MODELLING CONDITION MONITORING INSPECTION USING THE DELAY-TIME CONCEPT

By

WENBIN WANG

A thesis presented for the Degree of Doctor of Philosophy at the University of Salford, U.K.

Department of Mathematics and Computer Science

1992

CONTENTS

LISTS OF TABLES

LISTS OF FIGURES

ACKNOWLEDGEMENTS

I would like to acknowledge my debt of gratitude to my supervisor, Professor A.H. Christer, for his suggesting the topic of this research, untiring interest, constant motivation, enthusiastic support and elegant polishing of the final thesis.

Thanks are also due to Dr. Rose Baker who lent her cheerful collaboration to our joint study and contributed particularly to Chapter 4 and 5 of this thesis.

I would also like to thank Miss Susan Sharples, for her enthusiastic help during the preparation of this thesis.

I also wish to thank my home University, Harbin Institute of Technology, which gave me the chance to study abroad at Salford University.

Finally, I am greatly indebted to my wife and daughter for their patience and understanding during the past two years.

ABSTRACT

In the literature on inspection modelling, the failure distribution traditionally plays a fundamental role in model construction in that it is assumed that system failures occur instantly at random time points from new with a known pdf. of time to failure. Numerous models have been built on this basis. However, Professor Christer challenged this traditional idea and proposed the concept of delay time. The idea, which is an essential part of most engineers' experience, assumes that defects do not just appear as failures, but are present for a while before becoming sufficiently obvious to be noticed and declared as failures. The time lapse from when a defect could first be identified at an inspection to consequential failure has been termed the "delay time". It is this idea which can be captured to reveal the nature and scope for preventive maintenance or inspection. It appears that the concept is now being taken up by many other authors.

In this thesis, various models for condition monitoring inspection are built on the basis of delay time analysis. Extensions and further developments are made here to enrich the delay-time modelling. Since the distribution of the delay time is important to delay time modelling, a new approach to estimate the delay time distribution is proposed. This technique, which contrasts with the previous subjective data estimation technique, is based upon objective data.

Assuming the distribution of the delay time is known, models of condition monitoring inspection are fully discussed for both perfect and imperfect inspections, and for infinite and finite time horizons. Based upon the models for perfect inspection, algorithms are presented to find the optimal solution. Numerical examples are presented in each Chapter to illustrate how models and algorithms work.

CHAPTER 1. INTRODUCTION

1.1 Maintenance, production and models

The production function manifests itself materially in the products that are manufactured. It is because of the use of means of production that maintenance is necessary, and consequently the costs of maintenance are to be attributed to the costs of the products manufactured. The maintenance models which aim at the optimization of maintenance must therefore be found in criteria which directly or indirectly minimize the cost of the products. This means that maintenance exists only by virtue of the fact that its function is derived from the production function through the need to output production. Seen in this way, it is possible to consider maintenance within an organization as a subsystem of production.

The role of maintenance in production has became paramount in many organizations as the equipment or systems in use has become more automated, high volume, and expensive. This means that on the one hand it is necessary to keep the systems maintained in appropriate running condition, whilst on the other hand it is necessary to reduce the maintenance cost in order to reduce the costs of production. It has been estimated that somewhere in the region of £12,000 million per annum is spent on maintenance of equipment in British industry. This may appear to be a large figure and to offer scope for considerable saving. Maintenance modelling has a significant contribution to make because it can help the decision-maker establish a cost effective balance between maintenance cost and equipment reliability.

As a result of technological take-up and improvement in industry, one of the consequences to maintenance of the advances in manufacturing technology and automated production is a change in emphasis from monitoring the quality of a product being manufactured to monitoring the quality and condition of the plant manufacturing the product. This leads to the main topic of this thesis, — condition monitoring inspection and its models.

1.2 Condition monitoring inspection and modelling

Industry world-wide understands the conventional time scheduled maintenance known as preventive maintenance, which is essentially based upon manufacturers' or suppliers' recommendations. Corrective maintenance is also well known, and followed when breakdowns occur or a mal-function develops. These conventional practices have some inherent deficiencies, the most important being excessive maintenance cost, and high manpower and inventory requirements. For the more complex plants, particularly process type, monitoring check devices and methods have been rapidly developed in recent years. This means that an alternative to preventive maintenance has emerged and has become widely used in industry, namely, condition based maintenance. The basic idea of condition based maintenance is to base maintenance decision making for plant upon monitored condition related information. There are two main types of condition based maintenance currently in use in industry, i.e., on-line condition self-scheduled monitoring and periodic condition monitoring inspection. The former is usually the continuous monitoring type, such as on-line condition monitoring of a flexible manufacturing cell. In this thesis we will focus on the latter, namely periodic condition monitoring inspection (for convenience, it will be simply called as inspection in the following text). This is the area with decision issue to raise and address and, therefore, presents a host of modelling opportunities for OR. and statistics. Like much maintenance, it is a relatively novel area for the actual applications of OR. and statistical modelling.

There is an important maintenance concept which has been developed by Christer *et al* since 1973, designed to reveal the nature and scope of preventive maintenance or inspection, namely the concept of delay time and delay time analysis which will be explained in more detail later. The important contribution of the delay time concept to maintenance theory and practice is that it models the failure process as a two stage process in which the first stage is from new to the time of a defect becoming first visible, and the second stage is from this visible time of a defect to a subsequent failure. The time length of this second stage is called the delay time. It is the existence of such a delay time for a defect that establishes the opportunity for effective preventive maintenance under an inspection system. Defects are identifiable during the delay time when an appropriate preventive maintenance action such as repair or replacement may be undertaken and so prevent equipment from incurring a failure.

In this thesis, various models and algorithms are constructed using the concept of the delay time and delay time analysis. In addition to condition monitoring inspection modelling within the scope of delay time analysis, methods developed to estimate the delay time distributions are also established by using the maximum likelihood method. Some of the models developed in this thesis can also be applied to more general areas and are not necessarily restricted to the field of condition monitoring inspection. The objective of our models here is to devise an inspection schedule so as to strike an appropriate cost or availability balance between the cost or time of inspection and the cost or time of additional or more serious failure which arises through not inspecting. It is well known that a defect can arise in a stochastic fashion which may or may not be resolved completely or partially by corrective action and the signal resulting from the monitoring checks may or may not be interpreted correctly by the decision maker. The decision problem relates to the choice of monitoring checks to apply, when to apply them, and the appropriate action to take subsequent to the monitoring check results. These issues will be addressed in our models.

Finally, we would like to point out that the general notation of equipment inspection, and the specific notation of condition monitoring inspection in particular, are very similar in concept. However whereas condition monitoring inspection adopts some specialized tools, such as via the analyses of oil, vibration, thermograph and ultrasonic, a general inspection is usually a manual and visual process. The former would usually be expected to give more accurate information on a component's condition than the latter, since equipment diagnostic techniques have rapidly developed during recent years. Whilst some special models in the following text may only apply to the case of condition monitoring inspection utilizing information of condition monitoring, most of models here can be generalized to the case of ordinary inspection.

CHAPTER 2. LITERATURE REVIEW

Over the last twenty-five years, OR. and mathematics have played a leading role in developing and exploring models of inspection situations. The target and objective of much of the modelling work has been to assist management by predicting the consequences of alternative sets of decision variables available to them. As the mathematical sophistication of inspection models has increased in parallel with the growth in the complexity of modern systems, numerous papers have been published in this field, among which is a series of major surveys made by McCall [1965], Pierskalla & Voelker [1975], Christer [1984], Thomas [1986], Valdez-Flores & Feldman [1989] and Cho & Parlar [1991]. These embrace several hundred papers. In addition to these surveys, bibliographic references can be found in Osaki & Nakagawa [1976], Sherif & Smith [1981], and Sherif [1982]. Some classical books on reliability and maintenance are also concerned with the contents of inspection, examples are Barlow & Proschan [1965], Jardine [1973], and Ascher & Feingold [1984].

Inspection models usually assume that the state of a system is completely unknown unless an inspection is performed. Every inspection is normally assumed to be perfect in the sense that it reveals the true state of the system without error. In general, at every inspection epoch there are two decision that have to be made. One decision is to determine what maintenance action to take, whether the system should be replaced or repaired to a certain state or whether the system should be left as is. The other decision is to determine when the next inspection is to occur. Thus, in general the decision space of a maintenance inspection problem is two dimensional. If however, a choice of what condition test to apply at the next inspection epoch also exists, the inspection decision space is three dimensional.

Different authors have produced many interesting and significant results for variations of inspection models. The different models developed depend on the assumptions made regarding the time horizon, the amount of information available, the nature of cost functions, the objective of models, the system's constraints, etc. The different models are, for the most part, however, very similar to a basic model presented by Barlow *et al* [1963]. This basic model is a pure inspection model for age replacement; i.e., no preventive maintenance is assumed, and the system is replaced only on failure.

The basic model assumes that (a) system failure is known only through inspection, (b) inspections do not degrade the system, (c) the system can not fail during inspection, (d) each inspection costs c_1, and the cost of leaving an undetected failure is c_2 per unit time, (e) inspection ceases upon discovery of failure. Hence the total cost per inspection cycle is given by

$$C(t,\mathbf{x}) = c_1 n + c_2(x_n - t), \qquad 2\text{-}1$$

where t is the time to failure , $\mathbf{x} = (x_1, x_2, \ldots)$ is the sequence of inspection times with $x_1 < x_2 < x_3 < \ldots$, and n is the inspection which detects the failure, that is $x_{n-1} < t \le x_n$. Usually the optimal inspection policy $\mathbf{x}^*$ is the one that minimizes $E[C(t,\mathbf{x})]$. However there two problems concerning this basic model. The first one is related to the detection of a failure. Since the model has assumed that the system can not fail during an inspection, this means that it either fails before or after an inspection. From the first assumption, we know that if there is a failure at time t ($x_{n-1} < t \le x_n$), it can only be identified at time x_n. However, in fact, this assumption would appear questionable in practice because if a failure occurs, the system would normally be examined and a repair or replacement be undertaken. It is generally impractical to leave the failed system until the next inspection. The second problem, which is actually related to the first one, is the possibility of obtaining the value of constant c_2 in practice. Obviously it is hard to define and estimate this value. However even with these practical drawbacks, many authors have made further contributions to this basic model because of their theoretical interest.

Using the basic model of Barlow *et al*, Beichelt [1981] determines the optimal times for the cases when replacement and no replacement of a failed system are permitted. He obtains the optimal inspection schedules when the lifetime distribution is partially unknown. Beichelt uses a minimax approach to find the optimal inspection for the case of partially unknown lifetime distribution but does not indicate a numerical procedure to obtain the optimal scheduling of inspections.

A different approach is used by Luss [1976] who looks at a system where a degree of deterioration can be observed through inspections. An inspection reveals that the system is in one of several intermediate states of deterioration. State-dependent maintenance policies are determined to minimize the long-run expected cost per unit time. He assumes that at inspection times the system may be found in any state $0,1,\ldots,L$. If the system is in state L, the failed state, it is immediately replaced at a higher cost than if it were replaced before a failure. The replacement cost at any other state is constant. Luss presents a very simple iterative procedure that finds the optimal control limit policy with control state α and the optimal inspection interval for states $0,1,\ldots,\alpha-1$. He assumes that the sojourn times, that is the time spent at any state $i=0,1,\ldots,L-1$, follow an exponential distribution with parameter λ.

A similar model is presented by Sengupta [1981]. He, however, lets the replacement cost be an increasing function of the deterioration states and allows a delayed replacement action. He shows that the policy that minimizes the long-run expected cost per unit time calls for inspection and delayed replacement intervals that are decreasing in the deterioration state. He also shows that the optimal solution is a control limit policy when replacements are made at inspection times. Sengupta gives an iterative algorithm that computes the optimal intervals.

Zuckerman [1980] examines a model very similar to Luss's model. Zuckerman presents a maintenance model in which the status of the system can be determined through inspection. At failure detection, the system should be immediately replaced by a new identical one. The costs incurred include inspection costs, operating costs, failure costs, and pre-planned replacement cost. He restricts the inspection policy to periodic inspections. The decision variables include the inspection interval and the scheduling of preventive replacements. The problem is to specify an inspection-replacement policy that minimizes the long-run expected cost per unit time. It is also assumed that a failure is discovered only by inspection. Zuckerman considers that the system is subject to a sequence of shocks with exponential distribution between occurrences and that each shock causes a random amount of damage that adds to the degradation of the system. The state of the system can then be any real non-negative number. He assumes that inspection and replacement are instantaneous. Zuckerman shows that the optimal replacement policy is a control limit policy, provided some conditions are satisfied. He does not present a general algorithm to compute the optimal policy, but notes that the difficulty in finding it depends heavily on the structure of the survival function of the system and the distribution of the magnitude of the shocks.

Abdel-Hameed [1987] generalizes the compound Poisson process used by Zuckerman and allows a more general damage structure.. Abdel-Hameed uses an increasing pure jump Markov process to model the deterioration. The system fails whenever the deterioration level is greater or equal to a threshold and is immediately replaced at a cost which is higher than the cost of replacing the system before failure. The deterioration level of the system is monitored periodically. He finds the optimal inspection period that minimizes the long-run expected cost per unit time.

A periodic inspection policy for Barlow's model, equation 2-1 is optimal when the failure distribution of the system is exponential, Barlow *et at* [1963]. For models that do not assume exponential failure times, a periodic inspection policy is not necessarily optimal. Rosenfield [1976] presents a model in which the system is considered to deteriorate according to a discrete time Markov chain. His major contribution is to prove that, under some condition on the transition probability matrix and the inspection, replacement and operating costs, a monotonic four-region policy is optimal. Rosenfield presents the models for both the long-run expected cost per unit time and the total expected discount cost. He does not present any specific algorithm, but the optimal solution can be obtained using standard policy iteration for Markov decision processes. White [1979,1978] investigates the same problem as Rosenfield and proves the same results under less restrictive conditions.

Kander [1978] considers inspection for a system that can be classified into discrete deterioration levels. Kander models the problem using semi-Markov processes to determine the optimal inspection schedule that minimizes the long-run expected cost per unit time. He considers three possible inspection policies called pure checking, truncated checking, and checking followed by monitoring. Under a pure checking inspection policy, successive check times are based on the last state of the system observed. Under a truncated checking inspection policy, the states of the system are essentially good or failed. If at an inspection time the system is in good state, a decision is made to determine the next inspection time; however, if the system is found in a bad state, the unit is replaced and the cycle is completed. For the checking followed by monitoring inspection policy, the states of the system are divided into two sets. If at inspection the system is in the set of states that is considered good, the next inspection time is determined as in the truncated checking case and no monitoring occurs. However, if the system is the state that belong to the set that is considered as not good, the system is continuously monitored at a certain cost until failure of the system occurs. When failure is detected the system is immediately replaced. Kander does not show a numerical procedure for obtaining the optimal policies. However, he gives an example in which the solution is found analytically, although this type of solution can not always be obtained.

An algorithm which is also adopted by Barlow & Proschan [1965], Jardine [1973], is presented by Brander [1963] in which an optimal inspection policy can be obtained for long-run expected total cost per unit time. He transforms the traditional formula of expected total cost per unit time into another one which can be simply minimized under some conditions with an extra parameter α. By varying α and x_1, an optimal inspection schedule can be obtained through a certain recursive relationship among x_i(i=2,3,...), x_1, and α. The cost function he adopted is the same as equation 2-1 and a numerical example is also presented in his paper.

A modified inspection model is proposed by Nakagawa [1984]. He considers a system that is checked periodically to see whether or not it needs to be replaced. If the system is not in good condition, it is immediately replaced. In this model the system has the same age after checks as before with probability p and is as good as new with probability q = 1 - p. He obtains the mean time to failure and the expected number of inspections before failure using a renewal-type equation. Nakagawa then investigates the properties of the mean time to failure and the expected number of inspections to failure when the failure rate of the system is increasing. He also derives the total expected cost and the expected cost per unit time until failure. Nakagawa notes that it is very difficult to obtain an analytic solution for the optimal inspection times, and suggests the use of a numerical search procedure to find them.

Menipaz [1979] considers an inspection model where inspection and downtime costs change over time. He finds that the optimal inspection policies for cases in which (a) the system is inspected at discrete points in time, and replaced as soon as a failure is detected, (b) the system is inspected up to a predetermined age and is replaced if it has not failed, and (c) if the system is inspected at discrete points of time t, it is continuously inspected from then on and replaced at failure. The optimal inspection policies that minimize the total discounted cost for the different inspection strategies described are obtained using algorithms that had been published by Kander and Naor [1969] and Luss [1977]. Luss [1977] studies an inspection model in which the duration of inspections and repairs (replacement) are not negligible. He presents two algorithms to solve the problem.

Wattanapanom and Shaw [1979] give algorithms for finding optimal inspection times in case of uniform and exponential failure time distributions. They assume that every inspection is hazardous and may degrade a good system. Their main contribution is the presentation of convergent algorithms for solving the optimization equations given to solve the basic model, equation 2-1. For a system subject to failure at random time, Keller [1982] presents an asymptotic solution for the inspection model with the cost of inspection small compared to the expected loss due to downtime. He shows that the limiting form of the equation needed to find the optimal inspection times is a nonlinear ordinary differential equation. Schultz [1985] presents an approximate periodic inspection solution to the basic model under a general failure distribution. He claims that this approximation is good as long as the cost of inspection is small relative to the cost of undetected failure. Furthermore, Schultz's approximation is easily computed and only requires knowledge of the mean time to failure.

Nakagawa and Yasui [1980] present an algorithm to compute near-optimal inspection policies for the case when the distribution to failure is not exponential. They give a numerical example that shows that the approximation is fairly good for a Weibull distribution. The procedure computes successive inspection times backwards by a recursive scheme. When the hazard rate is increasing, Munford [1981] has shown that inspection policies with decreasing intervals between successive inspections as a function of age are superior to periodic policies.

It has been mentioned that some inspection models are constructed under the assumption of Markov deterioration. The first original model of this type for a single deteriorating unit is described by Derman [1963]. It will be useful to briefly recall this paper. A unit is inspected every period and the state of the unit is ascertained. It is assumed that if nothing is done then the unit deteriorates according to a Markov chain on a finite set of states $\{0,1,2,\ldots,k\}$, where 0 denotes a new unit and state k means the unit has failed completely. At each period, once the state is known, a decision has to be made whether to replace the unit, perform a preventive maintenance overhaul, or do nothing. The difference between replacement and repair is the difference of the state that the unit returns to after the performance of that action. In most variants of the above model economic criteria are used, in which one tries to minimize the sum of the maintenance cost, the cost of replacement/repair due to failure, and the cost of preventive repair/replacement. Derman shows that if the probability of deterioration next period increases with the present state i, then a 'control limit' rule is optimal, so that one should repair or replace when the observed state i is greater than some limit i^*.

There have been a large number of extensions of this original model. Derman's model did not involve a maintenance cost, but Kolesar [1966] adds the maintenance cost without affecting the optimality of the 'control limit' rule. Ross [1969] extends the problem to a more general state space while Kao [1973] allows the time in each state to be random and proves similar results for a semi-Markov model.

Inspection is usually assumed to be perfect. However this assumption can be relaxed by allowing for imperfect inspection in the modelling. Anderson and Friedman [1977,1978] present a very theoretical model which involves the imperfect inspection case. They find the optimal inspection times by reducing the stochastic problem to a free boundary problem in analysis, which is then solved using iterative procedures. Furthermore, inspection may pose a hazard to the system to be checked. Chou and Butler [1983] and Butler [1979] have studied hazardous inspection models for aging systems. They found optimal policies that minimize the expected lifetime of the system under inspection. Their model assumed that each inspection either causes immediate failure or else increases the failure rate. However, on the other side, if the inspection contains the content of preventive maintenance, it may add some benefit to the system to be inspected. Baker [1990] considers this problem and presents a model to test the effect of such preventive maintenance. A similar problem is also discussed in a paper by Baker and Wang [1992], see Chapter 5 of this thesis for details.

There are parallel but completely different models for inspection problem which were first proposed and substantially studied by Christer *et al* [1973 - 1992] in which the delay time concept has been used to construct the models for both perfect and imperfect inspections. Christer notices that there is a time lapse between the time of the first noticeable event of a potential failure of a component (he calls a defect arisen) to the time of this failure resulting in a breakdown and a repair or replacement is imminent. In other words, the component failure process is actually a two-stage process in which the first stage is from new to the time a defect becomes first visible, and the second stage is from the time of this visible defect appearing to a consequential failure. The time lapse of this second stage is called the delay time of a defect. Clearly the concept of the delay time gives a more physical explanation of the principle of preventive maintenance and also overcomes two disadvantages of Barlow's basic model, equation 2-1. In the fundamental paper of Christer and Waller [1984a], they present a prototype model of the industrial maintenance problem using the delay time concept. It is proposed that the pdf. of the delay time is assessed via the collection of subjectively based data. It is shown how such a data base makes it possible to construct models for determining the optimal interval between inspections or the optimal replacement time that minimizes the expected downtime per unit time or the expected operating cost per unit time, Christer and Waller [1984b,c]. Subsequently, a series of papers followed extending the basic model and applications to cover more cases of maintenance modelling. Preliminary studies have been undertaken to investigate the applicability of delay time analysis and delay time models to condition monitoring problems be they discrete monitoring processes, Christer [1987, 1988] or continuous monitoring processes, Chilcott and Christer [1991], or a specific condition monitoring problem, namely the condition monitoring of a bearing wear process in production plant, Christer and Wang [1992]. In a more recent paper, Christer [1992] establishes models of condition monitoring inspection with irregular inspection intervals. It is worth pointing out that throughout this thesis, all models and algorithms constructed use the concepts of the delay time and delay time analysis, and contribute further developments of delay time modelling.

Christer and Redmond [1990a, b] consider the mathematics of delay time analysis and prototype models of inspection policies. The essential role of subjective estimation is indicated and the need for revisions of both the subjectively based prior delay time distribution and of the delay time model is highlighted. They notice that an unavoidable bias arises when estimating delay time distribution and present a mechanism for correcting this bias, based upon maximum likelihood consideration. Other applications of delay time modelling can also be found in the papers by Gebbie and Jenkins [1986], and Pellegrin [1992] who develops a graphical tool for maintenance decision making based upon the delay time concept.

Turco and Parolini [1984] also present a model which is very similar mathematically to delay time modelling [see Christer 1987]. They state the hypothesis of a different damage rate before and after the exceeding of the alarm threshold. However they did not mention how to estimate the distribution function of the time to the damage state while using the Weibull or Erlang distribution as the example.

In traditional delay time modelling, the distribution functions of the delay time are estimated by subjective data which is quite useful when there are no historic records. When historical data on inspections and failures are available, contrasting with the previous subjective assessment method, Baker and Wang [1992] propose an alternative method to estimate the delay time distributions using maximum likelihood theory and the Akaike information criterion(AIC), which will be explained in detail in Chapter 4 and 5 of this thesis.

So far in the inspection models mentioned above we have not distinguished between the general inspection modelling and condition monitoring inspection modelling. However, there are very few papers specifically concerned with condition monitoring problems compared with the considerable literature on general inspection modelling. This shows that modelling condition monitoring inspection is a quite novel area.

A case study presented by Sullivan [1991] describes a condition monitoring practice from the Parenco paper mill in Holland. They use the Mrüel and Kjær Systematic Machine Condition Monitoring concept to monitor the vibration spectrum at 6000 measurement points on their paper machines and power plant. The system features early fault detection, powerful fault diagnosis, and trend analysis to predict the lead time to breakdown. However the methods they used in their system monitoring consist of engineering judgments. There is no attempt to rationally relate the time to failure to the condition of the system monitored.

Chilcott and Christer [1991] propose a model based upon continuous condition monitoring for maintenance at the coal face within British Coal, again using the delay time concept. This model is used to predict the effectiveness of condition monitoring using the resultant downtime of machinery as the relevant measure. Numerical examples of the model developed are presented using data obtained from collieries in the course of a research program. A discrete type of condition monitoring model is also presented by Christer [1988] specifically designed to model major civil engineering structures. The delay time analysis is the main frame of model building, but in this model the cost of a repair may now change over the delay time period.

Usually in condition monitoring models, the equal space monitoring interval is a common assumption. However the best monitoring check interval may not be necessarily equal. An irregular inspection interval could give a better monitoring result than the equal interval scheduled one since if the irregular inspection policy is optimal, it must also be superior to the regular inspection one. Christer [1992] presents a prototype modelling of irregular condition monitoring of production plant for which a detailed description will be given in Chapter 6 of this thesis. The models which are based upon the delay time concept are designed to model the perfect and imperfect inspection cases with the objective of minimizing the long-run expected total cost per unit time. Although the paper gives numerical examples to show the method, it is mainly concerned with formulating models. It does not address the solution of the models, which will be the subject of Chapter 8 of this thesis.

Christer and Wang [1992] propose a particular model of condition monitoring inspection of production plant. They consider a bearing wear problem in which the wear process is regularly monitored by a special device. The detail of this paper will be given in the Chapter 7 of this thesis.

In most papers of inspection or condition monitoring modelling, the time horizons are either infinite or the system life cycle. In the infinite case, an asymptotic form of long-run expected total cost per unit time is used since it has a more simple form. In fact, this model assumes that the failure process is a renewal type in which the system is replaced by an identical one upon failure and the process resumes. For a model based upon the system life cycle, the modelling objective is to minimize the expected total cost over the system life time. That is, upon detection of failure the problem ends and no replacement or repair takes place. This model which has been addressed by many others may apply to the cases of detecting the occurrence of an event, such as the arrival of an enemy missile or the presence of some grave illness. However, we can not assume that a system is always replaced by an identical one because of rapid development of technology. Some replacement models over finite time horizons have to be proposed. Christer and Jack [1991] present a model in which an integral-equation approach is proposed to calculate the exact and asymptotic estimates of expected costs in stochastic replacement problem over a finite time horizon. Although the paper focuses upon the replacement problem, some insight can be gained through that paper into inspection modelling. Jack [1991] also considers a similar problem which involves imperfect repair on failure. There appear to be no papers which are directly concerned with a finite time horizon for condition monitoring modelling. Therefore, some models on Chapter 6 of this thesis will contribute to this problem.

The majority of inspection or condition monitoring modelling papers assume that the distribution of time to failure is known, and models are built on that basis. Much interest has been shown in the literature over the years in the estimation of life time distributions in the field of medical data analysis and reliability. Cox [1972] proposes proportional hazard modelling which has been used in inspection modelling by many authors. Lee and Pierskalla [1987] consider models of mass screening for contagious and non-contagious diseases, which are quite similar to the inspection models we mentioned above, see also Eddy [1980] and Brookmyer et al [1986]. Since the statistical analysis of the duration of life time has become a topic of considerable interest to statisticians and OR. workers in areas such as medicine, engineering, and the biological sciences, hundreds of papers which contribute to the areas have been published during recent years. Besides these papers, several books by Mann *et al* [1974], Gross and Clark [1975], Kalbfleisch and Prentice [1980], Cox and Oakes [1984], and Lawless [1982] deal extensively with estimation and analysis procedures for lifetime data.

To estimate the distribution function of lifetime, the maximum likelihood method is a most frequently used one, see Cramer [1946], Kaplan and Meier [1958], Cox and Hinkley [1974], and Zacks [1971]. Usually the goodness of values of parameters of a specific model can be measured by the expected log likelihood, namely the larger the expected log likelihood the better the values of parameters. We know that an increase in the number of model parameters can cause the fit to a given sample of data to improve the maximum log likelihood. However, the maximum log likelihood has a general tendency to over estimate the true value of the mean expected log likelihood, and this tendency is more prominent for models with a larger number of free parameters. This means that if we choose the model with the largest maximum log likelihood, a model with an unnecessary large number of free parameters is likely to be chosen. Akaike [1973] challenges the traditional idea to parameter estimation and proposes a criterion called AIC (Akaike Information Criterion) for model selection. The 'best' model is the one which minimizes AIC. Baker and Wang [1992] have used the AIC in their paper on the estimation of delay time distribution, which will also be explained in Chapter 4 and 5 of this thesis.

In all, papers and books published over recent years which contribute to inspection or condition monitoring modelling cover the full range of potential applications of mathematics to maintenance management problems varying from mathematically based techniques developed to solve a specific and recognized problem type, to the development of mathematical refinement to models which, while adopting the language of maintenance and reliability, are mainly of interest to mathematicians and have little or no pretension of applicability. For example, most of the published theoretical models of plant inspection problems mentioned above adopt the Markov approach, while few of them consider real application in industry. However, they provide ideas for possible model-building blocks along with some qualitative insight as to how an actual system might behave if only it would oblige the model assumptions, the major interest is in the solution procedure. Notwithstanding this, for inspection and condition monitoring problems, the outstanding problem now is not in solving models, but in producing and validating them. Thus in this thesis, the main topic will fall in the former category, i.e., to identify the problem and to make the models applicable to maintenance practice.

CHAPTER 3. THE DELAY TIME CONCEPT AND MODELLING

3.1 The delay time concept

In traditional modelling of maintenance, the failure distribution plays a fundamental role in model construction in that it is assumed system failures occur instantly at random time points from new with a known distribution function of time to failure. If a failure occurs, the event is classified as a breakdown and a replacement or repair is usually undertaken to restore the system to the normal working state. Numerous models in terms of probability measurements have been built on this failure distribution basis. However, the assumption of failures occurring instantly without any prodromal symptom is unlikely in maintenance practice because there will often be a time lapse from the first noticeable signal of a potential failure to the time when it develops into a breakdown. At the early stage of a potential failure, a system may be defective but still running. For example a small crack in a concrete structure progressively develops over time into a failure when the concrete structure breaks. Recognizing this situation, Christer has challenged the traditional idea of failure assumption and proposed the delay time concept.

The delay time concept, which was originally developed as a side issue in modelling building maintenance, Christer [1973, 1982], exploits the ideas of a "delay time" for a fault in building structures. In the fundamental paper by Christer and Waller [1984a], the idea which is essential to most engineers' experiences, and indeed most papers referenced above, is proposed in which they assume that defects do not just appear as failures, but are present for a while before becoming sufficiently obvious to be noticed and declared as failures. The time lapse from when a defect could first be reasonably expected to be identified at an inspection to a consequential repair or replacement being necessary has been termed the "delay time", usually denoted by h. The time length of a system from new to the initial point of a defect being first visible by some inspection devices is called the "initial time" of a defect, usually denoted by u. The basic idea is also illustrated in Figure 3-1.

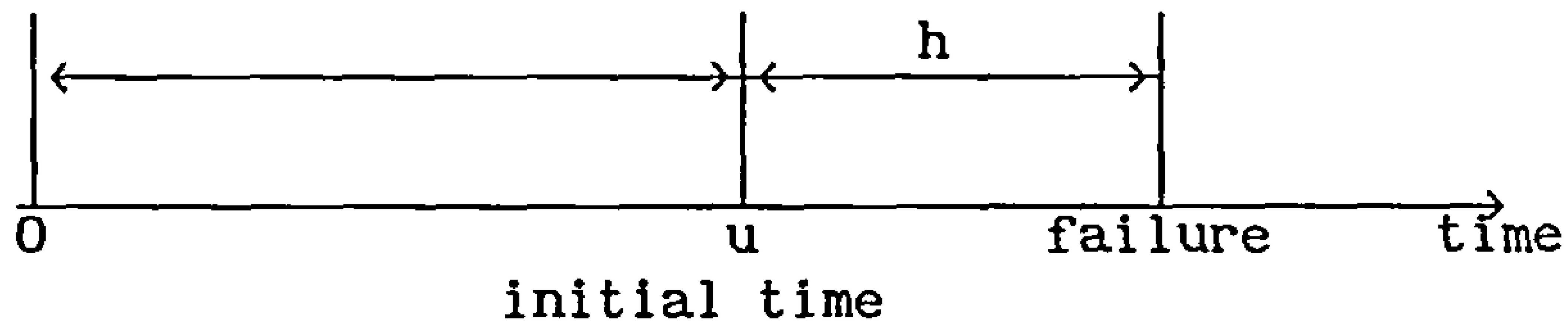


Figure 3-1 The delay time concept

In common with the notion of an initial time point of a defect, the idea of instant of failure is likewise simple enough to cope with mathematically, but needs careful consideration in practice since possible difficulties with the definition of a failure may arise. What actually constitutes a defect which necessitates immediate repair or replacement, that is a failure, can vary with both time and circumstances. So far, in all applied studies using the delay time concept, the moment of failure has been based upon the definition of operational practice, and no serious problems of definition have arisen. The potential robustness and value of delay time analysis lies in its very fundamental engineering-type view of the phenomena being studied. One of the important contributions of the delay time concept and analysis to maintenance theory and practice is that it reveals the nature and mechanism of underpinning preventive maintenance or inspection.

If the distributions of u and h are known (for convenience, we may simply call them delay time distributions in the following text), the failure behaviour of a system can in theory be determined under any specified maintenance policy. Also, if the consequences of defects before and after failure are known in terms of whatever variables are thought important, i.e. cost, downtime, output, risk, reliability, etc., then one expects that such consequent variables may be modelled as functions of the maintenance policy.

3.2 Delay time modelling

To introduce delay time modelling, suppose first that we are concerned with a particular component of plant which is associated with failures characterized by a delay time h with probability density function f(h) and an initial time u with probability distribution density function g(u). To simplify matters for the moment, think in terms of only one failure mode of the component in the plant which generates otherwise independent defects which are from the same population of delay times. Of course, this restriction can readily be lifted. Let a plant inspection be undertaken on a regular basis, with period t, which can be also relaxed later, and suppose for now that the inspection is perfect in that, if a defect is present at the time of an inspection, it will always be identified. Between inspections, say, (t_{i-1}, t_i), where i is the sequence of inspection (i=1,2,...) and $t_i = i \cdot t$, a defect can arise at a time u from new, say, and subsequently lead to a failure after time h if $h < t_i - u$, Figure 3-2 (a), and be identified at an inspection if $h \geq t_i - u$, see Figure 3-2 (b). We assume here that a defect identified at an inspection is repaired or replaced at that time and the process is resumed. First we model the probability that a defect results in a breakdown.

(a) Defect leads to a failure

(b) Defect is identified at an inspection

Figure 3-2 Failure and inspection mechanism using the delay time concept

Let the last perfect inspection be at time t_{i-1} and let the next inspection be at time t_i. Over the period (t_{i-1}, t_i), the probability of a defect arriving in the interval (u,u+du) is g(u)du, where $(t_{i-1} < u < t_i)$. A defect arising in (u,u+du) with a delay time $h < t_i - u$ will arise as a breakdown. Clearly, $P(h < t_i - u) = F(t_i - u)$, where $F(\cdot)$ denotes the cumulative distribution function of the delay time h. If we further assume that both u and h are mutually independent, we have, therefore, that the probability of a defect arising over period (u,u+du) and resulting a breakdown at u+h $(h < t_i - u)$ is $g(u)F(t_i - u)du$. Since u could be at any time over (t_{i-1}, t_i), integrating over (t_{i-1}, t_i), we find that probability $p_b(t_i)$ of a breakdown over (t_{i-1}, t_i) is

$$p_b(t_i) = \int_{t_{i-1}}^{t_i} g(u)F(t_i - u)du. \qquad 3\text{-}1$$

A situation which has been found to have practical significance occurs when the initial time is uniformly distributed with pdf. 1/T, where $T = n \cdot t$ and n is an integer. The probability of a breakdown given the inspection policy t, t ={t, 2t, 3t,...}, denoted by $p_b(t)$, is given here by

$$p_b(t) = \sum_{i=1}^{n} p_b(t_i) = \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \frac{1}{T} F(t_i - u)\, du.$$

Since $n = T/t$, and $\int_{t_{i-1}}^{t_i} F(t_i - u)du = \int_0^t F(t-u)du$,

We have

$$p_b(t) = \int_0^t \frac{F(t-u)}{t} du. \qquad 3\text{-}2$$

The function $p_b(t)$, or $p_b(t_i)$ is fundamental in delay time modelling. For example, suppose the following assumptions are valid for an inspection process, and that costs and downtime measures associated with inspection and plant failure and defect repairs are:

1. An inspection takes place every t time units, and requires d_i time units, with $d_i << t$.
2. Inspections are perfect.
3. Defects identified within an inspection will be repaired within the inspection period at an average cost of c_m.
4. Failures are repaired as soon as they arise at an average cost of c_b and downtime d_b, where $c_b > c_m$ and $d_b << t$.
5. Defects arise within the plant at a known instantaneous rate of $\psi(u)$ at time u after the last inspection, ie. the number of defects arising in the period $(u,u+du)=\psi(u)du$.
6. the cumulative distribution function $F(\cdot)$ of the delay time h is known.

Under these assumptions, a model of the expected downtime per unit time as a function of the inspection interval t may be obtained directly. The total expected downtime of an inspection cycle consists of the expected downtime associated with failures and the downtime due to an inspection. Since we have assumed that there is no additional expected downtime due to repairing defects identified at an inspection, the total expected downtime per unit time is

$$D(t) = \frac{B(t)d_b + d_i}{t + d_i}, \qquad 3\text{-}3$$

where $B(t)=\int_0^t F(t-u)\psi(u)du$, is the expected number of breakdowns arising over period $(0,t)$.

In a similar way we can also establish the costs measure as a function of the inspection period t.

The function $p_b(t)$ can readily be calculated for use in a criterion function such as $D(\cdot)$, and the first industrial use of delay time analysis entailed using the model 3-2 for $p_b(t)$ with the downtime function 3-3 for $D(\cdot)$, to model the downtime of a high-speed canning line, Christer & Waller [1984a]

So far, it has been assumed that inspections are perfect in that any defect present will be identified. This assumption will undoubtedly be simplistic in some applications and may be relaxed if necessary. Assuming still that the initial point u is uniformly distributed along (0,t), suppose that there is a probability $\beta \le 1$ that any defect present at an inspection will be identified at the inspection. It has been shown, Christer & Waller [1984b], that under these circumstance, the probability of a defect leading to a failure becomes

$$p_b(t) = 1 - \int_0^t \sum_{n=1}^{\infty} \frac{\beta}{t} (1-\beta)^{n-1}\{1-F(nt-u)\}du. \qquad 3\text{-}4$$

Interestingly, the only modelling changes in permitting imperfect inspection ($\beta \ne 1$) is that $p_b(\cdot)$ changes in form, but the criterion function such as $D(\cdot)$, given in equation 3-3, remains the same. A variation on this imperfect-inspection formulation with $\beta \ne 1$ was required for an application of delay time analysis modelling of the planned maintenance for a vehicle fleet, Christer & Waller [1984c].

Obviously, the distributions of u and h are vital to delay time modelling. In most previous applications of delay time modelling, the period of u is assumed to be uniformly distributed and f(h) is estimated through subjective data which are obtained by the survey of engineers who are responsible for maintaining the system to be modelled. Since no explicit data on u and h can be obtained in practice, subjective assessments of the distributions of u and h have been proved to work in applications of inspection modelling. However, if there are past records of inspections and breakdowns on the system or component of interest, especially if records show whether or not an inspection finds a defect in the system or component, then other estimating procedures become possible. The use of objective data in estimating distributions of u and h will be fully discussed and explained in the next Chapter.

CHAPTER 4. BASIC METHODS OF STATISTICAL ESTIMATION OF DELAY TIME DISTRIBUTIONS

The delay time concept developed by Christer *et al* [1973-1992], as mentioned in previous Chapters, has provided a useful means of modelling the effect of periodic inspections on the failure and operating consequences of repairable machinery. Obviously, the distribution functions of u and h and their parameterizations are vital to delay time modelling. In this chapter we explore basic ideas and methods of estimation of delay time distributions.

4.1 Current methods of estimation of delay time distributions

From the definitions of delay and initial time distributions, it is clear that it is most unlikely to be possible to measure directly either the delay time associated with a defect, or the initial time u. This is so even when there are past records of inspections and failures because they will at best only show the times of failures or inspections. However, what has proved possible, as established by Christer *et al* [1984, 1991], is to estimate the delay time for a set of specific faults and failures from subjective estimates obtained from the repairing engineers. Based upon this data, it is possible to deduce the location of the initial time, and to estimate both the delay time and the initial time distributions. This method has been successfully applied to several applied maintenance studies by Christer *et al* [1984a,b,c] and may be applied if there are no historical records of inspections and failures.

Essentially, the method works as follows. At any repair intervention, be it due to a breakdown or a fault identified at an inspection, two questions could be asked of the repairing engineers:

(i) How long ago (HLA) could the fault reasonably have been expected to first have been noticed had there been an inspection?

(ii) If the repair was not carried out, how much longer (HML) could it be postponed before a failure was likely, that is, a repair is essential?

The delay time estimate for the fault is taken as $\hat{h}$ = HLA + HML. Although in practice the questions, and lead-up to them, are a little more elaborate in order to focus and concentrate the mind of engineers supplying estimates, the general principle is as straightforward as indicated. It is to be noted that the assessment is subjective and that two people could not be expected in general to produce the same estimates. Questions are, however, asked under very specific and precise circumstances with the defect in question present. By accumulating sufficiently many estimates $\hat{h}$, an estimate can be made of the delay time distribution.

The definition of failure is important to the assessment of the delay time distribution. There is often confusion between a defect and a failure in traditional maintenance modelling. Here the definition of failure is quite practically oriented towards the organizations being studied. A defect is a failure if the organizations consider immediate repair to be essential, Christer and Waller [1984a, 1984b].

At any point in time T when a defect is being attended to, having an estimate of HLA provides at once an estimate of the initial time u, namely u = T - HLA. It is the set of such estimates that enables the distribution of the initial time u to be estimated. However for most papers of delay time applications, the main point of interest is the estimation of the delay time h, while it is assumed that over its range of interest, u is uniformly distributed.

One of the interesting aspects of previous delay time analysis is that it uses a synthesis of subjectively derived data. If under the current inspection policy of constant inspection interval of period t_0, after accumulating records of failure and inspection repairs, the probability that a defect arising over the period $(0, t_0)$ results in a failure, say, during the first interval $(0, t_0)$, may be estimated as p_{b0}. One may expect from equation 3-1 that

$$p_{b0} = p_b(t_0). \qquad 4\text{-}1$$

However, the chance of conditions such as 4-1 being satisfied is remote. The left-hand side is an objective observation of practice, and the right-hand side is a function based upon an aggregate of subjective estimates. In common with any process of decision analysis entailing subjective assessment, it is to be expected that some revision will need to be made to the prior distribution or perhaps to the prior model. Christer and Redmond [1990a] have addressed this problem and proposed a model for the revision of the delay time distribution in which a shearing transformation $h \Rightarrow z$, such as $z=\alpha h+\omega$, is used to correct the error, where the task is to determine the appropriate value of α and ω. However in their case, what they consider is the probability of failure under the assumptions of an uniform distribution of u and a constant inspection interval of t. They do not take account of the case of non-uniform distribution of u, which may lead to a completely different formula for the probability of failure as we have mentioned in Chapter 3.

In a recent paper, Christer and Redmond [1990b] recognize that an unknown bias is entering into the delay time estimation, since $\hat{h}_1$= HLA will produce an underestimate of h, and $\hat{h}_2$= HLA + HML will produce an overestimate of h for reasons associated with the waiting-time paradox, Feller [1970]. They also propose a method for coping with the bias. Christer and Redmond assume that there is an unknown parameter existing in the prior distribution of the delay time h, which is estimated by the subjective data, say, γ. Accepting this distribution, for purpose of correcting this bias, a likelihood function based upon observations of failures over time t has been built up to estimate this parameter so that bias can be corrected through this parameter. However there is no numerical example presented in their paper to show how to use this method.

It is to be expected that engineers making subjective judgment will take into account any objective data available. Besides the subjective data based estimating procedure for the delay time distributions, if sufficient historical records of failures and inspections for a particular component are available, estimation of the delay time distributions is, however, possible from such objective data by statistical inference. We now develop this later method in detail.

4.2 Assumptions, notation and likelihood calculation

It is at first glance at least plausible that both the distributions of u and h can be determined from such data as the dates and results of pre-planned inspections, and times of failures. If the delay time is typically very small, there will be very few inspections in which a defect is detected, since failure follows very soon after a defect becomes visible. Conversely, the fact of many successful detections of defects at inspections implies a long delay time. If a component is inspected upon failure the period between adjacent failures is clearly an estimate of the sum of u and h if there are no inspection replacements during this period, and so the addition of breakdown times offers the possibility of determining the distribution of u. More clearly, if under the assumption of perfect inspections, there is a failure occurring between inspection interval, say, (t_{i-1}, t_i), then we can deduce both that u must be less than t_i and greater than t_{i-1}, and h must be less than t_i - u, Figure 4-1(a). However, if at an inspection time t_i, a defect is identified, Figure 4-1(b), this implies that $t_{i-1} < u \le t_i$ and $h \ge t_i - u$.

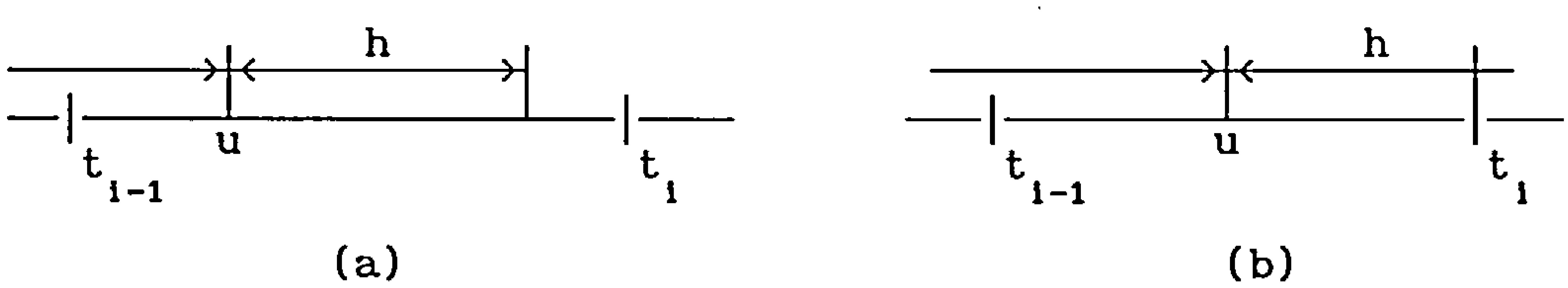


Figure 4-1 (a) Defect results in a failure
(b) Defect is identified at an inspection

With such information provided from failures and inspections records, we use state-of-the-art statistical techniques to recover the distributions of u and h as best we can. The interesting questions are perhaps: 1. 'Can this be done?'; 2. 'if so, are the errors in the estimates of g(u) and f(h) acceptable for practical purpose?'; and 3, 'Are subjective data still needed, and if so, how should they be combined with objective data?'. We focus on the first two questions in the thesis. To answer question 1, we must show both that the techniques work with Monte-Carlo data and with 'typical' real life data. The Monte-Carlo study alone can tell us that we are able to estimate known distributions of u and h to acceptable accuracy. A real life study is also essential to ensure that our method would work in practice. To answer question 2, a calculation of an optimal inspection schedule based upon the estimated delay time distributions will be carried out as an exploratory test. This will be discussed in Chapters 6 and 8.

4.2.1 Assumptions

The data comprise essentially of a history of breakdown times, and the results of inspections, which may be positive (defect found) or negative (no defect found). These data were available for key components of a sample of about 100 infusion pumps in Hope Hospital at Salford, which were originally collected for the use and guidance of technicians dealing with the maintenance of these machines occasions.

The assumptions of the basic methods of estimation of g(u) and f(h) are as follows:

1. The random variables u and h are assumed independent.
2. g(u) and f(h) are modelled as exponential or Weibull distributions.
3. The components of a machine are assumed independent, i.e. the failure of one will not affect the functioning of another. Further, each component is assumed to have only one failure mode.
4. Inspections are in general imperfect, i.e. they have a probability $\beta \leq 1$ of detecting a fault if it is present. When any component

fails, an inspection of the machine embracing all of its components is carried out, and these adventitious inspections have probability $\beta' \leq 1$ of detecting a fault if it exists. In general $\beta' \neq \beta$. There are no false positives, i.e. if a fault is not present one will not be identified. Probabilities β are assumed independent and constant between inspections.

5. Machines are assumed to behave identically and to have uniform usage.
6. Repair times are assumed to be negligible.
7. Repairs are taken as replacements, so that the faulty component is restored to an 'as-new' condition. The inspection is, therefore, effects a renewal point if fault is found and rectified.

Definitions of breakdown and the appearance of a visible defect are taken as operational, which absolves the modeller from the need to worry excessively about what these distributions actually measure. A breakdown is whatever engineers and users of the machine deem it to be. In this sense we follows the lead of Christer [1984a,b] and are constructing the user's model.

Consider first the simple case of a one-component machine where inspection is a perfect inspection process with replacement of defective components, and then progress to more general cases.

4.2.2 Notation and likelihood calculation

First we introduce the necessary terminology. The possible events that can contribute to a likelihood function are

- b a breakdown (failure).
- n an inspection where no defect is found.
- y an inspection where a defect is found.
- e the end of observation period, ie. censored by the data.

Event n will be referred to as a negative inspection, and conversely event y referred to as a positive inspection. In addition, the following are useful:

s the start of an observation period, ie. the renewal point of a component .

r a component replacement (following either b or y).

x to denote any event.

Event s is equivalent to that of an r event. We wish to write down the likelihood of observing a sequence of events $x_1, \ldots, x_n$ of types b, e, y and n at times $t_1, \ldots, t_n$. The key to doing this is the multiplication law of likelihood, i.e.

$$L = Px_1 \cdot Px_2|x_1 \cdot Px_3|x_1,x_2 \cdot, \ldots, \cdot Px_n|x_1,\ldots,x_{n-1}, \qquad 4\text{-}1$$

for the likelihood of n events, where x_i denotes event x at time t_i from the last renewal. The probability of an event is P, and $Px_2|x_1$ means the probability of event x_2 at t_2 given that event x_1 at t_1 has occurred.

After a replacement r, the likelihood does not depend on any event previous to r. Therefore the likelihood can be written as the product of terms conditional on subsequent events $rx_1x_2\ldots$ starting with the last renewal. Further, for an event x at time t following a sequence $n_1, \ldots, n_n$ of negative inspections at times $t_1, \ldots, t_n$ from the last renewal, see Figure 4-2, if we let $Px|rn_1n_2, \ldots, n_n$ denote the probability of x conditional on events $n_1, \ldots, n_n$, the fact that there was no defect visible on the last inspection t_n of the sequence is what determines the probability of the event x, i.e. $Px|rn_1n_2, \ldots, n_n = Px|rn_n$.

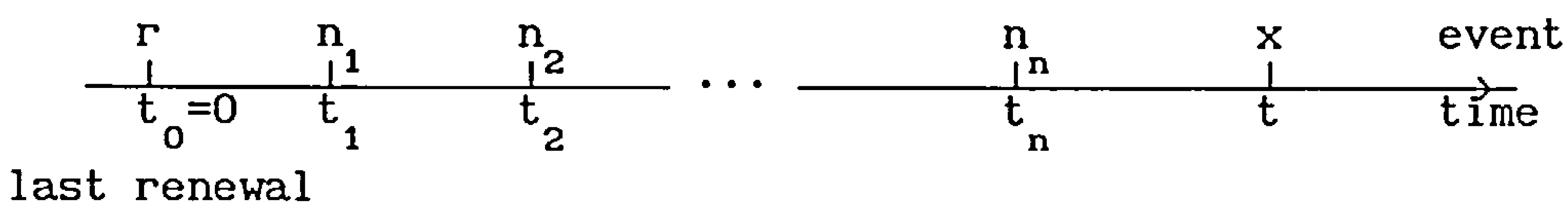


Figure 4-2 Inspection and renewal process

A replacement or renewal cycle could be defined here as the time from the last renewal (or new condition) to when the component is either replaced upon a failure or replaced after a defect has been identified at an inspection. In this sense, the cycle will consist of a series of negative inspections (possible null) followed by an event of type of b or y at the cycle end. We require to determine the likelihood of such events.

The contribution to the likelihood function from one replacement cycle with n negative inspections and a breakdown or positive inspection at time t is the product of the probability of each event conditional on events previous to it from the last renewal. Then from equation 4-1, we have

$$L_{cycle}=Pn_1 Pn_2|rn_1 Pn_3|rn_1,n_2 \ldots Pn_n|rn_1,n_2,\ldots,n_{n-1} Px|rn_1,n_2,\ldots,n_n,$$

where x=b or y and L_{cycle} denotes the contribution to the likelihood function from one replacement cycle.

Since we know that $Px|rn_1,n_2,\ldots,n_n=Px|rn_n$ and $Px|rn_n= P(x,n_n)/P(n_n)=Px/Pn_n$, where Px is the unconditional probability of event x at the cycle end and Pn_n is the probability of the last negative inspection before x, we have,

$$L_{cycle}=Pn_1\frac{Pn_2}{Pn_1}\cdot\frac{Pn_3}{Pn_2}\cdot\ldots\cdot\frac{Pn_n}{Pn_{n-1}}\cdot\frac{Px}{Pn_n}=Px, \qquad x=b \text{ or } y$$

The same argument can also be extended to the case of censored data, that is x=e. Then it turns out that only three unconditional key probabilities need be considered. The likelihood can be built up from these three. This greatly simplifies the process. The key probabilities are now established below.

Let $P_b(t_n,t)$ denote the pdf of a sequence of negative inspections, of which the last occurs at time t_n from last renewal, and a breakdown at time t from last renewal. The sequence of negative inspections may be null, in which case $t_n=0$.

This use of notation reflects the fact that an inspection made at the instant of renewal must be negative with probability unity. Hence one can always 'smuggle in' as such a notional inspection without altering the likelihood, and hence from the definition, we have

$$P_b(t_n,t) = \int_{t_n}^{t} g(u)f(t-u)\, du. \qquad 4\text{-}2$$

Since g(u) is the pdf that a defect arises at time u, and f(h) is the pdf that a breakdown occurs a time h later, g(u)f(t-u) is the pdf of a failure at t arising from a defect at u, and the integration sums over all possible times u. These can only occur after the last moment that there was known to be no defect, t_n, and before the breakdown time t.

Let $P_e(t_n,t)$ denote probability of a (possibly null) sequence of negative inspections of which the last is at t_n, and no breakdown before observation ceases at time t from last renewal. We have in a similar way

$$P_e(t_n,t) = 1 - G(t_n) - \int_{t_n}^{t} g(u)F(t-u)\, du, \qquad 4\text{-}3$$

where $G(\cdot)$ and $F(\cdot)$ are the cumulative distribution functions of u and h respectively.

This expression is simpler to interpret in its alternative form

$$P_e(t_n,t) = 1 - G(t) + \int_{t_n}^{t} g(u)(1-F(t-u))\ du.$$

The first part, 1-G(t), is the probability that no defect arises before time t, and the second contribution to the probability of no failure represents the event that a defect does arise at some time $u > t_n$, but does not lead to a failure before time t. The product g(u)(1-F(t-u)) is the pdf that a defect arises at u and that there is no failure before time t. The integration sums over all possible times u, after the last negative inspection at t_n and before time t.

Let $P_y(t_n,t)$ denote the probability of a sequence of negative inspections of which the last occurs at t_n, followed by a positive inspection at time t from last renewal. In the same way, we have

$$P_y(t_n,t) = \int_{t_n}^{t} g(u)(1 - F(t-u))\ du. \qquad 4\text{-}4$$

To see this result, it is simplest to be understood that the pdf for a fault arising at time u is g(u), and the probability of no breakdown before t is (1-F(t-u)). The integration sums over all possible times of fault origin u.

The three key probabilities of the proposed method are conditional on the last renewal. With Weibull distributions for u and h, the probabilities are calculated by substituting:

$$G(u) = 1-\exp\{-(\alpha_1 u)^{\beta_1}).$$

$$g(u) = \beta_1\alpha_1^{\beta_1}\ u^{\beta_1-1}\exp\{-(\alpha_1 u)^{\beta_1}).$$

$$F(h) = 1-\exp\{-(\alpha_2 u)^{\beta_2}).$$

$$f(h) = \beta_2\alpha_2^{\beta_2}\ u^{\beta_2-1}\exp\{-(\alpha_2 u)^{\beta_2}).$$

where α_1,α_2 are scale parameters and β_1,β_2 are shape parameters. The likelihood is calculated by accumulating the product of these three terms. Each renewal may be followed by a sequence of negative inspections, and this must terminate in an event of type b, e or y. Event e is really 'no event'. The likelihood L for a total of n_b breakdowns at times t_i, n_e 'no failure before observation ceases' events at times t_j, and n_y positive inspections at times t_k, is

$$L = \prod_{i=1}^{n_b} P_b(t_i^*, t_i) \prod_{j=1}^{n_e} P_e(t_j^*, t_j) \prod_{k=1}^{n_y} P_y(t_k^*, t_k), \qquad 4\text{-}5$$

where the notation t_i^*, t_j^*, t_k^* denotes the time of the latest negative inspection, or failing that the latest renewal such that $t_i^* < t_i$, and so on.

In the more general case of several identical machines, the likelihoods corresponding to individual machines are multiplied together.

4.2.3 Multi-component case

To see how the argument may be extended to systems of components, we now focus on the case of a machine comprising two components. They are assumed to be mutually independent in that the state of either component is assumed not to affect that of the other. There are two possible scenarios: when component A fails, component B is either not inspected (case 1) or inspected and replaced if visibly defective (case 2). Happily both case are tractable.

In case 1, the two components are completely independent---nothing that happens to either of them can affect the other, and the likelihood factorizes. The log-likelihood is the sum of log-likelihoods for each component, $\log(L) = \log(L_A) + \log(L_B)$. In case 2, they are no longer independent, because a failure of A will cause the replacement of B, if B iS visibly defective, and vice versa. Happily, the likelihood can still be written in factored form, even although the components are not now independent. A failure of either component (A, say) simply generates an inspection event (n or y) for the other, at the failure time t_A. These extra inspection events mean that the log-likelihood for A contains extra contributions at times determined by the behaviour of B, and vice versa. The computer analysis is simple, as the program merely has to insert these extra inspections into the record before further analysis, and then proceed with the calculation for each component separately, as long as the components have no parameters in common.

The argument generalizes immediately to arbitrarily many components.

4.3 Imperfect inspection

So far it has been assumed that inspections always find a visible defect if it is there. In the case of imperfect inspections, there is a probability $\beta \le 1$ that a defect is found if it exists. This is equivalent to saying that a (perfect) inspection is carried out with probability β, and that with probability $1-\beta$ the inspection is 'omitted'.

When a defect is found, the likelihood is simply multiplied by β, the probability of the observed event, as there is a probability β of carrying out a perfect inspection. However, when a defect is not found, the state of the system is not known. It could be either a defect existing there but has not being found, or no defect exists. Since it is assumed that inspections are imperfect, the meaning of negative inspections differs from the one we used in perfect inspection case. Negative inspections under the assumption of imperfect inspection merely mean that no defects is found at an inspection, but does not mean that no defect exists. Therefore, a negative inspection under the assumption of imperfect inspection would imply two possibilities, either a 'real' negative inspection with probability β, or a 'false' negative inspection (equivalent to an omitted inspection) with probability $1-\beta$.

Here the conditional probabilities mentioned for b, y and e events are used, and event x denotes any one of these. Note that the probability of event x is conditional on the appropriate preceding event---r or n. After m negative inspections, there are 2^m terms. For example if let $Pn_i|r(t_j,t_i)$ denote the conditional probability of a real negative inspection at time t_i from renewal given that the last real negative inspection or renewal is at time t_j. And let $Px|r(t_j,t)$denote the conditional probability of event x at time t from renewal given that the last real negative inspection or renewal is at time t_j. For example, consider the formulation of likelihood of two negative inspections followed by event x, or more precisely, rn_1n_2x. In this case, we have four possible combinations of the following joint events, namely, (a) two real negative inspections at t_1 and t_2; (b) one false negative inspection at t_1 and one real negative inspection at t_2; (c) one real negative inspection at t_1 and one false negative inspection at t_2; and finally (d) two false negative inspections at t_1 and t_2. The likelihood of event x at time t after two negative inspections at t_1 and t_2 is the sum of four possible terms, that is

$$L = \beta Pn_1|r(0,t_1)\beta Pn_2|r(t_1,t_2)Px|r(t_2,t) + (1-\beta)\beta Pn_2|r(0,t_2)Px|r(t_2,t)$$

$$+ \beta Pn_1|r(0,t_1)(1-\beta)Px|r(t_1,t) + (1-\beta)^2 Px|r(0,t).$$

Fortunately this simplifies considerably, as, given that the nth negative inspection occurs, the occurrence or otherwise of previous events does not change the likelihood. Thus

$$Pn_1|r(0,t_1)Pn_2|r(t_1,t_2)Px|r(t_2,t) = Px(t_2,t),$$

$$Pn_2|r(0,t_2)Px|r(t_2,t) = Px(t_2,t),$$

$$Pn_1|r(0,t_1)Px|r(t_1,t) = Px(t_1,t).$$

Hence,

$$L = \beta\, Px(t_2,t) + \beta(1-\beta)Px(t_1,t) + (1-\beta)^2Px(0,t).$$

For the event $rn_1n_2,\ldots,n_mx$ (m consecutive negative inspections) this suggests a simpler way of deriving such likelihoods as a sum of m+1 terms, each having the last real negative inspection carried out at t_j, where $t_0=0$. The probability of the jth real negative inspection is $\beta Pn_j{\rm ir}(0,t_j)$, and as the succeeding m-j inspections must not occur by definition, they contribute a factor of $(1-\beta)^{m-j}$. The likelihood is then

$$L = \beta Px(t_m,t)+\beta(1-\beta)Px(t_{m-1},t)+\ldots+\beta(1-\beta)^{m-1}Px(t_1,t)+(1-\beta)^mPx(0,t)$$

or

$$L = \beta\sum_{j=1}^{m} (1-\beta)^{m-j}Px(t_j,t) + (1-\beta)^mPx(0,t). \qquad 4\text{-}6$$

For details of this formulation, see Appendix C.

Note that the coefficients (weights) in equations 4-6 sum to unity. If the event x is y, a positive inspection, then L is multiplied by an extra factor of β, as the probability of observing a positive outcome is βPx.

4.4 Selection of the fitted distributions ---the Akaike information criterion (AIC)

Distribution functions with ever more parameters may be applied, and tests of fit carried out to assess adequacy, while the increase in likelihood per parameter added can be used to test whether that parameter was needed. Twice the increase in log-likelihood is asymptotically distributed as $\chi^2(f)$, where f is the number of new parameters added, Cox & Hinkley [1974]. However, Akaike has challenged this traditional approach to parameters estimation, Akaike [1984], on the grounds that it is not appropriate to set up a series of null hypotheses H_0 that each fresh parameter should not be needed. Why should we adopt such a conservative H_0, clinging to the assumption that no parameter is needed until compelling to include it, when we have no a prior idea of what the distribution should look like?

The Akaike Information Criterion (AIC) is $-2\log(L)+2f$, where f is the number of parameters estimated from the data. The 'best' distribution function minimizes this AIC. AIC is actually designed for the purpose of correcting the bias of the maximum likelihood. We use the AIC as a criterion of the 'best' distribution function, which then enables us to choose which of several possible parameters should be nonzero. In the simple distribution functions used in the analysis of the infusion pump data, the AIC was used to choose among four possible combinations of distribution functions, as both f(h) and g(u) could have pdf's of either exponential or Weibull type. It is thought by statisticians that the use of the AIC results in a slight 'overfitting' of data (too many parameters are fitted), and other selection criteria are also used: we chose the AIC since it is the simplest.

4.5 Infusion pump data

The medical physics department of Hope hospital in Salford, which maintains a large amount of medical equipment, records the history of breakdowns and repairs carried out via 'history cards' for each individual item of departmental equipment.

Information available to us included purchase date, dates of ppms and failures, and some description of the work carried out. There were no costs recorded, and the record was purely designed to guide technicians dealing with the machine on future occasions. Therefore much was implicit, but it was easy to recognize ppms, and other entries were usually either an initial acceptance test or a failure repair. The repairs done were described, and there might be no repair necessary or recorded if a reported fault turned out to be a false alarm. Sometimes the record was complex, as a fault was noted on ppm by one technician, the machine went to the workshop for repairs, and on further testing, perhaps by a third technician, further faults were discovered, and so on, resulting in several entries perhaps spanning a number of days. It was usually clear that all these entries pertained to the same ppm or failure. These infusion pumps were under warranty for their first year, but some details of repairs carried out by the manufacturer were still recorded.

Following discussions with the chief technician, it seemed best for our exploratory study to focus on the following data, to ensure samples of similar machine types, under heavy and constant use, with usefully long histories of failure, and with reasonably well-defined modes of failure. The items of interest were infusion pumps.

1. There were 105 volumetric infusion pumps. The most frequent failure mode was the failure of the pressure transducer (TX).

2. There were 35 peristaltic pumps in all, from the Intensive care, Neurosurgery and Heart care units. The most frequent failure modes were batteries and door-pads.

Table 4-1 shows the frequencies of the b, e, y and n events.

component	breakdowns	+ inspections	- inspections	no-event
door-pad	4	49	231	34
battery	36	18	230	34
transducer	80	20	323	155

Table 4-1. Number of breakdowns, positive and negative inspections and end-of-observation 'event' for the components studied. Unfailed pressure transducers were replaced with a later model, given rise to the large number of 'no events' for this component.

A number of minor problems arose when marrying theory to the real-world data. Here we list some of the problems and their solutions:

1. If a failure occurred just before a scheduled ppm, the ppm was carried out as part of a failure repair rather than at the scheduled time. The likelihood function does not need altering to cope with this, as it is conditional on the observed ppm timings. However, simulations of failures and inspections, and predictions of reliability and optimum ppm schedules, would need to allow for this effect in order to be completely realistic.

2. False alarms pose a similar problem, not of determination of model parameters, but of prediction. Brief inspections are carried out when faults are reported, even if the faults were non-existent. To make predictions of reliability, etc., one would need to model the frequency of false alarms, probably as a homogeneous Poisson process.

3. Acceptance tests and repairs need to be allowed for in our model: the machine should not need repairs when it is brand new. Acceptance tests and repairs are ignored, and we assume that the machine is in a perfect state and at time t_0 after the test.

4. For two-component systems, the case arises where one component fails, and in the resulting inspection accompanying replacement of the failed component, the other is also found to be faulty and is replaced. This is a b event for one component, and a y event for the other. However, it sometimes happens that in such cases where a failure has occurred, and both components have visible defects, it may not be known which component has actually failed. This ambiguity was not completely resolvable from the records in a few cases.

One of the strengths of the likelihood approach is that it can cope with such losses of information---it is only necessary to sum over all cases giving rise to the observed events. However, we had a few cases where this effect was a problem, and so we simply judged as best we could in conjunction with the chief technician.

5. All of a batch of volumetric pumps had pressure transducers replaced with a later version by the manufacturer, which could be an example of reliability growth. However, this has no influence on our likelihood function and could be modelled
   (a) Inserting an e event into the likelihood, to give the probability that no failures had occurred between the last renewal and the time of replacement.
   (b) Inserting a renewal r event after the e event.

   In a more elaborate model, it would be possible to use a multiplier parameter for the scale factor of the distributions of u and h for the new component to allow for its changed reliability.

6. Some repairs were not replacements, e.g. recharging of batteries. In this preliminary study, we classified these as negative inspections.

The log-likelihood corresponding to the likelihood in equation 4-5 can be written down as the sum of log-likelihoods, one of the three terms just derived for each event of type b, e or y, for each machine in the sample;

$$\log(L) = \sum_{i=1}^{n_b} \log P_b(t_i^*, t_i) + \sum_{j=1}^{n_e} \log P_e(t_j^*, t_j) + \sum_{k=1}^{n_y} \log P_y(t_k^*, t_k). \qquad 4\text{-}7$$

A computer program was written to read in the series of event types and machine ages (actually the date of the start of the observation period s and further dates convertible to machine ages) at each event. These data were stored as a list of elements of the following form:

1. Type of event (b, e, or y).

2. Machine age at last renewal.
3. Period t since last replacement or machine age if not replaced.
4. Time t_n of the last negative inspection after last renewal, or time from last renewal itself if there had been no inspection since then.

This is simple to program: one reads in machine histories, updating an array of the above information, updating t and t_n by the interval from the last event, resetting t and t_n to zero on a renewal or the start of a new machine record, and resetting t_n to the current time from last renewal on a negative inspection.

It is then possible to step through each event in the array, calculating and incrementing the log-likelihood, log(L), by the appropriate term for each event. In the case of imperfect inspections, we have seen that the times $t_1,\ldots,t_n$ of all negative inspections preceding an event x are needed. The data structure needed is just a little more complex, and a list of the elements is given below:

1. Type of event (b, e, or y).
2. Machine age at last renewal.
3. Period t since last replacement or machine age if not replaced.
4. Number of previous negative inspections since last renewal.
5. Pointer to machine age at the last n event (negative inspection) since most recent renewal, in an array of all n event times, or zero if there was no negative inspection after renewal.

FORTRAN was chosen as a suitable language, mainly because of the excellent NAG, Hopkins and Philips [1988], library of numerical routines available for the 386-PC. The NAG function minimizer E04JBF was used to minimize minus the log-likelihood. (Take heed that this minimizer, we now learn, is soon to be withdrawn). Because scale and shape parameters must by definition take positive values, only the logarithms of these parameters were seen by the minimizer. The sub-routine which evaluated the likelihood function should exponentiate them before use in order to make these parameters be the original values. This procedure meant that no bounds needed to be set on the parameters, and also ensured faster convergence owing to a more natural step size. Similarly, when varying the probability β of detecting an existing defect on ppm, the logit of β

$$\text{logit}(\beta) \equiv \log \frac{\beta}{1-\beta}$$

was used.

Integrations were performed using the Gauss-Legendre method, with n=20 abscissae. The optimum weights w_i and abscissae x_i such that

$$\int_A^B f(x)\,dx \cong \sum_{i=1}^{n} w_i f(x_i)$$

were found e.g. with A=-1, B=1 and then used to evaluate integrals between different sets of limits, e.g. a, b, using the result

$$\int_a^b f(x)\,dx \cong \frac{a-b}{A-B} \sum_{i=1}^{n} w_i f(y_i),$$

where $y_i = \frac{a-b}{A-B} x_i + \frac{b \cdot A - a \cdot B}{A - B}$.

When various distribution functions were fitted using the Akaike information criterion AIC, we chose as starting values those parameter values from the best previously fitted distribution function that was a subset of the current distribution function. Thus the order of fit for g(u) and f(u) was: Exponential/Exponential, Exponential/Weibull (scale parameters from last fit, shape set to 1), Weibull/Exponential (scale parameters from fit 1, shape parameter set to 1), Weibull/Weibull (starting values from fit 2 or 3, choosing that fit with smaller AIC, remaining shape parameter set to 1). The calculation of the covariance matrix on fitted parameters poses an important practical problem. Errors on fitted parameters must be propagated through any subsequent modelling to give error bars on quantities of ultimate interest, such as recommended intervals between ppms. Now in the next section, a reasonably accurate estimate of the covariance matrix is proposed and discussed.

4.6 Calculating the covariance matrix on fitted parameters

Denoting minus the log-likelihood as F, the covariance matrix C is commonly estimated as the inverse of the Hessian matrix H, where

$$H_{ij} = \partial^2 F/\partial v_i \partial v_j,$$

and v_i, v_j are the ith and jth parameters. Asymptotic arguments now indicate that this is a better estimate of covariance matrix C than the expectation $E\{H^{-1}\}$, Efron and Hinkley [1978]. One can either invest effort in calculating the Hessian analytically, or, as we chose, in estimating it from the likelihood by numerical methods.

Because of the non-quadratic nature of the valley bottom around the minimum value of F, great care had to be exercised in calculating covariance matrix C. If v_i^0 denotes the value of v_i that minimizes F, a central difference approximation to H_{ij} is

$$H_{ij} = \begin{cases} \left[F(v_i^0+\delta_i, v_i^0+\delta_j) - F(v_i^0+\delta_j, v_j^0-\delta_j) + F(v_i^0-\delta_i, v_j^0-\delta_j) - F(v_i^0-\delta_i, v_j^0+\delta_j)\right]/4\delta_i\delta_j & \text{, if } i \neq j \\ \left[F(v_i^0+\delta_i) + F(v_j^0-\delta_i) - 2F(v_i^0)\right]/\delta_i^2, & \text{if } i = j. \end{cases}$$

H was calculated using this formula with the δ_i set to .01---.1 of the estimated standard deviation of the v_i, and diagonalized. As H is symmetric, only one triangle was actually computed. This initial crude estimate of H was not always positive definite, i.e. one or more eigenvalues λ_i were negative. The matrix S_1 that diagonalizes H is such that $S_1^T \cdot H \cdot S_1 = A$, where A is the diagonal matrix of eigenvalues of H. The columns of S_1 are the eigenvectors of H, and a further set of central differences were taken along directions defined by the eigenvectors as unit vectors. This second set of differences was a much better approximation to H. The matrix of these differences was again diagonalized, by a matrix S_2, and the eigenvalues λ_i replaced with the difference approximation to the second derivatives of F along the new eigenvectors. Finally, H^{-1} was found as

$$H^{-1}_{ij} = \sum_{k=1}^{n} (S_1 S_2)_{i,k} (S_1 S_2)_{j,k} / \lambda_i .$$

Any surviving negative eigenvalues were given an inverse of zero.

The logic behind this procedure is that it is much easier to estimate second differentials by difference approximations along the principal axes of H. However, these principal axes are themselves found using a difference approximation for H, and so two steps are needed to converge on the true axes. Finally, $C = H^{-1}$ may be obtained as $S \cdot A^{-1} \cdot S^T$.

4.7 Results of fitted delay time distributions

4.7.1 The estimated delay time distributions: test calculation

Table 4-2 shows fitted parameter values and estimated coefficients of variation for 9 test simulations, in which 4 combinations of distribution functions were fitted by taking both f(h) and g(u) to be either Weibull or exponential. In all cases, the minimum AIC criterion successfully selected the correct distribution functions, which was a Weibull distribution for u and an exponential distribution for h. It can be seen from Table 2 that the ML estimates are unbiased within the standard error of the mean of the simulated results, and if there is a bias, it is much less than the random error on the estimates. In general one expects a bias of at worst O(1/n), where n is the number of events, and this would appear to be negligible for sample sizes large enough to make the analysis worth while.

simulation	α_1	β_1	α_2	CV α_1	CV β_1	CV α_2
1	.298	2.041	.852	.073	.122	.234
2	.293	1.570	.631	.092	.118	.235
3	.258	1.755	.816	.085	.124	.238
4	.246	2.036	.755	.076	.120	.246
5	.266	1.499	.473	.100	.122	.264
6	.265	1.710	.778	.085	.123	.238
7	.253	1.372	.664	.110	.136	.253
8	.298	1.397	.338	.105	.121	.285
9	.272	2.037	.364	.130	.189	.300
mean	.2722	1.680	.6633			
bias	-.0013	.0640	.0380			
SEM	.0074	.0940	.0586			

Table 4-2. Simulation results. α_1 is the scale parameter of the Weibull distribution of fault origin times, set to .2735, β_1 its shape parameter, set to 1.616, α_2 the scale parameter of the delay-time distribution, set to .625. SEM is the error (standard deviation) of the mean Monte-Carlo estimate, and the bias is the difference between the mean of the estimates and the true value. SEM is thus the error bar on the bias. CV is the covariance of parameters. Simulation 9 is an imperfect inspection simulation with β set to 0.7, and is not included in the averaging.

Table 4-3 shows that the standard deviation of the parameter estimates derived from the covariance matrix agrees acceptably well with the observed standard deviation between the 8 simulations of perfect inspections. The standard deviation from the covariance matrix is the root mean variance from the covariance matrix. Thus in this example at least, estimated error bars can be relied on as acceptably accurate estimates of the true error.

	α_1	β_1	α_2
observed s.d.	.021	.265	.179
from covariance matrix	.0251	.209	.166

Table 4-3. Simulation results. Standard deviation of the three fitted parameters as derived from the variance of the 8 simulations, and as calculated from the estimated covariance matrix.

Turning now to real data on infusion pumps, Table 4-4 shows the AIC and parameter estimates for the 4 combinations of distributions fitted under the assumption of perfect inspection, that is set β=1. It can be seen that distributions vary considerably. Delay-time distributions tend to be J-shaped, so that once a defect has become visible, the hazard of failure remains constant or decreases with time. Presumably the interpretation is that if a visible defect such as a crack has not caused failure for some time, it is relatively unlikely to do so in the immediate future. The occurrence of both IFOM and DFOM (increasing and decreasing force of mortality or hazard) distributions vindicates the use of the Weibull as a suitable parameterization.

For the imperfect inspection simulation the estimate of β did converge to near the true value of 0.7, and the error became smaller and smaller as the Monte-carlo sample size was increased, giving us confidence in the correctness of the Monte-Carlo program, the statistical theory, and the fitting program. For example, when the Monte-Carlo sample size is 50 breakdowns, 22 positive inspections, 181 negative inspections and 25 censored data, the estimation of β is 0.73. The AIC criterion again selected the correct distributions in all cases, out of a total of eight simulation data fitted: the four previously mentioned, each with β=1 and also with β allowed to vary.

Components	model	AIC	α_1	β_1	α_2	β_2
Door-pad	E/E	380.1	.00074	1	.00044	1
*	E/W	376.2	.00074	1	.000002	.31
C.Vs			.14	0	4.8	.56
	W/E	381.9	.00073	.944	.00043	1
	W/W	377.9	.00073	.924	.000002	.308
Battery	E/E	754.1	.0007	1	.0075	1
	E/W	756.1	.0007	1	.0073	.90
*	W/E	706.7	.00073	2.41	.009	1
C.Vs			.06	.11	.25	0
	W/W	708.6	.00073	2.41	.0093	1.20
Transducer	E/E	1400.7	.0014	1	.015	1
	E/W	1402.7	.0014	1	.0149	.997
*	W/E	1390.4	.0017	1.42	.0174	1
C.Vs			.081	.091	.215	0
	W/W	1391.9	.0017	1.43	.0174	.842

Table 4-4. Results of fitting to infusion pump data. The model selected by the minimum AIC is marked with an asterisk, and the coefficient of variation of the distribution function parameters appears below it. α_1 etc. are as defined in Table 4-2. Model types are e.g. E/W, exponential for g(u), Weibull for f(h). Units of time are days.

For real data, only the door-pad data gave a value of probability of detecting a fault $\beta < 1$. An examination of the data confirmed that failure often occurred soon after negative inspection. However, with only four breakdowns, very small sample size, the fitted value which was $\beta=0.17$ should not be taken too seriously. The maximum likelihood estimate will be biased, and the calculated confidence limits on β, which are derived assuming a Gaussian distribution of the estimator, will be too small.

4.7.2 Assessing fit of delay time distributions to infusion pump data

There are two strands to evaluating distribution fit. One approach is to embed the distribution in a more general distribution, i.e. add extra parameters and test whether they are needed, as described in e.g. Cox [1983]. This approach was not adopted, since any such extended distribution that was contemplated would be fitted, and the distribution adopted if the AIC was the lowest.

The other approach is to carry out a test of fit against a broad alternative hypothesis. The chi-squared test is simply such a (likelihood ratio) test, where the alternative hypothesis is derived by dividing the range of the distribution into classes or 'bins', and assuming a multinomial distribution of events among classes. The asymptotic form of this likelihood ratio is the familiar chi-squared, and is distributed according to the chi-squared distribution.

To apply this logic to our problem, we had to overcome two small difficulties. The log-likelihood of equation 4-7 corresponded to the continuous case, where breakdown times were not classified into bins, and the second difficulty was the existence of the time of most recent inspection. It is worthwhile sketching out our solution to the test-of-fit problem, as this is a case where the mathematical derivation gives insight into what a test of fit really is. This analysis and discussion is given in Appendix A.

Now with our real data, Figure 4-3 shows failures of peristaltic pump batteries, Figure 4-4 shows our predictions and data for positive inspections of door-pads, and Figure 4-5 shows breakdowns of volumetric pump transducers. The fit seems to be adequate for the first two distributions, with $\chi^2[6]=3.3$, $\chi^2[7]=4.8$ respectively. Both visually and numerically something is clearly wrong with the fit of volumetric pump transducers in Figure 4-4, where $\chi^2[8]=72.5$. It is clear that the precise number of degrees of freedom to be allocated to each histogram is unimportant. Some further development is clearly needed for transducers. In this pilot study we did not carry this out: our aim was merely to apply the simplest feasible method and evaluate its performance. The answer may lie with the quality of the data since components were replaced without failure by the manufacturer, some of 'no event' or e events may have involved a malfunction, and so should have been classified as breakdowns. Either way, further data collection add verification is required to resolve the situation.

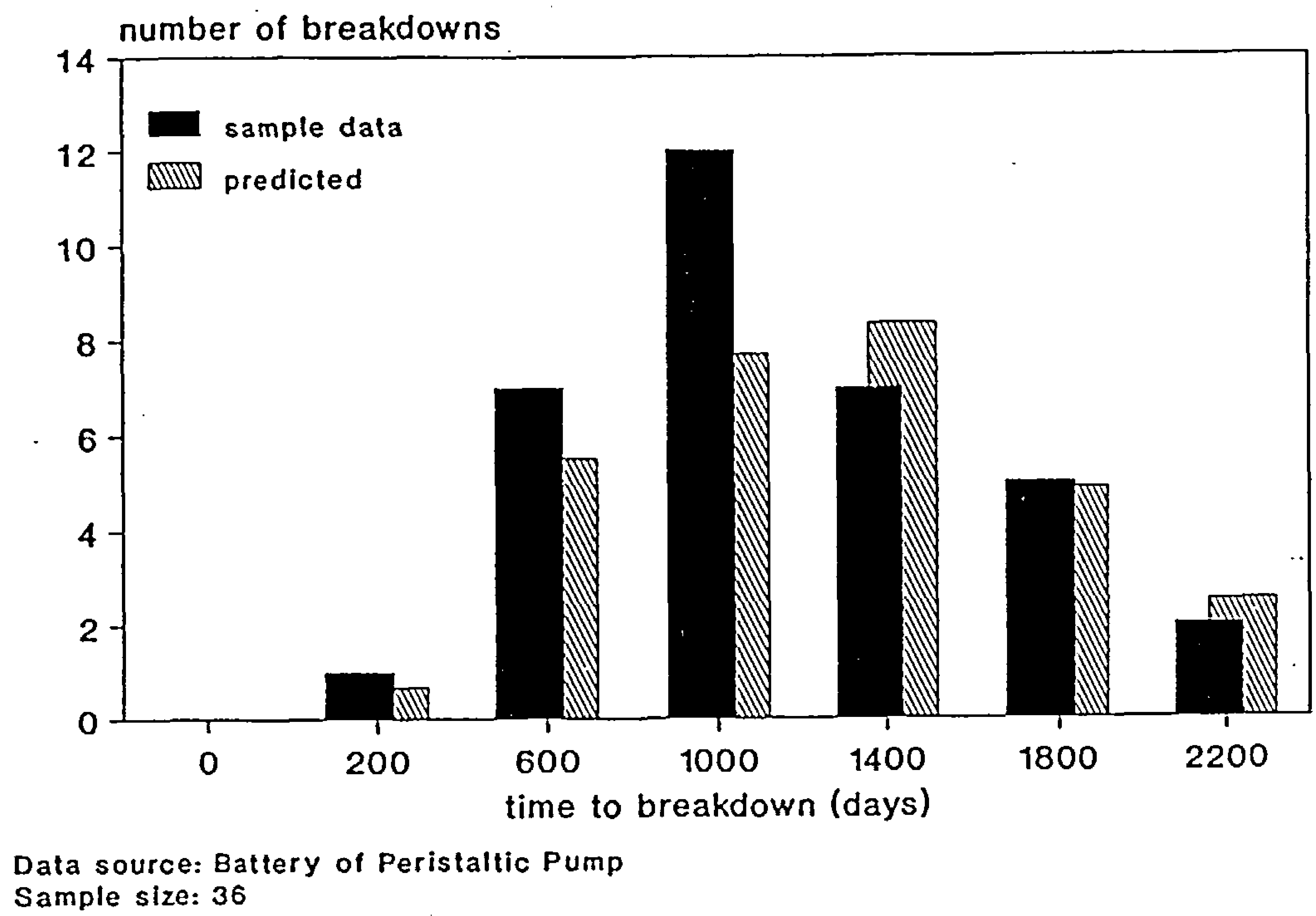
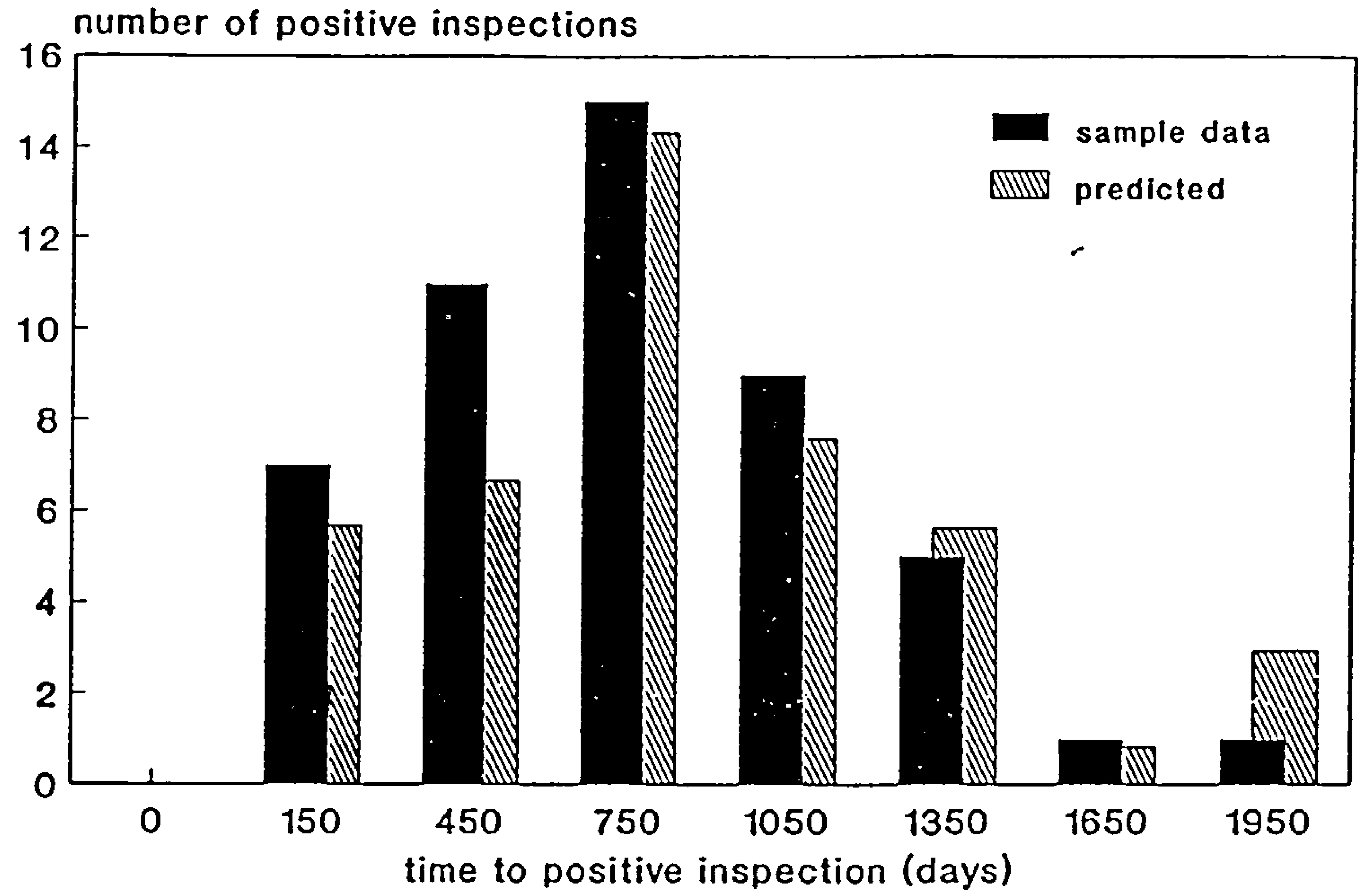
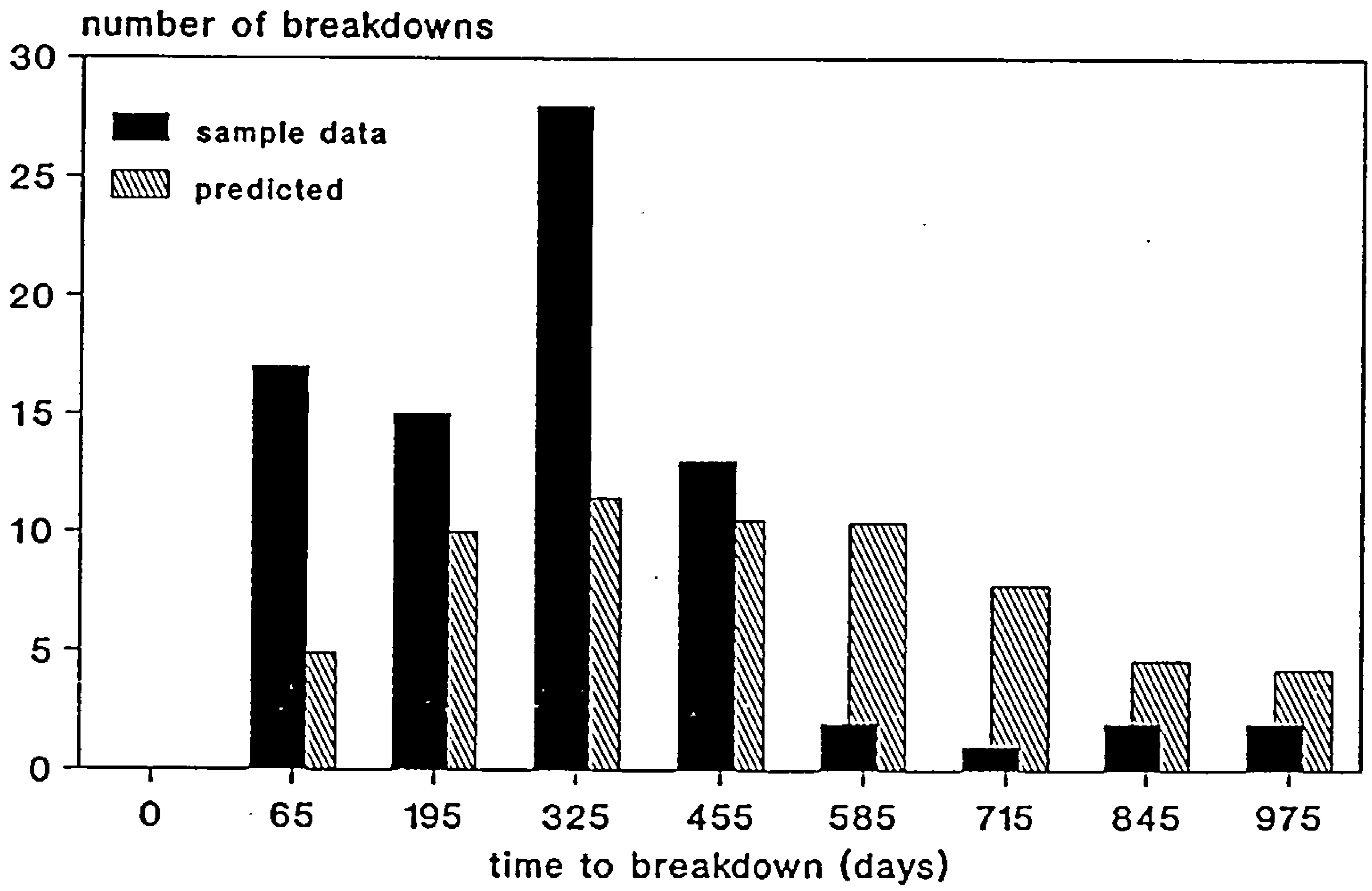


Figure 4-3 Histogram of Breakdowns



Data source:Door-pad of Peristaltic Pump
Sample size: 49

Figure 4-4. Histogram of Positive Inspections



Data source: Transducer of Vol. Pump
Sample size: 80

Figure 4-5. Histogram of Breakdowns of Volumetric Pumps

The fact that a method which works well for two components does not work so well for a (seemingly very similar) third component exemplifies the need for tests of fit.

4.8 Conclusions and further developments

This work has tested out our a priori conjecture that parameters of distributions based on the delay-time concept can be determined from 'typical' maintenance data, rather than predominantly from subjective data, and that these distributions can then be used for useful predictions. This was very much a pilot study, but some strong conclusions do emerge. We briefly review how far we might be said to have succeeded in our aim, enter some caveats, and conclude with a short list of future projects which will be explained in next Chapter.

The Monte-Carlo studies were limited in scope, but we have demonstrated the feasibility of using maximum-likelihood estimation of distributions parameters, and their error bars, using the AIC to select the optimum distributions. Even under imperfect inspection, the correct distributions could be found, and values of distribution parameters recovered.

The first attempt to fit distributions to real-life data and to make predictions based on that distribution would seem encouraging. As expected, there were large error bars on the estimated mean delay-time itself, resulting from the strong censoring imposed by periodic inspections. Tests of fit showed acceptable predictions of time to breakdown and numbers of positive inspections in two out of three applications. In the third, unfailed components had been replaced by the manufacturer, and some of these 'unfailed' replacement events may have been actual failure replacements.

Clearly more detailed modelling may be needed in particular cases, trying alternative distributions to the Weibull, and perhaps embracing general assumptions. The method needs some 'TLC' (Tender Loving Care), as messy real-world features of the data intrude rudely to spoil the elegance of theory. Fortunately, the likelihood method can cope well with censored and incomplete data. Methods of developing further need to be created, and evolved using several data sets. For example, fitting a piece wise exponential distribution would enable a histogram of the hazard function to be displayed corresponding to the pdf g(u), and this could be used to suggest fruitful parameterizations of g(u).

A caveat is that just any old information logged by technicians may not be adequate, if crucial information is missing, e.g. whether an event was a scheduled ppm or a breakdown repair. However, the method will work, given only information that lies readily to hand and can be recorded and computerized. If only some sufficiently complete database is maintained, the method described here, or a tailored version of it, can be the tool to convert a mass of lifeless data into cost-saving recommendations.

In conclusion, we have shown the feasibility of using objective data in the estimation of delay time distributions, and our experience suggests several areas for future extensions. These include developments of the method and development of diagnostic tools, such as the already mentioned use of piece wise functions as an exploratory tool in model development.

In the next Chapter, some further developments to the basic methods presented here will be explored to model more realistic and complicated cases based upon the same objective data. In comparison with the basic methods, the AIC is also chosen to select the 'best' distribution and see whether the complex distribution will improve the fit to data.

CHAPTER 5 EXTENSIONS TO THE BASIC METHODS IN CHAPTER 4

In the last Chapter, a basic method of the selection of distribution functions and parameter estimation for distributions of u and h was developed based upon objective data, i.e. the historic data of failures and inspections. The method contrasts with the previously advocated technique of assessment of distributions using mainly subjective data. Instead, the criterion of minimum Akaike Information Criterion (AIC) was utilized for delay time distributions fitting to data, and was successfully applied to real world data. It is therefore evidently possible to use routinely collected data (collected by technicians for purposes of maintenance) to estimate the parameters of the delay-time distributions, and to then use the distributions in inspection model to predict the optimum interval between ppms, for example. This latter calculation, unlike the fitting process, can be carried out by Monte-Carlo methods, and is the end-product and raison d'etre of the modelling process.

In more generality, as stated in the last Chapter, the AIC is -2log(L)+2f, where L is the total likelihood, and f the number of fitted parameters. The minimum AIC model is designed to be the 'best' distribution for predictive purposes on fresh data. The rationale behind it is the concept that although increasing the number of distribution parameters causes the fit to a given sample of data to become better, the likelihood function calculated by applying this distribution to a fresh sample is smaller. However, since the parameters estimated are not the true population values, there must be errors involved in the estimated parameters. On correcting the large-sample log-likelihood from a fit to data, one obtains the AIC.

The AIC is a complicated function of the distribution parameters. Hence it is difficult to assess the goodness of fit to data by plotting residuals. A global likelihood ratio test of fit is described in the last Chapter, but such tests, which have no specific alternative hypothesis, lack power to detect any specific weakness of the distribution. However, the creation of more elaborate distributions which generalize the simpler distribution does provide a means of ensuring the adequacy of distribution. The fitting of these more complex distributions is thus always beneficial; either a better fit to data is obtained, or one is given an assurance that the simpler distribution is adequate. In this Chapter some natural extensions are made to the basic methods, to model more realistic and complicated cases, and to make the technique more robust.

The new developments investigated in this Chapter are:

1. The functions g(u) and f(h), hitherto parameterized as Weibull distributions, may depend on the machine age t. For instance, this can be achieved by letting the scale parameter of the distributions be a function of the age t at the moment the component was renewed, e.g. by taking the scale factor $\propto e^{\lambda t}$.
2. Inspection may have a beneficial or adverse effect on a component's performance. We assume that the inspection exerts this influence by adding or subtracting a period Δ to the effective age of the component. Δ may be estimated along with the other parameters through the maximum likelihood method.
3. Often in practice, machines from which the data are collected have different usages and ages, and can not be treated as identical. The maximum likelihood principle can be extended to cope with a population, via the 'Empirical Bayes' method.
4. Two mechanisms which can induce correlations between the periods u and h are discussed. One mechanism, which gives rise to positive correlations, invokes a population of components. The other, which gives rise to a negative correlation, requires a two-stage failure process, with an additional delay after the completion of the first stage before a fault becomes visible.
5. Some miscellaneous topics are also discussed, e.g. imperfect inspections and the parameterization of g(u) and f(h).

Distribution functions including the extensions enumerated are derived, and tested on real world data and compared with the results by using simple distributions in the last Chapter.

5.1 The effect of machine age

New components introduced as replacements into an aging machine may fail more quickly than they would if the machine were new, because of operating in an aging environment. Both the time u to visibility of a defect, and the time h to subsequent failure, may be shortened.

Let the hazard functions of u and h be defined by $\psi(u)$ and $\phi(h)$ where $\psi(u)=g(u)/\{1-G(u)\}$ and $\phi(h)=f(h)/\{1-F(h)\}$. This effect can be parameterized by increasing the hazard $\psi(u)$ of a defect becoming visible, and the hazard $\phi(h)$ of subsequent failure. Alternatively, the times u and h could be scaled up. For the Weibull distribution, these proportional hazards and accelerated life models are functionally equivalent, Cox [1983]. We chose to multiply the scale factors α of the Weibull distributions used to parameterize g and f by $e^{\lambda t}$, where t is the machine's age at the time the component was renewed. This means that λ can take any value from $-\infty$ to $+\infty$, with a zero value of λ if the machine age has no effect on hazards $\psi(u)$ and $\phi(h)$. Other parameterizations were also tried, e.g. $\alpha \propto 1+\lambda t$. However, $\alpha \propto t^{\lambda}=\exp\{\lambda\log(t)\}$ is unacceptable, as the hazard of the first failure would be zero.

5.2 Hazardous or beneficial inspections

An inspection could have other effects besides the replacement of visibly defective components; a hazardous inspection might damage components and increase their hazard of subsequent failure, and on the other hand, a beneficial inspection might reduce the hazard of failure. This could happen if the inspection included an overhaul, e.g. via some adjustment or lubrication. In the context of medical screening, Lee [1987], procedures such as X-raying could increase the risk of the cancer they attempt to detect.

Such effects could be parameterized in very many ways, but a simple and economical approach where the hazard of failure increases with age would be to regard the inspection as either aging or rejuvenating the component by some period Δ. Here the parameter Δ means the age effectively removed from the component by overhauling it. This parameterization is attractive because only one parameter is required, and it has a simple interpretation.

More precisely, the hazard $\psi(u)$ of developing a visible fault at age u would, after inspection, jump back to its (lower) value at age $u-\Delta$.

Common sense suggests that this simple idea needs some modification. Two consecutive overhauls could only produce the rejuvenating effect of the first: once lubrication or adjustment had been carried out, no further improvement could result from immediately repeating the process. On the other hand, a second hazardous overhaul could well produce still more damage to components. Hence the age removed from a component should not exceed either the period from its last overhaul, or its total age. This last restriction means that overhauls can never restore a machine to 'better than new'.

Given $g_0(u)$ a pdf. for a defect arising at time u, and the corresponding distribution function $G_0(u)$ without any inspection influence, it is straightforward to calculate the pdf. of u and distribution function G(u) obtaining at time u, after n such hazardous or beneficial inspections with parameter Δ. If the hazard of developing a defect at age t is $\psi(t)$, after i-1 inspections at times $t_0, \ldots, t_{i-1}$, where $t_0=0$,

$$t \Rightarrow t_{effective} = t - \sum_{j=1}^{i-1} \mathrm{Min}\{t_j - t_{j-1}, \Delta\},$$

for $t_{i-1} < t \le t_i$, and $\psi(t) \Rightarrow \psi(t_{effective})$. The sum is nugatory if $j > i-1$, i.e. if $i=1$, so that no inspection has yet occurred.

The survival function $S(u)=1-G(u)$ proves unexpectedly complicated when $\Delta \ne 0$. Let S_0 be the survival function when $\Delta=0$. The equation

$$S_0(u) = e^{-\int_0^u \psi(t)dt} \qquad 5\text{-}1$$

is the key to calculating $S(u)$. For $t_{i-1} < t < t_i$, the hazard is $\psi(t - \Sigma_{j=1}^{i-1} Min\{t_j - t_{j-1}, \Delta\})$. The integral $\int_0^u \psi(t_{effective})dt$ must then be carried out piece wise, and is

$$\int_0^u \psi(t_{effective})dt = \sum_{i=1}^{n+1} \int_{t_{i-1}}^{t_i} \psi(t - \sum_{j=1}^{i-1} Min\{t_j - t_{j-1}, \Delta\})dt, \qquad 5\text{-}2$$

where a total of n inspections have been carried out by time u from renewal, $t_0=0$, $t_{n+1}=u$.

It is now possible to write down the survival function S, using the equation

$$e^{-\int_{t_{i-1}}^{t_i} \psi(t)dt} = S_0(t_i)/S_0(t_{i-1}),$$

derived from equation 5-1. Treating each term in the summation in equation 5-2 in this way, and remembering that $S_0(0)=1$, we finally have

$$S(u) = \prod_{i=1}^{n+1} \frac{S_0(t_i - \Sigma_{j=1}^{i-1} Min\{t_j - t_{j-1}, \Delta\})}{S_0(t_{i-1} - \Sigma_{j=1}^{i-2} Min\{t_j - t_{j-1}, \Delta\})}, \qquad 5\text{-}3$$

where u appears on the right hand side in the guise of t_{n+1}. Clearly, for exponential distributions the additional terms due to Δ cancel as they must, because when the hazard ψ is a constant, rejuvenation can have no effect upon it.

The pdf. $g(u) = -dS(u)/du$ is obtained by differentiating equation 5-3

$$g(u) = \psi(u - \sum_{j=1}^{n} \mathrm{Min}\{t_j - t_{j-1}, \Delta\})S(u),$$

for $u > t_n$. In terms solely of the original survival function S_0 and pdf. g_0, the pdf. is

$$g(u) = g_0(u - \sum_{j=1}^{n} \mathrm{Min}\{t_j - t_{j-1}, \Delta\})S(u)/S_0(u - \sum_{j=1}^{n} \mathrm{Min}\{t_j - t_{j-1}, \Delta\}),$$

where S is as defined in equation 5-3.

It is now possible to compute G(u) and g(u) when Δ is nonzero, if the original distribution function $G_0(u)$ and pdf. $g_0(u)$ can be computed.

Whether or not rejuvenation would be an improvement would depend on whether the hazard of a defect developing was increasing or decreasing with age --- restoring the machine to an earlier and more unreliable state would not be an advantage. The basic concept of changing the component's effective age is still valid for such DFOM (decreasing force of mortality) distributions, but here it is the increase in age that must be restricted. It is simplest to write

$$t \Rightarrow t_{effective} = t + \sum_{j=1}^{i-1} \mathrm{Min}\{t_j - t_{j-1}, \Delta\},$$

and to define Δ as the increase in age conferred by the inspection. However, for DFOM distributions the rationale of this approach, namely the notion of restoration to a younger and more reliable state, is not appropriate.

5.3 Population of machines and/or components

Initially all machines in the sample were treated as identical. However, usage of machines may vary, as well as intrinsic robustness. Modelling a population of machines with varying 'frailty is at once a means of generalizing the method, and of testing the hypothesis that machines are identical, by examining the decrease in AIC on introducing a population.

The following logic can be regarded as simply an application of maximum likelihood estimation, where it is usual to sum the likelihood over all events that could have given rise to the observations. However, it also falls under the heading of 'Empirical Bayes' methods, Maritz and Lwin [1989], as the pdf. of the population of frailties can be regarded as a Bayesian prior distribution; however, this is a 'prior' whose parameters can be estimated, hence the qualification of 'empirical'.

Let the scale factors of g(u) and f(h) be proportional to a frailty λ, characteristic of each machine. λ is assumed to be a random variable from a distribution such as log-normal or Gamma. Without loss of generality, it can be assumed the mean of λ to be unity. Then the pdf. of frailty is e.g.

$$p(\lambda|\gamma) = \frac{\gamma(\lambda\gamma)^{\gamma-1}e^{-\lambda\gamma}}{\Gamma(\gamma)}, \qquad 5\text{-}4$$

where the variance of λ is γ^{-1}. The likelihood function corresponding to each machine is conditional on λ, and must be integrated over all the (unknown) values of λ, i.e.

$$L(x|\gamma) = \int_0^\infty L(x|\lambda)p(\lambda|\gamma)\, d\lambda,$$

where x is the vector of observations. $L(x|\gamma)$ must be maximized for γ as well as for the parameters of f(h) and g(u), which are now parameters at average frailty. The total likelihood is the product of the likelihoods corresponding to individual machines.

As $\gamma \Rightarrow \infty$ and the variance of the population tends to zero the likelihood reverts to that calculated assuming identical machines.

5.4 Correlation between u and h

The concept of a two-stage failure process leads naturally to the assumption of independence of u and h. The condition of the component deteriorates to a point where it is held to be defective, and then further deterioration ensues until the component is regarded as failed. The determinant of time to failure would then be merely the fact of the machine's defective condition and not the time taken to deteriorate to that condition. Hence u and h should be independent.

However, there are several reasons why this simple notion might need modification. One is that there could be a population of components, some with longer periods both for a defect becoming visible and for subsequent failure, and some with shorter periods for both events. The two periods would then be correlated. Formally, one mode of generating correlated values of u and h would then be to consider a distribution of scale factors λ such that components have a hazard $\lambda\psi(u)$ of becoming visibly defective and hazard $\lambda\phi(h)$ of failing. Integrating the joint pdf. g(u)f(h) over the population of unobserved values of λ would give a correlated bivariate distribution of u and h. This is precisely the same 'Empirical Bayes' logic considered earlier, and generates only positive correlations. When g(u) and f(h) both have exponential distributions, and λ a Gamma distribution as in equation 5-4, the integral can be evaluated analytically. Here $g(u)=\alpha_1 e^{-\alpha_1 u}$, $f(h)=\alpha_2 e^{-\alpha_2 h}$, and

$$g(u)f(h) \Rightarrow \int_0^\infty \lambda^2 \alpha_1 \alpha_2 \exp(-\lambda\{\alpha_1 u-\alpha_2 h\})f(\lambda|\gamma)d\lambda.$$

This yields

$$g(u)f(h) \Rightarrow p(u,h) = \frac{(\gamma+1)\alpha_1\alpha_2}{\gamma\left[1+\dfrac{\alpha_1 u + \alpha_2 h}{\gamma}\right]^{\gamma+2}},$$

the Bivariate Pareto distribution for u and h. As $\gamma \Rightarrow \infty$, by virtue of the limiting result

$$\mathrm{Lim}_{n \Rightarrow \infty}(1 + x/n)^n= e^x,$$

it can be seen that p(u,h) reverts to its original form, as it must.

When g(u) and f(h) are Weibull variables, u and h are replaced by their powers, if the hazards are scaled by λ (the proportional hazards assumption). The Takahasi-Burr distribution resulting has survival function:

$$P(U>u,\ H>h) \equiv S(u,h) = \frac{1}{\left[1+\dfrac{(\alpha_1 u)^{\beta_1} + (\alpha_2 h)^{\beta_2}}{\gamma}\right]^{\gamma}}. \qquad 5\text{-}5$$

The bivariate pdf is

$$p(u,h) \equiv \partial^2 S/\partial u \partial h = \frac{(\gamma+1)\alpha_1\alpha_2\beta_1\beta_2(\alpha_1 u)^{\beta_1 -1}(\alpha_2 h)^{\beta_2 -1}}{\left[1+ \dfrac{(\alpha_1 u)^{\beta_1} + (\alpha_2 h)^{\beta_2}}{\gamma}\right]^{\gamma+2}}. \qquad 5\text{-}6$$

These and similar distributions resulting from components sharing the same environment are described by Hutchinson and Lai [1990]. If the accelerated life model is used, u and h are scaled by λ rather than u^{β_1} and h^{β_2}, and the integration must be carried out numerically.

Stone [1978], as quoted by Lawless [1982], measured times to development of microscopic faults in electrical cable insulation, and subsequent times to failure. The Spearman correlation between u and h for these data is ρ_s=0.583, which suggests that the Takahasi-Burr distribution could be fitted. The results of this fit are quoted later in this Chapter.

Another mechanism by which f(h) might become dependent on u, i.e. become f(h|u). would be if the component condition regarded as defective were to vary from inspection to inspection. Since the total time u+h to failure would be unaltered by the precise definition of 'defective', longer periods u would be associated with shorter periods h. This would induce a negative correlation between u and h.

The epidemiological analogy makes the logic clearer. Death from a disease is a two-stage process, with the period from birth to infection being the first stage, and the period from infection to death the second. There will in general be a lag t between infection and diagnosis of the disease. The period u corresponds to time to diagnosis, and h is the period from diagnosis to death.

If the distribution of diagnosis lag t, ie. from infection to diagnosis of the disease, is q(t), the bivariate distribution of u and h changes from g(u)f(h) to p(u,h), such that

$$p(u,h) = \int_0^u g(u-t)f(h+t)q(t)dt.$$

The periods u and h are now correlated, unless the distribution f of delay time is exponential, when f(h+t) factorizes, and u and h are still uncorrelated.

5.5 The parameterization of the distributions g(u) and f(h)

The Weibull distribution with survival function $S(t)=e^{-(\alpha t)^{\beta}}$ was used in the simple model of the last Chapter, as it can model both increasing and decreasing hazard distributions with only two parameters, and has been very widely applied in failure-time problems. An obvious drawback is that for $\beta > 1$ the hazard $\alpha\beta(\alpha t)^{\beta-1}$ is zero when t is zero. This seems restrictive, and a 3-parameter Weibull distribution with survival function

$$S(t) = e^{(\alpha\delta)^{\beta} - (\alpha(t+\delta))^{\beta}}$$

where $\delta > 0$ allows the initial hazard to be nonzero. As will be seen, it is necessary in fitting data to compute the distribution functions G(u) and F(h). Use of the Weibull distribution, where G and F can be evaluated without resorting to numerical integration, is an advantage. Error bars on the fitted distribution f(h) are higher than those on g(u), because under regular perfect inspection, a fault that has become visible can not remain so for longer than the period between inspections. The tail of the f(h) distribution is thus undetermined. Hence in this study the f(h) distribution was parameterized more simply than was the g(u) distribution. With Weibull distributions for u and h, the required probabilities are:

$$G(u) = 1 - \exp\{(\alpha_1\delta)^{\beta_1} - (\alpha_1(u+\delta))^{\beta_1}\},$$

$$g(u) = \beta_1\alpha_1^{\beta_1}(u+\delta)^{\beta_1-1}\exp\{(\alpha_1\delta)^{\beta_1} - (\alpha_1(u+\delta)^{\beta_1}\},$$

$$F(h) = 1 - \exp\{-(\alpha_2 h)^{\beta_2}\},$$

$$f(h) = \beta_2\alpha_2^{\beta_2} h^{\beta_2-1}\exp\{-(\alpha_2 h)^{\beta_2}\},$$

where α_1, α_2 are scale parameters and β_1, β_2 are shape parameters.

As will be shown, bivariate distributions $p(u,h)=g(u)f(h|u) \neq g(u)f(h)$, where u and h are positively or negatively correlated can be built up from the independent distributions described here.

5.6 Assumptions for the estimation of g(u) and f(h)

Assumptions for the model to estimate the parameters of g(u) and f(h) are relaxed considerably from the list quoted in the last Chapter, section 4.2. They are now as follows:

1. The components of a machine are assumed independent, i.e. the failure of one will not affect the functioning of another.
2. Components with identified defects are repaired immediately and instantaneously.
3. Repair times are assumed to be negligible.
4. The pdfs of u and h are modelled as 2 or 3-parameter independent Weibull distributions, which are subsequently built into a bivariate distribution of pdf. p(u,h), assuming neither, either, or both of the two correlation-inducing mechanisms described.
5. Imperfect inspections: inspections are in general imperfect, i.e. they have a probability $\beta \leq 1$ of detecting a fault if it is present. When any component fails, an inspection of the machine embracing all of its components is carried out, and these

adventitious inspections have probability $\beta' \leq 1$ of detecting a fault if it exists. In general $\beta' \neq \beta$. There are no false positives, i.e. if a fault is not present one will not be identified. Probabilities β are assumed independent between inspections. In addition to this parameter β, the bivariate form p(u,h) allows some faults to be undetected before breakdown.

6. Machines are assumed to be members of a population, with varying hazards both of defects becoming visible and of these visible defects causing a breakdown. This allows for unequal usage or differing intrinsic machine 'frailties'. The distribution of hazard-scale factors was taken as Gamma in this study.
7. Repairs are no longer taken as replacements, so that the faulty component is restored to an 'as-new' condition. Instead, the hazards of failure etc. are a function of machine age.
8. Inspections are no longer assumed to simply imply replacement of defective components. Instead, they may also be either hazardous or beneficial, effectively rejuvenating or aging components.

Clearly, attempts have been made to extend the simple distribution to include likely real-life features, such as failure-rate changing with machine age, side-effects of the inspection process, and correlation between time to visibility of a defect and subsequent time to breakdown of the component. The assumptions of component independence, immediate repair and negligible repair times stand out as unconditional assumptions, rather than descriptions of parameterizations which, although themselves assumptions of a sort, are in fact attempts to relax previous more stringent assumptions.

In this study it was known that repair-times were small enough to be regarded as negligible, and that identified defects were immediately rectified.

Component dependency was not modelled, except in that breakdown and subsequent repair of any component caused an 'adventitious' inspection of all other components. The approach used to deal with hazardous inspection could be used to treat component dependency; breakdown of (or even appearance of a visible defect in) any component could exert an aging effect on all other components of the machine. The amount of aging could be taken as a universal constant, or as a matrix of pair wise increments, where Δ_{ij} would be the increase in effective age of the ith component caused by breakdown of the jth. Thus, although our data were inappropriate for developing such distributions of pair wise dependency, all the necessary mathematics has been derived, and the fitting of these extra parameters would present no difficulty for our approach.

5.7 Calculation the likelihood

For reasons of which we have stated in the last Chapter, it again turns out that only three key probabilities need be considered; the likelihood can be built up from these three, and others which are special cases of them. The general case which allows negatively correlated periods u and h is now considered. A defect arises at time w with pdf g(w), distribution function G(w), becomes visible at time u with pdf q(u-w), distribution function Q(u-w) and causes breakdown of the component at time t=u+h with pdf f(h), distribution function F(h).

When the pdf. of the lag u-w between a defect arising and becoming visible is a Dirac delta-function so that these events occur simultaneously, the model presented here reduces to the simpler one proposed in the last section where a defect arises and becomes visible at time u with pdf g(u).

The probability expressions are quite similar to the ones in the last Chapter. But since we have introduced an extra period between the time of a defect arising and the time of this defect becoming first visible as a more general case, expressions 4-2, 4-3, and 4-4 need to be slightly modified. For details see Appendix B.

The three key probabilities $P_b(t_n,t)$, $P_e(t_n,t)$ and $P_y(t_n,t)$ are conditional on the last renewal. The corresponding expressions in the simpler case when there is no time-lag between a fault arising and becoming visible are derived by setting Q(t) equal to unity everywhere it appears, irrespective of the value of t.

Similar to that in previous Chapter, the likelihood is calculated by accumulating the product of these three terms. Each renewal may be followed by a sequence of negative inspections, and this must terminate in an event of type b, e, or y. Event e is really 'no event'. The likelihood L for a total of n_b breakdowns at times t_i, n_e 'no failure before observation ceases' events at times t_j, and n_y positive inspections at times t_k, is

$$L = \prod_{i=1}^{n_b} P_b(t_i^*, t_i) \prod_{j=1}^{n_e} P_e(t_j^*, t_j) \prod_{k=1}^{n_y} P_y(t_k^*, t_k) \qquad 5\text{-}7$$

where the notation t_i^*, t_j^*, t_k^* denotes the time of the latest negative inspection or, failing that the latest renewal, such that $t_i^* < t_i$, and so on. The likelihood under imperfect inspection can also be written in terms of the three key probabilities, as described in the last Chapter.

5.8 Results of fitting the extended distributions to data

Data comprised historical data on inspection results and breakdown times of the three infusion-pump components described in the last Chapter. In addition, the data of Stone, Lawless [1982] provided direct measurements of the periods u and h for failures of epoxy insulation of cables.

Turning to this data set first, it comprised 17 measurements of u, h pairs, and three censored cases in which no defect had appeared after a long time interval, and for which consequently only a lower bound on u was quoted. The likelihood to be maximized is the product of terms

$$\prod_{i=1}^{17} \frac{\partial^2 S(u_i,h_i)}{\partial u \partial h} \prod_{j=1}^{3} S(u_j,0)$$

where the survival function S(u,h) is given by equation 5-5. The bivariate pdf. has been written in terms of the survival function. A PC 386-based FORTRAN program was written, which called the NAG routine minimiser E04JBF to minimize the resulting AIC. The fitted value of γ was 0.39, corresponding to a coefficient of variation of $\gamma^{-1/2}$=1.6 for the Weibull scale factor, a very large variation in 'frailty' between samples. Several other parameterizations were also investigated, i.e. the use of a log-normal distribution for the Weibull scale factors rather than the Gamma, and the use of an accelerated life model rather than a proportional hazards model. The likelihood proved to be very insensitive to the different parameterizations.

To compute the Pearson correlation predicted by the fitted model, values of u were assumed censored at t=1740 hours, the lowest censored u value, and the distribution resulting from the conditional survival function

$$S'(u,h) = \frac{S(u,h) - S(u_0,h)}{1 - S(u_0,h)}$$

The Pearson correlation of the 17 data values was 0.505, and the correlation from the fitted distribution was 0.530, showing that the method has reproduced the observed correlation between the variables. The bivariate distribution does not fit better merely because it chances to fit the marginal distributions for u and h better than does the product of two unvariable distributions.

However, the population method also predicts a long tail to the distribution of u, and allows the distribution to successfully fit data points that would undoubtedly have been regarded as outliers and ignored under the hypothesis of independent u and h. The log-likelihood decreases by 4.7 on adding the mixing parameter γ, which corresponds to a decrease in goodness of fit chi-squared of 9.4. The criterion of minimum AIC would definitely require this parameter to be present. With populations of component 'frailties', one expects both correlation between u and h, and also long-tailed marginal distributions.

Turning to the infusion-pump data where the observations are indirect, the FORTRAN program used for the earlier study was considerably enlarged. A total of ten parameters can now be fitted, i.e.

1. The three Weibull parameters α_1, β_1 and δ for the distribution g(w) or g(u).
2. The two Weibull parameters α_2 and β_2 for the distribution of f(h).
3. The probability β that an inspection detects an existing defect.
4. The age Δ removed or added by a beneficial or hazardous inspection.
5. The (compound) rate of increase with machine age λ of hazards of defects arising and causing failure.
6. The scale parameter η of the exponential distribution of the delay between a fault arising and becoming visible. The distribution function is $Q(t)=1-e^{-\eta t}$.
7. The standard deviation $\gamma^{-1/2}$ of the log-normal distribution of the hazard scaling factors for different machines in the population.
8. Ditto, for the population of components.

Of these eleven parameters, only ten could be fitted at once, as the idea of a population of machines and a population of components simultaneously would require a double integration, and hence in total a triple integration for the likelihood, which would be prohibitively slow on a 16 MegaHertz 386-PC.

As before, the minimum value of the Akaike Information Criterion (AIC) was sought.

Since cycling through all possible subsets of the full distribution was a very time-consuming process, the program was modified to permanently keep any parameter that decreased the AIC, but to continue evaluating other parameters from distributions both with and without a parameter that failed to decrease the AIC. This strategy allows for the fact that some parameters can and were seen to potentiate others. Also, some parameters are known to exert no effect at all in the absence of others, and hence certain combinations did not need to be explored.

For example, the third Weibull parameter δ has no effect for exponential distributions where the shape parameter $\beta=1$, and neither does the rejuvenation parameter δ. The delay in visibility of a defect does not produce a correlation between u and h if the distribution of h is exponential (although it could in theory improve the distribution of g(u). Regarding imperfect inspections, there would be little chance of obtaining a superior fit with $\beta < 1$ if the likelihood, obtained with $\beta \cong 0.99$ and all other distribution parameters fitted, was increasing with the value of β. A population would also tend to zero variance if the likelihood tended to increase as variance decreased, for small variances. Avoiding these cases speeded up the computations to the point of feasibility.

Data for three components were fitted, and the values of fitted parameters and their error bars are shown in Table 5-1. The third Weibull parameter δ lowered the AIC for transducers, but the 2-parameter distribution was adequate for the other two components.

Increasing machine age increased the hazard of failure of batteries, to the extent that a lower AIC was obtained when the hazards of the g(u) and f(h) distributions were allowed to increase exponentially with machine age. This parameterization gave a better fit than a linear increase. Over the 10-year life-span of the machine, the value of λ of 0.0002 would give a twofold increase in hazard. New batteries inserted into an old machine would have a significantly shorter time to failure.

Transducer			
Parameter	Description	Values	Stan. Dev.
α_1	g(u) scale factor	0.0009	0.0009
β_1	g(u) shape factor	2.90	2.50
α_2	f(h) scale factor	0.0171	0.0036
δ	g(u) third parameter	518.0	1051
Battery			
α_1	g(u) scale factor	0.0007	0.00004
β_1	g(u) shape factor	2.72	0.3
α_1	f(h) shape factor	0.008	0.002
λ	machine age	0.0002	0.0001
Door-pad			
α_1	g(u) scale factor	0.049	0.003
β_1	g(u) shape factor	2.41	0.25
α_1	f(h) scale factor	0.0002	0.0001
β_1	f(h) shape factor	2.58	0.91
β	imperf. insp. para.	0.175	0.023

Table 5-1 Fitted parameter values and their standard deviations for minimum AIC delay time distributions of historical breakdown and inspection data for three infusion pump components.

Allowing inspection to be imperfect improved the fit for door-pads. The data do show some very short times from negative inspection to failure, as well as much longer times, so this result is not surprising. It could well be that door-pads progressively deteriorate, and that no well-defined defect is evident. It is interesting that the imperfect inspection parameterization fitted the data much better than the 'Q' distribution approach in which there is a time delay from a fault arising to its becoming visible.

None of the other parameters were able to produce a better fit as defined by one having a smaller AIC. Table 5-2 shows the parameter values obtained. Although the populations of machines and components seemed to have quite a large variability in hazard of failure, this increased the likelihood function of the fit only slightly, so that the AIC did not decrease. There must be a much higher variability before a population method will significantly improve the fit to data, as evidenced by the 300% coefficient of variation of the hazard scaling factor for the Stone data.

parameter	ηdays^{-1}	σ_m	σ_c	Δ days
transducer	0.88	0	0	473
battery	2.01	0.127	0.170	-60
door-pad	0.93	0.29	0	-82

Table 5-2 Values of parameters that would not be introduced into delay time distributions because they did not reduce the AIC in any of the three cases considered. Here ηdays^{-1} is the scale factor of the 'Q' distribution, σ_m the standard deviation of the hazard scale factor of the machine age, σ_c that of the component population, and Δ days the rejuvenation conferred by maintenance.

For our data, it seems that inspection is neither significantly hazardous or beneficial to the components studied. Also, the simple distributions of uncorrelated times to visibility of a defect, and subsequent time to breakdown cannot be improved upon. This is now known, because the parameter η that would import a negative correlation between u and h was not required, nor were the population parameters that would have given a positive correlation.

5.9 Conclusions

In this Chapter, a number of extensions to the relatively simple method of the estimation of delay-time distributions proposed in last Chapter have been described. It proved quite feasible to fit such extended delay time distributions to data, at the cost of writing a 1500-line FORTRAN program, and of waiting a few hours for it to run on a slowing 386-PC.

Some of the extensions, such as a third Weibull parameter, a hazard of component failure increasing with machine age, and a population of components of differing 'frailties' did improve the fits to data, but only for particular components.

We suggest that such extended methods are useful, as goodness-of-fit tests are insensitive towards particular defects of a fit to data, and because diagnostic plots have not been devised for this situation. Therefore, to be certain that a simpler method is adequate, parameterizations that relax the assumptions of the simpler method are needed.

After fitting the gamut of delay time distributions described in this Chapter, to continue developing a delay-time model, one would focus attention on those parameters whose inclusion drastically lowered the AIC, and extend the parameterization in that area. For example, where a population of components was indicated by a large value of the coefficient of variation of the Weibull scale factors, one would vary the compounding distribution from log-normal, and introduce other distributions, perhaps with more parameters.

The extensions to the basic method proposed here were intended to relax as many assumptions as possible, while economizing on the number of fresh parameters to be fitted. The extensions which were not required in order to fit these data were the concepts of a time lag in the visibility of a defect, and of hazardous or beneficial inspections. The practical usefulness of these extensions will only become clear after more data have been fitted. However, it is hard to believe (for example) that inspections are always neutral, and never exert a beneficial or hazardous effect.

5.10 Discussion on the methods of estimating the delay time distributions

In Chapter 4, section 4.1, we briefly introduce the basic methods of estimating the delay time distributions by subjective data developed by Christer [1984-1991]. Then, we intensively described the approaches used in the estimation of the delay time distributions using objective data. In fact, both methods are useful in certain instances which depend upon what kind of data are available. If there are no historical records of breakdowns and inspections in the past available, the only method we can use is the subjective assessment of the delay time distributions. Our experience also shows that such estimation can be close to the reality. When we have obtained some objective data, since we know that both the delay time and initial time distribution functions are estimated from indirectly observed data, (i.e., in general, no one can directly observe the delay time and initial time), and therefore the sample size should be big enough to make the estimation accurate. We especially need to know whether defects are found or not at inspections.

Another problem in using objective estimation is that the objective data may contain no information on some parameters, for example, the tail of the delay time distribution. Because the information we can obtained through the records of breakdowns and inspections can only tell us that the delay time is either shorter or greater than the inspection interval, if the inspections are carried out strictly according to a pre-scheduled inspection interval, there is limited information contained in such data. However, an engineer may know the consequence of introducing a very long inspection interval. Here one would be extrapolating objective data well into the tail of the delay time distribution, determined purely from data of small delay times. It may also be that subjective data, which gives estimates of u and h for each failure, might allow the probability of detecting a defect, β, to be determined more accurately, Christer and Waller [1984c].

It is also noted from our data that the records kept in the maintenance department are not usually as accurate or complete as desired. For example the transducer data records show that all the transducers were replaced after one year by the manufacture because of the design problem, but there are no indications as to which were replaced on failures and which were not. In this case, obviously the use of all available data is needed, both subjective and objective data. The likelihood is then the product of the likelihoods for the two types data. Distribution forms for subjective data can be parameterized, and the best parameters found by minimizing the AIC.

However, in general, the objective data assessment of the delay time distributions are reliable and accurate because it comes from the real world data. Compared with the subjective data estimation, it also gives more confidence since we can directly carry out the goodness of fit test to confirm the models since even after the estimation of the delay time distribution by the subjective data, we still need objective data (if available) to revise the distributions.

When subjective and objective data conflict, the question arises as to whether we can salvage anything useful from subjective and objective data? Clearly this is an interesting new area, full of unsolved questions.

CHAPTER 6. MODELS OF CONDITION MONITORING INSPECTIONS

An early paper, Christer and Waller [1984a] presented prototype models of the inspection of industrial equipment which were constructed utilizing the concept of delay time and delay time analysis. In a more recent paper, Christer [1992] extended the basic model to embrace condition monitoring tests where the test is assumed to be a (0,1) type in that it records there is either nothing wrong, or a defect requiring repair is identified. The model developed in Christer's paper [1992] is essentially for the case where the time horizon is infinite, but the case of a finite time horizon is also briefly discussed. Based upon that paper, extensions are made here to discuss more cases of condition monitoring inspection modelling, particularly models over a component life cycle or over a finite time horizon. For the sake of continuity, models over infinite time horizon are also included here as a part of the discussion because they share the common notation. Compared with the paper by Christer [1992], different format of modelling is adopted here in order to make the presentation consistent with the previous Chapters. As usual, it is convenient to first introduce the basic assumptions and notation of models.

6.1 General assumptions and notation

Condition monitoring inspection models vary according to the chosen time horizon, whether be it finite or infinite, the assumption of perfect or imperfect inspections, the nature and content of condition information obtained, and the decision criterion. For convenience, before we proceed to discuss specific models, we introduce some general assumptions for the models established in this Chapter. They are as follows:

1. Time is measured from the origin $t_0=0$, and t_i is the time to the ith inspection from t_0.
2. The initial point of a defect is at time u after t_0, and the pdf.

of u is known and denoted by g(u).

3. The delay time h of a defect has a known pdf. f(h).
4. The delay time h is independent of its initial time u.
5. Inspections are ordered as a sequence of t_i, i=1,2,....
6. Defects identified at an inspection are repaired or replaced as part of the inspection. Failures arising are rectified at once as breakdown repair or replacement.
7. Repair is equivalent to replacement in that it restores the component to the as good as new status.
8. Whenever there is a renewal (repair or replacement), the inspection process starts again from time zero.
9. Inspectionsare benign in that the process of inspection will not in itself induce defects.
10. Inspections are either perfect or imperfect.
11. Here we consider a single component with one failure mode.

At the same time, the notation which will be used in this and subsequent Chapters are also defined here.

Now let t denote the inspection policy, $t = (t_0, t_1, t_2, \dots, t_n)$, where $t_0=0$. Figure 6-1. presents the relative positioning of inspections when the repair times are negligible.

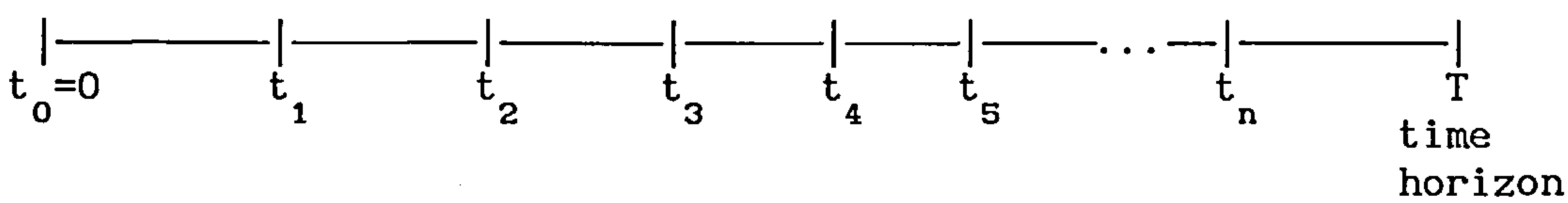


Figure 6-1. Inspection policy t

Again Let $p_b(t_i)$ denote the probability of a failure occurring in the inspection interval of (t_{i-1}, t_i).

$p_m(t_i)$ denote the probability of a defect being identified at an inspection time t_i.

$p_b(t)$ denote the probability that a defect will result in a failure under the inspection policy t.

$P_m(t)$ denote the probability that a defect will be identified at an inspection time under the inspection policy of t.

c_b, c_i, c_m denote the costs of a failure, an inspection without finding defect, and an inspection at which defect is found respectively. For convenience they are assumed to be constants.

And finally, let

d_b, d_i, d_m denote the times spent on a failure repair (or replacement), an inspection without finding defect, and an inspection at which defect is found respectively. Again these values are assumed to be constants.

Obviously, since the failure process is a typical stochastic process, the key functions to determine for subsequent inspection modelling are the probability measures defined above. Now, we first try to derive the key probability measures used in our models.

6.2 Key probability expressions

Since whether inspections are to be assumed perfect or not influences greatly the formulation of probability measures, it is both convenient and appropriate to start from the simplest case, namely perfect inspections with the downtime of inspections and breakdowns are assumed negligible.

6.2.1 Case of perfect inspections

The first probability measure derived is the probability of a failure occurring in an inspection interval, say, in (t_{i-1}, t_i) where time is measured from the last replacement or as new instant. If we assume that the inspection is perfect in that any defect present at an inspection will always be identified, then we must have $t_{i-1} < u < t_i$ and $h < t_i -$

u if there is a failure in (t_{i-1}, t_i). Since we have assumed that u and h are independent, then the probability of a defect arising in time interval $(u,u+du)$, $(t_{i-1} < u < t_i)$ and resulting in a failure before t_i is $g(u)F(t_i-u)du$, where $F(t_i-u)=\int_0^{t_i-u}f(h)dh$. Integrating over all possible u in (t_{i-1},t_i), we have

$$p_b(t_i) = \int_{t_{i-1}}^{t_i} g(u)F(t_i-u)\,du. \qquad 6\text{-}1$$

From the probability law of summation of all possible independent events, we have the probability of failure over the time interval (0,T) given the inspection policy **t** is given by,

$$p_b(\mathbf{t}) = \sum_{i=1}^{n} p_b(t_i) + \int_{t_n}^{T} g(u)F(t_n-u)du, \qquad 6\text{-}2$$

where n is the sequence number of the inspection which is just performed before the time horizon T. The last term in equation is due to the fact that if t_n is less than T, there is still a chance for a failure occurring in (t_n,T). But usually this term is very small if T is large. In the case an infinite time horizon, the expression for $p_b(\mathbf{t})$ is obtained taking the limit of equation 6-2 as $n \to \infty$.

Consider now the probability of an inspection repair where a defect is identified at an inspection, say, t_i, and then repaired. This will be derived with a similar way. Under the perfect inspection assumption, if a defect is identified at an inspection time point from new, say, t_i, the initial time interval $(u, u+du)$ of this defect must lie in (t_{i-1},t_i), and the delay time h should be longer than t_i-u. Since the probability of this event is $g(u)du\{1-F(t_i-u)\}$ under the assumption of independent u and h, we have, integrating over u from t_{i-1} to t_i,

$$p_m(t_i) = \int_{t_{i-1}}^{t_i} g(u)\{1-F(t_i-u)\}du. \qquad 6\text{-}3$$

Summing over the probabilities of these events, we have, the probability of an inspection repair is, under the inspection policy t.

$$p_m(t) = \sum_{i=1}^{n} p_m(t_i), \qquad 6\text{-}4$$

where as before, n is the index of the last inspection before the time horizon T of interest.

Finally, the probability of neither a failure nor an inspection repair arising before the finite time horizon T, denoted by $p_n(T)$, is

$$p_n(T) = 1 - p_b(t) - p_m(t). \qquad 6\text{-}5$$

Introducing equation 6-1 and 6-3 into equation 6-2 and 6-4, we have

$$\begin{aligned} p_b(t)+p_m(t) &= \sum_{i=1}^{n}\Big(\int_{t_{i-1}}^{t_i} g(u)F(t_i-u)du + \int_{t_{i-1}}^{t_i} g(u)\{1-F(t_i-u)\}du\Big) + \int_{t_n}^{T} g(u)F(t_n-u)du \\ &= \sum_{i=}^{n}\Big(\int_{t_{i-1}}^{t_i} g(u)du\Big) + \int_{t_n}^{T} g(u)F(t_n-u)du \\ &= \int_0^{t_n} g(u)du + \int_{t_n}^{T} g(u)F(t_n-u)du. \end{aligned}$$

Then finally we have

$$p_n(T) = 1 - \int_0^{t_n} g(u)du - \int_{t_n}^{T} g(u)F(T-u)du. \qquad 6\text{-}6$$

Now we turn to the case of imperfect inspection.

6.2.2 Case of imperfect inspections

If inspections are not perfect, there is a probability β that a defect present at an inspection will be identified. This means that some defects may pass the inspections without being discovered. In this case the above formulations of $p_b(t_i)$ and $p_m(t_i)$ and $p_n(T)$ will need to be modified whilst the form of $p_b(t)$ and $p_m(t)$ will be as before. For the sake of simplicity, we still assume that the downtime of inspections and breakdowns are negligible.

Consider the probability of a failure arising in the inspection interval (t_{i-1}, t_i). If a defect arises at u in the first inspection interval $(0, t_1)$, it will result in a failure in (t_{i-1}, t_i) provided the delay time is long enough, that is $t_{i-1}-u < h < t_i-u$, and at each intervening inspection t_j, $j=1,\ldots,i-1$, the defect is not observed. From the probability law of joint events, we have as before, the probability of this joint event would be

$$g(u)du(1-\beta)^{i-1} \{F(t_i-u)-F(t_{i-1}-u)\},$$

where $(1-\beta)$ has a power of $i-1$ implies there are $i-1$ inspections at which the defect is present yet not identified.

Integrating above probability expression over all u in $(0, t_1)$ we obtain the probability for a defect arising in $(0, t_1)$ and resulting in a failure in (t_{i-1}, t_i), namely

$$\int_0^{t_1} g(u)(1-\beta)^{i-1} \{F(t_i-u)-F(t_{i-1}-u)\}du.$$

This can be easily generalized into the case of a defect arising in an arbitrary inspection interval (t_{j-1}, t_j), $j=1,2,\ldots,i-1$, and resulting in a failure in (t_{i-1}, t_i). Figure 6-2 illustrates this situation.

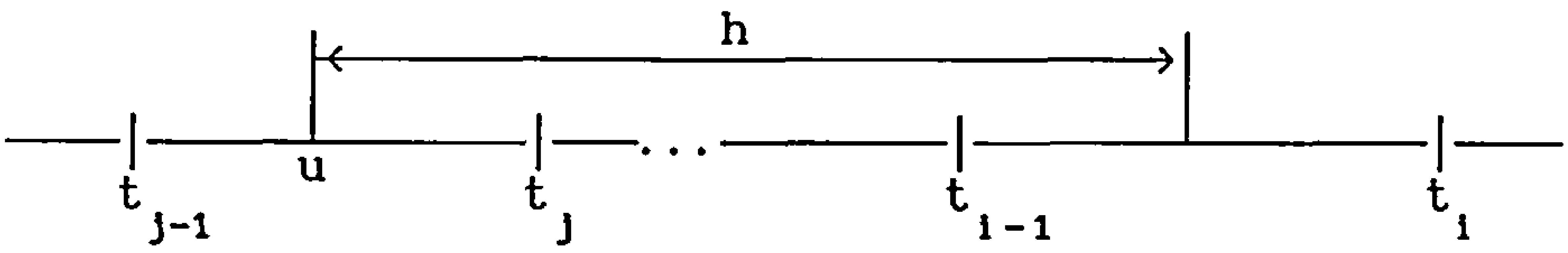


Figure 6-2 Imperfect inspection process

Consider the probability of a defect arising in (u,u+du) in (t_{j-1}, t_j) and resulting in a failure in (t_{i-1}, t_i), $i > j$. Since we know that the probability of this event is $g(u)du(1-\beta)^{i-j}\{F(t_i-u)-F(t_{i-1}-u)\}$. Integrating u over (t_{j-1}, t_j), we have for the probability of a defect arising in (t_{j-1}, t_j) and leading a failure in (t_{i-1}, t_i)

$$\int_{t_{j-1}}^{t_j} g(u)(1-\beta)^{i-j}\{F(t_i-u)-F(t_{i-1}-u)\}du.$$

In the case $u \subset (t_{i-1}, t_i)$, probability of a defect arising and resulting a failure in (t_{i-1}, t_i) is the same from equation 6-3, namely

$$\int_{t_{i-1}}^{t_i} g(u)\{1-F(t_i-u)\}du.$$

Summing over all the possible inspection intervals containing the initial point u, we finally have for the probability of a breakdown occurring in the ith inspection interval

$$P_b(t_i)=\sum_{j=1}^{i-1}\int_{t_{j-1}}^{t_j} g(u)(1-\beta)^{i-j}\{F(t_i-u)-F(t_{i-1}-u)\}du\Big]+\int_{t_{i-1}}^{t_i} g(u)F(t_i-u)du. \quad 6\text{-}7$$

For the probability of failure over the time horizon (0,T) given that the inspection policy is t, namely $p_b(t)$, in a similar way, it is given by

$$p_b(t) = \sum_{i=1}^{n+1} p_b(t_i), \qquad 6\text{-}8$$

where t_{n+1}=T and

$$P_b(T\ )=\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j} g(u)(1-\beta)^{i-j}\{F(T-u)-F(t_n-u)\}du\Big]+\int_{t_n}^{T} g(u)F(t_i-u)du.$$

We now formulate the probability of an inspection repair in which a defect is identified at an inspection, say, t_i, and then repaired. As before, consider a defect which arises in (t_{j-1},t_j), $j < i$, and is identified at inspection t_i. The following joint events arise, i.e. $t_{j-1}< u \le t_j$, $h \ge t_i-u$, and there is one successful inspection, and i-j unsuccessful inspections. We have, therefore, that the probability of a defect arising in (t_{j-1},t_j) and being identified at t_i is given by

$$\int_{t_{j-1}}^{t_j} g(u)\beta(1-\beta)^{i-j}\{1-F(t_i-u)\}du.$$

In the case where u arises in the last inspection interval, (t_{i-1},t_i), the probability of the defect being detected at inspection time t_i is

$$\int_{t_{i-1}}^{t_i} \beta g(u)\{1-F(t_i-u)\}du.$$

This is the same form as the above probability expression when j=i, and so the formulation may be extended to all interval $j \le i$. Summing over all $j=1,\dots,i$, we have for the probability of a defect being identified at inspection point t_i,

$$p_m(t_i) = \sum_{j=1}^{i} \int_{t_{j-1}}^{t_j} g(u)\beta(1-\beta)^{i-j}\{1-F(t_i-u)\}du. \qquad 6\text{-}9$$

The probability of having an inspection repair under the inspection policy t is the same structure as equation 6-4, but equation 6-9 should be used instead of equation 6-3 under the assumption of imperfect inspection.

The structure of the probability that no event arisen over T, $p_n(T)$, is similar to equation 6-5, but the component terms need modification because of the imperfect inspection influence. A defect may arise at any time before T and not be identified till T without causing a failure. In fact, introducing equation 6-8 and 6-9 into equation 6-5; we have

$$\begin{aligned} p_n(T) = 1 - \sum_{i=1}^{n} \Big[& \sum_{j=1}^{i-1} \int_{t_{j-1}}^{t_j} g(u)(1-\beta)^{i-j}\{F(t_i-u)-F(t_{i-1}-u)\}du \Big) + \\ & \int_{t_{i-1}}^{t_i} g(u)F(t_i-u)du + \sum_{j=1}^{i} \int_{t_{j-1}}^{t_j} g(u)\beta(1-\beta)^{i-j}\{1-F(t_i-u)\}du \Big) \Big] - \\ & \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} g(u)(1-\beta)^{n-j}\{F(T-u)-F(t_n-u)\}du + \int_{t_n}^{T} g(u)F(T-u)du. \end{aligned} \qquad 6\text{-}10$$

With these probabilities measures, cost or downtime modelling associated with an inspection process can be undertaken. We now discuss this modelling.

6.3 Models for an infinite time horizon

Infinite time horizon models are developed in two sub-sections since as before the formulations for perfect inspection are quite different from imperfect inspection case.

6.3.1 Perfect inspection

We first discuss the cost model

6.3.1.1 Cost model

Assuming perfect inspections, there are two events that could be associated with a defect, namely either a failure or an inspection repair. Suppose for the moment we are interested in maintenance costs only, and that the downtime of repairs and inspections may be neglected. We may further assume that the process is a renewal type in that either a failure repair or inspection repair restores the system as good as new, and that the inspection policy is restarted upon a renewal. The consequences would, of course, be different in the case where the inspection policy is continued. Now under the assumption of an infinite time horizon, one of the appropriate objective functions would be the asymptotic form of expected total cost per unit time, denoted by CT(t). That is

$$CT(t) = \frac{\text{Expected cost per repair cycle}}{\text{Expected repair cycle length}} . \qquad 6\text{-}11$$

Let C(t) denote the expected cost per repair cycle given the inspection policy is t.

Let T(t) denote the expected repair cycle length given the inspection policy is t.

We have, assuming concern is only with the maintenance costs,

$$C(t) = Ex(\text{ cost per repair cycle })$$
$$= Ex(\text{cost of a failure repair cycle} \mid \text{failure repair cycle}) \cdot p_b(t) + Ex(\text{cost of an inspection repair cycle} \mid \text{inspection repair cycle}) \cdot p_m(t)$$

Consider the first term on the right hand side. Since we know that

$$Ex\begin{bmatrix}\text{cost of failure} \\ \text{repair cycle}\end{bmatrix}\begin{vmatrix}\text{failure repair} \\ \text{cycle}\end{bmatrix} \cdot p_b(t) = \frac{Ex\begin{bmatrix}\text{cost of failure} \\ \text{repair cycle}\end{bmatrix}}{p_b(t)} \cdot p_b(t)$$

$$= Ex\begin{bmatrix}\text{cost of failure} \\ \text{repair cycle}\end{bmatrix},$$

and

$$Ex\begin{bmatrix}\text{cost of failure} \\ \text{repair cycle}\end{bmatrix} = \Sigma\left[\begin{pmatrix}\text{total cost ocurred in} \\ \text{a failure repair cycle}\end{pmatrix}\begin{pmatrix}\text{probability of} \\ \text{this failure}\end{pmatrix}\right],$$

where $\sum$ means summing all possible failure repair cycles.

It turns out that only unconditional expected values are necessary. The same argument can also be extended to the case of an inspection repair cycle.

For a failure repair cycle terminated at u+h in (t_{i-1}, t_i), the total cost up to and including the final failure cost is $(i-1)c_i+c_b$. Likewise, for an inspection cycle ended at t_i by an inspection repair, the total cost would be $(i-1)c_i+c_m$. Multiplying events by their correspond probabilities and summing over all $i=1,\ldots\infty$, we have for the expected cycle cost

$$C(t) = \sum_{i=1}^{\infty}\left[\{(i-1)c_i + c_b\}\cdot p_b(t_i) + \{(i-1)c_i + c_m\}\cdot p_m(t_i)\right]. \qquad 6\text{-}12$$

Introducing equations 6-1 and 6-3 into above equation, equation 6-12 becomes

$$C(t) = \sum_{i=1}^{\infty}\left[\{(i-1)c_i + c_m\}\int_{t_{i-1}}^{t_i} g(u)du + (c_b - c_m)\int_{t_{i-1}}^{t_i} g(u)F(t_i - u)du\right]. \qquad 6\text{-}13$$

Adopting the same principle, the expected repair cycle length would be given by

T(t) = Ex(repair cycle length)
= Ex(failure repair cycle length) + Ex(inspection repair cycle length).

Consider the first term on the right hand side first. If a failure occurs in (t_{i-1}, t_i), the contribution to the expected failure repair cycle length due to this failure is

$$\int_{t_{i-1}}^{t_i}\int_0^{t_i - u}(u+h)g(u)f(h)dhdu.$$

Since the contribution to the expected time to an inspection repair cycle for an inspection repair at t_i, would be $t_i p_m(t_i)$, we have, summing all the possible values of i, expected cycle length T(t) is

$$T(t) = \sum_{i=1}^{\infty}\left[\int_{t_{i-1}}^{t_i}\int_0^{t_i - u}(u+h)g(u)f(h)dhdu + t_i\cdot p_m(t_i)\right]. \qquad 6\text{-}14$$

Again, introducing equations 6-3 into equation 6-14, after some manipulation, equation 6-14 becomes

$$T(t) = \sum_{i=1}^{\infty} \left(t_i \int_{t_{i-1}}^{t_i} g(u)du - \int_{t_{i-1}}^{t_i} g(u)\bar{F}(t_i-u)du\right), \qquad 6\text{-}15$$

where $\bar{F}(\cdot) = \int_0^{\cdot} F(x)dx$.

The model objective is, of course, to minimize C(t)/T(t) with respect to t, that is

$$\underset{t}{\text{Min}} \left[\frac{C(t)}{T(t)}\right].$$

A similar expression can be constructed for downtime per unit time.

6.3.1.2 Downtime model

If now we consider the down time measure, we can not neglect the time spent on a failure repair or an inspection repair. Figure 6-3 shows the relationship among t_i, d_i, d_m, and d_b.

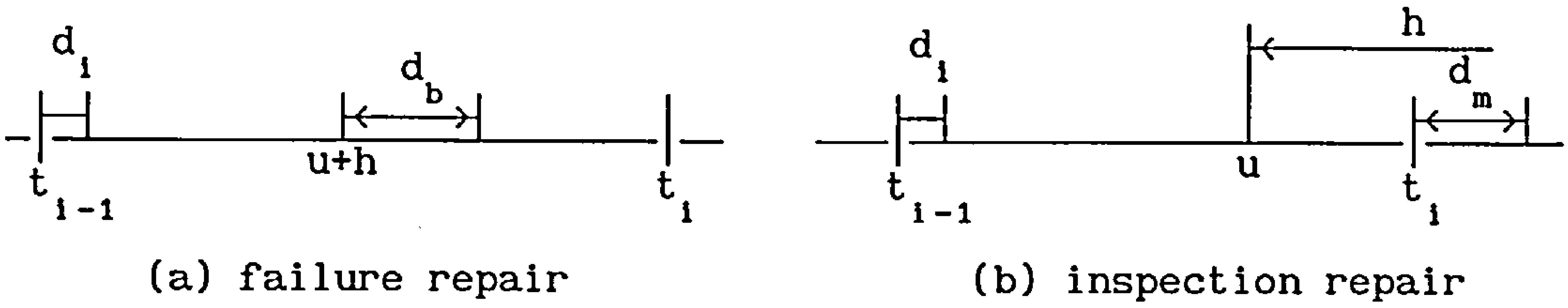


Figure 6-3 Failure and inspection times are not negligible

Let DT(t) denote the expected total downtime per unit time associated with failure repairs and inspection repairs given the inspection policy is t.

Let D(t) denote the expected downtime per repair cycle given the inspection policy is t.

We have the percentage availability measure of downtime, denoted by PA(t), under the inspection policy t is

$$PA(t) = DT(t)\cdot 100\%, \qquad 6\text{-}16$$

where $$DT(t)=\frac{D(t)}{T(t)}.$$

The form of T(t) will have changed since we now count the time spent on failure or inspection repairs. In a similar way to developing equation 6-12, D(t) can be written as

$$\begin{aligned} D(t) &= Ex(\text{downtime per repair cycle}) \\ &= Ex(\text{downtime of failure repair cycle}) + Ex(\text{downtime of inspection repair cycle}). \end{aligned}$$

If we assume that defects or failures do not occur at an inspection, we have

$$\begin{aligned} D(t) &= \sum_{i=1}^{\infty} \left[\{(i-1)d_i+d_b\}p_b(t_i) + \{(i-1)d_i+d_m\}p_m(t_i)\right] \\ &= \sum_{i=1}^{\infty} \left[\{(i-1)d_i+d_m\}\int_{a_i}^{t_i} g(u)du + (d_b-d_m)\int_{a_i}^{t_i} g(u)F(t_i-u)du\right], \qquad 6\text{-}17 \end{aligned}$$

where $a_i=t_{i-1}+d_i$ for i=2,3,...while i=1 $a_i=t_0$. This is because under the assumption that u can not arise during an inspection time d_i, the integration of u should start from $t_{i-1}+d_i$.

We have stated that equation 6-14 for T(t) needs a slight change here. From Figure 6-3, we can write down T(t) as

$$T(t)=\sum_{i=1}^{\infty}\Big(\int_{a_i}^{t}\int_{0}^{t_i-u}(u+h+d_b)g(u)f(h)dhdu+(t_i+d_m)\int_{a_i}^{t_i}g(u)\{1-F(t_i-u)\}du\Big), \quad 6\text{-}18$$

where a_i is as before.

This can be further simplified as

$$T(t) = \sum_{i=1}^{\infty}\Big((t_i+d_m)\int_{a_i}^{t_i}g(u)du+(d_b-d_m)\int_{a_i}^{t_i}g(u)F(t_i-u)du-\int_{a_i}^{t_i}g(u)\bar{F}(t_i-u)du\Big). \quad 6\text{-}19$$

Clearly, if we consider the times of failure repairs and inspections, equation 6-15 should be replaced by equation 6-19 when using equation 6-10 of the expected cost per unit time measure and the lower integration limit in equation 6-12 should be replaced by a_i.

As before, minimizing DT(t)=D(t)/T(t) in terms of t gives the optimal inspection policy t which makes the expected downtime per unit time to be the smallest.

6.3.2 Imperfect inspections

If inspections are not perfect and there is a probability β that a defect present at an inspection will be identified, the above formulation will need to be modified. However, it is known that functions of the perfect inspection case will clearly provide respectively lower and upper bounds for the imperfect inspection case.

This property could prove useful in dismissing a particular and ineffective inspection option without the need for the more complex modelling of the imperfect case. We do, however, now formulate the non perfect inspection case.

Consider first the expected cost per unit time measure. Since the format of equation 6-12 is still structurally correct for the imperfect inspection case, we simply introduce the non perfect inspection probabilities of equation 6-7 and 6-9 into equation 6-12 and obtain

$$C(t) = \sum_{i=1}^{\infty}\Big[\{(i-1)c_i + c_b\}\Big(\sum_{j=1}^{i}\int_{t_{j-1}}^{t_j} g(u)(1-\beta)^{i-j}\{F(t_i-u)-F(t_{i-1}-u)\}du\Big) + \{(i-1)c_i + c_m\}\Big(\sum_{j=1}^{i}\int_{t_{j-1}}^{t_j} g(u)\beta(1-\beta)^{i-j}\{1-F(t_i-u)\}du\Big)\Big], \qquad 6\text{-}20$$

where for j=i, we define $F(t_{i-1}-u)=0$ and if the inspection time is not negligible, t_{j-1} and t_{i-1} should be replaced by $t_{j-1}+d_i$ and $t_{i-1}+d_i$ respectively except j=1.

When $\beta = 1$, equation 6-20 reduces to equation 6-12 of the perfect inspection case as, indeed, it must. However, equation 6-20 can not be further simplified because the second term involves an extra β. This will pose no problem for calculating by a computer.

The expected length of per repair cycle, T(t), is as before, given by

T(t) = Ex(failure repair cycle) + Ex(inspection repair cycle).

If d_b, d_i, and d_m are not negligible, the first term, for a defect arising at time u in $(t_{j-1}+d_i, t_j)$ from new, it will arise as a breakdown at u+h in $(t_{i-1}+d_i, t_i)$ provided that it is not identified at t_k, k=j,...,i-1, and the delay time satisfies $t_{i-1}+d_i-u < h < t_i-u$.

Then the contribution to the expected failure cycle length is

$$\int_{t_{j-1}+d_i}^{t_j}\int_{t_{i-1}+d_i-u}^{t_i-u}(u+h+d_b)(1-\beta)^{i-j}g(u)f(h)dhdu.$$

Summing from j=1 to i, we have the total contribution to the expected cycle length when failure occurs at u+h in (t_{i-1}, t_i) is

$$\sum_{j=1}^{i}\int_{a_j}^{t_j}\int_{b_i}^{t_i-u}(u+h+d_b)(1-\beta)^{i-j}g(u)f(h)dhdu,$$

Where as before, we define $a_j=t_{j-1}+d_i$ (j=2,3,...i) and $b_i=t_{i-1}+d_i-u$ (i=2,3...). For j=1 and i, we set $a_j=t_0$ and $b_i=0$ respectively.

For the contribution of the expected cycle length when a defect is identified at t_i, we can easily write it down here as

$$\sum_{j=1}^{i}(t_i+d_m)\int_{a_j}^{t_j} g(u)\beta(1-\beta)^{i-j}\{1-F(t_i-u)\}du,$$

where a_j is defined as above.

Now summing all i from 1 to infinity, we have

$$T(t) = \sum_{i=1}^{\infty}\left[\sum_{j=1}^{i}\int_{a_j}^{t_j}\int_{b_i}^{t_i-u}(u+h+d_b)(1-\beta)^{i-j}g(u)f(h)dhdu + \sum_{j=1}^{i}(t_i+d_m)\int_{a_j}^{t_j} g(u)\beta(1-\beta)^{i-j}\{1-F(t_i-u)\}du\right]. \quad 6\text{-}21$$

In a similar way, replacing c_i, c_b, and c_m in equation 6-20 by d_i, d_b, and d_m, and let a_j and b_i be defined as before, we can write down D(t) as

$$D(t)=\sum_{i=1}^{\infty}\Big[\{(i-1)d_i+d_b\}\Big(\sum_{j=1}^{i}\int_{a_j}^{t_j}g(u)(1-\beta)^{i-j}\{F(t_i-u)-F(b_i)\}du+\Big)$$
$$\{(i-1)d_i+d_m\}\Big(\sum_{j=1}^{i}\int_{a_j}^{t_j}g(u)\beta(1-\beta)^{i-j}\{1-F(t_i-u)\}du\Big)\Big]. \qquad 6\text{-}22$$

6.4 Models for finite time horizons

In the last section we have discussed models of condition monitoring inspection over an infinite time horizon. For the long term expected cost per unit time measure, equation 6-10 gives the asymptotic form if the time horizon tends to infinity. One of the advantages of using equation 6-11 is that it has the simplest form and can be easily computed. However, we know that in practice the time horizon of use may be large, but it is always bounded. The solution obtained using this asymptotic criterion are , therefore, limiting approximations to reality of usually unknown error. In other words, we need to consider the more realistic case where the time horizon is finite. Now in this section, we discuss two cases which are related to the finite inspection time horizons. Starting from the simple one, we first consider models over a component life time.

6.4.1 Models over a component life time

In these models we assume that upon the detection of defect or the occurrence of failure the inspection process ends. There is an obligation to maintain the equipment cost effectively until it fails or a defect is identified at an inspection, at which point it will not be replaced. As an example, consider the problem of detecting the occurrence of an event (say, the presence of some grave illness such as cancer or the arrival of an enemy missile) when the time of occurrence is not known in advance. Each inspection, be it is an inspection or an inspection with a defect found, involves a cost so that we do not wish to check too often. On the other hand, there would be a cost associated with the occurrence of failure, which is higher than the inspection cost, so that we wish to check often enough to avoid a catastrophic failure. This kind of models is particularly useful in medical study since the checking process of a disease will usually stop upon the detection of it. Still under the general assumptions made in the last section, we hope to find an inspection policy which minimizes the expected cost or downtime measures during the component life time.

In fact, since what we want to minimize is the expected cost or downtime per cycle, therefore equations 6-13 and 6-17 for the perfect inspection case and equations 6-20 and 6-22 for the imperfect inspection case can be directly used here to establish the criteria functions. But, since it is assumed that there is no replacement at a cycle end, there would be no cost or downtime associated with an end cycle replacement. However, at this point one must expect some cost or downtime to be incurred because of the failure or the recognition of a defect. In this situation, in order to be consistent with the previous notation of costs and downtimes parameters, it is convenient to define c_m, c_b and d_m, d_b in general as follows.

Let c_m and d_m denote the cost and downtime associated with a defect if it is identified at an inspection.

c_b and d_b denote the cost and downtime associated with a failure.

Notation c_i and d_i have the same meaning as before, namely the cost and downtime respectively for an inspection. Clearly it is expected that $c_b > c_m > c_i$ and $d_b > d_m > d_i$.

For the sake of simplicity, if d_i, d_m and d_b are assumed to be negligible, we have from 6-13 and 6-20, the cost models are

$$C(t)=\begin{cases}\sum_{i=1}^{\infty}\Big[\{(i-1)c_i+c_m\}\}\int_{t_{i-1}}^{t_i}g(u)du + (c_b-c_m)\int_{t_{i-1}}^{t_i}g(u)F(t_i-u)du\Big]. & \text{perfect inspection}\\ \sum_{i=1}^{\infty}\Big[\{(i-1)c_i+c_b\}\Big(\sum_{j=1}^{i}\int_{t_{j-1}}^{t_j}g(u)(1-\beta)^{i-j}\{F(t_i-u)-F(t_{i-1}-u)\}du\Big) + \\ \quad \{(i-1)c_i+c_m\}\Big(\sum_{j=1}^{i}\int_{t_{j-1}}^{t_j}g(u)\beta(1-\beta)^{i-j}\{1-F(t_i-u)\}du\Big)\Big]. & \text{imperfect inspection}\end{cases}$$

6-23

If we count the times associated with inspections and failure, the downtime models, from equation 6-17 and 6-22 are given by

$$D(t)=\begin{cases}\sum_{i=1}^{\infty}\Big[\{(i-1)d_i+d_m\}\int_{a_i}^{t_i}g(u)du + (d_b-d_m)\int_{a_i}^{t_i}g(u)F(t_i-u)du, & \text{perfect inspection}\\ \sum_{i=1}^{\infty}\Big[\{(i-1)d_i+d_b\}\Big(\sum_{j=1}^{i}\int_{a_j}^{t_j}g(u)(1-\beta)^{i-j}\{F(t_i-u)-F(b_i)\}du +\Big) \\ \quad \{(i-1)d_i+d_m\}\Big(\sum_{j=1}^{i}\int_{a_j}^{t_j}g(u)\beta(1-\beta)^{i-j}\{1-F(t_i-u)\}du\Big)\Big], & \text{imperfect inspection}\end{cases}$$

6-24

where a_j and b_i are as before.

An optimal inspection policy is a specification of successive inspection times $t_1 < t_2 < t_3, \ldots$ for which C(t) or D(t) is minimized.

6.4.2 Models over finite time horizons

Now we consider a decision making problem of conditional monitoring inspection for a component over finite time horizons. In this section, the time horizon of interest, T, is finite. After a replacement(either a failure replacement or an inspection replacement) the inspection process resumes. In this sense, there may be several replacements over the time length (0,T). Figure 6-4 shows an example of this situation.

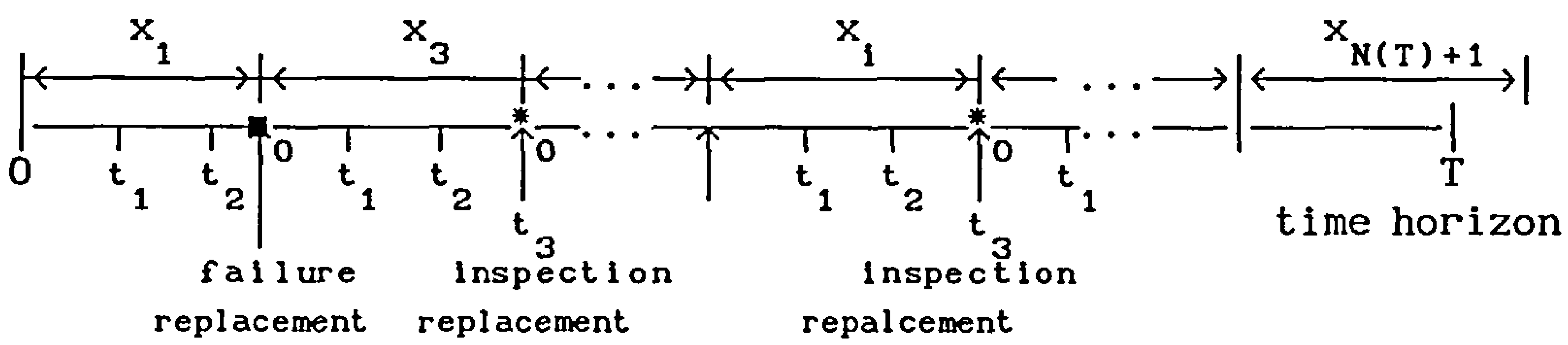


Figure 6-4 Inspection starts from time zero, upon failure or inspection replacement, the process re-starts, where x_i denotes the time between replacements(either a failure replacement or an inspection replacement), t_i denotes the time to the ith inspection from last renewal, ■ denotes the failure replacement, and * denotes the inspection replacement.

Clearly from Figure 6-4, the modelling of this kind of problem is not an easy job because T is finite. But since the process repeats and the inspection policy is assumed not to change over time T, the time between replacements follows a renewal process with identical and independently distributed cycle. Renewal theory in conjunction with delay time modelling can be used here to formulate models of the inspection process.

For the sake of simplicity, we only consider here the case under the assumption of perfect inspection, and are content to point out the same method, but a more complicated formulation, can be generalized into the case of imperfect inspection.

First we introduce the notation which will be used in this section, which is consistent with that commonly used in renewal theory.

Let N(T) denote the total number of replacements in (0,T).
Let H(T) denote the expected number of replacements occurring in (0,T).

H(T) is usually called a renewal function.

let C(T) denote the total cost occurring in (0,T).

Let μ_x and σ_x^2 denote the mean and variance of the inter-arrival times x between replacements.

Let μ_y denote the mean of costs occurring in a replacement interval.

Since a renewal corresponds to either a failure replacement or an inspection replacement which could occur in one of the inspection intervals, it follows from Figure 6-4 that the jth inter-renewal time is

$$x_j = \begin{cases} u+h, & t_{i-1} < u < t_i,\ h < t_i - u \\ t_i, & t_{i-1} < u \le t_i,\ h \ge t_i - u. \end{cases} \qquad i=1,2,\ldots$$

Obviously x_j is a combination of two types of random variables, ie. continuous over (t_{i-1}, t_i) and discrete at t_i (i=1,2,...). The cumulative distribution function for continuous part is

$$Q(x)=p(X\le x \mid t_{i-1}<x<t_i) = \sum_{j=1}^{i-1}\left[p(t_{j-1}<X<t_j)+p(X=t_j)\right] + p(t_{i-1}<x<t_i).$$

Since we know that $p(t_{j-1}<X<t_j)=p_b(t_j)$ and $p(X=t_j)=p_m(t_j)$, from equation 6-1 and 6-3 we have

$$\sum_{j=1}^{i-1}\left[p(t_{j-1}<X<t_j)+p(X=t_j)\right]=\int_0^{t_{i-1}} g(u)du,$$

and

$$p(t_{i-1}<x<t_i) = \int_{t_{i-1}}^{t_i} g(u)F(t_i-u)du.$$

Finally we have

$$Q(x)=p(X \le x)=\begin{cases} \int_0^x g(u)F(x-u)du, & 0 < x < t_1 \\ \int_0^{t_1} g(u)du + \int_{t_1}^x g(u)F(x-u)du, & t_1 < x < t_2 \\ \vdots & \\ \int_0^{t_{i-1}} g(u)du + \int_{t_{i-1}}^x g(u)F(x-u)du, & t_{i-1} < x < t_i \\ \vdots & \end{cases}$$

And the discrete part, the probability of having an inspection replacement at t_i is

$$p(x=t_i) = p_m(t_i) = \int_{t_{i-1}}^{t_i} g(u)\{1-F(t_i-u)\}du. \qquad i=1,2,...$$

From renewal theory, if we assume that the first replacement occurs at x, then from well known equation that $Ex\{N(T)|x_1=x\} = 1 + H(T-x)$ if $x \le T$, it follows that

$$H(T) = Ex\{N(T)\} = \int_0^T Ex\{N(T)|x_1=x\}dQ(x)=\int_0^T \{1+H(T-x)\}dQ(x).$$

i.e. $$H(T) = Q(T) + \int_0^T H(T-x)dQ(x). \qquad 6\text{-}26$$

Since we know that the first renewal must be either a failure replacement or an inspection replacement, from the expression of Q(x), we know that

$$dQ(x) = \begin{cases} d\int_0^x g(u)F(x-u)du, & 0 < x < t_1 \\ d\int_{t_1}^x g(u)F(x-u)du, & t_1 < x < t_2 \\ \vdots & \\ d\int_{t_{i-1}}^x g(u)F(x-u)du, & t_{i-1} < x < t_i \\ \vdots & \end{cases}$$

If we assume that $x_i \in (t_{i-1}, t_i]$, i=1,2,...n+1, where n=sup{n such that $t_n < T$}, $t_0=0$, from equation 6-26, we have

$$\begin{aligned} H(T)=&\int_0^{t_1}\{1+H(T-x)\}d\int_0^x g(u)F(x-u)du+\{1+H(T-t_1)\}p_m(t_1) \\ &+\int_{t_1}^{t_2}\{1+H(T-x)\}d\int_{t_1}^x g(u)F(x-u)du+\{1+H(T-t_2)\}p_m(t_2) \\ &+\ldots \\ &+\int_{t_{i-1}}^{t_i}\{1+H(T-x)\}d\int_{t_{i-1}}^x g(u)F(x-u)du+\{1+H(T-t_i)\}p_m(t_i) \\ &+\ldots \\ &+\int_{t_n}^{T}\{1+H(T-x)\}d\int_{t_n}^x g(u)F(x-u)du. \end{aligned} \qquad 6\text{-}27$$

Hence since $dQ(x) = d\left[\int_{\bullet}^x g(u)F(x-u)du\right] = \int_{\bullet}^x g(u)f(x-u)dudx$, then changing the integration sequence and letting x=u+h, after some manipulation, we finally have

$$H(T)=G(T)+\sum_{i=1}^{n}\left[\int_{t_{i-1}}^{t_i}\int_{0}^{t_i-u}H(T-u-h)g(u)f(h)dhdu+H(T-t_i)\int_{t_{i-1}}^{t_i}g(u)\{1-F(t_i-u)\}du\right],$$
$$+\int_{t_n}^{T}\int_{t_n}^{T-u}H(T-u-h)g(u)f(h)dhdu, \qquad 6\text{-}28$$

where $G(T)=\int_0^T g(u)du$.

Equation 6-27 is equivalent to equation 6-25 which is termed 'the integral equation of renewal theory' and is based upon what is known as the 'renewal argument'. This basically means that the probabilistic structure of the process begins anew after the moment of the first renewal, x_1.

The renewal function H(T) is fundamental in renewal theory since it forms the basis of renewal reward processes which is of interest in our modelling. However, as can be seen from equation 6-27, it is not generally possible to derive an exact solution analytically since it is a double integral equation. However, according to the renewal theory and the structure of equation 6-27, since it satisfies the renewal assumption of identical and independently distributed inter-renewal times, it is possible to derive an asymptotic solution to H(T).

We now discuss this asymptotic solution.

Suppose now that Z(T) is some expectation related to the renewal process, and that $a(\cdot)$ is a known non-negative function. The integral equation

$$Z(T) = a(T) + \int_0^T Z(T-v)dQ(v). \qquad 6\text{-}29$$

is called the 'generalized renewal equation'and its solution is given by the following

$$Z(T) = a(T) + \int_0^T a(T-v)dH(v). \qquad 6\text{-}30$$

Suppose now that the asymptotic form of Z(T) is of interest. If a is directly Riemann integrable and the inter-renewal times x_1 are non-lattice, then Smith's [1958] key renewal theorem gives

$$\lim_{T \to \infty} Z(T) = \frac{1}{\mu_x} \int_0^\infty a(v)dv, \qquad 6\text{-}31$$

where $\mu_x = Ex(x_i)$, the mean inter-renewal time.

Now let $Z(T)=H(T)-T/\mu_x$, using equation 6-26 and 6-29, after some manipulation and let $T \to \infty$, we have the well known formula

$$H(T) = \frac{T}{\mu_x} + \frac{\sigma_x^2 - \mu_x^2}{2\mu_x^2} + o(1). \qquad 6\text{-}32$$

For reference see Tijms [1986].

The above discussion can now be easily extended to renewal-reward process.

Suppose now that a cost y_i is occurred in the ith renewal interval, this will comprise the inspection costs that accumulate during the cycle together with the replacement cost of a new component at the end of the cycle (either failure replacement or inspection replacement). We assume that $y_1, y_2, \ldots$ to be independent and identically distributed non-negative random variables with mean $\mu_y < \infty$. This is true in our case considered here. y_i may depend on x_i, but we suppose that the pairs (x_i, y_i), i=1,2,... are independent and identically distributed. In fact, this sequence defines a renewal reward process.

The total cost of the inspection policy up to time T, y(T) is then given by

$$y(T) = \sum_{i=1}^{N(T)} y_i + \varepsilon(T)$$

where N(T) is the number of renewals occurring in (0,T], and $\varepsilon(T)$ is the inspection cost that accumulates in $(x_{N(T)}, T)$. Since we know that N(T)+1 is a stopping time for the sequence $(x_1, x_2, \ldots)$, it follows that N(T)+1 is also a stopping time for the dependent sequence $(y_1, y_2, \ldots)$. We have therefore from Wald's equation that

$$\begin{aligned} C(T) &= Ex\Big(\sum_{i=1}^{N(T)+1} y_i\Big) - Ex(y_{N(T)+1}) + Ex\{\varepsilon(T)\} \\ &= \{1+H(T)\}\mu_y - Ex(y_{N(T)+1}) + Ex\{\varepsilon(T)\}. \qquad 6\text{-}33 \end{aligned}$$

It is known that C(T) can also be defined by an integral equation, Christer [1978], Christer [1987], Christer and Jack [1991], and Jack [1991], which raises the possibility of utilizing the above asymptotic results and generalized integral equation solution to the renewal-reward process. Conditioning on the time of the first renewal it follows that, if v is the time to the first replacement,

$$Ex\{y(T) \mid x_1 = v\} = \begin{cases} \alpha(v) + C(T-v) & v \le T \\ \gamma(T) & v > T \end{cases}$$

where $\alpha(v) = Ex\{y(v) \mid x_1 = v\}$ and $\gamma(T) = Ex\{y(T) \mid x_1 > T\}$. Hence

$$C(T) = \int_0^\infty Ex\{y(T)|x_1=v\}dQ(v) = \int_0^T \{\alpha(v)+C(T-v)\}dQ(v) + \int_T^\infty \gamma(T)dQ(v),$$

i.e.

$$C(T) = M(T) + \int_0^T C(T-v)dQ(v), \qquad 6\text{-}34$$

where $M(T)=\int_0^T \alpha(v)dQ(v)+\gamma(T)\{1-Q(T)$.

Knowing the solution, equation 6-30, of the generalized renewal equation 6-28, we have the solution of the cost equation 6-34 as

$$C(T) = M(T) + \int_0^T M(T-v)dH(v).$$

The alternative form for the above equation which will be used later can be derived by integration by parts and using the fact that M(0)=H(0)=0, namely

$$C(T) = \int_0^T \{1+H(T-v)\}dM(v). \qquad 6\text{-}35$$

Let $Z(T)=C(T)-(\mu_y/\mu_x)T$, then using equation 6-34 and the key renewal theorem 6-31, after some manipulation, we have

$$C(T) = \frac{\mu_y}{\mu_x} T + \frac{\sigma_x^2+\mu_x^2}{2\mu_x^2}\mu_y - \lim_{T\to\infty} Ex(y_{N(T)+1}) + \lim_{T\to\infty} Ex\{\varepsilon(T)\}, \qquad 6\text{-}35$$

where $\lim_{T\to\infty} Ex(y_{N(T)+1})=\frac{1}{\mu_x}\int_0^\infty \alpha(v)vdQ(v)$,

and $\lim_{T\to\infty} Ex\{\varepsilon(T)\}=\frac{1}{\mu_x}\int_0^\infty \gamma(v)\{1-Q(v)\}dv$.

See Heyman and Sobel [1982], and Jack [1991] for detail.

From equation 6-36 we have, the expected cost per unit time measure

$$\frac{C(T)}{T} = \frac{\mu_y}{\mu_x} + \left[\frac{\sigma_x^2+\mu_x^2}{2\mu_x^2}\mu_y - \lim_{T\to\infty} Ex(y_{N(T)+1}) + \lim_{T\to\infty} Ex\{\varepsilon(T)\}\right]/T. \qquad 6\text{-}37$$

Clearly as $T \to \infty$, equation 6-37 becomes μ_y/μ_x for which we have used in the previous sections.

Christer [1978] used an alternative approach to obtain result similar to equation 6-37 under the assumption that the renewal cost occurs at the end of the renewal cycle. Under such assumption the term lim $Ex\{\varepsilon(T)\}$ is no longer needed. Christer substituted the asymptotic form of H(T), result 6-32, into equation 6-33 to give, as $T \to \infty$

$$C(T) = \frac{\mu_y}{\mu_x} T + \frac{\sigma_x^2 + \mu_x^2}{2\mu_x^2}\mu_y - \lim_{T\to\infty} Ex\{y_{N(T)+1}\}. \qquad 6\text{-}38$$

Christer [1978] refers to result 6-38 without term $\lim Ex\{y_{N(T)+1}\}$ as the 'refined' asymptotic form for C(T), with μ_y/μ_x being the corresponding 'crude' asymptotic form. However, even if we use equation 6-37 as a solution of renewal reward processes, it is also a limiting result because we use the property $T \to \infty$. If T is not sufficient large, there must be an unknown error involved by using equation 6-37. The exact solution should be obtained by solving equation 6-34 or 6-35 of the integral equation form without using the key renewal theory. But, comparing the cost of complicated calculation of equation 6-34 or 6-35 with the relative small error (see following numerical example) by using simple form of equation 6-37, it is clear that equation 6-37 is preferable. For the present we leave the discussion on the exact solution to equation 6-34 or 6-35 to a future study. To be consistent with the term 'refined' asymptotic form, equation 6-37 is called 'accurate' asymptotic form.

Now turning to our problem, since equation 6-28 is equivalent to equation 6-26 in principle we can, therefore, in a similar way, also establish the expression for C(T) of our problem using the integral equation method as follows.

Knowing that

$$\alpha(v) = Ex[y(v)|x_1=v\}=\begin{cases} \begin{cases} c_b, & 0 < v < t_1 \\ c_m, & v = t_1 \end{cases} \\ \begin{cases} c_i+c_b, & t_1< v < t_2 \\ c_i+c_m, & v = t_2 \end{cases} \\ \vdots & \vdots \\ \begin{cases} (i-1)c_i+c_b, & t_{i-1}< v < t_i \\ (i-1)c_i+c_m, & v = t_i \end{cases} \\ \vdots \\ nc_i+c_b. & t_n < v \leq T \end{cases} \qquad 6\text{-}39$$

and also from the expression of Q(x), we have

$$\begin{aligned}\int_0^T \alpha(v)dQ(v) &= c_b\int_0^{t_1}\int_0^v g(u)f(v-u)dudv + c_m\int_0^{t_1} g(u)\{1-F(t_1-u)du \\ &+ (ic_i+c_b)\int_{t_1}^{t_2}\int_{t_1}^v g(u)f(v-u)dudv + (ic_i+c_m)\int_{t_1}^{t_2} g(u)\{1-F(t_2-u)\} \\ &+ \ldots \\ &+ \{(i-1)c_i+c_b\}\int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^v g(u)f(v-u)dudv+\{(i-1)c_i+c_m\}\int_{t_{i-1}}^{t_i} g(u)\{1-F(t_i-u)\} \\ &+ \ldots \\ &+ (nc_i+c_b)\int_{t_n}^T\int_{t_n}^v g(u)f(v-u)dudv.\end{aligned}$$

Inter-change the integration sequence and let h=v-u. Since we know that $\gamma(T)=Ex(\ y(T)|x_1)=nc_i$, and

$$1 - Q(T) = 1 - G(T) - \int_{t_n}^{T} g(u)\{1-F(T-u)\}du,$$

then from the expression of M(T) (see equation 6-34), after some manipulation, we obtain

$$M(T) = \sum_{i=1}^{n}\left[\{(i-1)c_i+c_m\}\int_{t_{i-1}}^{t_i} g(u)du + (c_b-c_m)\int_{t_{i-1}}^{t_i} g(u)F(t_i-u)du\right] + (nc_i+c_b)\int_{t_n}^{T} g(u)F(T-u)du + nc_i\{1-G(T)+\int_{t_n}^{T} g(u)\{1-F(T-u)\}du\}. \quad 6\text{-}40$$

Then from equation 6-35, we have the cumulative cost expression as

$$C(T) = M(T) + \int_0^T H(T-v)dM(v). \quad 6\text{-}41$$

Similar to the expression for M(T), M(v) is defined as $\int_0^v \alpha(v)dQ(v)$ since $\gamma(v)=0$ for $v \le T$. Then from the expression of $\alpha(v)$, we write down dM(v) as

$$dM(v) = \begin{cases} c_b\int_0^v g(u)f(v-u)dudv, & 0 < v < t_1 \\ (c_i+c_b)\int_{t_1}^v g(u)f(v-u)dudv, & t_1 < v < t_2 \\ \vdots & \vdots \\ \{(i-1)c_i+c_b\}\int_{t_{i-1}}^v g(u)f(v-u)dudv, & t_{i-1} < v < t_i \\ \vdots & \vdots \\ (nc_i+c_b)\int_{t_n}^v g(u)f(v-u)dudv. & t_n < v < T \end{cases}$$

When $v=t_i$, $i=1,2,\ldots,n$, since $Q(v)$ is discrete at these points, but we know probabilities of event $v=t_i$, then similar to $Q(v)$, the contribution to $dM(v)$ at these discrete points are

$$\{(i-1)c_i+c_m\}\int_{t_{i-1}}^{t_i} g(u)\{1-F(t_i-u)\}du. \qquad i=1,2,\ldots,n$$

Then inter-change the integration sequence and let $v=u+h$, we have

$$\begin{aligned} C(T) = M(T) + \sum_{i=1}^{n}\Bigg[& \int_{t_{i-1}}^{t_i} H(T-u-h)\{(i-1)c_i+c_b\}\int_{t_{i-1}}^{t_i} g(u)f(h)dhdu + \\ & H(T-t_i)\{(i-1)c_i+c_m\}\int_{t_{i-1}}^{t_i} g(u)\{1-F(t_i-u)\}du\Bigg] + \\ & \int_{t_n}^{T} H(T-u-h)(nc_i+c_b)\int_{t_n}^{T} g(u)f(h)dhdu. \end{aligned} \qquad 6\text{-}42$$

Equation 6-42 is equivalent to equation 6-35 in principle and can be solved in a similar way.

We have that $\mu_x=T(t)$, $\mu_y=C(t)$, and

$$\begin{aligned} \sigma_x^2 &= \int_0^{\infty}(x-\mu_x)^2 dQ(x) \\ &= \sum_{i=1}^{\infty}\Bigg[\int_{t_{i-1}}^{t_i}(u+h)^2\int_0^{t_i-u} g(u)f(h)dhdu) + t_i^2(1-\int_{t_{i-1}}^{t_i} g(u)[1-F(t_i-u)]du\Bigg]-\mu_x^2 \end{aligned}$$

Therefore from equation 6-36, we can calculate the value of $C(T)/T$, if we can derive the formula for

$$\lim_{T \to \infty} Ex(y_{N(T)+1}) = \frac{1}{\mu_x} \int_0^{\infty} \alpha(v) v dQ(v), \qquad 6\text{-}43$$

and

$$\lim_{T \to \infty} Ex\{\varepsilon(T)\} = \frac{1}{\mu_x} \int_0^{\infty} \gamma(v)\{1-Q(v)\}dv. \qquad 6\text{-}44$$

In our case, using equation 6-39, equation 6-43 becomes

$$\lim Ex\{y_{N(T)+1}\} = \frac{1}{\mu_x} \sum_{i=1}^{\infty} \left[\{(i-1)c_i + c_m\} \int_{t_{i-1}}^{t_i} \int_0^{t_i - u} (u+h) g(u) f(h) dh du \right.$$
$$\left. + \{(i-1)c_i + c_m\} t_i \int_{t_{i-1}}^{t_i} g(u)\{1-F(t_i - u)\} du \right]. \qquad 6\text{-}45$$

Since

$$\gamma(v) = Ex\{y(v) \mid x_1 > v\} = \begin{cases} 0, & 0 \le v < t_1 \\ c_i, & t_1 \le v < t_2 \\ i \cdot c_i, & t_2 \le v < t_3 \\ \vdots & \\ (i-1) \cdot c_i, & t_{i-1} \le v < t_i \\ \vdots & \end{cases}$$

and because we know that

$$1 - Q(v) = \int_v^{\infty} g(u) du + \int_{t_j}^{v} g(u)\{1-F(v-u)\} du,$$

where t_j is the inspection time point just before v, equation 6-44 becomes

$$\lim \mathrm{Ex}\{\varepsilon(T)\}=\frac{1}{\mu_x}\sum_{i=1}^{\infty}\left[\int_{t_{i-1}}^{t_i}(i-1)c_i\{\int_v^{\infty}g(u)du+\int_{t_{i-1}}^{v}g(u)[1-F(v-u)]du\}dv\right]. \quad 6\text{-}46$$

This completes our formulations, namely, the 'accurate' asymptotic formulation, equation 6-37, the 'refined' approximation, equation 6-38 without the last term, and the 'crude' approximation, equation 6-11.

To confirm our formulations we now give a numerical example to illustrate the method.

Let the initial time u be an exponentially distributed variable with scale factor α=0.5822, and let the delay time distribution be exponential with scale factor λ=0.7633, namely $g(u)=\alpha e^{\alpha u}$ and $f(h)=\lambda e^{\lambda h}$. The cost values are c_i=15, c_b=200, and c_m=150 units. Suppose further the inspection policy is regular with inspection interval Δt=2.0. For time horizons T=10, 15, 20, 25, 30, 35, and 40, the results of the renewal reward function of equation 6-37 is shown in Table 6-1.

'accurate' asymptotic	time horizon T: 10	15	20	25	30	35	40	∞
C(T)/T	55.63	56.20	56.48	56.66	56.77	56.85	56.92	57.35

Table 6-1 Results of the 'accurate' asymptotic model over finite time horizons.

The expected total cost per unit time over infinite time horizon in our example is 57.345 which clearly shows the difference between the cost measures over finite and infinite time horizons. As T increases, Table 6-1 shows that C(T)/T increase as well as it must be.

As a matter of interest, the expected number of renewals in time horizon T using equation 6-32 is also illustrated in Table 6-2.

	time horizon T						
	10	15	20	25	30	35	40
H(T)	3.46	5.51	7.55	9.59	11.64	13.68	15.72

Table 6-2 Expected number of renewals in time horizons

It is possible to obtain the exact solution to C(T) if we can solve integral equation 6-35. Numerical analysis can provide us a tool to handle this problem within the required accuracy. However, as can be seen from the above expressions, the formulations are very complicated since they involve the double integration. As the time horizons in our example are not very long, we want to know how good the approximations are, that is, whether there is significant difference between the true value and the asymptotic one. To do this, we ran a thousand simulations on each time horizon with the parameters of our example, Since the sample size is big enough (1000), we feel this provides a good estimate of the true value of C(T). The comparison between the results of our 'accurate' asymptotic formulation and simulation is illustrated in Figure 6-5, which shows the agreement between these two results and the trend that as T increases, these two results are tending to be equal and approaching μ_y/μ_x=57.34.

This result is quite encouraging because it shows that even in the case where T is relatively short, we can still use the asymptotic approach to obtain the approximated solution to our problem without losing much accuracy. Note also that in this example the difference between the results of finite and infinite time horizon is not very large, which illustrates the possibility of using μ_y/μ_x as an objective function instead of C(T)/T because the former can be more ready calculated and optimized.

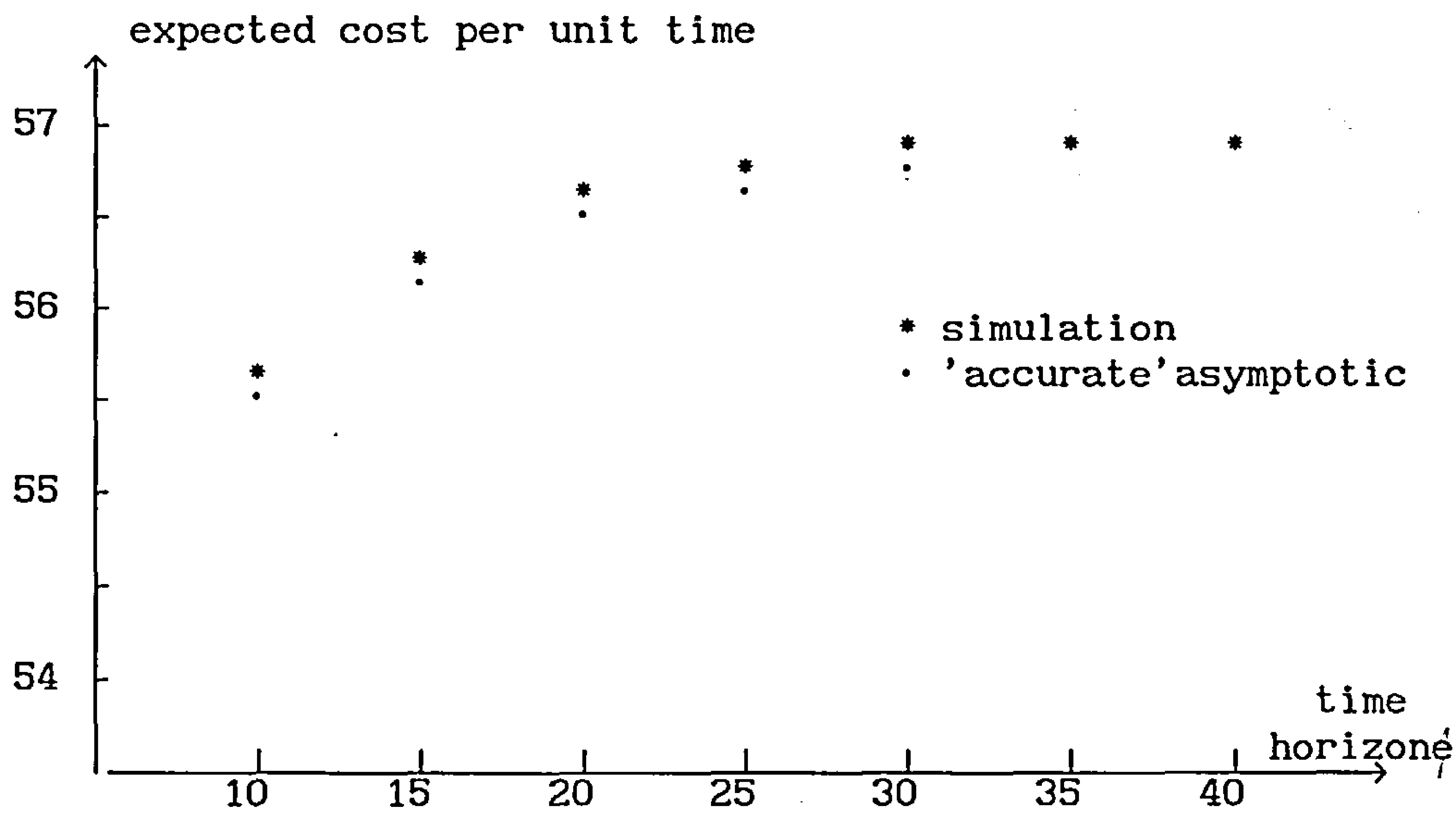


Figure 6-5 Comparison of the results between simulation and 'accurate'asymptotic solution

6.5 Conclusions

Delay time analysis has already proved useful in the rudimentary applications made so far. Its scope for development has still to be really explored. The purpose of this Chapter was not so much to develop specific models of condition monitoring inspection as to show such activities could be modelled using the concept of delay time and to both discuss and highlight some of the issues and modelling options involved. Measuring the condition of plant by some process has, intuitively, a connection with estimating the delay time of a defect. But there is also a difference. Whereas the delay time measure entails consideration of defect prognosis and, therefore, can be and is used in analysis to model consequences of maintenance actions, the current state of much of condition monitoring is embryonic in that it is crudely of a (0,1) nature. A monitoring test will often indicate whether or not a defect exists, but yields little additional information of a prognostic type. Once an abnormality is identified at a monitoring check, corrective action is often taken as though a breakdown were imminent. This is essentially the situation captured in models of this Chapter, where no attempt has been made to exploit the unexpired delay time that may be available to seek more cost effective repair schedules and perhaps, exploit opportunistic events.

Hopefully, however, the industrial situation will improve with technological advancements in monitoring techniques to the extent that monitoring test results will be allied with a quantified prognosis, and perhaps the monitoring policy t will become consequential on the results of previous inspections, that is the timing of t_i and the nature of test at t_i will depend upon the results of the previous test up to t_{i-1}. Such a dynamic monitoring regime could perhaps be arrived at through modelling where the delay time distribution for a defect is conditional upon the previous condition measures at t_j, $j < i$. Proportional hazard modelling (PHM), Cox [1972], could be a tool to model this situation. An initial effort has been made to model the condition monitoring inspection by using PHM and delay time modelling, which is essentially based upon the idea of using the historic data of monitoring checks. Since much work needs to be done in this modelling and many ideas need to be tested and developed, for the time being this topic is not included in this thesis. However, clearly this is an area worth exploring.

As has been stated, models developed in this Chapter are based upon the prototype model proposed by Christer [1992]. Extensions are made here to discuss more cases of condition monitoring inspection modelling. It is shown in the models over finite time horizons that delay time modelling can be used to model a renewal reward process provided that after each replacement the process resumes. Also, in the models over finite time horizons, for the sake of simplicity, only models for perfect inspection case are discussed. However, the same argument can be generalized to the case of imperfect inspection at the price of more complicated formulation. We leave the modelling of condition monitoring inspection over finite time horizons for the imperfect case to a future study.

Finally, the modelling of inspection efficiency has been through the probability β that a defect present at inspection will be identified. This simplification is appropriate to the prototype modelling being discussed here. Another form of error is the probability ρ that a non-defect will be identified as a defect. This is just as real a problem as the β type of error. A related 'false alarm' problem is that of non-defects reported as defects to the maintenance staff. When collecting the data of infusion pumps in Hope Hospital at Salford, it is found that in some records equipment is reported faulty, but actually no defect was found. User errors could be the main reason for this kind of false alarm, but it involves the cost of inspections. Clearly the ρ type errors, β type errors as well as false alarms exist and require study within the context of monitoring models.

CHAPTER 7. A SPECIAL CASE OF CONDITION MONITORING INSPECTION MODELLING

7.1 Preliminary

This chapter is concerned with modelling a practical decision problem relating to an area of growing significance to production engineers, namely condition monitoring of production plant. The situation may be presented in its simplest form as follows: there is a critical component such as a bearing, say, used in production machines for which a condition test is capable of indicating the extent of wear. If the wear is less than a certain amount "H", the bearing is functioning satisfactorily with no immediate risk of failure. If, however, the wear is greater than "H", the level is regarded as critical and a replacement is initiated as soon as possible. Should the lining reach the zero level, a costly failure occurs.

In order to avoid the inconvenience and the costly consequence of a breakdown of the production plant, a condition monitoring inspection check is used to test the wear level at regular intervals. If the wear is below the "H" level, no action is taken, but if it is above the "H" level, the bearing is replaced immediately before a breakdown occurs. It is possible such an inspection replacement could be undertaken during a non production time.

In this situation, the test result is effectively of a (0,1) type in that it signals either all is ok or that a failure is imminent. The main decision problem relates to how and when to schedule condition checks. If tests are very frequent, there is an increased chance of detecting a defect before it leads to a failure. However, there is a proportional increase in the cost of condition inspections and, if the plant is required to be stopped for condition testing, an increase in lost production time. The problem is to select an economic condition monitoring inspection schedule to balance the cost and downtime to be expected due to breakdowns and inspections.

Here we will follow the prototype modelling suggested by Christer [1992], which is discussed in the last Chapter, and develop a delay time based inspection model for production plant with a particular type of wear pattern. Wear can not be assumed to be either a uniform or a determinate process and it is the need to establish an appropriate wear model in any particular case that causes the real problem in modelling and scheduling condition-monitoring inspection. However, we consider here the case when the pattern of wear varies linearly with time, but where the constant of proportion is a random variable, see Figure 7-1. This particular pattern of wear was suggested by B. Gits of Eindhoven University as a wear pattern appropriate to some production equipment in a steel plant.

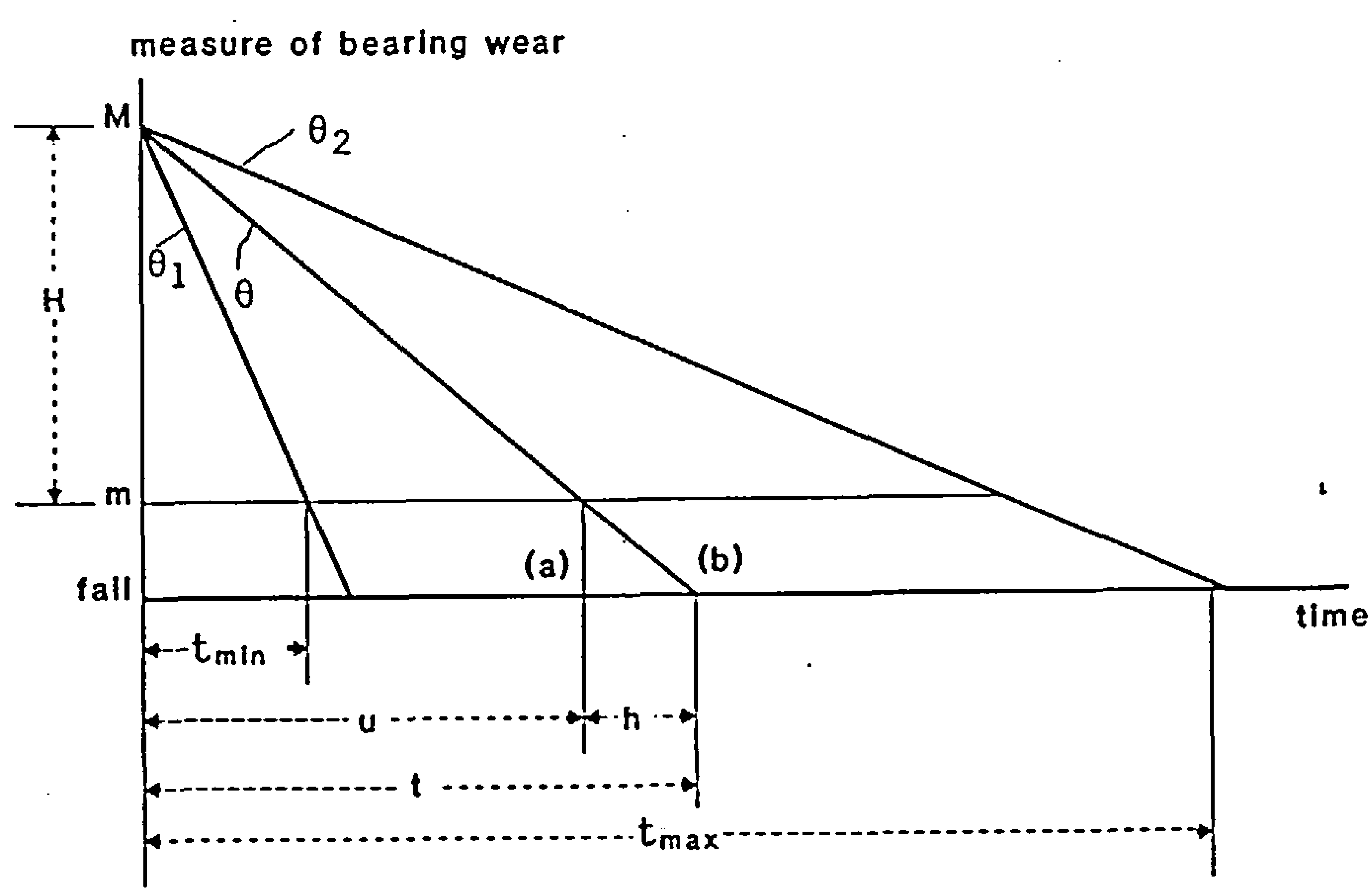


Figure 7-1 Presentation of the wear mechanism

To fit ideas, we assume that the initial or "as new" thickness of lining in a bearing, say, is M. When the lining is known to have degraded to m or below, a replacement would be made. Clearly, $H = M-m$ is the tolerance of the wear. The angle of wear θ, Figure 7-1, is a random variable assumed bounded between limiting angles θ_1 and θ_2 so that the unknown random variable θ satisfies $\theta_1 \le \theta \le \theta_2$.

Assuming for now the level "m" has been set in the condition monitoring device which essentially records (0,1) as the output. Although simplistic in form, this modelling of a condition monitoring device embraces a number of forms of monitoring including looking for a trace element in oil, or a special frequency in vibration analysis, which may or may not be present. It is seen from Figure 7-1 that the first opportunity to detect a pending fault given wear angle θ is at point (a) after time u when the remaining thickness of the bearing lining is m. If the defect is not detected at point (a) or subsequently, it will lead to a failure after time t at point (b) where the lining thickness reaches zero.

In the context of delay time analysis, Christer and Waller [1984], Christer [1992], see also Chapter 3 and 6, u represents the initial point of the defect, that is when it first becomes detectable. The delay time of this defect from its initial point to the failure time if left unattended is h = t-u, where the failure occurs after time t at point (b) when the lining thickness of the bearing reaches zero, Figure 7-1.

If t_1 denotes time to the first inspection from new, a minimum and maximum time range of t_1 exist for usefully implementing the condition monitoring inspection, namely

$$t_{min} = (M-m)\tan(\theta_1) \text{ and } \quad t_{max} = M \cdot \tan(\theta_2).$$

If $t_1 < t_{min}$, no defect will be identified and the monitoring is pointless. However, if $t_1 \ge t_{max}$, and if a failure hasn't already occurred, the monitoring check will identify the component wear to be greater than the tolerant level H, i.e. the remaining lining thickness is below m and the check will always result in a replacement. Under this condition the test is again not necessary to the decision process, that is $t_{min} \le t_1 < t_{max}$.

We are assuming that the angle of bearing wearing θ is a random variable which has a known distribution with the lower bound θ_1 and upper bound θ_2, and our task is to construct a model of the condition monitoring process where the decision variables are the number and frequency of inspection checks and the critical wear level m. Consequence variables are the operating costs, probability of failure and available production time.

7.2 The basic model

7.2.1 Assumptions and notation

Since the model developed here is different from the general models described in Chapter 6, it is convenient to introduce some special assumptions related to this particular model. The assumptions of the model are:

1. Bearing wear is linear with time, when at time $t \le t_{max}$, the wear level is given by $t \cdot \cot(\theta)$.
2. The angle of wear θ is uniformly distributed over (θ_1, θ_2).
3. n condition monitoring inspections are scheduled on a regular cyclic basis with the initial inspection taking place at a time greater than or equal to t_{min}, $n \ge 1$.
4. Inspections are assumed to be perfect and benign in that the results of condition monitoring are true and accurate and inspections do not induce defects.
5. The monitoring check produces binary information on the bearing condition, i.e.(0) or (1). (0) means the bearing wearing is less than H and all is assumed well, while (1) means bearing wear is greater than H and a replacement is urgently required.
6. Once the wear is recognized as being greater than H, a bearing

replacement is undertaken immediately as a preventive replacement.

In addition to costs and time parameters, there are certain probabilistic events which are important elements in developing a condition monitoring model and require specific definition. It is convenient to introduce them here along with other terms.

1. Let c_b, c_i and c_m denote the average cost of failure replacement, condition monitoring inspection and inspection replacement respectively as before.
2. Let d_b, d_i, and d_m denote the average time duration of failure replacement, condition monitoring inspection and inspection replacement. Clearly, we expect $c_b > c_m > c_i$ and $d_b > d_m > d_i$.
3. Let CT(n) denote the asymptotic total expected maintenance cost per unit time when the number of planned inspections is n.
4. Let AT(n) denote the asymptotic expected percentage availability of the bearing per unit time when the number of planned inspections is n.
5. Let C(n) denote the expected total cost arising over a bearing's life time when the number of planed inspections is n. We assume here it consists of the contributions from the cost of breakdown replacement, inspection replacement and condition inspection monitoring. With slight modification, downtime cost can also be included within C(n).
6. Let T(n) and A(n) denote the expected total time and available production time to the first completed bearing replacement of a new bearing under the current condition based maintenance policy with n planned inspections.
7. Let t_n denote the length of the regular monitoring inspection interval when the number of planned inspections is n, that is $t_n = (t_{max} - t_{min})/(n+1)$.
8. Let $t_{i,n}$ denote the time to the ith planned monitoring inspection from new when the number of planned inspections is n, that is $t_{i,n} = t_{min} + i \cdot t_n$, $1 \le i \le n$. It will be both convenient and consistent to define $t_{n+1,n} = t_{max}$.

9. Let $p_b(i,n)$ be the probability of a failure occurring during the ith monitoring inspection interval when the number of planned inspections is n.
10. Let $p_m(i,n)$ be the probability of a defect being identified at the ith monitoring inspection time point when the number of planned inspections is n.
11. Let $p_b(n)$ denote the probability of a failure occurring during the bearing life time when the number of planned inspections is n.
12. Let $p_m(n)$ denote the probability of a defect being identified at any one of the monitoring inspection time points when the number of planned inspections is n.

Clearly, we have

$$p_m(n) = \sum_{i=1}^{n} p_m(i,n),$$

and

$$p_b(n) = 1 - p_m(n).$$

7.2.2 Model criteria

Assuming as criteria the asymptotic cost and availability formulation per unit time measured over numerous bearing replacement cycles, as discussed in the last Chapter, we have

$$CT(n) = \frac{C(n)}{T(n)}, \qquad 7\text{-}1$$

and

$$AT(n) = \frac{A(n)}{T(n)} \; (100\ {}^{o}/o). \qquad 7\text{-}2$$

Now, C(n) = (expected contribution to total cost from cycle ending in a breakdown replacement) + (expected contribution to total cost from cycle ending in an inspection replacement),

that is,

$$C(n) = \sum_{i=1}^{n+1} \{ (i-1)\cdot c_i + c_b \}\cdot p_b(i,n) + \sum_{i=1}^{n} \{ (i-1)\cdot c_i + c_m \}\cdot p_m(i,n). \quad 7\text{-}3$$

Again, T(n) = (expected contribution to total time from cycle ending in a breakdown replacement) + (expected contribution to total time from cycle ending in an inspection replacement),

that is $$T(n) = \sum_{i=1}^{n+1} (T_{i,n}+d_b)\cdot p_b(i,n) + \sum_{i=1}^{n} (t_{i,n}+d_m)\cdot p_m(i,n), \quad 7\text{-}4$$

where $T_{i,n}$ denotes time to the breakdown which occurs in the ith inspection interval when the total number of planned inspection is n.

And finally,

A(n) = (expected contribution to total available production time from cycle ending in a breakdown replacement) + (expected contribution to total available production time from cycle ending in an inspection replacement),

that is,

$$A(n) = \sum_{i=1}^{n+1} (T_{i,n}-(i-1)\cdot d_i)\cdot p_b(i,n) + \sum_{i=1}^{n} (t_{i,n}-(i-1)\cdot d_i)\cdot p_m(i,n). \quad 7\text{-}5$$

These expectation expressions in conjunction with criterion 7-1 and 7-2 establishes the basic format of models which may be optimized with respect to the decision variable n and m. It remains, however, to establish expressions for the probabilities in equations 7-3 — 7-5.

7.2.3 Expressions for the probabilities

Adopting the concept of delay-time, see Chapters 3 and 6 and also Christer and Waller [1984], we have for a breakdown to occur during the ith monitoring inspection interval given n perfect inspections in total, the initial point u must lie in $(t_{i-1,n}, t_{i,n})$, and the delay time satisfy $h \subset (0, t_{i,n}-u)$, as shown in Figure 7-2 (a).

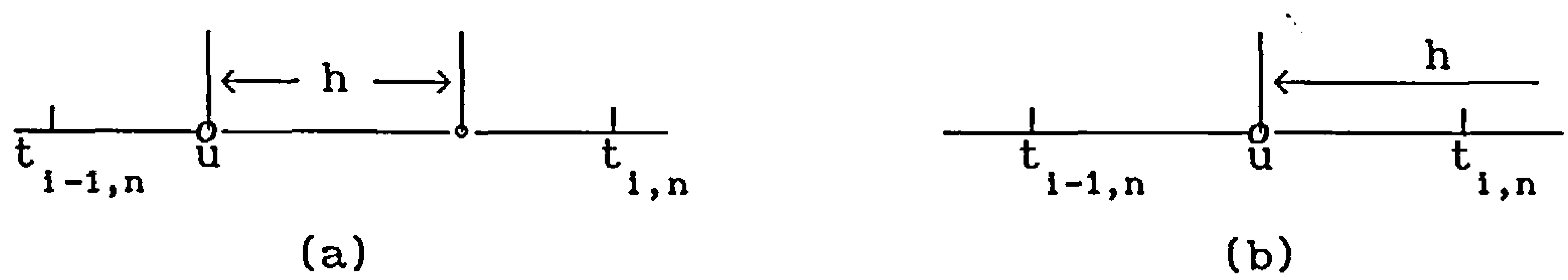


Figure 7-2 The delay time concept

Then under the assumption of perfect inspections, we have

$$p_b(i,n) = p\{\ u \subset (t_{i-1,n}, t_{i,n}),\ h \le (t_{i,n}-u)\ \}. \qquad 7\text{-}6$$

Similarly, for a defect to be identified at the ith monitoring inspection time point given n inspections in total, we must have the initial point $u \subset (t_{i-1,n}, t_{i,n})$ and the delay time satisfy $h \subset (t_{i,n}-u, \infty)$, see Figure 7-2 (b). That is,

$$p_m(i,n) = p\{\ u \subset (t_{i-1,n}, t_{i,n}),\ h > (t_{i,n}-u)\ \}. \qquad 7\text{-}7$$

From Figure 7-1 and assumption 1, it is clear that u and h are actually related here and satisfy the relationship

$$u = (M - m)\cdot \tan(\theta) \quad \text{and} \quad h = m \cdot \tan(\theta),$$

and therefore,

$$h = \frac{u \cdot m}{M - m}. \qquad 7\text{-}8$$

Introducing the relationship 7-8 into expressions 7-6 and 7-7, it follows that

$$p_b(i,n) = p_r\{ t_{i-1,n} < u < (\frac{M-m}{M})\cdot t_{i,n} \}, \qquad 7\text{-}9$$

and

$$p_m(i,n) = p_r\{ (\frac{M-m}{M})\cdot t_{i,n} \le u \le t_{i,n} \}. \qquad 7\text{-}10$$

Since we have defined g(u) as the pdf. of u, then expressions 7-9 and 7-10 become

$$p_b(i,n) = \int_{t_{i-1,n}}^{(\frac{M-m}{M})t_{i,n}} g(u)\ du, \qquad 7\text{-}11$$

and

$$p_m(i,n) = \int_{(\frac{M-m}{M})t_{i,n}}^{t_{i,n}} g(u)\ du, \qquad 7\text{-}12$$

respectively.

Clearly $t_{i-1,n}$ should be less than or equal to $\frac{M-m}{M} \cdot t_{i,n}$, that is the lower limit of integration in equation 7-11 should be less than the upper limit, otherwise slight modifications would be required. Since the situation can arise where this condition is not satisfied, this case will be considered in more detail later.

Our task here is to find the appropriate form of g(u) for the current problem. Let G(u) denote the probability function of u. From assumption (1), which establishes u as functions of θ ($\theta_1 \le \theta \le \theta_2$), we have, using Figure 7-1,

$$G(u) = P\{ (M-m)\cdot\tan(\theta) \le u \}$$
$$= P\{ \theta \le \tan^{-1}(u/(M-m) \}. \quad 7\text{-}13$$

Because we have assumed θ to be uniformly distributed over (θ_1,θ_2), it follows from equation 7-13 that,

$$G(u) = \{ \tan^{-1}(u/(M-m) - \theta_1 \}/(\theta_2 - \theta_1). \quad 7\text{-}14$$

Returning to the probability expressions 7-11 and 7-12, we now have

$$p_b(i,n) = G\{ t_{i,n}\cdot(\tfrac{M-m}{M}) \} - G\{ t_{i-1,n} \}$$

$$= \{ \tan^{-1}(t_{i,n}/M) - \tan^{-1}(t_{i-1,n}/(M-m) \}/(\theta_2 - \theta_1), \quad 7\text{-}15$$

and

$$p_m(i,n) = G\{ t_{i,n}\} - G\{ t_{i,n}\cdot(\tfrac{M-m}{M}) \}$$

$$= \{ \tan^{-1}(t_{i,n}/(M-m)) - \tan^{-1}(t_{i,n}/M) \}/(\theta_2 - \theta_1). \quad 7\text{-}16$$

The significance of the angles in these two probability expressions, which are illustrated in Figure 7-3, in that they represent bounding angles on θ for a defect to be detected at the ith monitoring inspection time point, or for a failure to occur in the ith monitoring inspection interval.

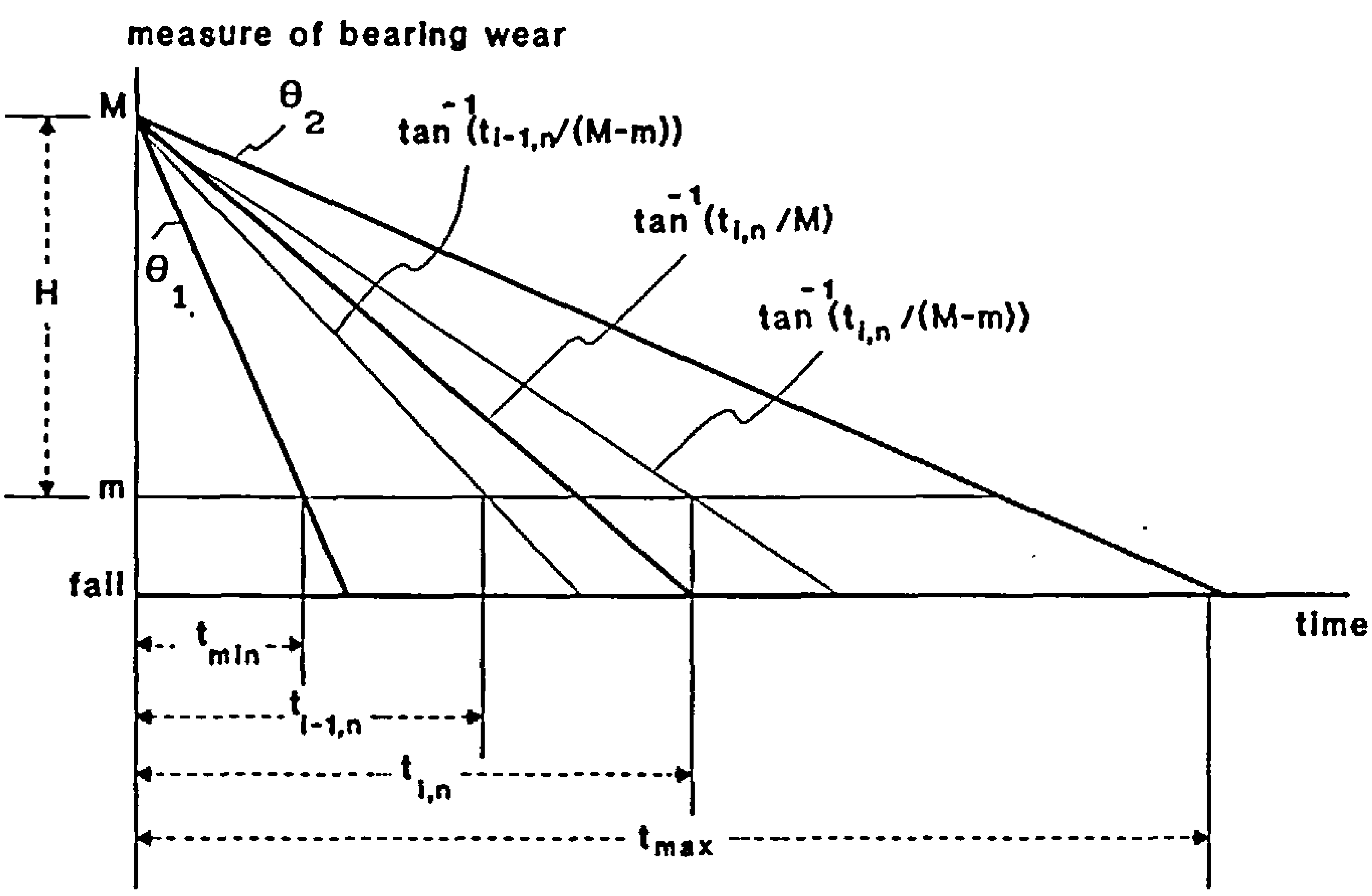


Figure 7-3 Relationships between θ, $p_b(i,n)$ and $p_m(i,n)$

As already indicated, equations 7-15 and 7-16 are key probability expressions in developing a condition monitoring inspection model. However, before we proceed further to develop this model, there are a number of pragmatic modelling considerations to raise associated with bounds on parameters if inspections are to be intelligently scheduled. We now formally list these for clarification and convenience.

1. The smallest delay time in our problem is $m \cdot \tan(\theta_1)$. When the regular inspection interval is less than $m \cdot \tan(\theta_1)$, a detect will always be detected when the wear angle is θ, $\theta_1 \le \theta \le \theta_2$. Consequently there is nothing to be gained by inspecting on a period less than $m \cdot \tan(\theta_1)$, and the maximum number of monitoring

inspections n to consider for this problem is bounded above by

$$INT(\frac{t_{max}-t_{min}}{m\cdot \tan(\theta\ )}),$$

where INT(x) denotes the smallest integer value which is greater or equal to x. Beyond this number of inspections, the monitoring effort would be wasted on unnecessary inspections.

2. It can be seen from Figure 7-4 that if $t_{i-1,n}/(M-m) \geq t_{i,n}/M$ and the wear angle $\theta > \tan^{-1}(t_{i-1,n}/(M-m))$, then if the remaining part of the bearing lining reaches the critical level m after the (i-1)st monitoring check, it must be identified at the ith check point, that is no failure will occur before the ith check. For this case, therefore, $p_b(i,n) = 0$. At the same time, the limits of integration of equation 7-12 for $p_m(i,n)$ become $t_{i-1,n}$ and $t_{i,n}$, see Figure 7-4.

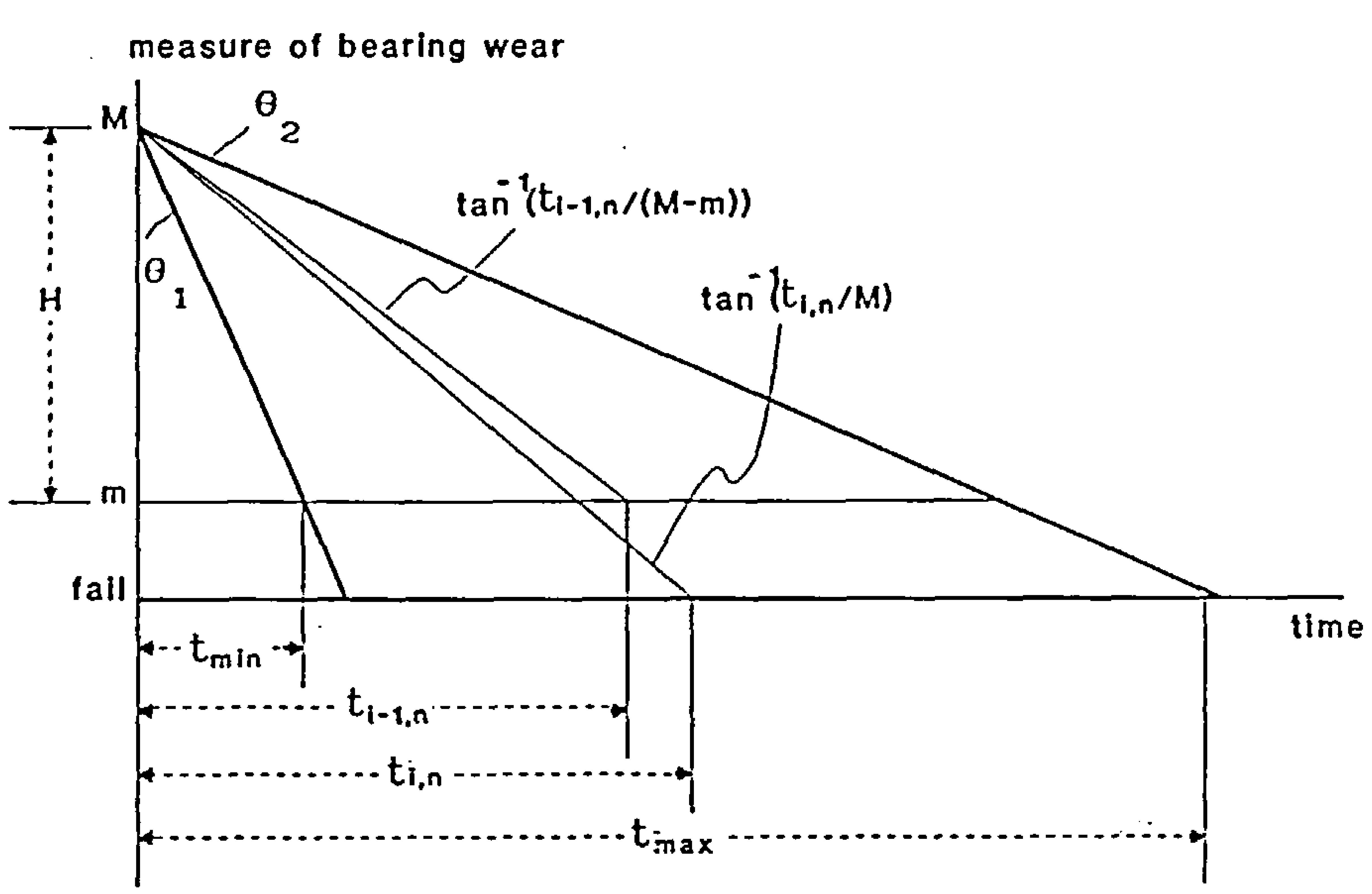


Figure 7-4 Situation if $t_{i-1,n}/(M-m) \geq t_{i,n}/M$

Here the appropriate expression for equation 7-16 becomes

$$p_m(i.n) = \{ \tan^{-1}(t_{i,n}/(M-m)) - \tan^{-1}(t_{i-1,n}/(M-m)) \}/(\theta_2 - \theta_1). \quad 7\text{-}17$$

3. If a monitoring check takes place at time $(M-m)\cdot\tan(\theta_2)$, see Figure 7-5, then no subsequent failure can arise. In other words, if $t_{i,n} \geq (M-m)\cdot\tan(\theta_2)$, we must have $p_b(j,n) = p_m(j,n) = 0$ for all $j > i$. In this case, the integration limits on equation 7-12 become $t_{i,n}\cdot(M-m)/M$ and θ_2 and equation 7-16 becomes

$$p_m(i,n) = \{ \tan^{-1}(\theta_2) - \tan^{-1}(t_{i,n}/M) \}/(\theta_2 - \theta_1). \quad 7\text{-}18$$

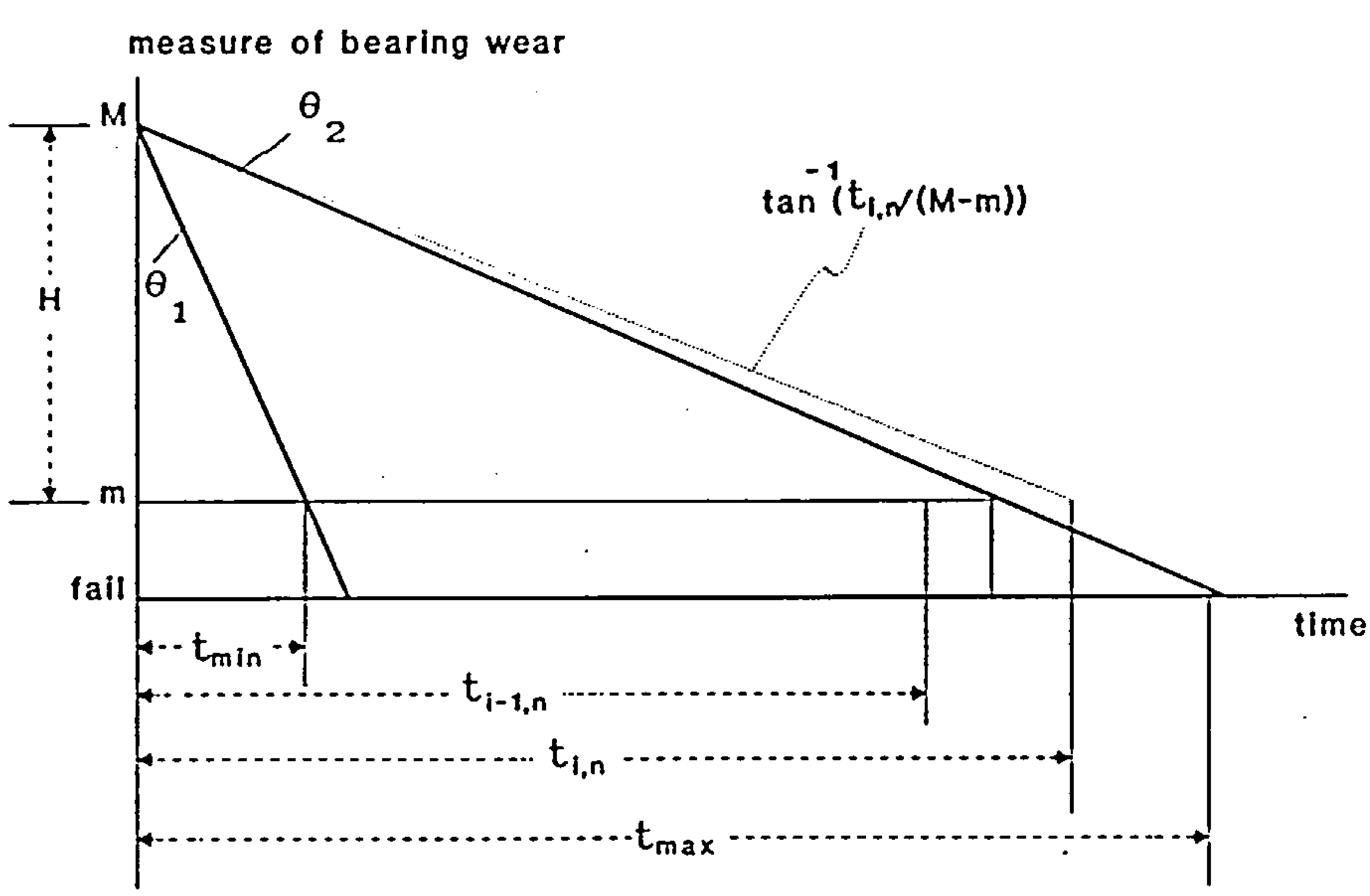


Figure 7-5 Situation if $t_{i,n} > (M-m)\cdot\tan(\theta)$

4. Finally, if the bearing survives into the time interval $(t_{n,n}, t_{max})$, then a failure must occur in this interval. Let $p_b(n+1,n)$ denote the probability of this event, then we have

$$p_b(n+1,n) = \{ \theta_2 - \tan^{-1}(t_{n,n}/(M-m)) \}/(\theta_2 - \theta_1). \qquad 7\text{-}19$$

This completes the formulation of the appropriate parts of equation 7-3 for the expected cycle cost.

We now establish from equations 7-4 and 7-5 the corresponding expressions for the expected cycle time to replacement, and the availability over a cycle. It is noted that the time to a breakdown in the ith inspection interval, $T_{i,n}$, can be represented by u+h. Using this result in conjunction with equation 7-11, equations 7-4 and 7-5 can be rewritten as

$$T(n) = \sum_{i=1}^{n+1} \int_{t_{i-1,n}}^{\left(\frac{M-m}{M}\right) t_{i,n}} (u+h+d_b) g(u) du + \sum_{i=1}^{n} (t_{i,n} + d_m) \cdot p_m(i,n), \qquad 7\text{-}20$$

and

$$A(n) = \sum_{i=1}^{n+1} \int_{t_{i-1,n}}^{\left(\frac{M-m}{M}\right) t_{i,n}} (u+h-(i-1) \cdot d_i) g(u) du + \sum_{i=1}^{n+1} (t_{i,n} - (i-1) \cdot d_i) \cdot p_m(i,n). \qquad 7\text{-}21$$

Differentiating expression 7-14 of the probability function G(u) to obtain the density function g(u), and using relationship 7-8 between u and h, expressions 7-20 and 7-21 for T(n) and A(n) become, after some re-arrangement and integration,

$$T(n) = \sum_{i=1}^{n+1}\left[\frac{M}{2(\theta_2-\theta_1)}\log\left(\frac{(M-m)^2 + (1-m/M)^2 t_{i,n}^2}{(M-m)^2 + t_{i-1,n}^2}\right) + d_b\cdot p_b(i,n)\right]$$

$$+ \sum_{i=1}^{n} (t_{i,n}+d_m)\cdot p_m(i,n), \qquad 7\text{-}22$$

and

$$A(n) = \sum_{i=1}^{n+1}\left[\frac{M}{2(\theta_2-\theta_1)}\log\left(\frac{(M-m)^2 + (1-m/M)^2 t_{i,n}^2}{(M-m)^2 + t_{i-1,n}^2}\right) + (i-1)d_i\cdot p_b(i,n)\right]$$

$$+ \sum_{i=1}^{n} (t_{i,n} - (i-1)d_i)\cdot p_m(i,n), \qquad 7\text{-}23$$

respectively.

Expressions for $p_b(i,n)$ and $p_m(i,n)$ are given by equations 7-15 and 7-16 along with special cases 7-17——7-19. We still need to pay attention to special situations, namely:

1. Since equations 7-22 and 7-23 are derived from equation 7-4, 7-20, 7-5 and 7-21 and are summed over each inspection interval, if in the kth inspection interval $p_b(k,n) = 0$, then, since the expected value is obtained through the correspondent probability measures , the corresponding part in the equations 7-22 and 7-23 should be zero as well.

2. If both $p_b(k,n)$ and $p_m(k,n) = 0$, then for the same reason as above, we should set the relevant part in equation 7-22 and 7-23 which involves $p_b(k,n)$ and $p_m(k,n)$ to zero before the evaluation of equations 7-22 and 7-23.

7.3 Numerical example

Assume $M = 20$, $m = 2$,

$\theta_1 = \pi/8$, $\theta_2 = \pi/3$,

$c_b = 350$, $c_i = 15$, $c_m = 150$ cost units,

$d_b = 0.60$, $d_i = 0.014$, $d_m = 0.42$ time units,

With the above data, the numerical results of the model when n = 1 and 2 are shown in Table 7-1 and 7-2. These results are intuitively reasonable in that they show that as the number of inspection checks increases from 1 to 2, the probability of a failure decreases, the maintenance cost decreases, and both the expected cycle length, T(n), and available production time , A(n), over a cycle decrease.

The expected total maintenance cost per unit time, the expected percentage availability and the probability of failure when n=1,2,...,32 are shown in Figure 7-6, 7-7 and 7-8, (max(n)=32 in this example). These figures display the expected tradeoffs between increasing the number of inspections and reducing the probability of failure, or decreasing the availability of plant, or increasing the maintenance cost. If it were possible to conduct monitoring checks during downtime periods, there would be greater plant availability. The interpretation of these figures for decision making would be considerably influenced by the value of plant operating time and considerations associated with risk if the plant were not available and risk should the plant breakdown.

number of inspection n	C(n)	T(n)	A(n)	$p_b(n)$	$p_m(n)$	CT(n)	AT(n)
1	338.21	19.25	18.66	0.920	0.080	17.56	0.969
2	327.70	19.15	18.57	0.884	0.156	17.11	0.970

Table 7-1 Result of the numerical example when n=1 and 2

number of inspection n	sequence of inspection i	$p_b(i,n)$	$p_m(i,n)$	probability of failure in the last interval $p_b(n+1,n)$
1	1	0.6390	0.0080	0.3530
2	1	0.4547	0.0796	0.1372
	2	0.2516	0.0769	

Table 7-2 Values of $p_b(i,n)$ and $p_m(i,n)$ when n=1 and 2

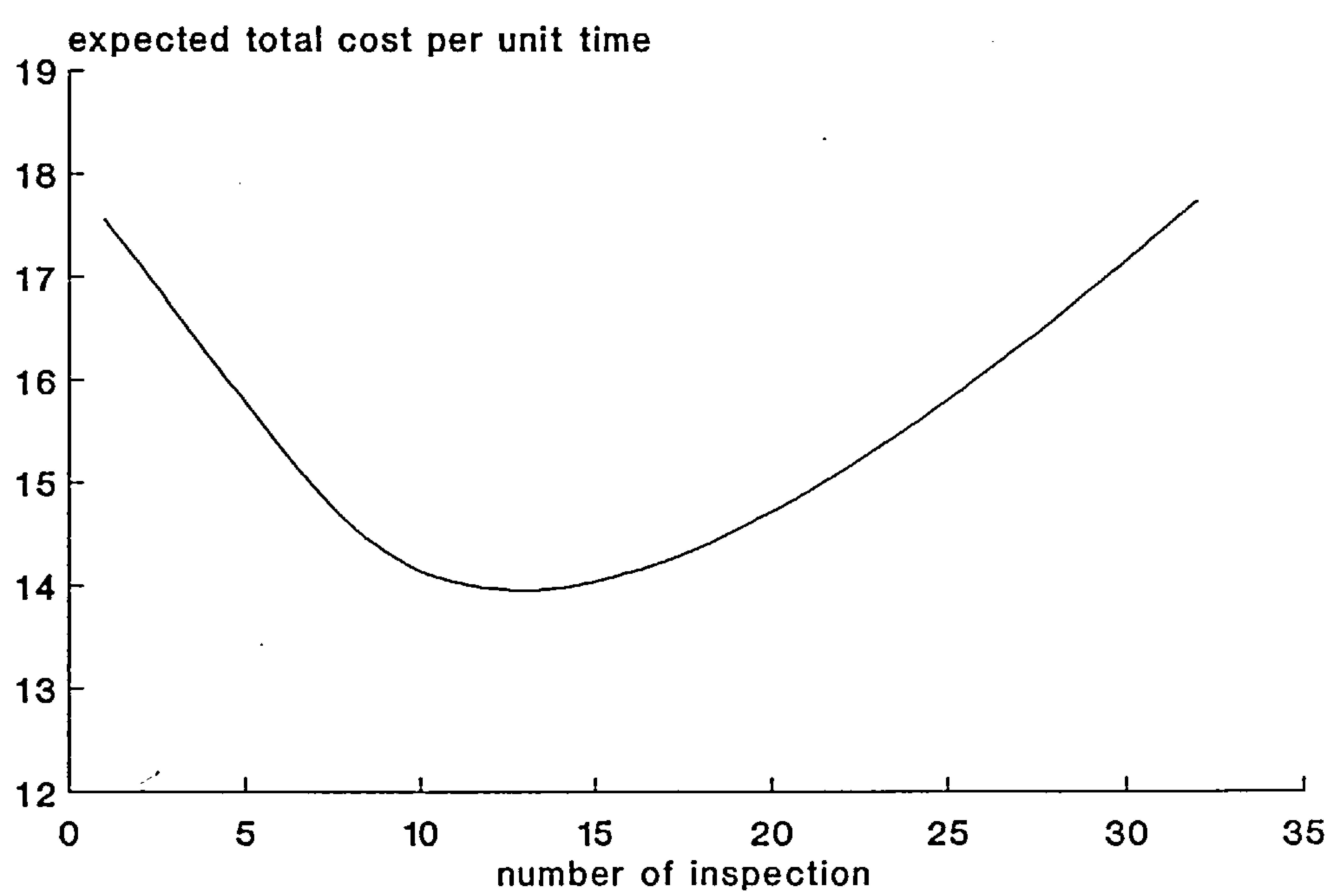


Figure 7-6 Expected total cost per unit time

7.4 Extensions to the basic model

Several extensions to the above inspection model could be considered. For instance, the schedule of monitoring checks has been assumed to be regular. It could be that an irregular schedule has sufficient advantage for some equipment to justify the additional effort of operating on a non-regular basis. This can only be checked by considering the consequence of a variable period between monitoring checks and the readily formulated model solved as a multi-dimensional optimization problem. This will be discussed in next Chapter.

Again, if at each check there was detail information available on the wear, and not just binary data as assumed in the basic model, then depending upon the confidence that can be placed in the measure, a model could be constructed to help decide when to carry out the next inspection. We now take a brief look at this particular point.

Suppose that the past records of condition monitoring times and information are available in more than binary form, and that at each inspection time point the detail information of a wear measurement of the bearing has been recorded, see Figure 7-9.

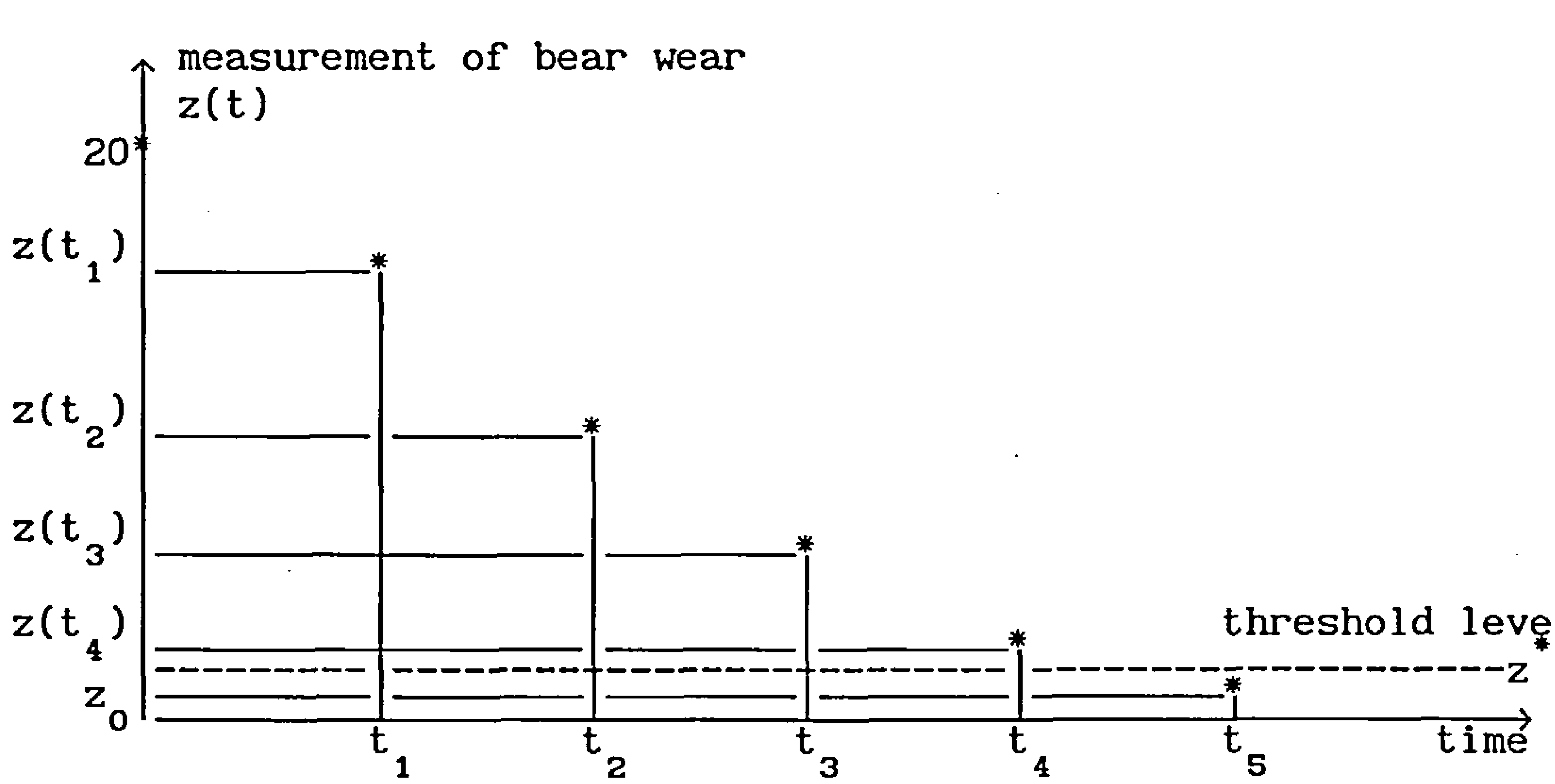


Figure 7-9 Bearing wear process

The problem is to decide when to conduct the next condition inspection and what action to take conditional upon the previous results of condition monitoring checks.

There are two approaches which could be used to model the above situation. The first one, as have stated in Chapter 6, section 6.5, will utilize proportional hazard modelling (PHM) to model the consequence of condition monitoring inspection using the full information of previous monitoring checks. The second one, perhaps a more generalized one, will adapt the stochastic theory of wear process to model building. To see this, consider the case when failure is dependent upon the wear of the bearing. Let z(t) be the value of the wear of the bearing at time t (age). Clearly {z(t)} is a stochastic process of specified structure. Suppose further that there are two fixed critical levels of wear, namely z^* and z_0. That is, if $z_0 < z(t) \le z^*$, the component is defective, where z^* denotes the threshold level of wear in defective state. And then, if $z(t) \ge z_0$, a failure occurs. The initial time u of a defect being first detected is now the first passage time of the stochastic process {z(t)} across the barrier z^*. See Cox [1962] for a definition of the first passage time. For suitably simple {z(t)} the distribution of u can be determined explicitly. For example, suppose that wear is produced by a series of 'blows' occurring in a Poisson process, or more generally in a renewal process. Suppose also that the wear at the ith blow is a positive random variable w_i, the sequence $\{w_i\}$ being independent identically distributed random variables of the renewal process. Mercer [1961] gives a special case of this model where the wear per blow has a Gamma distribution.

Another variant of the problem is where other forms of distribution of the wear rate parameter θ might be considered more appropriate, and where this distribution is revised after each inspection result becomes known. The consequence of wear models, other than the linear one adopted here, could also be considered. Such generalizations will pose no problem of principle, though they would obviously lead to a more complex formulation.

7.5 Conclusions

The above discussion has described a specific model of condition monitoring based on a linear wear pattern and presented numerical results. Both the model in general and the results in particular feature the tradeoffs that are to be expected of a condition monitoring situation. In the numerical example, both the plant availability and maintenance costs are optimized at between 10 and 15 monitoring checks over the variable monitoring range. The fact that the optimal range is common for both availability and cost considerations is simply fortunate in this case. As the number of monitoring checks increases, the probability of a failure decreases as expected.

Possible extensions to the basic model have been briefly addressed in the last section. It is expected that some of them will be discussed in detail in due course. Particularly, the modelling of condition monitoring inspection based upon records of condition information z at the past. The effect of this later method depends upon the correct and accurate estimation of the pdf of u conditionally on z.

The benefit of this and related studies at this stage is that they give some quantitative insight into the order of magnitude of the various effects and gains that might be expected from condition monitoring of equipment. This in turn gives a measure of the effort in data collection and experimentation that can justifiably be spent in establishing appropriate wear models. As indicated in the reference of Christer [1992], a key issue in justifying quantitatively the implementation of condition monitoring is an understanding of the prognosis of plant behaviour subsequent to a condition monitoring reading, such as the revision of a failure distribution in the light of the monitoring result. The technology of condition monitoring appears currently short of this ideal.

It is hoped that attempting to model quantitatively condition monitoring, attention can be focused on the issues, and the contribution that engineers, statistician and OR scientists can jointly make be highlighted and research in the area stimulated.

CHAPTER 8. ALGORITHMS FOR CONDITION MONITORING INSPECTION MODELLING

8.1 Introduction

As we have stated before that the importance of equipment conditional monitoring inspections in many organizations is increasing markedly as manufacturing plant becomes more automated, high volume, and expensive. In response, more condition monitoring check devices and methods are being developed and implemented in industry. The purpose of undertaking conditional monitoring inspections is to monitor the condition of plant manufacturing the products either continuously or periodically so that appropriate responsive maintenance actions may be undertaken. Accordingly, management expect to maintain the plant in adequate working condition with an acceptably low level of production loss. Obviously a well structured condition monitoring inspection schedule is important because the decision of how and when to carry out monitoring checks may not only be vital to the performance of plant, but may also be of considerable significance to the cost and efficiency of maintenance performance. Traditionally, the inspection schedule is usually made on a regular basis, that is, the inspection interval is equally spaced. However, inspections may not be regular. It is possible that irregular inspection intervals could give a better result. Clearly, if an irregular inspection schedule is optimal it must also be superior to the regular inspection schedule since the latter is a special case of the former. Therefore, in this Chapter, we will focus upon the problem of irregular conditional monitoring inspection scheduling and take the regular inspection scheduling as a special case of it.

In Chapter 6, we have discussed models of conditional monitoring inspection originally developed by Christer and Waller [1984a,b] to model the inspection of industrial equipment using the concepts of delay time and delay time analysis, see Chapter 3. In a more recent paper, Christer [1992] presented prototype models of irregular condition monitoring inspection, but did not discuss in any detail methods of obtaining their solution. In this Chapter, we develop algorithms to obtain the optimal condition monitoring inspection schedule based upon the models addressed in Chapter 6. For convenience, only cost measures are taken as the objective function to be minimized. However, the same method can be easily generalized to the case of downtime models. In Chapter 6, three kinds of models are discussed, namely, models over an infinite time horizon, models over a component life time, and models over a finite time horizon. For reason of simplicity, only algorithms for the first two models, ie. models over an infinite time horizon and models over a component life time, are established here because these two have relatively simple structures. We leave the development of algorithms of the finite time horizon case to a future study. Numerical examples are given to illustrate the methods developed here.

8.2 Model assumptions

Since we have listed the basic assumptions of models in Chapter 6, most of them still hold in this Chapter, but, as the algorithms developed here are for special cases, it is convenient to repeat them here. We now formally list them as follows

1. Inspections are perfect, that is the condition monitoring check information is accurate in that any defect present will be identified at an inspection.
2. If a defect is identified at an inspection, a repair or replacement is undertaken as a part of the inspection, which restores the component to as good as new condition. For convenience, we simply refer to it as an inspection replacement(we regard repair as the replacement). If a failure occurs, the component is also repaired or replaced immediately. This is simply refer to a failure replacement.
3. Times taken to conduct a monitoring inspection, inspection replacement and failure replacement are assumed negligible compared with the inspection interval. Without difficulty, we can generalize

to the case where such times are not negligible.

4. Inspection intervals need not be equal, but could be any time length and are decision variables of our model.
5. Inspections are benign in that the process of inspection will not in itself induce defects.
6. The task is to determine the optimal inspection schedule t which minimizes, say, the expected total cost per life cycle or the expected total cost per unit time over an infinite time horizon.
7. The initial time u and the delay time h are independent with pdfs g(u) and f(h) and cdfs G(u) and F(h) respectively.

We now first discuss algorithms of the model over a component life time.

8.3 Algorithm for irregular conditional monitoring inspection over a component life cycle

8.3.1 Algorithm for continuous case: Algorithm 8-1

Adopting the perfect inspection model, we established in section 4 of Chapter 6, the expected total cost per life cycle, denoted by C(t), as

$$C(t) = \text{Ex(cost of failure replacement)} + \text{Ex(cost of inspection replacement)}.$$

That is, from equation 6-23 of Chapter 6,

$$C(t) = \sum_{i=1}^{\infty}\left[\big((i-1)c_i + c_b\big)p_b(t_i) + \big[(i-1)c_i + c_m\big]p_m(t_i)\right], \qquad \text{8-1}$$

where

$$p_b(t_i) = \int_{t_{i-1}}^{t_i} g(u)F(t_i - u)du, \qquad \text{8-2}$$

and

$$p_m(t_i) = \int_{t_{i-1}}^{t_i} g(u)\{1-F(t_i-u)\}du. \qquad 8\text{-}3$$

The notation is that of Chapter 6 in which $p_b(t_i)$ is the probability of a failure occurring in (t_{i-1},t_i) and $p_m(t_i)$ is the probability that a defect is identified at t_i.

Introducing equations 8-2 and 8-3 into equation 8-1, we have

$$C(t) = \sum_{i=1}^{\infty}\left[\{(i-1)c_i+c_m)\}\int_{t_{i-1}}^{t_i} g(u)du + (c_b-c_m)\int_{t_{i-1}}^{t_i} g(u)F(t_i-u)du\right]. \qquad 8\text{-}4$$

Equation 8-4 is the objective function which will be minimized with respect to the decision variables $t=(t_1,t_2,\ldots,t_i,\ldots)$, where $t_0=0$.

As we know, a necessary condition that a sequence $\{t_i\}$ be a minimum cost inspection procedure is that $\partial C(t)/\partial t_i=0$ for all i. Hence using equation 8-4 and noting that

$$\frac{\partial}{\partial t_i}\left(\int_{t_{i-1}}^{t_i} g(u)F(t_i-u)du\right) = \int_{t_{i-1}}^{t_i} g(u)f(t_i-u)du,$$

we obtain for all i=1,2,3,...

$$-c_i g(t_i) + (c_b-c_m)\left(\int_{t_{i-1}}^{t_i} g(u)f(t_i-u)du - g(t_i)F(t_{i+1}-t_i)\right) = 0.$$

That is, for the optimal inspection policy $t = \{t_i\}$,

$$F(t_{i+1}-t_i) = \frac{(c_b-c_m)\int_{t_{i-1}}^{t_i} g(u)f(t_i-u)du - c_i g(t_i)}{(c_b-c_m)g(t_i)}. \qquad 8\text{-}5$$

Equation 8-5 defines an implicit recursive relationship among t_{i-1}, t_i and t_{i+1}. Since $t_0=0$, if we know the value of t_1, a sequence of $\{t_i\}$ may be recursively obtained through equation 8-5.

We now discuss how to choose the appropriate value of t_1. Our initial objective is to find some bounds upon t_1 to simplify subsequent numerical analysis.

Suppose that the first inspection is scheduled at time τ, then the expected cost is

$$c_m\int_0^{\tau} g(u)du+(c_b-c_m)\int_0^{\tau} g(u)F(\tau-u)du + C(\tau^*), \qquad 8\text{-}6$$

where $C(\tau^*)=\sum_{i=2}^{\infty} \{(i-1)c_i+c_m)\int_{t_{i-1}}^{t_i} g(u)du+(c_b-c_m)\int_{t_{i-1}}^{t_i} g(u)F(t_i-u)du\}$ and τ^* denotes the optimal inspection policy starting from $t_1=\tau$.

If an additional inspection is performed at time x before τ, the expected cost is

$$c_m\int_0^{x} g(u)du+(c_b-c_m)\int_0^{x} g(u)F(x-u)du+(c_i+c_m)\int_x^{\tau} g(u)du+(c_b-c_m)\int_x^{\tau} g(u)F(\tau-u)du$$

$$+\sum_{i=3}^{\infty} \left[\{(i-1)c_i+c_m\}\int_{t_{i-1}}^{t_i} g(u)du + (c_b-c_m)\int_{t_{i-1}}^{t_i} g(u)F(t_i-u)du\right]. \qquad 8\text{-}7$$

We note that $\sum_{i=3}^{\infty}\Big[\{(i-1)c_i+c_m)\}\int_{t_{i-1}}^{t_i}g(u)du + (c_b-c_m)\int_{t_{i-1}}^{t_i}g(u)F(t_i-u)du\Big]$ can be

be expressed as

$$c_i\sum_{i=3}^{\infty}\int_{t_{i-1}}^{t_i}g(u)du+\sum_{i=2}^{\infty}\Big[\{(i-1)c_i+c_m\}\int_{t_i}^{t_{i+1}}g(u)du+(c_b-c_m)\int_{t_i}^{t_{i+1}}g(u)F(t_{i+1}-u)du\Big]$$

$$= c_i\int_{\tau}^{\infty}g(u)du+\sum_{i=2}^{\infty}\Big[\{(i-1)c_i+c_m\}\int_{t_i}^{t_{i+1}}g(u)du+(c_b-c_m)\int_{t_i}^{t_{i+1}}g(u)F(t_{i+1}-u)du\Big],$$

and note also that the last two integrals are over adjacent inspection times. Now if we further require that we adopt the optimal inspection policy starting from τ, then equation 8-7 becomes

$$c_m\int_0^x g(u)du+(c_b-c_m)\int_0^x g(u)F(x-u)du+(c_i+c_m)\int_x^{\tau}g(u)du+(c_b-c_m)\int_x^{\tau}g(u)F(\tau-u)du$$

$$+ c_i\int_{\tau}^{\infty}g(u)du + C(\tau^*). \qquad 8\text{-}8$$

Thus having the first inspection performed at time $t_1=\tau$ is preferable to having the the second performed at time τ if the equation 8-8 minus equation 8-6 is greater than or equal to zero, that is

$$c_i\int_x^{\infty}g(u)du+(c_b-c_m)\{\int_0^x g(u)F(x-u)du+\int_x^{\tau}g(u)F(\tau-u)du-\int_0^{\tau}g(u)F(\tau-u)du\} \geq 0$$

or

$$s(x,\tau) = \frac{\int_0^x g(u)\{F(\tau-u)-F(x-u)\}du}{\int_x^{\infty}g(u)du} \leq \frac{c_i}{c_b-c_m}, \qquad 0 \leq x < \tau. \qquad 8\text{-}9$$

By varying τ and calculating $s(x,\tau)$ for all x, $x < \tau$, we can find the appropriate range of τ for which condition 8-9 is satisfied. In other words, if the range of τ is $(0,t_{1u})$ where t_{1u} is the upper limit of τ, then for the optimal inspection policy $t^*=\{t_i^*\}$, t_2^* should be greater than or equal to t_{1u}, ie. no additional optimal inspection point should be scheduled before τ where $0 \le \tau \le t_{1u}$.

Now consider an example. Suppose the initial time is Weibull distributed with scale factor α=0.1722 and shape factor β=1.68, and that the delay time distribution is exponential with scale factor λ=0.6633. Suppose further that c_b=200, c_m=50, and c_i=15. We have the following computed values of $s(x,\tau)$ for additional inspections, which are also plotted in Figure 8-1.

Clearly from Table 8-1 and Figure 8-1, the appropriate range of τ satisfying condition 8-9 is in (0,4.1) since $c_i/(c_b-c_m)$=0.1.

τ=2.5		τ=3.0		τ=4.0		τ=4.1		τ=4.5	
x	s(x,τ)	x	s(x,τ)	x	s(x,τ)	x	s(x,τ)	x	s(x,τ)
0	0	0	0	0	0	0		0	0
0.5	0.011	0.6	0.015	0.7	0.027	0.8		0.9	0.033
1.0	0.027	1.2	0.039	1.4	0.067	1.6	0.086	1.8	0.083
1.5	0.037	1.8	0.054	2.1	0.097	2.4	0.100	2.7	0.121
2.0	0.032	2.4	0.049	2.8	0.090	3.2	0.098	3.6	0.144
2.5	0	3.0	0	3.5	0	4.0	0.019	4.5	0

Table 8-1 Values of s(x,τ)

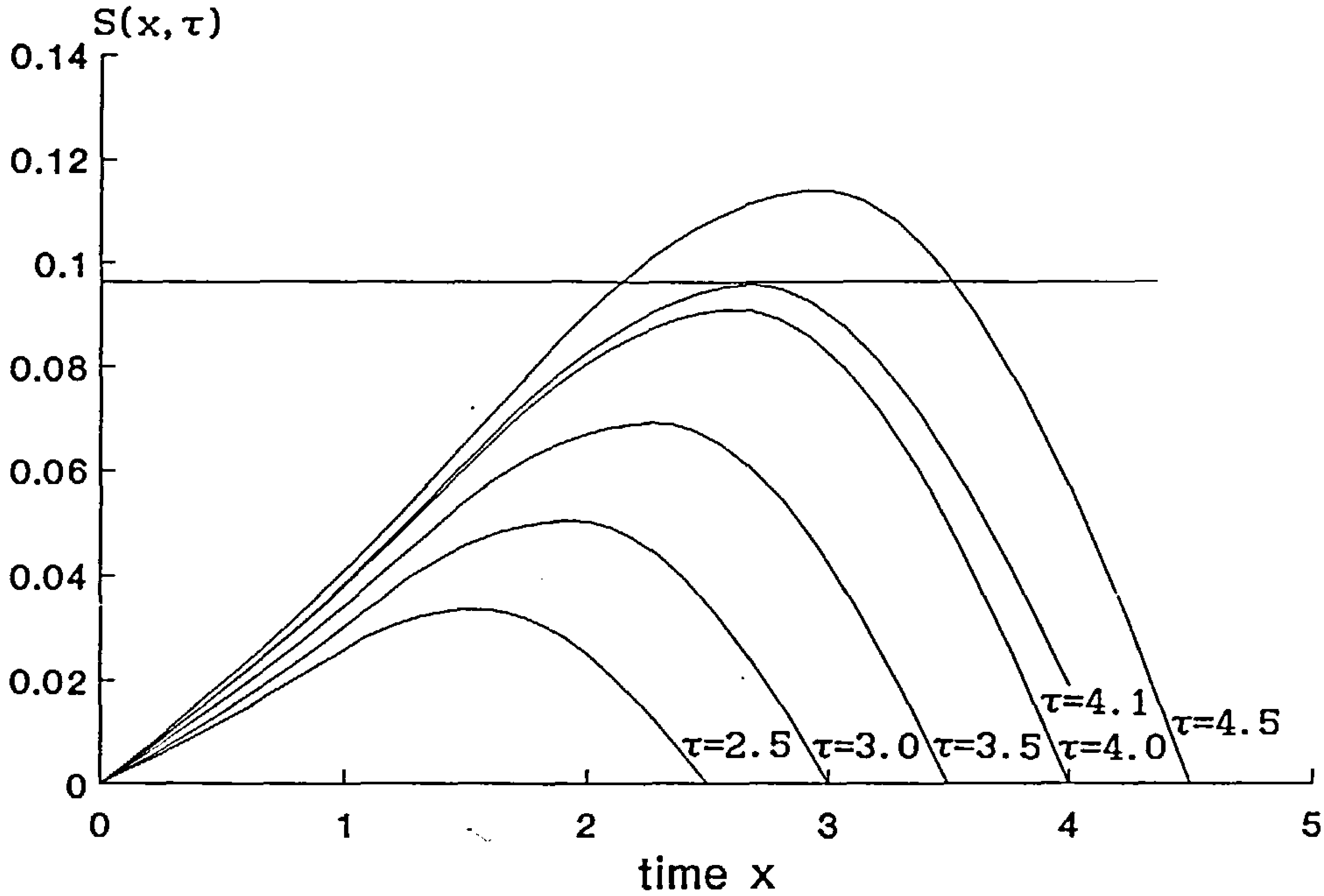


Figure 8-1 Values of $s(x,\tau)$

However, there is another issue we must pay attention to. Since $0 \le F(t_{i+1}-t_i) \le 1$, we have, the necessary condition for a solution to equation 8-5 existing is

$$\frac{\int_{t_{i-1}}^{t_i} g(u)f(t_i-u)du}{g(t_i)} \ge \frac{c_i}{c_b - c_m}, \qquad 8\text{-}10$$

and

$$\int_{t_{i-1}}^{t_i} g(u)f(t_i-u)du \le \left(1 + \frac{c_i}{c_b - c_m}\right) g(t_i), \qquad i=1,2,\ldots \qquad 8\text{-}11$$

In order to produce a sequence of $\{t_i\}$ for which $C(t)$ can be minimized, we require $t_1 < t_2 < t_3$ ....However, experience has shown that equation 8-5 can not be used recursively forever. It will stop at a certain stage of n because expressions 8-10 or 8-11 can no longer be satisfied. In this case, it means that no more inspections are needed beyond t_n when t_1 is fixed. However, if t_n is not sufficient large, there may exist the possibility of a defect arising after t_n and resulting in a failure. In this sense, we need to consider the probability of a failure after t_n in our cost model, equation 8-1. Suppose now that using equation 8-5, we obtain the optimal inspection policy $t=\{t_1,t_2,...,t_n\}$, then the total expected cost per cycle would be,

$$C(t) = \sum_{i=1}^{n} \left[\{(i-1)c_i+c_m\}\int_{t_{i-1}}^{t_i} g(u)du + (c_b-c_m)\int_{t_{i-1}}^{t_i} g(u)F(t_i-u)du \right] + (nc_i+c_b)p(t_n), \qquad 8\text{-}12$$

where $p(t_n) = 1 - G(t_n)$ is the probability of a failure after t_n.

Since by using the recurrence equation 8-5, the multi-dimensional optimization problem is changed into one dimensional one because we only need to determine the value of t_1, we can ready specify the computing procedure for obtaining the optimal inspection schedule.

Algorithm 8-1

1. Choose the range of t_1 satisfying the bounds of equation 8-9, denoted by (t_{11},t_{1u}), where in general t_{11}=0.
2. For the given range of t_1, set the step of Δt. Then for $t_1=t_{11}+j\cdot\Delta t$ till t_{1u}, calculate $t_2,t_3,...,t_n$ recursively from equation 8-5.
3. For each $t=\{t_1,t_2,...,t_n\}$, compute C(t) according to equation 8-12.
4. Select the optimal one from all C(t).

Consider now our numerical example described above, introducing the distribution of f(h) into equation 8-5, we have

$$t_{i+1} = \frac{\log\left[1 - \frac{(c_b - c_m)\int_{t_{i-1}}^{t_i} g(u)\lambda\exp\{-\lambda(t - u)\}du - c_i g(t_i)}{(c_b - c_m)g(t_i)}\right]}{-\lambda} + t_i. \qquad 8\text{-}13$$

Since we known that range of t_1 is (0,4.1), enumerating all the possible t_1 in (0,4.1) by setting Δt=0.005, we found that the optimal $C^*(t)$ is 141.17 and the optimal inspection policy is shown in Table 8-2.

inspection sequence	inspection time point	inspection interval	inspection sequence	inspection time point	inspection interval
1	3.23	3.23	9	12.49	0.95
2	4.83	1.60	10	13.44	0.95
3	6.17	1.34	11	14.39	0.95
4	7.38	1.21	12	15.37	0.98
5	8.50	1.12	13	16.43	1.06
6	9.55	1.05	14	17.66	1.23
7	10.56	1.01	15	19.32	1.66
8	11.54	0.98	16	23.94	4.62

Table 8-2 Optimal inspection policy by using algorithm 8-1

Note from Table 8-2 that the optimal inspection interval decreases till the 10th inspection time point then increases after that, which is not what we expected. We will give a brief discussion about this problem in section 8-5.

For the purpose of illustration, values of C(t) when t_1=2.5,...,4.1 are also shown in Figure 8-2, which clearly shows that t_1=3.23 is the optimal one

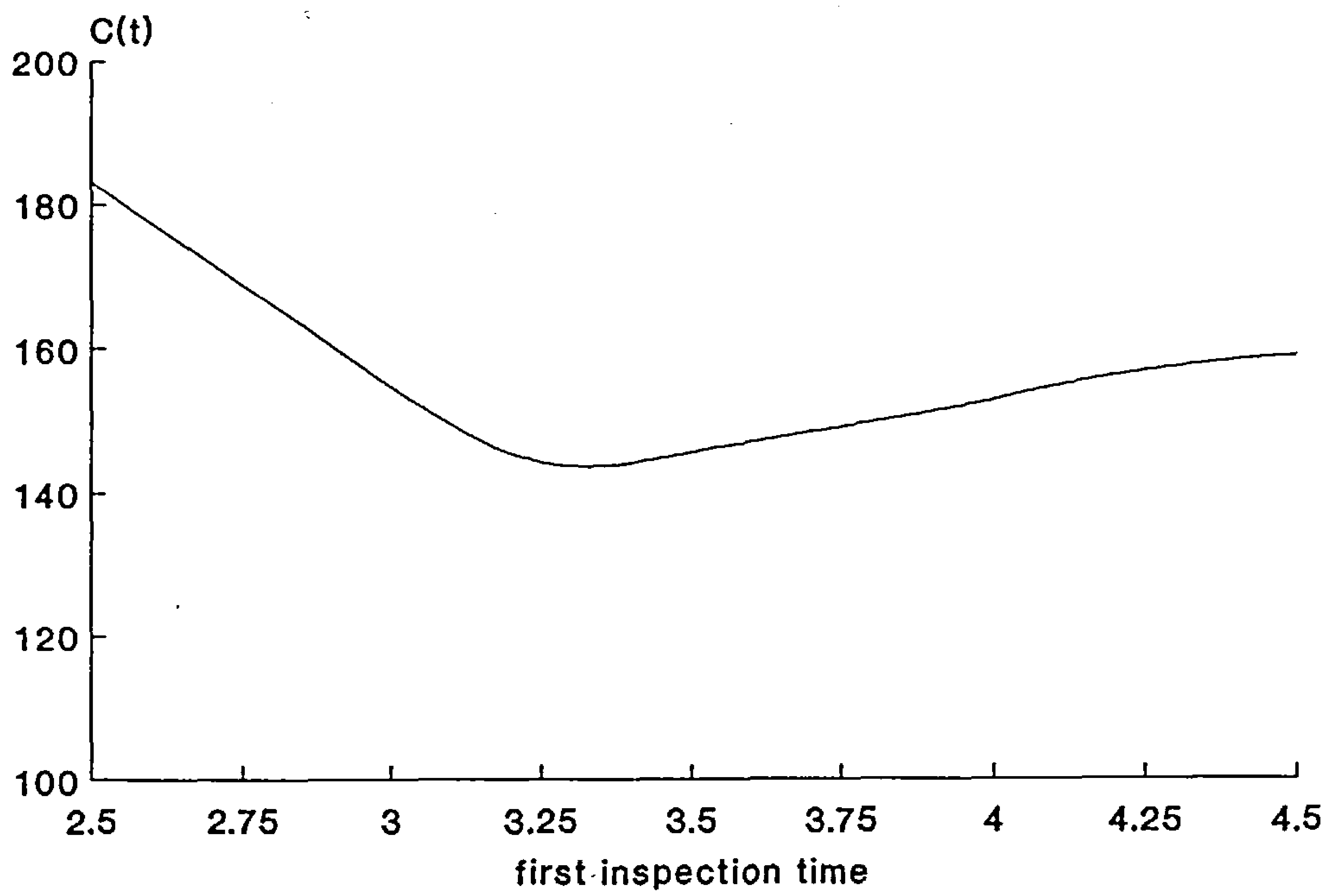


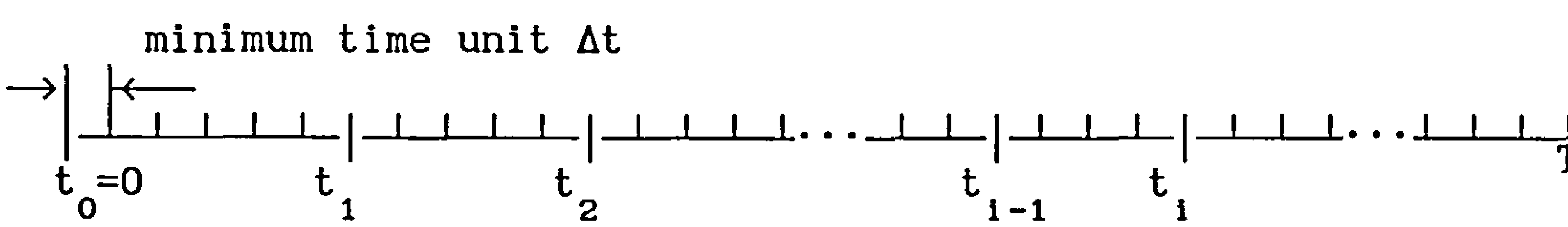
Figure 8-2 Values of C(t) while t_1 is from 2.5 to 4.1

The above algorithm gives the optimal inspection schedule for the continuous case. However in practice maintenance engineers usually adopt a convenient integer time unit as the inspection interval, for example days, weeks, months or years. Of course, we can still use the above algorithm and adjust the inspection interval to be equal to the near integers of the optimal inspection time points. But clearly we can not expect this to also be an optimal solution. However, according to the structure of equation 8-4, if we assume that the inspection interval is taken as the integer times a specified minimum time unit, we can use an alternative approach to derive the optimal inspection schedule for which the inspection interval is an integer multiple of a convenient minimum time unit.

We now discuss this method

8.3.2 Algorithm for discrete case: Algorithm 8-2

Consider the perfect inspection case. If we assume that the inspection interval is an integer multiple of a specified minimum time unit, see Figure 8-3, we can construct the algorithm as follows.



where T is the time horizon. T should be sufficiently large so that no defect could arise after it, that is $G(T) \to 1$.

Figure 8-3 Inspection process

Since we know from equation 8-4 for C(t) that if we fix one of the inspection time points, say t_i, then equation 8-4 can be divided into two parts which are mutually independent, namely

$$C(t)=\sum_{j=1}^{i}\left[\{(j-1)c_i+c_m\}\int_{t_{j-1}}^{t_j}g(u)du+(c_b-c_m)\int_{t_{j-1}}^{t_j}g(u)F(t_j-u)du\right]+$$

$$\left[\sum_{k=i+1}^{n}\left[\{(k-1)c_i+c_m\}\int_{t_{k-1}}^{t_k}g(u)du+(c_b-c_m)\int_{t_{k-1}}^{t_k}g(u)F(t_k-u)du\right]+\right.$$

$$\left.(nc_i+c_b)\{1-G(t_n)\}\right],$$

where t_n denotes the last inspection before T.

This property provides us with the chance to use dynamic programming technique to compute the optimal inspection policy.

Algorithm 8-2

Suppose now that we start from n=1, i.e. only one inspection is performed. Then the optimal expected total cost, $C^*(t)$, for convenience denoted this by $C^*(1)$, is given by

$$C^*(1) = \min_{t_1}\{ C(0,t_1) + C(t_1,T) \}, \qquad t_1=k\Delta t,\ k=1,\dots,n_{max},$$

where $C(0,t_1)$ and $C(t_1,T)$ are contributions to the expected values for which

$$C(0,t_1)=c_m\int_0^{t_1}g(u)du + (c_b-c_m)\int_0^{t_1}g(u)F(t_1-u)du$$

and $$C(t_1,T)=(c_i+c_b)p(t_1) = (c_i+c_b)\{1 - G(t_1)\}.$$

The maximum number of possible inspection points is given by, n_{max}=INT(T/Δt).

Now, for n=2, the optimal expected total cost, denoted as $C^*(2)$, is

$$C^*(2) = \min_{t_2,t_1}\{ C(0,t_1) + C(t_1,t_2) + C(t_2,T)\},$$

where $t_1=k\Delta t$, $k=1,\dots,n_{max}-1$ and $t_2=j\Delta t$, $j=k,\dots,n_{max}$.

Since we have known the values of $C(0,t_1)$ in last step calculation, what we need here is to calculate the values of $C(t_1,t_2)$ and $C(t_2,T)$ for which

$$C(t_1,t_2) = (c_i+c_m)\int_{t_1}^{t_2} g(u)du + (c_b-c_m)\int_{t_1}^{t_2} g(u)F(t_1-u)du,$$

and

$$C(t_2,T) = (2\cdot c_i+c_b)\{1-G(t_2)\}.$$

Now let

$$C^*(0,t_2) = \min_{t_1}\{\ C(0,t_1) + C(t_1,t_2)\ \},$$

where t_2=jΔt, j=2,...,n_{max} and t_1=kΔt, k=1,...,j-1.

We now proceed to n=3. Clearly, the optimal expected total cost, $C^*(3)$, is

$$C^*(3) = \min_{t_2,t_3}\{\ C^*(0,t_2) + C(t_2,t_3) + C(t_3,T)\ \}$$
$$= \min_{t_2}\left[\ C^*(0,t_2) + \min_{t_3}\{\ C(t_2,t_3) + C(t_3,T)\ \}\right],$$

where t_3=jΔt, j=3,...,n_{max} and t_2=kΔt, k=2,...,j-1.

As before, since we have known the value of $C^*(0,t_2)$, we need only compute $C(t_2,t_3)$ and $C(t_3,T)$ for which

$$C(t_2,t_3) = (2c_i+c_m)\int_{t_2}^{t_3} g(u)du + (c_b-c_m)\int_{t_2}^{t_3} g(u)F(t_3-u)du,$$

and $$C(t_3,T) = (3c_i+c_b)\{1 - G(t_3)\}.$$

Similarly, let

$$C^*(0,t_3) = \min_{t_2}\{ C^*(0,t_2) + C(t_2,t_3) \},$$

where t_3=jΔt, j=3,...,n_{max} and t_2=kΔt, k=2,...,j-1.

We can proceed to n=4 and so on.

Suppose now that n=i and we have known the value $C^*(0,t_{i-1})$, then we have

$$C^*(i) = \min_{t_{i-1},t_i} \{ C^*(0,t_{i-1}) + C(t_{i-1},t_i) + C(t_i,T) \}, \qquad 8\text{-}14$$

where t_i=jΔt, j=i,...,n_{max} and t_{i-1}=kΔt, k=i-1,...,j-1.

In a similar way we can obtain all the values $C^*(i)$, i=1,...,n_{max} among which we can select the minimum one which minimizes the expected total cost over the component life time. This completes our algorithm.

The principle behind algorithm 8-2 is quite straight forward since it is the principle of dynamic programming. In each step we change the last inspection time point, then add its contribution to the expected value to the last step local optimal result which must remain optimal as the latter are independent of the former. It is this recurrence relationship , as presented in equation 8-14, which provides the basis of our algorithm.

Now we consider to use algorithm 8-2 on our numerical example which has been used in algorithm 8-1. With the same data used in algorithm 8-1, let T = 20, which G(20)=0.9994 is quite close to 1, and set Δt=0.5. We have n_{max}=INT(20/0.5)=40. The computed results of $C^*(i)$ are shown in

Figure 8-4. The optimal result is n=17 with $C^*(17)=141.49$ for which the optimal inspection policy is shown in Table 8-3.

inspection sequence	inspection time point	inspection interval	inspection sequence	inspection time point	inspection interval
1	3	3	9	12	1
2	5	2	10	13	1
3	6	1	11	14	1
4	7	1	12	15	1
5	8	1	13	16	1
6	9	1	14	17	1
7	10	1	15	18	1
8	11	1	16	19	1
			17	20	1

Table 8-3 Optimal inspection policy using algorithm 8-2

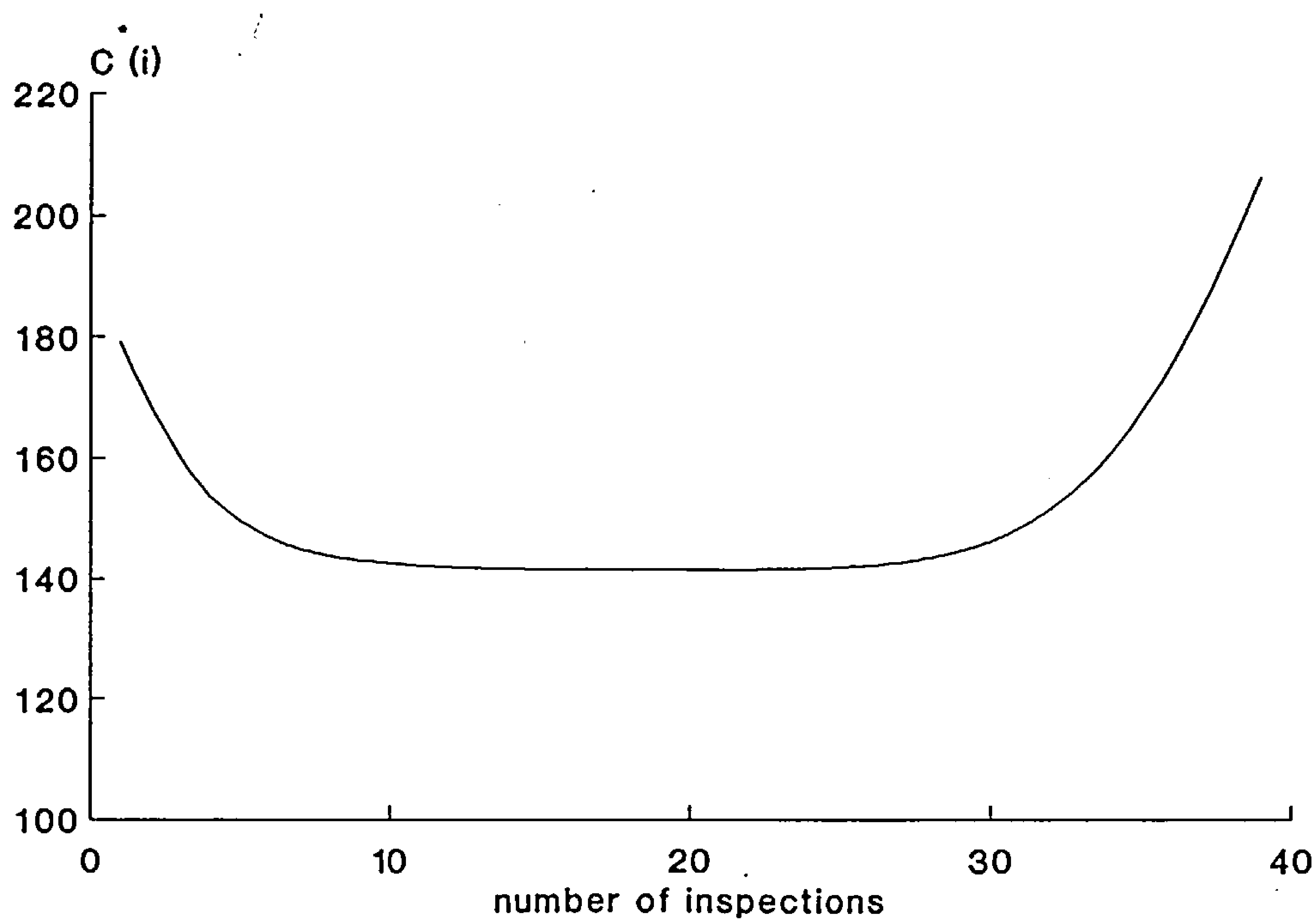


Figure 8-4 Values of $C^*(i)$

Comparing Table 8-2 with 8-3, it is shown that the two optimal policies are quite similar, as indeed they should be. The difference between the optimal cost results of these two algorithms is only 0.33. From the point of view of the theory, algorithm 8-1 gives the exact optimal solution no matter what time unit is used. However algorithm 8-2 can only give the optimal one when the inspection intervals are integer multiples of a specified and chosen calendar time unit. When the step or time unit in algorithm 8-2 is very small, it could give an answer which is quite close to the one obtained from algorithm 8-1, but will be associated with the cost of more computing time. However from the point of view of maintenance practice, algorithm 8-2 is preferable because the result obtained from it is easy to use and more consistent with maintenance practice.

For the purpose of comparison, the expected cost in the regular inspection case, that is with constant inspection interval ΔT, is also computed. The results are that the optimal $\Delta T=1.8$ and the expected total cost is 148.43. Compared with the irregular inspection, $C^*(t)=141.17$, the latter clearly gives the optimal result, as indeed it must. It should be noted, however, that the extra cost of managing a variable policy over a constant one has not been costed, and could make the constant period policy more attractive.

8.4 Algorithm for irregular conditional monitoring inspection over an infinite time horizon

In the last section we have derived optimal algorithms under the assumption that the time span extended only until the detection of a defect or a failure. In many situations, however, the process is continuing after a renewal (resulting from repair or replacement). For example a machine produces units continuously, the performance of the machine is inspected periodically to determine whether the machine is functioning satisfactorily or not. Upon detection of the malfunction, a repair or replacement is made, production then resumes, and inspection continues.

In this situation, the optimal inspection schedule should be the one that minimizes a measure of the expected total cost per unit time, or usage, as we have discussed before. Now in this section, an algorithm is presented which was first suggested by Brender [1963] to derive an optimal inspection schedule. However since the models we adopted here are based upon the concepts of delay time and delay time analysis and are, therefore, different from Brender's model, necessary revisions are made and numerical analysis techniques used to solve the infinite time horizon model presented in Chapter 6.

The expected total cost per unit time over an infinite time horizon, using the asymptotic formulation, is

$$CT(t) = \frac{C(t)}{T(t)}, \qquad \text{8-15}$$

where C(t) and T(t) denote the expected cost per cycle and expected cycle length given an inspection policy t.

From Chapter 6. section 6.3.1, we know that

$$C(t) = \sum_{i=1}^{\infty}\left[\{(i-1)c_i+c_m\}\int_{t_{i-1}}^{t_i} g(u)du + (c_b-c_m)\int_{t_{i-1}}^{t_i} g(u)F(t_i-u)du\right], \qquad \text{8-16}$$

and

$$T(t) = \sum_{i=1}^{\infty}\left[t_i\int_{t_{i-1}}^{t_i} g(u)du - \int_{t_{i-1}}^{t_i} g(u)\bar{F}(t_i-u)du\right], \qquad \text{8-17}$$

where $\bar{F}(\cdot)= \int_0^{\cdot} F(h)dh$.

The task is to find an optimal inspection schedule t which minimizes the expected total cost per unit time measure, CT(t)=C(t)/T(t).

8.4.1 The algorithm in general

In an early paper of Brender [1963], an algorithm, which was also quoted by Barlow and Proschan [1965] and Jardine [1973], is developed to solve equation 8-15. The idea is to transform equation 8-15 into another form with an extra parameter α, that is, to define

$$D(\alpha, t) = C(t) - \alpha \cdot T(t). \qquad 8\text{-}18$$

Brander established the necessary condition for minimizing CT(t) as:

1. first find t minimizes $D(\alpha, t)$ and,
2. require $D(\alpha, t) = 0$.

For a proof, see Brender [1963].

The first condition can be satisfied by minimizing equation 8-18 in terms of t_i. Partial differentiation will produce a recursive function among t_{i-1}, t_i, t_{i+1} and α which, since $t_0 = 0$, can be used to solve recursively for {t, minD(α, t)} if t_1 and α are known. Therefore, in fact, t is a function of α and t_1 . By varying α and t_1 to make the $D(\alpha, t)$ zero, an optimal solution could be obtained. That is, to minimize CT(t) is equivalent to finding appropriate t and α which minimize $D(\alpha, t)$ and then require that $D(\alpha, t)$ must be zero. Thus in Brender's algorithm, the first step, is for given α, to find $t(\alpha)$ which minimizes $D(\alpha, t)$; this can be achieved by varying different t_1. Suppose it is t_1^*. Then we vary α until we find an α^* for which α^* and t_1^* minimize $D(\alpha, t)$ and for which $D(\alpha^*, t(\alpha^*, t_1)) = 0$. However for the sake of saving computation time, the algorithm needs to be revised and also satisfy the two optimal conditions presented in Brender's paper. That is, for a given t_1 we first find an α which makes $D(\alpha, t)=0$, (finding the root of equation 8-18). Repeat this procedure for all possible t_1 we may obtain a series of $D(\alpha, t)$ which all are zero. Among them the one which has the smallest α minimizes CT(t). We will give the proof of this revised algorithm in section 8.4.2. Clearly, in general, finding the root of equation 8-18 is much quicker than finding the minimum of the same equation from the point of view of numerical calculation. And actually we only need one time minimum in the revised algorithm. Also, since the cost and time functions of C(t) and T(t) were quite simple in Brender' model, the problem could be solved analytically. But in our model, it can be seen from equations of C(t) and T(t) that the structures of C(t) and T(t) are more complex and involve double integrations and an analytic solution is not generally possible.

In principle, numerical analysis techniques enable us to solve this optimization problem to the required accuracy with the assistant of a computer. Now, in the following, the chosen optimal algorithm will follow the principle of Brender's algorithm, but with a different approach to solve our particular problem.

8.4.2 Revised algorithm: Algorithm 8-3

It was stated above that since it includes many minimizing processes Brender's algorithm needs more computation time that is necessary. We present here a slightly different method to solve the problem, though using the same general principle.

Now algorithm 8-3.

1. For given t_1, by varying α find an $\bar{\alpha}$ which makes $D(\bar{\alpha}, t(\bar{\alpha}, t_1))=0$.
2. Repeat above procedure for all possible t_1, then find a t_1^* which makes $D(\alpha^*, t(\alpha^*, t_1^*))=0$ and $\alpha^*=\min\{\ \bar{\alpha}\ \}$, where $\{\ \bar{\alpha}\ \}$ denotes all the values of $\bar{\alpha}$ obtained in step (1). Then $t(\alpha^*, t_1^*)$ minimizes CT(t).

The logic behind this algorithm is as follows: If for given α and t_1, we can obtain $t=\{t_0, t_1, ...\}$, where $t=t(\alpha, t_1)$ which minimizes $D(\alpha, t)$, when t_1 is fixed. Thus, by varying α for given t_1, we can obtain a series of t which minimize $D(\alpha,t)$. Of these α, we may find an $\bar{\alpha}$ which makes $D(\bar{\alpha},t(\bar{\alpha},t_1))=0$. By definition, we know that $t(\bar{\alpha},t_1)$ minimizes $D(\bar{\alpha},t(\bar{\alpha},t_1))$, then for any t which has the same t_1 as above we have

$$D(\bar{\alpha},t(\bar{\alpha},t_1)) \le D(\bar{\alpha},\bar{t}_1),$$

where $\bar{t}_1$ denotes an inspection policy whose t_1 is fixed as above and the others are arbitrary.

Since $D(\bar{\alpha},t(\bar{\alpha},t_1)) =0$ implies $C(t(\bar{\alpha},t_1))/T(t(\bar{\alpha},t_1)) = \bar{\alpha}$, this leads to

$$\frac{C(\bar{t}_1)}{T(\bar{t}_1)} \ge \bar{\alpha}.$$

If we repeat this procedure for all possible t_1, we can obtain a series of $D(\bar{\alpha},t(\bar{\alpha},t_1))$ which all are equal to zero. Among them we can at least find a $D(\alpha^*,t(\alpha^*,t_1))$ which has the smallest value of $\bar{\alpha}$. Since

$$\frac{C(t(\alpha^*,t_1))}{T(t(\alpha^*,t_1))} = \alpha^* \le \frac{C(t(\bar{\alpha},t_1))}{T(t(\bar{\alpha},t_1))} = \bar{\alpha},$$

then we have, for all possible t_1

$$\frac{C(t(\alpha^*,t_1))}{T(t(\alpha^*,t_1))} \le \frac{C(t)}{T(t)}$$

Now there are two problems which must be resolved. The first one is how to find a t which minimizes $D(\alpha,t)$. The second one is how to vary α so that $D(\alpha,t)=0$. These lead to the follows;

8.4.3 Necessary condition for minimizing $D(\alpha,t)$

A necessary condition for a minimum for $D(\alpha,t)$ as t is varied is obtained by setting each $\partial D(\alpha,t)/\partial t_i=0$ if α is fixed. According to the structure of $D(\alpha,t)$, this will implicitly produce a recursive relationship among t_{i-1}, t_i, t_{i+1}, and α. We now discuss this procedure below.

Since

$$D(\alpha,t) = C(t) - \alpha\cdot T(t),$$

therefore

$$\partial D(\alpha,t)/\partial t_i = \partial C(t)/\partial t_i + \alpha\cdot\partial T(t)/\partial t_i.$$

Now for convenience, we derive $\partial C(t)/\partial t_i$ and $\partial T(t)/\partial t_i$ separately, and then sum them together. From equation 8-16, 8-17, we have

$$\partial C(t)/\partial t_i=\partial/\partial t_i\left[\sum_{i=1}^{\infty}\Big[\{(i-1)c_i+c_m\}\int_{t_{i-1}}^{t_i}g(u)du+(c_b-c_m)\int_{t_{i-1}}^{t_i}g(u)F(t_i-u)du\Big]\right]$$

$$=\partial/\partial t_i\Big[\{(i-1)c_i+c_m\}\int_{t_{i-1}}^{t_i}g(u)du+(c_b-c_m\}\int_{t_{i-1}}^{t_i}g(u)F(t_i-u)du$$

$$+(ic_i+c_m)\int_{t_i}^{t_{i+1}}g(u)du+(c_b-c_m)\int_{t_i}^{t_{i+1}}g(u)F(t_{i+1}-u)du\Big]. \qquad 8\text{-}19$$

Simplifying equation 8-19 and differentiating it with respective to t_i, since we know that $\partial/\partial t_i(\int_{t_{i-1}}^{t_i} g(u)F(t_i-u)du) = \int_{t_{i-1}}^{t_i} g(u)f(t_i-u)du$, equation 8-19 becomes

$$\partial C(t)/\partial t_i=(c_b-c_m)\left[\int_{t_{i-1}}^{t_i} g(u)f(t_i-u)du - g(t_i)F(t_{i+1}-t_i)\right] - c_i g(t_i). \qquad 8\text{-}20$$

In a similar way but with slightly more complicated differentiation, we have,

$$\partial T(t)/\partial t_i=\partial/\partial t_i\left[\sum_{i=1}^{\infty}\left[t_i\int_{t_{i-1}}^{t_i} g(u)du - \int_{t_{i-1}}^{t_i} g(u)\bar{F}(t_i-u)du\right]\right]$$

$$= G(t_i)-G(t_{i-1}) - \int_{t_{i-1}}^{t_i} g(u)F(t_i-u)du_i$$

$$+ g(t_i)(t_i-t_{i+1}) + g(t_i)\bar{F}(t_{i+1}-t_i), \qquad 8\text{-}21$$

where $\bar{F}(\cdot) = \int_0^{\cdot} F(x)dx$.

Since we have assumed that we know the distribution density functions of u and h, and the values of c_i, c_m, c_b and α, it is clear from equations 8-20 and 8-21 that by letting $\partial D(\alpha,t)/\partial t_i=0$, we have,

$$(c_b-c_m)\left[\int_{t_{i-1}}^{t_i} g(u)f(t_i-u)du - g(t_i)F(t_{i+1}-t_i)\right] - c_i g(t_i) -$$

$$\alpha\left[G(t_i)-G(t_{i-1})-\int_{t_{i-1}}^{t_i} g(u)F(t_i-u)du+g(t_i)(t_i-t_{i+1})+g(t_i)\bar{F}(t_{i+1}-t_i)\right]=0.$$

8-22

If t_{i-1}, t_i and α are known, t_{i+1} may be implicitly obtained through equation 8-22 recursively. However, as can be seen from equation 8-22, it is impossible in general to derive an analytical solution to t_{i+1} if the pdf. of h is a complicated function such as Weibull or Gamma. A numerical analysis tool has to be used, for example, NAG routine C05ADF(routine to find the roots of a function), to obtain the numerical solution. Now supposing equation 8-22 is solvable, repeating this procedure, we can find $t=(t_0,t_1,t_2,...)$ for which the necessary condition of minimizing $D(\alpha,t)$ is satisfied.

However, it is interesting to note that, under the assumption of exponentially distributed delay time f(h), equation 8-22 becomes tractable. For example, introducing $f(h)=\rho e^{-\rho h}$, after some manipulation, equation 8-22 becomes

$$t_{i+1}-t_i-\log\left[1+\frac{c_i}{c_b-c_m-\alpha/\rho}-\rho\frac{\int_{t_{i-1}}^{t_i}g(u)e^{-\rho(t_i-u)}du}{g(t_i)}\right]/(-\rho)=0. \qquad 8\text{-}23$$

This is actually an explicit function for t_{i+1}, which can be easily calculated if we know the values of t_{i-1}, t_i, and α.

For convenience, let equation 8-22 be defined in general by

$$S(t_{i-1},t_i,t_{i+1},\alpha)=0, \qquad i=1,2,.... \qquad 8\text{-}24$$

Then, for a given value of α, since we know that $t_0=0$, then, once t_1 is selected, $t_2,t_3,...$ may be obtained recursively from 8-24 by numerical method, that is, to the root of $t_{i+1}$of equation 8-24 is known when t_{i-1}, t_i, and α are known. By either varying t_1 or α or both we can find a series of $t(\alpha,t_1)$ which all achieves the minimum value for $D(\alpha,t)$.

As discussed in section 8-3, difficulty may occur when we solve equation 8-24 because for given t_1 and α, equation 8-24 may show no roots beyond a certain value of n and no value of t_j $(j > n)$ exists. To this situation, since we know that equation 8-24 is the necessary condition for minimizing $D(\alpha, t)$, this means there is no such $t_j (j > n)$, for which $D(\alpha, t)$ is minimized. In other words, for the given values t_1 and α, the inspection policy is $t=\{t_1, t_2, \ldots, t_n\}$. Then as in section 8.3, C(t) and T(t) should be re-written as

$$C(t) = \sum_{i=1}^{n}\left[\{(i-1)c_i+c_m\}\int_{t_{i-1}}^{t_i} g(u)du + (c_b-c_m)\int_{t_{i-1}}^{t_i} g(u)F(t_i-u)du\right]+ (nc_i+c_b)\{1-G(t_n)\}.$$

and

$$T(t) = \sum_{i=1}^{n}\left[t_i\int_{t_{i-1}}^{t_i} g(u)du - \int_{t_{i-1}}^{t_i} g(u)\bar{F}(t_i-u)du\right] + \int_{t_n}^{\infty}\int_{0}^{\infty}(u+h)g(u)f(h)dhdu.$$

Since
$$\int_{t_n}^{\infty}\int_{0}^{\infty}(u+h)g(u)f(h)dhdu = \int_{t_n}^{\infty}ug(u)du + \int_{t_n}^{\infty}\{g(u)\int_{0}^{\infty}hf(h)dh\}du$$
$$= \int_{t_n}^{\infty}ug(u)du + \{1-G(t_n)\}\int_{0}^{\infty}S(h)dh,$$

where $S(h) = 1 - F(h)$,

we have, T(t) becomes

$$T(t)=\sum_{i=1}^{n}\left[t_i\int_{t_{i-1}}^{t_i} g(u)du-\int_{t_{i-1}}^{t_i} g(u)\bar{F}(t_i-u)du\right]+\int_{t_n}^{\infty}ug(u)du+\{1-G(t_n)\}\int_{0}^{\infty}S(h)dh.$$

However if t_n is sufficiently large, the last term in C(t) and T(t) would be small since $g(t_n)$ is small and $G(t_n)$ tends to 1, so that the last term is expected to be negligible. However, as in section 8.3.1, if $g(t_n)$ is not small or $G(t_n)$ is not close to 1 we can not ignore the last term in C(t) and T(t) and should use the full formulas. Similarly, for the purpose of reducing the computing time of the algorithm, we will also discuss the boundary problem of t_1 and α in the numerical example section.

8.4.4 Numerical computation procedure: Algorithm 8-4

From algorithm 8-3 and the last section, it is clear that from equation 8-18 and 8-24, $D(\alpha,t)$ is a function of α and t_1, that is, once α and t_1 are selected, we can obtain $t=(t_1,t_2,...)$ by using equation 8-24 recursively if it exists, and then $D(\alpha,t)$ is minimized. Since we know that equation 8-24 is a necessary condition for minimizing equation 8-18, then according to algorithm 8-3, if t_1 and α are found for which equation 8-18 is sequentially minimized and then zeroed, $t=(t_1,t_2,...)$ minimizes CT(t).

It has already been stated that the complication of equation 8-20 and 8-21 implies a numerical analysis technique must be used here to solve the problem. For convenience, NAG, Hopkins and Philips [1988], library of numerical routines available for 386-PC was chosen as the numerical analysis tool in which routines D01AJF, C05ADF and E04JBF were utilized to fulfill the purpose of calculation.

Now the complete computing procedures, algorithm 8-4 as follows:

1. Set the range of α and t_1.
2. Start from the low limit of t_1.
3. Call C05ADF (routine to find the root of a function) repeat to find $t=(t_0,t_1,t_2,...)$
4. Call C05ADF again to find the root of α in $D(\alpha,t)=0$ when t_1 is selected.

5. Call E04JBF (routine to minimize a function of variables), Here the function is CT(t) and the variable is t_1. During this call we call C05ADF again with different parameters, this call is used to obtain the root of α in $D(\alpha,t)$ while t_1 is passed from the E04JBF and α is send back to E04JBF as the function value of CT(t). Noted here that, on each iteration, C05ADF is called repeatedly to find t. We now have that α^* and t_1^* for which $D(\alpha,t)$ is minimized and equal to zero.
6. Setting $\alpha=\alpha^*$ and $t_1=t_1^*$, call C05ADF again to find $t^*=(t_1^*,t_2,\ldots)$.
7. Print out the result:

   the optimal inspection schedule: $t = t^*$ and

   the expected total cost per unit time: $CT(t) = \alpha^*$.

8.4.5 Numerical example

Assume the initial time distribution is Weibull and the delay time distribution is Exponential. The parameters and the other data are as used in the previous sections, which, for convenience, are also shown in Table 8-4.

distributions	parameters	values of cost
$g(u)=\lambda\beta(\lambda u)^{\beta-1}\exp(-(\lambda u)^{\beta})$	λ = 0.1722 β = 1.6800	c_i = 15.00 c_m = 50.00
$f(h)=\rho\exp(-\rho h)$	ρ = 0.6633	c_b = 200.00

Table 8-4 Distribution density functions of u and h and their parameters

The first step is to set the range of α and t_1. Since we know that $D(\alpha,t)=0$ means that $\alpha=CT(t)$ and the optimal irregular inspections must include the regular cases, then a simple way to set the range of α is to run a program to obtain the optimal $CT(\Delta t^*)$, where Δt^* denotes the optimal regular inspection interval by enumerating. Clearly, $\bar{\alpha}=CT(\Delta t^*)$ can be the upper limit for α since α^* must be less than $CT(\Delta t^*)$. For the lower limit of α, $\underline{\alpha}$, zero can be selected because α must be always greater than zero. Now we try to derive the range of t_1.

From equation 8-23, we know that the necessary condition for t_{i+1} existing is

$$0 \le 1 + \frac{c_i}{c_b - c_m - \alpha/\rho} - \rho \frac{\int_{t_{i-1}}^{t_i} g(u) e^{-\rho(t_i - u)} du}{g(t_i)} \le 1.$$

This is equivalent to

$$\frac{c_i}{c_b - c_m - \alpha/\rho} \le \rho \frac{\int_{t_{i-1}}^{t_i} g(u) e^{-\rho(t_i - u)} du}{g(t_i)} \le 1 + \frac{c_i}{c_b - c_m - \alpha/\rho}.$$

Let i=1, and let

$$\begin{cases} f1 = \dfrac{c_i}{c_b - c_m - \alpha/\rho}, \\[2ex] f2 = \rho \dfrac{\int_0^{t_1} g(u) e^{-\rho(t_1 - u)} du}{g(t_1)}, \\[2ex] f3 = 1 + \dfrac{c_i}{c_b - c_m - \alpha/\rho}. \end{cases}$$

Then by introducing $\bar{\alpha}$ and $\underline{\alpha}$ into f3 and f1, and letting f2=f3, f1=f2, using numerical methods we can obtain the value of t_1 which would be the upper and lower limits of t_1 respectively. The result is shown in Figure 8-5.

Before using algorithm 8-4 in our example, it is found by trial and error that equation 8-23 is not sensitive to α but very sensitive to t_1. Since for a given α, equation 8-23 is continuous in terms of t_1, by trying equation 8-23 with different values of α below $\bar{\alpha}$, only $t_1 > 3$ is seen to give lower CT(t). Therefore in the following calculation, we set t_1 in the range of (3,4) which is also shown in Figure 8-5 below.

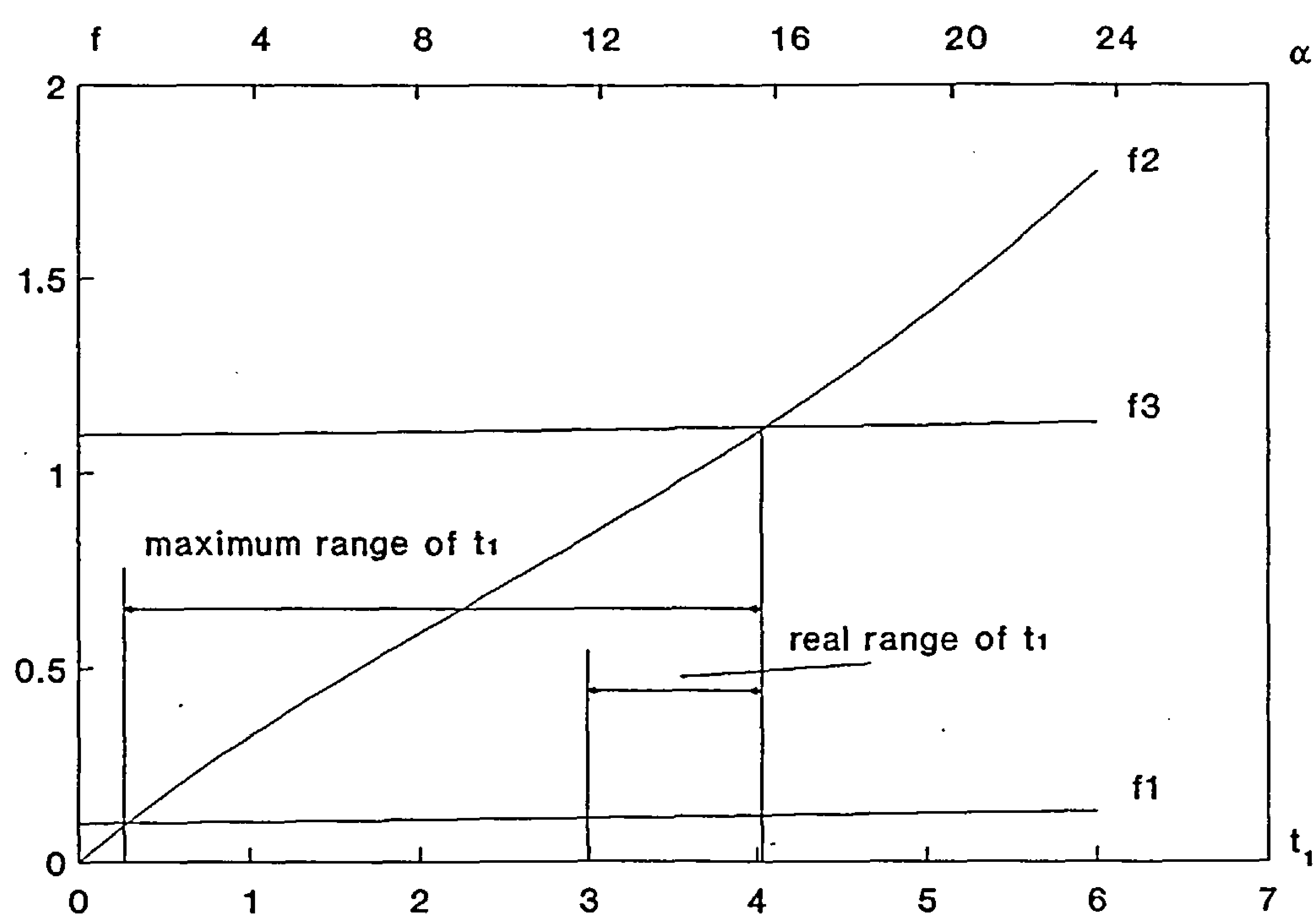


Figure 8-5 Relationship among $\bar{\alpha}, \bar{t}_1, \underline{\alpha}$, and $\underline{t}_1$.

The numerical example problem was programed in Fortran 77 in conjunction with NAG library routines E04JBF, C05ADF and D01AJF. The result shown in Table 8-5 below required only a few minutes of computing time.

inspection sequence	inspection time points	inspection intervals	$CT(t^*)$
1	3.7499	3.7499	
2	5.5488	1.7989	
3	7.0544	1.5056	
4	8.4064	1.3522	
5	9.6607	1.2543	
6	10.8485	1.1878	24.2812
7	11.9925	1.1440	
8	13.1140	1.1215	
9	14.2392	1.1252	
10	15.4095	1.1703	
11	16.7070	1.2975	
12	18.3500	1.6430	
13	21.6444	3.2944	

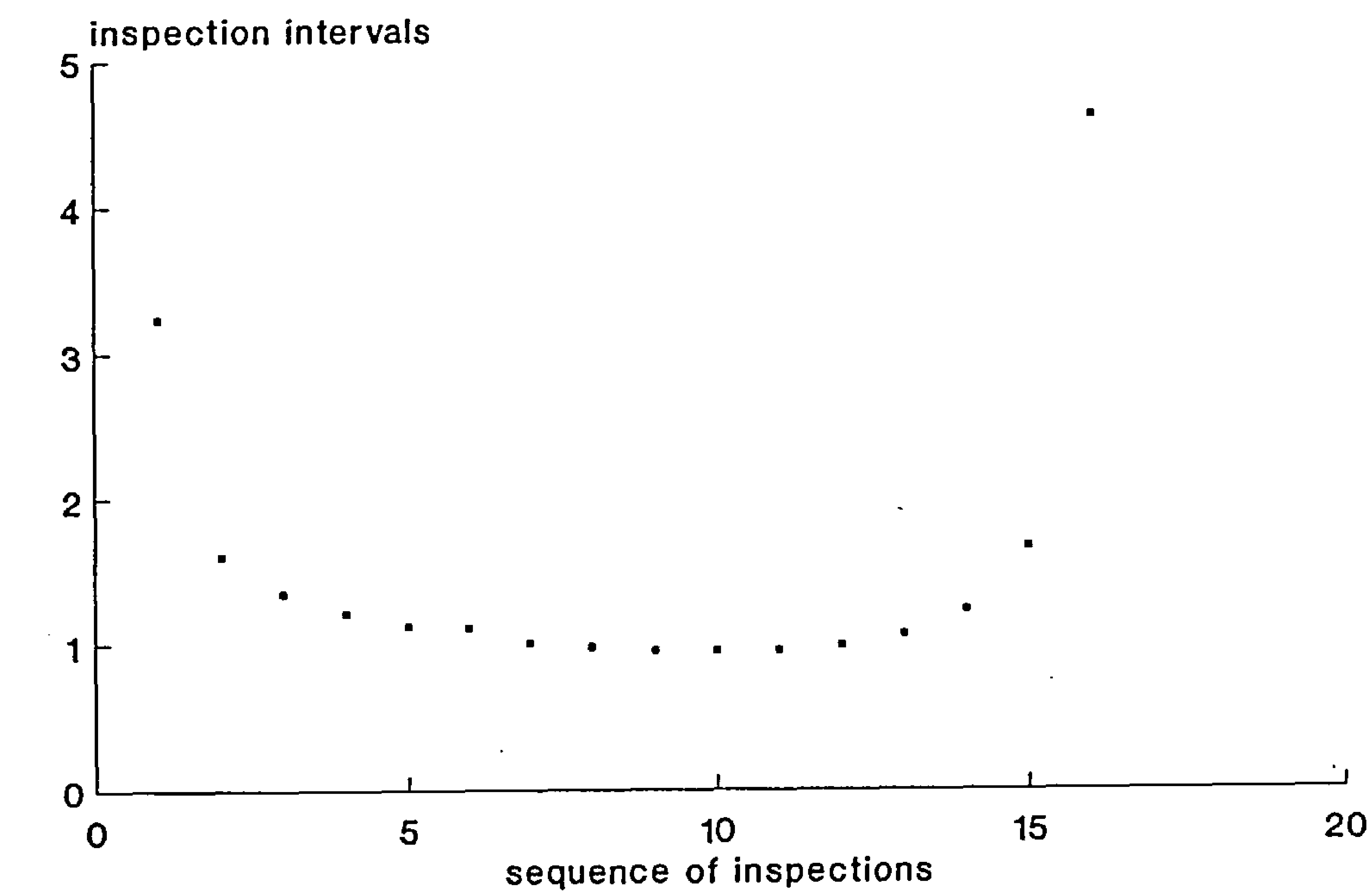
Table 8-5 Optimal inspection policy

As a matter of interest, if we assumed that the inspection interval is ΔT, constant, with the enumeration method under the same pdf. g(u), f(h), and values of costs, the optimal result are $\Delta T=2.212$ and $CT(\Delta t) = 26.30$. Compared with the optimal result of irregular inspections, the difference is about 2 units which shows that irregular inspections do give a better mathematical solution, though perhaps not necessarily a valid operational solution. The usefulnesses depends upon the valuation of the gain versus the additional effort.

8.5 Discussion and conclusions

The algorithms derived in this Chapter are both for the perfect inspection cases over a component life time and over an infinite time horizon. The objective functions are to minimize the cost measures. Obviously they can be easily extended to the case of optimizing downtime or reliability measures. The excellent Fortran Library Routine NAG has been adopted as the numerical analysis tool and the computer programs coded in Fortran are run and tested on numerical examples. The results on these examples are satisfactory and confirm our algorithms. However there are a few problems which may need attention and a further study.

1. Noted from Table 8-2 and 8-5 that the optimal inspection intervals are decreasing at the early stage of inspections, then increasing again, that is, they are U-shaped, see Figure 8-6.



data taken from Table 8-2

Figure 8-6 Inspection intervals

In general, it has been proved, Barlow et al [1965], that if the pdf. of time to failure is a PF_2 (Pólya frequency density of order 2) density function, then the optimal inspection interval is non-increasing. A density function is PF_2 if and only if it satisfies

$$s(x) = \frac{\text{pdf. of } x}{p\{\text{ failure in } (x, x+\Delta)\}}, \qquad (x, \text{ time to failure}).$$

is increasing in x.

Taking the numerical example of algorithm 8-1 in section 8.3 as an example here. Let $x=t_i$ and $\Delta=t_{i+1}-t_i$, then $x+\Delta=t_{i+1}$ for $i=1,\ldots.$ Since we know that

$$\text{pdf. of } x = \int_{t_i}^{x} g(u)f(x-u)du, \qquad x=t_{i+1},\ i=0,\ldots$$

and

$$p\ \{\text{ failure in } (t_i,t_{i+1})\} = \int_{t_i}^{t_{i+1}} g(u)F(x-u)du, \qquad i=1,\ldots$$

we have

$$s(t_i) = \frac{\int_{t_i}^{t_{i+1}} g(u)f(x-u)du}{\int_{t_i}^{t_{i+1}} g(u)F(x-u)du}, \qquad i=1,\ldots$$

With the data in Table 8-2, we have values of $s(t_i)$ as shown in Figure 8-7.

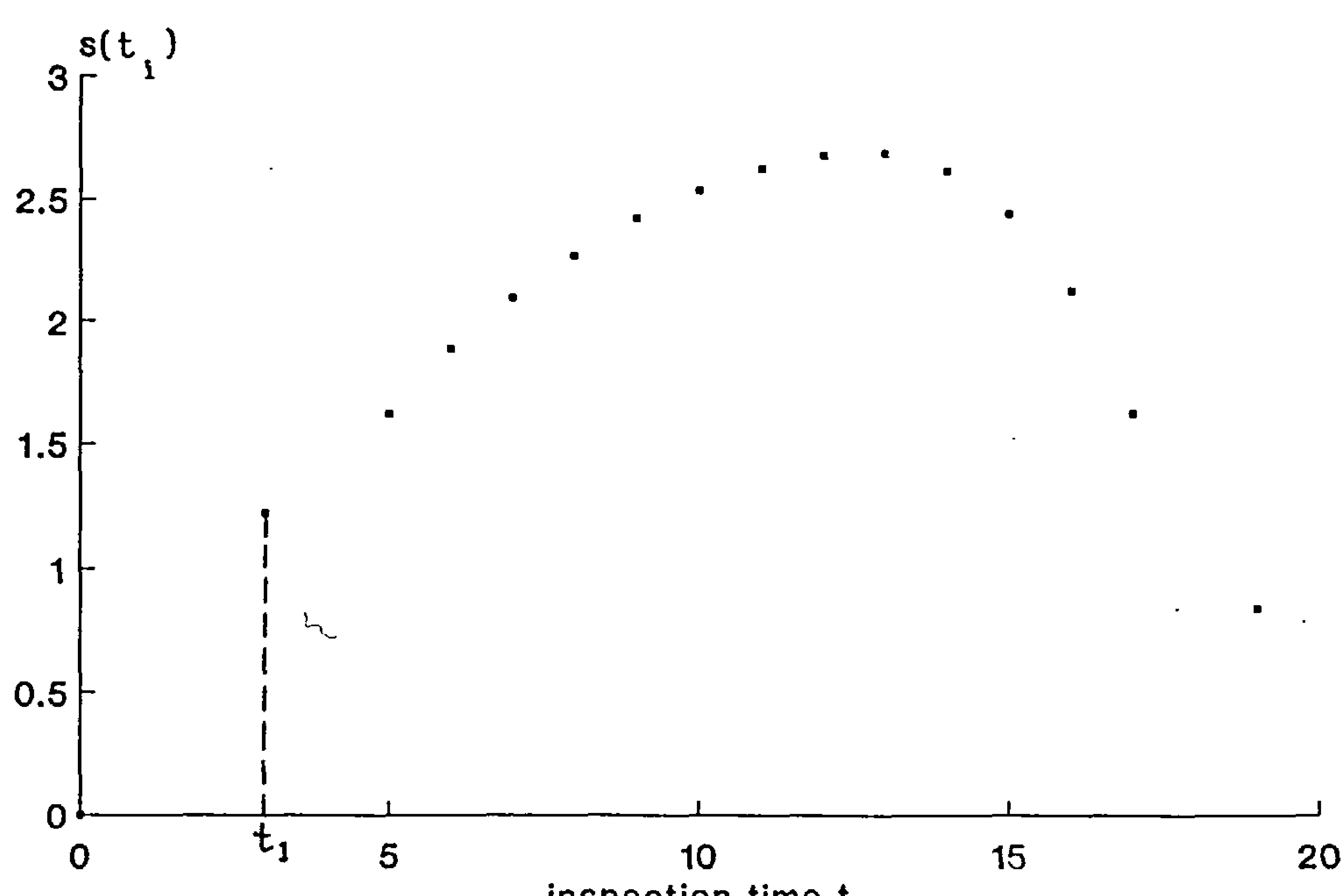


Figure 8-7 Values of $s(t_i)$

Clearly from Figure 8-7, we see that $s(t_i)$ is increasing from i=1 to i=12, then decreasing till i=16. We have stated that if s(x) is increasing in x, the optimal inspection interval is non-increasing. Figure 8-6 confirms this statement when i=1,...,12. It is very interesting to note that from i=12 to 16, $s(t_i)$ is decreasing and the optimal inspection interval is increasing. Although there is no mathematical proof that if s(x) decreases, then the optimal inspection interval increases, our example shows this trend. This topic could be worth exploring

2. The algorithms developed here are only for the perfect inspection case. Models of imperfect inspection developed in Chapter 6 are seen to be more complicated. Since we can not obtain the simple recursive relationship among t_{i-1}, t_i, and t_{i+1}, we could not directly use the algorithms presented in this Chapter. To develop an algorithm for imperfect inspection case would be another research topic.

3. In Chapter 6, we also construct a model over finite time horizon using renewal theory. However there is no appropriate approach to derive an optimal irregular inspection schedule on this model. Similar to the last problem, it could be a potential future research subject.

4. Numerical examples presented in this Chapter do show that irregular inspections are numerically superior compared to regular inspections, but there was not very much difference in the example taken. Since the purpose of the example is to show how our approach works, and also the model output depends heavily on the distributions of u and h and the values of costs, it is dangerous to generalize here since other examples may well produce different results.

APPENDIX A GOODNESS OF FIT TEST

We first derive the 'multinomial' likelihood. The multinomial log-likelihood log L_m is given by

$$\log(L_m) = \sum_{l=1}^{b_b} nb_l \log(Pb_l) + \sum_{m=1}^{b_e} ne_m \log(Pe_m) + \sum_{n=1}^{b_y} ny_n \log(Py_n), \qquad \text{A-1}$$

where

$$\sum_{l=1}^{b_b} Pb_l + \sum_{m=1}^{b_e} Pe_m + \sum_{n=1}^{b_y} Py_n = 1.$$

Here b_b, b_e and b_y are the numbers of bins for breakdown, 'no event' and positive inspection events respectively, and nb_l is the number of breakdowns in the lth breakdown bin. The origin of time is a renewal, and the size of each bin Δ_l is arbitrary. Probabilities must sum to unity, as there must be some outcome, if we include 'no event' as an outcome.

This multinomial likelihood is maximized (e.g. by Lagrangian multipliers) when probabilities P are set to the values $\hat{P}$, where e.g.

$$\hat{P} = nb_l / \left(\sum_{l=1}^{b_b} nb_l + \sum_{m=1}^{b_e} ne_m + \sum_{n=1}^{b_y} ny_n \right). \qquad \text{A-2}$$

The test of fit statistic is $S=\log(L_m)-\log(L)$, and by asymptotic theory, for large samples 2S is distributed as $\chi^2(b_b+b_e+b_y-1-f)$, where f is the number of model parameters fitted to the data.

To convert the log-likelihood of equation 4-7 in Chapter 4 to a log-likelihood where events are grouped into classes or bins corresponding to equation 4-9, we need the probability that an event falls into a particular bin. If the bins are narrow, we may approximate the probability $\int_a^{a+\Delta} P(y)dy$ by $P(x)\cdot\Delta$, where x is the observed time of the event. This simply puts back discrete versions of the differential coefficients that belong to the likelihood, but are customarily omitted. Finally, all terms corresponding to events belonging to a particular bin l must be added, to give for example

$$nb_l \log(\bar{P}b_l) = \sum_{i\in\Delta_l} \log\{ P_b(t_i^*, t_i)\cdot\Delta_i\},$$

where log-likelihood terms for the nb_l breakdown events in the lth breakdown bin have been added. Note that $\bar{P}$ is the geometric mean of the nb_l probabilities involved, as

$$\bar{P}b_l = (\prod_{i\in\Delta_l} \{P_b(t_i^*, t_i)\cdot\Delta_i\})^{1/nb_l}.$$

These factors of Δ are not required for positive inspections and 'no event' or e events.

Aimed with the $\bar{P}$ probabilities, which are predictions, one simply needs to form S as indicated above.

$$S = \sum_{l=1}^{b} {}^{b} nb_l \log(\hat{P}b_l/\bar{P}b_l) + \sum_{m=1}^{b} {}^{e}\, ne_m \log(\hat{P}e_m/\bar{P}e_m) + \sum_{n=1}^{b} {}^{y}\, ny_n \log(\hat{P}y_n/\bar{P}y_n). \quad \text{A-3}$$

Maclaurin series expansion of the logarithm now shows that 2S can be approximated by the usual definition of a chi-squared for large samples:

$$2S = N\{ \sum_{l=1}^{b_b} \frac{(\hat{P}b_l - \bar{P}b_l)^2}{\bar{P}b_l} + \sum_{m=1}^{b_e} \frac{(\hat{P}e_m - \bar{P}e_m)^2}{\bar{P}e_m} + \sum_{n=1}^{b_y} \frac{(\hat{P}y_n - \bar{P}y_n)^2}{\bar{P}y_n} \}, \qquad \text{A-4}$$

where N is the total number of renewals,

$$N = \sum_{l=1}^{b_b} nb_l + \sum_{m=1}^{b_e} ne_m + \sum_{n=1}^{b_y} ny_n .$$

Thus the likelihood-based test of fit requires the summation of the chi-squareds for three histograms: breakdowns, positive events, and 'no event's. To obtain a useful graphical picture of the fit, the histograms and predictions can be exhibited.

Strictly speaking, one can not partition degrees of freedom between the three histograms comprising the total chi-squared, as the multinomial prediction loses one degree of freedom in total since all probabilities must sum to unity, and the model has f degrees of freedom. However, if one can neglect this effect, or subtract the (f+1)/3 degrees of freedom from each histogram.

This approach has the drawback that when there are few events, bins must be wide, and it is then really necessary to integrate the probability over the whole bin. This however causes a further problem, in that as soon as one looks at probabilities for e.g. breakdown at times later than actually observed, it is unclear what value of the time from renewal t^* should be used. The occurrence of the actual observed breakdown itself imposes a renewal, which one would have to ignore, but the sequence of future inspection times if the component had not failed is unknown. In the absence of completely regular inspections, which would have enabled us to evaluate the integral, we preferred to side-step that problem.

APPENDIX B KEY PROBABILITY MEASURES IN THE LIKELIHOOD FUNCTION

Let $P_b(t_n,t)$ is the pdf. of a sequence of negative inspections, of which the last one occurs at time t_n from last renewal, and a breakdown (b-event) at time t from last renewal.

$$P_b(t_n,t) = \int_0^{t_n} g(w)\{1-Q(t_n-w)\}f(t-w)dw + \int_{t_n}^{t} g(w)f(t-w)dw. \qquad \text{B-1}$$

The first term is the pdf of a breakdown at t, from a defect that arises at $w < t_n$. It arises at time w with pdf g(w), has a probability $1-Q(t_n-w)$ of being unobserved in inspections to time t_n, and causes a failure at time t, a time t-w after arising, with pdf f(t-w). The contribution of such defects to the breakdown pdf. is then the integral over all possible times $w < t_n$.

The second term is the pdf. of breakdown at time t from defects arising after time t_n. Here we do not care whether the defect is visible or not, and the contribution to the pdf. of breakdown is the pdf. g(w) of a defect arising at w, multiplied by the pdf. of breakdown t-w later, and integrated over all $w > t_n$ and less than t.

Let $P_e(t_n,t)$ is the probability of a (possible null) sequence of negative inspections of which the last one is at time t_n, and no breakdown before observation ceases at time t from last renewal. This is referred to as an e-event as before.

$$P_e(t_n,t)=1-G(t)+\int_0^{t_n} g(w)\{1-Q(t_n-w)\}\{1-F(t-w)\}dw+\int_{t_n}^{t} g(w)\{1-F(t-w)\}dw. \qquad \text{B-2}$$

The first term, 1-G(t), is the probability that no defect arises before time t, the second contribution to the probability of no failure is the probability that a defect does arise at time $w < t_n$, but is unobserved by time t_n and does not lead to a failure before time t. The third contribution is the probability that a defect arises after time t_n, and does not lead to failure before time t. As before, one must integrate over all possible (unknown) times w at which the fault might have arisen.

Let $P_y(t_n,t)$ is the probability of a sequence of negative inspections of which the last occurs at time t_n, followed by a positive inspection at time t from last renewal. This is called a y-event (y for yes).

$$P_y(t_n,t)=\int_0^{t_n} g(w)\{Q(t-w)-Q(t_n-w)\}\{1-F(t-w)\}dw+\int_{t_n}^{t} g(w)Q(t-w)\{1-F(t-w)\}dw. \quad \text{B-3}$$

The first term is the contribution to the probability of a positive inspection at t from faults arising at times $w < t_n$, the second term the contribution from faults arising at times $w > t_n$ and less than the time of breakdown t.

A fault arising at time w with pdf. g(w) has probability $Q(t-w)-Q(t_n-w)$ of becoming visible after the last inspection at time t_n, and before the final positive inspection at time t. The probability that no failure occurs before time t is $1-F(t-w)$, and the product of these three terms is the probability of a positive inspection from defects arising at time w. This must be integrated over all $w < t_n$. Faults arising after the last negative inspection at t_n contribute similarly to the probability of a positive inspection, but must now simply not be visible or cause a breakdown by time t.

APPENDIX C FORMULATION OF THE LIKELIHOOD FUNCTION FOR IMPERFECT INSPECTION

For the events $rn_1n_2,\ldots,n_mx$ at times $t_0,t_1,t_2,\ldots,t_m,t$, see Figure C-1,

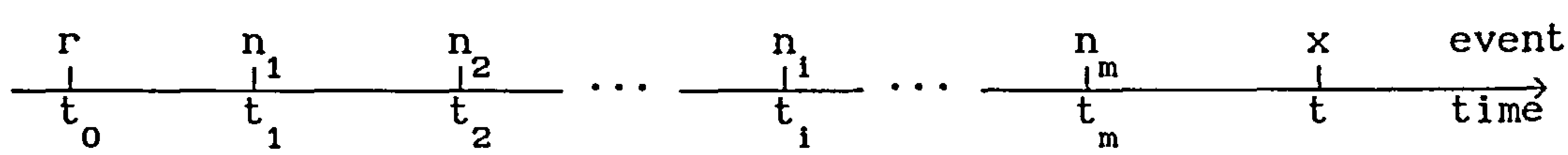


Figure C-1 Inspection process

we have, in theory, 2^m terms contribute to the likelihood of this series of events. However, from the property of conditional probability $Px|r(t_j,t)$, as defined in section 4.3, the likelihood function can be simplified considerable to only m+1 terms needed as shown in equation 4-6. We now use induction to prove the formulation, equation 4-6.

Since when m=2, we have derived the formula for a likelihood of event x at time t after two negative inspections at time t_1 and t_2 from the last renewal in section 4.3, namely

$$L = \beta Px(t_2,t) + \beta(1-\beta)Px(t_1,t) + (1-\beta)^2Px(0,t). \qquad \text{C-1}$$

Suppose for now that when $m \le n$, the likelihood of event x at time t after n negative inspections, from equation 4-6, is given by

$$L=\beta Px(t_n,t)+\beta(1-\beta)Px(t_{n-1},t)+\ldots+\beta(1-\beta)^{n-1}Px(t_1,t)+(1-\beta)^nPx(0,t). \qquad \text{C-2}$$

We will prove that when m=n+1, the likelihood of event x at time t after n+1 negative inspections will be

$$L=\beta Px(t_{n+1},t)+\beta(1-\beta)Px(t_n,t)+...+\beta(1-\beta)^n Px(t_1,t)+(1-\beta)^{n+1}Px(0,t). \quad C-3$$

For convenience, assume that one extra negative inspection is performed at time t_{n+1} before t. From equation C-2, the likelihood of a negative inspection at t_{n+1} after n negative inspection is

$$L = \beta Pn(t_n,t_{n+1})+\beta(1-\beta)Pn(t_{n-1},t_{n+1})+...+\beta(1-\beta)^{n-1}Pn(t_1,t_{n+1})$$
$$+(1-\beta)^n Pn(0,t_{n+1}). \quad C-4$$

Since event n at t_{n+1} may be either a real negative inspection or a false negative inspection, then the likelihood of event x at time t ($t > t_{n+1}$) after n+1 negative inspections with event n at t_{n+1} being a real negative inspection, denoted by L_r, is given by

$$L_r= \left(\beta Pn(t_n,t_{n+1})+\beta(1-\beta)Pn(t_{n-1},t_{n+1})+...+\beta(1-\beta)^{n-1}Pn(t_1,t_{n+1})\right.$$
$$\left.+(1-\beta)^n Pn(0,t_{n+1})\right)\beta Px|r(t_{n+1},t). \quad C-5$$

For definition of $Px|r(t_{n+1},t)$, see section 4.3.

From the definition of $Pn(t_i,t_{n+1})$, for any i=0 to n, we have

$$Pn(t_i,t_{n+1})\cdot Px|r(t_{n+1},t) = Px(t_{n+1},t). \quad C-6$$

Note also that the sum of all coefficients in equation C-4 is unity. Then equation C-5 becomes

$$L_r = \beta Px(t_{n+1}, t). \qquad \text{C-7}$$

If the inspection at t_{n+1} is a false negative one, the state at time t_{n+1} is not known. But the possibility that event n at time t_i ($i < n+1$) could be a real negative inspection exists. Since if event n at t_i is a real negative inspection, we can confirm that there is no defect before t_i. Then if let L_i denote the likelihood of event n at time t_i after i-1 negative inspections, we have

$$L_i = \beta Pn(t_{i-1}, t_i) + \beta(1-\beta)Pn(t_{i-2}, t_i) + \ldots + \beta(1-\beta)^{i-2}Pn(t_1, t_i)$$

$$+(1-\beta)^{i-1}Pn(0, t_i). \qquad \text{C-8}$$

Let L_f denote the likelihood function of event x at time t after n+1 negative inspections with event n at t_{n+1} being a false negative inspection. The task now is to formulate L_f. For convenience, we consider the situation in a backward order. Start first from i=n. If the inspection at t_n is a real one, we have the contribution to the likelihood L_f of

$$L_n \beta(1-\beta) Px|r(t_n, t),$$

where $(1-\beta)$ means the inspection at t_{n+1} is a false one.

If the inspection at t_n is a false one, the inspection at t_{n-1} may be a real one. If so, the contribution to L_f, when inspections at t_{n+1} and t_n are false ones and the inspection at t_{n-1} is a real one, is

$$L_{n-1}\beta(1-\beta)^2 Px|r(t_{n-1},t).$$

In a similar way, we have the contribution to the likelihood L_f when inspections at $t_{n+1}, t_n, \dots, t_{i+1}$ are false ones and the inspection at $t_i (i=1,2,\dots,n)$ is a real one, is

$$L_i\beta(1-\beta)^{n+1-i} Px|r(t_i,t).$$

Summing all i from i=1 to n, we have

$$L_f = \sum_{i=1}^{n} L_i\beta(1-\beta)^{n+1-i}\ Px|r(t_i,t) + (1-\beta)^{n+1}Px(0,t), \qquad \text{C-9}$$

where the last term denotes the contribution to the likelihood L_f when all previous n+1 inspections are false ones.

From equation C-6 and the property that the sum of all coefficients in equation C-8 is unity, we have

$$L_f = \sum_{i=1}^{n} \beta(1-\beta)^{n+1-i}\ Px(t_i,t) + (1-\beta)^{n+1}Px(0,t). \qquad \text{C-10}$$

Sum equation C-7 and C-10 we obtain C-3. Thus, assuming C-2 is true for all $m \le n$, we find it is true for $m \le n+1$, and since C-2 is true for m=2, it is true for all m. This completes our formulation.

REFERENCES

Abdel-Hameed, M., [1987], "Inspection and maintenance policies of devices subject to deterioration", *Advances in Applied Probability*, 10, 917-931.

Akaike, H., [1973], "Information theory and an extension of the maximum likelihood principle", *2nd International Symp.on Information Theory*, (Eds. Petrov. B.N., and Csaki, F.), Akademiai Kiado, Budapest.

Anderson, R.F., and Friedman, A., [1977], "Optimal inspection in a stochastic control problem with costly observations", *Mathematics of Operations Research*, 2, 155-190.

Anderson, R.F., and Friedman, A., [1978], "Optimal inspection in a stochastic control problem with costly observations II", *Mathematics of Operations Research*, 3, 67-81.

Ascher, H., and Feingold, H.,[1984], "Repairable system reliability", New York, *Marcel Dekker, Inc.*

Baker, R.D., [1991], "Testing the efficacy of preventive maintenance", *Journal of the Operational Research Society*, 42, 494-503.

Baker, R., and Wang, W., [1991], "Developing and testing the delay time model", *Operational Research Conference*, 20, Dec., 1991, Salford.

Baker, R., and Wang, W., [1992], "Determining the delay-time distribution of faults in repairable machinery from failure data", *IMA Journal of Mathematics Applied in Business and Industry*. Vol. 4, No. 4.

Barlow, R.E., Hunter, L.C., and Proschan, F., [1963], "Optimal Checking procedures", *Journal of the Society for Industrial and Applied Mathematics*, 4, 1078-1095.

Barlow, R.E., and Proschan, F., [1965], "Mathematical theory of reliability", New York, *Wiley*.

Beichelt, F., [1981], "Minimax inspection strategies for single unit system", *Naval Research Logistics Quarterly*, 28, 375-381.

Birolini, A., [1985], "On the use of stochastic process in modelling reliability problem", Berlin, *Springer-Verlag*.

Brender, D.M., [1963], "A surveillance model for recurrent events", *IBM*

Watson Research Centre Report.

Brookmyer, R., Day, N.E., and Moss, S., [1986], "Case-control studies for estimation of the natural history of preclinical disease from screening data", *Statistics in Medicine,* 5, 127-138.

Butler, D.A., [1979], "A hazardous-inspection model", *Management Science,* 25, 79-89.

Cerone, P., [1991], "On a simplified delay time model of reliability of equipment subject to monitoring", *Journal of the Operational Research Society*, 42, 505-511.

Chilcott, J., and Christer, A.H., [1991], "Modelling of condition based maintenance at the coal face", *International Journal of production Economics*, Vol.22, No. 1.

Cho, D.I., and Parlar, M., [1991], "A survey of maintenance models for multi-unit systems", *European Journal of Operational Research*, 51, 1-23.

Chou, C-K., and Butler, D.A., [1983], "Assessment of hazardous inspection policies", *Naval Research Logistics Quarterly,* 30, 171-177.

Christer, A.H., [1973], "Innovatory decision making", *The Role and Effectiveness of Theories of Decision in Practice* (Ed. White, D.L., and Brown, K.C.), Hodder & Stoughton, London, 368-377.

Christer, A.H., [1978], "Refined asymptotic costs for renewal reward processes", *Journal of Operational Research Society,* 29, 577-584.

Christer, A.H., [1982], "Modelling inspection policies for building maintenance", *Journal of Operational Research Society,* 33, 723-732.

Christer, A.H., [1984], "Operational research applied to industrial maintenance and replacement. In: *Developments in Operational Research* (Ed. Egless and Rand). Pergamon Press, 31-58.

Christer, A.H., [1987a],"Delay-time of reliability of equipment subject to inspection monitoring", *Journal of Operational Research Society,* 38, 329-334.

Christer, A.H., [1987b], "Comments on finite-period application of age-based replacement models", *IMA Journal of Mathematics in Management*, Vol. 1, 111-124.

Christer, A.H., [1988], "Condition-based inspection models of major civil engineering structures", *Journal of Operational Research Society*, 39, 71-82.

Christer, A.H., [1992], "Prototype model of irregular condition monitoring inspection", *IMA Journal Mathematics Applied in Business and Industry*, Vol. 3, No. 3, 219-240.

Christer, A.H., and Jack, N., [1991], "An integral-equation approach for replacement modelling over finite time horizons", *IMA Journal of Mathematics Applied in Business and Industry*, 3, 31-44.

Christer, A.H., and Redmond, D., [1990a] "Revising maintenance models and subjective data", *Proceedings of IFORS. Conference,* Athens, (Ed. Bradley, H.E.,).

Christer, A.H., and Redmond, D., [1990b], "A recent mathematical development in maintenance theory", *IMA Journal of Mathematics Applied in Business and Industry*, 2, 97-108.

Christer, A.H., and Waller, W.M., [1984a], "Reducing production downtime using delay-time analysis", *Journal of Operational Research Society,* 35, 499-512.

Christer, A.H., and Waller, W.M., [1984b], "Delay-time models of industrial inspection maintenance problems", *Journal of Operational Research Society,* 35, 401-406.

Christer, A.H., and Waller, W.M., [1984c], "An OR approach to planned maintenance modelling P.M. for a vehicle fleet", *Journal of Operational Research Society,* 35, 967-984.

Christer, A.H., and Wang, W., [1992], "A model of condition monitoring inspection of production plant", *International Journal of Production Research,* in press.

Cox, D.R., [1962], "Renewal theory", London, *Chapman and Hall Ltd.*

Cox, D.R., [1972], "Regression models and life-tables (with discussion)", *Journal of Royal Statistics Society,* B, 34, 187-220.

Cox, D.R., and Hinkley, D.V., [1974], "Theoretical statistics", *Chapman & Hall*, London.

Cox, D.R., and Oakes, D., [1983], "Analysis of survival data", London, *Chapman and Hall.*

Cox, D.R., and Lewis, P.A.W., [1966], "The statistical analysis of series of events", London, *Wiley.*

Cramer, H., [1946], "Mathematical methods of statistics", *Priceton University Press*, Princeton, New Jersey.

Derman, C., [1963], "On optimal replacement rules when changes of state are Markovian", *In Mathematical optimization Techniques* (Ed. Bellman, R.), University of California Press, Berkeley, 1963, 201-210.

Drury, M.R., Walker, E.V., Wrightman, D.W., and Bendell, A., [1988], "Proportional hazard modelling in the analysis of computer system reliability", *Reliability Engineering and system safety*, 21, 197-214.

Eddy, D.M., [1980], "Screening for cancer: theory, analysis and design", *Prentice-Hall*

Efron, B., and Hinkley, D.V., [1978], "Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information", *Biometrika*, 65, 457-487.

Feller, W., [1970], "An introduction to probability theory and its application", Vol. II, New York, *Wiley*.

Gebbnie, R., and Jenkins, L., [1986], "A decision support system for maintenance management", *IEE Inter. Maintenance Conference Proceedings*, Oct., 27-29, Chicago, Illinois.

Gross, A.J., and Clark, V.A., [1975], "Survival distributions: reliability applications in the biomedical sciences", New York, *Wiley*.

Heyman, D.P., and Sobel, M.J., [1982], "Stochastic models in operations research, Vol. I, stochastic processes and operating characteristics", New York, *McGraw-Hill*.

Hopkins, T., and Philips, C., [1988], "Numerical methods in practice using the NAG library", Wokingham, *Addison-Wesley*.

Hutchinson, T.P., and Lai, C.D., [1990], "Continuous bivariate distributions, emphasizing application", South Australia, Adelaide, *Rumsby Scientific Publishing*.

Jack, N., [1991], "Repair replacement modelling over finite time horizons", *Journal of Operational Research Society,* 42, 759-766.

Jardine A.K.S., [1973], "Maintenance, replacement, and reliability", London, *Pitman Publishing*.

Jardine, A.K.S., Ralsons, P., Reid, N., and Stafford,J., [1989], "Proportional hazard analysis of diesel engine failure data", *Quality and Reliability Engineering Int.*, 5, 207-216.

Kalbleisch, J.D., and Prentice, R.L., [1980], "The statistical analysis of failure time data", New York, *Wiley*.

Kander, Z., [1978], "Inspection policies for deterioration equipment characterized by N quality level", *Naval Research Logistics Quarterly*, 25, 243-255.

Kander, Z., and Naor, P., [1969], "Optimization of inspection policies by classical method", *Proceedings of the Annual Conference in Operational Research,* Tel Aviv, Israel.

Kao, E., [1973], "Optimal replacement rules" *Operations Research,* 21, 1231-1249.

Kaplan, E.L., and Meier, P., [1958], "Nonparametric estimation from incomplete observations", *Journal of American Statistics Association*, 53, 457-481.

Keller, J.B., [1982], "Optimum inspection policies", *Management Science,* 28, 447-450.

Kolesar, P., [1966], "Minimum cost replacement under markovian deterioration", *Management Science*, 12, 694-706.

Lawless, J.F., [1982], "Statistical models and methods for lifetime data", New York, *Wiley*.

Lee H.L., and Pierskalla W.P., [1987], "Theory and general models mass screening for contagious and non-contagious diseases", *Operational Research' 87* (Ed. Rand G.K.), 428-441.

Less, H., [1976], "Maintenance policies when deteriorating can be observed by inspection", *Operational Research,* 24, 359-366.

Luss, H., [1977], "Inspection policies for a system which is inoperative during inspection periods", *AIIE Transaction,* 9, 189-194.

Makis, V., and Jardine, A.K.S., [1991], "Optimal replacement of a production system — a proportional hazard model", *Proceedings of Inter. Conference of Production Research,* Hefei, China.

Makis, V., and Jardine, A.K.S., [1992], "Computation of optimal policies in replacement models", *IMA Journal of Mathematics Applied in Business and Industry*. Vol. 3, No. 3, 169-176.

Mann, N.R., Schafer, R.E., and Singpurwalla, N.D., [1974], "Methods for statistical analysis of reliability and life data", New York, *Wiley.*

Maritz, J.S., and Lwin, T., [1989], "Empirical Bayes methods", Second Edition, London, *Chapman and Hall.*

Mazzuchi, T.A., and Soyer, A., [1989], "Assessment of machine tool reliability using a PHM", *Naval research Logistics*, 36, 765-777.

McCall, J.J., [1965], "Maintenance policies for stochastically failling equipment: a survey", *Management Science*, 11, 493-524.

Menipaz, E., [1979], "Cost optimization of some stochastic maintenance policies", *IEEE Transactions on Reliability,* R-28, 133-136.

Mercer, A., [1961], "On wear-dependent renewal processes", *Journal of Royal Statistics Society*, B, 23, 368-376.

Motylenski, R.J., [1988], "Equipment failure information: how to use it effectively", *Paper MC-88-002, NPRA: Refinery and Petrol chemical Plant Maintenance Conference,* April, 26-29, U.S.A.

Nakagawa, T., [1984], "Periodic inspection policy with preventive maintenance", *Naval Research Logistics Quarterly*, 31, 33-40.

Nakgawa, T., and Yasui, K., [1980], "Approximate calculation of optimal inspection times", *Journal of Operational Research Society,* 31, 851-853.

Osaki, S., and Nakagawa, T., [1976],"Bibliography for reliability and availability of stochastic systems", *IEEE Transactions on Reliability*, R-25, 284-287.

Pierskalla W.P., and Voelker J.A., [1975], "A survey of maintenance models: the control and surveillance of deteriorating systems", *Naval Research Logistics Quarterly,* 23, 353-388.

Pellegrin, C.A., [1992], "A graphical procedure for an on-condition maintenance policy:imperfect-inspection model and interpretation", *IMA Journal of Mathematics Applied in Business and Industry*, Vol. 3, No. 3, 177-192.

Rosefield, D., [1976], "Markovian deterioration with uncertain information —— a more general model", *Naval Research Logistics Quarterly*, 23, 389-405.

Ross, S.M., [1969], "Markovian replacement model with a generalization to include stocking", *Management Science*, 15, 702-715.

Ross, S.M., [1970], "Applied probability models with optimization applications", San Francisco, *Holden-Day.*

Ross, S.M., [1980], "Stochastic processes", New York, *Wiley.*

Sakcmoto, Y., Ishiguro, M., and Kitagawa, G., [1986], "Akaike information criterion statistics", Tokyo, *KTK Publishing House.*

Schultz, C.R., [1985], "A note on computing periodic inspection policies", *Management Science*, 31, 1592-1596.

Sengupta, B., [1980], " Maintenance policies under imperfect information", *European Journal of Operational Research,* 5, 198-204.

Sherif, Y.S., [1982], "Reliability analysis:optimal inspection and maintenance schedules of failing systems", *Microelectronics and Reliability*, 22, 59-115.

Sherif, Y.S., and Smith, M.L., [1981], "Optimal maintenance models for system subject to failure —— a review", *Naval Research Logistics Quarterly*, 28, 47-74.

Smith, W., [1958], "Renewal theory and its ramifications", *Journal of the Royal Statistics Society*, B-20, 243-302.

Stone, G.C., [1978], "Statistical analysis of accelerated aging tests on solid electrical insulation", Msc. thesis, University of Waterloo.

Sullivan, M., [1991], "Systematic machine-condition monitoring-- a case study from Parenco paper mill in Holland", *Engineering Costs and Production Economics*, 21, 277-293.

Thomas, L.C., [1986], "A Survey of maintenance and replacement models for maintainability and reliability of multi-item systems", *Reliability Engineering,* 16, 297-309.

Tijm, H.C., [1986], "Stochastic modelling and analysis: a computational approach", New York, *Wiley.*

Turco, F., and Parolini, P., [1984] , "A nearly optimal inspection policy for productive equipment", *Inter. Journal of Production Research*, 22, 515-528.

Valdez-Flores. C., and Feldman. R.M., [1989], "A survey of preventive maintenance models for stochastically deteriorating single-unit systems", *Naval Research Logistics,* Vol. 36, 419-446.

Wattanapanom, N., and Shaw, L., [1979], "Optimal inspection schedules for failure detection in a model where tests hasten failures", *Operations Research,* 27, 303-317.

White, C.C., III, [1978], "Optimal inspection and repair of a production process subject to deterioration", *Journal of Operational Research Society*, 29, 235-243.

White, C.C., III, [1979], "Optimal control-limit strategies for a partially observed replacement", *International Journal of System Science*, 10, 321-331.

Zacks, S., [1971], "The theory of statistical inference", New York *Wiley*.

Zuckerman, D., [1980], "Inspection and replacement policies", *Journal of Applied Probability*, 17, 168-177.