MATHEMATICAL ASPECTS OF RELIABILITY-CENTERED MAINTENANCE

H. L. Resnikoff,

R & D Consultants Company

NOTICE

THIS DOCUMENT HAS BEEN REPRODUCED FROM THE BEST COPY FURNISHED US BY THE SPONSORING AGENCY. ALTHOUGH IT IS RECOGNIZED THAT CERTAIN PORTIONS ARE ILLEGIBLE, IT IS BEING RELEASED IN THE INTEREST OF MAKING AVAILABLE AS MUCH INFORMATION AS POSSIBLE.

MATHEMATICAL ASPECTS OF RELIABILITY-CENTERED MAINTENANCE

H. L. Resnikoff
R & D Consultants Company

Contract No MDA903-75-C-0349

Los Altos, CA: Dolby Access Press, 1978

TABLE OF CONTENTS

1. INTRODUCTION

1.1 The main purpose of this appendix is to provide a mathematical description of the Reliability-Centered Maintenance Program developed by United Air Lines and described in [6] and [7]. Although a mathematical formulation may not make it any easier to implement this program, by placing it in a broader context we hope to emphasize the generality of its underlying principles and encourage their application to complex systems other than commercial air fleet maintenance operations.

Another purpose of this appendix is to provide a brief but coherent introduction to those aspects of the theory of probability necessary for an understanding of the theoretical basis for the Reliability-Centered Maintenance Program. This account differs appreciably from the presentations usually found in textbooks on reliability theory: standard treatises concentrate on the functions associated with reliability and on their analytical manipulation. Here we focus on the underlying sets of items and events and on their mutual relationships. There are two principal reasons for this difference of approach, a difference which is in large measure fundamental to the philosophy underlying Reliability-Centered Maintenance.

The first reason is that collections of operational commercial and military gas-turbine-engined aircraft are among the most complex systems evolved by civilization. A single aircraft consists of tens of thousands of interrelated parts whose integrated and harmonious operation is necessary for successful completion of the aircraft's mission. These constituent parts, assemblies, and subsystems exhibit every extreme and intermediate aspect of reliability behavior. For this reason alone--complexity due to diversity--there can be no hope for a complete analytical description of reliability properties which could form the basis for development of an optimal maintenance policy. Aircraft, and aircraft systems, consist of sets of constituent parts--sets having a large number of elements, sets whose elements are related in complicated ways. Consequently, our attention must be primarily (although not exclusively) directed to consideration of aircraft and aircraft systems as sets.

The second reason is more subtle. It has been said that the principal problem facing the designer of a maintenance policy for aircraft operations is one of information. It would be more accurate to assert that the problem is one of lack of information. One of the most important contributions of the Reliability-Centered Maintenance Program is its explicit recognition that certain types of information heretofore actively sought as a product of maintenance activities are, in principle as well as in practice, unobtainable. The twentieth century has identified uncertainty as a fundamental principle on whose shifting sands profound and powerful theories have been erected: Gödel's Incompleteness Theorem in mathematical logic and Heisenberg's Uncertainty Principle in quantum physics stimulated rather than stifled progress, the spawn of the latter including microelectronics as well as nuclear science. The Reliability-Centered Maintenance Program extends these philosophical views to reliability engineering by elevating the unobtainability of information to a positive principle. This is a consequence of the following observation: the only information-bearing events which are of ultimate significance to the aircraft maintenance policy designer are failures, and among these the critical failures bear the greatest amount of information. Thus, the task of the maintenance policy designer is to minimize information. In most other comparable circumstances failure information is avidly sought, through prototype testing and sampling procedures, but those traditional approaches are inapplicable here. Fleets consist of a relatively small number of aircraft which are in a continuous state of evolution and modification and which are brought into operation in a serial rather than simultaneous manner. Hence sample sizes are generally too small for statistical procedures to carry much conviction, and for the leading edge of high-time aircraft they are always too small. In such an environment actuarial procedures are of relatively little use because the operating lifetime of an aircraft (in a fixed configuration) is relatively brief. Actuarial analyses provide interesting historical information about the effectiveness of maintenance policies and design features, but they cannot be a basis for maintenance policies.

Acquisition of the information most needed by maintenance policy designers--information about critical failures--is in principle unacceptable and is evidence of failure of the maintenance program. Critical failures entail potential (in certain cases, probable) loss of life, but there is no rate of loss of life that is acceptable to a common carrier or military organization as the price of failure information to be used for designing a maintenance policy. Thus the policy designer is faced with the problem of creating a maintenance system for which the expected loss of life will be less than 1 over the planned operational lifetime of the aircraft. This means that, both in practice and in principle, the policy must be designed without using experiential data which will arise from the failures the policy is meant to avoid.

Maintenance policy designers do have the advantage of experience gathered from operation of previous generations of aircraft. Although those aircraft are different, both in the design and fabrication of many of their constituent parts and in the relationships among those parts, it is nevertheless true that many constituents are unchanged, and most changes are minor and evolutionary rather than revolutionary. There is, consequently, a certain continuity from one generation of aircraft to the next which is utilized in an informal way by experienced maintenance engineers and aircraft designers. Although it is difficult to formulate this aspect of policy design in mathematical terms, the theoretician should not be deterred from the task because prior experience is probably the major single source of information which can be used for maintenance policy design.

In short, maintenance policy design is a problem of information and of statistics. N. Wiener [15] and A. N. Kolmogorov [5] were among the first to recognize the close relationship between statistics and information, particularly with regard to communication theory. C. Shannon [11] expanded and developed their ideas to create a rigorous and useful <u>information theory</u>. The application of Shannon's theory to maintenance policies requires that both of the concepts of information and reliability be formulated in terms of the structure of the sets of constituents of aircraft, and of functions defined on those sets. Again the desirability of a set-oriented presentation of reliability is underscored.

1.2 We will summarize the contents of the subsequent sections. Section 2, Elements of Probability, introduces the basic concepts and relationships employed throughout this work. The notion of a measurable space, which consists of a set whose elements are the items of interest, a distinguished collection of subsets called events, and a probability measure which measures the likelihood of an event, is central. Random variables are introduced as functions defined on the set of items and compatible with the structure specified by the collection of events. The distribution function associated with a random variable and probability measure is often the starting point in treatments of reliability theory. This necessitates a brief description of the three possible types of distribution functions. The remainder of the work is restricted to distribution functions which are linear combinations of absolutely continuous distributions (that is, those which have a corresponding density function) and discrete distributions. The discussion and notation are arranged in a sufficiently general manner to permit a unified treatment of both types of distributions as well as combinations of them. Combined distributions are not merely academic curiosities. Whenever a system is operated continuously over a period with numerous brief (discrete) intervals of peak stress having special characteristics, its survival distribution will be a linear combination of an absolutely continuous distribution corresponding to the continuous mode of operation and a discrete distribution corresponding to the peak-stress operation. A tungsten-filament light bulb provides a simple example. When operated continuously its survival characteristics are related to continuous filament evaporation. When the controlling switch is first turned "on," the cold filament is heated rapidly and undergoes thermal stresses. These loads evidently depend on the history of the switching activity and yield a discrete distribution. Similar phenomena occur in aircraft operation, particularly in hot areas of gas turbine engines.

These circumstances demand consideration of the Lebesgue-Stieltjes integral. The latter is not as commonly used in the literature as it should be. We present a brief and, we hope, readily accessible definition of this integral and description of those properties needed for the applications. The discussion is based on the integration-by-parts formula familiar from elementary calculus.

The derivative of an absolutely continuous distribution is called a density function. For instance, the normal density function is $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$. Discrete distributions do not have derivatives in the ordinary sense, so it is not possible to unify the treatment of densities of combined distributions without generalizing the concept of function. The required generalization is the generalized function known as the Dirac delta function, long used by engineers.

With these preliminaries in hand, conditional probabilities are defined and Bayes' Principle of Inverse Probability is introduced. Bayes' Principle is a consequence of a certain symmetry of roles played by observations and hypotheses, a symmetry most readily made evident by the set-theoretic formulation of these concepts. This symmetry, and Bayes' Principle, are of special importance to us because they provide the formal mechanism for the conversion of prior observations, e.g., survival data for constituents of a currently obsolete aircraft, into current hypotheses, e.g., initial specifications for hard-time maintenance. This application of Bayes' Principle is taken up in Section 7.

Section 3, Terminology of Reliability Theory, applies the general development of the previous section to the particular circumstances of reliability problems. The main features in this application are two: first, time t is a random variable, and the events are parameterized by t; and second, the event associated with t is interpreted as the set of items which failed prior to t. Failure and survival distributions are introduced, and it is shown how to calculate the mean time before failure. Failure density is defined and used to introduce the important concept of the hazard rate, also known as the failure rate. The hazard rate has two useful properties. The survival distribution can be expressed in terms of the hazard rate. Moreover, the hazard rate of a collection of independently failing items is the sum of the hazard rates of the individual items.

Section 4, Useful Survival Distributions, introduces five survival distributions which appear frequently in the literature: exponential, normal, Weibull, lognormal, and gamma. In each case the corresponding density and hazard functions are displayed. The survival characteristics of various jet engines or their subsystems are often accurately approximated by one of these distributions. An example of such an application is supplied for each one.

The exponential distribution plays a unique role among survival distributions. Since its hazard rate is constant, it separates the distributions which have increasing hazard rates from those which have decreasing hazard rates. Thus it also separates two fundamentally distinct classes of maintenance policies, since in the former case replacement of old by new items reduces failure rate and can, under certain circumstances, be cost-effective, whereas in the latter, replacement of old by new is only reasonable after failure.

Section 5, Simple and Complex Systems, considers infant mortality and wear out as components of the general hazard function. Simple systems, consisting either of one cell or of symmetrically interconnected replicas of one cell, are contrasted with complex systems. The principal conclusion is that complex systems are not amenable to complete mathematical reliability analysis.

Section 6, Reliability-Centered Maintenance, is the heart of this paper. Mathematical reliability analysis of an aircraft is impossible because the latter consists of tens of thousands of diverse parts. The United Air Lines Reliability-Centered Maintenance Program presents a method for grouping parts and assemblies into functionally related subsystems and systems, and for systematically eliminating certain of them from maintenance policy considerations. The purpose of Section 6 is to represent this procedure in mathematical terms.

The collection of types of items which are part of an aircraft is considered as a set with an associated survival distribution. This set is partitioned into maximally independent elements which loosely correspond to the partition described in the Reliability-Centered Maintenance Program. To each independent element is assigned a cost function which includes the direct and indirect estimated costs of a failure in addition to the costs associated with the maintenance program under consideration. The objective of the maintenance program designer is to minimize the sum of these cost functions.

Although this minimization problem is too complicated to admit a purely mathematical solution, it is nevertheless arranged in a form which makes it possible to recursively and systematically revise the maintenance policy so that total cost is reduced by each revision cycle. In fact, since the cost function is the sum of the cost functions associated with the elements of the maximally independent partition, it follows that any policy modification which reduces the cost function for one independent element while leaving the maximally independent partition unchanged must necessarily reduce the total cost function. Hence, iteration of this procedure of local cost reduction without changing the partition will lead, in the limit, to a local minimum of the total cost function. There is no way to prove that this local minimum will be the global minimum, nor is there as yet an analytical way to estimate or speed up the rate of convergence to the local minimum. Nevertheless, this procedure, which reflects the essence of the Reliability-Centered Maintenance Program, assures the maintenance policy designer that the program is self-improving.

The section closes with presentation of a geometrical model of the Reliability-Centered Maintenance Program. The maintenance/failure cost function, considered as a function of time and the policy parameters, defines a surface in a multi-dimensional space. The program defines an iterative procedure for locating a local minimum (as a function of time) on this surface.

Section 7, Information and Maintenance Policies, returns to the theme discussed earlier in this Introduction, that the most important information available to the maintenance policy designer is provided by failure experience. The designer cannot plan on the availability of such information. Three aspects of this problem are discussed in Section 7. First, the geometrical interpretation of the Reliability-Centered Maintenance Program presented at the end of Section 6 is elaborated in order to show why that program can succeed using only the small amount of information which is actually available. In essence, the program seeks valleys on the multi-dimensional surface defined by the maintenance/failure cost function. It achieves its objective by identifying a direction of decreasing cost on the surface at the point corresponding to the maintenance policy in effect, and then moving along the surface (i.e., modifying the policy) in that direction. If the distance moved is sufficiently small, iteration of this process converges to a valley point on the surface, that is, to a local minimum of the cost/maintenance function. The central fact is that relatively little information is needed to determine a direction of decreasing cost.

The difficult problem of optimizing the size of the policy change at each iteration of the program is discussed next. More information is needed to assess this 'step' size than to merely identify downward directions on the surface because the former depends on the magnitude of the derivatives of the functions defining the surface.

There follows a brief discussion of the applicability of statistical methods to complex long-lived systems having few replicas. The physical universe itself provides one example of such a system. Insofar as statistical methods are conceived as an analytical apparatus for describing sample variation, it appears that they cannot be relied on to monitor or analyze the reliability of complex systems. An alternative view, based upon Gibbs' concept of a virtual ensemble of systems, is presented. From this standpoint, statistics emerge as a selection principle which identifies a system among the virtual ensemble of its alternatives which are compatible with the non-statistical laws of nature.

Information plays a central role in the Reliability-Centered Maintenance Program and also in the discussions presented throughout this appendix but especially in Sections 5-7. In the final subsection, the quantity of information associated with a given survival distribution and inspection intervals is defined and then applied to the determination of the inspection intervals such that each interval produces the same amount of information. It is found, in agreement with expectations, that extension of replacement and/or inspection intervals is justified during periods of declining hazard rate.

A Glossary of Notation and Terminology follows Section 7.

2. ELEMENTS OF PROBABILITY

2.1 Theories of maintenance and reliability are ultimately based upon the theory of probability and on the properties of various distribution functions which have been found, either through repeated observation and experience, or by means of theoretical analyses, to occur frequently and play a role in the description and prediction of survival characteristics.

In this section we provide a brief summary of the concepts and mathematical structures used in the theory of probability in order to introduce the notations and techniques which will be used later, and also to delimit the range of our subject.

Probability theory is concerned with events and measures of the likelihood of their occurrence. These commonly used words are given precise meaning by introduction of the fundamental concept of a measurable space. Let $\underline{U}$ denote an arbitrary non-empty set and Ω a collection of subsets of $\underline{U}$ such that*

$$\underline{U} \in \Omega ; \tag{2.1.1}$$

$$\omega_1 \cap (\underline{U} - \omega_2) \in \Omega \text{ whenever } \omega_1 \in \Omega \text{ and } \omega_2 \in \Omega ; \tag{2.1.2}$$

$$\bigcup_{i=1}^{\infty} \omega_i \in \Omega \text{ whenever } \omega_i \in \Omega,\ i=1,2,3,\ldots . \tag{2.1.3}$$

The elements of Ω are called events. Eq. (2.1.1) states that the set of $\underline{U}$ (the "universe" of discourse) is an event; the meaning of eq. (2.1.2) is indicated by Figure 2.1 below, and eq. (2.1.3) asserts that any sequence of events can be combined to form an event.

*See the Glossary of Notations and Terminology for definitions of $\in$, $\cap$, $\cup$, etc.

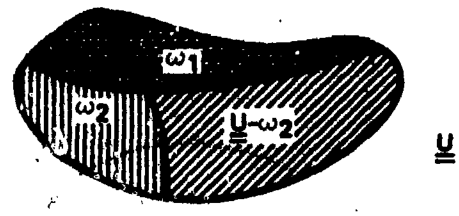


Figure 2.1. Illustrating Eq. (2.1.2)

We wish to assign some measure to the probability of occurrence of an event. This is achieved by considering a function

$$\mathbf{P} : \Omega \to [0,1] \tag{2.2}$$

which associates to each event a number between 0 and 1 inclusive such that

$$\mathbf{P}(\mathbf{U}) = 1; \tag{2.3.1}$$

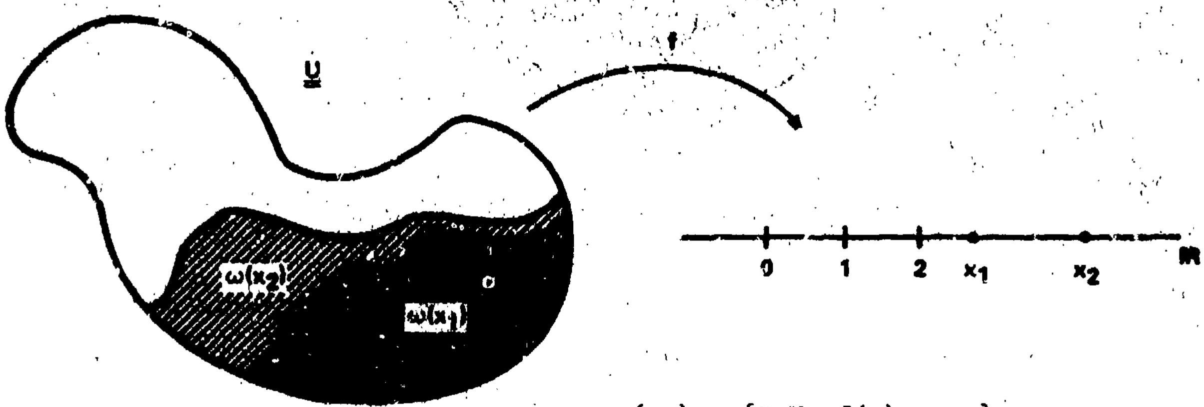
$$\mathbf{P}\left(\bigcup_{i=1}^{\infty} \omega_i\right) = \sum_{i=1}^{\infty} \mathbf{P}(\omega_i) \tag{2.3.2}$$

whenever the ω_i are disjoint events, that is, $\omega_i \in \Omega$ and $\omega_i \cap \omega_j = \emptyset$ for $i \neq j$. Such a function $\mathbf{P}$ is called a probability measure. In order to emphasize that a probability measure is a function defined on sets rather than on numbers, we use bold face type to denote it.

A probability measure is defined on the collection of events Ω, that is, on a collection of certain subsets of $\mathbf{U}$. It is also important to be able to consider functions defined on $\mathbf{U}$ itself, but not every such function can be effectively studied by analytical means, so it becomes necessary to identify a special family of functions on $\mathbf{U}$ which can be conveniently and effectively studied. These are called random variables and are specified as follows.

Suppose $f: \mathbf{U} \to \mathbb{R}$ is a real-valued function (see Figure 2.2). Each real number x can be used to specify a subset $\omega(x)$ of $\mathbf{U}$ by putting

$$\omega(x) = \{\xi \in \mathbf{U} : f(\xi) \le x\}. \tag{2.4}$$



$$\omega(x_1) = \{\xi \in \underline{U} : f(\xi) \le x_1\}$$

$$\omega(x_2) = \{\xi \in \underline{U} : f(\xi) \le x_2\}$$

$$x_1 < x_2 \text{ implies } \omega(x_1) \subset \omega(x_2)$$

Figure 2.2. Schematic Diagram of a Random Variable

The set $\omega(x)$ may be an event, that is, $\omega(x)$ may belong to Ω. If $\omega(x)$ is an event for every choice of a real number x, then the function f is called a random variable. The property of being a random variable depends on the collection of events Ω as well as on the particular function f.

The concept of integration plays an essential role in probability and statistics, and hence in the theory of reliability. A random variable f is called integrable if it can be integrable over the whole space $\underline{U}$ with respect to the probability measure $\underline{P}$. The integral is understood in the sense of Lebesgue (cp. [12], [15]). In many practical situations this integral can be expressed in terms of the ordinary Riemann integral and/or a series summation. We will have occasion to say more about this below.

The integral of a random variable f over the whole space $\underline{U}$ with respect to the probability measure $\underline{P}$ will be written

$$\int f \, d\underline{P}. \tag{2.5}$$

From eq. (2.3.1) it follows that

$$\int d\underline{P} = 1 . \tag{2.6}$$

If ω is an event, then the function c_ω defined by

$$c_\omega(\xi) = \begin{cases} 1 & \text{if } \xi \in \omega \\ 0 & \text{if } \xi \notin \omega \end{cases} , \tag{2.7}$$

called the <u>indicator function</u> of the event ω, is a random variable, and the product of $c_\omega f$ is also a random variable whenever f is.

Using this product, we define the integral of f over the event ω by

$$\int_\omega f d\underline{P} = \int c_\omega f d\underline{P} . \tag{2.8}$$

The integral of the random variable f over the whole space $\underline{U}$ is called the <u>mean value</u> of f or also the <u>expectation</u> of f, and will be denoted by $\bar{f}$:

$$\bar{f} = \int f d\underline{P} = \int_{\underline{U}} f d\underline{P} . \tag{2.9}$$

The number $\int_\omega f d\underline{P}$ can be thought of as the mean value of f on the event ω.

The variance of the random variable f (which is also the square of the standard deviation $\sigma(f)$ of f) is defined by

$$\sigma(f)^2 = \overline{(f - \bar{f})^2} , \tag{2.10}$$

that is

$$\sigma(f)^2 = \int (f - \bar{f})^2 d\underline{P} . \tag{2.11}$$

Notation is somewhat abused in this equation; the <u>number</u> $\bar{f}$ (the mean value of the random variable f) is used to stand for the random variable $\bar{f} \cdot \underline{1}$, where $\underline{1}$ is the random variable which takes on the value 1 for each element $\xi \in \underline{U}$.

2.2 In probability and statistics one is most often interested in the probability measure of the set of those events for which a random variable f satisfies $f(\xi) \leq x$ for all $\xi \epsilon \omega$, where x is some real number. Since f is a random variable, $\omega_f(x) = \{\xi \epsilon \underset{\sim}{U} : f(\xi) \leq x\}$ is an event, so by definition

$$P_f(x) = \int_{\omega_f(x)} d\underset{\sim}{P} \tag{2.12}$$

is a function of x which varies from 0 to 1. If the random variable f is fixed for the discussion, or otherwise understood, then the subscript f will be omitted and we will write $P(x)$ in place of $P_f(x)$. This function $x \to P(x)$ is called the <u>distribution function</u> of the random variable f; it is the distribution function customarily used in statistics. In order to emphasize that P is a function of a numerical variable rather than a set function it is printed in ordinary Roman type.

A distribution function has the following properties:

$x \to P(x)$ is a non-decreasing function; (2.13.1)

$$P(x) = P(x+0) \, , \tag{2.13.2}$$

where $P(x+0) = \lim_{h \searrow 0} P(x+h)$ (h approaches 0 through positive values; thus P is <u>continuous from the right</u>);

$$P(-\infty) = 0, \quad P(+\infty) = 1 \, , \tag{2.13.3}$$

where $P(\pm\infty) = \lim_{x \to \pm\infty} P(x)$

We started with a collection Ω of sets called events and associated a probability measure $\underset{\sim}{P}$ with these sets. The values assumed by $\underset{\sim}{P}$ are real numbers in the interval [0,1]. Then we defined random variables and their associated integrals relative to the probability measure. By means of the latter we have been able to construct an interplay between functions defined on sets, such as probability measures and random variables, and functions defined on real numbers. The distribution function $P_f : \mathbb{R} \to [0,1]$ provides a method for completing this transference by defining a measure on sets of real numbers which will correspond to $\underset{\sim}{P}$ and thereby enable us to express all of the integrals which occur in terms of integrals of functions of real numbers, rather than as integrals of functions of sets. This is important because the analysis of functions of real numbers is highly developed and well known. This nice property is achieved by defining a measure on $\mathbb{R}$ associated with the distribution function P_f in the following way.

If $(a,b] = \{x \in \mathbb{R} : a < x \leq b\}$, then define the measure

$$\mu_p((a,b]) = P(b) - P(a) \ , \tag{2.14}$$

where $P = P_f$ is the distribution function of the random variable f. Since $P(x) = P_f(x)$ is the probability that the random variable f assumes a value $\leq x$, it follows that $\mu_p((a,b])$ is the probability that f assumes a value in the half-open interval $(a,b]$. μ_p is a measure on the real line. The integral of a real-valued function $g : \mathbb{R} \to \mathbb{R}$ with respect to this measure is called the <u>Lebesgue-Stieltjes integral</u> of g with respect to μ_p, written

$$\int g(x) d\mu_p = \int_{-\infty}^{+\infty} g(x) dP_f(x) \ . \tag{2.15}$$

Use of the Lebesgue-Stieltjes integral unifies the treatment of discrete probability distributions and probability distributions which have density functions. Nevertheless, the Lebesgue-Stieltjes integral has not yet become a standard part of the education of those who use statistics nor an explicitly used tool in most reference books. This is no doubt due to the greater technical complications of developing the properties of this integral in the most general setting (cp., e.g. [12]). Fortunately, for the cases of interest to us, there is a simple way to express the Lebesgue-Stieltjes integral in terms of ordinary integrals and to obtain the properties of the former from the well-known properties of the latter. After some preparatory remarks we will introduce this approach, which will enable us to simplify and unify our discussion of reliability.

A distribution function P of a random variable can always be expressed as a convex sum of distribution functions of three types:

$$P = a_1 P^{abs} + a_2 P^{dis} + a_3 P^{sing} , \tag{2.16}$$

where $0 \le a_i \le 1$ and $a_1+a_2+a_3 = 1$. P^{abs} is called the <u>absolutely continuous</u> part of P, P^{dis} is the <u>discrete</u> part of P, and P^{sing} is the <u>singular</u> part of P. The absolutely continuous part P^{abs} can be differentiated with respect to x (therefore P^{abs} is a continuous function), so we can write

$$dP^{abs} = \frac{dP^{abs}}{dx} dx . \tag{2.17}$$

If $P = P^{abs}$, that is, if the distribution function of the random variable f is absolutely continuous, then

$$dP = \frac{dP}{dx} dx$$

and the function

$$p(x) =_{def} \frac{dP}{dx} \quad \text{(that is, equal by definition)} \tag{2.18}$$

is called the (<u>probability</u>) <u>density function</u> of the random variable f. The usual continuous distribution functions which appear in statistics text books are absolutely continuous and therefore possess density functions. The latter are usually the main topic of study rather than the more general but more complicated distribution functions.

The discrete part P^{dis} of the general distribution function P is a step function with at most a countable number of discontinuities. That is, if the discontinuities of P^{dis} occur at the numbers x_k, $k = \ldots-2,-1,0,1,2,\ldots$, then there are non-negative constants b_k such that

$$P^{dis}(x) = b_k \quad \text{if} \quad x_k \le x < x_{k+1} \tag{2.19}$$

where $0 \le b_k \le 1$.

Figure 2.3 gives an example of such a function.

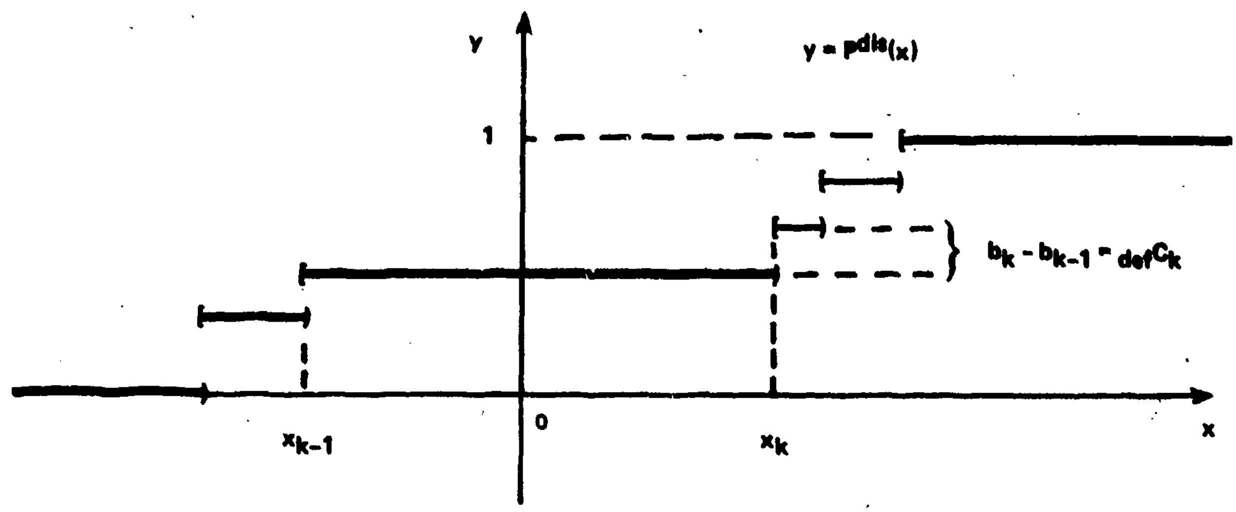


Figure 2.3. A Discrete Probability Distribution

Notice that sufficiently far to the left in the figure, $P^{dis}(x) = 0$, and sufficiently far to the right, $P^{dis}(x) = 1$; this will occur if the number of discontinuities is finite. Otherwise, in accordance with eq. (2.13.3), P^{dis} need only approach 0 (respectively, 1) in the limit as $x \to -\infty$ (respectively, $x \to +\infty$). In the figure the notation ⊢—⊣ means that the left-hand endpoint of the interval is included, whereas the right-hand endpoint is omitted. This means that P^{dis} is continuous from the right, and is the graphical interpretation of eq. (2.13.2) for P^{dis}. The quantity

$$c_k =_{def} b_k - b_{k-1} = P^{dis}(x_k) - \lim_{x \nearrow x_k} P^{dis}(x) \tag{2.20}$$

is the "jump" of the function P^{dis} for the discontinuity at $x = x_k$.

The third part of the general distribution function, the singular part P^{sing}, is of no practical importance. It is a function that is continuous everywhere and has a derivative equal to zero everywhere except on some event (subset) whose probability measure is 0. It is a remarkable fact that singular distribution functions exist. Such a function P^{sing} is non-decreasing; $P^{sing}(-\infty) = 0$; and $P^{sing}(+\infty) = 1$, which shows that $P^{sing}(x)$ actually increases as x increases; since its derivative is 0 almost everywhere, P^{sing} is constant almost everywhere. But it is also continuous; there are no "jumps." Although functions having these unusual properties can be constructed (cp. [15]), they are so complicated and pathological that they cannot play a role in practical applications of probability theory. Therefore singular distributions will be excluded from consideration in what follows: hereafter, a probability distribution will consist of a linear combination of an absolutely continuous probability distribution and a discrete probability distribution.

2.3 We are now prepared to express the Lebesgue-Stieltjes integral $\int_\alpha^\beta g(x)dP$ of a function g relative to such a probability distribution in familiar terms. The "integration by parts" formula

$$\int_\alpha^\beta g\,dP = g(x)P(x)\Big|_\alpha^\beta - \int_\alpha^\beta P\,dg \tag{2.21}$$

is valid for the Lebesgue-Stieltjes integral [12]. We will use it to define that integral in terms of the familiar integral for functions g which are differentiable. Thus, if dg/dx exists, define

$$\int_\alpha^\beta g\,dP = g(x)P(x)\Big|_\alpha^\beta - \int_\alpha^\beta P(x)\,\frac{dg}{dx}\,dx\,. \tag{2.22}$$

The integral on the right side is a conventional (Lebesgue or Riemann) integral. All the properties of the Lebesgue-Stieltjes integral can be obtained by interpreting the left side of eq. (2.22) in terms of the right side.

In particular, if $P = P^{dis}$ is the discrete distribution given by eq. (2.19) and if $x_{M-1} \le \alpha < x_M$, $x_N \le \beta < x_{N+1}$, then

$$\int_\alpha^\beta g\,dP^{dis} = \sum_{k=M}^{N} c_k g(x_k)\;, \tag{2.23}$$

where c_k is the jump of P^{dis} at the discontinuity x_k. This formula is verified by a simple calculation using eq. (2.22). Indeed, since $P(\beta) = b_N$ and $P(\alpha) = b_{M-1}$, by eq. (2.19),

$$\int_\alpha^\beta g dP = g(x)P(x)\Big|_\alpha^\beta - \int_\alpha^\beta P(x)\frac{dg}{dx}dx$$

$$= b_N g(\beta) - b_{M-1} g(\alpha) - \int_\alpha^\beta P(x)\frac{dg}{dx}dx . \qquad (2.24)$$

The last integral can be expressed as a sum of three parts (cp. Figure 2.4):

$$\int_\alpha^\beta P(x)\frac{dg}{dx}dx = \int_\alpha^{x_M} + \int_{x_M}^{x_N} + \int_{x_N}^\beta . \qquad (2.25)$$

Since P^{dis} is constant between successive discontinuities,

$$\int_\alpha^{x_M} P\frac{dg}{dx}dx = b_{M-1}\Big(g(x_M) - g(\alpha)\Big) , \qquad (2.26.1)$$

$$\int_{x_N}^\beta P\frac{dg}{dx}dx = b_N\Big(g(\beta) - g(x_N)\Big) , \qquad (2.26.2)$$

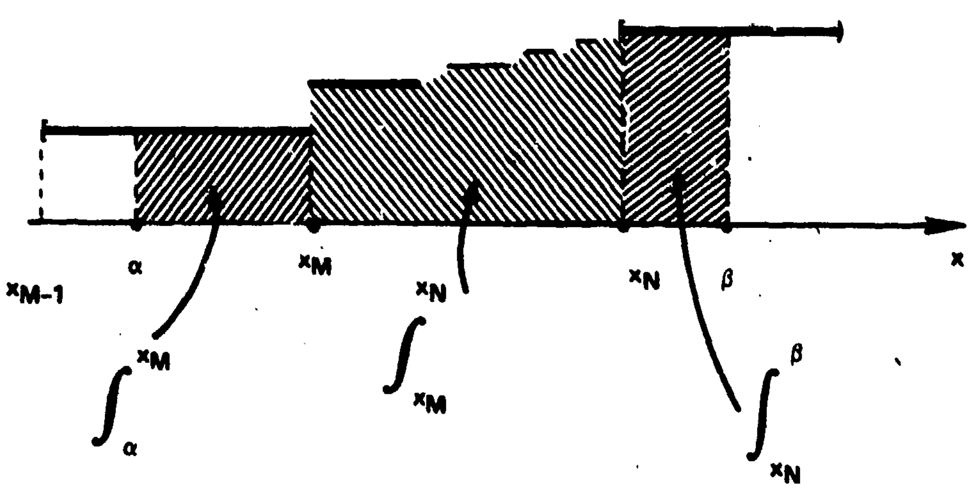


Figure 2.4. Calculation of Lebesgue-Stieltjes Integrals

$$\int_{x_M}^{x_N} P\frac{dg}{dx}\,dx = \sum_{k=M}^{N-1} \int_{x_k}^{x_{k+1}} b_k \frac{dg}{dx}\,dx$$

$$= \sum_{k=M}^{N-1} b_k\Big(g(x_{k+1}) - g(x_k)\Big) \ . \qquad (2.26.3)$$

Substitution of eq. (2.26) in eq. (2.24) yields

$$\int_\alpha^\beta g\,dP = b_N g(\beta) - b_{M-1} g(\alpha) - b_{M-1}\Big(g(x_M) - g(\alpha)\Big)$$

$$- b_N\Big(g(\beta) - g(x_N)\Big) - \sum_{k=M}^{N-1} b_k\Big(g(x_{k+1}) - g(x_k)\Big)$$

$$= b_N g(x_N) - b_{M-1} g(x_M) + \sum_{k=M}^{N-1} b_k g(x_k) - \sum_{k=M}^{N-1} b_k g(x_{k+1})$$

$$= b_N g(x_N) - b_{M-1} g(x_M) + \sum_{k=M}^{N-1} b_k g(x_k) - \sum_{k=M+1}^{N} b_{k-1} g(x_k)$$

(by relabelling the summation index in the last sum),

$$= \sum_{k=M}^{N} (b_k - b_{k-1}) g(x_k)$$

$$= \sum_{k=M}^{N} c_k g(x_k)$$

where $c_k = b_k - b_{k-1}$ is the jump of P^{dis} at x_k. Thus the Lebesgue-Stieltjes integral with respect to a discrete distribution reduces to the usual series sum. This means that both absolutely continuous and discrete distributions can be treated simultaneously and in a uniform manner.

Mixed distributions, that is, distributions which have both an absolutely continuous and a discrete component, are not uncommon. For instance, the failure distribution for light bulbs is of this type.

Another example, more closely related to the main theme of this work, is provided by jet aircraft engines. In particular, the failure distribution of turbine blades can be considered as a combined distribution. The absolutely continuous part is associated wtih failures which occur as a function of operating time or wear, and the discrete part is associated with the periodic stresses due to rapid temperature changes which occur in the blades during take-off operations.

Although a discrete distribution does not have a density function, there is the useful notion of a generalized density function which makes it possible to study densities of combined distribution functions in a unified way. We will use this concept in Section 3 and again in Section 6, but this is the logical place to introduce it.

Let $\delta(x)$ denote the Dirac delta function, a generalized function characterized by the property that if $\alpha < x_0 < \beta$ and $g(x)$ is a function, then

$$\int_{\alpha}^{\beta} g(x)\delta(x-x_0)dx = g(x_0) \quad . \qquad (2.27)$$

If c_k denotes the "jump" in $P^{dis}(x)$ at the discontinuity $x = x_k$, then

$$\int_{\alpha}^{\beta} g(x)\sum_{k} c_k\delta(x-x_k)dx = \sum_{k=M}^{N} c_k \int_{\alpha}^{\beta} g(x)\delta(x-x_k)dx$$

$$= \sum_{k=M}^{N} c_k g(x_k) \; ; \qquad (2.28)$$

hence, by eq. (2.23),

$$\int_{\alpha}^{\beta} g dP^{dis} = \int_{\alpha}^{\beta} g(x)\sum_{k} c_k\delta(x-x_k)dx \; , \qquad (2.29)$$

so

$$\frac{dP^{dis}}{dx} = \text{def} \sum_{k} c_k\delta(x-x_k) \qquad (2.30)$$

can be thought of as the generalized density function corresponding to the discrete distribution P^{dis}. This extension of the notation of density function makes it possible to study densities of combinations of absolutely continuous and discrete distributions in a uniform way.

2.4 The notion of independence of random variables will be of special importance in what follows because it will provide the means for reducing complex problems to tractable components. If $\underline{\underline{U}}_i$, $i=1,2,\ldots$ is a sequence of sets, Ω_i a collection of events on $\underline{\underline{U}}_i$, $\underline{\underline{P}}_i$ a probability measure on Ω_i, and f_i a random variable on $\underline{\underline{U}}_i$, then the products

$$\underline{\underline{U}} = \underline{\underline{U}}_1 \times \underline{\underline{U}}_2 \times \ldots , \tag{2.31.1}$$

$$\Omega = \Omega_1 \times \Omega_2 \times \ldots , \tag{2.31.2}$$

$$\underline{\underline{P}} = \underline{\underline{P}}_1 \times \underline{\underline{P}}_2 \times \ldots , \tag{2.31.3}$$

define a collection of events Ω on $\underline{\underline{U}}$ and an associated probability measure $\underline{\underline{P}}$. Notice that $\underline{\underline{P}}$ is merely a probability measure on sets and does not have anything to do with a particular random variable. The random variable f_i defined on $\underline{\underline{U}}_i$ can also be considered as a random variable on the product set $\underline{\underline{U}}$ by defining

$$f_i(\xi_1, \xi_2, \ldots, \xi_i, \ldots) = f_i(\xi_i) . \tag{2.32}$$

If f is a random variable on $\underline{\underline{U}}$ such that its distribution function P_f is the product of the distribution functions P_{f_i} of the random variable f_i on $\underline{\underline{U}}$, that is, if

$$P_f(\xi_1, \xi_2, \ldots) = P_{f_1}(\xi_1)P_{f_2}(\xi_2) \ldots , \tag{2.33}$$

then the f_i are said to be independent random variables. If $f_1, f_2, \ldots$ are independent random variables, then the mean of $f=f_1 f_2 \ldots$ is given by (cp. eq. (2.9))

$$\bar{f} = \overline{(f_1 f_2 \ldots)} = \int f d\underline{\underline{P}} = P_f(\infty) = P_{f_1}(\infty) P_{f_2}(\infty) \ldots$$

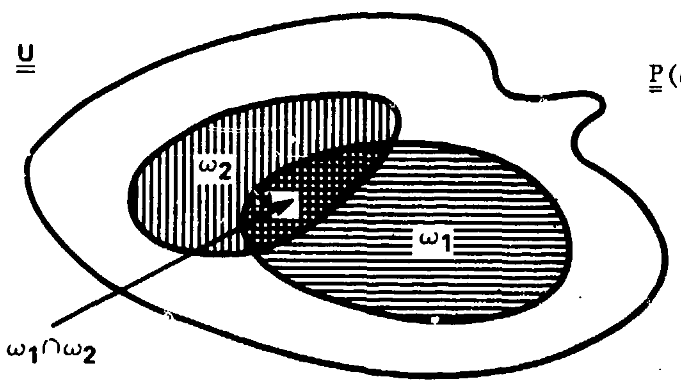
$$= \bar{f}_1 \bar{f}_2 \ldots ; \tag{2.34}$$

that is, the mean value of a product of independent random variables is the product of their mean values.

2.5 In the theory of reliability and maintenance the notion of the conditional probability of survival plays a central role. If ω_1 and ω_2 are two events and if $\underline{\underline{P}}(\omega_1) > 0$, that is, the probability of event ω_1 is positive, then the conditional probability of ω_2 given ω_1 is

$$\underline{\underline{P}}(\omega_2/\omega_1) =_{\text{def}} \frac{\underline{\underline{P}}(\omega_2 \cap \omega_1)}{\underline{\underline{P}}(\omega_1)} . \tag{2.35}$$

If areas of sets are used to represent probabilities, then, in Figure 2.5, $\underline{\underline{P}}(\omega_2/\omega_1)$ can be interpreted as the ratio of the area of the region $\omega_2 \cap \omega_1$ to the area of the region ω_1.



$$\underline{\underline{P}}(\omega_2/\omega_1) = \frac{\underline{\underline{P}}(\omega_2 \cap \omega_1)}{\underline{\underline{P}}(\omega_1)}$$

Figure 2.5. Conditional Probability

Related to this interpretation of conditional probability is the important Bayes' Principle of Inverse Probability. Suppose that $\omega_1, \omega_2, \omega_3$ are three events in $\underline{\underline{U}}$ which have a non-empty intersection. The situation is depicted in Figure 2.6.

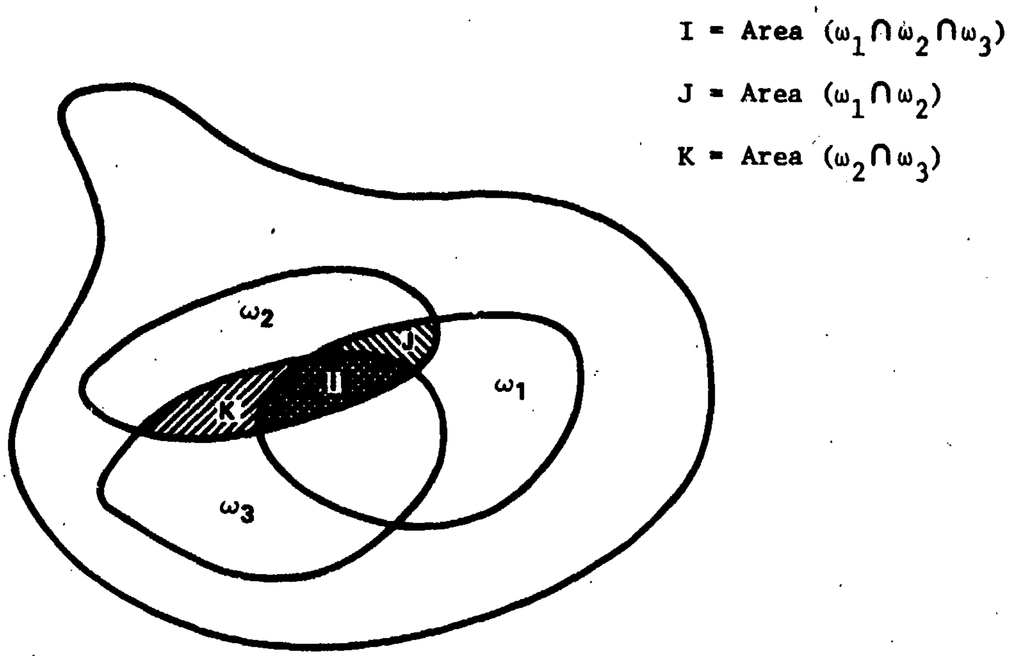


Figure 2.6. Bayes' Principle

Continuing to interpret the probability of an event ω as the "area" of the set ω in the figure, let

$$I = \text{Area } (\omega_1 \cap \omega_2 \cap \omega_3) = \underline{P}(\omega_1 \cap \omega_2 \cap \omega_3) ,$$

$$J = \text{Area } (\omega_1 \cap \omega_2) = \underline{P}(\omega_1 \cap \omega_2) ,$$

$$K = \text{Area } (\omega_2 \cap \omega_3) = \underline{P}(\omega_2 \cap \omega_3) .$$

The numerical ratio $I/(J \cdot K)$ can be expressed in two different ways:

$$I/(J \cdot K) = (I/J)/K = (I/K)/J , \tag{2.36}$$

that is,

$$\underline{P}(\omega_1 | \omega_2 \cap \omega_3)/\underline{P}(\omega_1 \cap \omega_2) = \underline{P}(\omega_3 | \omega_1 \cap \omega_2)/\underline{P}(\omega_2 \cap \omega_3) ,$$

or equivalently,

$$\underline{P}(\omega_1 | \omega_2 \cap \omega_3) = \frac{\underline{P}(\omega_3 | \omega_1 \cap \omega_2)\underline{P}(\omega_1 \cap \omega_2)}{\underline{P}(\omega_2 \cap \omega_3)} . \tag{2.37}$$

In order to interpret eq. (2.36) in a manner appropriate for our later needs, we will discuss two types of events: observations and hypotheses. Statistical inference generally proceeds from a collection of hypotheses, whose probabilities are assumed known, to assessments or predictions of the probability of various observations. This procedure can be inverted to provide assessments of the probability of various hypotheses when a collection of observations is given. Adopting the latter viewpoint, let $\{\omega_i : i \in \mathcal{J}\}$ be a fixed collection of hypotheses and set

$$\omega = \bigcap_{i \in \mathcal{J}} \omega_i \; ; \tag{2.38}$$

ω is the event which corresponds to the simultaneous validity of all hypotheses ω_i. Let H denote another hypothesis and σ an observation. Then eq. (2.37) can be rewritten, using this notation, as

$$\underline{\underline{P}}(H|\sigma \cap \omega) = \frac{\underline{\underline{P}}(\sigma|H \cap \omega)\underline{\underline{P}}(H \cap \omega)}{\underline{\underline{P}}(\sigma \cap \omega)} \tag{2.39}$$

The quantity $\underline{\underline{P}}(H|\sigma \cap \omega)$ is called the likelihood ratio for the hypothesis H given the observation σ and the fixed collection of hypotheses $\{\omega_i : i \in \mathcal{J}\}$. The likelihood ratio is proportional to the probability of the observation given the hypothesis H (and ω) multiplied by the a priori probability of H (and ω). The factor of proportionality is independent of the collection of alternative hypotheses H under consideration. Therefore, it is reasonable to select that H from among a collection of alternative hypotheses for which the product $\underline{\underline{P}}(\sigma|H \cap \omega)\underline{\underline{P}}(H \cap \omega)$, and hence the likelihood ratio $\underline{\underline{P}}(H|\sigma \cap \omega)$, is maximal. This is Bayes' Principle. It can be considered as a generalization of the well-known Maximum-Likelihood method of estimation of parameters [4], [13]. In the latter, the event H is a set of values of the parameters of a probability distribution; $\omega = H \cup \sigma$, and σ is the event which consists of independent observations $x_1, x_2, \ldots, x_n$ of a random variable x. If it is assumed that $\underline{\underline{P}}(H)$ is independent of H, then the likelihood ratio is proportional to

$$\underline{P}(\sigma|H) = \prod_{i=1}^{n} \underline{P}(x_i|H) \quad , \tag{2.40}$$

where the right-hand side uses a suggestive if not quite exact notation. The right-hand side is called the likelihood function. Bayes' Principle applied to this special case is the Maximum-Likelihood method.

Notice that Bayes' Principle is a consequence of the symmetry inherent in the definition of conditional probability (as exhibited in eq. (2.36) and the triple intersection displayed in Figure 2.6), and the symmetrical interpretation of hypotheses and observations as events. Thus there is a certain degree of interchangeability of hypotheses and observations. Hypotheses which remain unchallenged by observations assume, as experience accumulates, the role and properties of observations themselves, and observations (considered as events) can be converted to hypotheses in the right circumstances.

This interchangeability, or substitutability, plays an unheralded but substantial role in the practical analysis of the reliability of rapidly evolving complex systems, for which only small sample observations can ever be available. Modern commercial and military aircraft provide an example. The relatively small production runs and the very small number of aircraft of any one type which reach high operating times preclude the possibility of collecting extensive actuarial data for the assessment of hypotheses concerning reliability. This difficulty is mitigated to some extent by making hypotheses (concerned, e.g., with Hard Time maintenance intervals) based upon prior experience with similar although by no means identical equipment. In this way prior limited observational experience is transformed into current working hypotheses against which current observations, limited though they may be, are tested. In turn, these observations form the foundation for, and in the sense described above, are equivalent to, future hypotheses. Although this application of Bayes' Principle is rarely made explicit and quantitative — one speaks instead of the need for "experienced" reliability analysts — it nevertheless plays a major role in the practical analysis of complex systems which evolve with time, have a relatively brief life, and of which only a small number of replicas are fabricated.

3. TERMINOLOGY OF RELIABILITY THEORY

3.1 The basic concept in reliability theory is that of the probability of failure or, if one prefers a more sanguine outlook, the probability of survival, frequently called the reliability. For this application we may think of the set $\underline{\underline{U}}$ as a universe of items or components ξ whose failure characteristics are of interest to us. The subsets of $\underline{\underline{U}}$ which constitute events consist of those $\xi \epsilon \underline{\underline{U}}$ which have failed prior to a given time. Thus, if t denotes time, then the collection Ω of events consists of the subsets

$$\begin{aligned} \omega(t) &= \{\xi \epsilon \underline{\underline{U}} : \xi \text{ has failed prior to } t\} \; ; \\ \Omega &= \{\omega(t) : t \epsilon \mathbb{R}\} \; . \end{aligned} \qquad (3.1)$$

Recall that $\mathbb{R}$ denotes the set of all real numbers. It is evident that $t_1 < t_2$ implies $\omega(t_1) \subset \omega(t_2)$ since each item which failed prior to t_1 certainly failed prior to t_2. This is illustrated in Figure 3.1, which is the same as Figure 2.2 although it has a different interpretation. In this way the collection of events Ω is parametrized by the real-valued time variable.

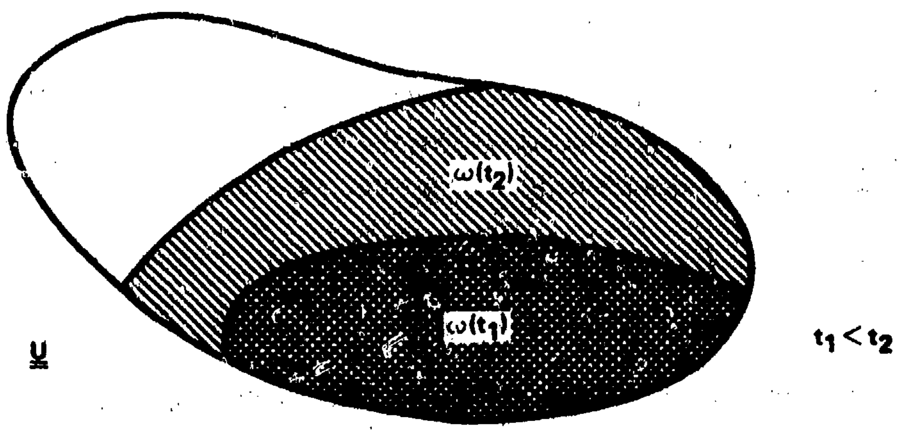


Figure 3.1. Sets of Failed Items

Associated with the universe $\underline{U}$ of items and the collection Ω of events is a probability measure $\underline{F}$ which expresses the probability of failure corresponding to events $\omega \in \Omega$ (We may think of $\underline{F}(\omega(t))$ as the "area" occupied by the event $\omega(t)$ if we interpret probabilities as areas (e.g., in Figure 3.1) and recall that the total "area" of $\underline{U}$ itself, considered as an event in Ω, must be equal to 1). With this interpretation, $\underline{F}(\omega(t))$ is the <u>probability of failure prior to time</u> t; the probability of survival associated with the event $\omega = \omega(t)$ is

$$\underline{R}(\omega(t)) = 1 - \underline{F}(\omega(t)) , \tag{3.2}$$

so $\underline{R}(\omega(t))$ is the <u>probability of survival until time</u> t, also called the <u>reliability</u>. $R(\omega(t))$ can be interpreted as the "area" of $\underline{U} - \omega(t)$.

If $\underline{U}$ consists of N items, if the number of items in $\omega(t)$ is $N(t)$, and if the measure $\underline{F}$ is counting measure, then, since $N(t)$ is the number of items which failed prior to t,

$$F(\omega(t)) = \frac{N(t)}{N} . \tag{3.3}$$

In order to transfer the above notions from the realm of sets to the realm of numbers, where the methods of calculus can be applied, we use the indicator function defined by eq. (2.7) to obtain a failure distribution function. Recall that

$$c_{\omega(t)}(\xi) = \begin{cases} 1 & \text{if } \xi \in \omega(t) \\ 0 & \text{if } \xi \notin \omega(t) . \end{cases} \tag{3.4}$$

The function which assigns to each event ω the number 1 is a random variable. The corresponding distribution function associated with the failure probability measure $\underline{F}$ is, by eq. (2.12),

$$F(t) = \int_{\omega(t)} d\underline{F} = \int c_{\omega(t)} d\underline{F} = \underline{F}(\omega(t)) . \tag{3.5}$$

Thus the failure distribution $F(t)$ is merely the probability measure $\underline{F}$ considered as a function of the time parameter t.

We can calculate the survival distribution R(t) in a similar way, but it is easier to use eq. (3.2) directly to obtain

$$R(t) = 1 - F(t) . \tag{3.6}$$

Notice that F really is a probability distribution in the sense specified by Eq. (2.13), but that R(t) has slightly different properties:

$$R(t) \text{ is non-increasing;} \tag{3.7.1}$$

$$R(t) = R(t+0) ; \tag{3.7.2}$$

$$R(-\infty) = 1 , \quad R(+\infty) = 0 ; \tag{3.7.3}$$

each of these properties follows immediately from the defining relation eq. (3.6) and the corresponding eq. (2.13). R(t) will be referred to as a survival distribution even though it is not a distribution in the technical sense.

The graph of the function $t \rightarrow R(t)$ is called a survival curve. Figure 4.2 of Section 4 exhibits a typical survival curve for an aircraft gas turbine engine.

In practice, measurements and observations are always discrete and finite in number. This means that actual worldly knowledge of survival and other probability distributions only supplies an approximation which (may be exact and) is a discrete distribution. On the other hand, theory and philosophical beliefs about the nature of reality often suggest that observations are discrete sets of values drawn from absolutely continuous distributions or from combinations of absolutely continuous and discrete distributions; moreover, the techniques of mathematical analyses are more highly developed for studying absolutely continuous distributions. Consequently, whenever it is possible to do so, it is desirable to suppose that observations have been drawn from ideal and hypothetical absolutely continuous distributions. The density functions corresponding to these distributions play a central role in most developments of the subject. The absolutely continuous distributions which have been found to be most useful in practice and are most extensively studied by theorists will be introduced in Section 4.

The generalized density function corresponding to the failure distribution $F(t) = F_f(t)$ (which may consist of both an absolutely continuous part and a discrete part) is denoted $\rho(t)$; that is

$$\rho(t) = \frac{dF}{dt} \tag{3.8}$$

is the <u>failure probability density</u>. Equation (2.30) must be used to express the generalized density function for the discrete part of F. From Eq. (3.6) we see that the <u>survival probability density</u> is given by

$$\frac{dR}{dt} = -\rho(t) \ . \tag{3.9}$$

Corresponding to survival and failure distributions are <u>conditional survival</u> and <u>conditional failure</u> distributions. First consider the conditional probability of survival. According to eq. (2.35), the conditional probability of the event ω_2 given ω_1 is

$$\underline{\underline{R}}(\omega_2|\omega_1) = \frac{\underline{\underline{R}}(\omega_2 \cap \omega_1)}{\underline{\underline{R}}(\omega_1)} \tag{3.10}$$

F distributions parametrized by survival time we consider two times, t_1 t_2. Then the definition of $\omega(t)$, eq. (3.1), implies $\omega(t_1) \subset \omega(t_2)$ so the complementary events satisfy the reverse inclusion, i.e.,

$$\underline{\underline{U}} - \omega(t_1) \supset \underline{\underline{U}} - \omega(t_2) \ .$$

The formula $R(t) = 1 - F(t)$ implies that $R(t) = \underline{\underline{R}}(\underline{\underline{U}} - \omega(t))$. Introduce $\omega_1 = \underline{\underline{U}} - \omega_1(t)$, $\omega_2 = \underline{\underline{U}} - \omega_2(t)$. Then $\omega_2 \subset \omega_1$: items in ω_2 have survived at least until t_2 whereas items in ω_1 have survived at least until t_1 (cp. Figure 3.2). Now we can compute the conditional probability that an item will survive at least until t_2 given that it has survived until t_1, where $t_1 < t_2$. From eq. (3.10), this is

$\omega_1 = \underline{\underline{U}} - \omega(t_1)$

$\omega_2 = \underline{\underline{U}} - \omega(t_2)$

$\omega(t_1) \subset \omega(t_2)$

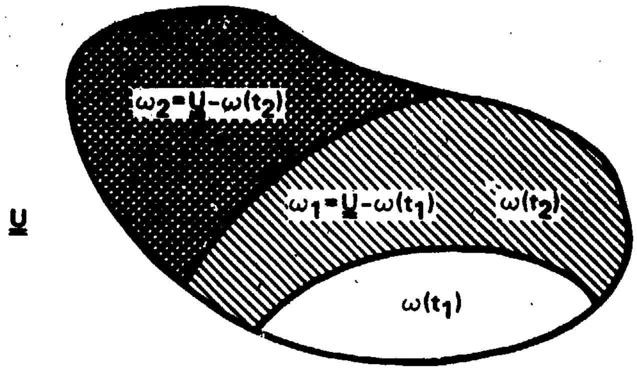


Figure 3.2. Conditional Probability of Survival

$$R(t_2|t_1) = \underline{\underline{R}}(\omega_2|\omega_1) \tag{3.11}$$

$$= \frac{\underline{\underline{R}}(\omega_2)}{\underline{\underline{R}}(\omega_1)} \quad \text{since} \quad \omega_2 \subset \omega_1 \ ,$$

$$= \frac{R(t_2)}{R(t_1)} \ .$$

We are to understand that t_1, and hence the condition $\omega(t_1)$, is held fixed and only t_2 varies (through values greater than t_1). The expression eq. (3.11) for the conditional probability amounts to the same as the assumption that the universe of items has been reduced from $\underline{\underline{U}}$ to ω_1 (cp. Figure 3.2), and that the probabilities have been renormalized by division by $\underline{\underline{R}}(\omega_1)$ so that the total measure of ω_1 is adjusted to equal 1.

The conditional survival density is therefore obtained by differentiating the numerator of eq. (3.11) at $t_2 = t$, which yields

$$\frac{dR(t|t_1)}{dt} = \frac{-\rho(t)}{R(t_1)} \ . \tag{3.12}$$

The conditional probability of failure must be treated slightly differently since in order to fail during the interval (t_1, t_2), an item must first have survived until t_1. Therefore, the conditional probability of failure prior to t_2 given survival until t_1 is

$$F(t_2|t_1) = \frac{F(t_2) - F(t_1)}{1 - F(t_1)} = \frac{F(t_2) - F(t_1)}{R(t_1)} \quad ; \tag{3.13}$$

the corresponding density at $t_2 = t_1 = t$ is usually called the hazard rate (also often the failure rate) and is expressed by

$$\eta(t) = \frac{\rho(t)}{R(t)} \quad . \tag{3.14}$$

By utilizing eq. (3.9) the hazard rate can be expressed in terms of the survival distribution as

$$\eta(t) = \frac{\rho(t)}{R(t)} = -\frac{d}{dt}\log_e R(t) \quad ,$$

(where $\log_e$ denotes the natural logarithm function), and the survival distribution is given in terms of the hazard rate by

$$R(t) = \exp\left(-\int_{-\infty}^{t} \eta(x)dx\right) \tag{3.16}$$

(where $\exp x = e^x$). Formulas (3.15) and (3.16) are valid for absolutely continuous survival distributions. For discrete distributions eq. (3.15) must be replaced by the corresponding finite-difference formulation.

Suppose that an item consists of various parts, and survives only if all of its parts survive. If the survival distribution of the item is $R(t)$ and that of the k^{th} part is $R_k(t)$, and if failure of the various parts is due to independent causes, then

$$R(t) = \prod_k R_k(t) \quad . \tag{3.17}$$

In this case the hazard rate is

$$\eta(t) = -\frac{d}{dt}\left(\log_e \prod R_k(t)\right)$$

$$= \sum_k -\frac{d}{dt} \log_e R_k(t) \quad ,$$

that is

$$\eta(t) = \sum_k \eta_k(t) \tag{3.18}$$

where

$$\eta_k(t) = -\frac{d}{dt} \log_e R_k(t) \tag{3.19}$$

is the hazard rate of the k^{th} part. Thus, the hazard rate is additive for independent causes of failure. This result is valid for discrete as well as absolutely continuous survival distributions. This convenient property permits independent assessment of constituent hazard rates and provides a simple method for combining them, by means of eqs. (3.18) and (3.16), to recover $R(t)$ itself.

3.2 If R is a survival distribution, then the area under the graph of $t \mapsto R(t)$ is the mean lifetime of the items $\xi \in \underline{U}$. In order to adapt our notation to items which begin life at $t = 0$, let us suppose that R is defined on $[0, \infty)$ and that $R(0) = 1$, $\lim_{t\to\infty} tR(t) = 0$. The area under the graph is $\int_0^\infty R(t)dt$. Integration by parts yields

$$\int_0^\infty R(t)dt = tR(t)\Big|_0^\infty - \int_1^\infty tdR$$

$$= -\int_1^0 tdR = \int_0^1 tdR \quad ,$$

since $\lim_{t \to \infty} tR(t) = 0$ by hypothesis and R varies from 1 to 0 as t varies from 0 to ∞. Now recall from eq. (2.9) that

$$\int_0^1 t\,dR = \int t\,d\underline{R} = \bar{t} ,$$

the mean value of the random variable t, thus the mean time before failure. Graphically this result amounts to nothing more than evaluating the area under the graph of $t \mapsto R(t)$ by integrating along the R-axis as indicated in Figure 3.3.

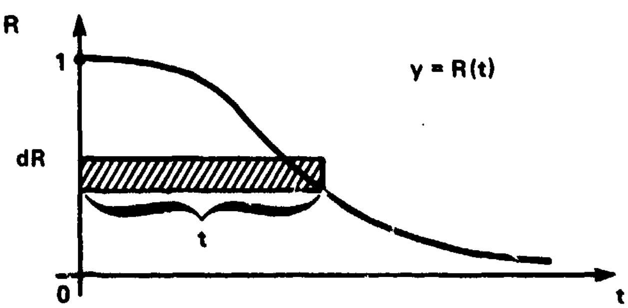


Figure 3.3. Calculation of Mean Time Before Failure

4. USEFUL SURVIVAL DISTRIBUTIONS

The probability-of-survival distributions most commonly used in the practical analysis of reliability data are also among those distributions which have been most intensively studied by theoreticians. They are the

1) Exponential,
2) Normal (also called Gaussian),

and 3) Weibull

distributions. In addition, the

4) Lognormal

and 5) Gamma

distributions have played significant roles. We will define each of these and derive the corresponding density and hazard functions. Since all of these distributions are absolutely continuous, the usual techniques of the calculus can be employed.

It will be assumed hereafter that a survival distribution is defined on some closed half-infinite interval, which will generally be $0 \le t < \infty$.

4.1 Exponential Survival Distribution

For this distribution the probability of survival to time t is

$$R(t) = \exp(-\lambda t), \quad \lambda > 0, \quad t \ge 0 \quad . \tag{4.1}$$

Observe that $\lim_{t\to\infty} R(t) = R(\infty) = 0$ implies $\lambda > 0$. Figure 4.1.1 illustrates the graph of a typical exponential distribution.

The exponential survival density corresponding to eq. (4.1) is

$$\rho(t) = -\frac{dR}{dt} = \lambda \exp(-\lambda t) \quad , \tag{4.2}$$

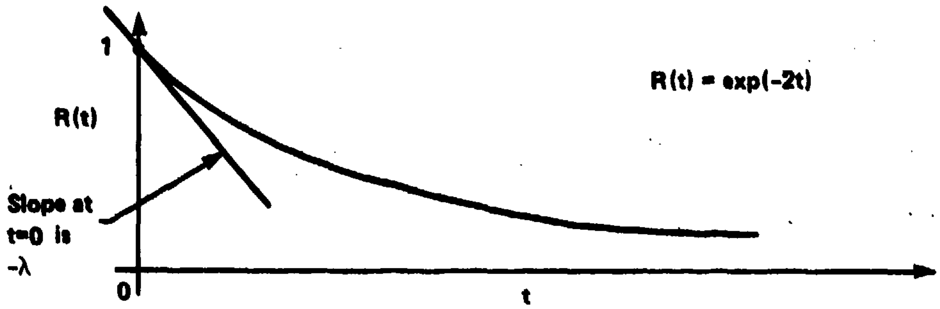


Fig. 4.1.1

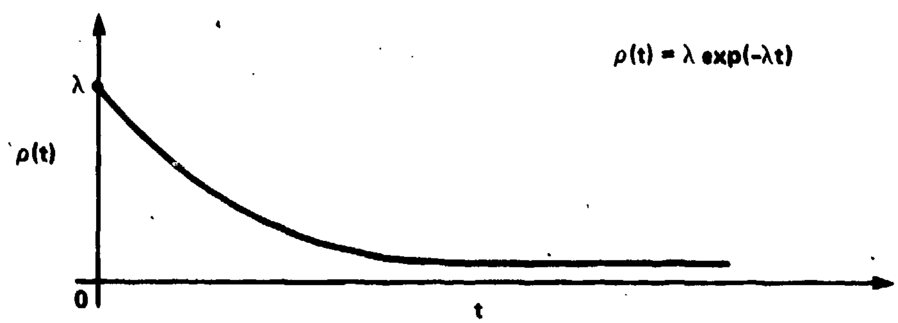


Fig. 4.1.2

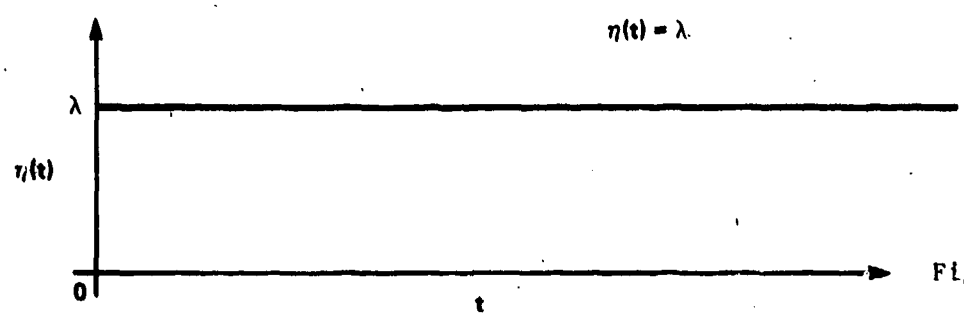


Fig. 4.1.3

Figure 4.1. Exponential Survival Distribution, Density, and Hazard Rate

and the hazard rate is

$$\eta(t) = -\frac{d\,\log_e R(t)}{dt} = \lambda\ ;$$

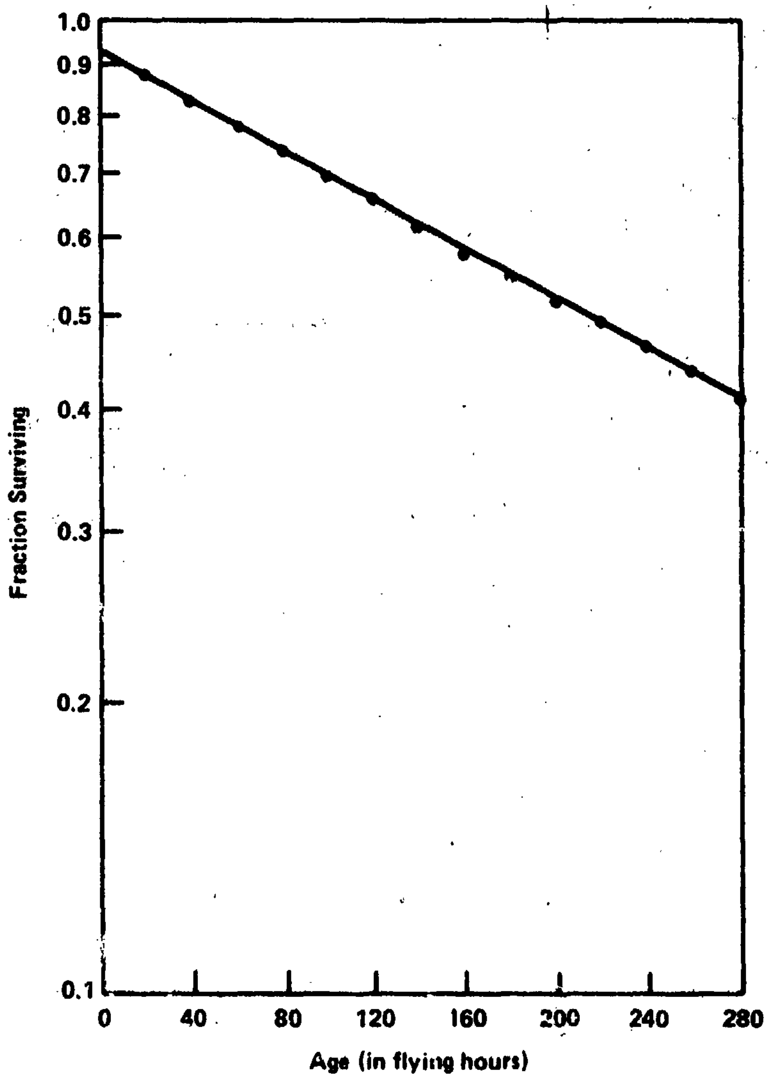
cp Figures 4.1.2 and 4.1.3.

Figure 4.2 displays survival data for the J65-W-3 jet engine. Semi-logarithmic graph paper is used so that the graph of an exponential distribution appears as a straight line. In this example the data points lie close to the line shown: the underlying distribution can be accurately approximated by an exponential. Were the exponential of the form eq. (4.1), then we would find $R(0) = 1$, but, the data indicates that at $t = 0$ (that is, upon initial operation) approximately 6% of the items were found to be in a failed condition. The variety of potential meanings and definitions of the term "failed condition" have been explored at length earlier in this volume. Regardless of the precise meaning attributed to the term, the phenomenon can be interpreted as reflecting manufacturing defects which have escaped test procedures as well as failures induced by pre-operational tests or aspects of the production process itself which are not detectable (or at least not detected) until initial operation is attempted at $t = 0$. This phenomenon is accommodated in the mathematical formalism by the simple expedient of replacing the time variable t by $t + t_0$, where t_0 can be thought of as corresponding to the duration of pre-operational exposure of the item. This problem is not confined to exponentially failing items; it is found for all types of distributions. The solution is always the same: renormalization of the zero time by replacement of t by $t + t_0$ for an appropriate positive t_0. Thus the renormalization exponential distribution (also called "shifted exponential distribution) is

$$R(t) = \exp(-\lambda\,(t + t_0)) \tag{4.4}$$

with corresponding density

$$\rho(t) = \lambda \exp(-\lambda\,(t + t_0)) \tag{4.5}$$



SOURCE: United States Air Force, ***Procedures for Determining Aircraft Engine (Propulsion Unit) Failure Rates, Actuarial Engine Life, and Forecasting Monthly Engine Changes by the Actuarial Method***, Technical Order, TO 00-25-128, October 20, 1959.

Figure 4.2. Typical Exponential Survival Distribution: J65-W-3 Jet Engine (semi-logarithmic graph paper)

and hazard rate

$$\eta(t) = \lambda \quad . \tag{4.6}$$

Note the important fact that the hazard rate is independent of t_0.

The exponential distribution plays a special role in the theory of reliability for two quite different reasons. The first is a practical one: it has been found that an exponential distribution characterizes the life history of a variety of equipment types, generally including electronic devices and complex equipment. The second reason is a theoretical one, and in some respects it is the more basic. Because it has a constant hazard rate, the exponential distribution separates the survival distributions which have an increasing hazard rate from those which have a decreasing hazard rate, and it therefore also separates two completely different types of maintenance policies.

It is clear that if an item has a non-increasing hazard rate, then there is no advantage gained in replacing that item by a new one at any time prior to its failure. Indeed, if the hazard rate is strictly decreasing with time, then replacement substitutes an item with a greater probability of failure for the one already in operation. If, however, the hazard rate is strictly increasing, then replacement of the item by a new one will increase the probability of survival. In this case the main issue is the cost of replacement maintenance, and a principal mathematical problem is to determine replacement intervals which are optimal with respect to some mix of acceptable failure rate and maintenance cost. The exponential distribution separates these fundamentally different classes of survival distributions and maintenance policies. This is specially fortunate because the exponential distribution has particularly simple mathematical properties which often make it possible to carry out technical analyses in complete and rigorous detail, thereby obtaining lower or upper bounds for the properties of general non-decreasing or non-increasing survival distributions. This is one main reason why a large portion of the reliability theory literature is devoted to the study of systems whose constituents have exponential distributions.

The above remarks can be made more precise. Let $R(t)$ be a survival distribution with corresponding hazard rate $\eta(t)$, and suppose that η is either non-increasing or non-decreasing for $0 \leq t \leq T$. Then

$$\eta(t) \leq \eta(0) \qquad \text{or} \tag{4.7}$$

$$\eta(t) \geq \eta(0) \tag{4.8}$$

according as η is non-increasing or non-decreasing, respectively. Hence

$$-\frac{d \log_e R(t)}{dt} = \eta(t) \begin{smallmatrix} \leq \\ \geq \end{smallmatrix} \eta(0) \tag{4.9}$$

implies

$$\log_e R(t) \begin{smallmatrix} \geq \\ \leq \end{smallmatrix} \int_0^t - \eta(0) dt = - \eta(0) t \tag{4.10}$$

for $t \leq T$, where the upper (respectively, lower) inequality corresponds to a non-increasing (respectively, non-decreasing) hazard rate. Since exponentiation preserves the direction of an inequality, we find

$$R(t) \geq \exp\,(- \eta(0)t) \quad , \quad 0 \leq t \leq T \tag{4.11}$$

if the hazard rate $\eta(t)$ corresponding to $R(t)$ is non-increasing for $0 \leq t \leq T$, and

$$R(t) \leq \exp\,(- \eta(0)t) \quad , \quad 0 \leq t \leq T \tag{4.12}$$

if $\eta(t)$ is non-decreasing on $0 \leq t \leq T$, as claimed.

That an easily analyzed survival distribution separates the two classes is an important and useful fact. But there is an unexpected bonus: the single parameter $\lambda = \eta(0)$, which specifies the exponential distribution, is equal to $\frac{dR}{dt}(0)/R(0)$ (even in the time renormalized form $R(t) = \exp\,(- \lambda(t + t_0))$, and consequently it can be estimated from data collected during the early life history of the item. Therefore, if there are reasons to believe that $\eta(t)$ is either non-increasing or non-decreasing, then R(t) can be bounded by an exponential distribution whose parameter (= hazard rate) can be realistically estimated.

4.2 Normal Survival Distribution

This distribution is also frequently encountered in applications. If the mean of the normal distribution is positive and large in comparison with the standard deviation, then truncation by restricting its domain to the set of non-negative t will not result in practical difficulties. Otherwise, the truncated distribution must be normalized to ensure that $R(0) = 1$. To do this, define

$$A = \frac{1}{\sigma^* \sqrt{2\pi}} \int_0^\infty \exp\left(-\frac{1}{2}\left(\frac{u - t^*}{\sigma^*}\right)^2\right) du \quad . \tag{4.13}$$

Then the truncated normal survival distribution is

$$R(t) = \frac{1}{A\sigma^* \sqrt{2\pi}} \int_t^\infty \exp\left(-\frac{1}{2}\left(\frac{u - t^*}{\sigma^*}\right)^2\right) du, \qquad t \geq 0 \tag{4.14}$$

where t^* and σ^* are, respectively, the mean and standard deviation of the untruncated normal distribution. The associated truncated normal survival density function is

$$\rho(t) = -\frac{dR}{dt} = \frac{1}{A\sigma^* \sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{t - t^*}{\sigma^*}\right)^2\right) \quad , \tag{4.15}$$

and the truncated normal hazard rate is

$$\eta(t) = \frac{\exp\left(-\frac{1}{2}\left(\frac{t - t^*}{\sigma^*}\right)^2\right)}{\int_t^\infty \exp\left(-\frac{1}{2}\left(\frac{u - t^*}{\sigma^*}\right)^2\right) du} \quad . \tag{4.16}$$

Notice that the hazard rate is independent of the truncation normalization factor A. These functions and an application are illustrated in Figures 4.3 and 4.4.

$$R(t) = \frac{1}{A\sigma\sqrt{2\pi}} \int_{t}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{u-\bar{t}}{\tau}\right)^2\right) du$$

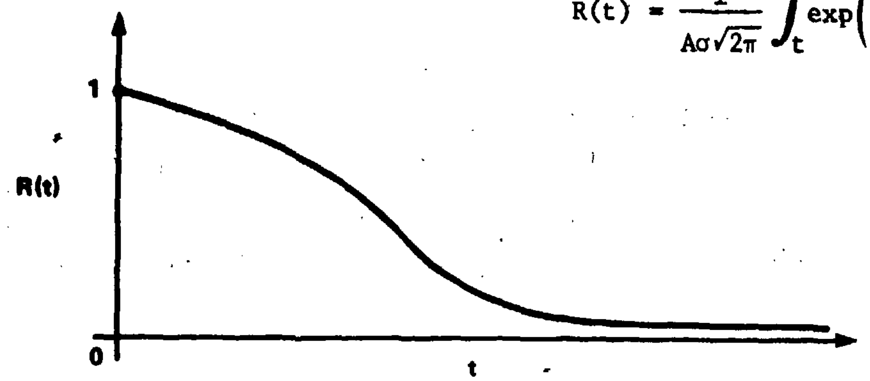


Fig. 4.3.1

$$\rho(t) = \frac{1}{A\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{t-\bar{t}}{\sigma}\right)^2\right)$$

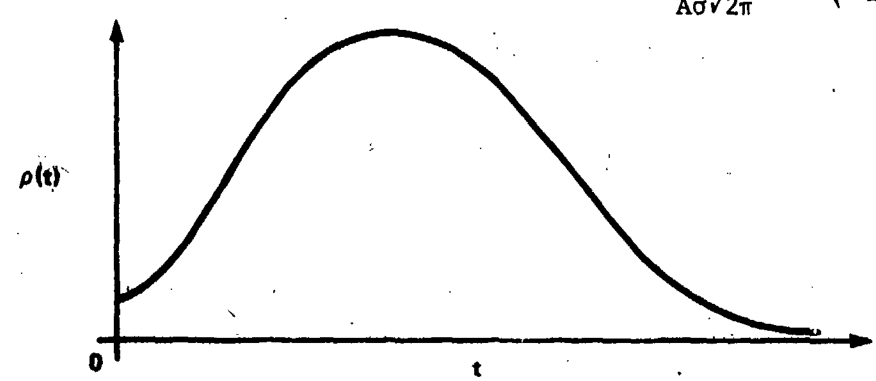
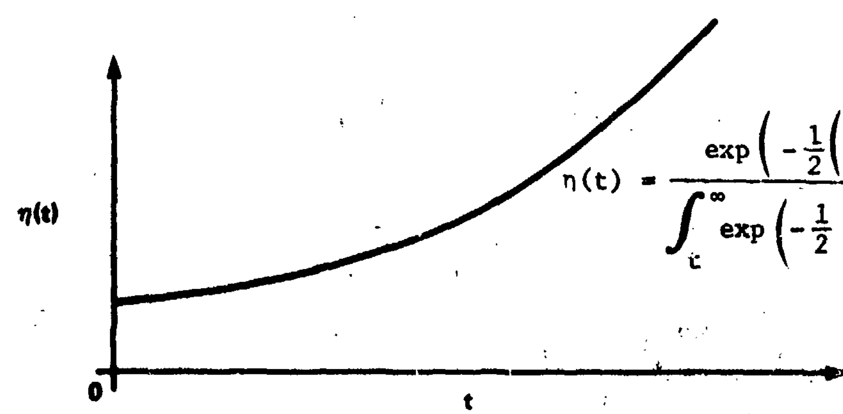


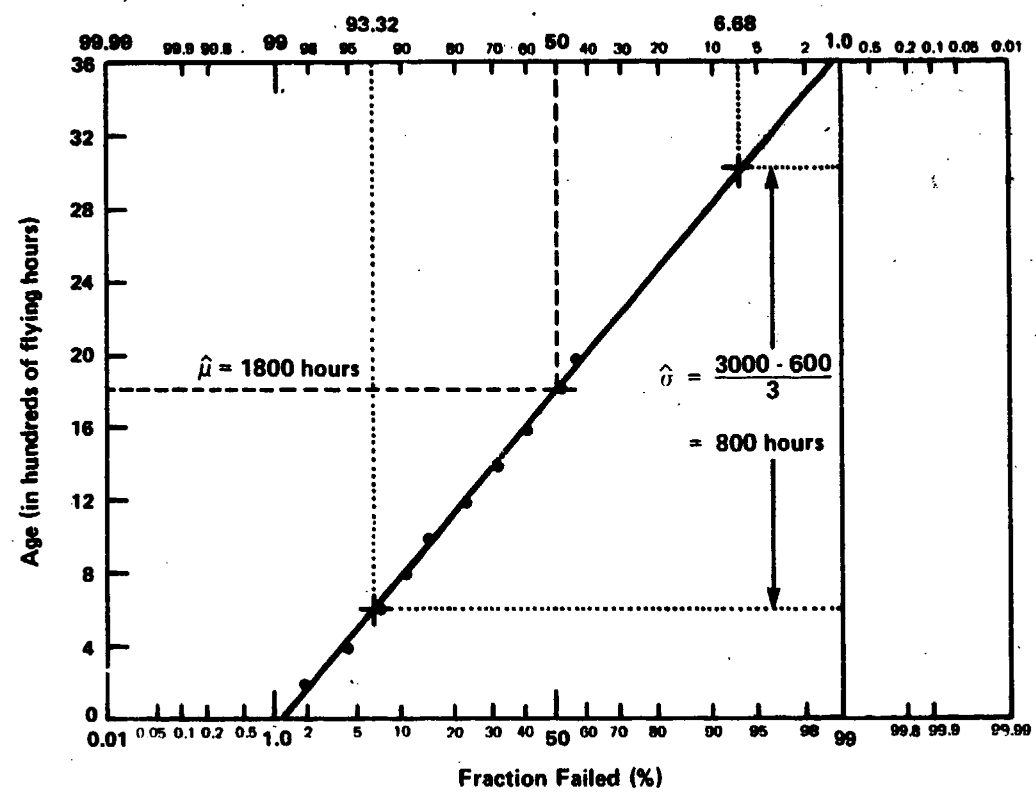
Fig. 4.3.2



$$\eta(t) = \frac{\exp\left(-\frac{1}{2}\left(\frac{t-\bar{t}}{\sigma}\right)^2\right)}{\int_{t}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{u-\bar{t}}{\sigma}\right)^2\right) du}$$

Fig. 4.3.3

Figure 4.3. Truncated Normal Distribution, Density, and Hazard Rate



SOURCE: United States Air Force, *Procedures for Determining Aircraft Engine (Propulsion Unit) Failure Rates, Actuarial Engine Life, and Forecasting Monthly Engine Changes by the Actuarial Method*, Technical Order, TO 00-25-128, October 20, 1959.

NOTE: $\hat{\mu}$ and $\hat{\sigma}$ are maximum likelihood estimates of μ and σ. The Appendix describes the techniques used to obtain them. Using these estimates, a chi-square goodness-of-fit test was performed. The hypothesis of normality could not be rejected at the 20-percent significance level. Results differing from these by less than 1 percent were given in a curve fitted by the rules of E. B. Ferrell, "Plotting Experimental Data on Normal or Log-Normal Probability Paper," *Industrial Quality Control*, Vol. 15, 1958, pp. 12-15.

Figure 4.4. Truncated Normal Survival Distribution: J57-F-59 and J57-P59 Jet Engines (Normal probability graph paper)

4.3 Weibull Survival Distribution

The Weibull distribution was introduced in 1951 by the Swedish statistician Waloddi Weibull in order to describe the tensile strength of steel [14]. It has since been applied to a variety of reliability problems. The Weibull survival distribution is defined on $0 \leq t < \infty$ and assumes the form

$$R(t) = \exp(-\lambda t^{s}) \quad , \quad \lambda > 0 \quad , \quad s > 0 \quad . \tag{4.17}$$

The corresponding Weibull survival density function is

$$\rho(t) = -\frac{dR}{dt} = \lambda s t^{s-1} \exp(-\lambda t^{s}) \quad , \tag{4.18}$$

and the Weibull hazard rate takes the form

$$\eta(t) = \lambda s t^{s-1} \quad . \tag{4.19}$$

Observe that if $s = 1$, then the Weibull distribution reduces to the exponential distribution with parameter λ. The hazard rate is increasing if $s > 1$ and decreasing if $0 < s < 1$. One can think of the Weibull hazard function as the best power-function approximation to an arbitrary continuous hazard rate in a neighborhood of $t = 0$.

Graphs of the Weibull distribution, density, and hazard rate and an application appear in Figures 4.5 and 4.6.

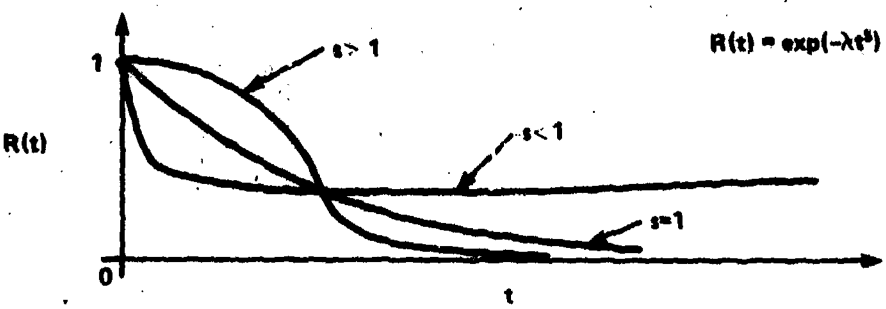


Fig. 4.5.1

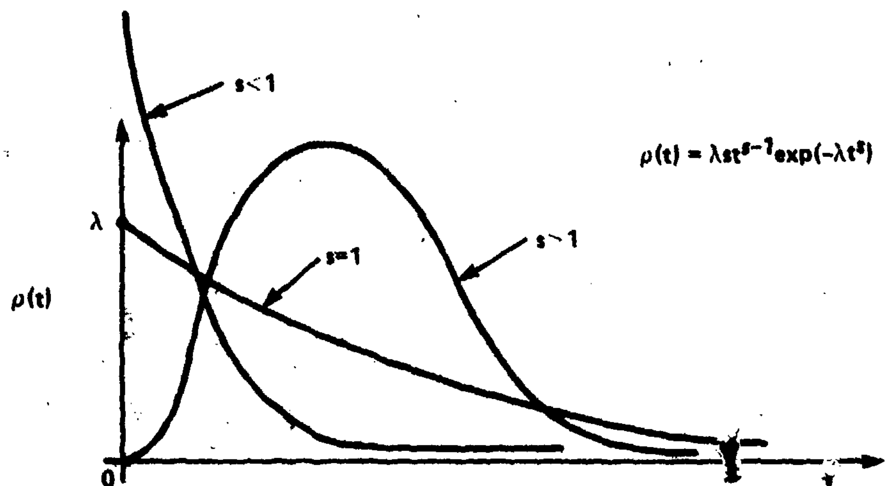


Fig. 4.5.2

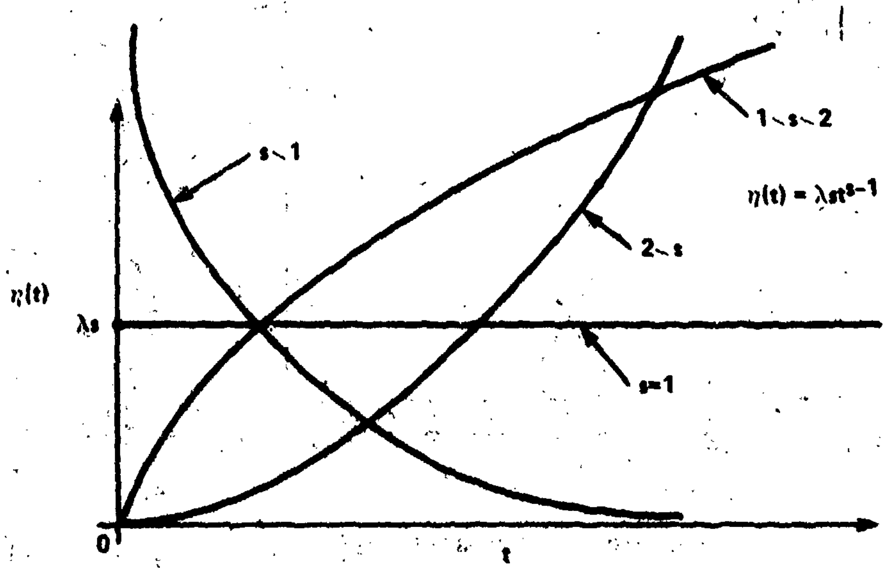
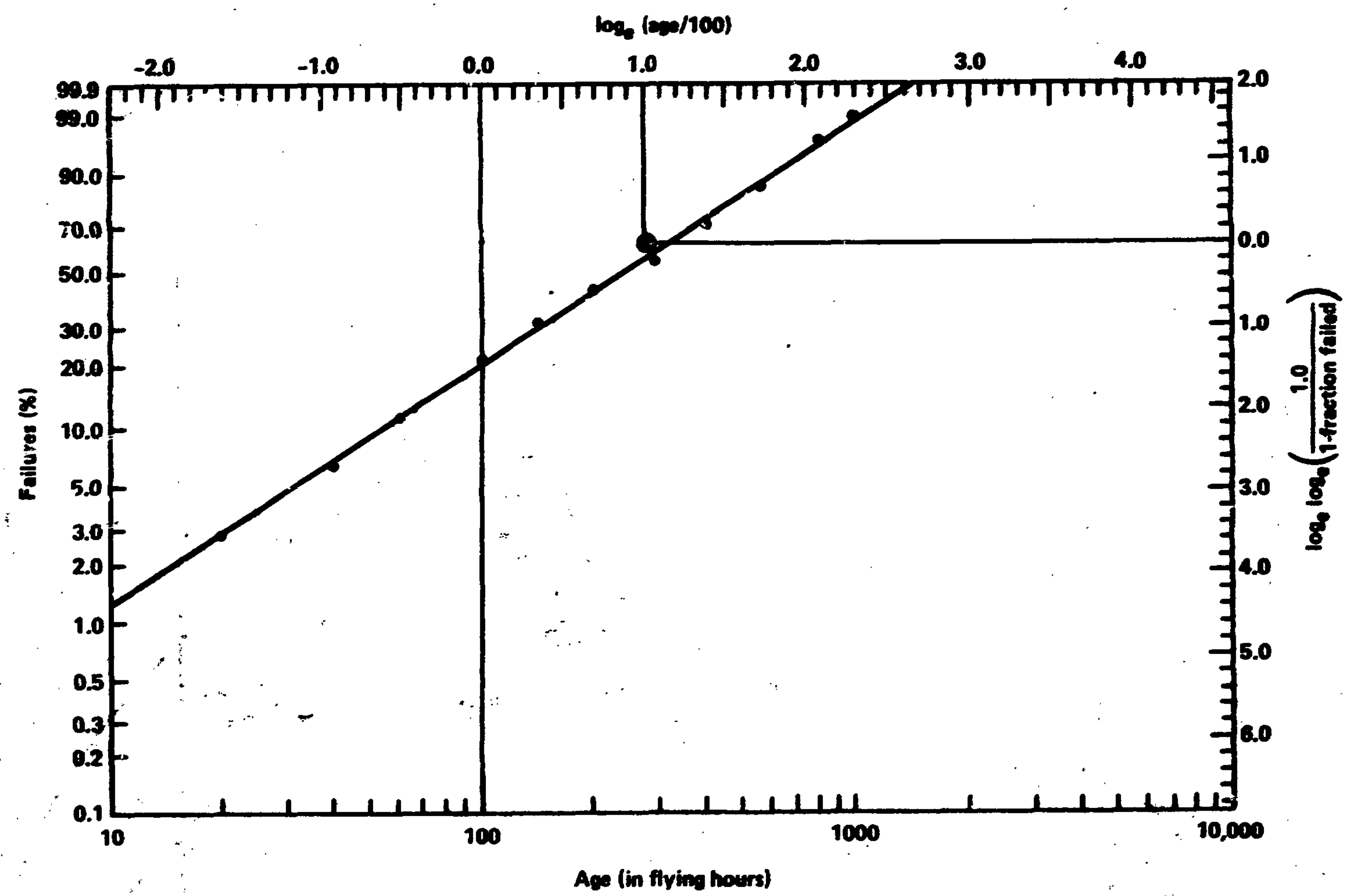


Fig. 4.5.3

Figure 4.5. Weibull Survival Distribution, Density, and Hazard Rate



SOURCE: United States Air Force, *Procedures for Determining Aircraft Engine (Propulsion Unit) Failure Rates, Actuarial Engine Life, and Forecasting Monthly Engine Changes by the Actuarial Method*, Technical Order, TO 00-25-128, October 20, 1969.

Figure 4.6. Weibull Survival Distribution: J47-GE-27 Jet Engines (Log-log graph paper)

4.4 Lognormal Survival Distribution

The lognormal survival distribution appears to be finding increasing favor as a candidate for the description of survival data. It has been applied to the description of crack growth as a function of time in primary aircraft structures [3] and to jet engine compressor bleed control data (cp. Figure 4.8).

The lognormal survival distribution is

$$R(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_t^{\infty} \exp\left(-\frac{1}{2}\left(\frac{\log_e u - \overline{\log_e t}}{\sigma}\right)^2\right)\frac{du}{u} , \qquad (4.20)$$

where $0 < t$, $\overline{\log_e t}$ is the mean of $\log_e t$, and σ is the standard deviation. The corresponding lognormal survival density is

$$\rho(t) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \frac{1}{t} \cdot \exp\left(-\frac{1}{2}\left(\frac{\log_e t - \overline{\log_e t}}{\sigma}\right)^2\right) , \qquad (4.21)$$

and the lognormal hazard rate is

$$\eta(t) = \frac{\exp\left(-\frac{1}{2}\left(\frac{\log_e t - \overline{\log_e t}}{\sigma}\right)^2\right)}{t\int_t^{\infty} \exp\left(-\frac{1}{2}\left(\frac{\log_e u - \overline{\log_e t}}{\sigma}\right)^2\right)\frac{du}{u}} . \qquad (4.22)$$

Graphs of these functions are displayed in Figure 4.7 [illegible] .8 exhibits an application to observations. Notice that [illegible] rate (Figure 4.7.4) increases at first, attains a maximum [illegible] decreases thereafter. While this behavior is not often observed, the example illustrated in Figure 4.8 suggests that the lognormal distribution may be appropriate for some special types of aviation failure phenomena.

$$R(t) = \frac{1}{\sigma\sqrt{2\pi}}\int_t^{\infty} \exp\left(-\frac{1}{2}\left(\frac{\log u - \overline{\log t}}{\sigma}\right)^2\right)\frac{du}{u}$$

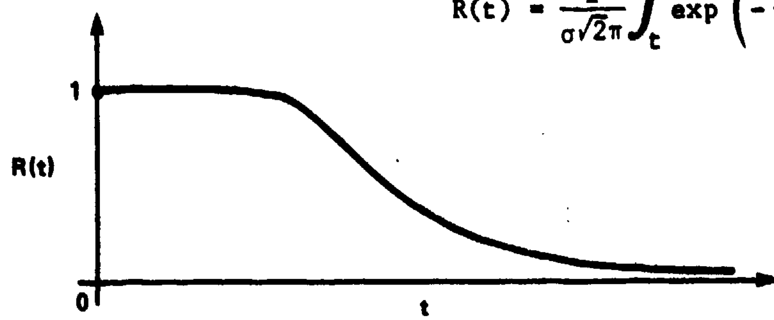


Fig. 4.7.1

$$\rho(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\log t - \overline{\log t}}{\sigma}\right)^2\right)$$

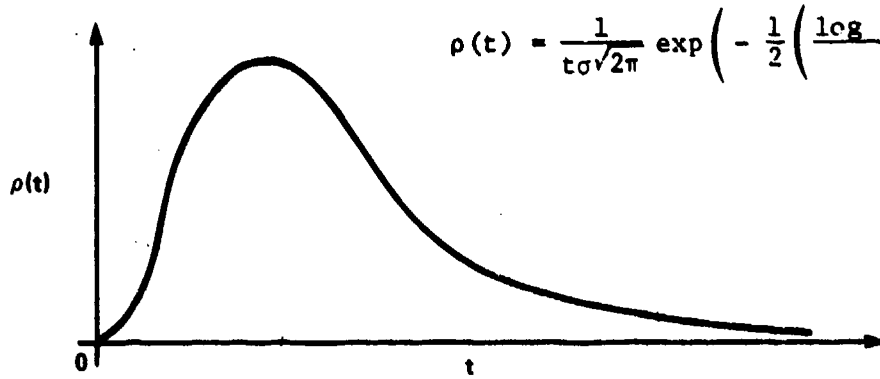


Fig. 4.7.2

$$\eta(t) = \frac{\exp\left(-\frac{1}{2}\left(\frac{\log t - \overline{\log t}}{\sigma}\right)^2\right)}{t\int_t^{\infty} \exp\left(-\frac{1}{2}\left(\frac{\log u - \overline{\log t}}{\sigma}\right)^2\right)\frac{du}{u}}$$

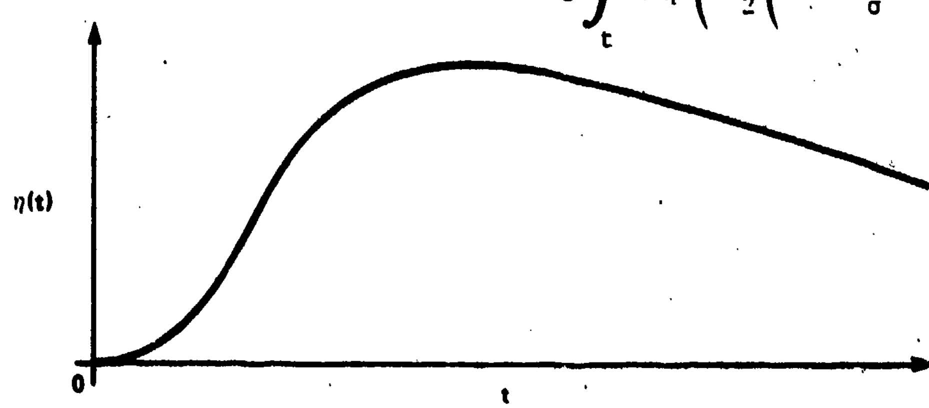
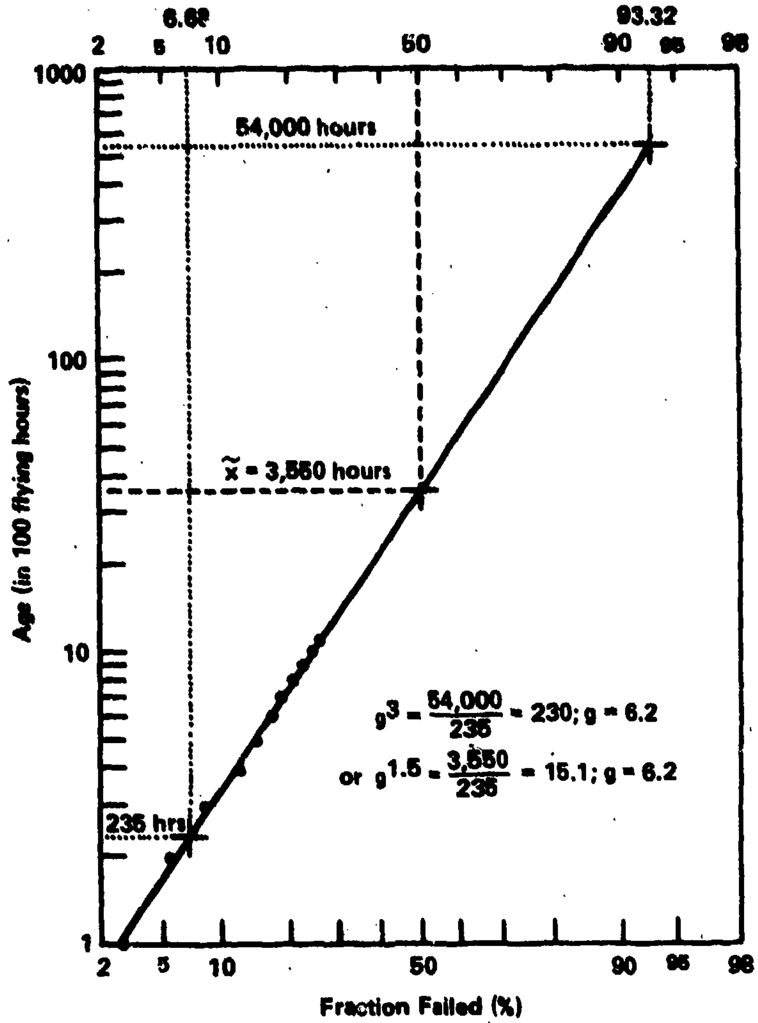


Fig. 4.7.3

Figure 4.7. Lognormal Survival Distribution, Density, and Hazard Rate



NOTE: Based on 49 failures in airline service.

Figure 4.8. Lognormal Survival Distribution: Jet Engine Compressor Bleed Control (Lognormal probability graph paper)

4.5 Gamma Survival Distribution

This distribution generalizes both the lognormal and the exponential distributions. The gamma survival distribution is

$$R(t) = \frac{1}{\Gamma(s)} \int_{\lambda t}^{\infty} e^{-u} u^{s-1} du \quad , \quad s > 0 \quad , \quad \lambda > 0 \tag{4.23}$$

where $t > 0$ and

$$\Gamma(s) = \int_0^{\infty} e^{-u} u^{s-1} du \tag{4.24}$$

is Euler's gamma function. If $s = 1$, then the gamma distribution reduces to the exponential distribution $R(t) = \exp(-\lambda t)$.

If $s = \frac{1}{2}$, then the substitution of variables

$$u = \frac{1}{2} \left(\frac{\log_e v - \mu}{\sigma} \right)^2 \tag{4.25}$$

transforms the gamma distribution Eq. (4.23) into a lognormal distribution relative to a pseudo time variable $\tau = e^{\mu + \sigma\sqrt{2\lambda t}}$. The gamma survival density is

$$\rho(t) = \frac{\lambda^s t^{s-1}}{\Gamma(s)} \exp(-\lambda t) \quad , \tag{4.26}$$

and the gamma hazard rate is given by

$$\eta(t) = \frac{\lambda^s t^{s-1} \exp(-\lambda t)}{\int_{\lambda t}^{\infty} e^{-u} u^{s-1} du} \quad . \tag{4.27}$$

Their graphs and an application are displayed in Figures 4.9 and 4.10.

$$R(t) = \frac{1}{\Gamma(s)} \int_{\lambda t}^{\infty} e^{-u} u^{s-1} du$$

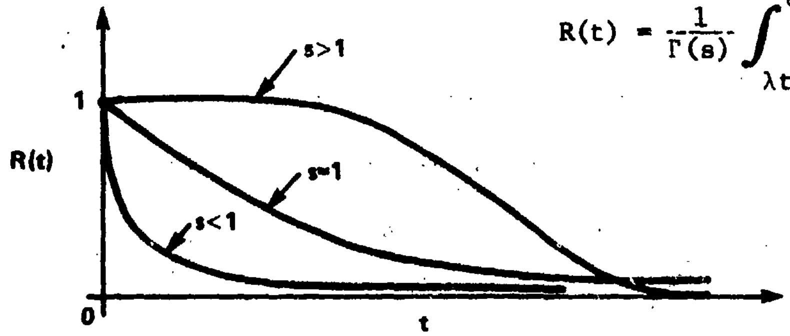


Fig. 4.9.1

$$\rho(t) = \frac{\lambda^s t^{s-1}}{\Gamma(s)} \exp\,(-\lambda t)$$

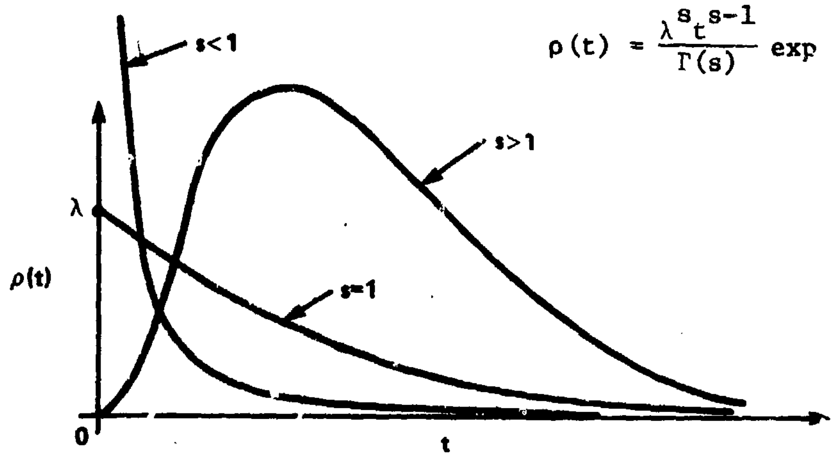


Fig. 4.9.2

$$\eta(t) = \frac{\lambda^s t^{s-1} \exp\,(-\lambda t)}{\int_{\lambda t}^{\infty} e^{-u} u^{s-1}\, du}$$

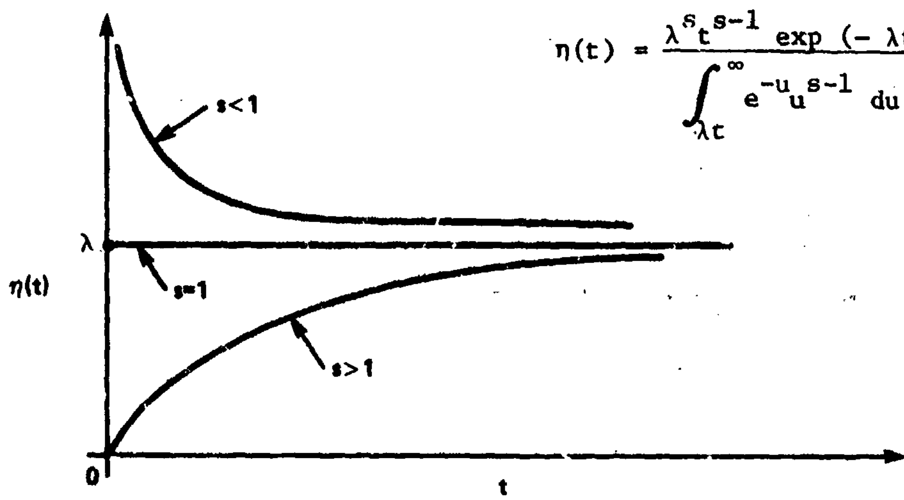
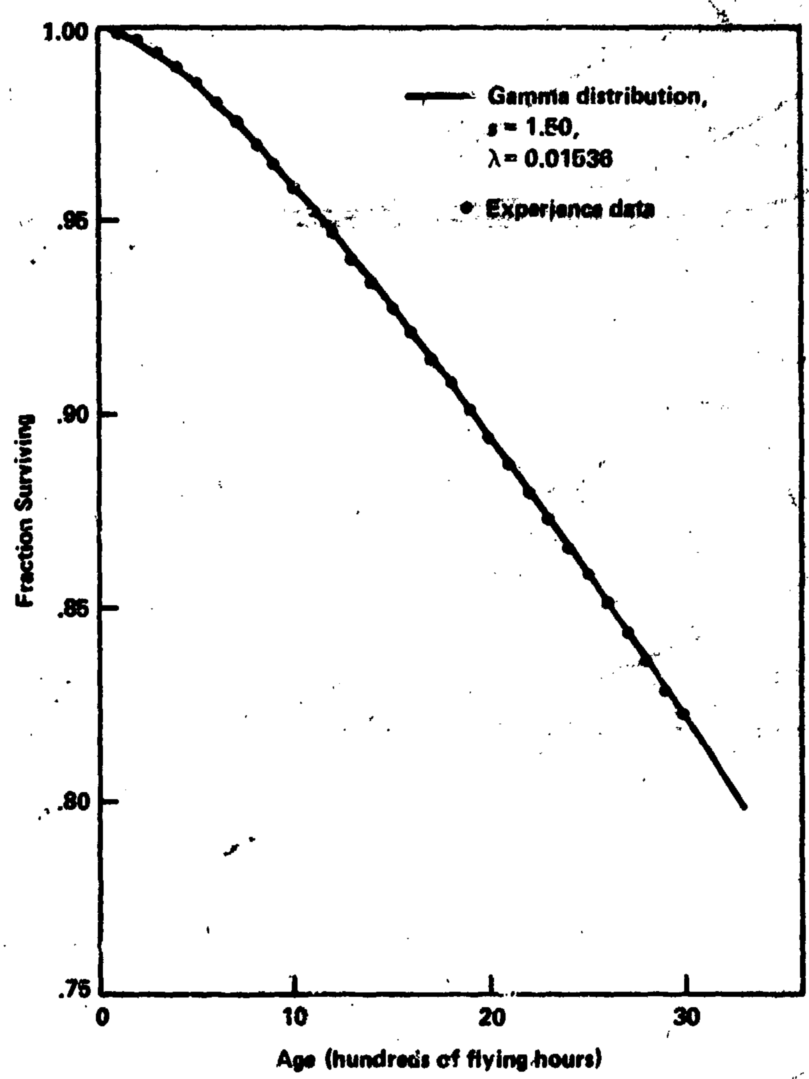


Fig. 4.9.3

Figure 4.9. Gamma Survival Distribution, Density, and Hazard Rate



SOURCE: United States Air Force, *Procedures for Determining Aircraft Engine (Propulsion Unit) Failure Rates, Actuarial Engine Life, and Forecasting Monthly Engine Changes by the Actuarial Method*, Technical Order, TO 00-25-128, October 20, 1969.

Figure 4.10. Gamma Survival Distribution: J69-T-25 Jet Engine

5. SIMPLE AND COMPLEX SYSTEMS

5.1 The statistical study of reliability had its origin in demography, and its terminology reflects this history. The survival distribution, which specifies the probability that an individual belonging to a homogeneous population will survive until time t, yields a hazard function which, as time increases from birth until death, initially decreases from large values during an interval of infant mortality, remains relatively constant for some time, and then, as the wear-out interval of old age is attained, once again increases. Thus the graph of the hazard function is a shallow U-shaped curve, frequently called the "bathtub curve" in reliability literature; cp. Figure 5.1.

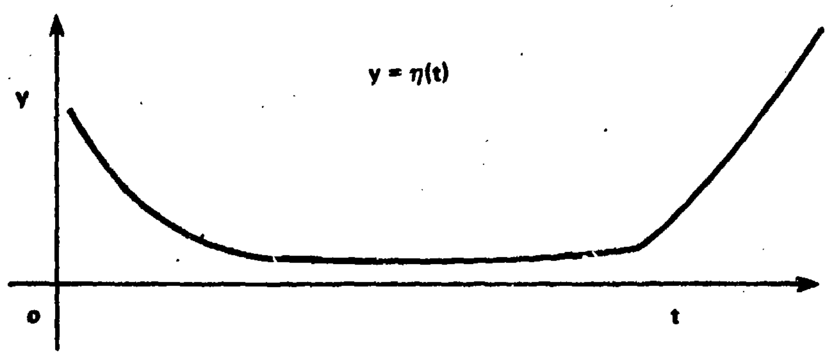


Figure 5.1. "Bathtub" Hazard Function

The reader will have noticed that none of the standard reliability distributions described in Section 4 gives rise to a hazard function whose graph is U-shaped. For instance, the Weibull distribution corresponds to a hazard function of the form (cp. eq (4.19))

$$\eta(t) = \lambda s t^{s-1} \quad , \quad \lambda > 0 \quad , \quad s > 0 \quad , \qquad (5.1)$$

which increases with increasing time if $s > 1$, is constant if $s = 1$, and decreases with increasing time if $s < 1$. The hazard function for the lognormal survival distribution increases to a maximum and then decreases as time increases. Such functions can be used to describe the infant mortality regime of a hazard function, or the wear-out regime, but not both. This remark has an important consequence: if an item displays infant mortality characteristics, that is, if $\eta(t)$ is a decreasing function for small positive t, then $\eta(t)$ can only be represented by one of the standard distributions (such as those described in Section 4) for epochs much earlier than typical wear-out epochs, since the infant mortality data is necessarily acquired first. There can be no solely mathematical method for gaining information about wear-out characteristics from data which includes infant mortalities.

The reason for this state of affairs is plain in human mortality characteristics. Although the statistical properties of infant mortality and wear-out at old age are separately highly regular and susceptible to statistical analysis, their causes and corresponding hazard functions are very different. There is no reason to believe that any one mathematically simple statistical distribution can be related to the underlying physical phenomena which correspond to both extremes. It thus becomes necessary to think of the U-shaped hazard function as the sum of (at least) two independent functions,

$$\eta(t) = \eta_0(t) + \eta_\infty(t) \quad , \tag{5.2}$$

where $\eta_0(t)$ describes the hazard due to infant mortality and $\eta_\infty(t)$ describes that due to wear. One could argue for including a third hazard function to describe hazard at intermediate ages, as is done in demographic analysis, but this will not be necessary for our present purpose.

With η decomposed as in eq (5.2) above, and supposing that both infant mortality and wear-out are present for the items under consideration, we may think of η_0 as a decreasing function which tends to a limit k_0, $0 \le k_0 = \lim_{t\to\infty} \eta_0(t)$. It is clear that without further detailed information about the (physical) characteristics of the item under consideration, analysis of early failure data cannot lead to any conclusions about wear-out <u>if</u> infant mortality persists for a significant period of time.

Reliability theoreticians are consequently constrained to study specific systems for which it is possible, on physical or other grounds, to determine η_0 and η_∞ independently, or to study systems for which either infant mortality or wear-out (or both) are negligible. They are faced with an additional difficulty. Complex systems whose constituents follow various distinct survival distributions, or the same distribution with a variety of parameter values, are not amenable to rigorous analysis. For these reasons the bulk of the theoretical literature concerned with reliability is devoted to <u>simple (one-celled) items</u> for which the hazard function is assumed either non-decreasing (wear-out) or non-increasing (infant mortality)--the constant hazard function, corresponding to the exponential survival distribution, is a special case of both--and to configurations of identical or closely related simple items which possess special symmetries, e.g., series- or parallel-connected simple items. With these constraints it may be possible to derive optimal maintenance policies if the family of policies considered is sufficiently structured. Perhaps the most popular structural policy constraint is maintenance periodicity.

Many simple items exhibit wear. If replicas of such an item are expected to be in service at a future date significantly greater than the lifetime of an individual item and if single items are producable at low cost and in great number, then age exploration, or life testing, will establish the hazard function from observations and thereby identify it as a standard hazard function, amenable to theoretical study, if it happens to be one. If an analytical expression is not known, an approximation can be obtained (e.g., following the prescription given in [9]), or numerical methods can be used to carry out the computations called for by theoretical analyses. In this case there is no problem in principle in applying standard methods of reliability theory.

If, however, the expected operational lifetime prior to obsolescence of the type of item is comparable with the expected lifetime of an individual item of that type, then, unless accelerated testing is possible, there will be no time for age exploration; wear characteristics must be derived from some more basic, usually physical, argument, or hypothesized based on related prior experience, or analyses founded on explicit knowledge of the hazard function must be forgone.

S[illegible] show that there are important categories of items for which the survival distribution is a standard distribution, and the parameter values can be estimated from actuarial analysis. An extensive analysis of survival distributions was reported by D. J. Davis [1]. Among his findings were that the exponential survival distribution was characteristic of such devices as

- commercial aircraft radio tubes,
- Linotype machines
- automated mechanical calculating machines
- ball bearings

All but the last are now obsolete. It has since been reported that most electronic systems and most 'complicated' systems also fall into this category. Aircraft engines, however, usually exhibit some degree of wear-out, i.e., their hazard function ultimately increases with time (cp. Figures 4.4, 4.6, and 4.10, but also Figure 4.2).

Typical studies of preventive maintenance policies for simple systems assume that the actual state of the item is known at all times prior to failure, including the associated survival distribution. The time of failure of the item is the only unknown. Moreover, typical maintenance actions are restricted to replacement of a given item by an identical zero-timed item, thus 'renewing' the system of which the item is a constituent. Generally, the problem treated is determination of the time of replacement (renewal) to minimize cost or meet a numerically expressed safety requirement, or to introduce redundancy (i.e., create a symmetrically interconnected collection of replicas of the simple item to form a simple system) in order to reduce the failure rate.

5.2 When the items which constitute a system are essentially identical and are interconnected in a symmetrical way (e.g., series--or parallel--interconnection) and when the survival distribution corresponding to each individual item is known, then it may be possible to perform a complete mathematical analysis of the reliability of the system. Systems for which one or more of these assumptions are invalid can be called complex systems. This definition differs in an inessential way from that given in Chapter 4, Section 2. The combined vehicle- and earth-based control systems for the Apollo and Viking projects are examples of one-time complex systems for which neither complete age exploration nor accelerated testing to determine survival characteristics was possible. This deficiency was compensated, to some extent, by the extensive use of redundancy. Nevertheless, it is clear that a complete mathematical reliability analysis for such a system is out of the question.

Commercial and military aircraft are examples of complex systems about which much more can be learned through testing, age exploration, and experience because there are, relatively, so many more of them and, ultimately, they are in operation for a long period of time. But for them also a complete mathematical analysis is out of the question because of the large number of diverse items, each with its own survival characteristics, and the complex and irregular interconnections and multiple uses and paths which have been designed into modern aircraft, or are unintentional consequences of the design. Moreover, aircraft are modified as time passes to incorporate new developments in assembly and subsystem design, and maintenance activities quickly ensure that the ages of various subsystems, both major and minor, bear little relationship to the nameplate age of the airframe.

Just as the (classical) properties of a gas cannot, in practice, be derived from knowledge of Newton's equations although the latter suffice in principle for the task, so too the survival characteristics of a complex system could not be obtained in practice even if complete knowledge of the survival characteristics of its constituent parts as well as the details of their interconnection were available. An alternative method is needed, less sensitive to the 'microscopic' structure of the complex system and therefore necessarily of insufficient power to treat all conceivable questions, but powerful enough nevertheless to guide the formulation of maintenance policy. To continue this simile, the relationship of a method for analysis of complex systems to the traditional method for analysis of simple systems can be likened to the relationship between statistical mechanics and newtonian mechanics: detailed knowledge about individual items and their interconnection will, in general, not play an explicit role, but the method will provide the decisive information which is used to formulate answers to the basic maintenance policy questions.

The Reliability-Centered Maintenance Program [6] described in this volume is a general method of designing maintenance policies for complex systems which requires very little explicit 'microscopic' knowledge of survival distributions and interconnections for the tens of thousands of constituents of a commercial aircraft. The next Section is devoted to a mathematical description of the structure of this Program.

6. RELIABILITY-CENTERED MAINTENANCE

6.1 The principal goal of a maintenance system is to ensure the highest practical standard of operating performance of the equipment being maintained. Criteria of operating performance are, however, quite varied, depending simultaneously upon the cost of maintenance and the consequences of failure. For circumstances where the consequences of failure are relatively minor it will generally be sufficient to focus on the reliability of the constituent items of the system, and to learn from experience as well as from testing whether component redesign is necessary and which maintenance policies are cost-effective. As such information accumulates, maintenance policies and system design evolve together to improve operating reliability.

Those systems for which the consequences of failure are serious, such as commercial aircraft, nuclear reactors, and military missile systems, must be considered from a different point of view. In each of these instances, the consequences of certain failures are unacceptable. Critical failures in the sense of Chapter 3.2 belong to this category. It will be convenient to refer to any unacceptable failure as a critical failure. The criteria of unacceptability may be quite complicated in any specific instance, although certain types of failure will normally be clearly unacceptable. For example, a failure in a military missile which destroyed its ability to complete its mission would be unacceptable, as would a failure of a nuclear power reactor situated in a densely populated region which could lead to an explosion.

In situations such as these there is the temptation to avoid failure "at all costs," but, since there are always practical limitations to the resources which can be brought to bear on any single problem, and also because in certain complicated circumstances it is not possible to obtain all of the potentially valuable information which would in principle be necessary to avoid failure, the attempt to avoid critical failures must inherently be a compromise between the imputed cost of the failure and the cost of procedures that would decrease the probability of failure.

For complex systems such as commercial aircraft, it would be prohibitively costly to devote serious and scheduled maintenance to each of its tens of thousands of parts. But of greater importance is the observation that intensive scheduled maintenance (be it "Hard time" or "On Condition;" cp. Chapter 5), regardless of cost, will not necessarily reduce the probability of critical failures. This suggests that the constituent items of a system should be analyzed with regard to the consequences of their failure rather than merely with regard to their reliability. If the consequences of failure are acceptable, then, in the absence of some other reason unrelated to criticality of failure, the maintenance policy designer need not and should not devote resources to scheduled maintenance of the item. The recognition of the importance of the functional role and consequences of failure of an item are basic principles of the Reliability-Centered Maintenance Program; cp. the extensive discussion in Chapter 3. Its main practical consequence in the case of commercial aircraft is that, of the tens of thousands of items which are part of an aircraft, only several hundred participate in critical failures and therefore the latter are the only candidates for scheduled maintenance procedures.

It may turn out that an item participates in critical failures but cannot benefit from scheduled maintenance. There may not be any way to detect reduced resistance to failure. One resolution of this dilemma is to redesign the item to avoid participation in critical failures or so that reduced resistance to failure can be detected by scheduled maintenance operations. The latter solution is an instance of another important principle of the Reliability-Centered Maintenance Program: items which participate in critical failures should be replaced by items which convert critical failures to non-critical failures or to a mode of reduced resistance to failure which can be detected by scheduled maintenance operations. One consequence of this policy is that it may lead to an increase in the number of failures or equipment replacements, thereby increasing maintenance costs; but, by reducing the probability of critical failures, it also reduces the total system operating costs which include the imputed large costs of critical failures. Thus, application of the Reliability-Centered Maintenance Program simultaneously

- Reduces the probability of critical failure;
- Reduces maintenance costs by reducing the number of items considered for scheduled maintenance;
- Increases maintenance costs by replacement of items whose reduced resistance to failure is unobservable by items whose reduced resistance is detectable by scheduled maintenance, or whose failure is non-critical.

The remainder of this Section provides a mathematical formulation of the preceding ideas. There are three main mathematical aspects. The first corresponds to the partition of the system into sets of items that are functionally related by means of the consequences of their failure (cp. Chapter 7). The second is the formal expression of the costs of maintenance and consequences of failure in common terms of direct and imputed costs. This maintenance/failure cost function is really the main object of study. The principal purpose of the maintenance policy designer is to minimize the maintenance/failure cost function. The third mathematical aspect models the iterative procedure used in the Reliability-Centered Maintenance Program to minimize the total cost function. The Decision Diagram approach of Chapter 8 is the main component of this part of the Program.

6.2 Every complex system is composed of many individual parts or items. These constituents are not necessarily in one-to-one correspondence with functions performed by the system. Most physically distinct parts perform no function at all in isolation; some may be cannily designed to participate in the performance of several distinct functions (as an airliner seat cushion is also a flotation device). Thus it is impossible to identify parts with functions or roles, and it may not even be possible to obtain complete agreement upon what constitutes the set of elementary, or irreducible, items of a complex system. We will assume that some choice has been made. The volume titled Reliability-Centered Maintenance [7] provides a detailed description of one procedure that can be followed to make this selection for commercial aircraft.

Let $\underline{\underline{S}}$ denote the set of items of some complex system and let s denote a typical item belonging to $\underline{\underline{S}}$. Items of a given type may occur more than once in the system; each occurrence is represented by a distinct element of $\underline{\underline{S}}$. We may think of the items which constitute the system as represented by points, and of $\underline{\underline{S}}$ as the set of those points; this interpretation is used in Figure 6.1

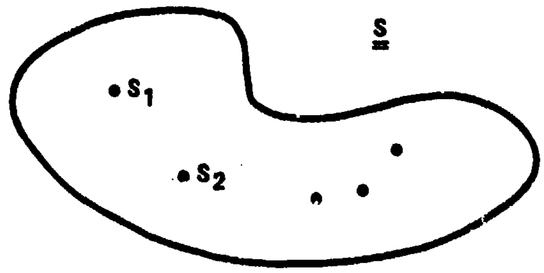


Figure 6.1. Set of Items of a Complex System

To each $s \in \underline{\underline{S}}$ there corresponds an associated survival distribution $t \mapsto R_s(t)$, where we suppose that some satisfactory definition of failure for s has been selected. The reader should recall the extensive discussion of this difficult problem in Chapter 3. With an appropriate definition of failure for the system $\underline{\underline{S}}$ itself, let $R_{\underline{\underline{S}}}(t)$ denote the survival distribution for $\underline{\underline{S}}$. If $R_{\underline{\underline{S}}}(t)$ could be readily expressed in terms of the $R_s(t)$, $s \in \underline{\underline{S}}$, then the problem of maintenance policy design would be reduced to the establishment of a maintenance procedure for each $s \in \underline{\underline{S}}$ which ensures that $R_{\underline{\underline{S}}}(t) > k$ (where k is a given minimal acceptable system reliability) and, subject to that constraint, costs least to implement. In other words, programs developed using the techniques of reliability-centered maintenance tend towards minimizing all costs that are a function of scheduled maintenance.

But $R_{\underline{S}}(t)$ cannot be explicitly expressed in terms of the $R_s(t)$ for complex systems consisting of numerous parts. The set $\{R_s(t) : s \in \underline{S}\}$ of survival distributions does not contain all the information necessary for the analytical solution of the problem because the components s of the system are in general interconnected and, therefore, at least some of the survival distributions $R_s(t)$ are not independent. Suppose, for the moment, that the probability of survival of each $s \in \underline{S}$ were independent of the probability of survival of the remaining items. Then

$$R_{\underline{S}}(t) = \prod_{s \in \underline{S}} R_s(t) \ , \qquad (6.1)$$

and this relationship would enable one directly to reduce all questions about system survival to questions about the survival characteristics of the elementary items, ignoring their interconnection. Since the $R_s(t)$ are, in general, dependent, we have the choice of studying the interconnection of the items or avoiding consideration of elementary items altogether. The first alternative is typical of the standard methods reported in the literature. The second alternative has received much less attention (a recent analysis which adopts this viewpoint is reported in [10]); it lies at the foundation of the Reliability-Centered Maintenance approach.

We need some terminology. If $\underline{S}$ is any set, then a <u>partition</u> of $\underline{S}$ is a collection of subsets Λ such that

$$\bigcup_{\lambda \in \Lambda} \lambda = \underline{S} \qquad (6.2.1)$$

and, if $\lambda \in \Lambda$, $\lambda' \in \Lambda$, then

$$\lambda \neq \lambda' \quad \text{implies} \quad \lambda \cap \lambda' = \emptyset \ ; \qquad (6.2.2)$$

eq. (6.2.1) asserts that the subsets λ exhaust $\underline{S}$, and eq. (6.2.2) states that no two of the subsets overlap. This situation is represented in Figure 6.2. A partition M of $\underline{S}$ is said to be a <u>refinement</u> of the partition Λ if each $\mu \in M$ is contained in some $\lambda \in \Lambda$. Thus, the refinement M further partitions the subsets λ. A refinement of the partition Λ exhibited in Figure 6.2 is designated by the dotted lines in Figure 6.3.

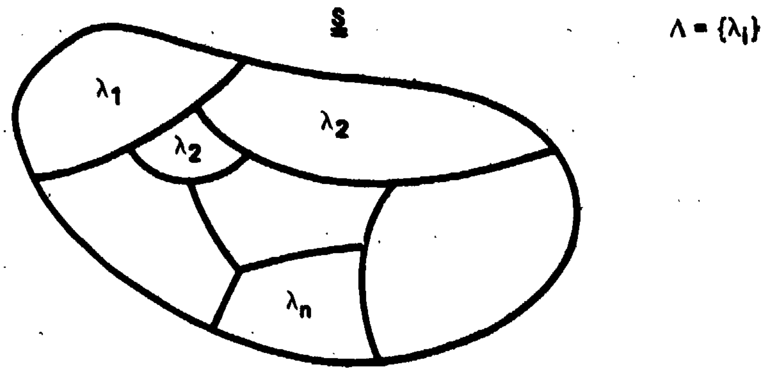


Figure 6.2. A Partition Λ of S

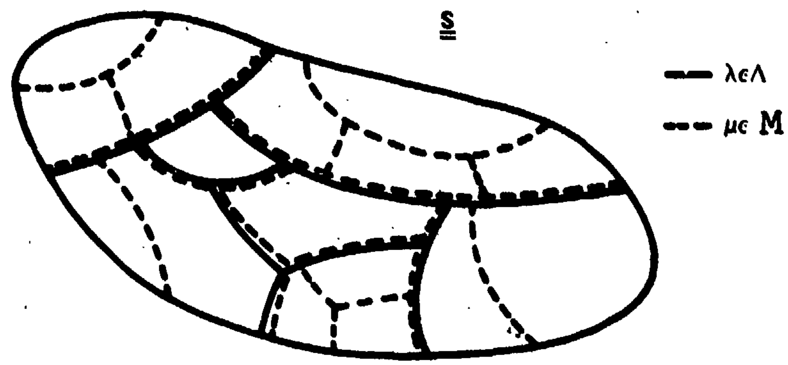


Figure 6.3. Refinement M of Partition Λ

The collection of subsets of the form $\{s\}$, where s runs through all the elements of S, is the finest partition of all; it is a refinement of every partition of S.

Now suppose that $\underline{\underline{S}}$ is the set of elementary items of a complex system and that Λ is a partition of $\underline{\underline{S}}$. Just as the survival distribution $R_{\underline{\underline{S}}}(t)$ is associated with the system itself, so too can a survival distribution $R_\lambda(t)$ be associated with each set λ belonging to the partition. Each λ is a collection of items, but $R_\lambda(t)$ will not in general be the product of the survival distributions of the constituent items $s \in \lambda$ because of their interconnections. Nevertheless there may be some partitions Λ for which the survival distributions assume a particularly convenient analytical form or, even if they cannot be explicitly identified, have particularly convenient properties.

One purpose of the decomposition and partition procedures discussed in Chapters 2 and 7 is to define a convenient partition of the set of parts of an airliner. The method described is applicable, in principle, to any complex system. Various parts are amalgamated by their interconnections and functional interdependence into components, subassemblies, assemblies, and subsystems. Each of these is a natural candidate for an element in a partition of $\underline{\underline{S}}$. If, for example, a partition Λ contains some subsystem λ, then the subassemblies which constitute λ, together with the other elements of Λ, define a refinement of Λ.

Let us suppose that Λ is a partition of $\underline{\underline{S}}$ such that the survival functions $R_\lambda(t)$ and $R_{\lambda'}(t)$ of any pair of distinct elements $\lambda \neq \lambda'$, $\lambda, \lambda' \in \Lambda$, are independent. Then

$$R_{\underline{\underline{S}}}(t) = \prod_{\lambda \in \Lambda} R_\lambda(t) \quad . \tag{6.3}$$

A partition enjoying this property always exists, because the <u>coarsest</u> partition, which consists of the single set $\underline{\underline{S}}$ itself, has this property. In principle, most is known about the survival characteristics of the elementary parts from which $\underline{\underline{S}}$ is ultimately constructed, and progressively less is known about increasingly complex amalgamations of the elementary parts. Therefore we seek that compromise partition whose constituent subsets are as simple as possible, i.e., as close to the elementary parts as possible, while still retaining the property that the survival distributions of the elements of the partition are independent, so that eq. (6.3) remains valid. That is, among all partitions of $\mathcal{S}$ for which eq. (6.3) holds, we seek a partition such that if M is any refinement of Λ, then eq. (6.3) does not hold for M. A partition Λ which has this property will be said to be maximally independent. It is clear that maximally independent partitions exist but are not necessarily unique; that is, there may be more than one way to select a maximally independent partition.

It is intuitively clear that a complex system such as an airliner can be partitioned into independent (or at least very nearly independent) subsystems according to this prescription. For instance, apart from a common interdependence on the power-plant as an energy source, the subsystem consisting of the collection of passenger reading lights is independent of the cabin pressurization subsystem, the landing gear assembly is independent of the flight control surfaces subsystem, and so forth.

Hereafter we will assume that some maximally independent partition Λ has been selected. The next task is to associate a cost function with this partition.

6.3 Let $C_\lambda(t)$ denote the sum of the expected cost of maintenance and imputed cost of failure of the partition element $\lambda \in \Lambda$ as a function of time t. $C_\lambda(t)$ includes the cost of Hard Time replacements, of On Condition inspections and replacements, of warehousing and distribution of replacement items, and all other costs attributable to the maintenance function. It also includes the imputed cost of failure of λ. For some partition elements the cost of failure is negligibly greater than the cost of renewal of λ. For instance, failure of the in-flight motion picture system, a fairly frequent occurrence on some airlines, is at worst an irritation which may influence passenger preference in a minor way and thereby affect future passenger load factors to some slight degree. Failure of other, safety-related, partition elements can entail costs far greater than the cost of renewal of λ. If failure aborts the mission or, in the case of airliners, causes loss of life and/or loss of the entire system, then the imputed costs of the failure constitute the principal driving force behind the design of the maintenance policy. $C_\lambda(t)$ includes such costs.

Certain costs are not included in $C_\lambda(t)$ in what follows, although they might find their place in a more comprehensive treatment of our subject. In the case of commercial airliners, revenue-producing costs, including advertising and non-maintenance personnel expenses, are excluded from $C_\lambda(t)$.

Without loss of generality we may suppose that $C_\lambda(t)$ is the sum of an absolutely continuous function (which represents, in part, the imputed cost of failure) and a discrete part (which includes the costs of periodic maintenance and renewal); recall the definitions given in Section 2. It follows that the cost function $C_\lambda(t)$ possesses a corresponding cost density (generalized) function $c_\lambda(t)$ (recall eq. (2.30) and the related discussion of the density associated with a discrete distribution). Then

$$C_\lambda(t) = \int_0^t c_\lambda(t)dt \quad . \tag{6.4}$$

If λ never fails, then $C_\lambda(t)$ essentially reduces to the cost of Hard Time and On Condition maintenance through time t, and can be approximated by a step function. If the probability of failure is not zero, then it will be more useful to express the maintenance/failure cost $C_\lambda(t)$ in terms of the failure distribution $F_\lambda(t)$. If λ is maintained in accordance with a Hard Time policy without inspection, then the associated renewal cost density will be proportional to the number of items which survive until the replacement time. If that time is t_k, then the corresponding cost density is proportional to $\delta(t-t_k)\,R(t)$, where $\delta(t-t_k)$ is the Dirac delta function (cp. Eq. (2.27)). If λ is maintained in accordance with an On Condition policy, then costs will be incurred at every inspection. If inspection times are $t_1, t_2, \ldots, t_i, \ldots$, then the cost density will be of the form

$$\sum_i c^m_{\lambda,i}(t)\ \delta(t-t_i)\,R_\lambda(t) \quad ; \tag{6.5}$$

here $c^m_{\lambda,i}(t)$ denotes the cost of maintenance of λ at time t_i. Finally, if λ is maintained in accordance with a Condition Monitoring process, or if λ actually fails, then the corresponding cost density will be proportional to the failure density $\rho_\lambda(t) = -\frac{dF_\lambda}{dt}$. It follows that the general cost density has the form

$$c_\lambda(t) = c^f_\lambda(t)\ \rho_\lambda(t) + \sum_i c^m_{\lambda,i}(t)\ \delta(t-t_i)\ R_\lambda(t) \quad . \tag{6.6}$$

The imputed costs of failure are represented by $c^f_\lambda(t)$. Eq. (6.6) shows that failure cost density is proportional to the failure density and that other maintenance costs are proportional to the survival distribution. In order to make these expressions comparable, we will express the survival distribution in terms of the hazard rate and the failure density. From eq. (3.14) we have $R_\lambda(t) = \rho_\lambda(t)/\eta_\lambda(t)$ so

$$\begin{aligned} c_\lambda(t) &= \left\{c^f_\lambda(t) + \sum_i \frac{c^m_{\lambda,i}(t)\ \delta(t-t_k)}{\eta_\lambda(t)}\right\}\rho_\lambda(t) \\ &=_{\text{def}}\ \gamma_\lambda(t)\ \rho_\lambda(t) \end{aligned} \tag{6.7}$$

where we have written

$$\gamma_\lambda(t) = c^f_\lambda(t) + \sum_i \frac{c^m_{\lambda,i}(t)\ \delta(t-t_k)}{\eta_\lambda(t)}$$

If the frequency of inspections is large compared with the frequency of failures, then the delta functions can be approximated by linear interpolation. This amounts to the assumption that inspection costs are expended uniformly with time rather than at a discrete set of times.

The functions $c^f_\lambda(t)$ and $c^m_{\lambda,i}(t)$ are <u>costs</u>, hence positive functions. This fact will be used in what follows.

The total maintenance/failure cost of the system $\underset{=}{S}$ as a function of time will be

$$C(t) = \sum_{\lambda\in\Lambda} C_\lambda(t) = \sum_{\lambda\in\Lambda} \int_0^t \gamma_\lambda(t)\,dF_\lambda(t) \quad , \tag{6.8}$$

where we have written $dF_\lambda(t)$ in place of $\rho_\lambda(t)dt$. As we have already remarked above, the main objective of the Reliability-Centered Maintenance Program is to minimize the value of $C(t)$ for each time t, given the history of the system for times $t' < t$.

6.4 The problem of minimizing $C(t)$ is still too complicated to admit a mathematical solution even if all the quantities involved were precisely known. Nevertheless, a simple observation provides the key for implementation of a systematic iterative procedure which acts to reduce $C(t)$ if the latter is not already a local minimum.

Since Λ is a maximally independent partition, it will not be possible to reduce $C(t)$ by passage to a refinement M for which the survival distributions $R_\mu(t)$, $\mu \in M$, are independent. This means that $C(t)$ need not be the globally minimal maintenance/failure cost for the system even though it may be minimal for the collection of all maximally independent partitions. Furthermore, the local minimum (subject to the constraint of maximal independence of the partition) depends on the choice of partition. There is no guarantee that minimization of $C(t)$ for the given maximally independent partition Λ is the same as minimization of $C(t)$ over the class of all maximally independent partitions of the system. Thus we must again conclude that the minimum which will be attained by the procedure about to be described, hence also the minimum attained by the Reliability-Centered Maintenance Program, is not necessarily global. Nevertheless, experience suggests that the minimum achieved in the application of the Program to commercial airline operations may be close to the global minimum and, in any event, partial minimization of $C(t)$ by application of the policies introduced below leads to significant reductions in the value of $C(t)$ in practical situations.

Returning now to eq. (6.8), observe that $\gamma_\lambda(t) \geq 0$. Then $C(t)$ is a finite sum (over the elements of the partition Λ) of integrals

$$\int_0^t \gamma_\lambda(\tau)\,\rho_\lambda(\tau)d\tau$$

whose integrands are products of non-negative functions. Consequently, the total cost $C(t)$ through time t will be reduced if one or more of the following three possibilities occurs:

I. For some $\lambda \epsilon \Lambda$ there is a maintenance policy which replaces the failure cost density $c_\lambda^f(t)$ by a failure cost density $c_\lambda^f(t)^*$ such that

$$c_\lambda^f(t)^* \leq c_\lambda^f(t) \quad \text{for all } t \text{ and}$$

$$c_\lambda^f(t)^* < c_\lambda^f(t) \quad \text{for } t \text{ in some open interval.}$$

II. For some $\lambda \epsilon \Lambda$ there is a maintenance policy which replaces the maintenance cost function $c_{\lambda,i}^m(t)$ by a maintenance cost function $c_{\lambda,i}^m(t)^*$ such that

$$c_{\lambda,i}^m(t)^* \leq c_{\lambda,i}^m(t) \quad \text{for all } t \text{ and}$$

$$c_{\lambda,i}^m(t)^* < c_{\lambda,i}^m(t) \quad \text{for } t \text{ in some open interval.}$$

III. For some $\lambda \epsilon \Lambda$ there is a maintenance policy which replaces the product $\gamma_\lambda(t)\ \rho_\lambda(t)$ by a product $\gamma_\lambda^*(t)\ \rho_\lambda^*(t)$ such that

$$\gamma_\lambda^*(t)\ \rho_\lambda^*(t) \leq \gamma_\lambda(t)\ \rho_\lambda(t) \quad \text{for all } t \text{ and}$$

$$\gamma_\lambda^*(t)\ \rho_\lambda^*(t) < \gamma_\lambda(t)\ \rho_\lambda(t) \quad \text{for } t \text{ in some open interval,}$$

and neither I nor II is applicable.

Maintenance policies of Type I occur when an item is redesigned to incorporate redundancy or other fail-safe design methods which act to reduce the cost of failure of the initial item without necessarily affecting its probability of failure. This type of policy change tends to apply to modifications of equipment design rather than to modifications of operational maintenance procedures.

Type II policies are indifferent to survival distributions and therefore are really independent of the properties of the equipment being maintained. They are principally managerial or organizational policies concerned with matters such as scheduling of periodic maintenance tasks, location of depots, provision of replacement parts in adequate number to reduce downtime revenue loss while avoiding costs associated with excessive replacement parts stock, and so forth. Optimal Type II policies can be difficult to identify and implement, but their nature and importance have always been understood by managers and cost accountants. Nevertheless, the large costs of critical failures cannot, in typical situations, be counterbalanced by efficiencies from Type II decisions, that is, without modification of the survival distribution or the cost of failure.

The most significant opportunities for the introduction of maintenance policies which reduce $C(t)$ are of Type III, which can be further categorized into three subtypes. Using the notations and constraints given in III, they can be expressed as follows:

IIIA. $\gamma_\lambda^*(t) \geq \gamma_\lambda(t)$ and $\rho_\lambda^*(t) \geq \rho_\lambda(t)$ for all t;

IIIB. $\gamma_\lambda^*(t) \leq \gamma_\lambda(t)$ and $\rho_\lambda^*(t) \geq \rho_\lambda(t)$ for all t;

IIIC. Neither of the above.

For either of the first two conditions there will be some open interval on which strict inequality obtains because of the condition

$$\gamma_\lambda^*(t)\ \rho_\lambda^*(t) < \gamma_\lambda(t)\ \rho_\lambda(t)$$

in III. It is possible that there will be some intervals where $\rho_\lambda^*(t) < \rho_\lambda(t)$ and others where $\rho_\lambda^*(t) > \rho_\lambda(t)$ compatible with III; these cases are subsumed under IIIC.

In circumstances where IIIA is applicable the reduction in the probability of failure density may result in an increase in maintenance costs. Nevertheless, if a failure of the item in question is critical with a corresponding large cost of failure density $c_\lambda^f(t)$, then the product $\gamma_\lambda(t)\ \rho_\lambda(t)$ will generally be reduced, often by a substantial amount. Maintenance policies of this type correspond to situations where a judicious additional investment in an appropriate maintenance action results in a significant decrease in the failure density for items which are associated with large failure costs. Essential for the effective introduction of Type IIIA maintenance policies is an evaluation of failure modes and the consequences of failure. Based upon such information, maintenance policies of Type IIIA can act to reduce $C(t)$ by introducing a redefinition of an unsatisfactory condition (cp. the discussion in Chapter 3)--that is, a failure--in order to convert functional failures (especially critical failures) into non-functional failures. This conversion will normally be accomplished by introducing instrumentation or various inspection and monitoring activities, each of which adds to maintenance cost, but the increase in maintenance cost is offset by the reduction in the expected cost of failure.

Policies of Type IIIB are particularly effective when applied to non-significant items (cp. the discussions of significant items and Condition Monitoring maintenance in Chapter 8). They decrease $\gamma_\lambda(t)$ while possibly increasing the failure density $\rho_\lambda(t)$ in a manner which decreases the product of these two functions. If the failures of an item are not significant, then there generally is no compelling reason to implement either a Hard Time or an On Condition maintenance policy. By placing such items in the Condition Monitoring category, Type IIIB cost reductions can be obtained. In effect, this means that the failure cost density $c^f_\lambda(t)$ reduces to the cost of replacing the failed item. If this is less than the cost of maintenance over the lifetime of the item, then the cost density product is reduced by implementing this policy. For example, a maintenance policy which periodically dismantled and renewed seat recliners would be relatively costly compared with the imputed cost of a recliner failure. Consequently, although the failure density might be increased thereby, a revised policy which merely monitored the condition of the recliners by establishing a mechanism to report users' complaints would certainly reduce $C_\lambda(t)$ and thus $C(t)$ itself. It is of particular importance to seek those elements of the partition for which scheduled maintenance policies result in greater values of $C_\lambda(t)$ than would Condition Monitoring, (i.e., surveillance) policies either because maintenance processes do not reduce the failure density (e.g., if the associated hazard rate is non-increasing) or because the cost of reduction by maintenance is greater than the imputed added cost of failure through lack of scheduled maintenance. The Decision Diagram technique of the Reliability-Centered Maintenance

Program provides an explicit means for identification of partition elements to which Type III B policies can be applied.

It may happen that application of policies I - III decreases C(t) but that the new cost function is not minimal. Less expensive conversions of functional to non-functional failures, longer inspection intervals and Hard Time renewal intervals may be recognized as beneficial at some subsequent time. New information may become available as a result of experience or testing or theoretical advances. Equipment will generally evolve, and constituent items will be replaced by others with different (but not always more favorable) reliability characteristics. Each of these occurrences may provide a cost-effective reason to apply the policies I - III again, thereby bringing the maintenance/failure cost function closer to a local minimum. The history of the iterated application of maintenance policies of Types I to III will typically, when conceived as one grand maintenance policy, be of Type IIIC: neither the cost densities nor the failure densities exhibit monotone decreasing behavior as time increases, but the policy nevertheless achieves an overall cost reduction at each stage of the iteration.

A simple geometric interpretation of this procedure can be readily visualized. [illegible]der the maintenance/failure cost function C(t) as a function of t[illegible]rious parameters which determine a maintenance policy. These would include Hard Time replacement intervals, the reliability distributions of the parts, and so forth. As a function of these variables and for each time t the maintenance/failure cost function determines a hypersurface in a multidimensional euclidean space. This surface has the property that the total cost function is positive for each time t. The maintenance policy designer seeks a curve on this surface which depends on t such that for each fixed value of t, the curve passes through the minimum point on the hypersurface corresponding to that time. In more picturesque language, the desired maintenance policy is represented by a curve which passes through the lowest points of the deepest valley of the cost hypersurface. The policies I - III are valley-seeking; with each application, they direct the curve further downward into a valley.

Although there is no assurance that the valley into which they direct the policy is the lowest of all, the Reliability-Centered Maintenance Program does ensure that the maintenance policy selected gravitates ever closer to the local valley floor.

7. INFORMATION AND MAINTENANCE PROGRAMS

7.1 Critical failures of large-scale complex systems are generally extremely costly; consequently, a maintenance policy which attempts to minimize total costs must also attempt to minimize the number of critical failures. Thus, an effective maintenance program will of necessity be reliability-centered. The more effective the program is, the fewer critical failures will occur, and correspondingly less information about operational failures will be available to the maintenance policy designer. It is in this sense that the objective of the maintenance policy designer can be thought of as an attempt to minimize information, and that the most successful policy yields no information whatsoever about critical failures because it precludes their occurrence. That the optimal policy must be designed in the absence of critical failure information, utilizing only the results of component tests and prior experience with related but different complex systems, is an apparently paradoxical situation. Moreover, the applicability of statistical theories of reliability to the very small populations of large-scale complex systems typically encountered in practice is questionable and calls for some discussion. Each of these distinct viewpoints leads to the conclusion that maintenance policy design is necessarily conducted with extremely limited information of dubious reproducibility, and we must consider why it is nevertheless possible, and how it can be done. The following two subsections take up these questions in turn.

7.2 Recall the geometric interpretation of the Reliability-Centered Maintenance Program given at the end of Section 6. For each fixed time t the maintenance/failure cost function can be considered as a function of the various parameters whose selection specifies a maintenance policy. This function defines a hypersurface in some multi-dimensional euclidean space. Since costs are necessarily non-negative, the cost function will attain its minimum value at some point(s) of the surface; we may say that such a point is the lowest point of a valley on the surface. The

Reliability-Cent   Maintenance Program is designed to seek the lowest point in some valley on the surface, for each time t.

Denote the sur.ace associated with time t by S_t. If the variation of t is identified with a variable point on a line, then the individual surfaces S_t can be stacked one next to other to form a set

$$S = \{S_t : 0 \leq t < \infty\} \quad ; \tag{7.1}$$

S need not be a smooth surface itself because discontinuous modifications of equipment may introduce discontinuities in S as t increases. For the sake of discussion, let us assume that S itself is a surface (of dimension 1 greater than the dimension of each S_t). The optimal maintenance policy at time t is one which corresponds to a local minimum, i.e., a lowest valley point, on S_t. Combining these as t varies, one obtains a lowest valley point on S_t for each t. These points need not trace out a curve on S because changes of maintenance policy can correspond to a "jump" from the lowest point in one valley on S_t to the lowest point in some other valley on S_t. Nevertheless, it is impossible to implement more than a finite number of policy changes in a finite time interval, so that an optimal Reliability-Centered Maintenance Program corresponds to a finite number of curves lying on S, each of the form $t \to f(t)$, with $f(t)$ a point in S_t which is the lowest point in some valley on S_t. Thus, as t increases, the point $f(t)$ which corresponds to a solution of the maintenance problem traces out a curve which runs along the floor of a valley in S possibly jumping, from time to time, from one valley to another.

The mathematical problem which corresponds to this description consists of locating the minima of S_t as t varies. If the equation which defines S is known, then this problem can in principle be solved by applying the methods of advanced calculus. In practice, were the defining equation known, the number of independent variables entering into it would be so great as to preclude an explicit analytical solution of the problem. In any event, for reasons already cited and discussed in detail throughout [6], the defining equation can not be known because the available information is insufficient.

The defining equation of S contains all possible relevant information about the consequences of all conceivable maintenance policies. This is surely much more information than is actually needed either to specify a locally minimal (valley floor) curve or even to locate one. Indeed, if p denotes a point of S and if any downward direction at p is known, i.e., a direction for which the directional derivative at p is negative, then a small displacement from p along the surface in that downward direction leads to a nearby point, say q, for which the maintenance/failure cost is strictly less than the maintenance/failure cost corresponding to p. Observe that this procedure merely requires information about the cost benefits of policies which differ little from the policy corresponding to p: we may say that this procedure only requires information about policies in a small neighborhood of the policy p. Such information is the most likely to be available, or estimable, in practice. Moreover, this procedure does not even require full information about all policies in a small neighborhood of p; it suffices to know one direction which leads to cost reduction. In this sense, we may construe the Reliability-Centered Maintenance Program as a well-defined procedure for identifying directions on S which tend downward, i.e., reduce maintenance/failure cost.

The rapidity with which the floor of the valley is reached by this process depends on the size of the step taken in the downward direction. If the step size is smaller than necessary, it will take more steps to reach the valley floor, so that greater than necessary maintenance/failure costs will be borne: unnecessary maintenance activities will have been supported, avoidable failures will have been experienced. If the step size is too large, then the maintenance policy may leap from one valley wall to another, unable to detect the floor; and producing an oscillating policy which can, in unfavorable circumstances, produce successively greater maintenance/failure costs and ultimately oscillate among local maxima. The choice of step size is critical, as has been implicitly recognized in the conservative federal guidelines concerning extension of

Hard Time replacement intervals in commercial airline maintenance policies. It is clearly preferable to select a step size smaller than optimal instead of one larger than optimal because the consequences of the former vary continuously with step size whereas small changes in the latter can produce large and unanticipated cost increases. It must also be recognized that the size of the optimal step depends on its location on the surface S, or, to put it more picturesquely, it depends on how "wrinkled" the surface is in the neighborhood of the point from which the step is taken. If the surface slopes gradually and gently downward toward the valley floor, then a larger step will be admissible than is the case when the step-off point lies at the top of a steep cliff overlooking the valley. Determination of the optimal step size is a more difficult problem than is determination of a direction in which the step should be taken because the former implicitly requires some estimate of the <u>magnitude</u> of the directional derivative whereas the latter merely utilizes the <u>sign</u> of that derivative. Suppose that there is reason to believe that the absolute value of the directional derivative is bounded by a known constant on the entire surface S. This information enables one to establish a maximum step size such that maintenance/failure cost increases as the result of over-stepping are held below some prearranged value. Hypotheses about the maximum absolute value of directional derivatives can be based upon prior experience; relative to a maximum step size determined this way, the assertion of some reliability engineers that "there are no cliffs" in hazard functions and other reliability measures is given a precise mathematical interpretation.

In summary, although the maintenance policy designer has little information at his disposal regarding the precise nature of the maintenance/failure cost surface, creation of an iterative minimum-seeking policy only requires enough information to identify downward-tending directions in the neighborhood of an existing policy, and to establish an upper bound for step size in order to avoid overstepping.

<u>7.3</u> It is generally impossible to adequately test most large-scale complex systems because so few replicas are built and the time needed to test one system at the desired confidence level often approximates the

expected lifetime of the system type prior to obsolescence. Simple systems are also subject to the latter problem if high reliability is demanded and technology is rapidly varying. For instance, MIL-STD-690A, "Life Test Sampling Procedures for Established Levels of Reliability and Confidence in Electronic Part Specifications," proposed in 1965, required zero test failures in 230 million part hours to meet a standard of 0.001% failures per thousand hours at the 90% confidence level. Testing as many as 10,000 parts simultaneously would require 4.5 years of testing 24 hours per day. But recent electronic technology has been consistently undergoing major revolutions at intervals of approximately 5 years. We must conclude that a product which has been adequately tested according to conventional standards may be obsolete by the time it satisfies the testing criteria. Thus, complexity of equipment and high performance requirements conspire to eliminate the possibility of observing the survival characteristics of system replicas in sufficient quantity for statistical analysis of sample variation to be a valuable guide.

Although it is common to view statistics as an analytical arsenal for the description of observed variations in large samples of homologous items subjected to similar environmental stresses, there is another, more profound, view introduced into statistical mechanics by J. Willard Gibbs. Prior to Gibbs, the application of statistical methods to the study of physical reality was beset with philosophical problems arising from the irrefutable observation that there is but one universe, not a s 'e of universes the variation of whose properties statistics would describe. It was Gibbs who conceived the fruitful notion of a virtual ensemble of potential universes upon which statistical analysis acted to select one – the one that exists – as a kind of solution to a variational problem, the problem of maximizing expectation. In this way statistics is applied as a cardinal principle in our model of nature, on somewhat the same footing as Newton's Laws, to determine which among the conceivable universes shall occur; it is not a descriptive tool to provide a measure of observed variation. Elevated to a principle, statistics nevertheless cannot determine the course of nature without additional information, just as application of Newton's Laws requires knowledge of the appropriate force function.

The statistics of traditional reliability theory has few points of contact with the Gibbsian interpretation; it is woven together with product sampling and age exploration. When these are not possible, when the system is complex, unreplicated, and rapidly becomes obsolete, then application of statistics as a means for the analysis of variation must yield to the Gibbsian role of statistics as a selection principle.

These remarks neither solve any problem of reliability nor yield profound insights. But they perhaps suggest a philosophical foundation upon which an acceptable theory of the application of statistics to the reliability of complex systems can be developed.

<u>7.4</u> Recalling the ideas and notations of Section 7.2, we recognize that the step size used in implementing the Reliability-Centered Maintenance Program depends on the policy selected and also on the time of selection of the policy. A point on the maintenance/failure cost surface S_t corresponding to time t is specified by the policy parameters, which will be collectively denoted by $p(t)$, and the corresponding cost, say $C(t, p(t))$. Thus the corresponding point on S_t has coordinates $(p(t), C(t,p(t)))$; and, when it is considered as a point on the full policy surface S, its coordinates are $(t,p(t), C(t,p(t)))$ with time as an explicit variable. Selection of a step is the same thing as selection of a pair of points on S, say $(t',p'(t'), C(t',p'(t)))$ and $(t'',p''(t''), C(t'',p''(t'')))$. The time variable plays its usual distinctive role since it is subject to unicursal variation: time always increases. This implies that of two applications of the minimizing maintenance policies (I) – (III) of Section 6, one will always be antecedent to the other; we can suppose, without loss of generality, that $t' < t''$. It may happen the policy p remains unchanged from t' until t'': that is, a review of policy may not bring forth sufficient reasons to implement a policy change. The process of review, and the process of implementation of a policy change, may be costly, which is an inducement to extend the interval $t''-t'$ between successive reviews or changes as much as possible. Counterbalancing this argument is the possibility that a review will lead to a substantial cost decrease, i.e., that there will be sufficient information to enable the size and direction of the next step in the

iterative minimization procedure to be determined. For these reasons the problem of determining the step size in the time variable, that is, of determining the interval $t''-t$ between successive applications of the policies (I) – (III) to the system, assumes a particularly significant role. An intensively studied special case of this problem is concerned with the extension of Hard Time replacement intervals for equipment as experience accumulates.

Determination of the optimal intervals for application of the Reliability-Centered Maintenance Program policies appears to be a particularly difficult problem, depending as it does on both the conversion of operating experience into information about the survival distributions of the elements of the partition of the system, and on the effect this information should or would have on those who bear the responsibiltiy for making policy changes such as increasing replacement or inspection intervals. We have already noted that larger than optimal step sizes can lead to wild oscillations in maintenance/failure costs and to an increasing number of critical failures, whereas smaller than optimal step sizes, which can also be called conservative estimates, merely reduce the rate of approach to the optimal policy. This is a persuasive argument for a conservative implementation of a maintenance program. Excessive conservatism, however, is often too costly and retards the evolution of related systems. It is therefore worthwhile to try to formulate the decision process in a manner which makes it subject to analysis.

One way to formalize the problem of interval determination is based upon its connection with information theory. Let $t_1, t_2, \ldots, t_n$ be a sequence of inspection or replacement times for samples of a type of item. Let $R(t)$ be the observed survival distribution and $\underline{\underline{U}}$ the universe of sample items. If $\xi \varepsilon \underline{\underline{U}}$ is an item, it will age and finally fail at some time $t(\xi)$. Let

$$\omega(t_i) = \{\xi : t_{i-1} \leq t(\xi) < t_i\} , \; i=1,\ldots,n, \text{ with } t_0=0 \quad (7.2)$$

that is, let $\omega(t_i)$ denote the set of items which fail before the i^{th} inspection but not before the $(i-1)^{th}$ inspection. The sets

$\{\omega(t_i): i=1,2,\ldots,n\} = \Omega$ constitute a partition of $\underline{\underline{U}}$. The probability that $\xi\epsilon\omega(t_i)$ is $R(t_{i-1}) - R(t_i)$. The <u>information</u> associated with the partition is (cp.[2] [11])

$$I(\Omega) = -\sum_i \Big(R(t_{i-1}) - R(t_i)\Big) \log_e\Big(R(t_{i-1}) - R(t_i)\Big) \quad . \tag{7.3}$$

Passing to continuous variables, this corresponds to (cp.[11])

$$I(\Omega) = -\int_0^\infty \rho(t)\log_e \rho(t)dt \quad . \tag{7.4}$$

Note that the information defined by eq. (7.4) depends on the coordinate system; it is not independent of transformations of the time variable, among which selection of zero time is included. In particular, the information corresponding to a continuous survival probability density may be negative. Information differences do have absolute meaning, independent of coordinate transformations.

It can be shown that, among all differentiable survival probability densitites which have the same mean time before failure T,

$$T = \int_0^\infty t\rho(t)dt \quad , \tag{7.5}$$

the exponential survival distribution, for which

$$\rho(t) = \frac{1}{T} \exp\ (-t/T) \quad , \tag{7.6}$$

maximizes the information eq. (7.4). A simple calculation shows that in this case

$$I = 1 + \log_e T. \tag{7.7}$$

The information corresponding to the exponential distribution and inspection intervals of equal duration can be easily calculated. Let the inspection times be

$$t_i = iqT \quad , \quad i=0,1,\ldots \tag{7.8}$$

where T denotes the mean time before failure and q is a positive constant. The inspection intervals have common duration $t_{i+1} - t_i = qT$, and the survival distribution is given by eq. (7.6). From the formulae

$$\sum_{i=1}^{\infty} x^i = \frac{x}{1-x}$$

and

$$\sum_{i=1}^{\infty} ix^i = \frac{x}{(1-x)^2} \quad ,$$

each valid for $-1 < x < 1$, we find, from eq. (7.3):

$$I = I(q) = -\sum_{i=1}^{\infty}\left(e^{-iq} - e^{-(i+1)q}\right) \log_e\left(e^{-iq} - e^{-(i+1)q}\right) \tag{7.9}$$

$$= (1-e^{-q})\left\{\sum_{i=1}^{\infty} qie^{-iq} - \log_e(1-e^{-q})\sum_{i=1}^{\infty} e^{-iq}\right\}$$

$$= \frac{q}{e^q-1} - e^{-q}\log_e(1 - e^{-q})$$

As the inspection interval tends to zero, the discrete formula eq. (7.3) does not pass over to formula eq. (7.4) corresponding to the information associated with an absolutely continuous distribution, so one should not expect that $\lim_{q \to 0} I(q)$ will reduce to eq. (7.7); instead we find

$$I(0) = \lim_{q \to 0} I(q) = \infty \quad ; \tag{7.10}$$

periodic inspection with zero interinspection interval produces infinite information for the exponential distribution. At the other extreme, if there are no inspections, which is equivalent to the condition that q is infinite, then

$$I(\infty) = \lim_{q \to \infty} I(q) = 0 \quad ; \tag{7.11}$$

the information gain is zero. These calculations agree with our intuitive assessment of the situation.

$I(q)$ decreases from infinity to zero as the interinspection interval increases from zero (continuous inspection) to infinity (no inspection). For inspection intervals equal to T we find

$$I(1) = \frac{1}{e-1} - \frac{1}{e} \log_e (1 - \frac{1}{e}) = 0.750^{+} \tag{7.12}$$

Our objective is to determine inspection intervals so that there is some desired relationship between them and the corresponding measure of information.

Let the inspection times $t_1, t_2, \ldots, t_{n-1}$ be given and let it be required to determine t_n. Moreover, consider further inspection times t_{n+k}, $k=1,2,\ldots$, corresponding to equal inspection intervals

$$t_{n+k+1} - t_{n+k} = qT, \; k=1,2,\ldots \quad ; \tag{7.13}$$

we will later let q approach infinity, and it will follow from the calculations previously given that the information corresponding to the latter intervals will be zero. The information corresponding to the partition induced by $\{t_1, t_2, \ldots\}$ is

$$I = -\sum P_i \log_e P_i \tag{7.14}$$

where we have abbreviated

$$P_i = R(t_{i-1}) - R(t_i) \quad . \tag{7.15}$$

If the desired relationship among the intervals is that each inspection interval produces the same amount of information, then the condition is

$$P_i \log_e P_i = \text{const.} = P_1 \log_e P_1$$

for all i. If the survival distribution is exponential and the first inspection time is t_1, then t_2 is determined by the equation

$$\left(1 - e^{-t_1/T}\right)\log_e\left(1 - e^{-t_1/T}\right) = \tag{7.17}$$

$$\left(e^{-t_1/T} - e^{-t_2/T}\right)\log_e\left(e^{-t_1/T} - e^{-t_2/T}\right)$$

This is equivalent to an equation of the form

$$(1 - e^{-x})\log_e(1 - e^{-x}) = \text{const.} \quad , \tag{7.18}$$

where $x=(t_2-t_1)/T$, and can be solved by numerical methods.

The left side of eq. (7.17) is known from observations obtained through time t_1. By monitoring $R(t)$ throughout the interval $t_1 \leq t$, one can always calculate when the incremental information satisfies eq. (7.16), which establishes t_2 and the successive intervals.

In general, if there is infant mortality, then $t_2-t_1 > t_1$; it will take longer to acquire additional information about the survival distribution after the epoch of infant mortality has been outlived. Similarly, should a wear-out period exist for aged items, inspection intervals established by the principle of equal information will be comparatively shortened to compensate for the increased hazard rate. However, the reduction may have a negligible practical effect because there may be too few items surviving until advanced ages to significantly affect total fleet maintenance cost. This is in accord with experience.

The policy outlined above <u>responds to failure</u>; cp. Chapter 6.

As items in the initial inventory fail, they may be renewed and returned to service, or replaced by new items of the same type. Additions to the operational inventory of items may also be made at various times and in varying quantities. As a consequence, the oldest items in operation are likely to constitute only a small fraction of the "fleet" even if the failure rate is low.

Additions to the operational inventory and renewal of failed equipment creates complex, unpredictable, and continually varying age distributions. Figure 7.1 illustrates an age distribution, with each renewed item returned to service treated as distinct from all others, including its pre-failure form. t is a measure of operational time and t^{chron} denotes chronological time. The total operating time until failure of item i is denoted by t_i. In this figure, item No. 4 might be a renewed version of the failed item No. 3; No. 5 is a non-initial acquisition which has not failed during the span of chronological time displayed in the figure.

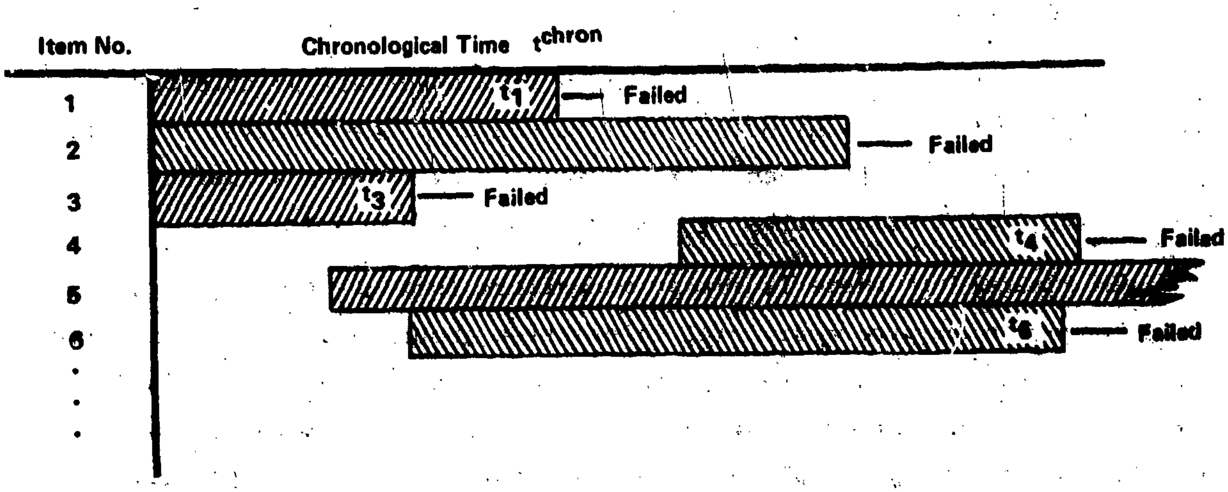


Figure 7.1. An Age Distribution - Chronological Display

If the information provided in the figure is displayed in terms of operating time t, then it can be arranged as shown in Figure 7.2.

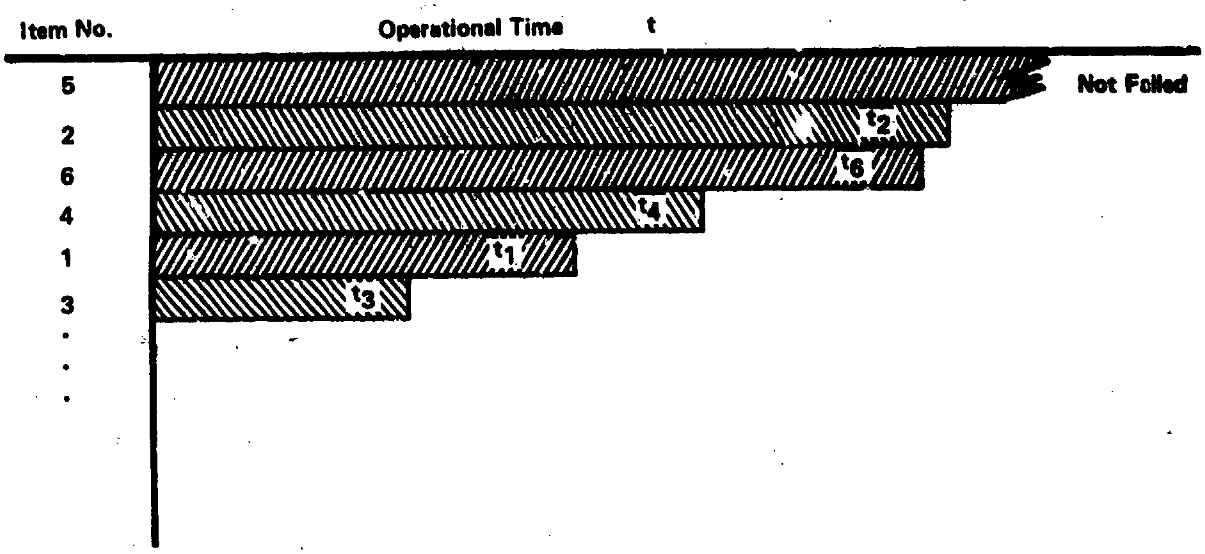


Figure 7.2. An Age Distribution - Operational Display

The survival distribution R(t) can be estimated from this data for all t not greater than the operational age of the oldest item in service. If failed items are renewed and returned to service, the sample size for estimation of R(t) for small t will generally be significantly larger than the total inventory of items since given renewed items share multiple operating histories. Since an estimate of R(t) is given by the fraction of items surviving until t, as experience accumulates, renewed items are returned to the operating inventory, and new items are acquired, the estimates of R(t) for small t can be repeatedly updated. As data accumulates, the estimates of R(t) will stabilize; thus, replenishment and expansion of the operating inventory only act to refine the estimate of R(t) and reduce its variance. Since the failure information measure I of eq. (7.3) is completely determined by R(t) and the inspection intervals, it follows that the estimate of I is independent of replenishment and expansion of the inventory except that as chronological time passes, the estimates of I for small operating times become increasingly reliable.

REFERENCES

1. Davis, D. J. An analysis of some failure data. JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION. 47(258):113-150; 1952 June.

2. Dolby, J. L. On the notions of ambiguity and information loss. INFORMATION PROCESSING AND MANAGEMENT. 13(1):69-77; 1977.

3. Eggwertz, Sigge; Lindsjo, Goran. STUDY OF INSPECTION INTERVALS FOR FAIL-SAFE STRUCTURES. 6th ICAS Congress, Munich, Germany; 1968 September 13; Stockholm: FFA, Flgtekniska Forsoksanstalten, Aeronautical Research Institute of Sweden; c1970 January. FFA-120.

4. Hoel, P. G. INTRODUCTION TO MATHEMATICAL STATISTICS, 2nd ed. New York: John Wiley & Sons; 1954.

5. Kolmogorov, A. Interpolation und Extrapolation von stationären zufälligen Folgen. BULL. DE L'ACAD. SCI. U.R.S.S. 30:13-17; 1941.

6. Nowlan, S. and Heap, H. RELIABILITY-CENTERED MAINTENANCE. Part I. To appear.

7. Nowlan, S. and Heap, H. RELIABILITY-CENTERED MAINTENANCE. Part II. To appear.

8. Parzen, E. MODERN PROBABILITY AND ITS APPLICATIONS. New York: John Wiley & Sons; 1960.

9. Resnikoff, H. L. On the psychophysical function. JOURNAL OF MATHEMATICAL BIOLOGY. 2:265-276; 1975.

10. Rosenthal, Arnie. Computing the reliability of complex structures: SIAM JOURNAL OF APPLIED MATHEMATICS. 32:384-393; 1977.

11. Shannon, E. E. A mathematical theory of communication. BELL SYSTEM TECHNICAL JOURNAL 27;379-423, 623-656; 1948.

12. Smirnov, V.I. A COURSE OF HIGHER MATHEMATICS. Vol V. Reading, MA: Addison-Wesley Publishing Company; 1964.

13. Trumpler, R. J. and Weaver, H. F. STATISTICAL ASTRONOMY. Berkeley, CA: University of California Press; 1958.

14. Weibull, W. A statistical distribution function of wide applicability. JOURNAL OF APPLIED MECHANICS. 18:293-297; 1951.

15. Wiener, N. EXTRAPOLATION, INTERPOLATION, AND SMOOTHING OF STATIONARY TIME SERIES, WITH ENGINEERING APPLICATIONS. Cambridge, MA: M.I.T. Press; 1949.

16. Wiener, N. NONLINEAR PROBLEMS IN RANDOM THEORY. New York: John Wiley & Sons; 1958.

GLOSSARY OF NOTATIONS AND TERMINOLOGY

Notations

$\in$	Set membership symbol. For '$x \in S$' read 'x is an element of S' or 'x belongs to S'.
$\cup$, $\bigcup$	Set union symbol. $S \cup T$ is the set whose elements belong to at least one of S and T.
$\cap$, $\bigcap$	Set intersection symbol. $S \cap T$ is the set whose elements belong to both S and T.
$-$	Set difference symbol. $S - T$ is the set whose elements belong to S but not to T.
$\subset$	Set inclusion symbol. $S \subset T$ signifies that each element of S is also an element of T.
$\emptyset$	Empty set.
$\gamma_\lambda(t)$	Maintenance/failure cost density of element λ of a partition Λ of system $\underline{\underline{S}}$ with respect to the failure distribution $F_\lambda(t)$; eq. (6.7).
$\delta(x)$	Dirac delta (generalized) function; eq. (2.27).
$\eta(t)$	Hazard rate, also called failure rate; eq. (3.14).
λ	Typical element of the partition Λ of system $\underline{\underline{S}}$; eq. (6.2).
Λ	Partition of a system $\underline{\underline{S}}$; eq. (6.2) and Figure 6.2.
μ	Typical element of the partition M of system $\underline{\underline{S}}$, where M is a refinement of Λ; Figure 6.3.
M	Partition of a system $\underline{\underline{S}}$ which refines another partition Λ; Figure 6.3.
$\rho(t)$	Failure probability density; eqs. (3.8), (3.9).

$\rho_\lambda(t)$	Failure probability density of element λ of partition Λ of system $\underline{\underline{S}}$; eq. (6.6).
ω	Event in a measurable space; eq. (2.1).
$\omega(t)$	Set of items which have failed prior to time t; eqs. (3.1), (7.2).
Ω	Collection of events in a measurable space; eq. (2.1).
$c_\lambda(t)$	Cost density with respect to time, corresponding to cost function $C_\lambda(t)$ for partition element λ; eq. (6.4).
$c_\lambda^f(t)$	Imputed cost density of failure of partition element λ per unit hazard rate of λ; eq. (6.6).
$c_{\lambda,i}^m(t)$	Cost density of maintenance of partition element λ corresponding to inspection time t_i; eq. (6.5).
$c_\omega(\xi)$	Indicator function of event ω; eq. (2.7).
$C(t)$	Maintenance/failure cost function for the complex system $\underline{\underline{S}}$; eq. (6.8).
$C_\lambda(t)$	Maintenance/failure cost function for the element λ of partition Λ of the complex system $\underline{\underline{S}}$; §6.3.
$F(t)$	Distribution function for failure prior to time t; eq. (3.5).
$F_\lambda(t)$	Distribution function for failure of partition element λ prior to time t; §6.3.
$\underline{\underline{F}}(\omega(t))$	Probability of failure prior to time t; eqs. (3.3), (3.5).
$I(q)$	Information corresponding to an exponential survival distribution and inspection intervals of duration qT with T the mean time before failure; eq. (7.9).
$I(\Omega)$	Information corresponding to discrete partition Ω; eq. (7.3).
$p(x)$	Probability density function corresponding to the probability distribution $P = P_f$ of the random variable f. The random variable is usually suppressed from the notation; eq. (2.18).

P	Distribution function of a fixed random variable (not indicated by the notation), relative to the probability measure $\underline{\underline{P}}$; eqs. (2.12), (2.13).
P_f	Distribution function of random variable f relative to the probability measure $\underline{\underline{P}}$; eq. (2.12).
P^{abs}	Absolutely continuous distribution function; eq. (2.16).
P^{dis}	Discrete distribution function; eq. (2.16).
P^{sing}	Singular distribution function; eq. (2.16).
$\underline{\underline{P}}$	Probability measure; eq. (2.1).
$\underline{\underline{P}}(\omega_2 \| \omega_1)$	Conditional probability of event ω_2 given event ω_1; eq. (2.35).
$R(t)$	Distribution function for survival until time t, also known as the reliability; eq. (3.6).
$R_{\underline{\underline{S}}}(t)$	Distribution function for survival of system $\underline{\underline{S}}$ until time t; eq. (6.3).
$R_\lambda(t)$	Distribution function for survival of partition element λ of $\underline{\underline{S}}$ until time t; eq. (6.3).
$\underline{\underline{R}}(\omega(t))$	Probability of survival until time t; eq. (3.2).
$\mathbb{R}$	Set of real numbers.
S	Maintenance/failure cost surface; eq. (7.1).
S_t	Maintenance/failure cost surface for time t; eq. (7.1).
$\underline{\underline{S}}$	Set of items which constitute a complex system; eq. (6.2).
T	Mean time before failure; Figure (3.3), eq. (7.5).

Terminology

Bathtub curve - Typical shape of a hazard function graph; Figure 5.1.

Bayes' Principle of Inverse Probability - Figure 2.6, eq. (2.39).

Condition Monitoring - One of the three primary maintenance processes, consisting of no scheduled preventive maintenance. Condition monitoring depends on the surveillance and analysis program for data collection and data analysis, upon which judgements can be made relative to maintaining items; see [6].

Conditional probability - eq. (2.35).

Conditional probability of failure - eq. (3.13).

Distribution function - eq. (2.13).

Event - eq. (2.1).

Exponential survival distribution - §4.1

Failure probability density - eq. (3.8).

Failure rate - Same as hazard function; eq. (3.14).

Gamma survival distribution - §4.5.

Hard Time - One of the three primary maintenance processes, requiring fixed-limit removal for overhaul or time limits; see [7].

Hazard function - Same as failure rate; eq. (3.14).

Information - A measure of the organization of elements of a set associated with some partition. Modern Information Theory was developed by Claude Shannon in connection with communication systems during the 1940s. Soon thereafter its relation to older ideas in statistical mechanics and statistics was recognized, and its fundamental role throughout the physical sciences was elaborated in numerous articles and books, among which those by L. Brillouin and E. Schroedinger are particularly worthy of mention. Measures of information are now systematically employed in fields as diverse as linguistics and psychophysics, biology and physics, communication engineering and library science. Although originally conceived in the context of transmission of sequences of symbols drawn from a finite inventory with fixed probabilities, the concept of information is more general and can be associated with any partition of a finite set, and in certain instances with infinite sets as well. See [6], especially eqs. (7.3) and (7.4), and references [2], [11].

Independent random variables - §2.4, eq. (2.33).

Lebesque-Stieltjes integral - §2.3, eq. (2.22). See also [12] for a more general and comprehensive development.

Likelihood ratio - eq. (2.39).

Lognormal survival distribution - §4.4.

Maximum-likelihood method of estimation - eq. (2.40).

Normal survival distribution - §4.2.

On Condition - One of the three primary maintenance processes, requiring repetitive inspections or tests to determine reduced resistance to failure for specific failure modes.

Probability density function - eq. (2.18).

Probability of failure - §3.1.

Probability of survival - eq. (3.2).

Random variable - Paragraphs following eq. (2.3) and Figure 2.2.

Weibull survival distribution - §4.3.