Adaptive Designs
IMS Lecture Notes - Monograph Series (1995) Volume 25

ADAPTIVE DESIGNS FOR OPTIMAL AGE REPLACEMENT POLICY

BY GIRISH ARAS

Indian Institute of Technology

Abstract

Consider a system that is subject to failure and must be replaced when this occurs. If it costs less to replace the system in advance before failure, it may be advantageous to use an age replacement policy. However, the optimal age to replace the system is unknown if the underlying model for machine failure is unknown. This paper reviews various schemes that balance the conflicting goals of gathering enough information about the lifetime distribution, and simultaneously controlling costs by reducing system failures.

1. Introduction. In general, optimal replacement policies are designed to reduce the number of system failures and minimize maintenance costs by adopting a schedule of planned replacements. A great deal of literature [see Thomas (1986) and Valdez-Florez and Feldman (1989) for review articles] is devoted to the study of optimal replacement policies. By far the greater part of this literature is concerned with finding the optimal policy when the underlying model for system failure is known. Much less work has been done to actually estimate these optimal policies based on maintenance history data. Most estimation procedures that have been developed are based on observing a

Received December 1992; revised April 1993.
AMS 1991 subject classifications. Primary 62N05, 62G05.
Key words and prhases. Censored survival data, Kaplan-Meier estimator, reliability.

fixed number of complete lifetimes. For systems and equipment that are long lived or costly, the luxury of obtaining a sample of complete lifetimes before a maintenance policy is implemented may not be practical or affordable.

An attempt to minimize costs and failures needs to be made even while collecting the data to be used for estimating an optimal policy. A practical approach will be to estimate adaptively after a complete lifetime pilot sample is obtained. By this we mean that, at replacement (whether it is due to system failure or stipulated by the current maintenance policy), the current estimate of the optimal policy is updated. The difficulty and challenge of successfully implementing such a scheme is to balance the conflicting goals of gathering enough information about the system lifetimes for estimation and simultaneously trying to control costs by preventing system failures.

Barlow and Proschan (1965) provide a general description of maintenance policies including the age replacement policy (ARP) that we shall describe now. A system is replaced at age φ (planned replacement) or at failure (unplanned replacement), whichever comes first. If the cost c_1 of an unplanned replacement is more than the cost c_2 of a planned replacement, ARP can lead to considerable savings. The usual criterion for choosing the optimal age φ^* of replacement is to minimize long run expected cost per unit time. Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed system lifetimes with distribution F supported on the positive half-line. The actual cost of the first n units under the ARP at φ is

(1)

$$C_n(\varphi) = \sum_{i=1}^{n} \{c_1 I(X_i < \varphi) + c_2 I(X_i \geq \varphi)\},$$

where $I(A)$ is the indicator function of set A. The amount of time the n units have functioned is $(X_1 \wedge \varphi) + (X_2 \wedge \varphi) + \cdots + (X_n \wedge \varphi)$, where $\wedge$ denotes the minimum.. Let $N_\varphi(t)$ be the number of replacements by time t so that

$$N_\varphi(t) = \sum_{i \geq 1} I\{(X_1 \wedge \varphi) + \cdots + (X_i \wedge \varphi) \leq t\}.$$

The cost $C_n(\varphi)$ described in (1) is a renewal process. Thus it can be easily verified that the long run expected cost per unit time is [see Barlow and Proschan (1965, p. 86)]

(2)

$$C(\varphi) = \lim_{t\to\infty} E\left(\frac{C_{N(t)}}{t}\right) = \frac{[c_1 F(\varphi) + c_2 S(\varphi)]}{\mu(\varphi)},$$

where $S \equiv 1 - F(\varphi)$ and

$$\mu(\varphi) = E(X_1 \wedge \varphi) = \int_0^{\varphi} S(y)\, dy.$$

An alternative interpretation of (2), is that for an ARP at age $\varphi, C(\varphi)$ is the long run average system unavailability, where c_1 is interpreted as the expected time to repair the system and c_2 is the expected time to replace the system with a new one [cf. Barlow (1978)]. Under broad conditions, there is a unique and finite time, say φ^*, where $C(\cdot)$ attains a global minimum [cf. Bergman (1979)]. We call φ^* the optimal replacement time and $C(\varphi^*)$ the optimal cost. Minimizing $C(\varphi)$ is only one possible cost criterion. See Ansell, Bendell and Humble (1984) for a discussion of alternative criteria.

In practice φ^* is unknown since F is unknown and must be estimated. Bather (1977) introduced a method of adaptively constructing an estimator φ_n of φ^* based on past experience. Using his estimator at the nth stage, if the unit fails prior to φ_n, then the cost c_1 is incurred; otherwise the unit is replaced at φ_n incurring cost c_2. The age of the unit at the nth replacement is then used to update the estimate φ_n, producing φ_{n+1}. The cost after n units is

(3)

$$C_n = \sum_{i=1}^{n} \{c_1 I(X_i < \varphi_i) + c_2 I(X_i \geq \varphi_i)\}.$$

Since φ_i's are typically not independent and identically distributed, C_n is not a renewal process which complicates the analysis of this ARP. Bather constructed an adaptive procedure so that under mild conditions on F, φ_n converges to φ^* with probability one. In this setting, the number of replacements by time t is equal to

(4)

$$N(t) = \sum_{i \geq 1} I\left((X_1 \wedge \varphi_1) + \cdots + (X_i \wedge \varphi_i) \leq t\right).$$

An important result of Bather's (1977) is that

(5)

$$\lim_{t \to \infty} \frac{C_{N(t)}}{t} = C(\varphi^*)$$

with probability one.

Several other innovative approaches have been investigated for estimating φ^* [cf. Bergman (1979), Ingram and Scheaffer (1976), and Arunkumar (1972)]. Graphical methods have also been discussed by Bergman (1977) and Barlow (1978). These approaches are based on fixed-sample (i.e., non-adaptive procedures that rely on observing independent identically distributed lifetimes and thus do not allow truncation of the lifetimes). The cost after n units is nc_1 and hence these procedures can't achieve (5). However, these procedures are not without merit. Typically, before the adaptive ARP is implemented, the researcher will conduct a small pilot sample of complete lifetimes and get an initial estimate.

Adaptive methods have been much less studied. Frees and Ruppert (1985) extended the results of Bather in two directions. They showed that any sequence $\{\varphi_n\}$ converging to φ^* almost surely and satisfying some mild measurability conditions implies (5). This result opens the door for other types of adaptive procedures. In particular, they introduced a stochastic approximation estimator φ_n and give conditions under which it converges to φ^*. They also proved $n^{\frac{1}{2}-\varepsilon}(\varphi_n - \varphi^*)$ converges in distribution to a normal random variable, where ε is a positive number depending only on the smoothness of F at φ^* and the choice of the kernel function used in the procedure.

Aras and Whitaker (1991) motivate and develop an adaptive procedure in which the sequence of estimators φ_n is based on a nonparametric maximum likelihood approach. Their treatment is similar to Bather's and follows his ideas closely.

In the following sections we review the above three procedures and discuss their merits and demerits. For the proofs of these results readers should refer to the original articles.

2. Bather's procedure. The procedure depends upon two sets of constants. Let $b_0 > b_1 > b_2 \cdots$ be such that $b_m \to 0$ as $m \to \infty$ and $\sum b_i$ diverges. Define a sequence of independent Bernoulli variables $\{\alpha_n\}$ such that $\alpha_n = 0$ or 1 with probability p_n or $1 - p_n$, respectively. The sequence $\{\alpha_n\}$ is assumed to be independent of $\{X_n\}$. When $\alpha_n = 1$, the complete lifetime X_n is observed; in other words, the age replacement policy is not implemented. It is assumed that $p_1 = 0$. Suppose we have constructed φ_n by using data obtained from the first n units. After recording the value of α_n,

$$\begin{aligned}
\beta_n &= \max\{b_m\colon b_m \le \varphi_n, m \ge 0\};\\
\eta_n &= \beta_n \text{ and } \xi_n = \varphi_n, \text{ if } \alpha_n = 0;\\
\eta_n &= \xi_n = \infty, \text{ if } \alpha_n = 1.
\end{aligned}$$

ξ_n is the ARP, which is the same as φ_n except when $\alpha_n = 1$. Note that φ_n and ξ_n can take any positive value, whereas η_n cannot. η_n is a discredited version of ξ_n along the decreasing sequence $\{b_n\}$. Bather introduces $\{\eta_n\}$ purely for technical reasons.

In order to complete our inductive specification of the decision procedure, it remains to construct a new estimate φ_{n+1}. For any $x \ge 0$, and $j = 0, 1, \ldots, n$, let

$$\begin{aligned}
Y_j(x) &= 1 \text{ if } (X_j \wedge \eta_j) > x, && \text{and 0 otherwise;}\\
Z_j(x) &= 1 \text{ if } \eta_j > x, && \text{and 0 otherwise.}
\end{aligned}$$

Then $S(x)$ is estimated by using the ratio

$$S_{n+1}(x) = \frac{\sum_{j=0}^{n} Y_j(x)}{\sum_{j=0}^{n} Z_j(x)}.$$

Note that $S_{n+1}(x)$ is a right continuous step function with values in the interval $[0, 1]$. It is non-increasing in x, except perhaps along $\{b_m\}$. The set of points $\{b_m\}$ and the random sequence $\{\eta_n\}$ with support $\{b_m\}$ were introduced in order to restrict the location of any upward jumps in $S_{n+1}(x)$ along a countable set $\{b_m\}$. Cost function C is then estimated by

$$K_n(x) = [c_1 F_{n+1}(x-) + c_2 S_{n+1}(x-)] / \widehat{\mu}(x),$$

where

$$\widehat{\mu}(x) = \int_0^x S_{n+1}(y)\, dy.$$

The estimate φ_{n+1} can be determined by minimizing $K_{n+1}(x)$ with respect to x. The ARP is now completely specified.

The Borel-Cantelli Lemma and the condition that $\sum p_n = \infty$ imply that there exists an infinite sequence $\{\gamma_n\}$ along which $\alpha_{\gamma_n} = 1$ (and in between it is 0). Since $p_n \to 0$, Cherbychev's inequality shows that one can conclude that $\gamma_n/n \to \infty$ as $n \to \infty$. Since $\{\alpha_n\}$ is assumed to be independent of $\{X_n\}$, Bather's procedure is equivalent to observing complete lifetimes along some predetermined 'megre' sequence $\{\gamma_n\}$, where $\gamma_n/n \to \infty$ as $n \to \infty$. Bather shows that φ_n converges to φ^* almost surely and satisfies (5) as desired.

3. The Frees-Ruppert estimator. Define the function

$$M(t) = (c_1 - c_2) f(t) \int_0^t S(u)\, du \; - \; S(t) \{c_1 F(t) + c_2 S(t)\},$$

where f is a probability density function corresponding to F. Note that

$$\frac{d}{dt} C(t) = K_t M(t),$$

where K_t is a positive function of t. Instead of assuming C is uniquely minimized at some point φ^*, Frees and Ruppert (1985) assume the slightly stronger condition that $M(t)(t - \varphi^*) > 0$ for $t \neq \varphi^*$. If C is assumed to be differentiable, then the above assumption is equivalent to the statement that C has no points of local relative minima. Let g be a known, strictly increasing smooth nonnegative function. Define ξ^* such that $\varphi^* = g(\xi^*)$. Note that ξ^* is unique minimum of $C \circ g$ and is thus a unique, finite zero of $g^{'} \cdot M \circ g$, where "$\circ$" denotes composition of two functions and "$\cdot$" denotes the product. The function g is introduced to lead to unconstrained recursive estimation of ξ^* rather than to estimate a strictly positive parameter φ^* [e.g., g could be $\log\{1+\exp(x)\}$]. After defining an estimate ξ_n of ξ^*, g could be used to calculate φ_n, an estimate of φ^*.

It is convenient to describe the Frees-Ruppert procedure with a paired sequence of independent identically distributed random variables $\{X_{i,n}\}, i = 1, 2$ with distribution function F rather than a single sequence $\{X_n\}$. Suppose ξ_1 is an initial estimate of ξ^* such that $\mathrm{E}(\xi_1^2) < \infty$, and $\{E_n\}$, $\{a_n\}$ and $\{d_n\}$ are sequences of random variables. For $i = 1, 2$, define the truncated observations $\{Z_{i,n}\}$ by $Z_{i,n} = \min\{Z_{i,n}, g(\xi_n + d_n)\}$. It is assumed that a_n and d_n are measurable with respect to the σ-algebra generated by ξ_1 and $Z_{i,j}, i, j \leq n-1$. Typically $\{a_n\}$ and $\{d_n\}$ are chosen such that for fixed $a, d > 0$ and $\gamma \in (0,1)$, $na_n \to a$ and $n^\gamma d_n \to d$.

Let β_0 be the class of all Borel functions that are bounded and equal to zero outside $[-1,1]$. For some positive integer define

$$B_1 \equiv \left\{ k\varepsilon \in B_0 : \int_{-1}^{1} y^j k(y)\, dy = \left\{ \begin{array}{l} 1 \text{ if } j = 0 \\ 0 \text{ if } j = 1, \ \ldots, \ r-1 \end{array} \right. \right\}.$$

Define $H(t) \equiv F(g(t))$ and let $H^{(j)}$ be the jth partial derivative of H. For $i = 1, 2$, let $F_{i,n}(t) \equiv I\{Z_{i,n} < t\}$ and $S_{i,n}(t) \equiv 1 - F_{i,n}(t)$. For $k \in B_1$,

$$h_n(t) = k\left[\left(g^{-1}(Z_{1,n}) - t\right)/d_n\right]/d_n$$

is an estimator of $H^{(1)}(t)$. The estimator of $g'(t)\, M(g(t))$ is

(6)

$$M_{g,n}(t) = (c_1 - c_2)\, h_n(t) \int_0^{g(t)} S_{2,n}(u)\, du$$
$$- g'(t)\{c_1 F_{2,n}(g(t)) + c_2 S_{2,n}(g(t))\}.$$

The estimator (6) is constructed so that the conditional expectation of $M_{g,n}(t)$, given knowledge up to epoch $n-1$ is sufficiently close to $g'(t)\,M(g(t))$. The estimators ξ_n are constructed by the recursive algorithm,

(7)
$$\xi_{n+1} = \xi_n - a_n M_{g,n}(\xi_n).$$

Under smoothness assumptions on F and g and growth conditions on $\{d_n\}$ and $\{a_n\}$, Frees and Rupert (1985) not only prove almost sure convergence, but also the asymptotic normality of φ_n.

Estimating F, which is an important intermediate step in Bather's procedure and in the Aras–Whitaker procedure, is not crucial in the Frees-Ruppert procedure. Instead, they concentrate on posing the problem in the framework of stochastic approximation so that they can apply well developed machinery of stochastic approximation theory. This approach pays rich dividends in terms of a central limit theorem for φ_n. Their estimator of F, namely $F_{i,n}$ is based only on the nth observation and not on earlier data. This does raise fears about possible slow convergence in practice. However, extensive simulations by Aras, Whitaker and Wu (1992) indicate that it does fairly well in practice, especially if the initial guess φ_1 is close to φ^*.

4. The Aras-Whitaker procedure. Suppose that when the nth unit is at risk the current ARP is ξ_n. Let $Z_i = X_i \wedge \xi_i$ be the age at the ith replacement, $\delta_i = I(X_i \le \xi_i)$ and $\Im_n$ be the σ-algebra generated by $\{(Z_i, \delta_i);\ i = 1, 2, \ldots, n\}$. While ξ_n is the ARP, φ_n is the estimate of φ^*. The data available after n replacements for constructing the estimator φ_n of φ^* is clearly right censored, although not randomly right censored since φ_n depends on earlier data. S is estimated by calculating the nonparametric maximum likelihood estimator (see Keifer and Wolfowitz (1958)). Since ξ_n is $\mathcal{F}_{n-1}$-measurable by successive conditioning arguments the nonparametric maximum likelihood estimate is seen to be that survival function S_n which maximizes the function

$$\prod_{i=1}^{n} \left(S(Z_i-) - S(Z_i)\right)^{\delta_i} \left(S(Z_i)\right)^{1-\delta_i}.$$

Thus S_n is the well-known product limit estimator of S:

$$S_n(x) = \prod_{\substack{i:Z_{(i)}\le x,\\ i\le n}} \left(\frac{n-i}{n-i+1}\right)^{\delta_{(i)}},$$

where $Z_{(1)}, \ldots, Z_{(n)}$ are the order statistics of $Z_1, Z_2, \ldots, Z_n$ and $\delta_{(1)}$, $\ldots, \delta_{(n)}$ are the corresponding sequence. As in Bather's procedure, an estimator of the cost function $C(\cdot)$ is

$$K_n(x) = \left[c_1 F_n(x) + c_2 S_n(x)\right] / \mu_n(x).$$

The current estimator φ_n and φ^* is where a minimum of $K_n(\cdot)$ is reached. In order to complete our inductive specification of the ARP, it remains to construct ξ_{n+1} after observing (Z_n, δ_n).

Obvious cost considerations demand that ξ_{n+1} should be 'close' to φ_n. Because φ_n is an estimator of φ^*, it is a reasonable choice to use as the ARP at the nth stage. Since we want to estimate φ^* consistently, it is necessary to allow the unit to operate beyond φ_n. Let $\epsilon > 0$. For $\xi_{n+1} = \varphi_n + \epsilon$, Aras and Whitaker (1991) show that $\varphi_n \to \varphi^*$. However $C_{N(t)}/t$, the actual average cost up to time t, converges to $C(\varphi^* + \epsilon)$, i.e. (5) is only approximately satisfied by this procedure.

It is an open question whether consistency is maintained if ϵ is replaced by a sequence ϵ_n that decreases to zero as n tends to infinity. This will allow (5) to hold. Obtaining a central limit theorem for φ_n is also an important open problem.

Though the Aras-Whitaker procedure is strongly consistent, in practice one should use the following modification. Let $\{a_n\}$ be a nonnegative, $\Im_n$ measurable sequence of random variables. Define $\xi_{n+1} = \varphi_n + a_n + \epsilon$. Of course, a_n should converge to zero. $\Im_n$ measurability allows the possibility of choosing $\{a_n\}$ adaptively. One could choose large values of a_n initially to get a better idea about the location of φ^*.

Aras, Whitaker and Wu (1993) use simulations to provide some guidelines for proper choice of the $\{a_n\}$ sequence. In general, when the underlying distribution is close to exponential, the value of a_n should be large and not decrease as rapidly as when F has an increasing failure rate. Aras, Whitaker and Wu also report simulation results comparing the Aras-Whitaker procedure with the Frees-Ruppert procedure. The former consistently had a lower average actual cost per unit time than the latter.

To conclude, none of the above procedures are totally satisfactory and there is definite need for researchers to come up with better procedures.

References

ANSELL, T., BENDELL, A. and HUMBLE, S. (1984). Age replacement under alternative cost criteria. *Management Science* **30** 358-367.

ARAS, G. and WHITAKER, L.R. (1991). Sequential nonparametric estimation of an optimal age replacement policy. Technical Report 158, University of California, Santa Barbara.

ARAS, G., WHITAKER, L.R. and WU, Y.H. (1993). Sequential nonparametric estimation of an age replacement policy: a simulation study. *Communications in Statistics - Simulation and Computation* **22** 1115-1134.

ARUNKUMAR, S. (1972). Nonparametric age replacement policy. *Sankya A* **34** 251-256.

BARLOW, R.E. (1978). Analysis of retrospective failure data using computer graphics. *Proceedings of the 1978 Annual Reliability and Maintainability Symposium,* 113-116.

BARLOW, R.E. and PROSCHAN F. (1965). *Mathematical Theory of Reliability.* New York: John Wiley.

BATHER, J.A. (1977). On the sequential construction of an optimal age replacement policy. *Bulletin of Institute of International Statistics* **47** 253-266.

BERGMAN, B. (1977). Some graphical methods for maintenance planning. *Proceedings of the 1977 Annual Reliability and Maintainability Symposium,* 467-471.

BERGMAN, B. (1979). On age replacement and the total time on test concept. *Scandinavian Journal of Statistics* **6** 161-168.

FREES, E.W. and RUPPERT, D. (1985). Sequential nonparametric age replacement policies. *The Annals of Statistics* **13** 650-662.

INGRAM, C.R. and SCHAEFFER, B. (1976). On consistent estimation of age replacement intervals. *Technometrics* **18** 213-219.

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many parameters. *The Annals of Mathematical Statistics* **27** 887-906.

THOMAS, L.C. (1986). A survey of maintenance and replacement models for maintainability and reliability of multi-item systems. *Reliability Engineering* **16** 297-309.

VALDEZ-FLOREZ, C. and FELDMAN, R. (1989). A survey of preventative maintenance models for stochastically deteriorating single unit systems. *Naval Research Logistics* **36** 419-446.

DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY
POWAI, BOMBAY 400076
INDIA