem 3.1, then it is better not to replace preventively at all. Alternatively, we can calculate a critical value r^* such that if you think that an upper bound r for the support of X_{n+1} is less than or equal to r^*, then it is optimal to replace at the optimal T obtained by Theorem 3.1, but if you think that $r > r^*$, it is better not to replace preventively.

Example 2.1 (ctd) To find the optimal replacement time for X_6 in the sense of minimising $\underline{C}_{X_6}(\cdot)$, using the renewal argument, the assumption $A_{(5)}$, and the data, we only have to calculate $\underline{C}_{X_6}(x_j^-)$, $j = 1, \ldots, 5$,

	x_1^-	x_2^-	x_3^-	x_4^-	x_5^-
	4^-	6^-	10^-	11^-	15^-
$\underline{C}_{X_6}(x_j^-)$	1/4	15/34	12/25	33/53	42/61

The optimal replacement time is $T = 4^-$, with corresponding lower costs $\underline{C}_{X_6}(4^-) = 1/4$. Figure 1 shows $\underline{C}_{X_6}(\cdot)$ and $\overline{C}_{X_6}(\cdot)$.

The critical value r^* equals 158. Hence, if you think that an upper bound for X_6 is less than or equal to 158, then the optimal replacement time remains $T = 4^-$, otherwise it is better not to replace preventively. In this last case the minimal costs are $51/(r + 46)$ with r the assumed upper bound for X_6.

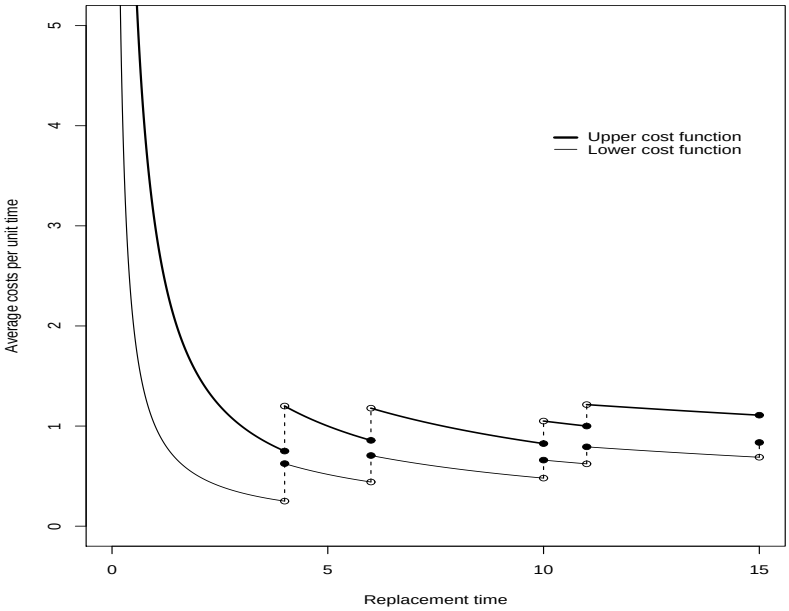

Figure 1: Lower and upper cost functions for Example 2.1

4 Use of percentiles for data

To discuss some features of our approach, assume that instead of n observed data we use the n equally spaced percentiles from a known distribution, so let x_j be the

$100j/(n+1)$-percentile, for $j = 1, \ldots, n$. For a distribution with survival function $S(\cdot)$, this implies $S(x_j) = \frac{n+1-j}{n+1}$, which agrees precisely, at these x_j-values, with our NPI-based predictive survival functions, as at these x_j-values our lower and upper predictive survival functions are the same and equal to this value. Therefore, if we apply our method to such percentiles of a known distribution, the actual underlying survival function is indeed everywhere between our NPI-based predictive lower and upper survival functions, and hence the corresponding theoretical cost function is between the NPI-based lower and upper cost functions. This is illustrated in Figure 2, where the theoretical model is a Weibull distribution[1], $W(2, 1)$, and the costs are $c_1 = 1$ and $c_2 = 10$, and we have used the percentiles corresponding to $n = 10$. We denote this theoretical cost function by $C(\cdot)$. Notice that, although these three survival functions are equal at these percentiles, the corresponding cost functions differ due to different values of the denominator in (1.1), as these survival functions differ in value at other points.

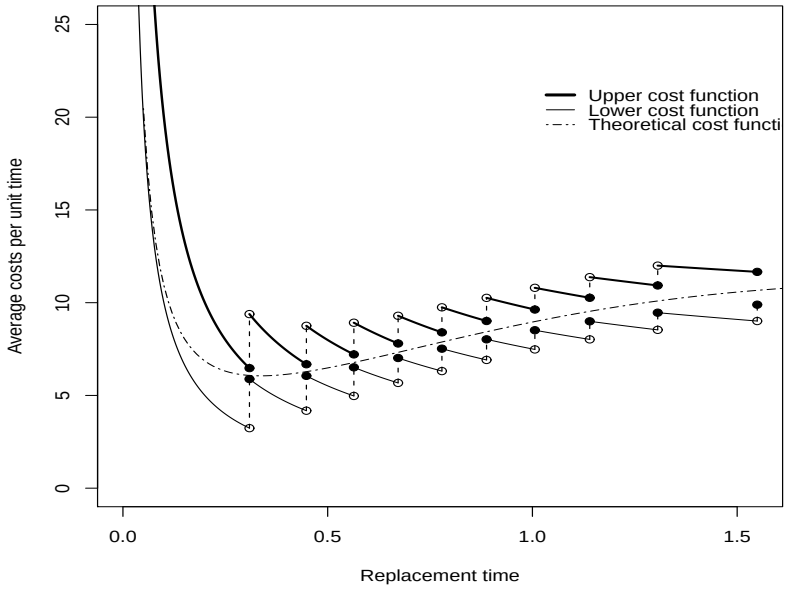

Figure 2: Cost functions with percentiles for data.

¿From Figure 2 we can conclude some interesting aspects. First, and most importantly, the fact that $C(\cdot)$ lies between our upper and lower cost functions does not imply a strong relation for their respective minima. In the situation in Figure 2, both the upper and lower cost functions happen to have their minimum in (or just before) the first included percentile, which is at 0.3087, whereas the theoretical cost function has its minimum at 0.3365. Indeed, we cannot generally conclude where the minimum of the underlying $C(\cdot)$ is on the basis of the minima of the upper and lower cost functions, even if the data were ideal reflections of the underlying distribution function as we assumed here.

[1] $W(\alpha, \beta)$ denotes a Weibull distribution with shape parameter α and scale parameter β.

If one would wish a strong general result on the location of the minimum of $C(\cdot)$ for such a situation with the appropriate percentiles replacing actual data, then the only achievable lower and upper bounds for the theoretical optimal replacement time T^* are those that follow logically from the fact that $C(T^*)$ must be between the minimum values of the upper and lower cost functions. However, since this is of little value when we work with actual data, we do not pursue this further. ¿From theoretical perspective this would be of interest, as in this way an interval can be derived that would contain the location of the minimum for the cost function (and similarly for the optimal value of the cost function) corresponding to each survival function between our upper and lower survival functions, so corresponding to the entire class of survival functions which are consistent with the NPI-based specification.

Further study of the performance of our method, when n such percentiles of a known distribution are used, revealed that, for increasing n, the optimal replacement times corresponding to our upper and lower cost functions converge to the theoretical optimum, which is a direct consequence of the fact that our upper and lower survival functions converge to the underlying survival function.

We also performed this analysis with an Exponential distribution, which has constant hazard rate, as underlying lifetime model, and with a $W(0.7, 1)$, which has decreasing hazard rate. For this latter Weibull model, the minima for both the lower and upper cost functions were assumed at the largest percentile, both for $n = 10$ and $n = 100$, and with $c_2 = 10$ and $c_2 = 50$ (always keeping $c_1 = 1$). For the Exponential distribution, we got a similar result for the lower cost function, while the minimum of the upper cost function (for these values for n and c_2) was always in one of the largest percentiles, but not the very largest. However, the cost functions related to the Exponential distribution are extremely flat in the right tail, so the differences in average costs per unit time between any large replacement times, or choosing not to perform preventive replacements at all, are very small indeed.

Considering the cost functions with these percentiles instead of the data is of interest as it illustrates some nice features of the approach, but it does not indicate how well our method performs when we get random data from an unknown underlying distribution. In the next section we report on simulation studies carried out to gain such insight. Of course, in such studies one has the advantage of actually knowing the assumed underlying distribution, and hence of being able to compare the theoretical optimal replacement times with the suggested replacement times following from minimisation of our NPI-based lower and upper cost functions, and also being able to compare the corresponding costs.

5 Simulations

In this section we present some simulation results to illustrate our method and discuss several of its features. All simulations are performed with the statistical package R [12]. The lifetimes are simulated from a known distribution, where we have restricted attention to Weibull distributions, so that we can 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 (1.1) for the distribution used in the simulation. It is well known that the theoretical optimal replacement time T^* is the unique solution to the equation [1]

$$h(T)\int_0^T S(x)dx - F(T) = \frac{c_1}{c_2 - c_1}, \tag{5.1}$$

with the hazard rate $h(T) = f(T)/S(T)$.

Let $T_{low}^{-*} = \text{argmin}\underline{C}_{X_{n+1}}(T)$, where this specific notation is used to indicate that the actual preventive replacement should take place just before the values T_{low}^* given in Table 3. Let $T_{up}^* = \text{argmin}\overline{C}_{X_{n+1}}(T)$, $\Delta_{low} = C(T_{low}^{-*}) - C(T^*)$, and $\Delta_{up} = C(T_{up}^*) - C(T^*)$. These Δ's indicate how good our optimum replacement times are compared to the theoretical optimum, for which the correct probability distribution needs to be known, judged by comparing the loss in long-run average costs per unit of time that would be incurred by using our optima instead of the theoretical optimum.

Table 1 gives the theoretical optimal replacement times T^* and the corresponding minimal costs $C(T^*)$ for the situation that the lifetime distributions are $W(2,1)$ (which has expected value 0.8862 and variance 0.1138), $W(3,1)$ (expected value 0.8930, variance 0.0098) and $W(1.2,1)$ (expected value 0.9407, variance 0.5639), and with $c_2 = 10$ or $c_2 = 50$. The row 'CASE' refers to the simulations reported in Table 3. The theoretical cost functions corresponding to these three Weibull distributions (with $c_2 = 10$) are presented in Figure 3. The limiting values for these cost functions, which relate to no preventive replacements being carried out (so $T = \infty$), are also given in Table 1, to assist comparison of the performance of our method in the simulations.

	$W(2,1)$		$W(3,1)$		$W(1.2,1)$	
	$c_2 = 10$	$c_2 = 50$	$c_2 = 10$	$c_2 = 50$	$c_2 = 10$	$c_2 = 50$
CASE	$1, 2$	3	$4, 5$	6	$7, 8$	9
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

Table 1: Theoretical results

It is well known that these three Weibull distributions, all with shape parameter greater than 1, model wearout, so indeed we expect finite optimal replacement strategies. These three distributions have similar expected values, but their variances are very different, implying e.g. that predicting when a unit fails for $W(3,1)$ would be far more accurate than for $W(1.2,1)$. This shows in the cost function values at the respective minima, and also when compared to the same cost functions if no preventive replacement would be carried out. Let us consider these distributions for the case $c_2 = 10$. The optimal strategy for $W(3,1)$ leads only to expected costs 3.9494, whereas for $W(1.2,1)$ the optimal expected costs are 10.0161, indicating that cases with little variance allow much more effective use of preventive replacements. This effectiveness

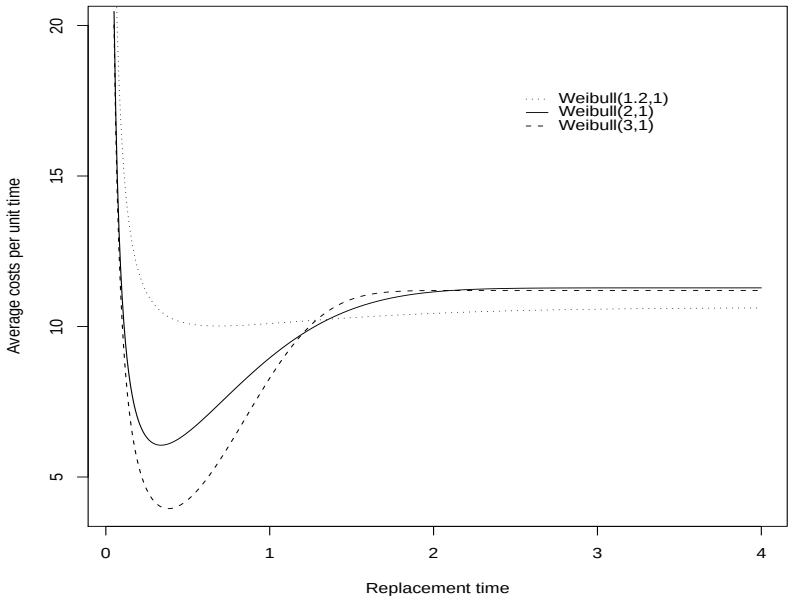

Figure 3: Theoretical cost functions for 3 Weibull distributions.

also shows when considering $C(\infty) - C(T^*)$, which is 7.2491 for $W(3,1)$, but only 0.6148 for $W(1.2,1)$.

Before presenting simulation results for our method, let us consider what the effect would be of wrongly assuming a fixed distribution which differs from the actual underlying distribution, where we again focus on the three Weibull distributions used above, and consider the case $c_2 = 10$. The results are presented in Table 2.

Actual distribution	$W(2,1)$		$W(3,1)$		$W(1.2,1)$	
Assumed distribution	$W(3,1)$	$W(1.2,1)$	$W(2,1)$	$W(1.2,1)$	$W(2,1)$	$W(3,1)$
Assumed optimal T	0.3825	0.6861	0.3365	0.6861	0.3365	0.3825
$C(T)-C(T^*)$|act.distr.	0.0458	1.3399	0.0601	1.5334	0.4828	0.3076

Table 2: Results when assuming a wrong underlying distribution

Suppose that the actual underlying distribution is $W(2,1)$, but we wrongly assumed it to be $W(3,1)$. Then we would have used the optimal strategy corresponding to $W(3,1)$, so $T = 0.3825$, which would have led to an increase of the average costs per unit of time of 0.0458. Similarly, had we wrongly assumed $W(1.2,1)$, the increase would be 1.3399, which is clearly a much worse mistake. If the underlying distribution is $W(3,1)$, the corresponding increases in costs would be 0.0601 if $W(2,1)$ were assumed, and 1.5334 if $W(1.2,1)$ were assumed. If the underlying distribution is $W(1.2,1)$, we get increases of 0.4828 ($W(2,1)$ assumed) and 0.3076 ($W(3,1)$ assumed). These last mistakes seem not too bad, but then again for this situation not applying any preventive

replacements only leads to an increase of 0.6148. The message seems to be that, if not using a method that takes data into account, but straightforwardly determining an optimal preventive replacement time based on an assumed distribution, overestimating the variance might lead to rather large increases in expected costs. The simulation studies below will show that our method, adapting the replacement time to the data, in most cases leads to increases in costs that are reasonably small.

In Table 3 the simulation results of 9 different cases are given. In each case we have simulated 10000 times. Each case differs in the number of observed lifetimes, or the distribution the lifetimes are simulated from, or the costs for corrective replacement.

	CASE 1: $n = 10$, $c_2 = 10$, $W(2,1)$				CASE 2: $n = 100$, $c_2 = 10$, $W(2,1)$			
	T^*_{low}	Δ_{low}	T^*_{up}	Δ_{up}	T^*_{low}	Δ_{low}	T^*_{up}	Δ_{up}
mean	0.4377	0.5845	0.5451	0.8207	0.3541	0.1864	0.3688	0.1855
median	0.4017	0.2917	0.5009	0.4828	0.3429	0.0894	0.3587	0.0912
sd	0.1993	0.7446	0.2312	0.9289	0.0920	0.2470	0.0933	0.2424
	CASE 3: $n = 100$, $c_2 = 50$, $W(2,1)$				CASE 4: $n = 10$, $c_2 = 10$, $W(3,1)$			
	T^*_{low}	Δ_{low}	T^*_{up}	Δ_{up}	T^*_{low}	Δ_{low}	T^*_{up}	Δ_{up}
mean	0.1640	0.9944	0.1933	1.2395	0.4823	0.5593	0.5480	0.8020
median	0.1546	0.4747	0.1832	0.6221	0.4700	0.2929	0.5331	0.4769
sd	0.0609	1.3409	0.0665	1.5902	0.1454	0.6992	0.1503	0.9057
	CASE 5: $n = 100$, $c_2 = 10$, $W(3,1)$				CASE 6: $n = 100$, $c_2 = 50$, $W(3,1)$			
	T^*_{low}	Δ_{low}	T^*_{up}	Δ_{up}	T^*_{low}	Δ_{low}	T^*_{up}	Δ_{up}
mean	0.4009	0.1267	0.4120	0.1348	0.2497	0.6572	0.2766	0.8703
median	0.3969	0.0592	0.4066	0.0630	0.2454	0.3208	0.2706	0.4390
sd	0.0708	0.1709	0.0715	0.1821	0.0662	0.8664	0.0666	1.1047
	CASE 7: $n = 10$, $c_2 = 10$, $W(1.2,1)$				CASE 8: $n = 100$, $c_2 = 10$, $W(1.2,1)$			
	T^*_{low}	Δ_{low}	T^*_{up}	Δ_{up}	T^*_{low}	Δ_{low}	T^*_{up}	Δ_{up}
mean	0.9564	0.7136	1.1078	0.3472	0.8969	0.3182	0.8206	0.2622
median	0.5339	0.3748	0.7775	0.1993	0.6183	0.1436	0.6480	0.1185
sd	1.0228	1.0584	0.9579	0.5113	0.9450	0.4720	0.6355	0.4016
	CASE 9: $n = 100$, $c_2 = 50$, $W(1.2,1)$							
	T^*_{low}	Δ_{low}	T^*_{up}	Δ_{up}				
mean	0.1747	2.9038	0.2367	1.8637				
median	0.1174	1.2872	0.1706	0.9281				
sd	0.2089	4.0133	0.2282	2.3655				

Table 3: Simulation results

All our mean T^*_{low}'s and T^*_{up}'s are larger than the corresponding theoretical T^*'s (see Tables 1 and 3). However, for individual cases it regularly occurs that our optima are a bit smaller than T^*. Generally, T^* is pretty small, for example quite smaller than the expected value of the corresponding distribution. In the simulation runs, most optima for our method are close to the T^*'s, but some are considerably larger, whereas much smaller times are rarely suggested, and of course they cannot be negative,

leading to the mean values exceeding the theoretical ones. Given the skewness of the distributions of our replacement times, the medians may be more natural indications of the performance of our method. Indeed, the medians are fairly close to the theoretical values. Of course, for $n = 100$ our method works better than for $n = 10$, when keeping $c_2 = 10$, the improvement is considerable as is shown in the median Δ values. We also studied (for $n = 100$) the effect of increasing c_2 to 50. Naturally, this leads to more cautious preventive replacement policies in the sense that such replacements take place considerably earlier than for $c_2 = 10$.

When considering the Δ values, the distributions are also very skewed to the right. Although large Δ values can occur if T^*_{low} or T^*_{up} is quite a bit smaller than T^*, the really large Δ values tend to appear for large T^*_{low} or T^*_{up}. In some simulation cases one or both of our optima, T^*_{low} and T^*_{up}, is at the largest observation x_n. This does not often happen for simulated data from $W(2, 1)$ and $W(3, 1)$, if it happens then normally only for small numbers of simulated data ($n = 10$ in this case), where it is possible that a simulated data set does not reflect wearout, but it happens fairly frequently for $W(1.2, 1)$. Although this will have affected the reported mean values in the table, it happens in fewer than half the cases, so did not affect the medians, which are best for judging the overall performance of our methods via simulations.

When comparing the T^*_{low} and T^*_{up} for individual cases (not shown in a table), these are often at the same x_i (so 'just before' it for the lower cost function). When this is not the case, T^*_{up} tends to be larger than T^*_{low}, but the reverse also occurs. It may perhaps be a bit surprising that there is no strong general result on the relation between T^*_{low} and T^*_{up}, but this is caused by the fact that the cost function is quite complex, as the optimum does not just follow from the survival function but also the hazard function plays a role, see (5.1). In our approach, the hazard function becomes effectively infinite just before, or at, the x_j, which leads to the fact that our upper and lower cost functions have local minima at (or just before) each x_j. In addition, intuitively one might perhaps expect that an upper survival function reflects a 'better unit' than a corresponding lower survival function, and that this would imply leaving such a unit longer in service. This would not be correct, as it is really the variation in the random lifetime that has most effect on the optimal replacement time. And, in this chosen criterion based on the renewal argument, we implicitly assume that a unit is replaced by a similar unit, so if a unit is 'better', the next one will also be good, so at a time of increased risk of a unit failing, it may make sense to replace it early by a new good unit. This just indicates that, when it comes to optimal age replacement strategies, intuition cannot really be relied on.

6 Concluding remarks

In this paper we concentrated on the predictive survival and cost functions for X_{n+1}. Future research will consider further adaptive aspects of NPI-based age replacement strategies, in particular how such strategies adapt to new data becoming available from the process, where a future observation, with strategy T^* in place as based on n observations, will either be a failure time less than T^*, or a right-censored observation

at T^* in case of preventive replacement. Besides using a renewal argument, we will also consider a 1-cycle criterion [11] to determine the optimal value of the control parameter T, which is appropriate if a different preventive replacement time is allowed for each future cycle, to be based on all process information available at the start of each cycle.

In this paper, we have mostly considered two fairly extreme cases, namely either using minimal modelling assumptions in our NPI-based methods, or assuming complete knowledge of the underlying lifetime distributions. Naturally, there are several other possibilities, including assumption of a parametric lifetime distribution, e.g. $W(\alpha, \beta)$, with parameters α and β not assumed to be known, but to be estimated from the data. For example, Mazzuchi and Soyer [11] take this approach, and learn about the parameter values via Bayesian updating of assumed prior distributions. Such methods are interesting, and successful as long as the assumed model is reasonably close to the unknown underlying model. If that is the case, such methods will perform better than our method, particularly for small data sets. On the other hand, if the underlying distribution is quite different from the assumed class of distributions, such methods offer less opportunity to adapt to the data available, and hence may lead to worse decisions. Detailed comparison of our method with such adaptive methods that use parametric models, e.g. via simulation studies, is an interesting topic for future research.

This paper has shown that our method tends to perform reasonably well in most cases, which is particularly interesting when one keeps in mind that no knowledge about underlying distributions has been included, so only the observed data were used. An obvious disadvantage of our method is that it explicitly requires data, which may not always be readily available. However, one could also use it to study, via simulations as reported here, adaptive policies based on data sets simulated from an assumed underlying distribution, for example if one would wish to conclude on robustness of a theoretical optimum. Although we considered minimisation of both $\underline{C}(\cdot)$ and $\overline{C}(\cdot)$, one could argue that, from a robust inference point of view, minimising $\overline{\overline{C}}(\cdot)$ would be the more relevant procedure.

Finally, although the Weibull distributions used in the simulation studies have quite different shapes, further simulation studies using other distributions, e.g. Lognormal or Gamma distributions, could be useful to get a more complete picture of the performance of our method.

Acknowledgement

This research is supported by the UK Engineering and Physical Research Sciences Council, grant GR/R92530/01.

References

[1] Barlow, R.E. and Proschan, F. (1965). *Mathematical Theory of Reliability*. Wiley, New York.

[2] Coolen, F.P.A. (2003). On the use of imprecise probabilities in reliability. *Quality and Reliability Engineering International*, this issue.

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

[4] Coolen, F.P.A. and Coolen-Schrijner, P. (2003). A nonparametric predictive method for queues. *European Journal of Operations Research*, **145**, 425-442.

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

[6] Coolen, F.P.A. and Newby, M.J. (1997). Guidelines for corrective replacement based on low stochastic structure assumption. *Quality and Reliability Engineering International*, **13**, 177-182.

[7] De Finetti, B. (1974). *Theory of Probability*. Wiley, London.

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

[9] Hill, B.M. (1988). De Finetti's Theorem, Induction, and $A_{(n)}$ or Bayesian nonparametric predictive inference. In *Bayesian Statistics 3*, J.M. Bernardo, et al. (eds.), 211-241.

[10] Hill, B.M. (1993). Parametric models for $A_{(n)}$: splitting processes and mixtures. *Journal of the Royal Statistical Society B*, **55**, 423-433.

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

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