Journal of the O. R. Society of Japan,
Vol. 7, No. 1, October 1964.

ON A PROBLEM OF PREVENTIVE MAINTENANCE

Y. ROLL and P. NAOR

Technion, Israel Institute of Technology, Haifa, Israel.

1. INTRODUCTION

Problems of preventive maintenance have recently drawn the attention of several researchers[1),2),3),4)].

Basically there are the following reasons for adopting a preventive maintenance procedure:

(a) To exercise control over production costs as compared with the best currently available alternative.

(b) To exercise control over the total maintenance costs including both scheduled and unscheduled attendance to, and operation of the equipment.

(c) To sustain a specified level of reliability.

For further development of this line of thought, it is necessary to define properly several terms which are used in the context of *preventive maintenance in the wide sense*:

Preventive replacement is concerned with the change or renewal of a relatively independent piece of equipment*.

Preventive maintenance (in the narrow sense**) is concerned with attendance to equipment during its lifetime, and carrying out such operations that are deemed necessary to keep it in running order.

---

* While, for reasons of convenience, we shall discuss the maintenance problem in terms of physical equipment, it is by no means limited to this. The same line of reasoning can be equally well applied to other "equipment" such as manpower, etc.

** As from here the term preventive maintenance will always refer to its narrow sense.

Preventive inspection is concerned with systematic observation, by some specified rule, which may possibly lead to further action such as preventive maintenance or replacement.

In actual practice, the operations of preventive replacement, maintenance and inspection are frequently interwoven.

The purpose of this paper is to describe and define a specific situation where both preventive replacement and preventive maintenance are of importance in connection with the minimization of total maintenance costs.

2. DEFINITION OF THE PROBLEM

We consider a piece of equipment rendering some service which is measurable in physical units such as time, volume of production, etc. For convenience, the physical unit chosen will be time. The equipment is subject of breakdowns which are random in the following sense: the life span, that is the time elapsing from the beginning of its operation to its termination (by breakdown), is a random variable governed by a density $f(t)$.

The level of preventive maintenance will be defined by the investment per unit of time, m say. It is then assumed that there exists a *response function* which links the level of preventive maintenance to the distribution of the life span, in particular to its expected value.

There are two distinct situations associated with the notion of replacement. In one, the old equipment is physically replaced by a new unit of the same type. In the other situation the equipment is subjected to a complete overhaul, reverting it to its initial condition. These two physically different situations are mathematically equivalent. As indicated in the definition of *preventive replacement* the word "replacement" will be applied to whichever method is in current use and no distinction will be made between them.

It is assumed that a cost $S(<0)$ is associated with replacement. A second cost arises in connection with unscheduled replacement. The stochastic character of the life span distribution is preserved even if the equipment is held at a high level of maintenance; hence it is liable to break down at unforeseen times and cause expenses additional to those connected with the replacement itself. We shall, then, assume a total cost $R(>S)$ to be associated with the process of unscheduled replacement. A third cost, previously mentioned, is proportional to the effort expended in preventive maintenance, and is designated by m. A policy will contain rules specifying:

(a) a period T (measured from the last replacement) at the end of which the old equipment is replaced.

(b) an expenditure m per unit of time to keep the equipment at a required maintenance level.

It is desired to determine a policy which minimizes the total cost of keeping the equipment in running and producing order. The set of feasible policies to be examined will contain all pairs (T, m) including the extreme cases $(T, 0)$ and (∞, m).

While the problem is now defined in terms of a cost criterion, it is necessary for the construction of the mathematical model to make specific assumptions regarding the life span distribution of the equipment as well as its response to preventive maintenance.

3. LIFE SPAN DISTRIBUTIONS

The life span of equipment is the period elapsing between the time it is put into operation (either new or renewed) and the time at which its activity is terminated by a breakdown. This quantity is a random variable and its variability may be represented by a function setting forth the probability of the life span exceeding time t; this is distribution function (cumulative to the right) $F^*(t)$. Other useful representations* are a) the density function $f(t)$ which is the negative derivative of $F^*(t)$, and b) the intensity function $\lambda(t)$ which is the negative logarithmic

* Assuming, of course, that the distribution function is continuous and differentiable everywhere in $(0, \infty)$

derivative of $F^*(t)$. The latter function, in particular, appears to be convenient for describing the notion of *reasonable life time distributions* since $\lambda(t)$ is a measure of the breakdown tendency as time proceeds. In principle the only restriction imposed on intensity functions is nonnegativity; however, in the context of mechanical equipment and its maintenance, non-decreasing intensity functions only appear of represent real phenomena.

A further requirement which we impose on the selection of our intensity functions in the number and specification of parameters. In this study we shall concern ourselves with two-parameter functions. On the one hand such functions are quite flexible and thus may be well abjusted to a wide range of possible real situations; on the other hand arbitrariness is rather circumscribed in the two-parameter families.

One of these parameters, α, will possess a physical dimension which, obviously, must be a power of the dimension of t; no generality is lost by arranging the dimension of α to be that of *reciprocal time.* The other parameter, β, will be a pure dimensionless number representative of rela-

Note on Weibull distribution: The conventional representation is not displayed in this Table; it may be obtained by an elementary transformation.

Note on Gamma distribution: Values of the functions associated with the gamma distribution are available in Pearson's Tables of the Incomplete Γ-function [5] and in Molina's Tables of Poisson's Exponential Limit [6].

Note on the truncated logistic distribution: The integral in the formula expressing the second moment is related to Euler's dilogarithm. This function has been discussed in some detail in a recent monograph [7]. Some numerical tabulations may be found there, but it is not difficult to derive numerical values by expansion into a series (setting $\beta=1+\epsilon$):

$$\int_1^\beta \frac{\ln\beta'}{\beta'-1}d\beta' = \int_0^\epsilon \frac{\ln(1+\epsilon')}{\epsilon'}d\epsilon' = \sum_{i=1}^{\infty}(-1)^{i-1}\frac{\epsilon^i}{i^2} \qquad (\epsilon<1)$$

$$\int_1^\beta \frac{\ln\beta'}{\beta'-1}d\beta' = \int_0^\epsilon \frac{\ln(1+\epsilon')}{\epsilon'}d\epsilon' = \frac{\pi^2}{6} + \frac{(\ln\epsilon)^2}{2} - \int_0^{1/\epsilon}\frac{\ln(1+\epsilon')}{\epsilon'}d\epsilon' =$$

$$= \frac{\pi^2}{6} + \frac{(\ln\epsilon)^2}{2} + \sum_{i=1}^{\infty}(-1)^i \frac{1}{\epsilon^i i^2} \quad (\epsilon\geq 1)$$

Table I. Some Distribution Functions of non-Negative Random Variables and their Properties

Families / Properties	Weibull Distribution	Gamma Distribution	Truncated Logistic Distribution
Distribution $F^*(t)$	$e^{-(\alpha t)^\beta}$	$\frac{\alpha^\beta}{\Gamma(\beta)}\int_0^t t'^{\beta-1}e^{-\alpha t'}dt' = \frac{\Gamma_{\alpha t}(\beta)}{\Gamma(\beta)}$	$\frac{\beta e^{-\alpha t}}{1+(\beta-1)e^{-\alpha t}}$
Density $f(t)$	$\alpha^\beta \beta t^{\beta-1}e^{-(\alpha t)^\beta}$	$\frac{\alpha^\beta t^{\beta-1}}{\Gamma(\beta)}e^{-\alpha t}$	$\frac{\alpha\beta e^{-\alpha t}}{[1+(\beta-1)e^{-\alpha t}]^2}$
Intensity $\lambda(t)$	$\alpha^\beta \beta t^{\beta-1}$	$\frac{\alpha^\beta t^{\beta-1}}{\Gamma_{\alpha t}(\beta)}e^{-\alpha t}$	$\frac{\alpha}{1+(\beta-1)e^{-\alpha t}}$
First Moment $E(t)$	$\frac{1}{\alpha}\Gamma\left(1+\frac{1}{\beta}\right)$	$\frac{\beta}{\alpha}$	$\frac{\beta\ln\beta}{\alpha(\beta-1)}$
Second Moment $E(t^2)$	$\frac{1}{\alpha^2}\Gamma\left(1+\frac{2}{\beta}\right)$	$\frac{\beta(1+\beta)}{\alpha^2}$	$\frac{2\beta}{\alpha^2(\beta-1)}\int_1^\beta \frac{\ln\beta'}{\beta'-1}d\beta'$

tive spread and (possibly) of other properties of the distribution function. Specifically in this study we are concerned with three families of distributions all associated with non-decreasing intensities (for a suitable choice of β): a) Weibull distributions, b) gamma distributions, and c) logistic distributions truncated on the left. Their relevant properties are summariszed in Table I.

The distributions as displayed in Table I are not necessarily in their conventional parametric form. Rather, they are being presented with a choice of parameters* such that a) they are significant in our context if, and only if, β does not fall short of 1, b) for $\beta=1$ all three families yield the exponential distribution with parameter α, and c) they are meaningful as distributions even for the interval $(0<\beta<1)$ albeit they possess decreasing intensity functions.

A distinguishing line may be drawn between the Weibull distributions and the two other families. In the Weibull distributions the intensity function tends to infinity with increasing time; in the other families there exists a limiting value which is approached asymptotically by the intensity function $\lambda(t)$.

4. RESPONSE FUNCTIONS

The response function represents the connection between the preventive maintenance effort and the life span distribution. The effort may be measured in expenditure per unit time or per unit of output.

We make one basic assumption regarding this connection: the response of the life span distribution is time-homogeneous. This means the following; whenever preventive maintenance at any level is applied, the form of the distribution function is not changed and the parameter β is not affected; the preventive maintenance makes its influence felt only through alteration of parameter α. In other words: on applying preventive maintenance the form of the survival function F^* (distribution function to the right) is preserved but the scale along the axis of time is homogeneously contracted.

Requirements of a reasonable response function are the following:

a. It is a positive, non-increasing function.

---

* We note, in passing, that our choice is associated with conveniently simple relations (precise or approximate) connecting the parameter β and the coefficient of variation, γ.

1) *Weibull distribution.*

Rough approximation: $\gamma^2 \approx \frac{1}{\beta^2}$

Good approximation: $\gamma^2 \approx \left(\frac{16}{\pi}-5\right)\frac{1}{\beta}+\left(6-\frac{16}{\pi}\right)\frac{1}{\beta^2}=\frac{.093}{\beta}+\frac{.907}{\beta^2}$

2) *Gamma distribution.*

Precise relation: $\gamma^2 = \frac{1}{\beta}$

3) *Truncated logistic distribution.*

Rough approximation: $\gamma^2 \approx \frac{1}{\sqrt{\beta}}$

Good approximation $\gamma^2 \approx \frac{1}{\sqrt{\beta}} + \frac{1}{18}\left(\frac{\beta-1}{\beta+1}\right)^2$

b. It is continuous and differentiable.

c. When m tends to infinity α_m approaches asymptotically a value $\alpha_\infty(\geq 0)$.

We do not impose the requirement that the second derivative of α is positive over the whole range; however, the actual response functions chosen possess this property which appears to us to represent real phenomena. Two relatively simple functions conforming to the above requirements are:

$$\alpha_m=(\alpha_0-\alpha_\infty)e^{-km}+\alpha_\infty=\alpha_0[\omega e^{-km}+(1-\omega)] \tag{4.1}$$

and

$$\alpha_m=\frac{(\alpha_0-\alpha_\infty)}{1+km}+\alpha_\infty=\alpha_0\left[\frac{\omega}{1+km}+(1-\omega)\right] \tag{4.2}$$

were k is the response constant of the function, and ω the fraction of α_0 which may be eliminated by applying preventive maintenance ($\omega=(\alpha_0-\alpha_\infty)/\alpha_0$).

There exists, of course, a multitude of functions other than (1) and (2) which possess the reasonable properties given above; however, a very few only possess the property of simplicity. In the following, (1) and (2) only make their appearance whenever explicit response functions are being utilized but it is not difficult to apply the theory to other functional forms.

5. COST FUNCTION AND OPTIMIZATION PROCEDURE

Consider the policy (T, m). The total expenditure in unit time, q, is then given by

$$q=\frac{R[1-F^*(T, m)]+SF^*(T, m)}{L(T, m)}+m \tag{5.1}$$

where $L(T, m)$ is the expected life span under preventive maintenance effort m and truncation T. We have then

$$L(T,\ m)=\int_0^T F^*(t,\ m)dt \tag{5.2}$$

The postulated property of time-homogeneity of the response enables us to utilize the various functions associated with the life span distribution in a dimensionless representation. In other words, it is possible to represent these functions in such a manner that they are made to depend on the product αt (and, of course, on the parameter β) throughout the optimization procedure. The moments of the distribution of αt are functions of β only, in particular

$$E(\alpha t)=\alpha E(t)=K_1(\beta) \tag{5.3}$$

and

$$E(\alpha t)^2=\alpha^2 E(t^2)=K_2(\beta) \tag{5.4}$$

For the purpose of comparing several distribution functions among themselves it is useful to introduce a standardized variable, τ say, whose expected value equals one. Thus we have

$$\tau=\frac{\alpha t}{K_1(\beta)} \tag{5.5}$$

and the coefficient of variation, γ, is given by

$$\gamma^2=V(\tau)=\frac{K_2(\beta)}{K_1^2(\beta)}-1 \tag{5.6}$$

Corresponding functions (on t, αt and τ, resp.) are related to each other in the following way

$$F^*(t;\ \alpha,\beta)=G^*(\alpha t;\ \beta)=\Phi^*(\tau;\ \beta) \tag{5.7}$$

$$f(t;\alpha,\beta)=\alpha\, g(\alpha t;\beta)=\frac{\alpha}{K_1(\beta)}\varphi(\tau;\ \beta) \tag{5. 8}$$

$$\lambda(t;\alpha,\beta)=\alpha\, h(\alpha t;\beta)=\frac{\alpha}{K_1(\beta)}\eta(\tau;\ \beta) \tag{5. 9}$$

$$L(t;\alpha,\beta)=\frac{1}{\alpha}\Lambda(\alpha t;\beta)=\frac{K_1(\beta)}{\alpha}\psi(\tau;\ \beta) \tag{5.10}$$

and we note again that these transformations are of an extremely simple type—all they involve is a linear change of scale in the variable.

If we define

$$\xi = \frac{R}{R-S} \tag{5.11}$$

relation (3) may be rewritten as

$$q = R\frac{1-\xi^{-1}F^*(T,\ m)}{L(T,\ m)} + m = R\alpha\frac{1-\xi^{-1}G^*(\alpha T)}{\Lambda(\alpha T)} + m =$$

$$= \frac{R\alpha}{K_1(\beta)} \cdot \frac{1-\xi^{-1}\Phi^*(\tau)}{\phi(\tau)} + m \tag{5.12}$$

Expressions similar to (14), where m was *not* considered a variable to be controlled, have been given by several authors[2)4)].

Partial differentiation of q with respect to both T and m, and setting the derivatives equal to zero, leads to the optimal policy (T^*, m^*). The previously introduced assumption of time-homogeneity of the response greatly simplifies further analysis. The partial derivative of q with respect to T (for constant and arbitrary m) is given by

$$\left(\frac{\partial q}{\partial T}\right)_m = \alpha\left(\frac{\partial q}{\partial(\alpha T)}\right)_m = \frac{\alpha}{K_1(\beta)}\left(\frac{\partial q}{\partial \tau}\right)_m$$

$$= R\alpha\frac{\Lambda(\alpha T)\xi^{-1}g(\alpha T) - [1-\xi^{-1}G^*(\alpha T)]G^*(\alpha T)}{\Lambda^2(\alpha T)}$$

$$= R\alpha\frac{\xi^{-1}\phi(\tau)\varphi(\tau) - [1-\xi^{-1}\Phi^*(\tau)]\Phi^*(\tau)}{T_1^2(\beta)\phi^2(\tau)} \tag{5.13}$$

Setting expression (5.13) equal to zero yields the optimal value—denoted by a star—of αT or, equivalently, of τ.

$$\Lambda(\alpha T)^*h(\alpha T)^* + G^*(\alpha T)^* = \phi(\tau^*)\zeta(\tau^*) + \Phi^*(\tau^*) = \xi \tag{5.14}$$

Thus the optimal value of τ (or, alternatively, of αT) does not depend on m. In other words, a quantity τ^* may be found which depends only on the functional form of the life span distribution (including, of course, the value of the parameter β) ond on the external conditions as expressed by ξ. Once the value of τ^* has been established an optimal $T_m{}^*$ may be found for any arbitrary m by the simple relation:

$$T_m^* = \frac{(\alpha T)^*}{\alpha_m} = K_1(\beta)\frac{\tau^*}{\alpha_m} \tag{5.15}$$

The next step leads to overall optimization. Combination of (5.12) and (5.14) results in

$$q_m^* = \frac{R\alpha h(\alpha T)^*}{\xi} + m = \frac{R\alpha\eta(\tau^*)}{K_1(\beta)\xi} + m \tag{5.16}$$

Minimization of $q_m{}^*$ with respect to m yields the expression

$$\left(\frac{d\alpha}{dm}\right)^* = \frac{-\xi}{Rh(\alpha T)^*} = \frac{-K_1(\beta)\,\xi}{R\eta(\tau^*)} = \frac{-K_1(\beta)}{(R-S)\eta(\tau^*)} \tag{5.17}$$

For the two alternative response functions (4.1) and (4.2) the optimal preventive maintenance expenditure, m^*, is derived as

$$m^* = \frac{1}{k}\ln[\alpha_0\omega k(R-S)h(\alpha T)^*] = \frac{1}{k}\ln[\alpha_0\omega k(R-S)\eta(\tau^*)K_1^{-1}(\beta)] \tag{5.18}$$

$$m^* = \frac{1}{k}\{[\alpha_0\omega k(R-S)h(\alpha T)^*]^{\frac{1}{2}} - 1\} =$$

$$= \frac{1}{k}\{[\alpha_0\omega k(R-S)\eta(\tau^*)K_1^{-1}(\beta)]^{\frac{1}{2}} - 1\} \tag{5.19}$$

We may obtain numerical values of m^* for the three life span distributions under study (as well as others) by inserting the optimal value of the intensity function $\eta(\tau^*)$—defined by (5.14)—in (5.18) or (5.19).

The optimal interval for preventive replacement (or renewal) is found on further developing (5.15)

$$T^* = \frac{(\alpha T)^*}{\alpha_{m^*}} = K_1(\beta)\frac{\tau^*}{\alpha_{m^*}} \tag{5.20}$$

Total expenditure per unit time under an optimal replacement and maintenance policy, q^*, for the two response functions under consideration are

$$q^*=\frac{1}{k}[\ln q_0{}^*k\omega+q_0{}^*k(1-\omega)+1] \tag{5.21}$$

and

$$q^*=\frac{1}{k}[2(q_0{}^*k\omega)^{\frac{1}{2}}+q_0{}^*k(1-\omega)-1] \tag{5.22}$$

where $q_0{}^*$ denoies the expenditure in unit time for an optimal replacement policy with *no* preventive maintenance ($m=0$),

$$q_0{}^*=(R-S)\alpha_0h(\alpha T)^*=(R-S)\alpha_0\eta(\tau^*)K_1{}^{-1}(\beta) \tag{5.23}$$

6. DISCUSSION

The approach of this study to prentive maintenance and replacement assumes detailed knowledge regarding both the life span distribution of the equipment under consideration and its response to the maintenance effort. In practice, the functional forms of the distribution and of the response function may not be available; again the functions may be known but the associated parameters must be estimated from experimental data. Typically the information on hand is rather scanty—usually a few samples of life spans under varying maintenance policies are available. This fact gives rise to a number of difficulties appearing on several levels.

It is usually proper to assume that m, R and S are constant and known. Furthermore, it is ordinarily possible to carry out relatively simple experiments determining the values of $K_1(\beta)/\alpha_0$ and of ω. However, it is frequently not feasible—economically and/or physically—to evaluate $\beta$with desirable precision; we take note that knowledge of the value of β is essential for utilization of key equation (16). It is also possibly rather expensive to obtain a good estimate of k, another prominent quantity in optimality considerations. In short, one basic difficulty in the utilization of the theory of this study is the fact that—*assuming the functional forms of distribution and response to be known*—important parameter values may be unknown and their substitution by estimates insufficient for the purposees of formulating optimal policies.

Another difficulty is related to the above: Suppose that a reasonable supply* of data exist for the estimation of β and of k, but that no prior information is available as to the functional form of the distribution and of the response; how are optimal policies formulated? Ordinarily, in such a situation the practitioner would select a functional form for the distribution which is 1) not too difficult to handle mathematically, and 2) compartible with the known mechanism of determination and breakdown as well as with the observations on hand; also he would proceed similarly with respect to the response function. This is certainly not an objectional procedure as long as one bears in mind that very same set of experimental data may be compatible with a number of different distribution and response forms. These may *possibly* lead to widely divegent optimal strategies. Indeed one would suspect that a fundamental difference exists between distributions whose intensity functions grow beyond all bounds with increasing time (e.g. the Weibull distribution) and distributions whose intensity functions tend to some asymptotic value with increasing time (e.g. the gamma and the logistic distribution). Clearly, in the latter category there exists-by (5.14)—a critical value of ξ ($\xi_{\text{crit}}=E(t)\lambda(\infty)=E(\alpha t)h(\infty)=E(\tau)\eta(\infty)$) such that for $\xi\geq\xi_{\text{crit}}$ the optimal policy is to have unscheduled replacement only. For the other hand, an optimal T^* can always be specifiied. While the distinction between these two categories is of theoretical interest it usually has no practical significance: if ξ is large enough, $\Phi^*(\tau)$, the probability of undertaking a

* By the expression "reasonable supply of data" roughly the following is meant: Given the functional form of the distribution and of the response, the experimental data suffice in order to construct confidence intervals (with given confidence coefficient) for k and β, smaller than prescribed intervals.

scheduled replacement, becomes very small so that the optimal strategy is (practically) associated with the rule of having unscheduled replacements only regardless of the class to which the intensity function belongs. We find it difficult to make *general* statements about the robustness of our model if only because the set of feasible distributions (all possessing identical expectations and identical variances) appears not to have a simple metric useful for our purposes: we have no obvious way of establishing whether two different distribution assumptions are near to each other or are far apart. On a more pragmatic level we found that combining a given response function with each of the tree specific investigations usually yielded similar results* in terms of q^*, though frepuently the T^*-s and the m^*-s were rather different. Moreover substituting one response function for the other introduced no serious change if good care was taken that the two response functions were close to each other within the interesting range.

The final (and possiby most serious) difficulty is concerned with the the assumed property of time-homogeneity of the response. In principle, this property is verifiable but frequently in practice this does not appear to be feasible. The mathematical analysis of this paper hinges on the assumption and ond one cannot expect to retain robustness of the model—in the sense of approximate validity of the final results-in the absence of this property. Indeed it is difficult to hold on to any "final results" under such circumstances since the notions of k and β—both appearing in final formulas—are defined only by means of this property. This fact places limitations on the application of the theory of this paper, but at least they are clear and easily recognizable.

It is our experience that the theory of preventive maintenance presented here closely reflects a class of real life phenomena. The area of its applicability is demarcated in the above Discussion.

---

* That is, the positive cost increment—brought about by using parameters optimized on assumptions other than those representing "reality"—is small compared with total optimal expenditure q^*.

REFERENCES

(1) Morse P.M. Queues, Inventories and Maintenance, John Wiley and Sons, New York, 1958.

(2) Barlow R. and Hunter L., "Optimum Preventive Maintenance Policies," Operations Research *8*, 90–100, 1960.

(3) Kimball G.E., "Reliability and Maintenance," Notes on Operations Research, The Technology Press-M.I.T., Cambridge, Mass., 1959.

(4) Senju S., "A Probabilistic Approach to Preventive Maintenance," Journal of the Operations Research Society of Japan, *1*, 49–58, 1957.

(5) Pearson K., Tables of the Incomplete Γ-Function, Cambridge University Press, Re-issue 1946.

(6) Molina E.C., Poisson's Exponential Binomial Limit, D. Van Nostrand, New York, Fourth Printing, 1947.

(7) Lewin L., Dilogarithms and Associated Functions, MacDonald, London, 1958.