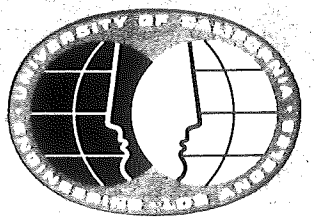
UCLA-ENG-7464
SEPTEMBER 1974



Prepared for the National Science Foundation
Under Grant GI-39416

"A General Evaluation Approach to Risk-Benefit
for Large Technological Systems and its
Application to Nuclear Power"

David Okrent, Project Director

MATHEMATICAL METHODS OF PROBABILISTIC SAFETY ANALYSIS

G.E. APOSTOLAKIS

UCLA • SCHOOL OF ENGINEERING AND APPLIED SCIENCE

published by

REPORTS GROUP
SCHOOL OF ENGINEERING AND APPLIED SCIENCE
UNIVERSITY OF CALIFORNIA, LOS ANGELES 90024

UCLA-ENG-7464
SEPTEMBER 1974

MATHEMATICAL METHODS OF PROBABILISTIC SAFETY ANALYSIS

by

George Apostolakis

Prepared for the National Science Foundation
Under Grant GI-39416

"A General Evaluation Approach to Risk-Benefit
for Large Technological Systems and its
Application to Nuclear Power"

David Okrent, Project Director

School of Engineering and Applied Science
University of California
Los Angeles, California 90024

PREFACE

The summer of 1974 concluded the first year of the National Science Foundation funded study at UCLA entitled "A General Evaluation Approach to Risk-Benefit for Large Technological Systems and Its Application to Nuclear Power" (NSF Grant GI-39416). The objectives of this project can be defined to include the following:

1) To make significant strides in the provision of improved bases or criteria for decision-making involving risk to the public health and safety (where a risk involves a combination of a hazard and the probability of that hazard).
2) To make significant strides in the structuring and development of improved, and possibly alternative, general methodologies for assessing risk and risk-benefit for technological systems.
3) To develop improvements in the techniques for the quantitative assessment of risk and benefit.
4) To apply methods of risk and risk-benefit assessment to specific applications in nuclear power (and possibly other technological systems) in order to test methodologies, to uncover needed improvements and gaps in technique, and to provide a partial, selective, independent assessment of the levels of risk arising from nuclear power.

The first year's effort has, to a considerable extent, involved reviews of some of the several fields of interest to the overall study. Beyond this, the work has largely been exploratory in nature and been concentrated in specific parts of the overall problem.

This UCLA Engineering report is one of a group of several which represent either completed reviews or interim reports on those segments of the exploratory research which have reached a stage suitable for publication. Publication of these reports has been expedited, accepting the possible loss of better editorial form and potential technical improvement, in order to make the information rapidly available.

TABLE OF CONTENTS

LIST OF FIGURES

Figures

Figures

Figures

LIST OF TABLES

1. INTRODUCTION

The aim of this report is to present in a systematic way the mathematical methods which are useful in reliability and safety studies. The bases upon which these methods are built are probability theory, statistics and logic.

In general, the problems that these studies deal with concern the prediction of the probability that a specified function will be performed satisfactorily over a period of time or per demand and the identification of events and their probabilities, which may lead to unfavorable circumstances endangering the health of the public.

The vast number of factors that influence the performance of the systems which perform the required functions and the impossibility of knowing a priori their time-, space- and magnitude-behavior suggest that the natural way to handle these problems is through probabilistic methods. The parameters appearing in the analysis are estimated from the past performance of similar systems functioning under similar conditions as the system under study. This estimation is, of course, done with the aid of statistical techniques.

Besides overcoming the problem of needing to know what happens exactly, a probabilistic analysis is very useful in that it assigns probabilities to the various possibilities whenever known uncertainties are encountered instead of assuming that the "worst" will happen. When many events appear to be possible at a certain stage of the analysis they are all included in it and they are combined with the use of elementary operations of mathematical logic. Such a case occurs frequently in the study of complex systems and it provides with a systematic way of identifying sequences of events that may lead to dangerous situations as well as their probabilities of occurrence. These probabilities form the basis of assessing the risk to the public (usually risk is defined as the product of the probability of an accident

times some measure of its consequences) and accordingly decide to eliminate or accept the risk. This decision is made with the use of a prespecified criterion, which is possible to establish only because the risk has been quantified.

Unfortunately, despite the merits of probabilistic methods, when one tries to apply them in real cases there are serious drawbacks that are revealed. The most important is the lack of statistically significant data upon which the models can be built. Many systems are new, thus no information regarding past experience is available; even standard components of the systems may operate under conditions which are unique to the case under study thus impairing the validity of existing data; people did not bother or did not have the necessary tools to make accurate measurements in the past (e.g., earthquake histories are, in general, very poor), and so on. Special statistical techniques are developed to deal with some specific situations, but the most common "solution" is to make drastic assumptions and use judgement which, of course, reduce our confidence on the methods and their results.

This report deals mainly with the mathematics involved in quantitative assessments. Summaries of illustrative applications are included and references to the literature are listed for further details. The second chapter describes the methods of handling problems involving one component or simple logical configurations and its contents are more or less what it is known as reliability theory. To make the report self-contained an introduction to the fundamentals of probability theory and statistics is given; then the modeling of the failure of components by statistical distributions is discussed followed by the mathematical description of various maintenance policies. Renewal theory and Markov processes are examined in detail, since they are important mathematical tools in safety studies.

The third chapter departs from conventional reliability theory in that it deals with the analysis of complex systems. The fault-tree methodology is developed in detail and its uses and limitations are investigated. Methods, like the failure modes and effects analysis are described even though they are not strictly mathematical, because they form an important part of a safety study. The special problems arising from software and human errors as well as the possibility of common mode failures are also discussed. Finally the use of statistical techniques to handle major natural phenomena and methods of dealing with systems without exploiting their logical structure are investigated. Throughout the report references are given where a more extensive discussion of the various topics can be found.

This report was written prior to the release of Draft WASH-1400 "Reactor Safety Study. An Assessment of Accident Risks in U. S. Commercial Nuclear Power Plants" and the methodological aspects and applications of WASH-1400 have not been factored into this report.

2. ANALYSIS OF COMPONENTS AND SIMPLE SYSTEMS

2.A ELEMENTS OF PROBABILITY THEORY AND STATISTICS

2.A.1 The Probability Concept

In the development of methods for Quantitative Safety Analysis extensive use of the concepts of probability Theory ("probability", "event" etc.) is made. It is essential that such terms are well understood for a successful application of the methods. For a complete analysis of probabilistic ideas the reader is referred to the book of Feller[1] and for more concise treatments to any standard textbook.[2,3,4] A brief summary of the important ideas is presented here.

There are various interpretations of probability from the strict mathematical formulation to the intuitive concept found in the average person. Except for the former the others are incomplete but it is worth mentioning them for they help to clarify things.

Subjective Interpretation: Probability is a measure of the belief that a person has to the truthfulness of a certain statement. In this sense probability reflects one's judgement and state of knowledge. Examples from everyday life are statements like "I am sure it will rain tonight," "The odds against X are four to one" etc.

Empirical Interpretation: The probability of an event A is the limit of the frequency $\frac{n}{N}$ as $N \to \infty$, where N is the number of times an experiment was repeated and n is the number of times the event A occurred. Although this interpretation is very common among applied scientists it is not sufficient. An obvious limitation is the requirement of having a large number of experimental data; in addition some questions regarding the existence of the above limit can be raised.

Classical Interpretation: If N is the number of all possible outcomes of an experiment and the event A can occur n times, the probability of A is $\frac{n}{N}$,

if all outcomes are equally likely. In this definition the experiment does not have to actually be performed. However the assumption of equally likely outcomes poses severe restrictions. In many cases it is difficult to establish its validity or, even worse, it is already known that the outcomes are not equally likely (e.g. an experiment with a loaded die). Other objections against the classical interpretation[2] include the fact that it is circular (the statement "equally likely" actually means "equally probable," but it is the meaning of "probable" that is attempted to be defined) and that in many experiments the number of all possible outcomes is infinite.

The difficulties which arise with the above definitions have led to the axiomatic formulation of the theory of probabilities. This in turn requires knowledge of the abstract ideas of measure theory. However it is possible to outline the formalism using elementary set theory. This will be done in the subsequent sections. At this point it must be emphasized that in the axiomatic treatment probabilities are assumed to be given parameters the actual numerical values of which are of no concern to the theory, in the same way that masses are treated in classical mechanics.[1]

2.A.2 Axiomatic Formulation and Basic Theorems

In the axiomatic formulation of any branch of mathematics there are certain concepts which are considered intuitive and remain undefined (e.g. the points of geometry). In probabilities such a concept is that of the sample space and sample point. Every possible outcome of an experiment is represented by one and only one sample point; the set of all sample points forms the sample space S. An event is defined as a collection of sample points, that is, it is a subset of S. For example, the sample space $S \equiv \{1,2,3,4,5,6\}$ represents all possible outcomes of throwing a die. The subset $A = \{1,3,5\}$ represents the event "the outcome of the experiment is an odd number."

From the previous paragraph it is clear that sets play an important role in this theory. Therefore some of their properties are of particular interest. All the sets will be assumed to be subsets of the sample space S.

i. The null set contains no sample points. It is denoted as ϕ and it represents an event that can never happen.

ii. The complement of an event A, written as $\overline{A}$, is the set which contains all the sample points of S not in A.

iii. The union of two events A and B is a third event C consisting of all the elements of either A or B or both (taken once). It is denoted as A+B = C or $A \cup B = C$.

iv. The intersection of two events A and B is a third event C, written as AB = C or $A \cap B = C$, which contains all the common elements of A and B. If the intersection of two events is the null set, i.e. $AB = \phi$, then the events are called mutually exclusive. The extension of the properties iii and iv to more than two events is straightforward.

Before proceeding to the definition of probabilities it is important to distinguish between two types of sample spaces, those consisting of a finite number of sample points and those with an infinite number of sample points. For the purposes of reliability analysis this distinction will suffice.

If the sample space contains a finite number of sample points (e.g. the die experiment) any subset of S is an event; to each event A a non-negative number P(A) is assigned such that P(S) = 1 and if the events $A_1, A_2, \ldots, A_n$ are mutually exclusive, i.e. $A_i A_j = \phi$ for all i and j, then $P(A_1+A_2+ \ldots +A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$. Observe that if the mutually exclusive events $A_1 \ldots A_n$ exhaust the sample space, that is, $S = \sum_{i=1}^{k} A_i$ the above axioms imply that $\sum_{i=1}^{n} P(A_i) = 1$ and hence $0 \leq P(A_i) \leq 1$.

The most important case in reliability of a sample space with infinitely many sample points is that of the real line. The time-to-failure of equipments and systems generates such a space. Events are all the points of the line, $t = t_1$, and all closed or open intervals $t_1 \le t \le t_2$ or $t_1 < t < t_2$ respectively. Then probabilities are assigned just as before with an additional axiom: if the events $A_1, A_2, \ldots, A_n$, are mutually exclusive that is $A_iA_j = \phi$ for $i,j = 1,2,\ldots n\ldots$, then $P\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

The probabilities of events are now given in terms of the density function $f(t)$. This is a non-negative integrable and bounded function such that $P(S) = \int_{-\infty}^{\infty} f(t)dt = 1$. The probability of the event $A = t_1 \le t \le t_2$ is given by $P(A) = \int_{t_1}^{t_2} f(t)dt$. Taking $t_2 = t_1 + \varepsilon$ and letting $\varepsilon \to 0$ it is easy to see that $P(t = t_1) = 0$. It is clear then that $P(t_1 \le t \le t_2) = P(t_1 < t < t_2)$.

One of the axioms in the definition of probability concerns the union of mutually exclusive events. In general the events are not mutually exclusive and the probability of their union is given by a different formula. For two events A and B this formula is

$$P(A+B) = P(A) + P(B) - P(AB). \tag{2.1}$$

This relation becomes complicated if more than two events are involved. Thus the probability that at least one of the events $A_1, A_2, \ldots, A_n$ occurs is

$$P(A_1+A_2+\ldots+A_N) = \sum_{j=1}^{N} P(A_j) - \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} P(A_iA_j) +$$

$$+ \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \sum_{k=i+2}^{N} P(A_iA_jA_k) - \ldots + (-1)^{N+1} P(A_1A_2\ldots A_N) \tag{2.2}$$

In practice such a formula is rarely used as it stands. Very often the main contribution to the sum comes from the first several terms. By truncating the series bounds are readily obtained, such as

$$P\left(\sum_{i=1}^{N} A_i\right) \leq \sum_{i=1}^{N} P(A_i)$$

$$P\left(\sum_{i=1}^{N} A_i\right) \geq \sum_{i=1}^{N} P(A_i) - \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} P(A_i A_j) \quad \text{etc.} \tag{2.3}$$

Finally the concept of conditional probabilities is introduced: If $P(B) \neq 0$, the conditional probability of the event A under the hypothesis B (or, given that B has occurred) is defined by $P(A/B) = \frac{P(AB)}{P(B)}$. All the theorems of probabilities hold also for conditional probabilities; for example

$$P(A+B/C) = P(A/C)+P(B/C) - P(AB/C)$$

If two events are mutually exclusive, then $P(A/B) = P(B/A) = 0$, since $P(AB) = 0$. If $P(AB) = P(A)P(B)$, the events A and B are called (stochastically) independent

As an example of the use of conditional probabilities, assume that the probability density function of the time-to-failure of an equipment is $f(t)$. Then the (a priori) probability that failure will occur in the interval $t_1 < t < t_2$ is $P(t_1 < t < t_2) = \int_{t_1}^{t_2} f(t)dt$. This is the probability of the event "the failure will occur after time t_1 and before time t_2". A question that could be asked now is "assuming that the equipment has already survived past the time t_1, what is the probability that it will fail before time t_2?" This is now a conditional probability the hypothesis being that $t > t_1$. In mathematical terms this is interpreted as $P(t_1 < t < t_2/t > t_1)$. But $P(t > t_1) = \int_{t_1}^{\infty} f(t)dt$, therefore $P(t_1 < t < t_2 \mid t > t_1) = \dfrac{\int_{t_1}^{t_2} f(t)dt}{\int_{t_1}^{\infty} f(t)dt}$.

If the events $A_1,\ldots,A_n$ are mutually exclusive and $A_1+\ldots+A_n = S$ (the sample space), an event B can be analyzed in a unique way as $B=BA_1+BA_2+\ldots+BA_N$.

From this it follows that $P(B) = P(B/A_1)P(A_1) + P(B/A_2)P(A_2) + \ldots + P(B/A_N)P(A_N)$.

Furthermore $P(A_i/B) = \frac{P(A_iB)}{P(B)} = \frac{P(B/A_i)P(A_i)}{P(B)}$ or, using the expression for P(B):

$$P(A_i/B) = \frac{P(B/A_i)P(A_i)}{\sum_{i=1}^{N} P(B/A_i)P(A_i)}, \text{ Bayes' theorem} \qquad (2.4)$$

2.A.3 Random Variables and Distribution Functions

We have identified each possible outcome of an experiment with a point in the sample space. Sets of sample points form events. One way of describing events is simply to state in words what they represent, e.g. the event "heads" in the experiment of tossing a coin, the event "failure occurs before time t" in the study of equipment failure etc.

This way of identifying events is inconvenient especially in cases where the description of an event is lengthy. To simplify things we assign a number to each sample point through a unique way, that is we define a function on the sample space. Such a function is called a random variable (or, variate).

As an example consider the failure of equipments; the sample space consists of two points: the equipment is functioning and the equipment is failed. A random variable which is often defined on this space is

$$X = \begin{cases} 1 & \text{the equipment is functioning} \\ 0 & \text{the equipment is failed} \end{cases}$$

In the experiment with the die an obvious random variable is simply the number showing on the die, that is X takes on the values $\{1,2,3,4,5,6\}$.

Such random variables which take on a countable number of values are called discrete random variables and they are defined on discrete sample spaces.

If the sample space is continuous (e.g. the real line) the random variable will also be continuous and it will take on any value in an interval. For

example, the time-to-failure of an item is a continuous random variable defined on the positive real axis.

The representation of events is now greatly simplified with the use of inequalities. Given a random variable X and a number x the notation $X \leq x$ implies the event consisting of all the sample points at which X takes on values less than or equal to x. With this event we associate a function which equals the probability of the event, that is

$$F(x) = P(X \leq x)$$

and we call it the (cumulative) distribution function of the random variable X.

From this definition the following properties of distribution functions are clear:

$$\lim_{x \to -\infty} F(x) = 0$$

$$\lim_{x \to \infty} F(x) = 1$$

Furthermore F(x) is a nondecreasing function of x.

This definition holds for both discrete and continuous random variables. However the information provided by the distribution function may not be enough; we may want to know the probability that $X = x$ (for a discrete random variable) or that X falls between x and $x + \Delta x$ (for a continuous random variable). For this kind of information we must distinguish between discrete and continuous random variable.

The probability function (or probability distribution) of a discrete random variable X is defined as

$$p(x) = P(X = x)$$

and is related to the distribution function by

$$F(x) = \sum_{\substack{\text{all} \\ x_i \leq x}} p(x_i)$$

Notice that

$$\sum_{\text{all } x_i} p(x_i) = 1$$

As an example we return to the familiar die experiment; assuming that all outcomes are equally likely we have the probability function

$$p(x_i) = \frac{1}{6} \quad , \qquad X = i = 1, 2, \ldots , 6.$$

Then $F(3) = P(X \leq 3) = p(1) + p(2) + p(3) = \frac{1}{2}$

For a continuous variate we define its probability density function (or simply, the density function) as (see also 2.A.2)

$$f(x) = \frac{dF(x)}{dx}$$

Then, clearly, $F(x) = \int_{-\infty}^{x} f(x)dx$ and $\int_{-\infty}^{\infty} f(x)dx = F(\infty) = 1$. The density function itself does not have a probabilistic meaning, however $f(x)\Delta x$ is the probability that the random variable x falls in the interval $(x, x+\Delta x)$.

Distributions of continuous random variable are used extensively in reliability as models of the time-to-failure of systems.[5,6,7,8,9] Thus, we use the symbol T for the random variable time-to-failure, which has range from 0 to ∞, and we define several quantities of interest in the study of failures.

If F(t) is the distribution function of the time-to-failure of an equipment, then the function

$$R(t) = 1 - F(t) = \int_{t}^{\infty} f(\tau)d\tau$$

is the probability that the equipment will not fail up to time t and it is called its reliability. F(t) is sometimes referred to as the unreliability of the equipment.

Another important function is the hazard rate (or instantaneous failure rate, or failure rate function) which is defined as

$$h(t) = \frac{f(t)}{1-F(t)} = \frac{f(t)}{R(t)} \tag{2.5}$$

Its interpretation is that if the equipment has not failed up to time t then the probability that it will fail between t and $t + \Delta t$ is $h(t)\Delta t$, that is, it is a conditional probability.

From the definition it follows that

$$F(t) = 1 - e^{\int_o^t h(\tau)d\tau} \tag{2.6}$$

and $$R(t) = e^{\int_o^t h(\tau)d\tau} \tag{2.7}$$

Notice also that $h(t) = -\frac{1}{R(t)}\frac{dR(t)}{dt} = -\frac{d\ln R(t)}{dt}$

There is one point which should be clarified here; the expression $h(t)\Delta t$ is a conditional probability of failure in $(t, t + \Delta t)$, but $h(t)$ should not be interpreted as a probability density.[2,10,11]

To see the difference we need the notion of a conditional probability density. From the discussion in (2.A.2) the conditional distribution of the time-to-failure of a system given that it has survived past the fixed time t_1 is

$$F(t/t_1) = \begin{cases} \dfrac{\int_{t_1}^{t} f(\tau)d\tau}{P(T > t_1)} = \dfrac{F(t)-F(t_1)}{1-F(t_1)} & , \ t \geq t_1 \\ 0 \quad , \quad t < t_1 \end{cases}$$

Then the conditional failure density is

$$f(t/t_1) = \begin{cases} \dfrac{dF(t/t_1)}{dt} = \dfrac{f(t)}{1-F(t_1)} & , t \geq t_1 \\ 0 \qquad t < t_1 \end{cases}$$

The conditional density has all the properties of a probability density, for example

$$P(t \geq t^* \ / \ t \geq t_1) = \int_{t^*}^{\infty} f(\tau/t_1)d\tau \quad , \quad t^* \geq t_1$$

is the probability that the system will fail after time t^* given that it has survived up to time t_1.

Observe that if the time t_1 is allowed to vary so as $t_1 = t$, then the hazard rate results, that is, $h(t) = f(t/t)$, but this is not a conditional density anymore because the condition changes with the variable. A consequence of this is that $f(t/t_1)$ being a density satisfies

$$\int_0^\infty f(t/t_1)dt = 1$$

while for the hazard rate we have

$$\int_0^\infty h(t)dt \to \infty$$

as it can be seen from the fact that

$$F(\infty) = 1 - \exp\left(-\int_0^\infty h(t)\ dt\right) = 1$$

2.A.4. Measures of Central Tendency and Dispersion

Of great interest in the study of populations are certain quantities which are not as detailed as the distributions but summarize important information about them and give a feeling to the analyst of the most important properties of the population.

The most widely used is the expected value (or mean, or arithmetic mean, or average), which is defined as

$$m = E[X] = \begin{cases} \int_{-\infty}^{\infty} xf(x)dx, & \text{for a continuous random variable} \\ \sum_{\text{all } x_i} x_i p(x_i), & \text{for a discrete random variable} \end{cases}$$

If the density function (or the probability distribution) is interpreted as a mass distribution, then the expected value corresponds to the center of gravity of the distribution.

Notice that it is possible that the random variable never takes on its expected value. For example, when tossing an ideal coin we may describe the outcomes through the random variable.

$$X = \begin{cases} 1 & \text{for "heads"} \\ 0 & \text{for "tails"} \end{cases}$$

Thus $p(1) = p(0) = 1/2$ and $E[X] = 1 \cdot p(1) + 0 \cdot p(0) = 1/2$ while the random variable can be only 0 and 1.

Besides the expected value there two other measures of central tendency, the median and the most likely value, which are rarely used in reliability.

The median is defined as that point x_m for which

$$F(x_m) = P[X \leq x_m] = 0.5$$

Thus for a continuous random variable it is defined by

$$\int_{-\infty}^{x_m} f(x)dx = 0.5$$

and for a discrete random variable by

$$\sum_{x_i \leq x_m} p(x_i) = 0.5$$

The most likely value (or mode) is defined for a discrete random variable as the value which has the highest probability and for a continuous random variable as the value at which the density $f(x)$ is maximum.

The quantities defined above do not give any information regarding other important properties of the distribution such as its spread, symmetry, peakedness, etc. These can be described with the use of the moments.

The n^{th} moment (or, moment about zero) of a distribution is defined as

$$m_n = E\left[X^n\right] = \begin{cases} \int_{-\infty}^{\infty} x^n f(x)\,dx & \text{for a continuous random variable} \\ \sum_{\text{all } x_i} x_i^n p(x_i) & \text{for a discrete random variable} \end{cases}$$

Notice that the expected value is the first moment (for n = 1).

These moments have the disadvantage that they depend on the origin which is arbitrarily set; once we find the mean it is more meaningful to work with the central moments, that is, moments defined about the mean as follows:

$$\mu_n = E\left[(X-m)^n\right] = \begin{cases} \int_{-\infty}^{\infty} (x-m)^n f(x)\,dx & \text{for a continuous random variable} \\ \sum_{\text{all } x_i} (x_i-m)^n p(x_i) & \text{for a discrete random variable} \end{cases}$$

The second central moment (the first is zero) is of particular interest to our work; it is called the variance of the distribution and it is a measure of its spread. From the above definition it follows that

$$\sigma^2 = \mu_2 = E\left[(X-m)^2\right] = \begin{cases} \int_{-\infty}^{\infty} (x-m)^2 f(x)\,dx & \text{, x continuous} \\ \sum_{\text{all } x_i} (x_i-m)^2 p(x_i) & \text{, x discrete} \end{cases}$$

It can be proved that the variance is related to the mean and the second moment about zero through

$$\sigma^2 = m_2 - m^2 = E(X^2) - E^2(X)$$

The square root of the variance is the standard deviation. It has the same units with the random variable and it is used extensively as a measure of dispersion. In the extreme case where the random variable can take on only

one value the standard deviation is zero, as a simple calculation reveals. In general it is not zero and, as an example, we mention that for the normal distribution (to be defined shortly) the probability that the random variable falls between $m-3\sigma$ and $m+3\sigma$ is 0.997. For distributions for which only the mean and the variance is known the very general Tchebycheff inequality gives lower bounds to the probability that the random variable will take on a value in some interval. This inequality states that

$$P[m-\nu < X < m + \nu] \geq 1 - \frac{\sigma^2}{\nu^2} \tag{2.8}$$

where m is the expected value, σ^2 the variance and ν an arbitrary number. If ν is measured in units of standard deviation, i.e. $\nu = k\sigma$, the inequality is written

$$P[m-k\sigma < X < m + k\sigma] \geq 1 - 1/k^2 \tag{2.9}$$

Therefore, for k=3 we have $1 - \frac{1}{3^2} = 0.889$ and we say that there is a probability of at least 0.889 that the random variable will fall in the interval $(m-3\sigma, m+3\sigma)$ for any distribution (for the normal distribution we found that this probability was 0.997).

The third and fourth central moments are related to the symmetry (skewness) and peakedness (Kurtosis) of the distribution respectively, but they are not used much in reliability.

2.A.5 Populations and Samples

Thus far the discussion concerned all possible outcomes of an experiment which we represented by points in the sample space and we described their properties with the use of the distribution functions; in short, we talked about populations and their characteristics.

In practice we never deal with whole populations, but with small samples from them. If we know the characteristics of the population it is a simple

matter to make statements about the sample. Thus, if we know the distribution of the time-to-failure of a certain type of valves, we can estimate the probabilities of failure as a function of time of a sample of n such valves.

However the problem we most often encounter is the inverse of the above; given the times at which the valves of the sample failed we wish to know the distribution of times-to-failure of the population.

First we need some terminology.[12] If the random variables $X_1,\ldots,X_n$ are independent and they have the same density function f(x), they are said to constitute a random sample. The n values of the above example are a random sample. A statistic is a function of one or more random variables that does not depend on any unknown parameter. If we estimate a characteristic of a population from a sample, the value from the sample is the statistic and the estimated characteristic is called a parameter.

Two very common statistics of a sample are its mean and variance. The mean is defined as

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2.10}$$

and the variance as

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \tag{2.11}$$

Out problem can now be stated in mathematical terms as follows: we know that the random variable X has a density $f(x; \theta_1,\ldots,\theta_n)$ of known form but with unknown parameters and we wish to estimate these parameters from sample data.

There are two methods of approach: 1) we can calculate appropriate statistics from the sample and use them as estimates of the parameters of the population (point estimation), and 2) using information from the sample we can

find ranges of the parameters $\theta_1,\ldots,\theta_n$, thus selecting a family of possible densities $f(x,\theta_1,\ldots,\theta_n)$. Each set of parameters yields a density which may be the density of the population (interval estimation).

Point Estimation

There are various way that statistics from the sample can be used as point estimates of parameters of the population. To select the appropriate one we have several criteria:

i) the statistic should be unbiased, that is its expectation value should be the population parameter. The sample mean is an unbiased statistic, while the sample variance is not; however with a slight modification it can become unbiased; the appropriate form is

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} \qquad (2.12)$$

ii) the statistic should have variance as small as possible (efficiency property).

Combining these two properties we call a statistic "best" if it is unbiased and its variance is smaller than or equal to the variance of any other unbiased statistic for the parameter (fact which is not always easy to establish).

The moments about zero are unbiased statistics (e.g. the sample mean) but the central moments are not (however we changed the sample variance slightly and it became unbiased).

Having these criteria in mind (especially the requirement that the statistic be unbiased) we describe two common methods of point estimation: the moment matching method and the maximum likelihood method.

In the moment matching method we calculate the theoretical moments of the distribution as functions of the unknown parameters and we equate them with

the moments of the sample. If there are n unknown parameters, this is done for the first n moments (about zero or central) thus yielding a system of n equations in n unknowns. Usually there are one or two unknown parameters and we use the mean or the mean and the (unbiased) variance respectively.

In the maximum likelihood method we take as estimate of each parameter the value which is most likely on the basis of the available data. In mathematical terms, suppose we have a random sample $X_1,\ldots,X_n$ from a density $f(x;\theta)$.

We form the likelihood function

$$L(\theta;\ x_1,\ldots,x_n) = \prod_{i=1}^{n} f(x_i;\theta) \tag{2.13}$$

We now consider the x_i's as constants and θ as a variable. The maximum likelihood estimate of θ is the value which maximizes $L(\theta;\ x_1,\ldots,x_n)$. As an example, suppose that the density function of the time-to-failure of a population is

$$f(t;\theta) = \frac{1}{\theta}\, e^{-t/\theta} \qquad \text{(exponential density)}.$$

and that we have the sample $T_1,\ T_2,\ldots,T_n$ (that is n items failed at these times).

Then the likelihood function is

$$L(\theta;\ T_1,\ldots,T_n) = \frac{1}{\theta^n} \exp\left[\left(-\sum_{i=1}^{n} T_i\right)/\theta\right]$$

The maximum of L is found by simple differentiation and the estimate of θ is

$$\hat{\theta} = \frac{\sum_{i=1}^{n} T_i}{n}$$

which is the sample mean.

Interval Estimation

Here we do not give an estimate of the parameters but a range of possible values. This defines a family of possible population distributions. The degree of certainty that we have in our assertion that the parameters lie in a certain interval is called confidence level. So, we may claim that the mean time-to-failure (MTTF) of a population is between 5×10^6 hr. and 9×10^6 hr. at a 90% confidence level; this means that if we test many samples from the population, then their MTTF will fall in the above interval 90% of the time.

The mathematics of interval estimation is quite involved and will be introduced in subsequent sections.

2.A.6 Some Useful Distributions

We examine briefly several distributions of discrete and continuous type which we will frequently encounter. Their particular uses in the study of failures and in reliability will be given in more detail in subsequent chapters; here we state only their definitions and the underlying assumptions in each one.

Discrete Distributions

1. The Binomial Distribution

$$p(r) = \binom{n}{r} p^r (1-p)^{n-r} = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}$$

where $r = 0, 1, \ldots, n$ and $0 \leq p \leq 1$

$$F(x) = \sum_{r=0}^{x} \binom{n}{r} p^r (1-p)^{n-r} \quad , \quad x = 0, 1, \ldots, n$$

$$m = np \quad , \quad \sigma^2 = np(1-p)$$

The binomial distribution is used when an experiment can have only two outcomes (which are, naturally, mutually exclusive and exhaustive, like success-failure, heads-tails etc.); the probability of, say, success is p and of failure 1-p. The experiment is repeated n times and p is assumed constant throughout

(Bernoulli trials); then p(r) gives the probability of exactly r successes in n trials and F(x) the probability of at most x successes.

If we set q = 1-p, it is useful to notice that

$$(p+q)^n = \sum_{r=0}^{n} \binom{n}{r} p^r q^{n-r}$$

Suppose that we have three identical units working in parallel; each has a probability p of functioning properly and a probability q of not functioning. The system is working if at least two units are "up". It is an easy matter to find the reliability of the system using the expansion

$$(p+q)^3 = p^3 + 3\,p^2\,q + 3\,pq^2 + q^3.$$

The first and second terms give the probabilities of no failure and one failure respectively. Thus the reliability is

$$R = p^3 + 3\,p^2\,q$$

while the probability of system failure (unreliability) is

$$F = 3\,pq^2 + q^3.$$

Finally we point out the binomial distribution has two parameters, the probability of "success" p and the number of trials n.

2. <u>The Hypergeometric Distribution</u>

$$p(r) = \frac{\binom{k}{r}\binom{N-k}{n-r}}{\binom{N}{n}} \qquad r = 0, 1, \ldots, n, \quad r \le k, \quad n - r \le N-k$$

$$F(x) = \sum_{r=0}^{x} p(r) \qquad x = 0, 1, \ldots, n$$

$$m = \frac{nk}{N} \quad , \quad \sigma^2 = \frac{nk(N-k)\,(N-n)}{N^2\,(N-1)}$$

Inherent in the binomial distribution is the assumption that the n trials are made when there is a possibility of infinite trials; for example, we can view the n coin tossings as a sample from an infinite number of tossings.

Suppose now that the underlying population is not infinite and the sample of n trials is drawn from a population of N possible trials of which k are "success", then the probability function p(r) of the hypergeometric distribution gives the probability of exactly r successes in the sample.

The word "trial" may be replaced by the word "unit" and then p(r) gives the probability of exactly r "good" units in a sample of n units drawn from a lot of N units of which k are "good".

The hypergeometric distribution has three parameters, N, n and k. When $n << N$ (so that the drawn sample can be considered as drawn from an infinite lot), it approaches the binomial distribution with parameters n and $p = \frac{k}{N}$.

3. The Geometric Distribution

$$p(r) = (1-p)^{r-1}p \qquad r = 1, 2, \ldots, \qquad 0 \le p \le 1$$

$$F(x) = \sum_{r=1}^{x} (1-p)^{r-1}p$$

$$m = \frac{1}{p} \quad , \quad \sigma^2 = \frac{1-p}{p^2}$$

Again we deal with Bernoulli trials; p(r) is the probability of exactly (r-1) failures preceding the first success (p is the probability of success and it is the only parameter of the distribution).

4. The Pascal Distribution

$$p(r) = \binom{r+s-1}{r} p^s (1-p)^r \quad , \quad r = 0, 1, \ldots, \quad ; \quad s = 1, 2, \ldots$$

$$0 \le p \le 1$$

$$F(x) = \sum_{r=0}^{x} p(r)$$

$$m = \frac{s(1-p)}{p} \quad , \quad \sigma^2 = \frac{s(1-p)}{p^2}$$

Once more we deal with Bernoulli trials in which the probability of success is p. Then p(r) is the probability of exactly r failures and s successes in a total of r + s trials where the last trial is a success. There are two parameters of the distribution: p and s.

Using the equality

$$\binom{r+s-1}{r} = (-1)^r \binom{-s}{r}$$

we can rewrite the probability function as

$$p(r) = \binom{-s}{r} p^s (-1+p)^r$$

If s is equal to one the Pascal distribution reduces to the geometric distribution (with a slight modification: r failures, instead of (r-1), precede the first success).

5. The Poisson Distribution

$$p(r) = e^{-\lambda t} \frac{(\lambda t)^r}{r!} \quad , \quad \lambda > 0 \quad , \quad r = 0, 1, \ldots$$

$$F(x) = \sum_{r=0}^{x} p(r) \quad , \quad x = 0, 1, \ldots$$

$$m = \lambda t \quad , \quad \sigma^2 = \lambda t$$

The Poisson distribution describes phenomena which are of different nature from the ones described by the other discrete distributions. We no longer need the notion of Bernoulli trials, in fact such a notion is not meaningful when we talk about events like radioactive disintegration, number of persons arriving randomly at a bus-stop etc. Assuming that the rate of occurrence of the events is a constant λ we interpret p(r) as the probability that exactly r events will take place in the interval t.

The model can also be used when we deal with random events in the plane or space as long as the density λ of the points is a constant (for example λ may be the number of flows per unit volume of a material).

The term

$$p(0) = e^{-\lambda t}$$

is of special interest, since it represents the probability of no occurrence in the interval t, while the probability of at least one event in t is $1 - e^{-\lambda t}$. (The term $e^{-\lambda t}$ can also be viewed upon as a continuous distribution of times t and then it is called the exponential distribution; it is very important in reliability and it will be examined in detail later).

Finally we note that the Poisson distribution can be used as an approximation to the binomial distribution with $\lambda t = np$ for $n \to \infty$ and $p \to 0$.

Continuous Distributions

1. The Normal (Gaussian) Distribution

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(t-m)^2}{2\sigma^2}\right]$$

$$-\infty < t < \infty \quad , \quad -\infty < m < \infty \quad , \quad 0 < \sigma < \infty$$

$$F(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{t} \exp\left[-\frac{(\tau-m)^2}{2\sigma^2}\right] d\tau$$

The normal distribution is one of the most widely used. It has two parameters: the mean m, which specifies its position on the real axis (location parameter) and the variance σ^2, which determines its spread (scaling parameter). The distribution is symmetric (bell-shaped) about the mean (Fig. 2.1).

If we define the new variable

$$y = \frac{t-m}{\sigma} \qquad \text{(standard normal random variable)}$$

we observe that its distribution is again normal, i.e.

$$f(y) = \frac{1}{2\pi} \exp\left[-\frac{y^2}{2}\right] \tag{2.14}$$

but now there are no unknown parameters in f(y) (or, equivalently, $\sigma = 1$). This is helpful since we can tabulate the distribution function F(y) (ref. 4, 12) and a simple change of variables from y to t will yield values of F(t).

The great applicability of the normal distribution is due to a very general central limit theorem, which states that, under very general assumptions, the mean of a sample of n independent random variables which follow the same or even different distributions with finite mean and variance is normally distributed for large n. Therefore, if a random variable can be considered as the result of many independent causes, none of which dominates, then it is normally distributed. This property is used in the study of wear-out of equipments.

2. <u>The Log-Normal Distribution</u>

$$f(t) = \frac{1}{\sqrt{2\pi}\beta t} \exp\left[-\frac{(\ln t-\alpha)^2}{2\beta^2}\right]$$

$$-\infty < \alpha < \infty \quad , \quad \beta > 0 \quad , \quad t \geq 0$$

$$m = e^{\alpha + \beta^2/2} \quad , \quad \sigma^2 = e^{2\alpha + \beta^2}\left(e^{\beta^2}-1\right)$$

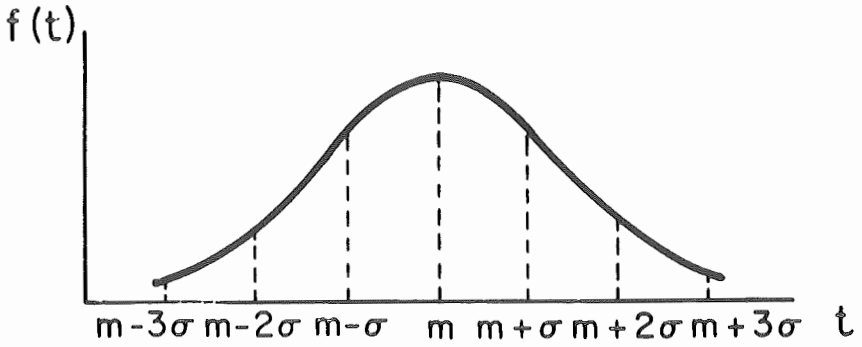


Figure 2.1. The Normal Distribution.

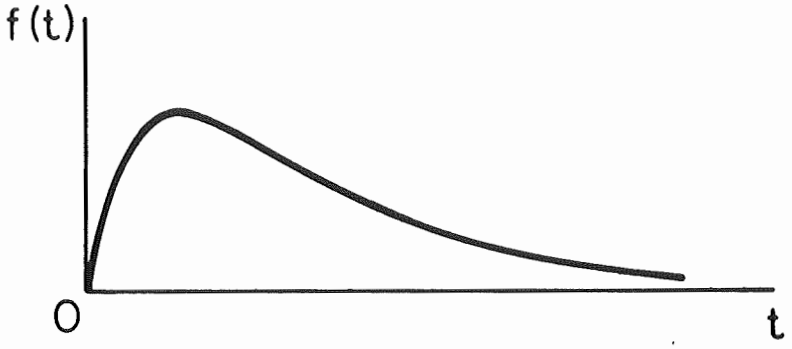


Figure 2.2. The Log-normal Distribution.

The random variable T has a log-normal distribution, if its logarithm follows a normal distribution. As shown in Fig. 2.2 the distribution is skewed to the right. It has two parameters: α specifying its scale and β specifying its shape. A location parameter can be introduced by substituting $t-\gamma$ for t in the density function. Then the range of t is $t \geq \gamma$.

The log-normal distribution is used as a model of failure and repair. The justification for using the log-normal distribution comes from another central limit theorem, which states that the product of n independent random variables is a log-normally distributed random variable for large n

3. The Gamma Distribution

$$f(t) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} t^{r-1} e^{-\lambda t} & , \quad t \geq 0, \quad r > 0, \quad \lambda > 0 \\ 0, \quad \text{otherwise} \end{cases}$$

$$F(t) = \frac{\lambda^r}{\Gamma(r)} \int_0^t \tau^{r-1} e^{-\lambda \tau} \, d\tau$$

$$m = \frac{r}{\lambda} \quad , \quad \sigma^2 = \frac{r}{\lambda^2}$$

and the gamma function is defined as

$$\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} \, dx$$

which, for r a positive integer, reduces to

$$\Gamma(r) = (r-1)!$$

The gamma distribution is obeyed by random variables which are defined on half the real axis. It has two parameters, λ and r, and it can take many shapes for various values of the parameters (Fig. 2.3). For $r \leq 1$ it is concave upwards while for $r > 1$ it is concave downwards with a maximum at $t = \frac{r-1}{\lambda}$.

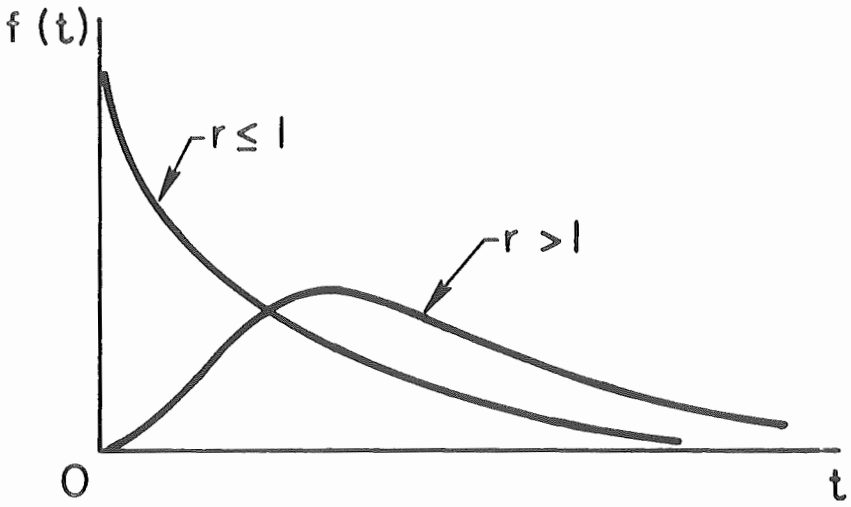


Figure 2.3. Shapes of the Gamma Distribution for the Two Different Ranges of r.

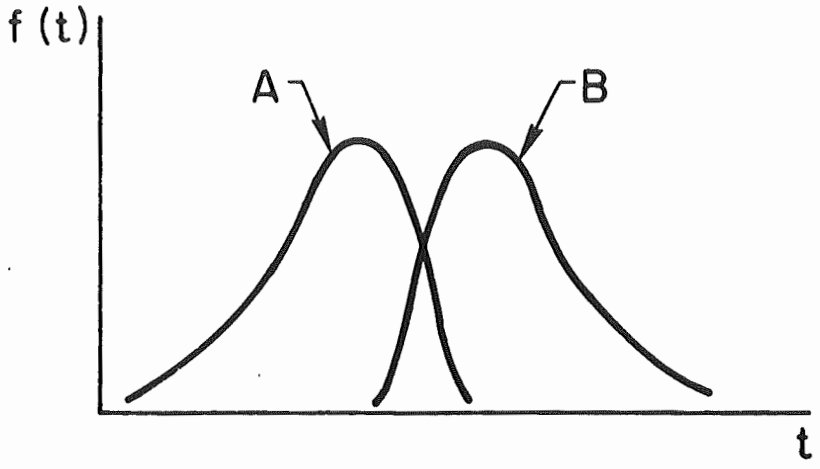


Figure 2.4. Extreme Value Distributions
A. Type I Asymptotic Distribution of Minimum Values.
B. Type I Asymptotic Distribution of Maximum Values.

The physical situations where the gamma distribution is used are when events are occurring at a constant rate λ and we are interested in the distribution of the time it takes for r events to occur (this interpretation implies that r is a positive integer, while the given definition does not require this restriction. A non-integer r could result from experimental data, although usually it is rounded off. When r is a positive integer the gamma distribution is also called Erlangian).

As the number r of events increases the distribution becomes more and more symmetrical and for large r it can be approximated by a normal distribution with the same mean and variance.

Tables of the distribution function F(t) (incomplete gamma function) can be found in Ref. 13.

4. The Chi-square Distribution.

$$f(t) = \begin{cases} \dfrac{1}{\Gamma\left(\frac{n}{2}\right) 2^{n/2}} t^{(n/2)-1} e^{-t/2} & t \geq 0 \quad , \quad n \text{ positive integer} \\ 0 \ , \quad \text{otherwise} \end{cases}$$

$$m = n \quad , \quad \sigma^2 = 2n$$

The chi-square is a special case of the gamma distribution for $\lambda = \frac{1}{2}$ and $r = \frac{n}{2}$, where n is a positive integer. The parameter n is quite arbitrarily called "degrees of freedom" and a usual notation is $\chi^2(n)$ meaning that the random variable has a chi-square distribution with n degrees of freedom.

Tables with values of the chi-square distribution can be found in many textbooks on statistics and reliability.[4,9,12,13,14] As an application we consider the goodness-of-fit problem, that is, we have experimental data and we wish to determine whether these data can be assumed to come from a theoretical distribution. The data come as observed frequencies of events and for the same events the assumed distribution predicts different, in general,

frequencies. A goodness-of-fit test determines whether these differences are due to chance or our assumption is wrong. One such test is as follows: from the theoretical frequency f_t^k and the observed frequency f_o^k of the k^{th} event (or category) we find the value

$$\chi^2 = \sum_{k=1}^{N} \frac{\left(f_t^k - f_o^k\right)^2}{f_t^k}$$

where N is the number of events. The degrees of freedom will be N-1, if no parameter used in the test is calculated from the test data (we use N-1 because the assumption of a distribution, from which the f_t^k are found, results to a loss of one degree of freedom). Having the χ^2 value and the degrees of freedom we find from tables what is the probability of such a value being due to chance and accordingly we accept or reject our assumption about the theoretical distribution. For example, suppose we toss a coin 100 times and we observe 42 times "heads" and 58 times "tails". We wish to check whether it is reasonable to assume that the coin is ideal, that is, whether there is a theoretical probability of 1/2 for heads and tails. Then we have $f_0^t = 58$, $f_0^h = 42$, $f_t^t = 50$, $f_t^h = 50$, $N = 2$ and

$$\chi^2 = \frac{(58-50)^2}{50} + \frac{(42-50)^2}{50} = 2.56.$$

The degrees of freedom are $n = N - 1 = 1$. From tables[9] we see that, the probability that a value of at least $\chi^2 = 2.56$ with $n = 1$ is due to chance, is only about 0.12. On this basis we would probably decide that the coin was not ideal.

5. The (Negative) Exponential Distribution

$$f(t) = \lambda e^{-\lambda t} \qquad \lambda > 0, \qquad t \geq 0$$

$$F(t) = 1-e^{-\lambda t}$$

$$m = \frac{1}{\lambda} \quad , \quad \sigma^2 = \frac{1}{\lambda^2}$$

This is also a special case of the gamma distribution for r = 1, that is, it is the distribution of the time for one event to occur, when the events take place at a constant rate λ. Changing the words a little we can also interpret it as the distribution of the time between successive events, when the events occur at a constant rate λ.

6. The Beta Distribution

$$f(t) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha-1} (1-t)^{\beta-1} & , \quad \alpha > 0, \ \beta > 0, \ 0 \le t \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

$$F(t) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^t \tau^{\alpha-1}(1-\tau)^{\beta-1} \, d\tau \qquad 0 \le t \le 1$$

$$m = \frac{\alpha}{\alpha+\beta} \quad , \quad \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

The beta distribution can be used when the random variable takes values in an internal. The formulas given above are for the interval [0,1]; if the random variable is limited in $[t_0,t_1]$, we map this interval on [0,1] by the transformation $t' = \frac{t-t_0}{t_1-t_0}$ and we use the above formulas.

The distribution function F(t) (incomplete beta function) has been tabulated in Ref. 15.

7. Extreme Value Distributions

In this category of distributions the random variable must be carefully specified.

Consider a random variable X, its density $\phi(x)$ and distribution function $\Phi(x)$. We select a sample of n values $x_1, x_2, \ldots, x_n$ from the domain of X and on this sample we identify the maximum (or minimum) value x_{max} (x_{min}). It is clear that by selecting another sample of n values the maximum (minimum) element in it will, in general, be different from that of the first sample.

This leads us to define a new random variable X_{max} (X_{min}), which is the largest (smallest) value of X in a sample of size n from an initial distribution $\Phi(x)$. We seek the distribution of this new random variable X_{max} (X_{min}).

As an example consider the floods from a river. The random variable X is the (average) daily discharge of the river. The sample is a year, that is, n = 365 days. In this sample we call the largest discharge a flood and it is the new random variable X_{max}. Our problem is to find the distribution of floods over the years.

Knowing the initial distribution $\Phi(x)$ it is easy to calculate the distribution of maximum values from samples of size n (Ref. 16). Recalling the probabilistic meaning of distribution functions we get

$$F(x_{max}) = \Phi^n(X_{max}) \tag{2.15}$$

and the density is

$$f(x_{max}) = \frac{dF(x_{max})}{dx_{max}} = n\Phi^{n-1}(x_{max})\ \phi(x_{max}).$$

Similarly for the minimum values we get

$$F(x_{min}) = 1 - [1 - \Phi(x_{min})]^n \tag{2.16}$$

and

$$f(x_{min}) = n\ [1 - \Phi(x_{min})]^{n-1}\ \phi(x_{min})$$

However, more general results can be obtained for large sizes, in which case only general properties of the initial distribution are required and not its exact form. These asymptotic results are very useful in practice and we examine them in detail.[4,16,17]

Type I asymptotic distribution of maximum values

The requirement we impose on the initial distribution $\Phi(x)$ is that it should be of exponential type, that is, it should tend to unity for increasing

x at least as rapidly as an exponential. This is a quite general condition and several common distributions satisfy it like the gamma (and, naturally, the exponential and the chi-square), the normal and the log-normal distributions. For such initial distributions the maximum values of large samples of independent values follow the distribution (we drop the subscripts max and min and we use the variable t)

$$f(t) = \alpha \exp\left[-\alpha(t-\beta) - e^{-\alpha(t-\beta)}\right]$$

for $-\infty < t < \infty$, $-\infty < \beta < \infty$, $\alpha > 0$

and $F(t) = \exp[-e^{-\alpha(t-\beta)}]$

Tables of values of F(t) can be found in Ref. 18. The floods mentioned before follow this distribution.

Type I asymptotic distribution of minimum values

The requirement here is that the initial distribution $\Phi(x)$ should tend to zero as $x \to -\infty$ at least as fast as an exponential. The normal distribution satisfies this condition. The distribution of the minimum values of large samples is then

$$f(t) = \alpha \exp[\alpha(t-\beta) - e^{\alpha(t-\beta)}]$$

for $-\infty < t < \infty$, $-\infty < \beta < \infty$, $\alpha > 0$

and

$$F(t) = 1 - e^{-e^{\alpha(t-\beta)}}$$

This distribution is used in series systems where the rule[16] "no chain is stronger then its weakest link" applies. For example, if n elements are connected in series the system will fail when the least reliable element fails.

For these distributions we can define the mean and variance as usual. Thus the mean for maximum values is $\beta + \frac{0.577}{\alpha}$ and for minimum values $\beta - \frac{0.577}{\alpha}$, while the variance is the same for both distributions and equal to $\frac{1.645}{\alpha^2}$.

The shape of the distributions can be seen in Fig. 2.4 and it is always the same, since there is no shape parameter (α is a scaling parameter and β a location parameter).

Notice that the asymptotic results are true for large samples of independent values. This may cause some hesitation occasionally, when the values are not independent; for example the discharge of a river one day is not always independent from the discharge of the previous day. But the samples are usually so large that another sample of truly independent values can be selected which will still be large enough for the asymptotic results to apply (we can select, for instance, 150 independent daily discharges from the 365 of the year).

Another useful extreme value distribution is examined in the following section.

8. The Weibull Distribution

$$f(t) = \begin{cases} \frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1} \exp\left[-\left(\frac{t}{\beta}\right)^{\alpha}\right] & \text{for } t \geq 0,\ \alpha > 0,\ \beta > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$F(t) = 1 - e^{-(t/\beta)^{\alpha}}$$

$$m = \beta\Gamma\left(\frac{1}{\alpha} + 1\right) \quad , \quad \sigma^2 = \beta^2\left[\Gamma\left(\frac{2}{\alpha} + 1\right) - \left[\Gamma\left(\frac{1}{\alpha} + 1\right)\right]^2\right]$$

This is also called Type III asymptotic distribution of minimum values. The initial distribution should be bounded at the left, like, for example, the gamma distribution. Therefore, we use the Weibull distribution for the distribution of the minimum values of large samples of independent values, when the initial distribution is gamma, while we use the Type I distribution of minimum values, when the initial distribution is of exponential type (normal).

The distribution, as given above, has two parameters, the scaling parameter β and the shape parameter α (see also Fig. 2.5). A location parameter γ can be introduced by substituting $t - \gamma$ for t in the given formulas. For $\alpha = 1$ the Weibull distribution reduces to the exponential distribution.

9. Student's t Distribution

In 2.A.5 we discussed the interval estimation of the parameters of a distribution. Also we gave formulas for the calculation of the mean $\bar{x}$ and the (unbiased) standard deviation s of a sample.

Student's distribution is used to find confidence intervals for the mean of a normal distribution. The means $\bar{x}$ from samples of size n are normally distributed (central limit theorem) with mean the population mean m and standard deviation $\frac{\sigma}{\sqrt{n}}$. Very often though σ is not known but estimated from the sample (s). Then the standard deviation of the means $\bar{x}$ is $\frac{s}{\sqrt{n}}$ but it is no longer accurate to assume their distribution to be normal. However, it can be proved that the variable

$$t = \frac{\bar{x} - m}{\frac{s}{\sqrt{n}}}$$

follows Student's t-distribution. This can be used to yield bounds for m and confidence levels, for example, how certain we are that the true population mean m lies in the interval

$$\bar{x} \pm t_{\alpha,r} \frac{s}{\sqrt{n}}$$

The value of $t_{\alpha,r}$ can be found from tables of the t distribution[4,9,12] once the confidence level $1-\alpha$ and the degrees of freedom r are given. The value of r is n-1. Some examples will clarify the procedure.

Suppose we calculate $\bar{x}$ and s from a sample of size n = 15. Then r = 14 and for a confidence level of $0.95 = 1 - \alpha$, that is $\alpha = 0.05$, we find from

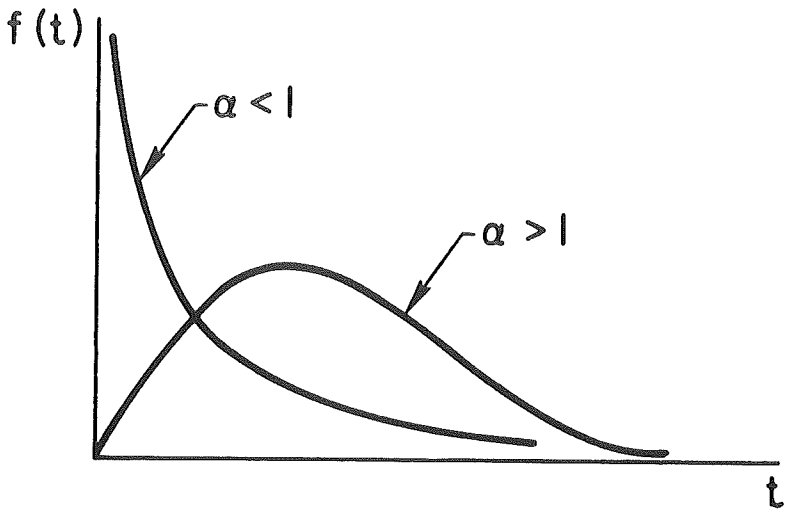


Figure 2.5. Shapes of the Weibull Distribution for the Two Different Ranges of α.

tables that $t_{0.05,14} = 2.145$. Then we say that if we take many samples and we calculate their mean, then 95% of the time the mean will fall in the interval.

$$\bar{x} \pm 2.145 \frac{s}{\sqrt{15}}$$

Instead of using so many words we simply say that the population mean lies in the above interval at a confidence level of 95%.

When we specify an interval for the mean we call the confidence level a two-sided confidence level. One-sided confidence levels can be found from similar table of $t_{\alpha,r}$ values and they will be of the form

$$\bar{x} - t_{\alpha,r} \frac{s}{\sqrt{n}}, \quad \text{lower bound}$$

and $$\bar{x} + t_{\alpha,r} \frac{s}{\sqrt{n}}, \quad \text{upper bound.}$$

In the above example we find that the value of $t_{0.05,14}$ for one-sided level is 1.761 and we claim that 95% of the time the sample mean will be greater than $\bar{x} - 1.761 \frac{s}{\sqrt{15}}$ or smaller than $\bar{x} + 1.761 \frac{s}{\sqrt{15}}$.

The t-distribution is also used to test hypotheses related to means as we will see in the following section.

2.A.7 Tests of Hypotheses

We have already encountered the two principal areas of statistical inference. In the point and interval estimation of parameters we described methods for estimating the parameters of distributions from information contained in samples. In the discussion of chi-square distribution the problem was of a different kind; we assumed a theoretical distribution and we performed a test to check whether it was reasonable to accept that the sample data came from that distribution. This problem falls in the category of statistical hypotheses.

Another example of a hypothesis could involve assumptions about the parameters of a known distribution. For example, a change in the production process of a certain equipment is expected to increase the mean m of a certain characteristic while the variance remains unaltered (assume normal distribution). Then we form two hypotheses:

H_0: $m \leq m_0$, the so-called null hypothesis expressing the possibility of no improvement, and

H_1: $m > m_0$, the alternative hypothesis.

Of course we will need a test which will enable us to accept or reject H_0. Thus we have the two definitions:[12] An assertion about the distribution of one or more random variables is called a statistical hypothesis. The acceptance or rejection of a hypothesis from information contained in a sample is based on certain rules which are called a statistical test.

From the above it is clear that there is a possibility that we may reject H_0 while actually it is true; this is called a type 1 error or the producer's risk. The reason for the last name is that when we test a lot of equipments for their quality, the null hypothesis H_0 is that the lot is "good", so by rejecting it the producer loses.

The other kind of error we can make is to accept H_0 while actually it is not true; this is the type 2 error or the consumer's risk, the last name being justified by arguments similar to those in the previous paragraph.

It is customary to give percentages for these errors. We say that the producer's risk is 100α per cent and the consumer's risk is 100β per cent and we mean that 100α per cent of the time we will reject H_0, while it is true, and 100β per cent of the time we will accept H_0, while it is wrong. By taking large samples we can be more confident about our judgement and thus reduce α and β.

A few examples will give a better picture of these concepts. For more complete discussions of the mathematical aspects of statistical hypotheses the reader is referred to Ref. 12; useful applications can be found in References 6 and 9.

Suppose we have a lot of N equipments of which S are "good". The number S is unknown and we take a sample of n equipments to decide whether we will accept the lot or not. In Ref. 6 (Tables 13.2 through 13.4) the hypergeometric distribution is used to produce tables which, for a lot of size N, sample size n and the hypothesis that the lot will be accepted if c or less equipments are found defective in the sample, give the percentages in the lots which will be accepted $(1-\alpha)100$ per cent of the time and 100β per cent of the time. The producer's risk is set at $\alpha = 0.05$ and the consumer's risk at $\beta = 0.10$. Thus for a lot of $N = 60$ equipments and a sample size $n = 10$ we decide to accept the lot if the number of defectives in the sample is less than or equal to $c = 1$. Table 13.3, then says that the producer has the risk to have the lot rejected 5 per cent of the time when it has 4.2% defectives. On the other hand the consumer has the risk of accepting the lot 10 per cent of the time with 32% defective items in it.

Another example involves the mean of a normal distribution. At a certain time we have reason to believe that the mean m has changed from its known value m_0, while the variance is the same but unknown. We take a sample of size n and we calculate the mean $\bar{x} = \frac{\Sigma x_i}{n}$ and the unbiased variance $s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$.

The null hypothesis is:

H_0: $m = m_0$ (the mean has not changed)

and the alternative hypothesis is:

H_1: $m \neq m$. (the mean has changed).

We seek a test to decide whether to accept H_o and we are willing to be in

error rejecting it (while it is correct) 5 per cent of the time (that is $\alpha = 0.05$). We do not specify the type 2 error. We know that the variable

$$t = \frac{\bar{x} - m_0}{\frac{s}{\sqrt{n}}}$$

follows Student's t-distribution with n-1 degrees of freedom. From tables we find the value $t_{\alpha,n-1}$ (two-sided) and this represents the maximum allowable value of (t) which can be due to chance at a confidence level of $(1-\alpha)100$ per cent. Thus the hypothesis H_0 is rejected if $(t) > t_{\alpha,n-1}$.

Other types of hypotheses and their handling will be found in later chapters.

2.B. THE FAILURE OF COMPONENTS

2.B.1 Introduction

It is a well-known fact that all devices or systems undergo failures of some kind due to various reasons, like manufacturing defects, very high stresses, unfavorable environmental conditions, degradation of strength due to aging, etc.

Of great importance in reliability and safety studies is the time-to-failure of a unit, that is the time it takes for a unit which is as good as new to fail (for repairable items the time between two successive failures is of interest, but we do not consider repair here). To predict the exact time of failure is a rather impossible task considering the many causes that can lead to it and the vast amount of information we would need. However a mathematical theory of failure can be developed with models which approximate the real situation. Of course, it comes as no surprise that such models are probabilistic in nature and rely heavily on Probability Theory and Statistics.

The above comments are very vague for quantitative analysis. To be precise we proceed to several definitions. When we talk about "units", or "elements", or "devices", or "components", we agree to view them as single entities and we completely ignore the fact that perhaps a particular unit consists of other parts. Every such component has been manufactured to perform a certain function; if, for any reason, the unit is not able to perform this specific function under its prespecified operational conditions, we will say that a failure has occurred.

To express the probability of failure as a function of time extensive use of distribution functions will be made. (see 2.A.3) The distribution $F(t)$ gives the probability that the unit will fail before t, while the density function $f(t) = \frac{dF(t)}{dt}$ is helpful to define the probability that the device will

fail between t and $t + \Delta t$, which is $f(t)\Delta t$. The quantity $R(t) = 1-F(t)$ is called the reliability of the item and it is the probability that the item will fail after time t.

Given a distribution function F(t) we define the hazard function (or hazard rate, or failure rate) as

$$h(t) = \frac{f(t)}{1-F(t)}$$

and we interpret it as follows: if the unit survives up to t, then the (conditional) probability that it will fail in $(t, t + \Delta t)$ is $h(t)\Delta t$.

The failure rate is very useful in helping to understand the physical phenomenon which the distribution function describes. A typical graph is shown in Fig. 2.6 (bathtub curve). In the burn-in period the hazard rate decreases due to the early failures of the units with manufacturing defects. This period is not of interest, because special "debugging" procedures are usually used to eliminate these defective elements (which, strictly speaking, do not belong to the population, since their manufacturing process deviated grossly from the design which would guarantee capability of performing the specified function).

The next two regions represent the two important phenomena that we will attempt to model. In the first one the failure rate is constant meaning that the probability of failure is independent of the items age. Failures are due to very high stresses due to chance (e.g. an accidental current surge which causes failure of a light bulb). In the next region the probability of failure increases with the item's age (after several thousand hours of operation the probability that the bulb will fail increases with time until, eventually, the bulb burns out). The failure is now due to wear, which is a generic term for the accumulating irreversible changes which weaken the strength of the equipment.

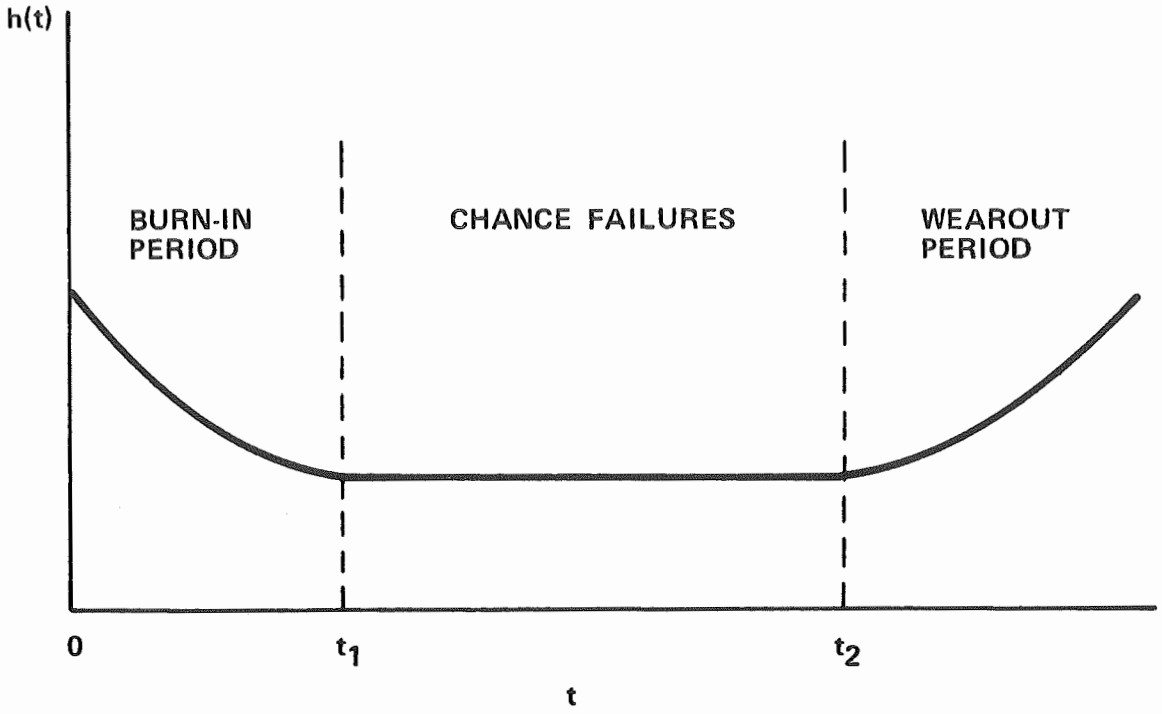


Figure 2.6. Typical Failure Rate as a Function of Time.

These two types of failure are examined in detail in the subsequent sections. Statistical distributions are presented with their mathematical properties and their applications as models of failure and other areas of safety.analysis. Then the general problem of analyzing sample data is considered with the associated statistical methods for determining the parameters of distributions.

Finally, we point out that we do not always talk about times-to-failure or times of successful operation. Time may not be the appropriate variable in cases of some units like switches or rotating devices and then we resort to such variables as cycles of operation, number of revolutions etc. Nevertheless, nothing essential changes and only slight modifications are needed in order to adjust the models to a particular situation.

2.B.2 Statistical Distributions

We present here the statistical distributions used to describe the lifetime of components, their properties and the physical situations which lead to them. A very detailed account is given in the book of Gertsbakh and Kordonskiy (Ref. 19); additional discussions can be found in many books on reliability,[4,5,6,7,8,9] and in Ref. 20 and 21. Some mathematical properties of discrete and continuous distributions were presented in Section 2.A.6 of this report.

1. The Exponential Distribution

$$f(t) = \lambda e^{-\lambda t} \quad , \quad \lambda > 0$$

$$F(t) = 1 - e^{-\lambda t} \quad , \quad R(t) = e^{-\lambda t}$$

$$h(t) = \lambda$$

$$m = \frac{1}{\lambda} \quad , \quad \sigma^2 = \frac{1}{\lambda^2}$$

The exponential distribution is the most widely used in reliability studies. What makes it distinctly different from the other distributions is the constancy of the hazard function, which qualifies it as the only distribution to describe the period of "chance failures" of an item.

Figure 2.7 helps to clarify the physical process that is modeled. The maximum stress that the item can withstand is S_{max} and it is constant with time. The actual stress applied is random in time and it is represented by the zig-zag curve. Clearly a failure occurs when the applied stress exceeds the maximum allowable stress. Such "peak" stresses (that is, greater than S_{max}) are assumed to follow the Poisson distribution.

$$p(r) = e^{-\lambda t} \frac{(\lambda t)^r}{r!} \quad , \quad \lambda > 0 \quad , \quad r = 0, 1, \ldots$$

where λ is the constant rate of occurrence of peak loads and $p(r)$ is the probability of exactly r peak loads occurring in an interval $(0,t)$. For this assumption to be valid the applied stress must have the following two properties[19]

i) asymptotic independence: the peak stresses are rare events, thus the time interval between any two of them is big enough, so that they can be considered as independent events.

ii) stationarity: this means that the stresses are homogeneously applied without a preferred direction (they do not gradually increase or decrease).

With the assumption of Poisson distributed peak stresses we can readily see that the component will not fail in the interval $(0,t)$ if no peak stresses occur in that interval, therefore its reliability is

$$R(t) = p(0) = e^{-\lambda t}$$

and its unreliability (failure distribution)

$$F(t) = 1 - R(t) = 1 - e^{-\lambda t}$$

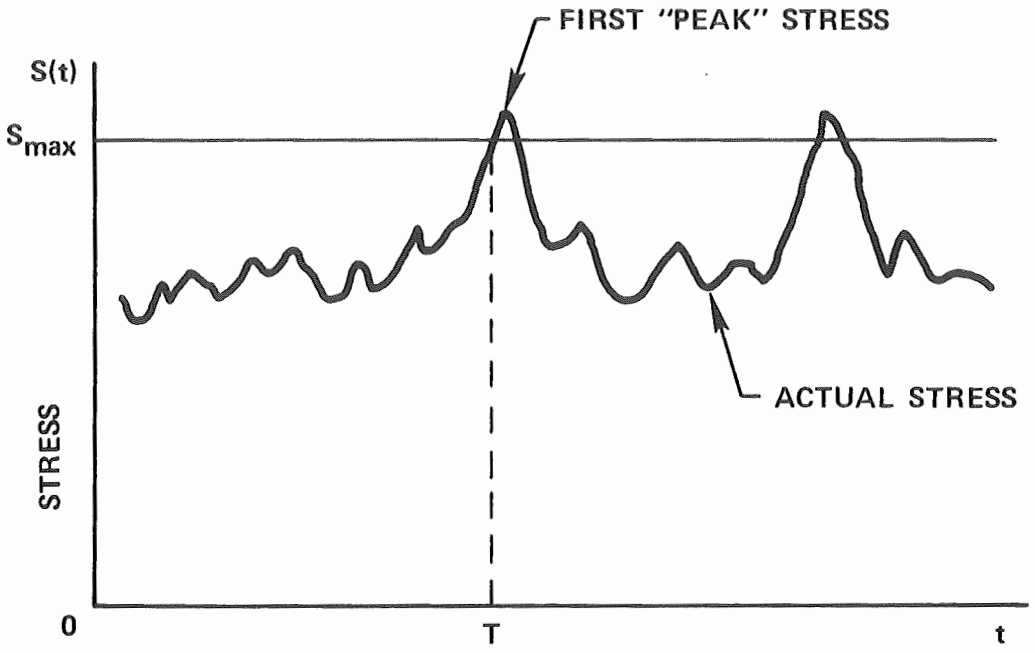


Figure 2.7. Stress as a Function of Time. Failure occurs at time T due to a "peak" stress exceeding the maximum strength of the component S_{max}.

The exponential distribution has been found particularly useful in the description of times-to-failure of electronic devices (electron tubes, etc.).

In most reliability applications the failure rate is very small (less than 10^{-4} hr^{-1}) and for not very large times t the exponential distribution can be approximated by

$$F(t) = 1 - e^{-\lambda t} \cong \lambda t.$$

It is interesting to see to what results the previous line of thought leads, when time is not the appropriate variable (e.g. we may be interested in the number of landings it will take for an airplane until a rough one occurs which leads to failure of some of its equipments,[19] or the number of startings of a Diesel engine until it fails, etc.)

In this case λ is not the mean rate of occurrence of peak stresses but the probability of the unfavorable event (rough landing, engine doesnot start). In lieu of the Poisson distribution we use the geometric distribution

$$p(r) = (1-\lambda)^{r-1}\lambda \quad , \quad 0 \leq \lambda \leq 1, \quad r = 1, 2, \ldots$$

which gives the probability of (r-1) favorable events before the unfavorable event occurs. The cumulative distribution function F(k) gives the probability that a failure occurs in the 1^{st} or 2^{nd} or ... k^{th} event

$$F(k) = \sum_{r-1}^{k} p(r) = 1 - (1-\lambda)^{k}$$

and the probability that in at most k events none will be unfavorable (i.e. the reliability) is

$$R(k) = (1-\lambda)^{k}$$

This is the probability that, for instance, either the (k+1)th, or the (k+2)th... landing is rough, or in other words, that at least k landings are "good". For λ small and k big we can approximate

$$R(k) = (1-\lambda)^k \cong e^{-k\lambda}$$

and the exponential distribution reappears, but with different units for λ confirming the comments made in the introduction.

2. The Gamma Distribution

$$f(t) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} t^{r-1} e^{-\lambda t} & , \quad t \geq 0,\ r > 0,\ \lambda > 0 \\ 0 & , \text{ otherwise} \end{cases}$$

$$F(t) = \frac{\lambda^r}{\Gamma(r)} \int_0^t \tau^{r-1} e^{-\lambda\tau}\, d\tau$$

or

$$F(t) = 1 - \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad \text{for} \quad r = 1, 2, \ldots$$

$$h(t) = \left[\int_0^\infty \left(1 + \frac{\tau}{t}\right)^{r-1} e^{-\lambda\tau}\, d\tau \right]^{-1}$$

or

$$h(t) = \frac{\lambda^r\, t^{r-1}}{(r-1)! \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!}} \quad , \quad \text{for } r = 1, 2, \ldots$$

$$m = \frac{r}{\lambda} \quad , \quad \sigma^2 = \frac{r}{\lambda^2}$$

The gamma distribution appears to be particularly suited for the study of failures, since the random variable is restricted on the positive real axis. The exponential distribution is a special case of the gamma for r = 1. The physical interpretation of the distribution is a natural extension of that for the exponential, i.e. the peak stresses are again Poisson distributed but now it takes r shocks for the failure to occur (this interpretation implies that r is a positive integer, while the given definition does not require this restriction. A non-integer r could result from experimental data, although usually it is rounded off. When r is a positive integer the distribution is also called Erlangian).

It follows from the above that the time-to-failure depends on how many shocks the device has suffered, that is, it depends on its age. Therefore the distribution is used in the wearout period of the components. The hazard function increases with time for $r > 1$ and it approaches λ for large values of t (Fig. 2.8). The density function is concave upwards for $r \leq 1$ and concave downwards for $r > 1$ with a maximum at $t = \frac{r-1}{\lambda}$ (Fig. 2.3).

The gamma distribution has two parameters: the scale parameter λ and the shape parameter r. A third parameter γ can be introduced by replacing t with $t-\gamma$ and the distribution holds for $t > \gamma$. This is called threshold of sensitivity or guarantee time, since, before γ, no damage occurs according to the model. Such a parameter can be introduced in all statistical models of failure which are bounded on the left.

3. The Normal Distribution

$$f(t) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(t-m)^2}{2\sigma^2}\right]$$

$$-\infty < t < \infty \quad , \quad -\infty < m < \infty \quad , \quad 0 < \sigma < \infty$$

$$F(t) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{t} \exp\left[-\frac{(\tau-m)^2}{2\sigma^2}\right] d\tau = N\left(\frac{t-m}{\sigma}\right)$$

$$\text{where} \quad N(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(-\frac{\tau^2}{2}\right) d\tau$$

The normal distribution is used as an approximation to the gamma for large r (approximately for $r > 12$). The physical assumptions remain the same.

At first it seems peculiar that a distribution the random variable of which is allowed to be negative may be useful in life studies. However, the normal distribution here approximates the gamma and, as such, the probability of negative times is negligibly small, that is, only the left tail of the

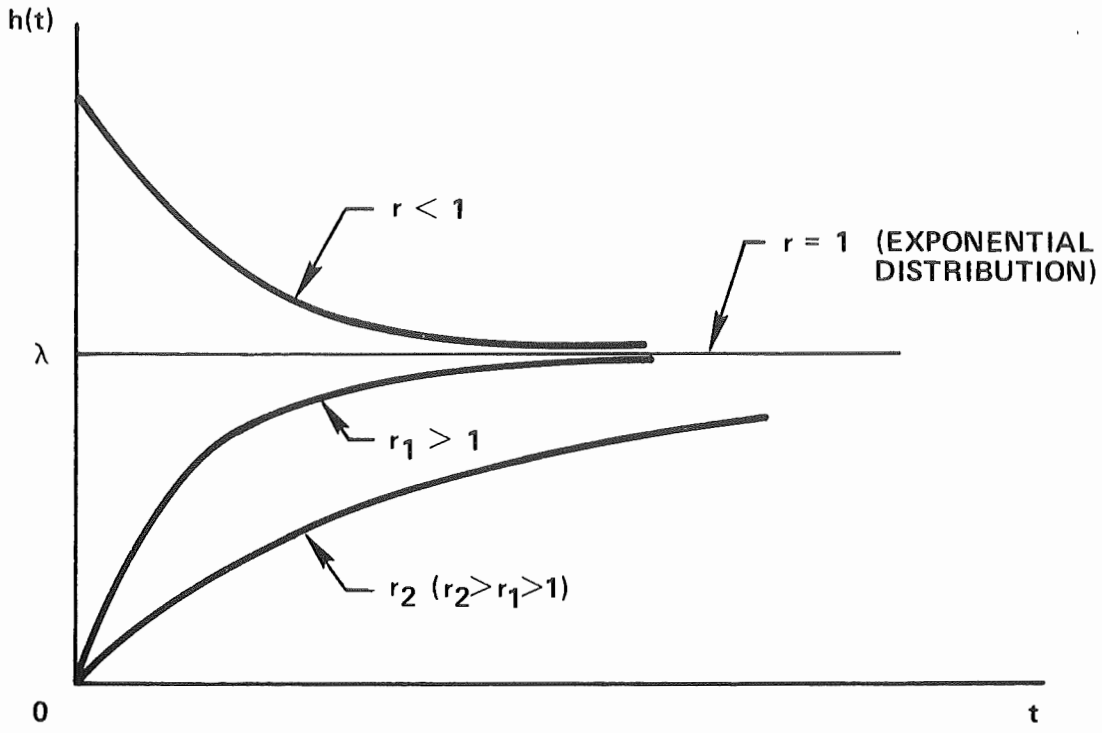


Figure 2.8. Hazard Function for the Gamma Distribution.

distribution covers the negative axis. Of course, the normal distribution can also be used for any random variable, which is result of many independent causes (central limit theorem).

The mean and variance of the distribution are $m = \frac{r}{\lambda}$ and $\sigma^2 = \frac{r}{\lambda^2}$. The hazard function is shown in Fig. 2.9; it is an increasing function of time and unbounded to the right.

4. The Log-Normal Distribution

$$f(t) = \frac{1}{\sqrt{2\pi}\,\beta t} \exp\left[-\frac{(\ln t-\alpha)^2}{2\beta^2}\right]$$

$$-\infty < \alpha < \infty \quad , \quad \beta > 0 \quad , \quad t \geq 0$$

$$m = e^{\alpha+\beta^2/2} \quad , \quad \sigma^2 = e^{2\alpha+\beta^2}\,(e^{\beta^2}-1)$$

The random variable t has a log-normal distribution, if its logarithm follows a normal distribution. The distribution is skewed to the right (Fig. 2.2) and it has two parameters: α specifying its scale and β specifying its shape. The hazard rate (Fig. 2.10) initially increases and for large times it tends to zero.

The usefulness of the log-normal distribution comes from a central limit theorem, which states that the product of n independent random variables is a log-normally distributed random variable for large n. Such a case arises in the study of failures from fatigue cracks. The random variable is the magnitude of the crack at successive times. At each time t_i the magnitude X_i is assumed to be proportional to the previous magnitude X_{i-1}, i.e.,

$$X_i = q_i\, X_{i-1}$$

where the q_i's are random variables (independent and not necessarily with the same distribution). Then, according to the above-mentioned central limit theorem, the random variable X_i is log-normally distributed for large i. In this

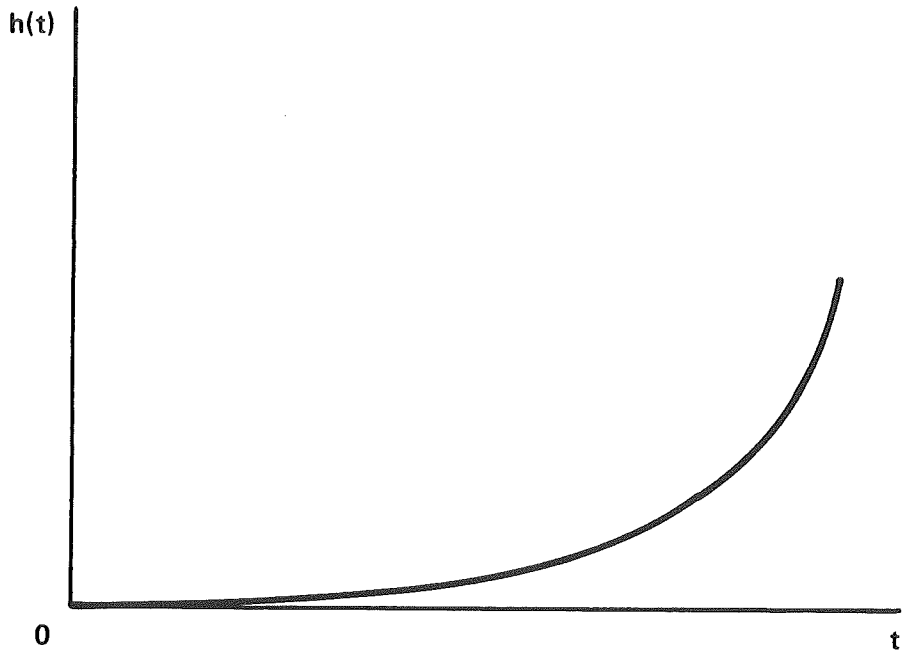


Figure 2.9. Hazard Function of a Normal Distribution.

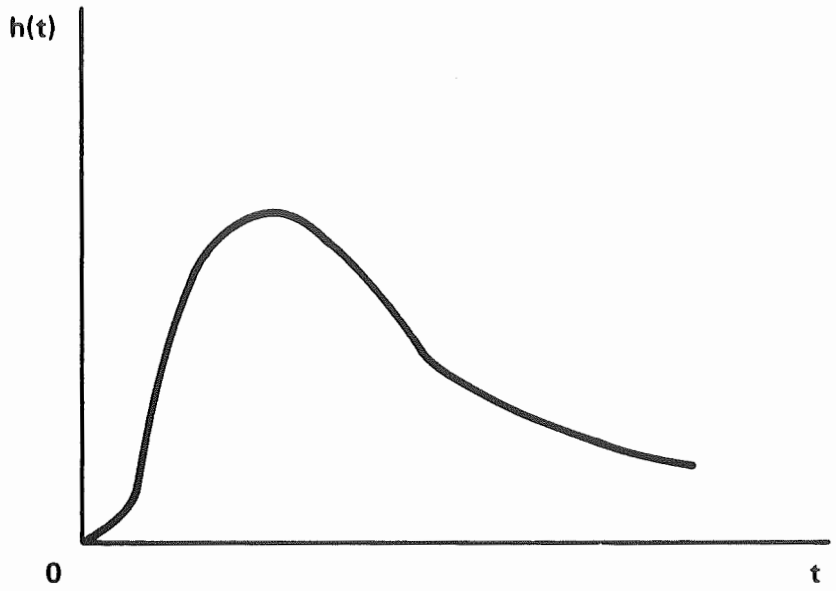


Figure 2.10. Hazard Function of a Log-normal Distribution.

respect, the lognormal distribution can be used to estimate the degree of deterioration of the object. Applications in life studies will be considered in subsequent sections.

5. The Weibull Distribution

$$f(t) = \begin{cases} \frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1} \exp\left[-\left(\frac{t}{\beta}\right)^{\alpha}\right], & t \geq 0 \quad \alpha > 0 \quad \beta > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$F(t) = 1 - e^{-(t/\beta)^{\alpha}}$$

$$h(t) = \frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1}$$

$$m = \beta\,\Gamma\left(\frac{1}{\alpha} + 1\right), \quad \sigma^2 = \beta^2\left[\Gamma\left(\frac{2}{\alpha} + 1\right) - \Gamma\left(\frac{1}{\alpha} + 1\right)\right]^2$$

The shape of the distribution can be seen in Fig. 2.5. It has two parameters: the shape parameter α and the scaling parameter β. Its hazard function is an increasing function of time for $\alpha > 1$ and a decreasing function for $\alpha < 1$ (Fig. 2.11). For $\alpha = 1$ the Weibull becomes the exponential distribution with constant failure rate.

The physical interpretation of the Weibull distribution is associated with the theory of extreme values. It is the distribution of the minimum value of large samples of independent values from a gamma distribution. Such a situation is encountered in devices which consist of many other components, the lifetime of which is given by the same gamma distribution; the device fails when any of its components fail, therefore its lifetime is a random variable which is the minimum of the lifetimes of the components. It has been found that the gamma distributions of the component lifetimes may be allowed to have slightly different parameters and still the Weibull distribution is applicable.

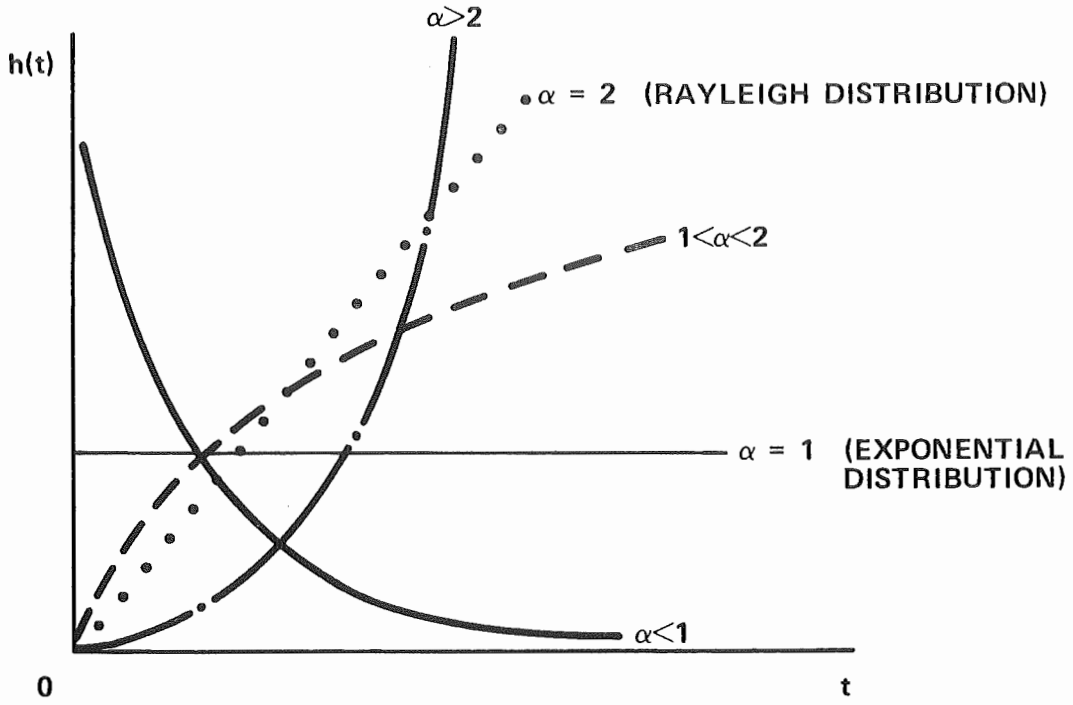


Figure 2.11. Hazard Function of the Weibull Distribution for Various Ranges of the Shape Parameter α.

The Weibull distribution is also called Type III asymptotic distribution of minimum values for obvious reasons. It has been used successfully to describe the time-to-failure of electron tubes,[22] ball-bearings,[23] et al.

6. Extreme Value Distributions

Type I Asymptotic Distribution of Maximum Values

$$f(t) = \alpha \exp\left[-\alpha(t-\beta) - e^{-\alpha(t-\beta)}\right]$$

$$-\infty < t < \infty \quad , \quad -\infty < \beta < \infty \quad , \quad \alpha > 0$$

$$F(t) = \exp\left[-e^{-\alpha(t-\beta)}\right]$$

$$h(t) = \frac{\alpha e^{-\alpha(t-\beta)}}{\exp\left[e^{-\alpha(t-\beta)}\right]-1}$$

$$m = \beta + \frac{0.577}{\alpha} \quad , \quad \sigma^2 = \frac{1.645}{\alpha^2}$$

Type I Asymptotic Distribution of Minimum Values

$$f(t) = \alpha \exp\left[\alpha(t-\beta) - e^{\alpha(t-\beta)}\right]$$

$$-\infty < t < \infty \quad , \quad -\infty < \beta < \infty \quad , \quad \alpha > 0$$

$$F(t) = 1 - e^{-e^{\alpha(t-\beta)}}$$

$$h(t) = \alpha\, e^{\alpha(t-\beta)}$$

$$m = \beta - \frac{0.577}{\alpha} \quad , \quad \sigma^2 = \frac{1.645}{\alpha^2}$$

Extreme value distributions deal with the distribution of the maximum or minimum value in large samples of independent values drawn from an initial distribution.[4,16,17]

Consider a sample of size n of values $x_1, x_2, \ldots x_n$ from a distribution function $\Phi(x)$. We define a new random variable $T = \min(x_1, x_2, \ldots, x_n)$ and we seek the distribution of T for all possible samples of size n. Knowing $\Phi(x)$ it is readily seen that

$$F(t) = P(T \leq t) = 1 - [1 - \Phi(t)]^n \qquad (2.16)$$

and the density function is

$$f(t) = \frac{dF(t)}{dt} = n\,[1 - \Phi(t)]^{n-1}\,\phi(t)\ .$$

If T is defined as the maximum value of the sample, its distribution will be

$$F(t) = \Phi^n(t) \qquad (2.15)$$

and $f(t) = n\,\Phi^{n-1}(t)\,\phi(t)$.

This approach requires knowledge of the initial distribution $\Phi(t)$ and of the sample size n. The usefulness of the asymptotic distributions lies in the fact that such detailed knowledge is not required; the samples should be large and the initial distribution should satisfy certain general requirements.

If the initial distribution $\Phi(x)$ tends to unity as $x \to \infty$ at least as fast as an exponential, the maximum values of large samples of independent values follow the Type I asymptotic distribution of maximum values. Common distribution which satisfy this requirement are the gamma (and, naturally, the exponential and chi-square), the normal and the log-normal.

The Type I asymptotic distribution of minimum values is applicable when the initial distribution is the normal (we have seen that when the initial distribution is the gamma then the minimum values of large samples of independent values follow the Weibull distribution).

One of the most well known applications of extreme value distributions is in the study of floods from a river.[16,17] The initial variate X is the average daily discharge of the river and its distribution is of exponential type (that is, $\Phi(x) \to 1$ for $x \to \infty$ at least as fast as an exponential). The sample size is one year (n = 365 days); the maximum discharge in one year is called a flood and the distribution of floods is the Type I asymptotic

distribution of maximum values. We have required that the values of the sample be independent and the discharge of the river in one day is not completely independent from the discharge of the previous day; however, another large sample of truly independent discharges can be selected (for example, we may consider only 150 independent daily discharges in lien of 365).

An interesting application of the theory of extreme values is in the study of strength of materials.[16,17,24] It has been found that the experimental strength is much smaller from the theoretical value derived from atomic considerations. This is attributed to the existence of flows in the material which initiate cracks that reduce its strength. The assumption is that there is a large number of such defects which are independent and they are randomly distributed in the material. We divide the material into a large number of volume elements and in each volume element, there is only one crack. The size of the crack is a random variable with a distribution of the exponential type (for example, exponential). The strength of the material in each elementary volume is decreased from the theoretical value by a quantity which is directly proportional to the crack size, i.e. $s_i = s_0 - cx_i$, where s_i the actual strength, s_0 the theoretical strength, c a proportionality constant and x_i the crack size. Therefore the strength of the material is a random variable; the breaking strength is the minimum of s_i and it corresponds to the maximum crack size x_i. But the maximum crack size has the Type I asymptotic distribution of maximum values and a simple change of variables reveals that the breaking strength follows the Type I asymptotic distribution of minimum values.

This example involved the distribution of the material strength; another example regarding times-to-failure concerns the failure of surfaces due to chemical corrosion.[21,25] Initially the surface has a large number of pits with random depths distributed according to the exponential distribution. Chemical

corrosion causes the depth of each pit to increase until failure occurs due to penetration of the surface. Assuming that the time of penetration is given by $t_i = c(H-h_i)$, where c is a constant, H the surface thickness and h_i the initial depth of the i^{th} pit, it is clear that the time-to-failure is equal to the minimum of $(t_1,\ldots,t_n)$, which, of course, corresponds to the maximum of $(h_1,\ldots,h_n)$. Just as before we find that the maximum initial depth follows the asymptotic distribution of maximum values and, by changing variables, the time-to-failure obeys the asymptotic distribution of minimum values.

Fig. 2.6 shows the asymptotic distributions and Fig. 2.12 shows the hazard functions of the extreme value distributions.

The extreme value distributions as presented above give the probability of the sizes of the maximum (minimum) value of large samples. Thus the asymptotic distribution of maximum values.

$$F(x) = \exp\left[-e^{-\alpha(x-\beta)}\right]$$

when applied to the study of floods enables one to state that the probability that the flood in any year is less than or equal to x is F(x).

However there are two other questions which remain unanswered, namely

1. How often does a flood of a certain size or greater occur?
2. What is the distribution of the floods in a period of m years?

To answer the first question the notion of the return period is introduced. For a given distribution function F(x) (not necessarily of the extreme value type) the quantity 1-F(x) is the probability that the random variable will take on a value at least x. Then the quantity

$$T(x) = \frac{1}{1-F(x)} \tag{2.17}$$

is called the return period and it is the average number of observations in which the random variable exceeds x once. For example, in the familiar

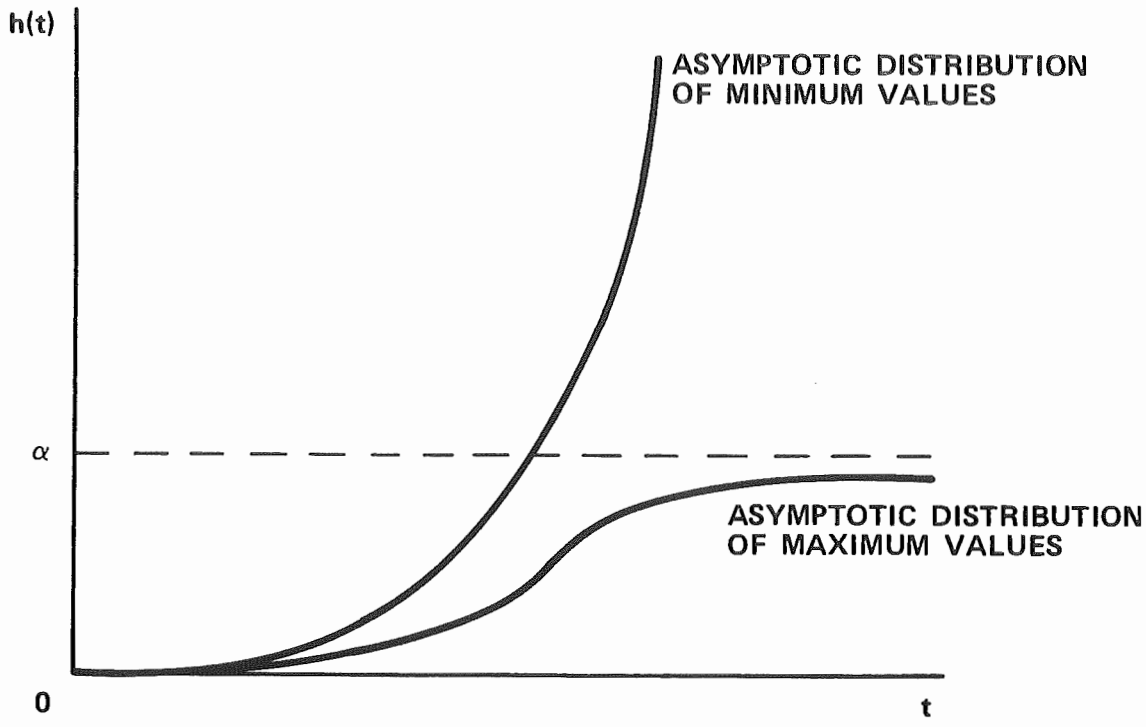


Figure 2.12. Hazard Functions of the Extreme Value Distributions.

experiment with an ideal die the probability of the outcome exceeding 4 is $1-F(4) = \frac{1}{3}$. Then $T(4) = 3$ and on the average the number showing on the die will be greater than 4 once every three trials, as it is intuitively clear.

When F(x) is the extreme value distribution and the maximum value of a number of observations in a year is the random variable, then the return period is the average number of years in which one observation of size greater than x will be made.

Restricting our attention upon the distribution of maximum values which is most useful in applications (floods, earthquakes), it can be shown that asymptotically the return period converges to

$$T(x) = e^{\alpha(x-\beta)} \tag{2.18}$$

As an example, consider the earthquakes occurring in a region.[26,38,39,40] The maximum annual magnitude has been found in many cases to be distributed according to the distribution for maximum values. In the notation common among earthquake engineers, it is written as

$$F(x) = e^{-\alpha e^{-\beta x}} \tag{2.19}$$

thus the most probable value is $\frac{\ln\alpha}{\beta}$ and in our notation we can rewrite

$$F(x) = e^{-e^{-\beta\left(x - \frac{\ln\alpha}{\beta}\right)}}$$

The return period is given by Eq. (2.17) and for large magnitudes by (using Eq. (2.18))

$$T(x) = \frac{1}{\alpha} e^{\beta x} \text{ (years)} \tag{2.20}$$

which means that it takes an average of T(x) years to observe an annual largest earthquake of magnitude at least x.

In a number N of annual largest earthquakes the number of the ones with magnitude at least x is

$$N(x) = N[1-F(x)] = \frac{N}{T(x)} \tag{2.21}$$

or, for large x, using Eq. (2.20),

$$N(x) = N\alpha\, e^{-\beta x} \tag{2.22}$$

Taking the logarithms of (2.22) we get

$$\ln N(x) = \ln(N\alpha) - \beta x \tag{2.23}$$

This equation has the same form as Richter's equation[41]

$$\log N(x) = a - bx \tag{2.24}$$

but the interpretation of the terms is different. In Eq. (2.23) N(x) is the number of annual largest earthquakes, while in (2.24) N(x) is the number of earthquakes with magnitude at least x which occurred in a given time interval. Details of the derivation of Eq. (2.24) may be found in the listed references.

Consider now the second question: the distribution of the maximum elements in a period of m years (for a nuclear reactor the distribution of floods and of large earthquakes in its lifetime is important; here m = 40 years)

This problem is again of extreme value type; we wish to find the distribution of the largest element of samples of size m, where the initial distribution is the asymptotic distribution of maximum values F(x). Using the method presented in the beginning of this section (Eq. (2.15)) we find

$$F_m(y) = F^m(y) = \exp\left(-me^{-\alpha(y-\beta)}\right) =$$

$$= \exp\left(-e^{-\alpha\left(y-\beta-\frac{\ln m}{\alpha}\right)}\right)$$

which is again the asymptotic distribution of maximum values

$$F_m(y) = \exp\left(-e^{-\alpha(y-\beta')}\right)$$

with $\beta' = \beta + \frac{\ln m}{\alpha}$

Therefore, in a period of m years the maximum flood will have a mean $\beta + \frac{0.577 + \ln m}{\alpha}$ and the most probable value will be

$$\beta' = \beta + \frac{\ln m}{\alpha}$$

A confidence interval for the return period can also be given. Gumbel calculates[16] that there is a probability $\frac{2}{3} = 0.68$ that the period will be in the interval $0.32T(x)$ and $3.13T(x)$. Therefore, if the return period of an earthquake of magnitude at least, say, 8 is 100 years, there is a probability 0.68 that such a big earthquake will occur in as short a period as 32 years or as long a period as 313 years.

Finally we notice that if in the distribution of minimum values we make the transformations $t = \ln t'$ and $\beta = \ln\beta'$ we get

$F(t') = 1 - e^{-(t'/\beta')^{\alpha}}$, that is, the Weibull distribution

Hence, the natural logarithm of a random variable which is Weibull distributed follows the extreme value distribution of minimum values (we have seen the same relation between the normal and log-normal distributions). This property is used when the problem of estimation of parameters is considered.

7. Superposition of Distributions

There are many situations where the need to combine different distributions arises.[19,21,22] In general we can distinguish two cases: 1) more than one independent causes of failure are present, 2) the population under study consists of several subpopulations of different characteristics.

When the equipments are under the parallel action of n independent causes with distributions $F_1(t)$, ..., $F_n(t)$, the distribution of the lifetime is given by

$$F(t) = 1 - \prod_{i=1}^{n} (1-F_i(t)). \qquad (2.25)$$

A simple example of such a case is the period of wearout of the items. The wearout may be modeled by an appropriate distribution (e.g. gamma, normal,

Weibull). Clearly the possibility of a failure due to chance (high stress) can not be ruled out therefore a superposition of the exponential distribution and the wearout model is necessary. For example, assume that the wearout distribution is the normal with parameters m and σ, i.e.

$$F_1(t) = N\left(\frac{t-m}{\sigma}\right)$$

$$\text{where } N(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(-\frac{y^2}{2}\right) dy. \tag{2.26}$$

For the exponential we have

$$F_2(t) = 1 - e^{-\lambda t}$$

Then the superposition of the two distributions leads to

$$F(t) = 1 - e^{-\lambda t}\left[1 - N\left(\frac{t-m}{\sigma}\right)\right]$$

with density

$$f(t) = e^{-\lambda t}\left\{\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(t-m)^2}{2\sigma^2}\right] + \lambda\left[1 - N\left(\frac{t-m}{\sigma}\right)\right]\right\}$$

Fig. 2.13 shows a plot of this density.

The second category of problems where superposition is necessary concerns heterogeneous populations. The population may consist of groups of components which have different characteristics due to various reasons.

As a first example we consider two groups of items; failures are due to chance, but the one group (proportion 100p per cent of the whole population) is weaker than the other (100(1-p) per cent). This situation naturally suggests the use of two exponential distributions with different failure rates λ_1 and λ_2 ($\lambda_1 > \lambda_2$, λ_1 the failure rate of the weak group). Then the distribution function of the population is

$$F(t) = p\left(1-e^{-\lambda_1 t}\right) + (1-p)\left(1-e^{-\lambda_2 t}\right)$$

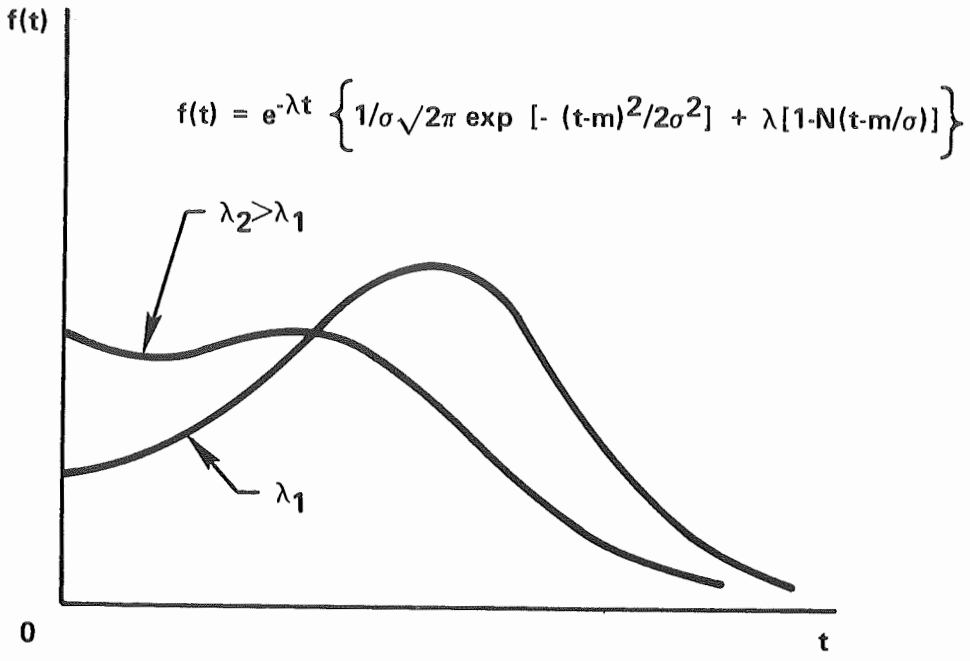


Figure 2.13. Probability Density Function of the Superposition of an Exponential and a Normal Distribution for Two Values of the Failure Rate of the Exponential.

with density

$$f(t) = p\lambda_1 e^{-\lambda_1 t} + (1-p)\,\lambda_2 e^{-\lambda_2 t}$$

In the second example the one group consists of components in the chance-failure period (exponential distribution) and the other group consists of aging items (gamma, Weibull et al). Assuming the gamma model for the aging components, we have

$$F(t) = p\left(1 - e^{-\lambda_1 t}\right) + (1-p)\left[1 - \sum_{k=0}^{r-1} \frac{(\lambda_2 t)^k}{k!} e^{-\lambda_2 t}\right]$$

with density

$$f(t) = p\lambda_1 e^{-\lambda_1 t} + (1-p)\,\frac{\lambda_2^r\, t^{r-1}}{(r-1)!}\, e^{-\lambda_2 t}$$

(r has been assumed an integer).

The density and the failure rate are shown in Figs. 2.14 and 2.15.

In the general case of n groups with failure distributions $F_1(t) \ldots F_n(t)$, the distribution for the population is

$$F(t) = \sum_{i=1}^{n} p_i\, F_i(t) \qquad (2.27)$$

where 100 p_i is the percentage of the i^{th} group in the population (and naturally $\sum_{i=1}^{n} p_i = 1$).

2.B.3 General Discussion of the Distributions

The distributions presented in the previous section are naturally idealizations of real situations. The failure of equipments is a very complicated phenomenon which can only be approximated under various assumptions by statistical distributions. Even experimental data do not always reveal the appropriate applicable model, because they are usually scattered in the region of central tendency of the distributions and, with the freedom provided by the

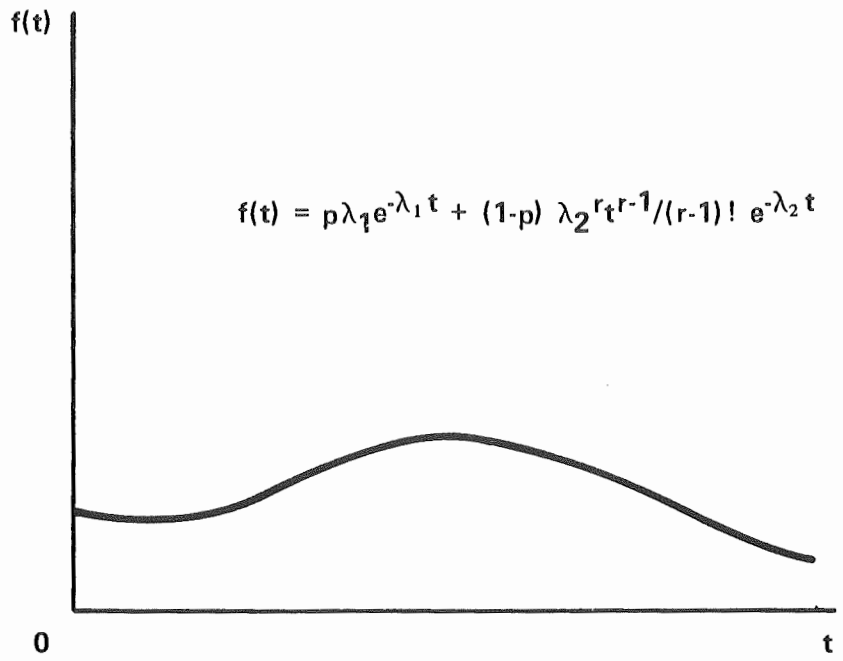


Figure 2.14. Probability Density Function of the Mixing of an Exponential With a Gamma Distribution.

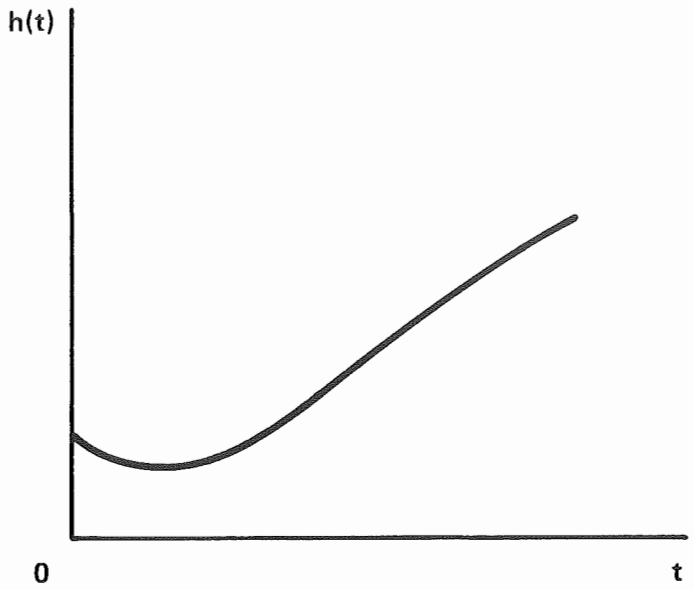


Figure 2.15. Nonmonotonic Hazard Function Resulting from Mixing an Exponential With a Gamma Distribution.

parameters of the distributions, several of them can be fitted to represent the data. In addition, the number of data points is not very large to ensure effective use of goodness-of-fit tests. Therefore, it is essential, before a distribution is selected, to understand the physics of failure and the various effects that influence it.

There are three quantities which govern the process of failure: 1) the initial strength 2) the loss of strength and 3) the limiting or reference strength. We describe briefly the nature of each of these quantities.

Initial Strength S_o

The term strength is used in a general sense. It may be the mechanical strength of an object (its resistance to tension, for example), the electrical strength (e.g. maximum voltage that a capacitor can withstand) or any quantity which determines the successful operation of the device (for example, the steepness of the characteristic - milliamperes/volt - of a lamp).

The initial strength is the value of the strength when the object is as good as new ($t = 0$). For a specific component it may be possible through some nondestructive test to assign a specific value to S_o. However, in the majority of cases the component is selected from a population of similar components, which makes S_o a random variable. The appropriate distribution is determined as follows.[27,28] A real object contains a large number of defects (flows, impurities) which cause departure from the perfect atomic structure of the material of the object. The object is visualized as consisting of a large number of volume elements each one having one defect. Depending on the size of the defect each element (link) has a certain strength. If the weak links receive support from the adjacent links (for example, strength of steel under tension) then the central limit theorem can justify the use of a normal distribution for S_o. Under special conditions (e.g. carefully controlled

manufacturing) the standard deviation of the normal distribution can become very small and it can be assumed that the initial strength is constant. When the links do not support each other the weakest link will fail first and naturally the distribution used is the Type I asymptotic distribution of minimum values or the Weibull (and its special case, the Raleigh, for $\alpha=2$; see Ref. 8 for an application).

Finally, we must allow for the possibility of an unusually large flow due to bad manufacturing (faulty weld, etc.). Then the size of the flow is the determining factor of the initial strength. The uniform distribution may be used for S_o on the basis that, since the flow is due to error, any size is equally likely.[27] If it is established that large size flows are more unlikely than small size ones an appropriate skewed distribution should be chosen (e.g. the exponential).

The Loss of Strength L

The loss of strength is a very complicated function of the environmental conditions and the applied stress. The stress, in turn, is a stochastic function of time which can exhibit various patterns of variation (stress is, again a generic term which may represent mechanical, electrical, thermal, and other stresses).

If the stress is constant in time (static stress) it may usually be well represented by a normal distribution (justified by the central limit theorem theorem [28,29]).

The more general case is when the stress is an arbitrary function of time. Then this stochastic function can be described by the two probability densities $\phi(s)$ and $\psi(t;s)$ where

$\phi(s)ds$: probability that the stress amplitude falls between s and s+ds

$\psi(t;s)dt$: probability that a stress of amplitude in the interval (s, s+ds) occurs in (t,t+dt).

A realization of the stress is shown in Fig. 2.16. Usually, the damage to the object occurs if the stress exceeds a limit s_1 (see Fig. 2.16). These peak stresses are assumed to follow a Poisson distribution with parameter depending on s_1, i.e.

$$p(r) = \frac{(\lambda(s_1)t)^r}{r!} e^{-\lambda t}$$

is the probability of r peak stresses in an internal of length t.

The effect of these loads to the loss of strength is studied by examining the behavior of the rate of change of L (Ref. 19, 30), i.e.

$$\frac{dL(t)}{dt} = v(t) \tag{2.28}$$

v(t) is again a stochastic function of time. Of course, such a detailed calculation of the stress and the rate of wear is impossible in practice. However, from general knowledge of the kind of stresses applied on the object, we can estimate the form of the time dependence of the rate of wear and this, hopefully, will eventually lead to the appropriate distribution for the life-time. To this end, we write the rate of wear as a product of two functions, i.e.

$$v(t) = \bar{v}(t)\ \rho(t) \tag{2.29}$$

where $\bar{v}(t)$ is the mean rate of wear and $\rho(t)$ a stationary function of time with constant mean and variance. Clearly, it is the mean rate of wear $\bar{v}(t)$, which determines the mean (permanent) loss of strength, i.e.

$$\bar{L}(t) = \int_0^t \bar{v}(t)dt. \tag{2.30}$$

In Fig. 2.17 the mean rate of wear is constant. The mean loss of strength is

$$\bar{L}(t) = \bar{v}t \tag{2.31}$$

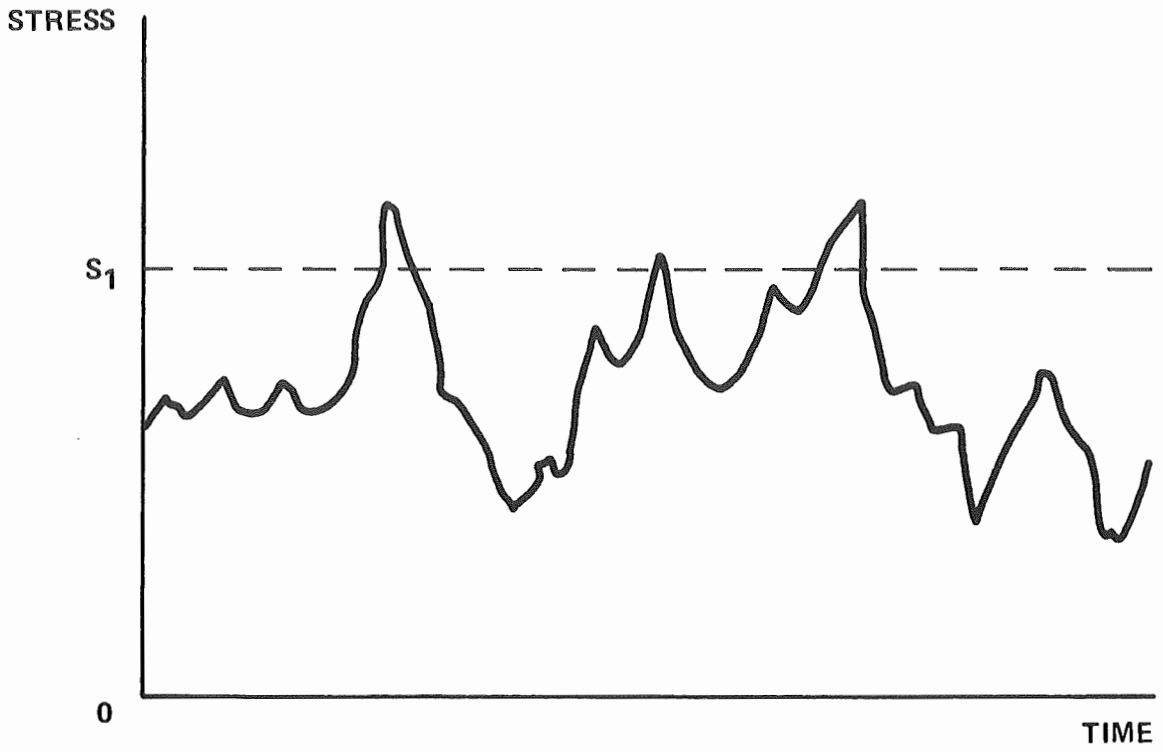


Figure 2.16. Realization of Stress as a Function of Time.

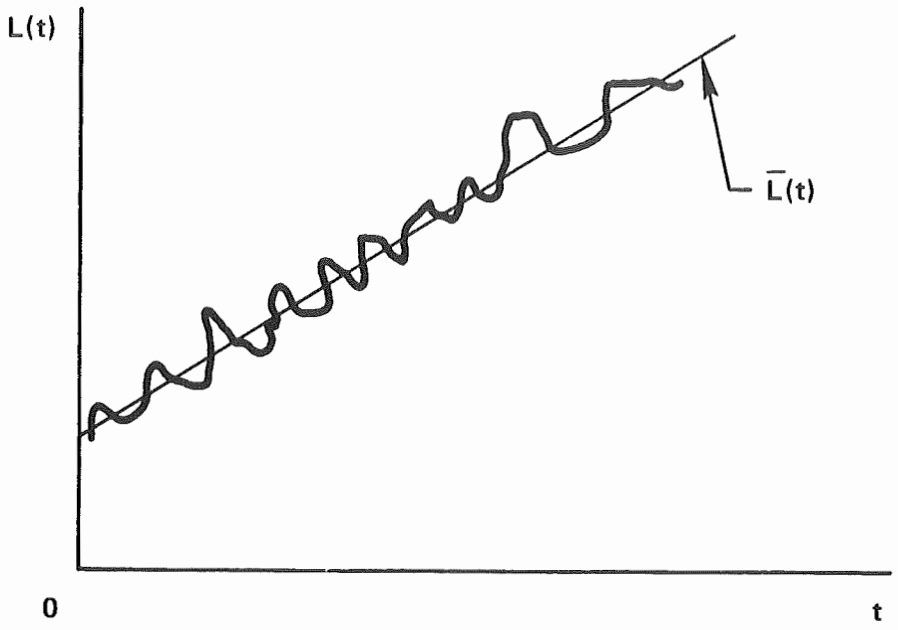


Figure 2.17. Loss of Strength When the Mean Rate of Wear is Constant.

and the actual loss of strength exhibits random variations about its mean value. This situation approximates the case where the stress is irregular and each peak stress as defined previously causes a certain amount of injury on the object (cumulative damage). The number of such injuries in a certain period of time is given by the gamma distribution.

In certain processes the mean rate of wear decreases with time (corrosion, creep of metals and general processes where the damage is caused by diffusion). Then the mean rate of wear is approximated as[19]

$$\bar{v}(t) = \frac{a}{b+t} \tag{2.32}$$

leading to a logarithmic increase of $\bar{L}$ with time.

Reference Strength S_R

The reference strength is defined as the limiting allowable value of the strength. Due to the continuous loss of strength the initial strength of the object changes continuously and when it gets beyond S_R failure occurs.

The reference strength is usually constant although it may change with time when the conditions of the use of the component change appreciably.

Having defined the three quantities governing the failure we can now state that the lifetime of the object is the solution of the equation

$$S_o - L(t) = S_R \tag{2.33}$$

Recalling the fact that S_o is a random variable and L(t) a stochastic function, it is clear that there is little hope of solving the equation exactly. However by assuming specific patterns of variation for the variables we can make useful predictions regarding the distribution of lifetime.

1. Constant Initial Strength and Constant Applied Stress

Constant initial strength can be assumed if it is possible to determine it for a specific object through some nondestructive test or if the manufacturing

process is of high quality and the variance of the distribution of the initial strength is very small.

Constant applied stress can be achieved in laboratory tests, where all the stresses can be carefully controlled. Also many electronic items are operated under essentially constant stresses. The constant applied stress can be constant in time or applied repeatedly on the object. In the first case the deterioration is continuous in time and in the second it is a function of the number of loadings (cyclic damage).

Even in this apparently simple situation the rate of wear can exhibit markedly different behavior. If the medium in which wear occurs is highly complex the loss of strength will again be a stochastic function of time. Such a case arises in the study of corrosion, aging and creep of metals,[19] where the medium is very complex and the propagation of the loss of strength from a region to another is affected by the properties of that particular region which are random in nature.

In certain cases the mean rate of loss of strength is constant and the picture of Fig. 2.17 applies (as an example we invoke the creep of metals in the region of steady state creep; in general, in the range of temperature between 0.4 and 0.7 of the melting point, metals creep with a strain rate nearly constant[31]). The stochastic nature of the rate of wear and the constancy of the mean rate of wear lead to a normal distribution for the lifetime. We recall that the normal distribution approximates the gamma distribution when the number of "shocks" is very large (here the "shocks" are not induced by peak stresses; they are equivalent to the successive accumulation of loss of strength and each shock occurs at a constant rate, since we assumed constant mean rate of loss of strength). A case of turbine blade failures due to creeping is reported in Ref. 28 and the lifetimes were found to be normally distributed.

An important case is when the mean rate of wear decreases with time (Fig. 2.18). This situation arises when strengthening of the object occurs. Then the loss of strength increases logarithmically with time. Such a phenomenon is observed when aging is due to some diffusion process, like diffusion of a metal into another, oxidation of a metal, where the oxide acts as a protective layer (Al), creeping of metals in the region of logarithmic creep, et al.

In this case the lifetimes are lognormally distributed. The justification is just as before (it takes a large number of shocks, i.e. elementary losses of strength, for failure and thus the loss of strength is normally distributed; but the loss of strength is a logarithmic function of time, therefore the lifetimes are lognormally distributed). Examples of several electronic items (resistors, transistors, diodes) where the lifetimes were found to fit the lognormal distribution under constant stresses are given in Ref. 32. Degradation was found to occur due to some diffusion process with strengthening (for thin resistance films, for example, the degrading process was oxidation of the film the rate of which was decreasing with time due to the acumulation of oxide).

The previous cases assumed that the stress was constant in time. When the stress is applied repeatedly the case of cyclic damage results. In general, we expect a lognormal distribution for the cycles to failure.[28,29,33] Examples of steel and aluminum wires whose lifetimes conformed with the lognormal distribution are given in Ref. 33. In Ref. 28 an example involving gear teeth failure under essentially constant stress is reported.

Finally, it is of interest to note that the variance of the rate of loss of strength is proportional to the square of the mean wear rate, as it can be easily shown by taking the variance of Eq. (2.29) and recalling that the

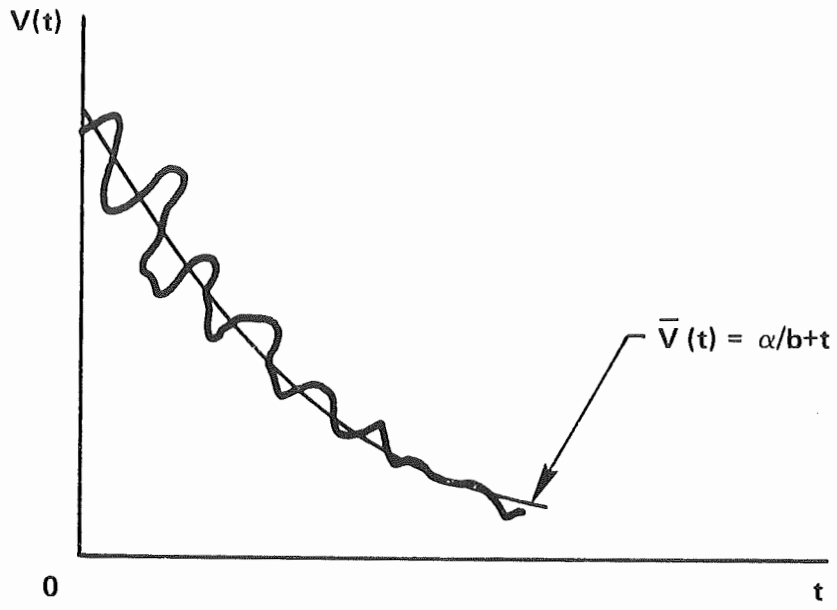


Figure 2.18. A Realization of the Rate of Loss of Strength With Decreasing Mean Rate of Wear.

variance of $\rho(t)$ is constant. The implication of this is that when the mean rate of wear is constant the same is true for the variance of $v(t)$, while for decreasing $\overline{v}(t)$ the variations of $v(t)$ about the mean value $\overline{v}(t)$ decrease (Fig. 2.18).

2. Variable Initial Strength and Constant Applied Stress

The previous discussion holds for objects of high quality the initial strength of which is practically constant. However, if the initial strength varies considerably more work is needed to determine the distribution of lifetimes.

Assume that in a lot of objects we can distinguish k groups, where the members of each group have approximately the same initial strength. Then knowing the distribution of time to failure $F_i(t)$ of each group we simply use superposition to find the distribution for the whole population, i.e. $F(t) = \sum_{i=1}^{k} p_i F_i(t)$, where p_i is the percentage of the lot belonging to the i^{th} group. Fig. 2.19 illustrates the situation when there are two groups. The number of objects in each group is the same (i.e. $p_1 = p_2 = 0.5$) and the first group has initial strength S_{01} and the second S_{02} (the reference strength is the same for both groups). The objects in each group fail according to the densities $f_1(t)$ and $f_2(t)$ (lognormal) and the failure density for the lot is $f(t) = \frac{1}{2} f_1(t) + \frac{1}{2} f_2(t)$.

If the initial strength is normally distributed we can use the previous method of superposition by dividing the normal distribution into several areas; each group will have as initial strength an average value representative of each area and the weighting factor will be the area itself. As an example, assume the distribution has a mean m and standard deviation σ; a possible division into areas is as follows:

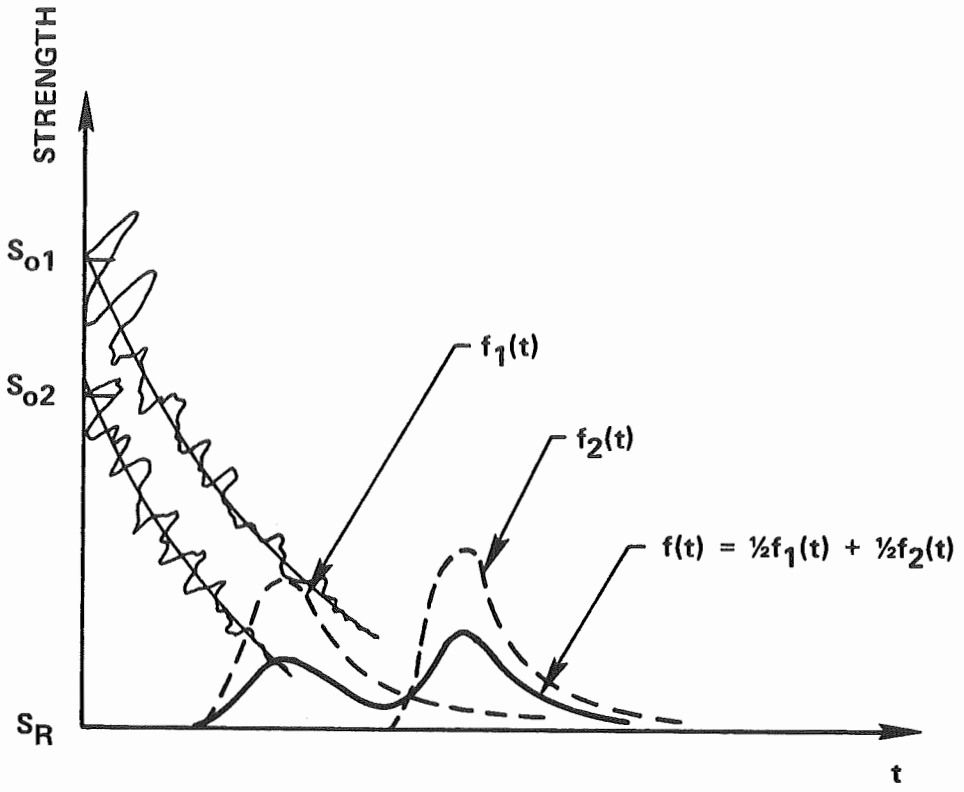


Figure 2.19. Failure Distributions for Objects With Different Initial Strength Under the Same Applied Stress.

group 1: $S_{01} = m$, $p_1 = 0.683$ (area under the curve between $m-\sigma$ and $m+\sigma$)

group 2: $S_{02} = m + \frac{3\sigma}{2}$, $p_2 = 0.135$

group 3: $S_{04} = m + \frac{5\sigma}{2}$, $p_3 = 0.021$

group 4: $S_{04} = m - \frac{3\sigma}{2}$, $p_4 = 0.135$

group 5: $S_{05} = m - \frac{5\sigma}{2}$, $p_5 = 0.021$

In the discussion of the initial strength of components it was stated that in many practical cases we can not assume that the defects present support each other but we must focus our attention to the weakest link of the object. For example, the dielectric of a capacitor contains impurities which are conductive; then the weakest link of the capacitor is the largest such impurity which causes a reduction of the breakdown voltage of the capacitor. The same situation appears when the strength of materials is primarily determined by the largest defect present, as it was mentioned in the discussion of extreme value distributions.

The initial strength is now described by the asymptotic distribution of minimum values. Assuming that the lifetime is linearly related to the initial strength, we can expect its distribution to be again the extreme value distribution of minimum values (see the section on extreme value distributions, example of failure of surfaces due to chemical corrosion in the presence of a large number of pits).

More generally, it may be assumed that the lifetimes of each link follows a gamma distribution and the lifetime of the object is described by the Weibull distribution.

Naturally in a real situation the initial strength will never be of exactly the normal or extreme-value form. As a result the lifetimes will have a distribution which will be between the lognormal and the Weibull distributions.

A plot of the data on Weibull and lognormal paper will help to choose the appropriate one.

3. Constant Initial Strength and Variable Applied Stress

The simplest case for failure under varying stress is that depicted in Figs. 2.7 and 2.16, where failure occurs instantaneously when a random peak occurs. The strength of the object is simply the maximum stress it can withstand and its failure is not due to aging. The distribution of lifetimes is naturally the exponential.

When the peaks add a single injury each time and they occur at a constant rate, the loss of strength will be again that of Fig. 2.17 and a gamma or normal distribution will describe the lifetimes, as it has already been mentioned.

If the distribution of lifetimes under constant stress is known and the applied stress is a random variable of known distribution the method of superposition may be used. In Fig. 2.20, the simple case of devices with the same initial strength is considered (i.e. resistors). At constant stress S_1 they fail according to the density $f_1(t)$. For a lower stress S_2 the obey the density $f_2(t)$ which is more spread out. If there is a probability p_1 of encountering the stress S_1 and $p_2 = 1-p_1$ of operating under the stress S_2, the lifetimes will follow the density $f(t) = p_1 f_1(t) + (1-p_1) f_2(t)$.

If the stress is normally distributed (as it can be usually assumed) we find the weighting factors as we did in the case of variable initial strength.

4. The General Case

When all the variables are random (initial strength, stress) the problem is extremely complicated. A detailed study of the particular situation is necessary in order to make simplifying assumptions which will hopefully lead to reasonable answers.

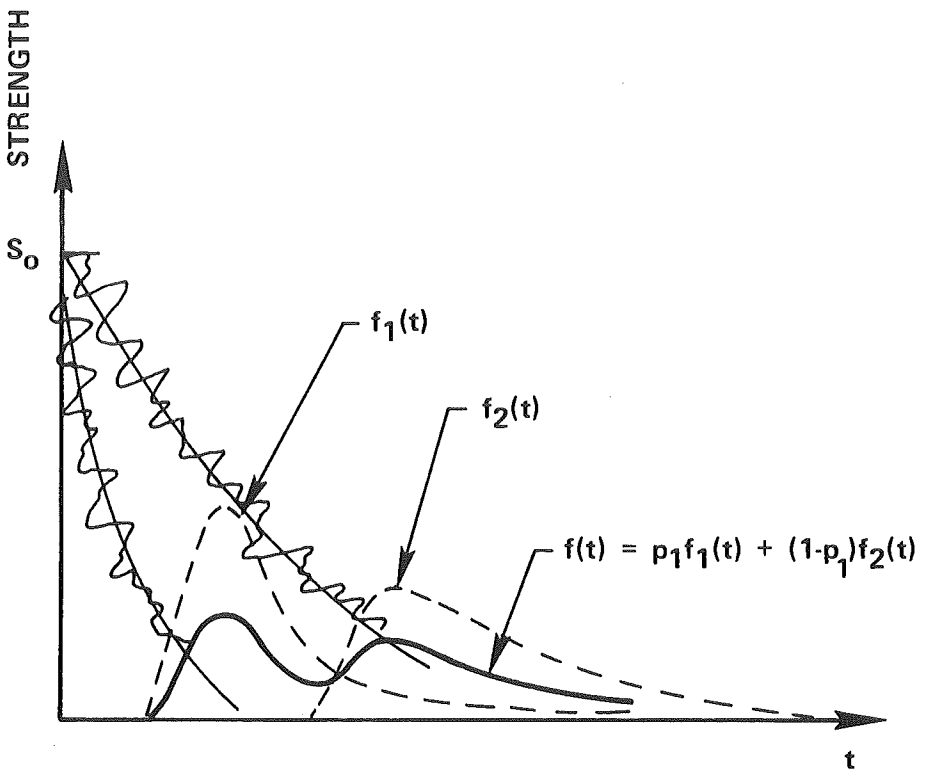


Figure 2.20. Failure Distributions for Objects With the Same Initial Strength Under Different Applied Stresses.

Discussions of various aspects of the problem can be found in Refs. 19, 27, 28, 30 and 34.

2.B.4 The Failure Rate

The distributions examined in the previous sections can be used to describe a specific behavior of the failure of devices, like failures due to chance or some aging mechanism. In some cases simple techniques can combine the distributions to describe more complex situations (e.g. superposition can combine failures due to chance and some aging law).

The general behavior of failures is very effectively studied with the use of the hazard function (failure rate). For convenience we repeat here some equations developed in Sections 2.A.3 and 2.B.1, namely

$$h(t) = \frac{f(t)}{1-F(t)} = \frac{f(t)}{R(t)} = -\frac{1}{R(t)}\frac{dR(t)}{dt} \tag{2.5}$$

Therefore we can express the quantities of interest as follows:

$$\text{reliability:} \quad R(t) = e^{-\int_o^t h(\tau)d\tau} \tag{2.7}$$

$$\text{failure density:} \quad f(t) = h(t)\, e^{-\int_o^t h(\tau)d\tau} \tag{2.34}$$

$$\text{failure distribution:} \quad F(t) = 1 - e^{-\int_o^t h(\tau)d\tau} \tag{2.6}$$

A generic hazard function is shown in Fig. 2.6. No common distribution function exists with failure rate exhibiting this behavior. However, the curve can be modeled mathematically as it will be shortly shown; then, using the assumed model for h(t) the reliability and the failure distributions can readily be found. Some possible models are the following:

i) Piecewise-linear model. The failure rate curve (Fig. 2.6) is divided into three distinct regions and a straight line approximates each region (an example is given in Ref. 8).

ii) Composite model. Each of the three regions is approximated by an appropriate distribution. Thus the burn-in period could be represented by a Weibull distribution with decreasing hazard function (i.e. $\alpha < 1$); from t_1 to t_2 (Fig. 2.6) an exponential model could be assumed and from t_2 to ∞ an appropriate distribution with increasing failure rate.

iii) Superposition. If it can be estimated what percentage of the population is more likely to fail due to early failures, chance failures and wearout, one can use superposition and write the density as

$$f(t) = p_b f_1(t) + p_c f_2(t) + p_w f_3(t)$$

where $f_1(t)$, $f_2(t)$, $f_3(t)$ are densities approximating the three regions and p_b, p_c, p_w are the weighting factors for burn-in, chance failures and wearout (naturally, $p_b + p_c + p_w = 1$). Difficulties may arise in determining the weighting factors.

When we have failure data we can use a slightly different approach. By plotting the data on probability paper we may be able to identify the various mechanisms of failure (for example, plotting on Weibull paper we may find distinct groups of points well fitted by straight lines, which means that several Weibull distributions represent the data; for an application see Ref. 35). Then we simply find the parameters of the distribution for each group and the sum of the hazard functions is the model for the hazard function of the population.

The distributions and the various models for the failure rate are useful tools in the hands of the analyst who may wish to study in detail the failures of certain objects. However in analyzing complex systems we seldom (if ever) use them; the exponential model is universally used and a constant failure rate is assigned to each component. This approximation is essential if methods for

the study of failures of complex systems, like fault trees, are to be advanced. Then it is natural to investigate how accurate such an approximation is.

Laboratory tests or operating experience yield a mean time to failure m, which is used in the exponential distribution. Of course, this MTTF is not the result of chance failures alone. If R(t) is the true reliability of the component, the following is true[5,36]

$$R(t) \geq e^{-t/m} \qquad \text{for} \qquad t \leq m \tag{2.35}$$

This inequality implies that the reliability of the component is underestimated when the exponential distribution is used with MTTF that of the aging component (in all practical applications the condition $t \leq m$ is satisfied).

Another implication of the inequality is that the exponential distribution predicts a shorter interval $(0, t_{exp})$ of successful operation at a given reliability level than the true interval $(0, t_{true})$ which would have been predicted by the true distribution. The percent error (PE) is defined as

$$PE = \frac{t_{true} - t_{exp}}{t_{exp}} 100$$

and it has been calculated for several distributions.[37] If the underlying distribution is the log-normal, this error is a function of the reliability level and the ratio $\frac{s}{m}$ where s is the standard deviation and m the mean for the log-normal distribution. It is given by

$$PE = -\frac{\left\{\left[1 + \left(\frac{s}{m}\right)^2\right]\frac{(2-1)}{2} + \ln R\right\}100}{\ell n R}$$

where R is the specified reliability and z is the solution of the equation

$$\frac{1}{\sqrt{2\pi}} \int_z^{\infty} \exp\left(-\frac{x^2}{2}\right) dx = R$$

Fig. 2.21 shows the behavior of PE for R = 0.50 (median life).

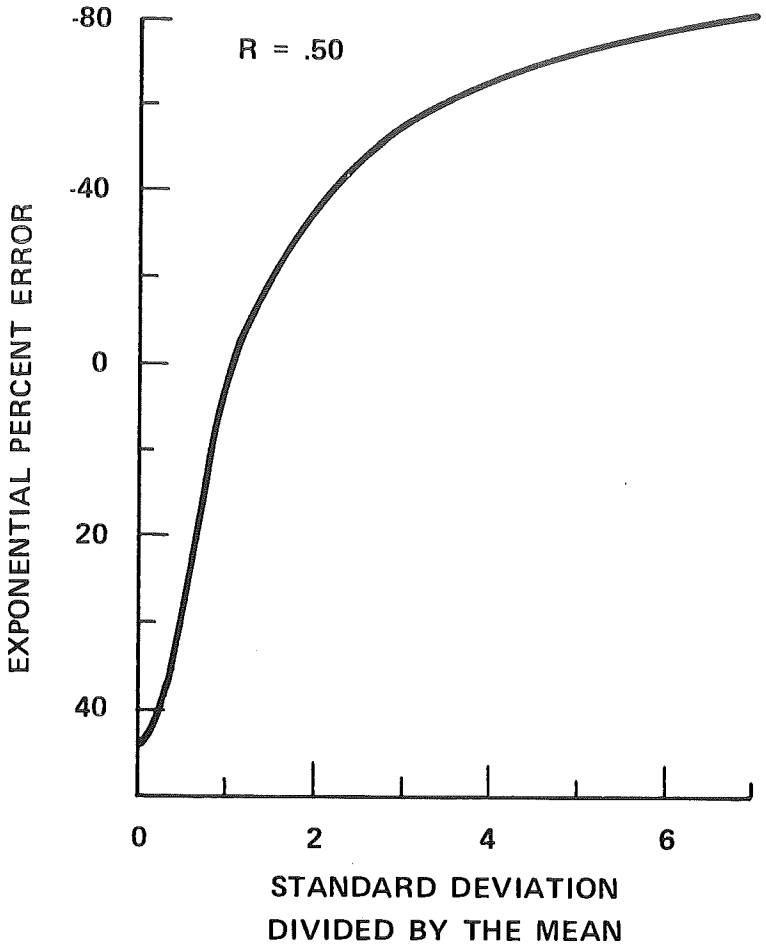


Figure 2.21. Percent Error vs the Log-normal Ratio s/m for the exponential median life. (Ref 37)

For the gamma distribution the percent error depends on the value of the shape parameter r. It is estimated by

$$PE = -\frac{\left(\frac{z}{r} + \ln R\right)100}{\ln R}$$

where

$$\frac{1}{\Gamma(r)} \int_{z}^{\infty} x^{r-1} e^{-x} dx = R$$

In Fig. 2.22 the percent error for the 90% reliable life (i.e. R = 0.90) is shown. Observe that for r = 1 the percent error is zero as expected, since the gamma distribution reduces to the exponential.

Finally, for the Weibull distribution PE depends on the shape parameter α as follows

$$PE = -\frac{\left[\frac{(-\ln R)^{1/\alpha}}{\Gamma\left(\frac{1}{\alpha} + 1\right)} + \ln R\right]100}{\ln R}$$

Fig. 2.23 shows PE for R = 0.90. Again, for α = 1 PE is zero, since the Weibull reduces to the exponential.

A final word of caution is in order here; when one uses the exponential model, it must be established that the stresses on the component are not much different than those under which the failure rate used was derived. If there is substantial difference the failure rate may be dramatically different and adjustments should be made. These adjustments are accomplished with the use of correction factors, which account for the different operational conditions of the device. Discussions of proposed models can be found in the references.[6,7,8,27,32,42]

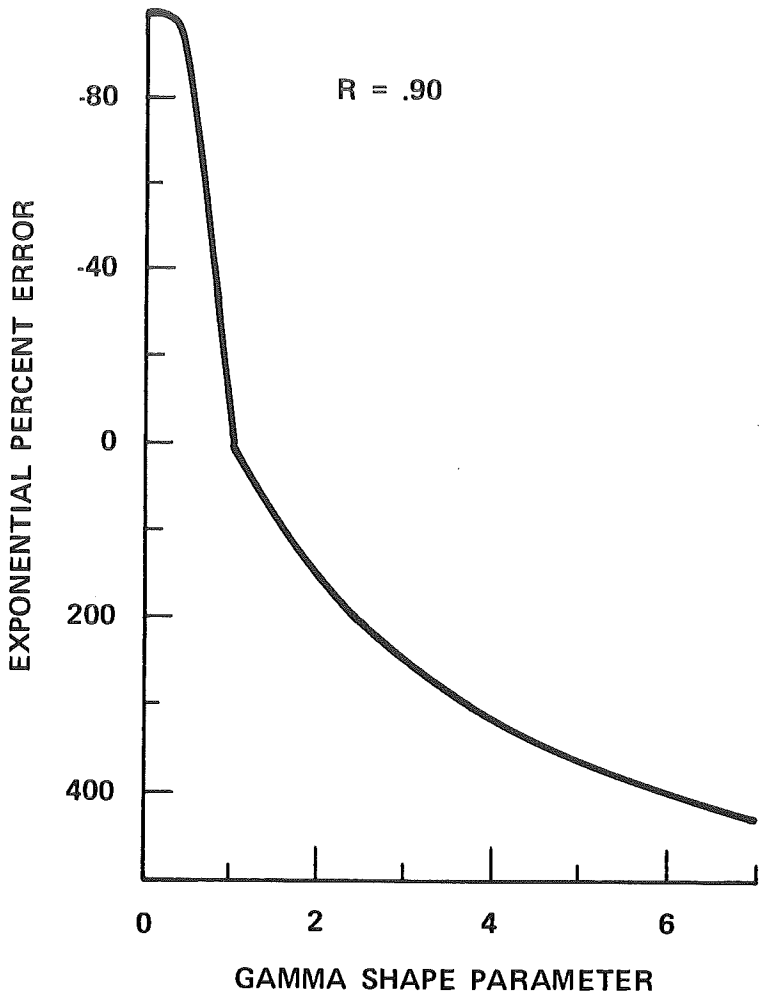


Figure 2.22. Percent Error vs Gamma Shape Parameter at a Reliability Level of 90% (Ref. 37).

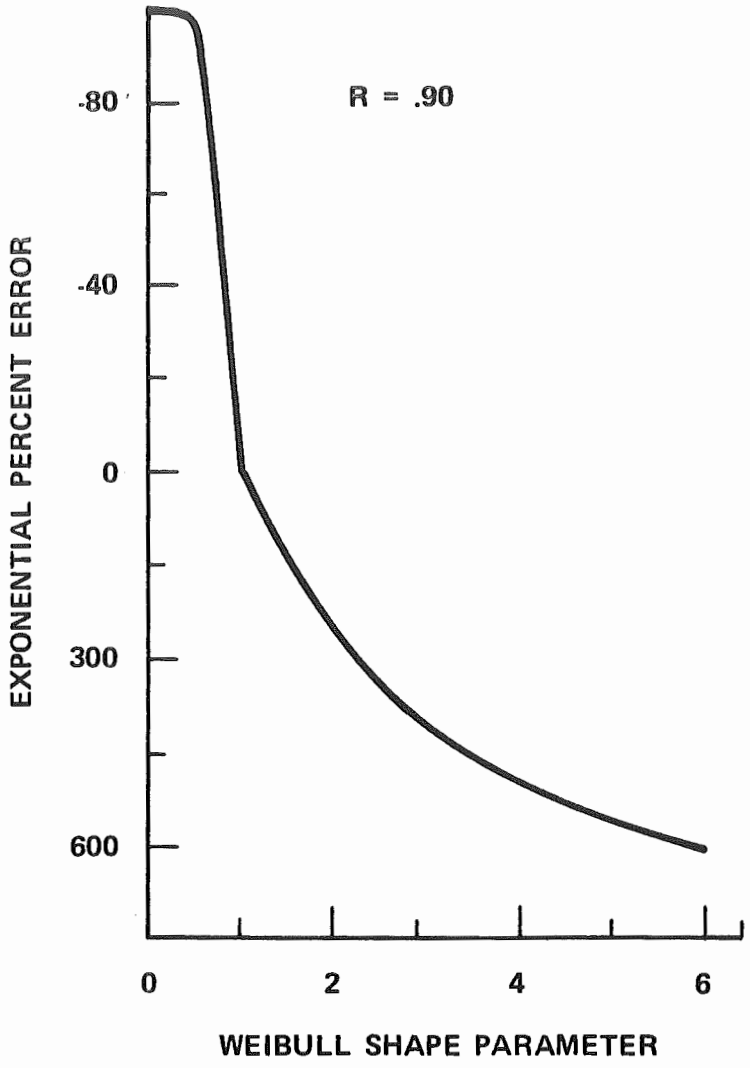


Figure 2.23. Percent Error vs Weibull Shape Parameter at a Reliability Level of 90% (Ref. 37).

2.B.5 Estimation of Parameters

In order to estimate the parameters of a distribution, which we believe from physical considerations that it represents the population, there are two questions that should be answered first, namely:

a) how was the data obtained, and

b) what method will be used

The first question arises in life tests, where n units are placed into operation and their times-to-failure are recorded. If all the units fail the data is called complete, otherwise it is incomplete. Incomplete data results if the test is terminated (truncated) at a time t_o while some units are still operating. Then the random variable is the number of failures r which have occurred. The test may also be terminated when the r^{th} failure occurs, in which case the random variable is the time of the r^{th} failure t_r. These types of data are called singly censored. Multiply censored data is obtained when observations are lost, units are withdrawn from the sample during the test, the units are put into operation at different times etc.

Finally, if the observations and censorship are made on the n components that were initially put into operation the test is called a nonreplacement test, while if the failed or censored components are replaced by new ones the test is termed as being one with replacement. Knowledge of these conditions is necessary for the correct treatment of the data and the extraction of the maximum possible amount of information from it.

The method to be used to estimate the parameters of the theoretical distribution is a very involved subject; details on the principles and criteria which are applied may be found in Section 2.A.5 of this report and in references 6,7,8,12 and 25.

Basically there are two methods of approach: 1) numbers are calculated for the unknown parameters by a specified method and these numbers are assumed to be the "best", in some sense, estimates of the parameters (point estimation), and 2) ranges of the unknown parameters are calculated and the probabilities that the true values of the parameters lie in these ranges are given (confidence intervals, interval estimation).

Point estimates are obtained either by the method of matching moments or the maximum likelihood method. The last method is more flexible and is considered superior to the first. Forming the likelihood function

$$L(t_1, t_2,\ldots,t_n; \theta_1, \theta_2,\ldots\theta_m) = \prod_{i=1}^{n} f(t_i; \theta_1,\ldots,\theta_m) \tag{2.13}$$

we can estimate any parameter θ_i by selecting that value which maximizes L or log L, since it is easier to work with the logarithm. This expression of the likelihood function is true for complete tests. For truncated or censored tests it must be changed to include the additional information. Thus, if the test is terminated at time t_o the likelihood function will be

$$L(t_1, t_2\ldots t_r; \theta_1, \theta_2,\ldots\theta_m) = \frac{n!}{(n-r)!} \prod_{i=1}^{r} f(t_i; \theta_1,\ldots,\theta_m) \left[1-F(t_0)\right]^{n-r} \tag{2.36}$$

where F(t) is the cumulative distribution function of the assumed probability density $f(t_i; \theta_1,\ldots,\theta_m)$. Thus the "best" estimate of θ_i is the solution of the equation

$$\frac{\partial \ln L}{\partial \theta_i} = 0 \tag{2.37}$$

This estimate $\hat{\theta}_i$ is a random variable (different samples will, in general yield different $\hat{\theta}_i$'s) and its variance is calculated from

$$\operatorname{var} \hat{\theta}_i = - \frac{1}{\dfrac{\partial^2 \ln L}{\partial \theta_i^{\,2}}} \tag{2.38}$$

for large n. Having the mean and the variance of the estimate we can find confidence intervals for θ_i either using Tchebycheff's inequality or using the fact that for large n the distribution of $\hat{\theta}_i$ tends to become normal. Of course, if the exact distribution of the estimator $\hat{\theta}_i$ is known (see exponential distribution below) we can find exact confidence intervals for samples of any size and we do not need to resort to these approximate methods.

Having point estimates and confidence intervals for the parameters of a distribution $F(t;\theta_i)$ we can attempt to answer two very important questions, namely:

a) Can we give confidence intervals for $R(t; \theta_i) = 1-F(t; \theta_i)$? If times-to-failure are studied, this is equivalent to assigning confidence intervals to the reliability of the units, e.g. What is a $100(1-\alpha)$ percent lower confidence limit on the probability that the units will survive a given interval of time $(0,t)$?

b) Reversing the problem in (a) we can seek confidence intervals for the time which corresponds to a given reliability, that is, can we make the statement "the probability that the units survive past the time t_γ is at least γ(pre-assigned reliability)" and be $(1-\alpha)100$ percent confident about the truthfulness of this statement? The quantity t_γ satisfies the equation $R(t_\gamma) = \gamma$ or $F(t_\gamma) = 1-\gamma$ and it is called the $(1-\gamma)$ fractile or quantile of $F(t)$.

Maximum likelihood point estimates of these quantities can be found by simply using the estimates for the parameters. Thus, for the extreme value distribution $F(t) = \exp[-\exp(-\alpha(t-\beta))]$ a point estimate for the reliability corresponding to a given time t is $\hat{R}(t) = 1-\exp[-\exp(-\hat{\alpha}(t-\hat{\beta}))]$. The $(1-\gamma)$ quantile is the solution of

$$e^{-e^{-\alpha(t_\gamma-\beta)}} = 1 - \gamma$$

hence $t_\gamma = \beta - \frac{1}{\alpha} \ln \left(\ln \frac{1}{1-\gamma} \right)$

and a point estimate is

$$\hat{t}_\gamma = \hat{\beta} - \frac{1}{\hat{\alpha}} \ln \left(\ln \frac{1}{1-\gamma} \right)$$

The estimation of confidence intervals for R(t) and t_γ is more complicated. If the distribution has only one parameter (e.g. the exponential), then the confidence limits on the parameter can be used directly to find confidence limits for R(t) and t_γ. However, when there are more than one estimated parameters this procedure is not applicable. Asymptotic results regarding the mean and variance of any function $G(\theta_i)$ of the parameters (like $R(t; \theta_i)$ and $t_\gamma(\theta_i)$) are given in Ref. 25; these can be used to find confidence intervals, as mentioned before. Since we will give the results for each distribution later we do not present the method of Ref. 25.

1. The Exponential Distribution

Point estimates and confidence intervals for the MTTF $\theta \left(= \frac{1}{\lambda} \right)$ have been found for all the types of tests. The results presented here are taken mainly from Ref. 43.

In all the tests $\hat{\theta}$ is estimated from

$$\hat{\theta} = \frac{T}{r} \qquad (2.39)$$

where T is the total operating time of censored, failed and unfailed components and r is the number of failures observed. The following results are obtained with the method of maximum likelihood.

For Tests Terminated at the r^{th} Failure: $(r \leq n)$

$$T = \sum_{i=1}^{r} t_i + (n-r)t_r \qquad \text{(without replacement)} \qquad (2.40)$$

$$T = nt_r \qquad \text{(with replacement)} \qquad (2.41)$$

$$T = \sum_{i=1}^{r} t_i + \sum_{j=1}^{k} t_j + (n-r-k)t_r \quad \begin{matrix}(k \text{ censored units at times } t_j, \\ \text{no replacement})\end{matrix} \tag{2.42}$$

(the expression for T can be modified along these lines to account for replacement of censored or failed components).

Two-sided 100(1-α) percent confidence interval:

$$\frac{2T}{\chi^2_{\frac{\alpha}{2},\,2r}} < \theta < \frac{2T}{\chi^2_{1-\frac{\alpha}{2},\,2r}} \tag{2.43}$$

one-sided 100(1-α) percent confidence interval:

$$\frac{2T}{\chi^2_{\alpha,\,2r}} < \theta \tag{2.44}$$

where $\chi^2_{\alpha,\,2r}$ is the upper α percentage point of the chi-square distribution obtainable from tables (notice that $\chi^2_{\alpha,\,2r}$ is the value which is exceeded with probability α; this clarification is important, since some tables give as $\chi^2_{\alpha,\,2r}$ the value which is not exceeded with probability α).

Since the distribution has only one parameter it is straightforward to find confidence intervals for the reliability and the quantiles. The point estimates are $\hat{R}(t) = e^{-t/\hat{\theta}}$ and $\hat{t}_\gamma = \hat{\theta} \log \frac{1}{\gamma}$.

Two-sided 100(1-α) percent confidence intervals:

$$\exp\left(-\frac{\chi^2_{\frac{\alpha}{2},2r}}{2T}\,t\right) < e^{-\frac{t}{\theta}} < \exp\left(-\frac{\chi_{1-\alpha/2,2r}}{2T}\,t\right) \tag{2.45}$$

$$\frac{2T \log \frac{1}{\gamma}}{\chi^2_{\alpha/2,2r}} < t_\gamma < \frac{2T \log \frac{1}{\gamma}}{\chi^2_{1-\alpha/2,2r}} \tag{2.46}$$

one-sided 100(1-α) percent confidence intervals:

$$\exp\left(-\frac{\chi^2_{\alpha,2r}}{2T}\,t\right) < e^{-t/\theta} \tag{2.47}$$

$$\frac{2T \log \frac{1}{\gamma}}{\chi^2_{\alpha, 2r}} < t_\gamma \qquad (2.48)$$

Example. Fifteen devices are put into operation and the test is terminated when three fail. Their times to failure are $t_1 = 83$ hr, $t_2 = 95$ hr and $t_3 = 110$ hr. Here $n = 15$ and $r = 3$ (no replacement). Therefore a point estimate for the MTTF is (using Eqs. (2.39) and (2.40))

$$\hat{\theta} = \frac{\sum_{i=1}^{3} t_3 + (15-3)t_3}{3} = \frac{1608}{3} = 536 \text{ hr.}$$

The one-sided 95% confidence interval is ($\alpha = 0.05$), Eq. (2.44),

$$\frac{2T}{\chi^2_{0.05,\ 6}} < \theta$$

From tables we find $\chi^2_{0.05,\ 6} = 12.6$, hence

$$\theta > \frac{2 \times 1608}{12.6} = 255 \text{ hr}$$

and we can state that we are 95 percent confident that the true MTTF is at least 255 hr (strictly speaking, we can only say that if we repeat the test many times then the estimate of θ will exceed 255 hr in 95 percent of the tests).

A point estimate of the reliability at 100 hr is $\hat{R}(100) = \exp(-100/536) = 0.83$ and the one-sided 95% confidence interval is

$$R(100) > e^{-100/255} = 0.67$$

Put into words, there is a probability of at least 0.67 that a unit will survive for 100 hr and this conclusion is true 95 percent of the time.

For the 0.10 quantile ($\gamma = 0.90$) a point estimate is

$$t_{0.9} = \hat{\theta} \log \frac{1}{0.90} = 258 \text{ hr}$$

and the 95% lower confidence limit

$$t_{0.9} > 255 \times \log \frac{1}{0.90} = 123 \text{ hr}$$

which means that $100\gamma = 90$ percent of the units of a lot will survive for at least 123 hr and this statement is true 95 percent of the time.

For Tests Terminated at Time t_o:

$$T = \sum_{i=1}^{r} t_i + (n-r)t_o \qquad \text{(without replacement)} \tag{2.49}$$

$$T = nt_o \qquad \text{(with replacement)} \tag{2.50}$$

$$T = \sum_{i=1}^{r} t_i + \sum_{j=1}^{k} t_j + (n-r-k)\, t_o \quad \text{(k censored units at times } t_j\text{, no replacement)} \tag{2.51}$$

two-sided $100(1-\alpha)$ percent confidence intervals: $(r < n)$

$$\frac{2T}{\chi^2_{\alpha/2,2r+2}} < \theta < \frac{2T}{\chi^2_{1-\alpha/2,2r}} \tag{2.52}$$

$$\exp\left(-\frac{\chi^2_{\alpha/2,2r+2}}{2T}\, t\right) < R(t) < \exp\left(-\frac{\chi^2_{1-\alpha/2,2r}}{2T}\, t\right) \tag{2.53}$$

$$\frac{2T \log \frac{1}{\gamma}}{\chi^2_{\alpha/2,2r+2}} < t_\gamma < \frac{2T \log \frac{1}{\gamma}}{\chi^2_{1-\alpha/2,2r}} \tag{2.54}$$

one-sided $100(1-\alpha)$ percent confidence intervals:

$$\frac{2T}{\chi^2_{\alpha,2r+2}} < \theta,\ \exp\left(-\frac{\chi^2_{\alpha,2r+2}}{2T}\, t\right) < R(t),\ \frac{2T \log \frac{1}{\gamma}}{\chi^2_{\alpha,2r+2}} < t_\gamma \tag{2.55}$$

Example: Ten units are tested for 100 hr (no replacement) and no failures are observed. Then $T = 10 \times 100 = 1000$ hr and a lower confidence interval can be given at 95% confidence level. From tables we find $\chi^2_{0.05,2} = 5.99$ hence

$$\frac{2 \times 1000}{5.99} \cong 334 \text{ hr} < \theta$$

A point estimate of θ cannot be given since no failures occurred (r=0). The 95% confidence intervals on reliability and the 0.1 fractile are

$$R(t) > \exp\left(-\frac{t}{334}\right)$$

and $t_\gamma > 334 \times \log \frac{1}{0.9} = 161$ hr

In some cases it may be impossible to find the total operating time T (e.g. when the times of failure of the devices t_i are not known). Then we can still give lower bounds to the reliability of the components using non-parametric methods. These results are not restricted to the case where the underlying distribution is the exponential, that is, they are distribution free.

For truncated tests (i.e., observations are made for an interval $(0, t_o)$ and r failures are observed) a lower bound to reliability at a $100(1-\alpha)$ percent confidence level is

$$R_L(t_o) = \frac{1}{1 + \left(\frac{r+1}{n-r}\right) F_{\alpha, 2r+2, 2n-2r}} \qquad (2.56)$$

where $F_{\alpha, 2r+2, 2n-2r}$ is the upper α percentage point of the F distribution with (2r+2) degrees of freedom in the numerator and (2n-2r) degrees of freedom in the denominator.

If the underlying distribution is assumed to be exponential, the above result leads to the following lower bound for the MTTF

$$\frac{t_o}{\ln\left[1 + \left(\frac{r+1}{n-r}\right) F_{\alpha, 2r+2, 2n-2r}\right]} < \theta \qquad (2.57)$$

Example: Consider the previous example with 10 units in operation for 100 hr with no failures. The non-parametric lower bound to the reliability at 95% confidence level is

$$R_L(100) = \frac{1}{1 + \frac{1}{10} F_{0.05,2,20}} = \frac{1}{1 + \frac{3.49}{10}} = 0.74$$

Assuming an exponential distribution the lower bound to the MTTF is

$$\frac{100}{\ln 1.349} = \frac{100}{0.3} \cong 333 \text{ hr} < \theta$$

which is very close to the bound found before (334 hr). This is expected, since no failures occurred, thus no information (failure times) was lost.

2. The Gamma Distribution

The method of maximum likelihood leads to a system of equations for the parameters of the distribution, which is too cumbersome to solve[44,45] and since the distribution is not used much in reliability studies we do not present the analysis here. Of course, when complete data are available a quick estimate of the parameters can be made by the method of matching moments or by probability plotting.

3. The Normal Distribution

For complete samples the mean and standard deviation are estimated from the sample mean and unbiased standard deviation, i.e.

$$\hat{m} = \frac{\sum_{i=1}^{n} t_i}{n} \tag{2.58}$$

$$\hat{\sigma} = \left[\frac{\sum_{i=1}^{n} (t_i - \hat{m})^2}{n-1}\right]^{1/2} = \left[\frac{n \sum_{i=1}^{n} t_i^2 - \left(\sum_{i=1}^{n} t_i\right)^2}{n(n-1)}\right]^{1/2} \tag{2.59}$$

two-sided $100(1-\alpha)$ percent confidence interval:

$$\hat{m} - t_{\alpha,r} \frac{\hat{\sigma}}{\sqrt{n}} < m < \hat{m} + t_{\alpha,r} \frac{\hat{\sigma}}{\sqrt{n}} \tag{2.60}$$

where $t_{\alpha,r}$ is found from tables of the t-distribution for two-sided confidence interval estimation and the degrees of freedom r is n-1.

One-sided 100(1-α) percent confidence intervals can be found using similar tables for $t_{\alpha,r}$ and they are of the form

$$\hat{m} - t_{\alpha,r}\frac{\hat{\sigma}}{\sqrt{n}} \qquad \text{lower bound} \tag{2.61}$$

and $$\hat{m} + t_{\alpha,r}\frac{\hat{\sigma}}{\sqrt{n}} \qquad \text{upper bound} \tag{2.62}$$

Since the distribution has two parameters, the calculation of confidence intervals for the reliability and the fractiles is not simple and tables must be used.

The (1-γ) quantile satisfies the relation

$$N\left(\frac{t_\gamma - m}{\sigma}\right) = \frac{1}{\sqrt{2\pi}}\int_0^{\frac{t_\gamma - m}{\sigma}} \exp\left(-\frac{\tau^2}{2}\right) d\tau = 1-\gamma$$

or $t_\gamma = m + y_\gamma \sigma$ where y_γ is the standard normal variate, such that $N(y_\gamma) = 1-\gamma$. Confidence limits on y_γ at a (1-α)100 percent confidence level are given in tables.[25,46] Given γ, α and the size of the sample n Table A.9 of Ref. 25 (or, Table 8.4 of Ref. 46) gives the one-sided lower confidence limit on y_γ as - k. (Notice that in the tables γ and α are what we call here (1-α) and (1-γ)). Then the corresponding limit on t_γ are $\hat{m} - k\hat{\sigma}$. Table 8.3 of Ref. 46 gives the two-sided confidence limits $\hat{m} - k\hat{\sigma}$ (the number k is called a tolerance factor). Notice that these limits are not the same with the ones given before which utilize the t-distribution; the latter give confidence limits for the mean only.

Confidence limits on the reliability of a given interval (0,t) are found by reversing the above procedure. Now $k = \frac{\hat{m}-t}{\hat{\sigma}}$ and from the same tables the value of γ is found given k, α and n. The following example will clarify the procedure.

Example: Ten units fail at times t_1 = 1500 hr, t_2 = 1550 hr, t_3 = 1625 hr, t_4 = 1715 hr, t_5 = 1750 hr, t_6 = 1785 hr, t_7 = 1800 hr, t_8 = 1865 hr, t_9 = 1900 hr and t_{10} = 1950 hr. Assuming a normal distribution we estimate

$$\hat{m} = \frac{\sum_{i=1}^{10} t_i}{10} = 1744 \text{ hr.}$$

$$\hat{\sigma} = 147 \text{ hr}$$

Since $t_{0.05,9} = 1.833$ (one-sided) the 95% confidence interval for the mean is

$$m > 1744 - 1.833 \times \frac{147}{\sqrt{10}} = 1659 \text{ hr}$$

We now seek a one-sided 95% (α=0.05) confidence limit on the 0.1 fractile of the distribution (γ=0.9). From Table A.9 of Ref. 25 we find for α=0.05, γ=0.09 (γ=0.95 and α=0.1 for the Tables) and n=10 that k=2.355 therefore

$$t_{0.9} > \hat{m} - k\hat{\sigma} = 1744 - 2.355 \times 147 = 1398 \text{ hr}$$

and we claim that in 95 percent of future tests the lifetime of 90 percent of the units will be at least 1398 hr.

To find confidence intervals for the reliability in an interval of 2000 hr we calculate

$$k = \frac{2000 - 1744}{147} = 1.74$$

We wish to find the 95% confidence interval, thus the tables for k = 1.465 and k = 2.355, γ = 0.95 and n = 10 yield α = 0.25 and 0.10. Using simple interpolation (another method is described in Ref. 25) we find that

$$\alpha = 0.1 + 0.15 \times \frac{2.355 - 1.74}{2.355 - 1.465} \cong 0.2.$$

In our notation $1 - \gamma = 0.2 \Rightarrow \gamma = 0.8$ and we make the statement that with confidence 95% the reliability over an interval of 2000 hr is at least 0.8.

For truncated tests the results become more complicated.[7,14] Assuming that when the test is terminated at t_o only r failures have occurred we define the degree of truncation as

$$h = \frac{n-r}{n} \tag{2.63}$$

and the quantity

$$y = \frac{r \sum_{i=1}^{r} (t_o - t_i)^2}{2\left[\sum_{i=1}^{r} (t_o - t_i)\right]^2} \tag{2.64}$$

Then the parameters of the distribution are found from

$$\hat{\sigma} = \frac{\sum_{i=1}^{r} (t_o - t_i)}{r}\, g(h,z) \tag{2.65}$$

(for an unbiased estimate use r-1 in the denominator) and

$$\hat{m} = t_o + z(h,y)\hat{\sigma} \tag{2.66}$$

The procedure is as follows: from the data and Eqs. (2.63) and (2.64) we calculate h and y. Table X of Ref. 14 then gives z(h,y) and the quantity $\psi'(z)$ which is used to find g(h,z) from the equation

$$g(h,z) = \frac{r}{(n-r)\psi'(z) - rz} \tag{2.67}$$

Having g(h,z) and z(h,y) the parameters $\hat{\sigma}$ and $\hat{m}$ can be calculated from Eqs. (2.65) and (2.66).

Confidence intervals for the mean can be found by utilizing the fact that for large samples $\hat{m}$ is approximately normally distributed with variance

$$\frac{\hat{\sigma}^2}{r}\, \mu_{11}(z)$$

where $\mu_{11}(z)$ is again found from Table X of Ref. 14 knowing z. Therefore, we have two-sided $100(1-\alpha)$ percent confidence interval

$$\hat{m} - k_{\alpha/2}\sqrt{\frac{\hat{\sigma}^2}{r}\mu_{11}(z)} < m < \hat{m} + k_{\alpha/2}\sqrt{\frac{\hat{\sigma}^2}{r}\mu_{11}(z)} \tag{2.68}$$

where $k_{\alpha/2}$ is the number of standard deviations of a normal distribution such that the value $m + k_{\alpha/2}\sigma$ is exceeded with probability $\frac{\alpha}{2}$.

Similarly the following one-sided confidence intervals are obtained

$$\hat{m} - k_{\alpha}\sqrt{\frac{\hat{\sigma}^2}{r}\mu_{11}(z)} < m \tag{2.69}$$

and $$m < \hat{m} + k_{\alpha}\sqrt{\frac{\hat{\sigma}^2}{r}\mu_{11}(z)} \tag{2.70}$$

Confidence intervals for the reliability and the fractiles can be found as before using the mean and the unbiased variance and tables

Example: Suppose that the test mentioned before was terminated at t_o = 1850 hr. Then only r = 7 failures are recorded. The degree of truncation is

$$h = \frac{n-r}{n} = \frac{3}{10} = 0.3$$

and y is calculated to be

$$y = 0.72$$

From Table X, Ref. 14, we find $z(0.3;0.72) = -0.54$, $\psi'(-0.54) = 1.156$ and $\mu_{11}(-0.54) = 1.141$, hence, from Eq. (2.67),

$$g(0.3, -0.54) = \frac{7}{3 \times 1.156 + 7 \times 0.54} = 0.965$$

Therefore

$$\hat{\sigma} = 121 \text{ hr.}$$

and $$\hat{m} = 1850 - 0.54 \times 121 = 1785 \text{ hr}$$

Also, since $k_{0.05} = 1.64$ (from tables of normal distribution), a lower bound for the mean at a confidence level of 95% can be given

$$m > 1785 - 1.64\sqrt{\frac{121^2}{7} \times 1.141} = 1707 \text{ hr}$$

4. The Log-Normal Distribution

Since the logarithm of the random variable is normally distributed it suffices to take the logarithms of the observations and use the methods for the estimation of parameters of a normal distribution.

5. The Weibull and Type I Extreme Value Distributions (Smallest Values).

The Weibull distribution is to the Type I extreme value distribution of minimum values as the log-normal distribution is to the normal distribution, that is, if in the extreme value distribution

$$F(t) = 1 - \exp\left(-e^{\alpha(t-\beta)}\right)$$

we make the transformation $t = \ln t'$ and $\beta = \ln \beta'$ we get the Weibull distribution

$$F(t) = 1 - e^{-(t'/\beta')^{\alpha}}$$

Therefore methods for estimating the parameters of the one distribution can also be applied to the other by a simple transformation.

The estimation of parameters leads to equations which cannot be solved analytically and iteration or Monte Carlo techniques are employed. A review and comparison of the various methods is given in Ref. 47. Estimation by the method of maximum likelihood is presented in Ref. 45 and 48.

Similarly, confidence intervals on the reliability and the quantiles cannot be found without the aid of the computer and usually tables are generated for certain parameters which help in determining such intervals. Thus, in Ref. 49 tables are given for the estimation of an exact lower confidence bound on the reliability (these tables can also be used for confidence intervals on the fractiles[47]). In Ref. 50 confidence intervals on the parameters are estimated and in Ref. 51 intervals for the reliability and quantiles are given; in both references complete data are considered and the maximim likelihood method is used.

We present here a method for estimating the parameters with the use of tables based on a method called best linear invariant estimation.[52,53] If a sample of n units is put into operation and the test is terminated when m units fail ($m \leq n$) the estimates for the parameters of the extreme value distribution are

$$\tilde{\beta} = \sum_{i=1}^{m} A_{i,m,n}\, t_i \tag{2.71}$$

and

$$\tilde{\alpha} = \left[\sum_{i=1}^{m} C_{i,m,n}\, t_i\right]^{-1} \tag{2.72}$$

where $A_{i,m,n}$ and $C_{i,m,n}$ can be found in the tables of Ref. 52 for $n = 2,\ldots,15$ and $m = 2,\ldots,n$ (for $n > 15$ see Ref. 52 and 49).

If the distribution is Weibull the above procedure is applicable of the failure times and the parameter $\tilde{\beta}' = \exp(\tilde{\beta})$.

The $(1-\gamma)$ quantiles are

$$t_\gamma = \beta + \frac{1}{\alpha} \ln \ln \left(\frac{1}{\gamma}\right) \qquad \text{(extreme value distribution)} \tag{2.73}$$

and

$$t_\gamma = \beta' \left(\ln \frac{1}{\gamma}\right)^{1/\alpha} \qquad \text{(Weibull distribution)} \tag{2.74}$$

Point estimates are found using the point estimates of α and $\beta(\beta')$. A lower bound on t_γ at confidence level $100(1-\alpha)$ percent is given by

$$t_\gamma \geq \tilde{\beta} - \frac{1}{\tilde{\alpha}} \left(V_\gamma\right)_{1-\alpha} \qquad \text{(extreme value distribution)} \tag{2.75}$$

and

$$t'_\gamma \geq \exp\left[\tilde{\beta} - \frac{1}{\tilde{\alpha}} \left(V_\gamma\right)_{1-\alpha}\right] \qquad \text{(Weibull distribution)} \tag{2.76}$$

where $\left(V_\gamma\right)_{1-\alpha}$ is found from tables in Ref. 53 for $\gamma = 0.90$, 0.95 and 0.99.

A lower bound to reliability is not readily available from these tables, since values of V are listed for only three values of γ. One calculates $\tilde{\alpha}(\tilde{\beta}-t)$ and if this value is approximately equal to a tabulated value of V_γ (for the fixed m and n and the specified confidence level), then the

corresponding γ is an approximate bound to the probability of survival past the time t(extreme value distribution). Of course, other methods may be used to find such a lower bound, like the one described in Ref. 49 or the non-parametric result given in the discussion of the exponential distribution (for tests terminated at a fixed time) or the asymptotic method described in Ref. Ref. 25.

Example: Ten units are tested and five fail at times (hr) $t_1 = 60$, $t_2 = 95$, $t_3 = 124$, $t_4 = 140$ and $t_5 = 160$. The underlying distribution is assumed to be Weibull. To find the parameters we first calculate the natural logarithms of the times to failure, $\ln t_1 = 4.1$, $\ln t_2 = 4.55$, $\ln t_3 = 4.83$, $\ln t_4 = 4.94$ and $\ln t_5 = 5.07$. Using the tables of Ref. 52 for $n = 10$ and $m = 5$ we find (Eqs. (2.71) and (2.72))

$$\tilde{\beta} = -0.1155 \times 4.1 - 0.0908 \times 4.55 - 0.0513 \times 4.83 + 0.0009 \times 4.94 + 1.2568 \times 5.07 = 5.259$$

and $$\tilde{\alpha} = \left[-0.1851 \times 4.1 - 0.1818 \times 4.55 - 0.1606 \times 4.83 - 0.1253 \times 4.94 + 0.6529 \times 5.07\right]^{-1} = 3.02$$

Therefore, the parameters of the Weibull distribution are $\tilde{\alpha} = 3.02$ and $\tilde{\beta}' = e^{5.259} = 192$ hr.

For $n = 10$, $m = 5$, $\gamma = 0.95$ and $1-\alpha = 0.95$ table 4 of Ref. 53 gives the value $(V_{0.95})_{0.95} = 8.39$ and we calculate

$$t_{0.95} \quad \exp\left[5.259 - \frac{8.39}{3.02}\right] = 12 \text{ hr.}$$

which means that the minimum life, for which the reliability is 0.95, is 12 hr with confidence 95 percent.

6. Type I Asymptotic Distribution of Maximum Values

The method of maximum likelihood will be used to estimate the parameters α and β. The results are taken from Ref. 54, 55 and 56.

If n observations are available the estimates of α and β are the solutions of

$$\left(\sum_{i=1}^{n} t_i \, e^{-\hat{\alpha} t_i}\right) \Big/ \left(\sum_{i=1}^{n} e^{-\hat{\alpha} t_i}\right) + \frac{1}{\hat{\alpha}} = \frac{\sum_{i=1}^{n} t_i}{n} \tag{2.77}$$

$$\hat{\beta} = \frac{1}{\hat{\alpha}} \ln \left(\frac{\sum_{i=1}^{n} e^{-\hat{\alpha} t_i}}{n} \right) \tag{2.78}$$

These equations are solved by some iterative procedure in a computer (notice that the first can be solved independently). As a first approximation the estimates resulting from the method of moments may be used, that is, the solutions of

$$\hat{\beta}_1 + \frac{0.577}{\hat{\alpha}_1} = \frac{\sum_{i=1}^{n} t_i}{n} = \bar{t} \tag{2.79}$$

and

$$\frac{1.645}{\hat{\alpha}_1^2} = \frac{\sum_{i=1}^{n} (t_i - \bar{t})}{n-1} = s^2 \tag{2.80}$$

An estimate of the return period is

$$\hat{T}(t) = e^{\hat{\alpha}(t-\hat{\beta})} \tag{2.81}$$

and in most applications it suffices to assume that the probability of T(t) being in the interval 0.32 $\hat{T}(x)$ and 3.13 $\hat{T}(x)$ is $\frac{2}{3}$ (= 0.68).

The $(1-\gamma)$ quantile (i.e., the solution of $\exp[-\exp(-\alpha(t_\gamma - \beta))] = 1-\gamma$) is

$$t_\gamma = \beta - \frac{1}{\alpha} \ln \left(\ln \frac{1}{1-\gamma} \right) \tag{2.82}$$

and a point estimate is

$$\hat{t}_\gamma = \hat{\beta} - \frac{1}{\hat{\alpha}} \ln \left(\ln \frac{1}{1-\gamma} \right) \tag{2.83}$$

Its variance is

$$\hat{\sigma}^2_{t_\gamma} = \frac{\left\{1 + 1.645 \left[1\text{-}0.577 - \ln\left(\ln \frac{1}{(1-\gamma)}\right)\right]^2\right\}}{\hat{\alpha}^2 \, n} \tag{2.84}$$

For large n we assume a normal distribution and we can claim that there is a probability $(1-\gamma)$ that the largest observed value in any year will be at most t_γ where t_γ lies in the interval $\hat{t}_\gamma \pm \hat{\sigma}_{t_\gamma}$ and this will occur 68.2 percent of the time (we could also consider the interval $\hat{t}_\gamma \pm 2\hat{\sigma}_{t_\gamma}$ in which case the confidence rises to 95.4 percent). In a similar way we can define one-sided bounds for t_γ; thus, there is a probability $(1-\gamma)$ that the largest value in any year will be at most $\hat{t}_\gamma + 1.64 \, \hat{\sigma}_{t_\gamma}$ with confidence 95 percent

The previous result may also be presented in another form: a point estimate of $t_y = \beta + \frac{y}{\alpha}$ (y is called the reduced variate) is

$$\hat{t}_y = \hat{\beta} + \frac{y}{\hat{\alpha}} \tag{2.85}$$

and its variance is

$$\hat{\sigma}^2_{t_y} = \left[1 + 1.645 \, (1\text{-}0.577 + y)^2\right] \Big/ (\hat{\alpha}^2 \, n) \tag{2.86}$$

We now wish to predict the maximum observation in the next m samples (years). We already know that if the distribution of the maximum in one year is known then the distribution of the maximum in the next m years will be again the asymptotic distribution of maximum values with parameters easily obtainable. However, now we only know estimates of the parameters of the distribution of maxima in one year and as a result the best we can do is find the mean and the variance of the maximum in m years. Writing τ_m for this maximum its mean is

$$\hat{m}_{\tau_m} \equiv E[\tau_m] = \hat{\beta} + \frac{0.577 + \ln m}{\hat{\alpha}} \tag{2.87}$$

and its variance

$$\hat{\sigma}^2_{\tau_m} \equiv \text{var}[\tau_m] = \frac{1.645}{\hat{\alpha}^2} + \left[1 + 1.645(1+\ln m)^2\right] \Big/ (\hat{\alpha}^2 n) \qquad (2.88)$$

Using Tchebycheff's inequality (Eq. (2.9)) we find that

$$P\left[\hat{m}_{\tau_m} - k\hat{\sigma}_{\tau_m} < \tau_m < \hat{m}_{\tau_m} + k\sigma\right] \geq 1 - \frac{1}{k^2}$$

and, as an example, the probability that the maximum observation in the next m years lies in the interval $\hat{m}_{\tau_m} \pm 3\hat{\sigma}_{\tau_m}$ is at least 0.889.

An approximate point estimate to the value $\hat{\beta} + \frac{c}{\hat{\alpha}}$ (c to be specified) which will be exceeded with probability γ in the next m years can also be given by estimating

$$c = -\ln\left(\ln\frac{1}{1-\gamma}\right) + \ln m + \frac{\left(1-\ln\frac{1}{1-\gamma}\right)}{2n}\left[1 + 1.645\left(1-0.577 - \ln\left(\ln\frac{1}{1-\gamma}\right) + \ln m\right)^2\right] \qquad (2.89)$$

If a point estimate of the probability γ is needed with given c the reverse procedure is applied (i.e., the above equation is solved for γ).

Example: In Ref. 57 the flooding hazard for a nuclear reactor (Monticello site) is studied. The floods for the years 1927-1970 (n = 44) have been recorded. Then the maximum likelihood estimates of the parameters of the extreme value distribution are found to be $\hat{\beta}$ = 15145 ft^3/sec and $\hat{\alpha}^{-1}$ = 6736 ft^3/sec. Using the method of matching moments the estimates are $\hat{\beta}$ = 15155 ft^3/sec and $\hat{\alpha}^{-1}$ = 6712 ft^3/sec.

Given these estimates we can make some probabilistic statements. The return period of a flood of size t = 50000 ft^3/sec. is, Eq. (2.81),

$$\hat{T}(50000) = \exp\left[\frac{50000-15145}{6736}\right] = 180 \text{ years}$$

and there is a probability of 0.68 that this flood will occur in as short a perior as 0.32 x 180 = 57.6 yr or as long a period as 3.13 x 180 = 564 yr.

Reversing the procedure we can find the flood t with return period $\hat{T}(t) = 1000$ yr by

$$t = 6736 \times \ln 1000 + 15145 = 61600 \text{ ft}^3/\text{sec.}$$

thus, every 1000 years we expect one flood of magnitude at least 61600 ft^3/sec. To find the confidence limits we have

$$t' = 6736 \times \ln 320 + 15145 = 53900 \text{ ft}^3/\text{sec.}$$

and $t'' = 6736 \times \ln 3130 + 15145 = 69300 \text{ ft}^3/\text{sec}$

therefore, every 1000 years there is a probability 0.68 that the largest flood which occurs is as low as 53900 ft^3/sec or as large as 69300 ft^3/sec.

We now ask the question: What is the maximum flood which has a probability 0.90 of occurring in any year? From the quantile with $1-\gamma = 0.90$ we get a point estimate

$$t_\gamma = t_{0.1} = 15145 - 6736 \times \ln\left(\ln \frac{1}{0.90}\right) = 30000 \text{ ft}^3/\text{sec.}$$

For confidence limits we need the standard deviation

$$\hat{\sigma}_{t_{0.1}} = \left\{1 + 1.645 \left[1\text{-}0.577 - \ln\left(\ln \frac{1}{0.90}\right)\right]^2\right\}^{1/2} \times 6736 \Big/ \sqrt{44} = 2560 \text{ ft}^3/\text{sec}$$

hence there is a probability 0.90 that the flood in any year is at most $30000 + 1.64 \times 2560 = 34200$ ft^3/sec. and this statement is correct 95 percent of the time.

Finally, a prediction for a period of m = 40 yr. will be made. The expected mean value of the floods over the forty years is

$$\hat{m}_{\tau_m} = 15145 + (0.577 + \ln 40) \times 6736 = 43900 \text{ ft}^3/\text{sec.}$$

and the standard deviation is found to be 10600 ft^3/sec. Therefore, there is a probability of at least 0.889 that a flood of size in the interval $43900 \pm 3 \times 10600$ or (12100, 75700) ft^3/sec. will occur in the next 40 years. A point estimate of the most probable flood in 40 years is $\hat{\beta} + \frac{\ln m}{\hat{\alpha}} = 39945$

ft^3/sec. To find an estimate of the flood which will be exceeded with proba probability 0.05 ($\gamma = 0.05$) we calculate $c = 7.055$, thus this flood is $15145 + 7.055 \times 6736 = 62700\ ft^3/sec$.

7. Superposition of Distributions

Combinations of distributions lead to many problems where the parameters must be estimated from test data. We present here a model involving two exponential distributions and references are given for other models of interest.

Suppose that components are subject to chance failures from two independent causes; then the compound exponential model applies, that is

$$F(t) = 1 - e^{-(\lambda_1+\lambda_2)t}$$

and $$f(t) = (\lambda_1+\lambda_2)e^{-(\lambda_1+\lambda_2)t}$$

Failure times are collected from n operating units from which r have failed when the test is terminated at t_o (truncated test). A particular unit fails due to either one of the causes and the cause of failure can be identified for a failed unit; thus r_1 units have failed due to the first cause and r_2 due to the second. We define t_{ij} to be the time of failure of the j^{th} unit due to the i^{th} (i=1, 2) cause. Writing $\bar{t}_i = \sum_{j=1}^{r_1} \frac{t_{ij}}{r_i}$, the maximum likelihood estimates for λ_1 and λ_2 are[58]

$$\hat{\lambda}_i = \left[\frac{r_1\bar{t}_1 + r_2\bar{t}_2 + (n-r)t_o}{r_i}\right]^{-1}, \quad i = 1, 2,\ r_1 + r_2 = r$$

An estimate for the overall failure rate is, naturally, $\hat{\lambda} = \hat{\lambda}_1 + \hat{\lambda}_2$. The variances of the estimators are given in Ref. 58 and in Ref. 59 the problem of estimation when the exact failure times are not known but the data is collected at certain times t_j is considered.

In Ref. 19 a method is described for the estimation of parameters of the compound model of an exponential and a normal distribution.

Consider now the case of a mixture of two or more distributions, i.e. $f(t) = \sum_{i=1}^{m} p_i f_i(t)$. The type of $f_i(t)$ is known and we wish to estimate its parameters and the mixing parameters p_i. References 60, 61 and 62 deal with the problem when the $f_i(t)$ are normal distributions and Ref. 22 and 63 deal with Weibull distributions.

2.B.6 Plotting Methods

Probability plotting is an easy and fast method of not only estimating the parameters of a distribution from sample data but also of checking how well the chosen distribution represents the observations. Detailed instructions as well as theoretical justification of the method are given in Ref. 4. This approach can be employed only for complete or singly censored samples.

A simple technique which may also be used in the case of multiply censored data has been developed recently. Instead of plotting the cumulative frequencies vs. the observations we plot the cumulative hazard function vs. the observations (hazard plotting method). The cumulative hazard is defined as

$$H(t) = \int_0^t h(\tau)d\tau \qquad (2.90)$$

and is related to the distribution function through

$$H(t) = \ln[1-F(t)] \qquad (2.91)$$

The theory and applications of hazard plotting are presented in Ref. 64, 65 and 66.

2.C. SIMPLE SYSTEMS

2.C.1. Introduction

The analysis of complex systems is very effectively performed with the use of path (cut) sets, that is, the identification of groups of events which together cause system success (failure). This procedure reduces a complex situation into a much simpler one, where events are interrelated in a manner which permits the direct use of probabilistic methods to calculate the probability of success (failure) of the system.

The simple configurations which will be presented here are not only useful in the analysis of complex systems. If a component is assigned the task of performing a certain function, the probability of successful performance (reliability) is improved with the use of more than one components which can perform the same function (redundancy).

In what follows we assume that the failures of the components are independent and that the probability of the i^{th} component being up is p_i and down q_i. Knowing the p_i's and the configuration we will calculate the probability P that the system is functioning properly. In time-dependent situations we will calculate the reliability of the system R(t) as a function of component reliability $R_i(t)$. Having the reliability we can calculate the mean time to failure of the system m using the equation

$$m = \int_0^\infty R(t)\,dt \ . \tag{2.92}$$

This equation is readily proved from the definition

$$m = \int_0^\infty tf(t)\,dt \quad , \quad (f(t) = \text{failure density})$$

where the substitution

$$f(t) = -\frac{dR(t)}{dt}$$

is made and the integral is evaluated by parts.

Useful expressions for the MTTF can also be given in terms of the Laplace transforms of R(t) and f(t). Defining the Laplace transformation as

$$\tilde{R}(s) = \int_0^\infty e^{-st} R(t)\,dt \tag{2.93}$$

it is immediately seen from Equation (2.92) and Equation (2.93) that

$$m = \tilde{R}(o) \quad . \tag{2.94}$$

Furthermore, if

$$\tilde{f}(s) = LT[f(t)]$$

then

$$-\frac{d\tilde{f}(s)}{ds} = LT[tf(t)]$$

thus

$$m = -\frac{d\tilde{f}(s)}{ds}\bigg|_{s=0} = \int_0^\infty tf(t)\,dt \quad . \tag{2.95}$$

2.C.2. Series System

A group of N components are said to be in series if all the components must function in order for the system to function (Fig. 2.24).

From the definition it follows that

$$P = \prod_{i=1}^{N} p_i \tag{2.96}$$

and for identical elements ($p_1 \equiv p_2 \equiv \ldots \equiv p_N$)

$$P = p^N \quad .$$

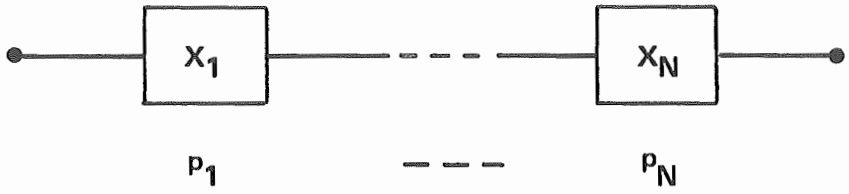


Figure 2.24. Series System.

The reliability of the system is

$$R(t) = \prod_{i=1}^{N} R_i(t) \tag{2.97}$$

and when $R_i(t) = e^{-\lambda_i t}$

$$R(t) = e^{-t \sum_{i=1}^{N} \lambda_i} \tag{2.98}$$

thus the failure rate of the system is

$$\lambda = \sum_{i=1}^{N} \lambda_i \tag{2.99}$$

and its MTTF

$$m = \frac{1}{\sum_{i=1}^{N} \lambda_i} \quad . \tag{2.100}$$

The series system is the only one in which components with constant failure rates induce a constant failure rate for the system. In all other configurations the reliability of the system is not exponential.

Since the system functions if all its components function, its reliability is smaller than the reliability of any of the components. Another way to look at this is by defining τ_i to be the time to failure of the i^{th} component. Then the system fails at a time t which is

$$t = \min\{\tau_1, \ldots, \tau_N\} \quad . \tag{2.101}$$

2.C.3. Parallel System

In a parallel (Fig. 2.25) configuration N components are performing the same function and any one component can successfully continue the operation (i.e., N-1 failures are allowed).

Since the system fails if all its elements fail we have

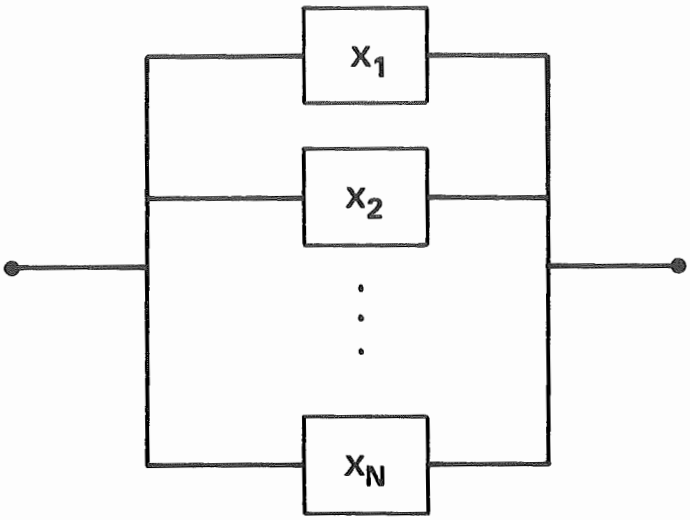


Figure 2.25. Parallel System.

$$Q = 1 - P = \prod_{i=1}^{N} q_i \tag{2.102}$$

or

$$P = 1 - \prod_{i=1}^{N} (1-p_i) \quad . \tag{2.103}$$

For identical elements

$$P = 1 - (1-p)^N \quad . \tag{2.104}$$

The reliability is

$$R(t) = 1 - \prod_{i=1}^{N} [1-R_i(t)] \tag{2.105}$$

and for exponential components

$$R(t) = 1 - \prod_{i=1}^{N} [1-e^{-\lambda_i t}]$$

and

$$m = \sum_{i=1}^{N} \frac{1}{\lambda_i} - \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{1}{\lambda_i + \lambda_j} +$$

$$+ \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \sum_{k=j+1}^{N} \frac{1}{\lambda_i+\lambda_j+\lambda_k} - \ldots + (-1)^{N-1} \frac{1}{\sum_{i=1}^{N} \lambda_i} \quad . \tag{2.106}$$

For identical elements

$$m = \sum_{n=1}^{N} \frac{1}{n\lambda} \quad . \tag{2.107}$$

As an example consider two units with failure rates λ_1 and λ_2. The reliability of the system is

$$R(t) = 1 - (1 - e^{-\lambda_1 t})(1-e^{-\lambda_2 t}) = e^{-\lambda_1 t} + e^{-\lambda_2 t} - e^{-(\lambda_1+\lambda_2)t}$$

and its mean time to failure

$$m = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} - \frac{1}{\lambda_1 + \lambda_2} \quad . \tag{2.108}$$

If $\lambda_1 \equiv \lambda_2 \equiv \lambda$,

$$R(t) = 2e^{-\lambda t} - e^{-2\lambda t} \tag{2.109}$$

and

$$m = \frac{1}{\lambda} + \frac{1}{2\lambda} = \frac{3}{2\lambda} \quad . \tag{2.110}$$

Since the system fails when all its elements fail, the time-to-failure of the system is related to the component failure times τ_i by

$$t = \max\{\tau_1, \ldots, \tau_N\} \quad . \tag{2.111}$$

2.C.4. r-out-of-N System

A generalization of the previous case is when N identical components function in parallel and r are needed (instead of only one).

The probability that any k of N components are functioning is (binomial distribution)

$$P_k = \binom{N}{k} p^k (1-p)^{N-k} = \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k} \quad .$$

Since the system is up if at least r components are functioning, the probability of successful performance is

$$P = \sum_{k=r}^{N} P_k = \sum_{k=r}^{N} \binom{N}{k} p^k (1-p)^{N-k} \quad . \tag{2.112}$$

For exponential components

$$R(t) = \sum_{k=r}^{N} \binom{N}{k} e^{-k\lambda t} (1-e^{-\lambda t})^{N-k} \tag{2.113}$$

and

$$m = \sum_{k=r}^{N} \frac{1}{k\lambda} \quad . \tag{2.114}$$

A common feature of the configurations studied above (series and parallel) is that the expressions for reliability are derived by merely replacing p_i with $R_i(t)$. The formulas for P may be interpreted as holding at every point in time and the state of the system is determined by the present state of its components. This is no longer true in the important case of a standby system as it will be seen shortly; the whole history of the system from t=0 must be considered.

2.C.5. Standby System

In a standby (or, sequential) system one component is functioning and when it fails it is replaced immediately by another component, which is not subject to failure until it is switched on ("cold" standby). Figure 2.26 shows such a system with N-1 standby units. The switch is assumed to be perfect.

Let τ_i be the failure time of the i^{th} component with failure density $f_i(\tau)$. Since the components are operated sequentially (i.e., the first operates until $t=\tau_1$, then the second from $t=\tau_1$ to $t = \tau_1+\tau_2$ etc.), the system fails at time $T = \sum_{i=1}^{N} \tau_i$. Clearly T is a random variable and its special feature is that it is the sum of N independent random variables. The density of T can be readily found with the use of the convolution theorem.

Convolution Theorem. Given two independent continuous random variables T_1 and T_2 with density functions $f_1(t)$ and $f_2(t)$ respectively, the density function f(t) of their sum $T = T_1 + T_2$ is the convolution of their densities $f_1(t)$ and $f_2(2)$, i.e.,

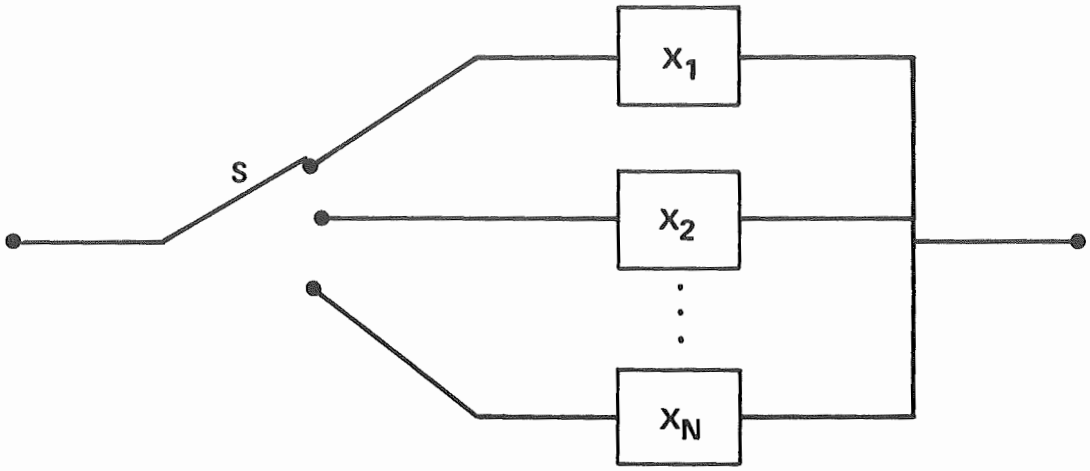


Figure 2.26. Standby System.

$$f(t) = f_1(t) * f_2(t) \equiv \int_{-\infty}^{\infty} f_1(x) f_2(t-x)\,dx \ . \qquad (2.115)$$

The proof may be found in any probability book.[1,2,3]

In the case of positive random variables the integral is defined from 0 to ∞. The convolution theorem from Laplace Transform theory is now invoked, i.e.,

$$\tilde{f}(s) = LT\ \{f_1(t) * f_2(t)\} = \tilde{f}_1(s)\tilde{f}_2(s) \ . \qquad (2.116)$$

Thus, the LT of the convolution of two functions is simply the product of their transforms.

Returning to the standby system we see that repeated application of the convolution theorem yields the LT for the density function of the time-to-failure of the system as the product of the LT of the failure densities of the components, that is,

$$\tilde{f}(s) = \prod_{i=1}^{N} \tilde{f}_i(s) \ . \qquad (2.117)$$

Then the reliability of the system is

$$R(t) = 1 - \int_0^t f(x)\,dx \ .$$

Let us apply the previous results in the case of N identical exponential components, i.e.,

$$f_i(t) = \lambda e^{-\lambda t} \ , \quad i=1,2,\ldots,N \ .$$

Then $\tilde{f}_i(s) = \frac{\lambda}{s+\lambda}$ and

$$\tilde{f}(s) = \frac{\lambda^N}{(s+\lambda)^N} \ . \qquad (2.118)$$

Inverting the transform we get

$$f(t) = \frac{\lambda^N t^{N-1}}{(N-1)!} e^{-\lambda t} \quad \text{(gamma density)} \tag{2.119}$$

and the reliability of the system is

$$R(t) = e^{-\lambda t} \sum_{k=0}^{N-1} \frac{(\lambda t)^k}{k!} \tag{2.120}$$

which is the well-known Poisson distribution. The MTTF is readily found to be (Equation (2.95))

$$m = -\left.\frac{d\tilde{f}(s)}{ds}\right|_{s=0} = \frac{N}{\lambda} \quad . \tag{2.121}$$

Another example involves two components with different failure rates λ_1 and λ_2. Then

$$\tilde{f}(s) = \frac{\lambda_1 \lambda_2}{(s+\lambda_1)(s+\lambda_2)}$$

and

$$f(t) = \frac{\lambda_1 \lambda_2}{\lambda_1 - \lambda_2} (e^{-\lambda_2 t} - e^{-\lambda_1 t}) \quad .$$

The reliability is

$$R(t) = \frac{\lambda_1 e^{-\lambda_2 t} - \lambda_2 e^{-\lambda_1 t}}{\lambda_1 - \lambda_2} \tag{2.122}$$

and the MTTF

$$m = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} \quad . \tag{2.123}$$

For N dissimilar components the MTTF is

$$m = \sum_{k=1}^{N} \frac{1}{\lambda_k} \quad . \tag{2.124}$$

An example which shows how the method presented above may be used in more general cases, involves a series system with M identical exponential elements for which there are N standby similar units. To find the reliability of the system we must determine the failure density of the series system first. Since

$$R_{ser}(t) = e^{-M\lambda t}$$

the failure density is

$$f_{ser}(t) = -\frac{dR_{ser}(t)}{dt} = M\lambda e^{-M\lambda t} \quad .$$

With N standby units the system is allowed to fail N times, therefore the LT of its failure density is

$$\tilde{f}(s) = [\tilde{f}_{ser}(s)]^N = \left[\frac{M\lambda}{s+M\lambda}\right]^N \tag{2.125}$$

hence

$$f(t) = \frac{(M\lambda)^N t^{N-1}}{(N-1)!} e^{-M\lambda t} \tag{2.126}$$

and the reliability of the system is

$$R(t) = e^{-M\lambda t} \sum_{k=0}^{N} \frac{(M\lambda t)^k}{k!} \tag{2.127}$$

which is the Poisson distribution again with rate of occurrence of events $M\lambda$. The MTTF is

$$m = \frac{N+1}{M\lambda} \quad . \tag{2.128}$$

2.C.6. Dependent Failures

All the previous models assumed independent failures, i.e., the failure of any element was not influenced by the failures of the other elements. There are cases, however, where such independence cannot be assumed.

In the study of the standby system we assumed that the units on standby were immune to failure. It is more realistic to assume that there is a finite probability that these units may fail. Consider, for example, one operating unit with failure density $f_1(t)$ and one standby unit with failure density $f_s(t)$. When the standby element is put on-line its failure density changes to $f_2(t)$. The convolution theorem can no longer be used to calculate the reliability of the system, because there is no independence of failures any more. We could use joint probability densities to attack the problem,[1,2,3,8] however, the following method helps to a better understanding of the sequence of events.

The system will perform its task in the interval (0,t) in either of the two mutually exclusive ways:

(i) Unit 1 does not fail in (0,t), or

(ii) Unit 1 fails in $(\tau,\tau+d\tau)$, where $0 < \tau < t$, unit 2 does not fail in $(0,\tau)$ while on standby and it operates successfully from τ to t.

The probability of the first case is simply the reliability of unit 1, i.e., $R_1(t) = 1 - \int_0^t f_1(x)\,dx$. The probability of the second case is the product of probability of the described events, i.e.,

$$\int_0^t [f_1(\tau)\,d\tau] \cdot [R_s(\tau)]\,[R_2(t-\tau)]$$

where

$$R_s(\tau) = 1 - \int_0^{\tau} f_s(x)\,dx$$

$$R_2(t-\tau) = 1 - \int_0^{t-\tau} f_2(x)\,dx$$

and we integrated over τ to cover all possibilities of τ in $(0,t)$.

Therefore, the reliability of the system is

$$R(t) = R_1(t) + \int_0^{\tau} f_1(\tau) R_s(\tau) R_2(t-\tau)\,d\tau \quad . \tag{2.129}$$

In the special case of exponential failure laws we have

$$f_1(t) = \lambda_1 e^{-\lambda_1 t}$$

$$f_s(t) = \lambda_s e^{-\lambda_s t}$$

$$f_2(t) = \lambda_2 e^{-\lambda_2 t}$$

and

$$R_1(t) = e^{-\lambda_1 t}$$

$$R_s(\tau) = e^{-\lambda_s \tau}$$

$$R_2(t-\tau) = e^{-\lambda_2 (t-\tau)}$$

therefore

$$R(t) = e^{-\lambda_1 t} + \int_0^t \lambda_1 e^{-\lambda_1 \tau}\, e^{-\lambda_s \tau}\, e^{-\lambda_2 (t-\tau)}\,d\tau =$$

$$= e^{-\lambda_1 t} + \frac{\lambda_1}{\lambda_1 + \lambda_s - \lambda_2} \left[e^{-\lambda_2 t} - e^{-(\lambda_1 + \lambda_s) t} \right] . \tag{2.130}$$

If $\lambda_1 \equiv \lambda_2 \equiv \lambda$ (the units have the same on-line failure rates) the reliability reduces to

$$R(t) = e^{-\lambda t} + \frac{\lambda}{\lambda_s}\left[e^{-\lambda t} - e^{-(\lambda+\lambda_s)t}\right]$$

or

$$R(t) = e^{-\lambda t}\left[1 + \frac{\lambda}{\lambda_s}\left(1 - e^{-\lambda_s t}\right)\right] \quad . \tag{2.131}$$

Observe that for $\lambda_s \equiv 0$ the above expression reduces to

$$R(t) = e^{-\lambda t}(1+\lambda t) \tag{2.132}$$

which is the known result for cold standby (L'Hospital's rule was used), Equation (2.117). If $\lambda_s \equiv \lambda$ the reliability becomes

$$R(t) = 2e^{-\lambda t} - e^{-2\lambda t}$$

which is the result for a parallel system, Equation (2.109).

A generalization of the above result is presented in Ref. 67. There is a total of M+N identical units of which M are required for successful system performance and N are on standby. The failure rate of the on-line units is λ and of the off-line units is λ_s. Defining for convenience the quantities

$$\alpha = M\frac{\lambda}{\lambda_s}$$

$$\beta = 1 - e^{-\lambda_s t}$$

and

$$R_u = e^{-\lambda t}$$

the reliability of the system is given by the following two equivalent expressions

$$R = \sum_{n=0}^{N} \binom{n+\alpha-1}{n} \beta^n (1-\beta)^{\alpha} \tag{2.133}$$

$$R = \sum_{n=0}^{N} \binom{\alpha+N}{n} \beta^n (1-\beta)^{\alpha+N-n} \quad . \tag{2.134}$$

Notice that α is not in general an integer. However, the binomial coefficient $\binom{x}{r}$ with x noninteger and r a positive integer may be defined as[1]

$$\binom{x}{r} = \frac{x(x-1)\ldots(x-r+1)}{r!}$$

or, in terms of gamma functions,

$$\binom{n+\alpha-1}{n} = \frac{\Gamma(n+\alpha)}{n!\,\Gamma(\alpha)} \quad .$$

The above distribution is called the Poisson-binomial distribution. When $\lambda_s \equiv 0$ it reduces to the Poisson distribution (M elements in series and N on "cold" standby), while when $\lambda_s \equiv \lambda$ it reduces to the binomial distribution (M-out-of-M+N system).

2.C.7. Imperfect Switching

There are many complexities which may be introduced to make the previous models more realistic. The methods presented though, are quite general and powerful so that only slight modifications will be needed to account for any additional features of the systems.

As an example, consider a standby system with one unit on-line (failure rate λ_1) and one on standby (on-line failure rate λ_2, cold standby). This problem was solved in 2.C.5 under the assumption of perfect switching and the reliability was found to be

$$R_{ps}(t) = \frac{\lambda_1 e^{-\lambda_2 t} - \lambda_2 e^{-\lambda_1 t}}{\lambda_1 - \lambda_2} \quad . \tag{2.122}$$

Suppose that there is a constant probability R_{sw} that the switching will be carried out properly. If the switch fails the reliability of the system is just the reliability of the on-line unit, i.e.,

$$R_{fs}(t) = e^{-\lambda_1 t} \quad .$$

Then in the presence of an imperfect switch the reliability of the system is

$$R(t) = (1-R_{sw})\, R_{fs}(t) + R_{sw} R_{ps}(t)$$

which yields

$$R(t) = e^{-\lambda_1 t} + \frac{R_{sw}\lambda_1}{\lambda_1 - \lambda_2} (e^{-\lambda_2 t} - e^{-\lambda_1 t}) \quad . \tag{2.135}$$

(Observe that the second term in the sum is the increase of reliability due to the standby unit.)

The same result could have been obtained with the method of 2.C.6. in a straightforward manner. In Reference 67 this method is applied to determine the reliability of a system consisting of M on-line elements (all required) with N standby units. The failure rate of a unit on-line is λ and on standby is λ_s. The probability of a successful switching is R_{sw} (constant). Furthermore, the standby units must be started when they are put on-line and the probability of a successful starting of any unit is R_{st} (initially the M on-line units are switched on but they have not been started). Then the reliability of the system is

$$R(t) = \sum_{k=0}^{\min(M,N)} \binom{M}{k} Q_{st}^{k} \, R_{st}^{M-k} \sum_{n=0}^{N-k} \binom{n+\alpha-1}{n} \beta^{n}(1-\beta)^{\alpha} \times$$

$$\times \sum_{x=0}^{N-k-n} \binom{N}{x} Q_{stsw}^{x} \, R_{stsw}^{N-x} \qquad (2.136)$$

where

$$\alpha = M \frac{\lambda}{\lambda_s}$$

$$\beta = e^{-\lambda t}$$

$$Q_{st} = 1 - R_{st}$$

$$R_{stsw} = R_{st} R_{sw}$$

$$Q_{stsw} = 1 - R_{stsw}$$

and $\binom{M}{k} = 0$ if $k > M$.

Further examples and discussions of redundant systems may be found in References 6,7,8,9.

2.D. MAINTENANCE MODELS

2.D.1. Introduction

The failure distributions and redundant systems which are studied in the previous sections attempt to predict the probabilistic aspects of the performance of a system that was built to satisfy a specified requirement. The system is put into operation and the reliability function gives the probability of successful operation for a given period of time.

In the present chapter we consider the problems which arise when the system is subject to maintenance policies. These may simply consist of replacement or repair of failed units (off-schedule maintenance) or of regular inspection and repair of redundant units according to a predetermined plan (preventive maintenance).

The first subject to be treated is the off-schedule maintenance. New mathematical tools are needed for the study of the problem and they are presented in the following sections.

2.D.2. Renewal Theory

A detailed exposition of renewal theory can be found in Cox.[68] The fundamentals of the theory with applications to reliability engineering are also presented in References 5,8,69 and 70.

A renewal process is defined to be a sequence of independent, non-negative identically distributed random variables $T_1, T_2, T_3, \ldots$. To visualize a situation which conforms with this definition, assume that an item is placed into operation in a socket. Its failure distribution (which is also the distribution of inter-arrival times) is $\Phi(t)$ and its failure density $\phi(t)$. The unit fails at a time $t_1 = T_1$ and it is instantaneously replaced by an identical unit, which fails after time T_2, i.e. at time $t_2 = T_1 + T_2$, and it

is replaced etc. Then the sequence T_1, T_2,... forms a renewal process (see Fig. 2.27). It is important to notice that the time t is counted from the beginning of the process, while the time T_i is the time interval between the $(i-1)^{th}$ and i^{th} replacement (inter-arrival time). By definition

$$P(T_i \leq t) = \Phi(t)$$

and

$$P(t \leq T_i \leq t + dt) = \phi(t)dt$$

(when all the inter-arrival times have the same distribution $\Phi(t)$ the process is called an ordinary renewal process; in some cases the first time T_1 has a distribution $\Phi_1(t)$ which differs from the distribution $\Phi(t)$ of T_2, T_3,... . Then we talk about a modified renewal process).

Having $\Phi(t)$ we will attempt to make statements about the following quantities:

a) $P(T_1+T_2+...+T_n < t)$: probability that the time of the n^{th} replacement (renewal) is less than t.

b) $N(t)$: the number of renewals in the interval (0,t)

c) $W(t) \equiv E[N(t)]$: the average number of renewals in the interval (0,t) (renewal function), and

d) $w(t) \equiv \frac{dW(t)}{dt}$: renewal density with interpretation:

$w(t)\Delta t$ = probability that a renewal occurs in the interval $(t, t+\Delta t)$.

$w(t)$ is a probability density, like $\phi(t)$, but it should not be confused with the latter; $\phi(t)$ concerns the failure of a specific unit which is placed in the socket while $w(t)$ refers to any failure (and thus a renewal) occurring in the socket. Also the time scales are different; time in $\phi(t)$ is counted from the moment the unit is placed into the socket, while time in $w(t)$ is counted from the beginning of the renewal process.

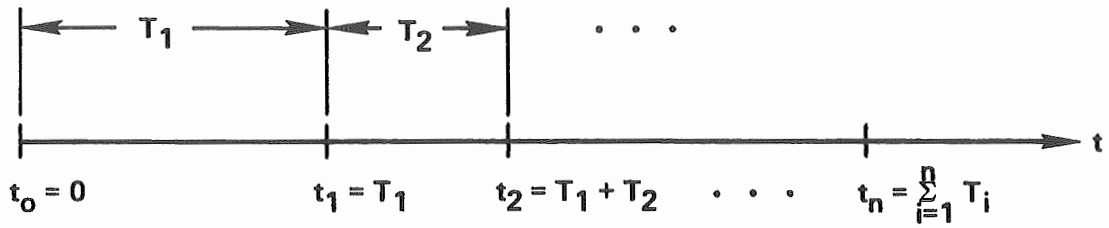


Figure 2.27. A Renewal Process.

To calculate the probability that $t_n = \sum_{i=1}^{n} T_i < t$ we invoke the convolution theorem (see Sec. 2.C.5). For a modified renewal process we define the following sequence of convolution integrals:

$$\begin{aligned}
\phi_1^{(1)}(t) &= \phi_1(t) \\
\phi^{(2)}(t) &= \phi^{(1)}(t) * \phi(t) = \int_0^t \phi^{(1)}(t-x)\,\phi(x)\,dx \\
&\vdots \\
\phi^{(n)}(t) &= \phi^{(n-1)} * \phi(t) = \int_0^t \phi^{(n-1)}(t-x)\,\phi(x)\,dx \quad . \qquad (2.137)
\end{aligned}$$

Then $\phi^{(n)}(t)$ is the density function of $t_n = \sum_{i=1}^{n} T_i$. By integrating over t we generate a sequence of convolution integrals for the distribution function, i.e.

$$\begin{aligned}
\Phi^{(1)}(t) &= \Phi_1(t) \\
\Phi^{(2)}(t) &= \Phi^{(1)}(t) * \Phi(t) = \int_0^t \Phi^{(1)}(t-x)\,\phi(x)\,dx \\
&\vdots \\
\Phi^{(n)}(t) &= \Phi^{(n-1)} * \Phi(t) = \int_0^t \Phi^{(n-1)}(t-x)\,\phi(x)\,dx \qquad (2.138)
\end{aligned}$$

Then

$$\Phi^{(n)}(t) = P[T_1 + T_2 + \ldots + T_n \leq t] \quad .$$

In the Laplace transform domain the expression for the density takes on a simple form; defining

$$\tilde{g}(s) = \int_0^\infty e^{-st}\, g(t)\,dt$$

we use the convolution theorem for Laplace Transforms to get

$$\tilde{\phi}^{(n)}(s) = \tilde{\phi}_1(s)\ [\tilde{\phi}(s)]^{n-1} \qquad (2.139)$$

Of course, for an ordinary renewal process we replace $\phi_1(t)$ with $\phi(t)$ and the results become simpler.

The distribution of the number of renewals N(t) is readily found from the previous results. Since the event N(t) = n is identical with the event $t_n \leq t \leq t_{n+1}$ we have

$$P[N(t) = n] = \Phi^{(n)}(t) - \Phi^{(n+1)}(t) \quad . \tag{2.140}$$

This result enables us to derive an equation for the renewal function W(t). By definition

$$\begin{aligned} W(t) &= \sum_{n=0}^{\infty} n[\Phi^{(n)}(t) - \Phi^{(n+1)}(t)] = \\ &= \sum_{n=1}^{\infty} \Phi^{(n)}(t) = \\ &= \Phi_1(t) + \sum_{n=2}^{\infty} \int_0^t \Phi^{(n-1)}(t-x)\,\phi(x)\,dx = \\ &= \Phi_1(t) + \int_0^t \sum_{n=1}^{\infty} \Phi^{(n)}(t-x)\,\phi(x)\,dx = \\ &= \Phi_1(t) + \int_0^t W(t-x)\,\phi(x)\,dx \quad . \end{aligned}$$

Therefore, we have the two equations

$$W(t) = \Phi_1(t) + \int_0^t W(t-x)\,\phi(x)\,dx \tag{2.141}$$

and, by differentiating,

$$w(t) = \frac{dW}{dt} = \phi_1(t) + \int_0^t w(t-x)\,\phi(x)\,dx \tag{2.142}$$

which are of the same form (renewal equation). We can interpret the terms of the equation as follows: $w(t)\Delta t$ is the probability that a failure (and thus

a renewal) occurs in the socket in the interval $(t,t+\Delta t)$. This event can occur in either of two mutually exclusive ways: a) the first unit that was placed in the socket fails in $(t,t+\Delta t)$, or b) a renewal took place at $(t-x)$, $0 < x < t$, and the unit that was placed in the socket than fails at $(t,t+\Delta t)$. These two events have probabilities equal to the first and second term of the right side of the renewal equation respectively.

To solve the renewal equation we use Laplace transforms and get

$$\tilde{w}(s) = \frac{\tilde{\phi}_1(s)}{1 - \tilde{\phi}(s)} \tag{2.143}$$

and

$$\tilde{W}(s) = \frac{\tilde{\Phi}_1(s)}{1 - \tilde{\phi}(s)} = \frac{\tilde{\phi}_1(s)}{s[1-\tilde{\phi}(s)]} \quad . \tag{2.144}$$

The solution may be obtained in closed form only in special cases. In the important case where the inter-arrival times are exponentially distributed, i.e.,

$$\phi_1(t) = \phi(t) = \lambda e^{-\lambda t}$$

inversion of Equations (2.143) and (2.144) yields

$$w(t) = \lambda \tag{2.145}$$

$$W(t) = \lambda t \tag{2.146}$$

which is expected, since the exponential model has no "memory". Furthermore, the density function for t_n is the convolution of n exponentials, thus

$$\phi^{(n)}(t) = \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} \quad \text{(gamma density)}$$

and

$$\Phi^{(n)}(t) = 1 - e^{-\lambda t}\left[1 + \lambda t + \ldots + \frac{(\lambda t)^{n-1}}{(n-1)!}\right]. \tag{2.147}$$

The probability of exactly n renewals in (0,t) is

$$P[N(t) = n] = \Phi^{(n)}(t) - \Phi^{(n+1)}(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

(Poisson distribution). (2.148)

Solutions in closed form can also be obtained for the more general gamma distribution, i.e.,

$$\phi_1(t) = \phi(t) = \frac{\lambda^r}{(r-1)!} t^{r-1} e^{-\lambda t} \quad , \quad r = 1,2,\ldots$$

$$\lambda > 0$$

In this case

$$P[N(t) = n] = \sum_{i=0}^{r-1} \frac{(\lambda t)^{nr+i}}{(nr+i)!} e^{-\lambda t} \tag{2.149}$$

$$W(t) = \frac{\lambda t}{r} + \frac{1}{r} \sum_{k=1}^{r-1} \frac{\theta^k}{1-\theta^k} [1-e^{-\lambda t(1-\theta^k)}] \tag{2.150}$$

where $\theta = \exp(\frac{2\pi i}{r})$, $i^2 = -1$.

In Reference 71 the renewal function is calculated in series form when the inter-arrival times are Weibull distributed. This solution was used in Reference 72 to derive graphs of the renewal density and function as functions of time for special values of the parameters of the Weibull distribution.

The previous results give the quantities of interest as functions of time. An asymptotic result is also of importance, namely

$$\lim_{t \to \infty} w(t) = \lim_{t \to \infty} \frac{W(t)}{t} = \frac{1}{m} \tag{2.151}$$

where m is the mean of $\Phi(t)$. This means that if the units are replaced as they fail, then the probability of a failure occurring at any time $(t,t+\Delta t)$ tends to a constant which is equal to the reciprocal MTTF of the unit (thus

in the modified process, the distribution $\Phi_1(t)$ does not affect the asymptotic behavior). The asymptotic value is reached after several MTTF's.

Observe that when $\phi_1(t) \equiv \phi(t) = \lambda e^{-\lambda t}$ we have

$$w(t) = \lambda = \frac{1}{m} \quad \text{for all } t, \tag{2.152}$$

that is, the asymptotic value is the exact solution. Of course, the hazard function is also $h(t) = \lambda$. This coincidence of numerical values has led to some confusion in the past and the renewal density has been treated as a renewal rate similar to the failure rate. This cannot be done, since the concepts are completely different: $h\Delta t$ is a conditional probability while $w\Delta t$ is a probability density.

2.D.3. Repair of a Single Unit

Renewal theory can be used directly in the study of failure and repair of components. Instead of replacements of units by new ones, we assume that when the component fails it undergoes repair which restores it to an "as-good-as-new" status. As renewal points we consider the times at which the unit enters the operating state either as new or after the completion of a repair. The inter-arrival times $T_1, T_2, \ldots$ are the sums of two independent random variables: the time of operation of the unit (or, the time spent in state 0) T_i^0 and the time it takes for the repair to be completed assuming it starts immediately after failure (or, the time spent in state 1) T_i^1, i.e.

$$T_i = T_i^0 + T_i^1 \quad . \tag{2.153}$$

If the failure density is $f(t)$ and the repair density $g(t)$ the convolution theorem (Equation (2.115)) gives the density of inter-arrival times as

$$\phi(t) = f(t) * g(t) = \int_0^t f(t-x)g(x)\,dx \tag{2.154}$$

or, in the Laplace transform domain,

$$\tilde{\phi}(x) = \tilde{f}(s)\tilde{g}(s) \quad . \tag{2.155}$$

The renewal density $w_r(t)$ then satisfies the renewal equation (ordinary renewal process, since at the beginning of the process the unit was new)

$$w_r(t) = \phi(t) + \int_0^t w_r(t-x)\phi(x)\,dx \tag{2.156}$$

and, in terms of Laplace Transforms,

$$\tilde{w}_r(s) = \frac{\tilde{\phi}(s)}{1-\tilde{\phi}(s)} = \frac{\tilde{f}(s)\tilde{g}(s)}{1-\tilde{f}(s)\tilde{g}(s)} \quad . \tag{2.157}$$

The expected number of repairs in (0,t) is then

$$W_r(t) = \int_0^t w_r(x)\,dx \tag{2.158}$$

with LT

$$\tilde{W}_r(s) = \frac{\tilde{f}(s)\tilde{g}(s)}{s[1-\tilde{f}(s)\tilde{g}(s)]} \quad . \tag{2.159}$$

In a similar manner we can calculate the probability of a failure in $(t,t+\Delta t)$ and the expected number of failures in (0,t). Now the renewal points are defined to be the points in time where the units fail. The inter-arrival times are again distributed according to $\phi(t)$ but now the process is a modified one; indeed, the unit starts as good as new and it fails at time T_1 which is distributed according to $\phi_1(t) = f(t)$ (the failure density of the unit). Thus, the renewal density $w_f(t)$ satisfies the equation

$$w_f(t) = f(t) + \int_0^t w_f(t-x)\phi(x)\,dx \tag{2.160}$$

hence

$$\tilde{w}_f(s) = \frac{\tilde{f}(s)}{1-\tilde{f}(s)\tilde{g}(s)} \quad . \tag{2.161}$$

The average number of failures in (0,t) is

$$W_f(t) = \int_0^t w_f(x)\,dx \tag{2.162}$$

and its transform

$$\tilde{W}_f(s) = \frac{\tilde{f}(s)}{s[1-\tilde{f}(s)\tilde{g}(s)]} \quad . \tag{2.163}$$

Using the convolution form of $\phi(t)$ we can write the renewal equation as

$$\begin{aligned} w_f(t) &= f(t) + \int_0^t w_f(x)\phi(t-x)\,dx \\ &= f(t) + \int_0^t dx\, w_f(x) \int_0^{t-x} g(t-x-\tau)f(\tau)\,d\tau \quad . \end{aligned} \tag{2.164}$$

The interpretation of this equation is as follows: a failure occurs in $(t,t+\Delta t)$ with probability $w_f(t)\Delta t$ which consists of the probabilities of two mutually exclusive events: the event that the unit fails for the first time in $(t,t+\Delta t)$ and the event that the unit failed in $(x,x+\Delta x)$, was repaired τ units of time later and it fails again in $(t,t+\Delta t)$. These two probabilities are the two terms on the right side of the equation.

The most important quantity in safety analysis is the availability $p(t)$ of the unit, which is defined as the probability that it is functioning at time t. Its complement is called the unavailability of the unit and they are related by

$$p(t) + q(t) = 1 \quad . \tag{2.165}$$

The initial condition is

$$p(0) = 1 \quad \text{(and, as a result, } q(0) = 0) \text{ .} \tag{2.166}$$

An integral equation can be written for p(t) as follows:

$$p(t) = 1 - F(t) + \int_0^t w_r(x)\ [1-F(t-x)]dx \tag{2.167}$$

and the interpretation is as usual (the first term on the right side is the probability of no failure in (0,t) and the second term is the probability of a repair at x and no failure from x to t).*

Taking Laplace transforms we get

$$\tilde{p}(s) = \frac{1-\tilde{f}(s)}{s[1-\tilde{f}(s)\tilde{g}(s)]} \text{ .} \tag{2.168}$$

It is shown in Barlow and Proschan[5] that the availability can also be calculated from

$$p(t) = 1 - [W_f(t) - W_r(t)] \text{ .} \tag{2.169}$$

Several asymptotic expressions find extensive use in applications. To find them we use the final value theorem of Laplace transform theory which states that

*Note: If $F(t) = 1 - e^{-\lambda t}$ and we multiply the equation by λ the following equation results

$$\lambda p(t) = \lambda e^{-\lambda t} + \int_0^t w_r(x)\ \lambda e^{-\lambda(t-x)}dx \text{ .}$$

It is easy to see that the right side is just the probability of a failure in $(t,t+\Delta t)$ (divided by Δt) which is w_f. Therefore we have derived the relation

$$w_f(t) = \lambda p(t) = \lambda[1-q(t)] \tag{2.170}$$

which is the same as Equation (12) in W. Vesely's paper, "A Time-Dependent Methodology for Fault Tree Evaluation," Nucl. Eng. and Design 13 (1970) 337-360. However, it holds only for constant failure rate and not in general; in addition, λ is a failure rate as conventionally defined and not as defined in the mentioned paper).

$$p_\infty \equiv \lim_{t \to \infty} p(t) = \lim_{s \to 0} [s\tilde{p}(s)] \quad . \tag{2.171}$$

Furthermore, for small values of s, Equations (2.94) and (2.95) lead to

$$\tilde{f}(s) \simeq 1 - ms \tag{2.172}$$

and

$$\tilde{g}(s) \simeq 1 - \tau s \tag{2.173}$$

where m is the mean of f(x) (called conventionally mean time between failures, MTBF) and τ is the mean time to repair, i.e.,

$$\tau \equiv \int_0^\infty t\, g(t)\, dt = MTTR \quad . \tag{2.174}$$

The asymptotic availability is then

$$p_\infty = \lim_{s \to 0} \left[\frac{1 - (1-ms)}{1-(1-ms)\ (1-\tau s)} \right] = \frac{m}{m+\tau} \quad , \tag{2.175}$$

Similarly, the asymptotic failure and repair renewal densities are

$$w_{f,\infty} = w_{r,\infty} = \frac{1}{m+\tau} \quad . \tag{2.176}$$

Applications

1. Exponential failure and exponential repair

Here $f(t) = \lambda e^{-\lambda t}$ and $g(t) = \mu e^{-\mu t}$ and their Laplace transforms are $\tilde{f}(s) = \lambda/s+\lambda$ and $\tilde{g}(s) = \mu/s+\mu$. Simple calculations yield

$$p(t) = \frac{\mu}{\mu+\lambda} + \frac{\lambda}{\mu+\lambda} e^{-(\lambda+\mu)t} \tag{2.177}$$

$$p_\infty = \frac{\mu}{\mu+\lambda} \tag{2.178}$$

$$w_f(t) = \frac{\lambda\mu}{\lambda+\mu} + \frac{\lambda^2}{\lambda+\mu} e^{-(\lambda+\mu)t} \tag{2.179}$$

$$w_f(t) = \frac{\lambda\mu}{\lambda+\mu} - \frac{\lambda\mu}{(\lambda+\mu)} e^{-(\lambda+\mu)t} \tag{2.180}$$

$$w_{f,\infty} = w_{r,\infty} = \frac{\lambda\mu}{\lambda+\mu} \quad . \tag{2.181}$$

The asymptotic results are reached after $\sim 3/\lambda+\mu$ units of time.

2. Exponential failure and fixed repair time

$$f(t) = \lambda e^{-\lambda t} \quad \text{and} \quad g(t) = \delta(t-\tau)$$

where τ is the constant time it takes for a repair to be completed, and $\delta(t-\tau)$ is the Dirac delta function.

The Laplace transforms are

$$\tilde{f}(s) = \frac{\lambda}{s+\lambda} \quad \text{and} \quad \tilde{g}(s) = e^{-s\tau}$$

hence

$$\tilde{p}(s) = \frac{1}{s+\lambda-\lambda e^{-s\tau}} =$$

$$= \frac{1}{(s+\lambda)\left[1 - \frac{\lambda}{s+\lambda} e^{-s\tau}\right]} =$$

$$= \frac{1}{s+\lambda} \sum_{n=0}^{\infty} \left(\frac{\lambda}{s+\lambda}\right)^n e^{-ns\tau} \quad . \tag{2.182}$$

The inverse transform is expressed in terms of the unit step function $U(t-n\tau)$ as follows

$$p(t) = \sum_{n=0}^{\infty} \frac{\lambda^n (t-n\tau)^n}{n!} e^{-\lambda(t-n\tau)} U(t-n\tau) \tag{2.183}$$

where

$$U(t-n\tau) = \begin{cases} 1 & \text{if} \quad t > n\tau \\ 0 & \text{if} \quad t < n\tau \end{cases} \quad . \tag{2.184}$$

The availability as a function of time is shown in Figure 2.28. In the first two time-intervals the availability is

$$p_1(t) = e^{-\lambda t} \quad , \qquad 0 < t < \tau \qquad \text{(reliability)}$$

$$p_2(t) = e^{-\lambda t} + \lambda(t-\tau)e^{-\lambda(t-\tau)} , \qquad \tau < t < 2\tau$$

etc.

The maximum value occurs in the second interval at

$$t_{max} = \tau + \frac{1-e^{-\lambda\tau}}{\lambda}$$

and it is

$$p_{max} = e^{-(1-e^{-\lambda\tau})} \quad .$$

The asumptotic availability is

$$p_\infty = \frac{1}{1+\lambda\tau} \tag{2.185}$$

and it is reached after $\sim 3\tau$.

The LT for the failure renewal density is

$$\tilde{w}_f(s) = \frac{\lambda}{s+\lambda-\lambda e^{-s\tau}}$$

and inverting as before, or using Equation (2.170), we get

$$w_f(t) = \lambda p(t) \ . \tag{2.186}$$

For the repair renewal density we have

$$\tilde{w}_r(s) = \frac{\lambda e^{-s\tau}}{s+\lambda-\lambda e^{-s\tau}}$$

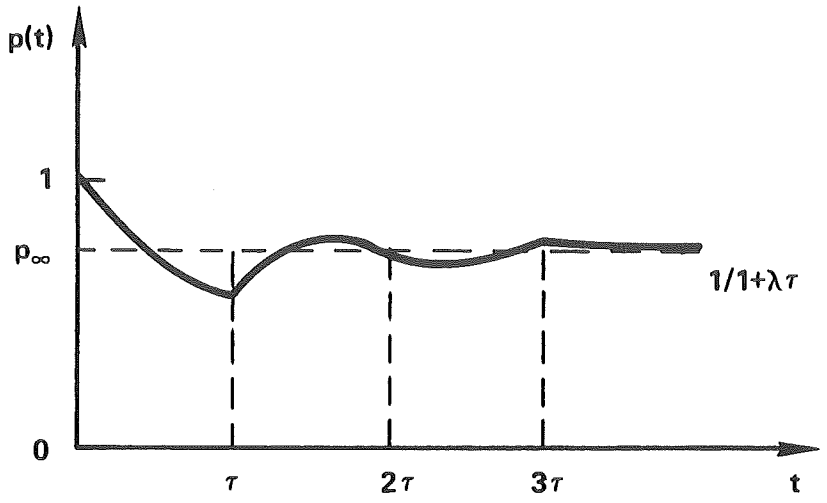


Figure 2.28. Availability as a Function of Time (Exponential Failure Distribution $F(t) = 1-e^{-\lambda t}$ and Constant Repair Time τ).

therefore

$$w_r(t) = \lambda p(t-\tau) \quad , \quad t > \tau$$
$$= 0 \qquad\qquad \tau < 0 \quad . \tag{2.187}$$

The asymptotic expressions are

$$w_{f,\infty} = w_{r,\infty} = \frac{\lambda}{1+\lambda\tau} \quad . \tag{2.188}$$

3. Exponential failure and gamma repair distribution

The description of repair by the gamma distribution is more realistic.[73,74] Here we assume that

$$g(t) = \mu^2 t\, e^{-\mu t}$$

thus

$$\tilde{g}(s) = \frac{\mu^2}{(s+\mu)^2} \quad .$$

To find the availability we have

$$\tilde{p}(s) = \frac{(s+\mu)^2}{s[s^2 + (2\mu+\lambda)s + 2\mu\lambda + \mu^2]} \quad . \tag{2.189}$$

The time-dependent availability may be found by inverting $\tilde{p}(s)$; this leads to expressions involving hyperbolic functions (with the condition that $\lambda > 4\mu$; if $\lambda < 4\mu$ trigonometric functions are used) which are quite complicated and of little use. The asymptotic availability though is readily found to be

$$p_\infty = \lim_{s \to 0} s\, \tilde{p}(s) = \frac{\mu}{\mu+2\lambda} \quad . \tag{2.190}$$

Furthermore

$$w_{f,\infty} = w_{r,\infty} = \frac{\lambda\mu}{\mu+2\lambda} \quad . \tag{2.191}$$

The time-dependent part of p(t) decays as $e^{-(\mu+\lambda/2)t}$, therefore the asymptotic expressions can be used after $\sim\frac{3}{\mu+\lambda/2}$ units of time.

4. Gamma failure and repair distributions

Now we have

$$f(t) = \lambda^2 t\, e^{-\lambda t} \quad \text{and} \quad g(t) = \mu^2 t\, e^{-\mu t}$$

thus

$$\tilde{f}(s) = \frac{\lambda^2}{(s+\lambda)^2} \quad \text{and} \quad \tilde{g}(s) = \frac{\mu^2}{(s+\mu)^2} \ .$$

Then

$$\tilde{p}(s) = \frac{(s+2\lambda)(s+\mu)^2}{s(s+\lambda+\mu)[s^2+(\lambda+\mu)s+2\lambda\mu]} \ . \tag{2.192}$$

The asymptotic availability is

$$p_\infty = \frac{\mu}{\lambda+\mu} \ . \tag{2.193}$$

The failure renewal density is given by

$$\tilde{w}_f(s) = \frac{\lambda^2(s+\mu)^2}{s(s+\lambda+\mu)[s^2+(\lambda+\mu)s+2\lambda\mu]} \tag{2.194}$$

and

$$w_{f,\infty} = \frac{\lambda\mu}{2(\lambda+\mu)} \ (= w_{r,\infty}) \ . \tag{2.195}$$

Since the failure rate of the gamma density is

$$\lambda(t) = h(t) = \frac{\lambda^2 t}{1+\lambda t} \quad \text{and} \quad \lim_{t\to\infty} \lambda(t) = \lambda \tag{2.196}$$

it is clear that the relation (Equation (2.170))

$$w_f(t) = \lambda(t)\, p(t)$$

does not hold here, as expected, since the failure distribution is not exponential.

A quantity which can serve as a measure of the performance of the unit is the downtime D(t) (References 5,75). It is defined as the total time the unit spends under repair in an interval (0,t). Similarly we can define the uptime U(t), which is the total time the units spends in the operating state. Therefore

$$D(t) + U(t) = t \tag{2.197}$$

The distribution of D(t) is given by complicated expressions and can be found in the references. However, the following asymptotic result is of interest

$$\lim_{t \to \infty} P\left[\frac{D(t) - q_\infty t}{\sqrt{\sigma_D^2 t}} \leq x\right] = N(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du \tag{2.198}$$

where

$$q_\infty = \frac{r}{m+r} \tag{2.199}$$

$$\sigma_D^2 = \frac{r^2\sigma_f^2 + m^2\sigma_r^2}{(r+m)^3} \tag{2.200}$$

and σ_f^2 and σ_r^2 are the variances of the failure and repair distributions respectively. In words, the above result states that for large t the downtime is normally distributed with mean the asymptotic unavailability times the time interval and variance σ_D^2. Notice that we only need to know the mean and the variance of the failure and repair distributions.

Finally, another quantity which also can serve as a measure of unit performance is the excess time.[76,77] It is defined as the total time B(τ) that the unit is under repair corresponding to τ operating units of time. Thus the

difference between D(t) and B(τ) is in the argument: t is the real time while τ is the time spent in state 0 only. Clearly

$$t = \tau + B(\tau) \quad . \tag{2.201}$$

The asymptotic behavior of B(τ) is

$$\lim_{t \to \infty} P\left[\frac{B(\tau) - m_B \tau}{\sqrt{\sigma_D^2 \tau}} \leq x\right] = N(x) \tag{2.202}$$

where

$$m_B = \frac{r}{m} \tag{2.203}$$

and

$$\sigma_D^2 = \frac{m\sigma_r^2 + r\,\sigma_f^2}{m^3} \quad . \tag{2.204}$$

2.D.4. Multiple-State Systems. Markov Approach

The preceding analysis referred to a single unit which could only be in two mutually exclusive states 0 (up) and 1 (down). For any failure and repair distributions we estimated the availability and other quantities of interest using renewal theory; this approach was possible because the regeneration points of the renewal process were readily identified.

Unfortunately this elegant method cannot be applied to more general situations, where the system may be in more than two states (e.g. units in series, parallel etc.). It is very difficult, if at all possible, to find the regeneration points. The use of Markov and Semi-Markov models makes it possible to obtain useful results in these cases.

The study begins with the identification of all the mutually exclusive states of the system. To make the discussion more concrete we will use as an example a system consisting of two units. We do not specify for the moment,

how they are logically interconnected. In addition there is one repair facility which restores a failed unit to an as-good-as-new status. The possible states of the system are the following:

0: both units are up

1: unit 1 is down and under repair, unit 2 is up

2: unit 2 is down and under repair, unit 1 is up

3: both units are down, unit 1 is under repair

4: both units are down, unit 2 is under repair .

These mutually exclusive states exhaust all possibilities. Observe that if two repair facilities were available, states 3 and 4 would be replaced by a single state: both units down and under repair.

The probability that the system will be in state i at time t is denoted as $P_i(t)$. To be able to write a system of equations relating these probabilities we define the transition probabilities as follows: the probability that the system will be in state j at $t+\Delta t$ given that it is in state i at t is $a_{ij}\Delta t$. The rates a_{ii} are defined as $a_{ii} = -\sum_{j\neq i} a_{ij}$ and $(1-a_{ii})\Delta t$ is the conditional probability that if the system is in state i at time t it will remain in that state in the next interval Δt. All the transition rates are assumed independent of time.

We can now write an equation for the change of $P_o(t)$ in Δt, i.e.,

$$P_o(t+\Delta t) = P_o(t)[1-(a_{01}+ a_{02})\Delta t] + P_1(t)a_{10}\Delta t + P_2(t)a_{20}\Delta t + o(\Delta t)$$

$$0,1,2,...$$

The first term on the right side is the probability that the system will remain in state 0, the second and third terms are the probabilities that a repair is completed in Δt on unit 1 or unit 2. These are the only terms which are of first order in Δt, terms of higher order are included in $o(\Delta t)$ (e.g.

transition from 0 to 3 or 4 requires the simultaneous failure of both units in Δt, the probability of which is of order $(\Delta t)^2$).

Dividing by Δt and letting $\Delta t \to 0$ yields

$$\frac{dP_o(t)}{dt} = -(a_{01} + a_{02})\, P_o(t) + a_{10}\, P_1(t) + a_{20} P_2(t) \quad .$$

Similarly we derive the equations

$$\frac{dP_1(t)}{dt} = a_{01}\, P_o(t) - (a_{10} + a_{13})\, P_1(t) + a_{41}\, P_4(t)$$

$$\frac{dP_2(t)}{dt} = a_{02}\, P_o(t) - (a_{20} + a_{24})\, P_2(t) + a_{32}\, P_3(t)$$

$$\frac{dP_3(t)}{dt} = a_{13}\, P_1(t) - a_{32}\, P_3(t)$$

$$\frac{dP_4(t)}{dt} = a_{24}\, P_2(t) - a_{41}\, P_4(t) \quad . \tag{2.205}$$

Figure 2.29 shows graphically the states and the transition rates.

The system can be written in compact form by defining the row vector

$$\underline{P}(t) \equiv (P_o(t),\ P_1(t),\ P_2(t),\ P_3(t),\ P_4(t)) \tag{2.206}$$

and the matrix

$$A \equiv \begin{bmatrix} -(a_{01}+a_{02}) & a_{01} & a_{02} & 0 & 0 \\ a_{10} & -(a_{10}+a_{13}) & 0 & a_{13} & 0 \\ a_{20} & 0 & -(a_{20}+a_{24}) & 0 & a_{24} \\ 0 & 0 & a_{32} & -a_{32} & 0 \\ 0 & a_{41} & 0 & 0 & -a_{41} \end{bmatrix} \tag{2.207}$$

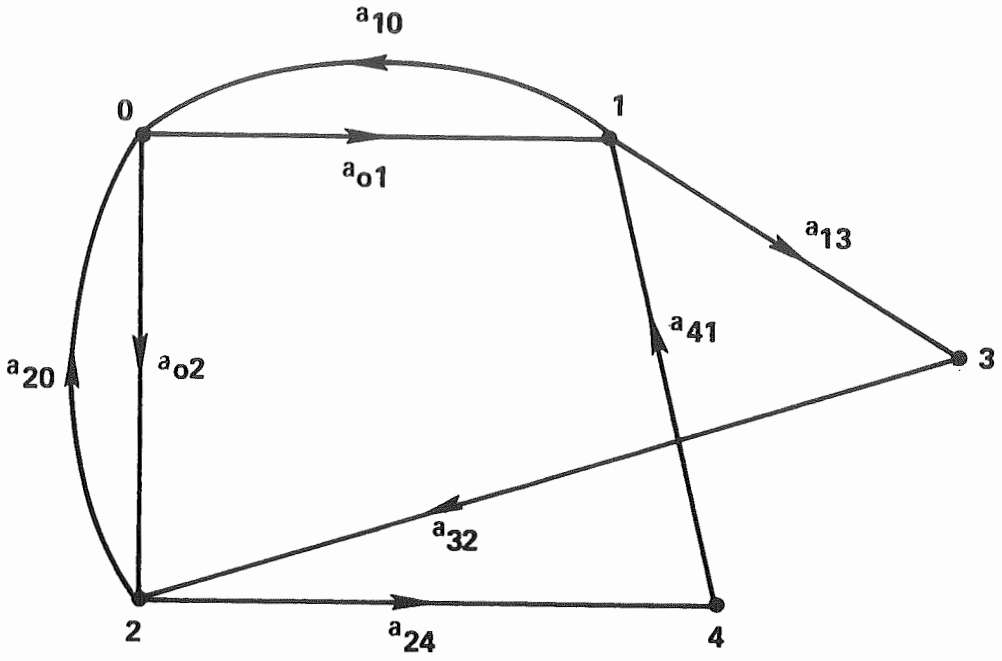


Figure 2.29. States and Transition Rates for a System With Two Dissimilar Units and One Repairman.

then we have

$$\frac{d\underline{P}(t)}{dt} = \underline{P}(t)\ A \tag{2.208}$$

and if some initial conditions $\underline{P}(0)$ are specified the system admits a unique solution.

The stochastic process $\underline{P}(t)$ has a very important feature, namely, it suffices to know one value of $\underline{P}(t)$ at a certain time (e.g. $\underline{P}(0)$) to determine completely the future (and past) behavior of the process. This is a direct consequence of the assumed constancy in time of the transition matrix A. The future depends only on the present and not on the history of the system (i.e. there is no memory). It is clear that such a model can be applied if and only if the failure and repair distributions are exponential, since this is the only distribution with lack of memory. The process we described is called Markov and has been studied extensively.[1,3,70,81] The theory and its application to reliability problems is presented in References 5 and 8 and the book by Sandler[78] is devoted exclusively to the Markov approach of reliability problems.

The system of differential equations (2.208) describes the general two-unit system with one repairman available. Knowledge of the logical interconnection of the units permits us to calculate the elements of the transition matrix A and to proceed to estimate various reliability quantities. Thus we distinguish the following cases:

Series system

We assume that the failure rates of the units are λ_1 and λ_2 and the repair rate is μ. A further assumption is that the units work independently and failure of one does not affect the performance of the other. Such a case arises, for instance, when two engines are connected in series and when one

fails the other continues to work. The situation is different, however, if two resistors are connected in series, then failure of one interrupts the current, therefore the other cannot fail. In this case states 3 and 4 are impossible. Then it is easy to see that

$$a_{01} = a_{24} = \lambda_1$$

$$a_{02} = a_{13} = \lambda_2$$

$$a_{10} = a_{20} = a_{32} = a_{41} = \mu \quad .$$

Since both units are required for the system to function, the only acceptable state is state 0. Therefore, the availability of the system is

$$p(t) = P_o(t) \quad .$$

Parallel system

Now only one unit is needed for successful system performance, therefore the availability is

$$p(t) = P_o(t) + P_1(t) + P_2(t) \quad .$$

Standby system

Unit 2 is initially on standby with zero failure rate and λ_2 on-line failure rate. In this case we must also re-examine the definition of the states. We assume that if both units are up it is unit 1 that is always on line (all the switchings are instantaneous). If this is not the case, state 0 should be split into two other states: unit 1 on-line, unit 2 on standby and unit 2 on-line, unit 1 on standby. Except for this nothing essential changes. We always assume that the standby unit cannot fail.

Then the transition rates are

$$a_{01} = a_{24} = \lambda_1$$

$$a_{02} = 0$$

$$a_{13} = \lambda_2$$

$$a_{10} = a_{20} = a_{32} = a_{41} = \mu.$$

The system is up if either or both units are up, therefore its availability is determined by

$$p(t) = P_0(t) + P_1(t) + P_2(t) \quad .$$

These three examples illustrate how the method is adjusted to cover the special features of a particular problem. The logical interconnection of the units which comprise the system determines the elements of the transition matrix in terms of the failure and repair rates and which states are the working states of the system so that the availability function can be calculated.

A drawback of the Markov model in applications is the large number of states that even a simple system can have. A great reduction in the number of states occurs, however, if all the units are identical. The two-unit system with one repairman can be in one of the following three states:

0: both units up

1: one unit up and one under repair

2: one unit down and the other under repair .

(The number of states is reduced by two in this example, but the reduction is must greater in more complex systems.)

The transition rate matrix A has the following form when the elements are in series or parallel

$$A = \begin{pmatrix} -2\lambda & 2\lambda & 0 \\ \mu & -(\lambda+\mu) & \lambda \\ 0 & \mu & -\mu \end{pmatrix} \tag{2.209}$$

with working state for the series system the state 0 and for the parallel system states 0 and 1. The graph is shown in Figure 2.30.

For the standby system the transition rate matrix is

$$A = \begin{pmatrix} -\lambda & \lambda & 0 \\ \mu & -(\lambda+\mu) & \lambda \\ 0 & \mu & -\mu \end{pmatrix} \tag{2.210}$$

and again the working states are 0 and 1.

Thus far the analysis concerned the estimation of the availability function. In many cases it is desirable to have expressions for the reliability of the system, its mean time to failure (MTTF) and the mean up (or down) time MUT. In the case of a single unit with repair the reliability and MTFF were not affected by the repair process, for redundant systems however they are improved. In general the reliability is less than or equal to the availability.

In order to calculate the above quantities a further discussion of the states of the system is necessary. A characteristic of the Markov model with the previously defined transition rate matrix A was that the system could visit and leave all the states or, in other words, in whatever state the system was initially, it would visit all other states after a finite time. These states are said to communicate. If a state cannot be left once it is entered it is called an absorbing state. The state i is absorbing if and only if $a_{ii} = -\sum_{j \neq i} a_{ij} = 0$. In our example if there were no repair, states 3 and 4 would be identical (both units down) and absorbing. Furthermore the states would not communicate anymore since it would be possible to go from 0 to 1 to 2 to 3 (or 4) but not back, i.e. states 0,1 and 2 cannot be re-entered once left.

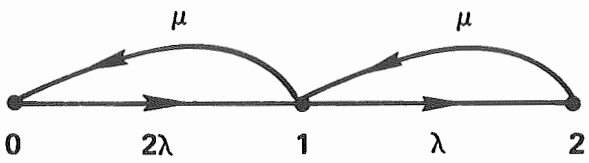


Figure 2.30. States and Transition Rates for a System With Two Identical Units in Series or Parallel and One Repairman.

For each logical interconnection of the units we have identified the working states and the failed states. Assuming that initially the system starts in a working state its reliability at time t is the probability that none of the failed states has been entered. This suggests that we change the failed states into absorbing ones and then calculate the probability that the system is in one of the working states. This probability is the system reliability, since we are sure that the system cannot leave the failed states. For example, consider the system of two identical units with one on standby. The unacceptable state is 2, therefore the transition rate matrix is modified to be

$$A = \begin{pmatrix} -\lambda & \lambda & 0 \\ \mu & -(\lambda+\mu) & \lambda \\ 0 & 0 & 0 \end{pmatrix} \tag{2.211}$$

and the system reliability is

$$R(t) = P_0(t) + P_1(t) \ .$$

The MTFF is given by

$$MTFF = \int_0^\infty R(t)\, dt \ .$$

2.D.5. Solution of the Markov System

The solution of the system

$$\frac{d\underline{P}(t)}{dt} = \underline{P}(t)A$$
$$\underline{P}(0) = \underline{C} \qquad \left(\sum_{i=0}^{n} C_i = 1 \right) \tag{2.212}$$

is discussed in many textbooks. A detailed exposition of the various methods can be found in Reference 79. The solution can be obtained by taking Laplace transforms, i.e.

$$s\,\tilde{\underline{P}}(s) - \underline{C} = \tilde{\underline{P}}(s)A$$

hence

$$\underline{P}(s) = \underline{C}(sI-A)^{-1} \quad , \qquad s \neq \omega_0, \omega_1, \ldots \omega_n \tag{2.213}$$

where I is the matrix with unities in the diagonal and all other elements zero. The Laplace transform variable s must not be equal to any of the (possibly multiple) eigenvalues ω_i of A for the inverse matrix to exist. The inverse is of the form

$$(sI-A)^{-1} = \frac{B(s)}{\det(sI-A)^{-1}} \tag{2.214}$$

where det $\equiv$ determinant and B(s) is a matrix whose elements are polynomials in s of degree at most (n). The inverse Laplace transform may be found by any of the standard methods (e.g. using tables, partial fraction expansions, etc.). If there are k distinct eigenvalues of A with multiplicities m_i $\left(\sum_{i=1}^{k} m_i = n+1\right)$ the inverse transform will be of the form

$$\underline{P}(t) = \underline{C} \sum_{i=1}^{k} \sum_{j=0}^{m_i-1} Y_{ij} \frac{t^j}{j!} e^{\omega_i t}$$

where Y_{ij} are constant matrices. Once the probability vector $\underline{P}(t)$ has been obtained the availability function will be the sum of the components of $\underline{P}(t)$ corresponding to the working states of the system. We can write an expression for the availability by defining a (n+1)-dimensional vector $\underline{V}$ whose i^{th} element is unity or zero depending on whether the i^{th} state is acceptable or not. Then the availability is

$$p(t) = \sum_{i=0}^{n} P_i(t) V_i \equiv \langle \underline{P}(t), \underline{V} \rangle \, . \tag{2.215}$$

It is clear that the eigenvalues of A play an important role. Due to its special structure $\left(\text{i.e., } a_{ii} = -\sum_{\substack{j=0 \\ j\neq i}}^{n} a_{ij} \text{ and } a_{ij} \geq 0\right)$ the following properties of the eigenvalues can be proven:

i. Zero is always an eigenvalue and the corresponding eigenvector has non-negative elements.

ii. All other eigenvalues have negative real parts.

From these properties it follows that a steady state solution to the system always exists. If there are no absorbing states, the asymptotic solution is found from

$$\begin{aligned} &\underline{\Pi}\, A = 0 \\ &\sum_{i=0}^{n} \Pi_i = 1 \end{aligned} \tag{2.216}$$

and it is independent of the initial vector $\underline{C}$.

The elements Π_i of the vector $\underline{\Pi}$ can be expressed in terms of the determinants of A using Cramer's rule. Define D_i to be the determinant of the matrix formed by striking out the i^{th} row and the i^{th} column of -A, then it is easily seen that

$$\Pi_i = \frac{D_i}{\sum_{i=0}^{n} D_i} \quad . \tag{2.217}$$

If there are k transient and n+1-k absorbing states we write A as

$$A = \begin{pmatrix} A_1 & A_2 \\ 0 & 0 \end{pmatrix} \tag{2.218}$$

where A_1 is a k×k matrix consisting of the transient states. We also split the probability vector $\underline{P}(t)$ into two parts

$$\underline{P}(t) = (\underline{P}_t, \underline{P}_a) \tag{2.219}$$

where

$$\underline{P}_t = (P_0, P_1, \ldots, P_{k-1})$$

$$\underline{P}_a = (P_k, P_{k+1}, \ldots, P_n) \quad .$$

The steady state solution for $\underline{P}_t$ is $\underline{\Pi}_t = \underline{0}$, that is, the system will enter the absorbing states and will stay there regardless of the initial conditions.

We now turn our attention to the estimation of reliability and various mean times of interest (in addition to the listed references see also Reference 80). As indicated in the previous section all the failed states are converted into absorbing ones and the working states are lumped in A_1. (Equation (2.218)) The initial probability vector is also written as $\underline{C} = (\underline{C}_t, \underline{c}_a)$ where $\underline{C}_t = (C_0, C_1, \ldots, C_{k-1})$ and $\underline{C}_a = (C_k, \ldots, C_n)$. We assume that the system is in one of the working states at time zero, i.e.

$$\sum_{i=0}^{k-1} C_i = 1$$

and, as a result, $C_i = 0$, $i = k, k+1, \ldots n$. The original system is now written as

$$\begin{aligned} \frac{d\underline{P}_t(t)}{dt} &= \underline{P}_t(t) A_1 \\ \frac{d\underline{P}_a(t)}{dt} &= \underline{P}_t(t) A_2 \\ \underline{P}_t(0) &= \underline{C}_t \\ \underline{P}_a(0) &= \underline{0} \quad . \end{aligned} \tag{2.220}$$

As before the solution is in the Laplace Transform plane

$$\underline{\tilde{P}}_t(s) = \underline{C}_t(sI-A_1)^{-1} \tag{2.221}$$

$$s\underline{\tilde{P}}_a(s) = \underline{C}_t(sI-A_1)^{-1}A_2 \quad . \tag{2.222}$$

Defining a k-dimensional vector $\underline{w}_k$ whose elements are unities we have for the reliability

$$R(t) = \langle \underline{P}_t(t), \underline{w}_k \rangle \tag{2.223}$$

which expresses the fact that the reliability is the sum of the probabilities that the system is in any one of the working states. Since $\underline{\tilde{P}}_a(s)$ is the LT of the probability vector for the failed states, $s\underline{\tilde{P}}_a(s)$ is its derivative, therefore the failure density of the system is

$$\tilde{f}(s) = \langle s\underline{\tilde{P}}_a(s), \underline{w}_{n+1-k} \rangle \tag{2.224}$$

and using the fact that the row sums of A are zero, i.e. $A_1\underline{w}_k + A_2\underline{w}_{n+1-k} = \underline{0}$, we derive the expression

$$\tilde{f}(s) = -\langle \underline{C}_t(sI-A_1)^{-1} A_1, \underline{w}_k \rangle \quad . \tag{2.225}$$

Of course, the reliability may also be found from

$$R(t) = 1 - \int_0^t f(t)\,dt \quad .$$

The MTTF can be determined from

$$MTTF = \int_0^\infty R(t)\,dt$$

or, from (Equation (2.95))

$$\text{MTTF} = \lim_{s \to 0} \left(- \frac{d}{ds} \tilde{f}(s) \right) = < \underline{c}_t (-A_1)^{-1}, \underline{w}_k > \qquad (2.226)$$

which leads to interpreting the elements of $(-A_1)^{-1}$, say m_{ij}, as the expected time that the system spends in the j^{th} state before absorption if it starts in the i^{th} state. The previous formula shows that these times are weighted by the probabilities of starting in any of the working states and then they are summed. The mean times m_{ij} can be determined in terms of determinants of $-A_1$. Thus we define M_{ij} to be the determinant of the matrix resulting from striking out the i^{th} row and j^{th} column of $-A_1$ and we have

$$m_{ij} = \frac{(-1)^{i+j} M_{ji}}{\det(-A_1)} \quad . \qquad (2.227)$$

If state 0 is when all the units are up, we define the system MTFF (mean-time-to-first-failure) as the average time the system will spend in the working states before failure. In terms of the previous quantities

$$\text{MTFF} = \sum_{j=0}^{k-1} m_{oj} \quad . \qquad (2.228)$$

The matrix formulation presented here is a useful description of the Markov model, however it should not be implied that every Markov system must be solved in this manner. Very often the system of equations is simple enough (three or four equations) to allow direct solution without reducing it to matrix form. The matrix approach will be particularly useful in handling systems with many states.

Applications

We present here several applications involving identical units; this case is important in practice and many such problems are treated in Sandler.[78] The

treatment of the problem as a birth-and-death process is presented in Barlow and Proschan.[5]

Standby System

We solve the two-unit system completely to show an application of the previous results. The states of the system are

0: both units up, one on-line, one on standby

1: one unit on-line, one under repair

2: both down and one under repair (i.e., there is only one repairman).

The matrix A is

$$A = \begin{array}{c} \begin{matrix} 0 & 1 & 2 \end{matrix} \leftarrow \text{states} \\ \downarrow \\ \begin{pmatrix} -\lambda & \lambda & 0 \\ \mu & -(\lambda+\mu) & \lambda \\ 0 & \mu & -\mu \end{pmatrix} \begin{matrix} 0 \\ 1 \\ 2 \end{matrix} \end{array} \quad . \tag{2.210}$$

We assume that initially both units are up, i.e.

$$\underline{c} = (1,0,0) \quad .$$

First we find $(sI-A)^{-1}$:

$$(sI-A)^{-1} = \begin{pmatrix} s+\lambda & -\lambda & 0 \\ -\mu & s+\lambda+\mu & -\lambda \\ 0 & -\mu & s+\mu \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} s^2+s(2\lambda+\mu)+\mu^2 & (s+\mu)\lambda & \lambda^2 \\ (s+\mu)\mu & (s+\lambda)(s+\mu) & (s+\lambda)\lambda \\ \mu^2 & (s+\lambda)\mu & s^2+s(2\lambda+\mu)+\lambda^2 \end{pmatrix} \times \frac{1}{\det(sI-A)} \tag{2.229}$$

and

$$\det(sI-A) = s(s-\omega_1)(s-\omega_2)$$

where

$$\omega_1 = -(\lambda+\mu) - \sqrt{\lambda\mu}$$

$$\omega_2 = -(\lambda+\mu) + \sqrt{\lambda\mu}$$

(0, ω_1, and ω_2 are the eigenvalues of A).

The Laplace transform of the probability vector is then, from Equations (2.213) and (2.229),

$$\underset{\sim}{\tilde{P}}(s) = \underline{C}(sI-A)^{-1} = (s^2 + s(2\lambda+\mu) + \mu^2, \quad (s+\mu)\lambda, \quad \lambda^2) \times \frac{1}{\det(sI-A)} \quad .$$

The transform of the availability is

$$\tilde{p}(s) = \tilde{P}_0(s) + \tilde{P}_1(s)$$

but it is easier to work with the unavailability

$$\tilde{q}(s) = 1 - \tilde{p}(s) = \tilde{P}_3(s) = \frac{\lambda^2}{s(s-\omega_1)(s-\omega_2)}$$

hence

$$q(t) = \lambda^2 \left(\frac{1}{\omega_1\omega_2} + \frac{1}{\omega_1(\omega_1-\omega_2)} e^{\omega_1 t} + \frac{1}{\omega_2(\omega_2-\omega_1)} e^{\omega_2 t} \right). \qquad (2.230)$$

The steady-state unavailability is then

$$q_\infty = \frac{\lambda^2}{\lambda^2+\lambda\mu+\mu^2} = \Pi_3 \qquad (2.231)$$

hence

$$p_\infty = \frac{\mu^2+\lambda\mu}{\lambda^2+\lambda\mu+\mu^2} = \Pi_1 + \Pi_2 \quad . \qquad (2.232)$$

To find the reliability of the system we convert the failed state 2 into an absorbing state and we have, (Equation (2.218)),

$$\begin{pmatrix} A_1 & A_2 \\ 0 & 0 \end{pmatrix} = \left(\begin{array}{cc|c} -\lambda & \lambda & 0 \\ \mu & -(\lambda+\mu) & \lambda \\ \hline 0 & 0 & 0 \end{array}\right) .$$

Again we determine the inverse matrix

$$(sI - A_1)^{-1} = \begin{pmatrix} s+\lambda & -\lambda \\ -\mu & s+\lambda+\mu \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} s+\lambda+\mu & \lambda \\ \mu & s+\lambda \end{pmatrix} \times \frac{1}{\det(sI-A_1)}$$

where

$$\det(sI-A_1) = s^2 + s(2\lambda+\mu) + \lambda^2 = (s-s_1)(s-s_2)$$

and

$$s_1 = \frac{-(2\lambda+\mu) + \sqrt{\mu^2+4\lambda\mu}}{2} < 0$$

$$s_2 = \frac{-(2\lambda+\mu) - \sqrt{\mu^2+4\lambda\mu}}{2} < 0 .$$

Then the probability vector of the working states is given by Equation (2.221), i.e.

$$\tilde{\underline{P}}_t(s) = (1,0)(sI-A_1)^{-1}$$

$$= (s+\lambda+\mu,\lambda) \times \frac{1}{\det(sI-A_1)}$$

thus the transform of the reliability is

$$\tilde{R}(s) = \frac{s+2\lambda+\mu}{(s-s_1)(s-s_2)}$$

hence

$$R(t) = \frac{s_1+2\lambda+\mu}{s_1-s_2} e^{s_1 t} + \frac{s_2+2\lambda+\mu}{s_2-s_1} e^{s_2 t}$$
$$= \frac{s_1 e^{s_2 t} - s_2 e^{s_1 t}}{s_1 - s_2} \quad . \tag{2.233}$$

For the mean times of failure we have

$$(-A_1)^{-1} = \begin{pmatrix} \lambda+\mu & \lambda \\ \mu & \lambda \end{pmatrix} \times \frac{1}{\lambda^2} \quad .$$

The MTFF is calculated using Equation (2.228), i.e.

$$MTFF = \frac{2\lambda+\mu}{\lambda^2} \tag{2.234}$$

which is this problem coincides with the MTTF because the system was initially in state 0. If it starts in state 1 (i.e. $\underline{c} = (0,1,0)$) its MTTF is

$$MTTF = \frac{\mu+\lambda}{\lambda^2} \quad .$$

Consider now a more general system with n identical units of which n-1 are on standby and one repairman. The matrix A is (the number of each state indicates the number of units which are down)

$$A = \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & 2 & 3 & \ldots. & (n-1) & (n) \end{array} \leftarrow \text{states} \\ \downarrow \\ \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \ldots. & 0 & 0 \\ \mu & -(\lambda+\mu) & \lambda & 0 & \ldots. & 0 & 0 \\ 0 & \mu & -(\lambda+\mu) & \lambda & \ldots. & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots. & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots. & \mu & -\mu \end{pmatrix} \end{array} \begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \\ n \end{array} \tag{2.235}$$

The asymptotic availability is

$$p_\infty = 1 - \Pi_n = 1 - \left[\sum_{i=0}^{n} \left(\frac{\mu}{\lambda} \right)^i \right]^{-1} . \qquad (2.236)$$

For n=1 we get

$$p_\infty = \frac{\mu}{\lambda+\mu}$$

which is the result we have found for a single unit (Equation (2.178)). For n=2 we find the expression for the two-unit system, Equation (2.232).

The MTFF is

$$MTFF = \frac{1}{\lambda} \sum_{i=0}^{n-1} \frac{n!}{(i+1)(n-i-1)!} \left(\frac{\mu}{\lambda} \right)^i . \qquad (2.237)$$

If the number of repairmen is n the corresponding expressions are

$$p_\infty = 1 - \left[\sum_{i=0}^{n} \frac{n}{i!} \left(\frac{\mu}{\lambda} \right)^{n-i} \right]^{-1} \qquad (2.238)$$

and

$$MTFF = \frac{1}{\lambda} \sum_{i=0}^{n-1} \frac{(1+\mu/\lambda)^i}{i+1} . \qquad (2.239)$$

Simple expressions for the reliability function cannot be found and in a particular problem the technique with the absorbing states may be employed. In the present case of identical units (for which matrix A is tridiagonal) the method has been systematized with the introduction of polynomials with special properties.[82,5] However, the algebra is quite involved.

Parallel System

When the number of units is equal to the number of the repairmen the availability function can be found using the binomial distribution without

solving the Markov model. This is justified by the fact that each unit functions and is repaired independent of the others. The steady-state availability and unavailability for each unit are $\mu/\mu+\lambda$ and $\lambda/\mu+\lambda$ respectively. Therefore, for a system with n units where m are needed the steady-state availability is

$$P_{\infty} = \sum_{i=m}^{n} \binom{n}{i} \left(\frac{\mu}{\lambda+\mu}\right)^{i} \left(\frac{\lambda}{\lambda+\mu}\right)^{n-i} \quad . \tag{2.240}$$

If, however, the number of repairmen is less than the number of units, the performance of each unit is not independent of the others. Assume, for example, that only one repairman is available. The solution of the Markov system yields the following simple expressions for the asymptotic probabilities of the system being in the various states (the number of each state again indicates how many units are down)

$$\Pi_{n} = \left[\sum_{k=0}^{n} \frac{1}{k!} \left(\frac{\mu}{\lambda}\right)^{k}\right]^{-1} \tag{2.241}$$

$$\Pi_{i} = \frac{1}{i!} \left(\frac{\mu}{\lambda}\right)^{i} \Pi_{n} \quad . \tag{2.242}$$

Thus if at least m units are needed for successful operation the asymptotic availability of the system will be

$$P_{\infty} = \sum_{i=0}^{n-m} \Pi_{i} \quad . \tag{2.243}$$

2.D.6. Non-Markovian Systems

A basic assumption in Markov processes is the constancy of the transition rates. In the framework of failure and repair of equipments a Markov model can be used only if the failure and repair distributions are exponential, since this is the only distribution in which the past does not affect the future,

that is, the resulting transition rates are independent of time. Therefore, in the important cases in applications, where the equipments age or the repair is not exponential, other methods of attacking the problem must be used. We will focus our attention upon the problem of exponential failure but arbitrary repair, which arises more often in practice.

The general problem can be treated as a Semi-Markov process, as we will see later. There are situations, however, where it can be formulated as a Markov process with the introduction of artificial ("dummy") states, so that all the transitions among states occur at a constant rate. A typical example useful in applications is when the failure is exponential and the repair is described by the gamma distribution

$$G(t) = 1 - e^{-\mu t} - \mu t e^{-\mu t}$$

with density

$$g(t) = \mu^2 t e^{-\mu t} \quad .$$

But we know that the gamma density is the convolution of exponential densities and in the present case

$$g(t) = (\mu e^{-\mu t}) * (\mu e^{-\mu t}) \quad .$$

This property suggests that the time-to-repair can be thought of as the sum of two independent random variables each exponentially distributed with the same hazard function μ. Therefore the repair process can be considered as performed in two identical stages. If there is only one unit we define the following states:

0: the unit is up

1: the unit is in the last stage of repair

2: the unit is in the first stage of repair.

The system of equations in (see also Figure 2.31)

$$\begin{aligned} \frac{dP_0}{dt} &= -\lambda P_0 + \mu P_1 \\ \frac{dP_1}{dt} &= -\mu P_1 + \mu P_2 \\ \frac{dP_2}{dt} &= \lambda P_0 - \mu P_2 \end{aligned} \qquad (2.244)$$

thus

$$A = \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \leftarrow \text{states} \\ \begin{pmatrix} -\lambda & 0 & \lambda \\ \mu & -\mu & 0 \\ 0 & \mu & -\mu \end{pmatrix} \begin{array}{c} \downarrow \\ 0 \\ 1 \\ 2 \end{array} \end{array} .$$

This system along with some initial conditions can be solved as before. The steady-state availability is

$$p_\infty = \Pi_o = \frac{\mu}{\mu+2\lambda} \quad . \qquad (2.245)$$

It is clear that the technique can also be used when the failure distribution is gamma distributed and for redundant configurations.

More generally, if only the mean τ and the variance σ^2 of the time-to-repair is known, this method of exponential stages can be used to render the system amenable to Markovian treatment (Reference 83 and, for an application to reliability problems, Reference 84). The coefficient of variation is defined as

$$v = \frac{\sigma}{\tau} \quad . \qquad (2.246)$$

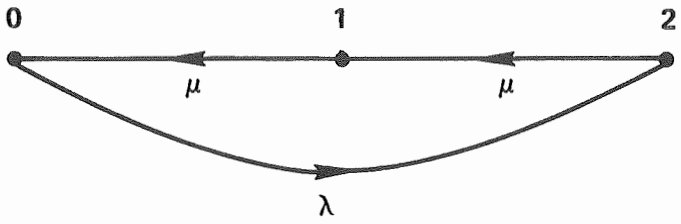


Figure 2.31. States and Transition Rates for One Unit With Exponentially Distributed Failure and Gamma Distributed Repair.

If $v < 1$ the repair process is modeled as a series of k successive identical exponential stages (Figure 2.32) each with repair rate μ. Then the repair density is the gamma (Erlangian) density

$$g(t) = \frac{1}{k!} \mu^k t^{k-1} e^{-\mu t} \quad . \tag{2.247}$$

The parameters k and μ are determined by equating the mean and the coefficient of variation of g(t) to the true ones, i.e.

$$\frac{k}{\mu} = \tau$$

$$\frac{1}{\sqrt{k}} = \frac{\sigma}{\tau} \qquad \text{(k is selected as the nearest integer satisfying this equation).} \tag{2.248}$$

If $v > 1$ the stages are connected in parallel. Assuming two stages for simplicity, 1 and 2, the failed unit enters either stages 1 or 2 with probabilities ρ and $(1-\rho)$ respectively. The repair rates are $2\rho\mu$ for stage 1 and $2(1-\rho)$ for stage 2 (always $0 < \rho \leq 0.5$). The repair density is then

$$g(t) = 2\rho^2 \mu e^{-2\rho\mu t} + 2(1-\rho)^2 \mu e^{-2(1-\rho)\mu t} \tag{2.249}$$

where ρ and μ are determined as before, i.e.

$$\frac{1}{\mu} = \tau$$

$$\sqrt{1 + \frac{(1-2\rho)^2}{2\rho(1-\rho)}} = \frac{\sigma}{\tau} \quad . \tag{2.250}$$

Figure 2.33 shows the states and transition rates.

Besides the method of stages, another powerful method of handling non-exponential failure or repair is that of semi-Markov process. Pyke[85,86] has studied in detail the semi-Markov model; presentations of the theory may also be found in References 5 and 87, while interesting applications appear in References 88,89,90 and 91.

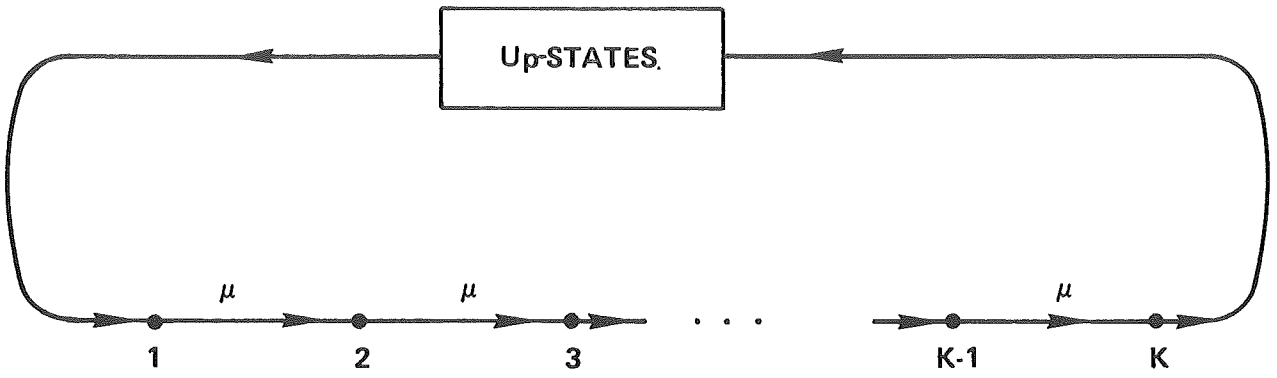


Figure 2.32. Model of a Repair Process as a Series Connection of K Exponential Stages.

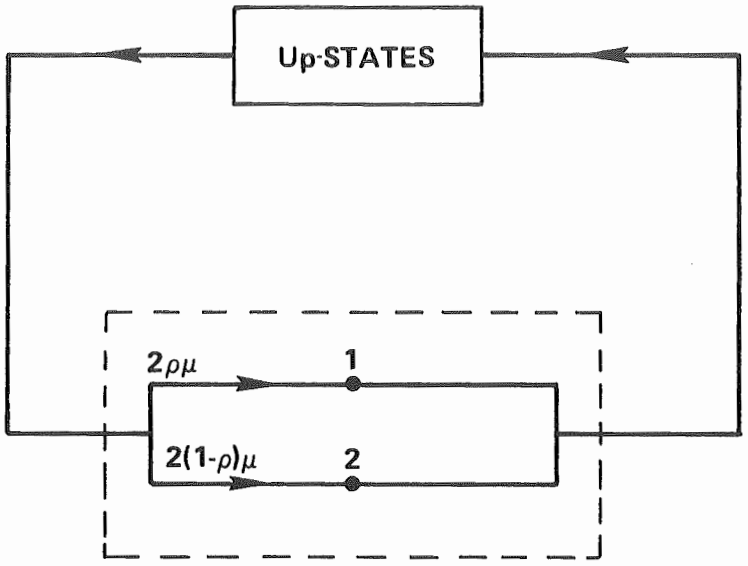


Figure 2.33. Model of a Repair Process as a Parallel Connection of Two Exponential Stages.

A semi-Markov process (or Markov renewal process) is a mixing of renewal theory and Markov processes. The system can be in any of a finite number of states at time t. If it enters state i at time t, it will spend there a random amount of time, say T, before visiting state j. Thus we define

$$F_{ij}(t) = P[\text{time spent in state i before visiting state j is less than t}]. \tag{2.251}$$

Notice that in a Markov process this distribution is the exponential; here it can be any distribution. When the system is in state i there is a probability p_{ij} that the next transition is to state j. The unconditional probability that the next transition is to state j before time t has elapsed is then

$$Q_{ij}(t) = p_{ij}\, F_{ij}(t), \qquad i \neq j \tag{2.252}$$

and the unconditional probability that an exit from state i will occur before time t is

$$Q_i(t) = \sum_{\substack{j \\ j\neq i}} Q_{ij}(t) \quad . \tag{2.253}$$

The mean time the system spends in state i is

$$\mu_i = \int_0^\infty (1-Q_i(t))\,dt = \sum_{\substack{j \\ j\neq i}} p_{ij}\mu_{ij} \tag{2.254}$$

where μ_{ij} is the mean of $F_{ij}(t)$.

The states are classified (absorbing, etc.) as in the case of Markov processes. Then equations for the time dependent probability of being in state i and the probability of absorption before t can be written down. These involve complicated convolution integrals which can be handled with Laplace

Transforms, although numerical problems arise due to the complicated character of the equations (see Reference 20 for an example). We will not present the whole theory here, but only several simple asymptotic results which are very useful in practice. These can be expressed in terms of the p_{ij} and μ_i alone.

Let P be the matrix with elements p_{ij} (the element p_{ii} is defined as $p_{ii} = 1 - \sum_j p_{ij}$). If all states communicate there exists a steady-state probability vector $\underline{\Pi} \equiv (\Pi_0, \Pi_2, \ldots, \Pi_n)$ where Π_i is the asymptotic probability that the system is in state i. To find Π_i we define the determinants D_i as

$D_i \equiv$ determinant of the matrix I-P with the i^{th} row and i^{th} column deleted.

Then it can be shown that

$$\Pi_i = \frac{D_i \mu_i}{\sum_i D_i \mu_i} \quad . \tag{2.255}$$

As an example we consider one unit operating and one identical unit on (cold) standby. The failure of both units is exponential with failure rate λ.

There is one repair facility and the repair time is fixed τ. The states of the system are

0: one unit is operating, the other is on standby

1: one unit is operating, the other is under repair

2: one unit is under repair, the other is down.

The conditional distribution functions $F_{ij}(t)$ and the transition probabilities must be specified. Clearly

$$F_{01} = 1 - e^{-\lambda t}, \qquad p_{01} = 1, \quad p_{02} = 0$$

since the system cannot go from state 0 to 2. The mean time spent in state 0 is then

$$\mu_o = \sum_{i=1}^{2} p_{oi}\,\mu_{oi} = \frac{1}{\lambda} \quad .$$

Once the system enters state 1 there are two possibilities, either it goes back to 0 if the repair is completed before the on-line unit fails, or it visits state 2, if the failure occurs before the completion of repair (i.e., before time τ has elapsed). Therefore p_{10} is the probability that the failure of the on-line unit occurs after time τ (measured from the moment the system enters state 1), thus

$$p_{10} = e^{-\lambda\tau} \quad .$$

The probability p_{12} is then

$$p_{12} = 1 - e^{-\lambda\tau} \quad .$$

To find the mean time spent in state 1 we use the unconditional probability $Q_1(t)$, which is

$$Q_1(t) = 1 - e^{-\lambda t} \qquad \text{if} \quad t < \tau$$

$$= 0 \qquad \text{if} \quad t > \tau \quad \text{(since the failed unit is repaired exactly after } \tau\text{)}$$

therefore,

$$\mu_1 = \int_o^\infty [1-Q_1(t)]dt = \int_o^\tau e^{-\lambda t}dt = \frac{1 - e^{-\lambda\tau}}{\lambda} \quad .$$

Finally, we must find the corresponding quantities for state 2. Since the only transition possible is to state 1 we have $p_{21} = 1$. To determine μ_2 is not as simple though. When state 2 is entered one unit is under repair and the other one has just failed. The time spent in state 2 is thus the time remaining for the completion of repair on the unit which is already in the repair facility.

The probability density that a unit fails in $(t_1, t_1 + dt)$ is $\lambda e^{-\lambda t_1}$ and the probability that it fails before τ is p_{12}, than the probability density of failure before τ is

$$\frac{\lambda e^{-\lambda t_1}}{p_{12}} = \frac{\lambda e^{-\lambda t_1}}{1-e^{-\lambda\tau}} \quad .$$

The time remaining until repair is completed is $t_2 = \tau - t_1$ with density

$$\frac{\lambda e^{-\lambda(\tau-t_2)}}{1-e^{-\lambda\tau}}$$

therefore, the mean time spent in state 2 is

$$\mu_2 = \frac{\lambda e^{-\lambda\tau}}{1-e^{-\lambda\tau}} \int_0^\tau t_2 e^{\lambda t_2} \, dt_2$$

$$= \frac{\tau\lambda - 1 + e^{-\lambda\tau}}{(1-e^{-\lambda\tau})} \quad .$$

To calculate the steady-state probabilities we need the determinants D_i. The matrix I-P is

$$I-P = \begin{pmatrix} 1 & -1 & 0 \\ -e^{-\lambda\tau} & 1 & -1+e^{-\lambda\tau} \\ 0 & -1 & 1 \end{pmatrix}$$

therefore

$$D_0 = e^{-\lambda\tau}, \quad D_1 = 1, \quad D_2 = 1 - e^{-\lambda\tau} \quad .$$

Obviously the availability of the system is

$$p_\infty = \Pi_0 + \Pi_1 = 1 - \Pi_2$$

$$= 1 - \frac{D_2\mu_2}{\sum_{i=0}^{2} D_i\mu_i}$$

$$= \frac{1}{\tau\lambda + e^{-\lambda\tau}} \quad . \tag{2.256}$$

It is evident from the previous calculations how extremely complicated the method is when more than two units are involved. Several general models are listed in Barlow and Proschan (Reference 5) and in Reference 9 the two-unit cold standby system with one repairman and general failure and repair distributions is analyzed. All these results are too complex to be reproduced here, however, the following two expressions are particularly simple:

1. For a system with one unit operating, n-1 units on standby and n repair facilities, where all the units are identical with failure rate λ and all repair facilities identical with general repair distribution G(t) with mean τ, the asymptotic probability that the system is in state i is

$$\Pi_i = \frac{(\lambda\tau)^i/i!}{\sum_{i=0}^{k} \frac{(\lambda\tau)^i}{i!}} \tag{2.257}$$

(the state number indicates how many units are down). Thus, for two units the availability of the system is

$$p_\infty = \Pi_0 + \Pi_1 = 1 - \Pi_2$$

$$= \frac{2(1 + \lambda\tau)}{2 + 2\lambda\tau + (\lambda\tau)^2} \quad . \tag{2.258}$$

2. If n components are in series and the MTTF of the i^{th} component is m_i and its MTTR is τ_i, the availability of the system is (Reference 92)

$$p_\infty = \Pi_0 = \left[1 + \sum_{i=1}^{n} \frac{\tau_i}{m_i}\right]^{-1} . \qquad (2.259)$$

The problem of determining the reliability of the various configuration has also been studied. The review paper by Osaki (Reference 93) contains the theory for two-unit systems and many references. More recent papers on the subject are References 89,91 and 94.

2.D.7. Inspection Intervals

The corrective actions discussed thus far belong to the class of off-schedule maintenance procedures: when the system or parts of it fail they are restored to the functioning state by some repair process.

As is often the case with redundant systems a failure can only be detected when the system is inspected. The system may still operate, but it is not as reliable any more and the probability for a total failure is enhanced. By planning to inspect at certain time intervals we can increase the availability of the system.

In this section we consider the following model (more general models of preventive maintenance will be presented later):

a. The system has a failure distribution F(t).

b. It is inspected every τ_i units of time.

c. Failures are detected only when the system is inspected. The probability of uncovering a failure at inspection is unity.

d. At each inspection the system is renewed either by repair or replacement of the failed parts.

We assume that the time it takes to inspect and repair or renew the system is on the average τ_r and that failure cannot occur during inspection.

Our objective is to estimate the availability of the system under this policy and to select the "best" τ_i and/or τ_r according to some criterion. This model is widely used in engineering applications and various aspects of it are discussed in References 9,95,96,97 and 98.

Since the model is periodic it suffices to examine one period from 0 to $\tau_i + \tau_r$ (Figure 2.34). In the interval $\tau_i + \tau_r$ the system is inoperable for a time τ_r plus the time interval between its failure and inspection. If it fails at t ($0 \leq t \leq \tau_i$) this last interval is $\gamma = \tau_i - t$. The average time the system is down due to failure is then

$$E[\gamma] = \int_0^{\tau_i} (\tau_i - t)\, dF(t)$$

$$= \int_0^{\tau_i} F(t)\, dt \quad . \tag{2.260}$$

Therefore the average system availability is

$$p = 1 - \frac{\tau_r + E[\gamma]}{\tau_i + \tau_r}$$

$$= \frac{\int_0^{\tau_i} R(t)\, dt}{\tau_i + \tau_r} \tag{2.261}$$

where $R(t) = 1 - F(t)$ is the system reliability.

If τ_r is fixed and we wish to calculate the optimum τ_i, so that p is maximum, we set the derivative of p with respect to τ_i equal to zero, that is.

$$\frac{dp}{d\tau_i} = (\tau_i + \tau_r)\, R(\tau_i) - \int_0^{\tau_i} R(t)\, dt = 0 \quad . \tag{2.262}$$

A simple example involves a system consisting of one component with exponential failure distribution. Then

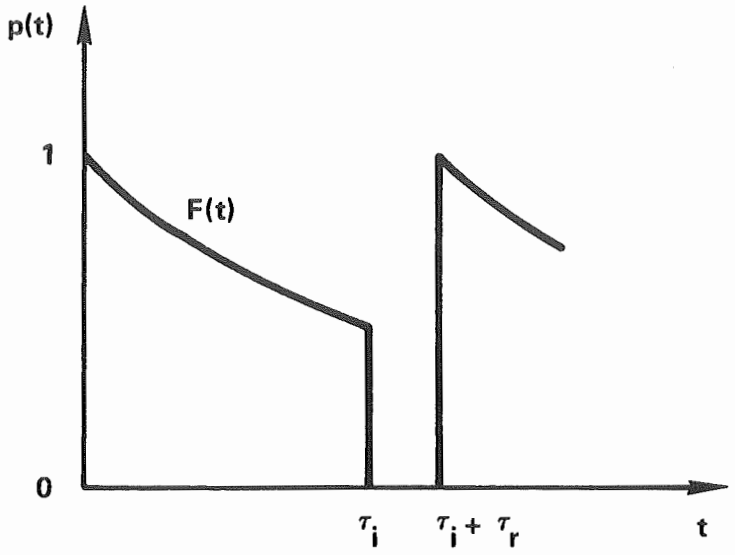


Figure 2.34. Availability of a System Under Inspection and Repair.

$$p = \frac{1 - e^{-\lambda\tau_i}}{\lambda(\tau_i + \tau_r)} \tag{2.263}$$

and the optimum τ_i is the solution of

$$e^{\lambda\tau_i} = 1 + \lambda(\tau_i + \tau_r) \quad . \tag{2.264}$$

If we assume that

$$\lambda(\tau_i + \tau_r) << 1 \tag{2.265}$$

the availability reduces to

$$p = \frac{\tau_i}{\tau_i + \tau_r} - \frac{\lambda \tau_i^2}{2(\tau_i + \tau_r)}$$

$$\cong 1 - \frac{\lambda\tau_i}{2} \quad , \quad \text{if} \quad \tau_r << \tau_i \tag{2.266}$$

and the optimum inspection interval is determined from

$$1 + \lambda\tau_i + \frac{(\lambda\tau_i)^2}{2} = 1 + \lambda(\tau_i + \tau_r)$$

which yields

$$\tau_i = \sqrt{\frac{2\tau_r}{\lambda}} \quad . \tag{2.267}$$

For systems consisting of two or more components the failure distribution is determined by the individual distributions and the logical interconnection of the units. Furthermore, the test can be conducted in either of two ways: if simultaneous testing is used all the components are checked at the same time, while in staggered tests the components are tested in different times and, as a result, at any instant of time the units have been in operation for different times. The type of testing employed also affects the system availability.

In redundant systems it would be meaningless to test all components at the same time, thus rendering the system inoperable. In the case of simultaneous testing the units are actually checked consecutively, so that the system can perform its task at all times. In order to see the effect of testing on a redundant system we analyse the case of two units in parallel under simultaneous testing. The test is performed every τ_i units of time and the checking and repair time for each unit is τ_r (the units are identical with failure rate λ). Then the period is $\tau_i + 2\tau_r$. To estimate the unavailability q of the system it is convenient to write

$$q = q_1 + q_2 \tag{2.268}$$

where q_1 is the average unavailability during the interval τ_i due to undetected failures of the system, and q_2 is the average unavailability during the interval $2\tau_r$ due to failures of the system while one component is under testing.

Again we assume that

$$\lambda(\tau_i + 2\tau_r) << 1 \tag{2.269}$$

so that for each component the failure distribution is approximated by

$$F_c = 1 - e^{-\lambda t} \cong \lambda t \quad . \tag{2.270}$$

During τ_i the failure distribution of the system is (one-out-of-two system)

$$F_1(t) = (\lambda t)^2 \tag{2.271}$$

therefore

$$q_1 = \frac{1}{\tau_i+2\tau_r} \int_o^{\tau_i} (\lambda t)^2 dt$$

$$= \frac{(\lambda\tau_i)^3}{3\lambda(\tau_i+2\tau_r)} \quad . \tag{2.272}$$

Usually $\tau_i >> 2\tau_r$ whence

$$q_1 = \frac{(\lambda\tau_i)^2}{3} \quad . \tag{2.273}$$

During $2\tau_r$ the failure distribution of the system is

$$\begin{aligned} F_2(t) &= \lambda t \quad , \quad \tau_i < t < \tau_i + \tau_r \\ &= \lambda[t-(\tau_i + \tau_r)], \quad \tau_i + \tau_r < t < \tau_i + 2\tau_r \end{aligned} \tag{2.274}$$

since at $\tau_i + \tau_r$ the operating component starts as good as new, therefore,

$$q_2 = \frac{1}{(\tau_i+2\tau_r)} \left[\int_{\tau_i}^{\tau_i+\tau_r} \lambda t \, dt + \int_{\tau_i+\tau_r}^{\tau_i+2\tau_r} \lambda[t-(\tau_i+\tau_r)]dt \right] \tag{2.275}$$

which, under the assumption $\tau_i >> 2\tau_r$, simplifies to

$$q_2 = \lambda\tau_r \quad . \tag{2.276}$$

Therefore, the average unavailability of the one-out-of-two system under simultaneous testing is

$$\begin{aligned} q &= q_1 + q_2 \\ &= \frac{(\lambda\tau_i)^2}{3} + \lambda\tau_r \end{aligned} \tag{2.277}$$

under the assumptions

$\tau_r << \tau_i$ the average time to inspect and repair is much shorter than the inspection interval,

and

$\tau_i+2\tau_r << \frac{1}{\lambda}$: the mean time to failure for each component is much longer than the testing period.

We now proceed to estimate the unavailability in the case when the checking of each component is staggered over the interval τ_i. Figure 2.35 shows the new situation: each unit is on-line for τ_i units of time. The only difference is that if one unit starts operating at t=0 the other starts at $t = k\tau_i + \tau_r$, where $0 \leq k \leq 1$. In a period $(0, \tau_i + \tau_r)$ the following situations arise:

$(0, k\tau_i)$: one unit has been on line for a time t and the other for a time $t + (1-\gamma-)\tau_i$

$(k\tau_i, k\tau_i+\tau_r)$: one unit is operating with age t, the other is down

$(k\tau_i+\tau_r, \tau_i)$: one unit is operating with age t and the other with age $t-(k\tau_i+\tau_r)$

$(\tau_i, \tau_i+\tau_r)$: only one unit of age $t-(k\tau_i+\tau_r)$ is operating.

The unavailability is then

$$q = \frac{1}{(\tau_i+\tau_r)} \left[\int_o^{k\tau_i} \lambda^2 t[t+(1-k)\tau_i]dt + \int_{k\tau_i}^{k\tau_i+\tau_r} \lambda t \, dt + \right.$$

$$\left. + \int_{k\tau_i+\tau_r}^{\tau_i} \lambda^2 [t-(k\tau_i+\tau_r)]t \, dt + \int_{\tau_i}^{\tau_i+\tau_r} \lambda[t-(k\tau_i+\tau_r)]dt \right]. \qquad (2.278)$$

Performing the integrations and assuming that $\tau_r << \tau_i$ we get

$$q = \frac{(\lambda\tau_i)^2}{3} + \lambda\tau_r - \frac{(\lambda\tau_i)^2}{2} k(1-k)$$

which attains its minimum for k = 1/2 (symmetrical or uniformly staggered test), thus

$$q = \frac{5}{24} (\lambda\tau_i)^2 + \lambda\tau_r \quad . \qquad (2.279)$$

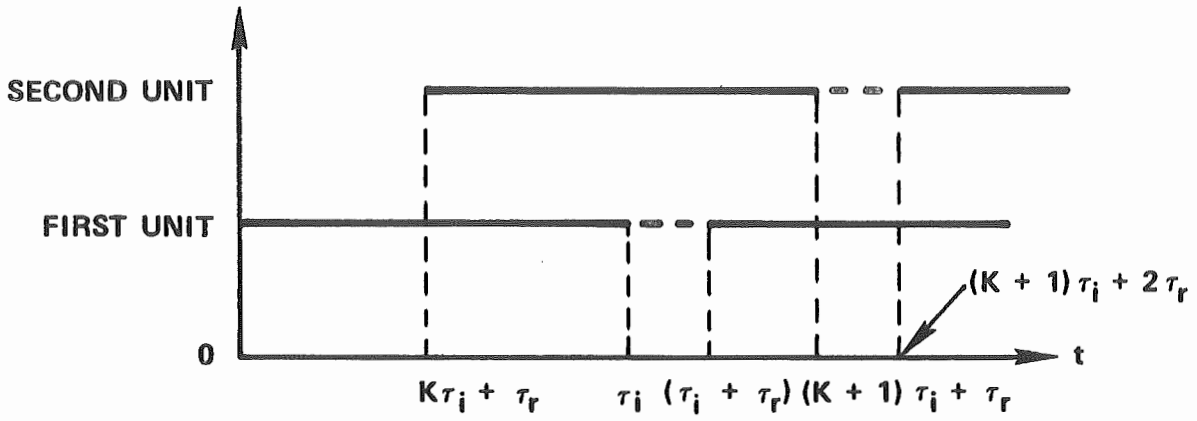


Figure 2.35. **Staggered Testing of a one-out-of-two System. The first unit starts new at t = 0 and is working until τ_i (continuous line) and then it is tested for τ_r units of time (broken line). The second unit starts at $K\tau_i + \tau_r$ and works for τ_i units of time and then is put under testing for τ_r units of time, etc.**

This expression may again be considered as the sum of two unavailabilities q_1 and q_2 as defined before, with

$$q_1 = \frac{5}{24} (\lambda\tau_i)^2$$

and

$$q_2 = \lambda\tau_r \quad .$$

Notice that q_1 is greater for a uniformly staggered test than for a simultaneous test. This is true for any configuration.

These formulas can be generalized to other more complex situations. For instance, under the same assumptions as before and for a parallel system with n units (one required) the unavailabilities are:

$$q = \frac{(\lambda\tau_i)^n}{n+1} + \lambda^{n-1} \tau_i^{n-2} \tau_r \ , \qquad \text{simultaneous test} \qquad (2.280)$$

$$q = \frac{n!(n+3)(\lambda\tau_i)^n}{4n^n(n+1)} + \frac{(n-1)!(\lambda\tau_i)^{n-2}\lambda\tau_r}{n^{n-2}} \ , \text{ perfectly staggered test}$$

(each unit tested every τ_i/n, $n \leq 5$)

(2.281)

For some common configurations Table 2.1 gives the unavailability due to failures (i.e., q_1) for simultaneous and symmetrically staggered testing (Reference 99).

This discussion was confined to estimating the effects of testing and renewal on the availability of the system. The results then may be used to estimate τ_i and τ_r so that a specific reliability goal can be achieved.

Assume that the unavailability of the system should not be greater than q. Then for a nonredundant system the time interval between tests is found from Equation (2.266), i.e.

TABLE 2.1 UNAVAILABILITY AS A FUNCTION OF LOGIC CONFIGURATION AND TESTING SCHEDULE. (Ref. IEEE STD 352-1972)

LOGIC m/n	SIMULTANEOUS TESTING	PERFECTLY STAGGERED TESTING
1/2	$(1/3)\ (\lambda\tau_i)^2$	$(5/24)\ (\lambda\tau_i)^2$
2/2	$\lambda\tau_i$	$\lambda\tau_i$
1/3	$(1/4)\ (\lambda\tau_i)^3$	$(1/12)\ (\lambda\tau_i)^3$
2/3	$(\lambda\tau_i)^2$	$(2/3)\ (\lambda\tau_i)^2$
3/3	$(3/2)\ (\lambda\tau_i)$	$(3/2)\ (\lambda\tau_i)$
1/4	$(1/5)\ (\lambda\tau_i)^4$	$(251/7680)\ (\lambda\tau_i)^4$
2/4	$(\lambda\tau_i)^3$	$(3/8)\ (\lambda\tau_i)^3$
3/4	$2\ (\lambda\tau_i)^2$	$(11/8)\ (\lambda\tau_i)^2$
(1/2) × 2	$2/3\ (\lambda\tau_i)^2$	$(5/12)\ (\lambda\tau_i)^2$

$$\frac{\lambda\tau_i}{2} = q$$

hence

$$\tau_i = \frac{2q}{\lambda} \quad . \tag{2.282}$$

For redundant systems we must have

$$q = q_1 + q_2 \quad .$$

Hirsh (Reference 97) used the criterion that during testing and repair the unavailability of the system should be equal to its unavailability during normal operation, i.e.

$$q_1 = q_2 \tag{2.283}$$

whence

$$q_1 = \frac{q}{2} \quad \text{and} \quad q_2 = \frac{q}{2} \quad . \tag{2.284}$$

These two relations allow one to calculate both the testing interval τ_i and the allowable repair time τ_r. The expression for q_1 can be found in Table 2.1 or must be calculated. Similarly q_2 must be calculated. Several common cases are given below:

<u>one-out-of-n system</u>:

simultaneous testing:

$$\left.\begin{aligned} \frac{(\lambda\tau_i)^n}{n+1} &= \frac{q}{2} \\ \lambda^{n-1}\tau_i^{n-2}\tau_r &= \frac{q}{2} \end{aligned}\right\} \Rightarrow \begin{aligned} \tau_i &= \frac{1}{\lambda}\sqrt[n]{\frac{(n+1)q}{2}} \\ \tau_r &= \frac{\lambda\tau_i^2}{n+1} \end{aligned} \tag{2.285}$$

Uniformly staggered test ($n \leq 5$)

$$\left.\begin{aligned} \frac{n!(n+3)(\lambda\tau_i)^n}{4n^n(n+1)} &= \frac{q}{2} \\ \frac{(n-1)!(\lambda\tau_i)^{n-2}\lambda\tau_r}{n^{n-2}} &= \frac{q}{2} \end{aligned}\right\} \Rightarrow \begin{aligned} \tau_i &= \frac{1}{\lambda}\sqrt[n]{\frac{2\,q\,n^n(n+1)}{n!(n+3)}} \\ \tau_r &= \frac{q}{2\lambda}\,\frac{n^{n-1}}{(n-1)!(\lambda\tau_i)^{n-2}} \end{aligned} \qquad (2.286)$$

<u>two-out-of-three system</u>:

simultaneous testing:

$$\left.\begin{aligned} (\lambda\tau_i)^2 &= \frac{q}{2} \\ 3\lambda\tau_r &= \frac{q}{2} \end{aligned}\right\} \Rightarrow \begin{aligned} \tau_i &= \frac{1}{\lambda}\sqrt{\frac{q}{2}} \\ \tau_r &= \frac{q}{6\lambda} \end{aligned} \qquad (2.287)$$

uniformly staggered test:

$$\left.\begin{aligned} \frac{2}{3}(\lambda\tau_i)^2 &= \frac{q}{2} \\ 3\lambda\tau_r &= \frac{q}{2} \end{aligned}\right\} \Rightarrow \begin{aligned} \tau_i &= \frac{1}{\lambda}\sqrt{\frac{3q}{4}} \\ \tau_r &= \frac{q}{6\lambda} \end{aligned} \qquad (2.288)$$

<u>two-out-of-four system</u>:

simultaneous testing:

$$\begin{aligned} \tau_i &= \frac{1}{\lambda}\sqrt[3]{\frac{q}{2}} \\ \tau_r &= \frac{q}{4\lambda\left[2\sqrt[3]{\frac{q}{3}} - \sqrt[3]{\left(\frac{q}{2}\right)^2}\right]} \end{aligned} \qquad (2.289)$$

uniformly staggered test:

$$\tau_i = \frac{2}{\lambda} \sqrt[3]{\frac{q}{6}}$$

$$\tau_r = \frac{2q}{\left[11 \sqrt[3]{\frac{8q}{6}} - 3\left(\sqrt[3]{\frac{8q}{6}}\right)^2\right]} \qquad (2.290)$$

one-out-of-two twice system

simultaneous testing:

$$\tau_i = \frac{1}{\lambda} \sqrt{\frac{3q}{4}}$$

$$\tau_r = \frac{q}{4\lambda} \qquad (2.291)$$

uniformly staggered test:

$$\tau_i = \frac{1}{\lambda} \sqrt{\frac{6q}{5}}$$

$$\tau_r = \frac{q}{4\lambda} \qquad (2.292)$$

Also in Reference 97 nomographs are presented for graphical estimation of τ_i and τ_r in these cases.

2.D.8. Maintenance Policies

The model analyzed in the preceding section is only one of the many different maintenance policies that can be employed. Presentation of other general models can be found in References 5,69,95,100 and 101. Here we discuss several generalizations which illustrate the various approaches to maintenance.

A general model which allows for imperfect checking, distinction between checking time and repair time and failure during checkout is as follows:[102]

a. The system fails according to the exponential distribution $F(t)=1-e^{-\lambda t}$.

b. Inspection is performed every τ_i units of time.

c. Inspection takes τ_c units of time.

d. The probability that a failure will be detected is θ.

e. The probability of a false alarm (i.e., calling a good system bad) is α.

f. Inspection introduces stresses on the system and the probability that the system will fail during the checkout period is β.

g. The probability that the failure, which occurs during the checkout period, occurs before the actual testing is γ.

h. If a failure is detected, the duration of repair is on the average τ_r.

Under these assumptions the availability of the system is found to be

$$p = \frac{\theta\left(1-e^{-\lambda\tau_i}\right)}{\lambda(\tau_i+\tau_c)\left\{1 + e^{-\lambda\tau_i}\left[\beta(1-\alpha+\alpha\gamma-\gamma\theta) - (1-\theta)\right]\right\} + \theta\lambda\tau_r\left[1-(1-\beta)(1-\alpha)e^{-\lambda\tau_i}\right]} . \tag{2.293}$$

If the system cannot fail during checkout ($\beta\equiv 0$) and if no false alarm is possible ($\alpha\equiv 0$), then the availability is

$$p = \frac{\theta\left(1-e^{-\lambda\tau_i}\right)}{\lambda(\tau_i+\tau_c)\left[1 - e^{-\lambda\tau_i}(1-\theta)\right] + \theta\lambda\tau_r\left(1-e^{-\lambda\tau_i}\right)} . \tag{2.294}$$

If in addition the detection of failure is perfect ($\theta\equiv 1$) we have

$$p = \frac{1 - e^{-\lambda\tau_i}}{\lambda\left[\tau_i + \tau_c + \tau_r\left(1 - e^{-\lambda\tau_i}\right)\right]} . \tag{2.295}$$

This equation is similar to Equation (2.263). In the latter the checkout time and the repair time are lumped in the constant τ_r, while in Equation (2.295) these two times are separated and the time required for repair is multiplied

by the probability of the system being down at the end of τ_i, since only then is repair undertaken.

As a last maintenance policy we discuss that of marginal testing. A component can now be in more than two (up-down) states which are divided into three groups: A (good), B (marginal, the component still operates satisfactorily but it is expected to fail soon) and C (bad, the component is failed). When the component reaches C its failure is detected immediately and it is replaced in negligible time. Furthermore, at regular time intervals a test is performed to determine whether the component is in A or B (it cannot be in C, since failure is detected immediately). If it is found to be in A it passes the test, while if it is in B it fails the test and it is replaced in negligible time, thus starting operation in state A. This model is discussed (along with other maintenance policies) in Reference 101 using the theory of semi-Markov processes and in Reference 103 it is analyzed in detail under the assumption that the process is Markovian. In this case the transition rates λ_{ij} during normal operation are assumed to be known. Assuming that the component is initially good (i.e. in state 0) and that the test is performed in intervals of T units of time integral equations for the following quantities are derived:

1. the expected number of failures in [0,t]
2. the expected number of preventive removals in [0,t]
3. the reliability function R(t;x), i.e. the probability of no failure in an interval of duration t following component age x.

These equations are too complex to be reproduced here.

3. SAFETY ANALYSIS OF COMPLEX SYSTEMS

3.A. LOGIC DIAGRAMS

3.A.1 Introduction

In this chapter we study the methods which can be used for a quantitative safety analysis of multicomponent systems. There are various reasons as to why the approach of statistical distributions fails in this case. The foremost of these is that each such system is unique in the sense that there are no other identical systems (same components interconnected in the same way and operating under the same conditions) for which failure data have been collected, in order to make a statistical analysis possible. Furthermore, it is not only the probabilistic aspects of failure of the system which are of interest but also the initiating causes and the combination of events which can lead to a particular failure.

It is already apparent that the methods we will develop will be event c oriented, that is, they will not be limited to analyzing a system failure in terms of component failure alone, but they will also include other events, such as human errors, which may influence the performance of the system.

The natural way to attack a problem of this nature, where many events interact to produce other events, is to relate these events using simple logical relationships (intersection, union etc) and methodically to build a logical structure which represents the system. In fact, this is the underlying principle in all the methods, which deal with complex systems. An indispensable tool of the analysis is the logic diagram, which depicts the events and their logical relationships. We have already encountered the simplest form of a logic diagram when we examined series and parallel systems in Chapter 2. The logic diagram is different from the system diagram, which simply shows the physical connection of the components of the system, although it will be recognized that the former draws much information from the latter.

The various methods can be classified as qualitative and quantitative or as inductive and deductive. Usually it is the combination of these approaches which leads to a successful safety analysis. A method is qualitative if its main goal is to discover how a particular event can occur or to what consequences it leads. A quantitative method, on the other hand, attempts to describe probabilistically the phenomena and clearly it must be preceded by a qualitative method.

A more important distinction is that between inductive and deductive methods. An inductive method starts from a particular event and proceeds to uncover its consequences, while a deductive method proceeds backwards to identify the causes of the event. Both approaches may give qualitative and quantitative results.

These introductory remarks already show how important the tools of mathematical logic are in the study of complex systems. An elementary introduction to the subject is presented in the next section.

3.A.2 Logic

A fundamental notion in logic is that of a proposition or statement. It is best introduced by examples, like "valve fails closed", "the power is off" etc. It is the meaning of such sentences which is called a proposition, without regard to the actual words used or to any subjective meaning the sentence may have for the speaker or listener (this definition and much of what follows may be found in Ref. 104). The propositions that interest us are the ones that are either true or false. Then we may assign an indicator or truth value to each proposition. The indicator is (arbitrarily) set equal to 0 if the proposition is false and equal to 1 if the proposition is true. The most common proposition in safety studies is "component i is failed", which implies that the component can be in either of two states, good or bad. If the component

can be in more than two states (e.g., a valve may be good, fail open or fail closed) nothing changes, except that for each mode of failure there will be a proposition which can be true or false, and these propositions will be included separately in the study. Thus, for the valve the two propositions are "the valve fails closed" and "the valve fails open". Notice however that in any one of these is false, it doesnot necessarily mean that the valve is good, since it may have failed according to the other mode of failure.

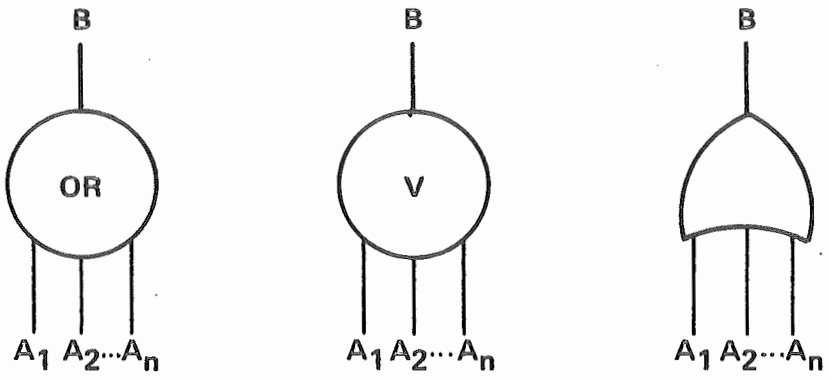
From given propositions we can derive new ones by applying simple operations. There are three fundamental operations, the union, the product and the complement. Their graphical representation (logic gates) is shown in Fig. 3.1

The union of n propositions A_1, $A_2 \ldots A_n$ is a new proposition B formed as $B = A_1$ or A_2 or ... or A_n meaning that B is true if any one of the A_i is true (including the possibility that more than one A_i may be true). Alternatively, B is false if all the A_i are false. Symbolically we write $B = A_1 \text{ V} A_2 \text{ V} \ldots \text{V} A_n$, or $B = A_1 + A_2 + \ldots + A_n$. The corresponding operation in set theory is the union of sets.

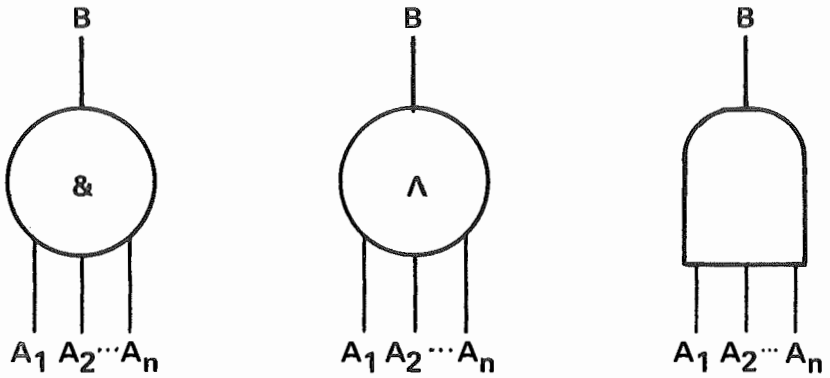
The product of n propositions $A_1, \ldots, A_n$ is a new proposition formed as $B = A_1$ and A_2 and ... and A_n meaning that B is true if all the A_i are true. We write $B = A_1 \, A_2 \ldots A_n$. It corresponds to the intersection operation in set theory.

Finally, the complement of a proposition A is a new proposition B, which is true if A is false and false if A is true. We write $B = \overline{A}$. Notice that this operation can be performed on one proposition only.

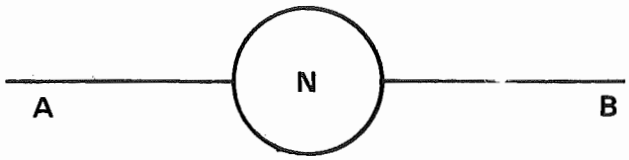
These three operations are the fundamental ones and any other operation on a finite number of propositions may be expressed in terms of products, unions and complements. Of course, given a logical function of a number of propositions we would like to know its truth value. A theorem which enables



OR GATE: $B = A_1$ or A_2 or . . . or $A_n = A_1 V A_2 V \ldots V A_n$



AND GATE: $B = A_1$ and A_2 and . . . and $A_n = A_1 A_2 \ldots A_n$



NOT GATE: B = not A = $\overline{A}$

Figure 3.1. Logic Gates.

us to do so states that the truth value b of a proposition B which is a function of the propositions A_i, is the same function of the truth values a_i of the propositions A_i. For example, the truth value of the AND gate is $b = a_1 a_2 \ldots a_n$. If we use the indicator values 0 and 1 as defined before, we see that b is 1 (true) if and only if all the a_i are 1(true), which is the definition of an AND gate. Therefore, the truth value of the logical function can be found by simple operations on the truth values of the original propositions. However, since the truth value a of a proposition A can be either 0 or 1 (binary logic), the following properties should be observed for the results to be consistent:

1. Complement

$$\bar{0} = 1$$
$$\bar{1} = 0 \qquad (3.1)$$

2. Union

$$a + a = a \qquad \text{(hence, } 1 + 1 = 1) \qquad (3.2)$$
$$1 + a = 1 \qquad (3.3)$$
$$\bar{a} + a = 1 \qquad (3.4)$$

3. Product

$$aa = a \qquad (3.5)$$
$$a.0 = 0 \qquad (3.6)$$
$$a\bar{a} = 0 \qquad (3.7)$$

Other operations with 0 and 1 follow the rules of arithmetic (e.g. 0 + 1 = 1, 0.1 = 0, etc.). Two laws which are often useful in the study of logical functions are the involution law (compare with Eq. (3.1)):

$$\bar{\bar{a}} = a \qquad (3.8)$$

and de Morgan's law:

$$\overline{a + b} = \overline{a}\,\overline{b} \tag{3.9}$$

$$\overline{ab} = \overline{a} + \overline{b} \tag{3.10}$$

Eq. (3.9) states that the complement of the output of an OR gate with inputs A and B is the output of and AND gate with inputs $\overline{A}$ and $\overline{B}$. Similarly, Eq. (3.10) states that the complement of the output of an AND gate is the output of an OR gate with inputs $\overline{A}$ and $\overline{B}$.

A convenient way to find the truth value of a logical function is the truth table. It lists the propositions and all the combinations of their truth values with the corresponding truth value of the outcome. Table 3.1 exhibits the truth tables for an OR and AND gate with two inputs each and for a NOT gate.

With the help of the three basic logical operations that we have introduced we can form any other logic gate which may be useful in a particular problem. The most common of these is the r-out-of-n gate (Fig. 3.2). The output B is true if any r or more inputs are true. The AND and OR gates are special cases of this gate for r = n and r = 1 respectively. As an example, consider the 2-out-of-3 gate. Its output is

$$B = A_1\,A_2 + A_2\,A_3 + A_3\,A_1 \tag{3.11}$$

and its truth table is shown in Table 3.2.

The operations with 0 and 1 presented above belong to a formal mathematical theory called Boolean algebra. It concerns the algebra of a set S with elements $a_1, a_2, \ldots$ (in the previous case the set consisted of the elements 0 and 1 only) in which the union (sum) and product are defined to obey the following axioms:

commutative law

$$a_1 + a_2 = a_2 + a_1 \tag{3.12}$$

$$a_1\,a_2 = a_2\,a_1 \tag{3.13}$$

TABLE 3.1 TRUTH TABLES FOR OR, AND AND NOT GATES

A_1	A_2	$B = A_1 + A_2$	$C = A_1 A_2$
0	0	0	0
1	0	1	0
0	1	1	0
1	1	1	1

A	$B = \overline{A}$
0	1
1	0

TABLE 3.2 TRUTH TABLE FOR A 2-OUT-OF-3 SYSTEM

A_1	A_2	A_3	$B = A_1A_2 + A_2A_3 + A_3A_1$
0	0	0	0
1	0	0	0
1	1	0	1
0	1	0	0
0	0	1	0
1	0	1	1
1	1	1	1
0	1	1	1

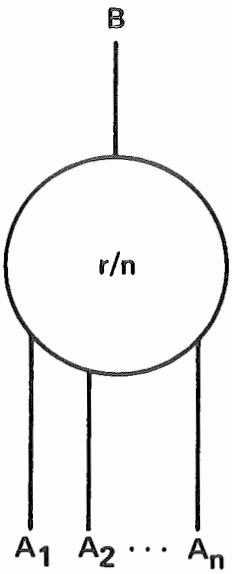


Figure 3.2. r-out-of-n Logic Gate.

associative law

$$(a_1 + a_2) + a_3 = a_1 + (a_2 + a_3) \tag{3.14}$$

$$(a_1\ a_2)\ a_3 = a_1(a_2\ a_3) \tag{3.15}$$

distributive law

$$a_1 + (a_2\ a_3) = (a_1 + a_2)\ (a_1 + a_3) \tag{3.16}$$

$$a_1(a_2 + a_3) = a_1\ a_2 + a_1\ a_3 \tag{3.17}$$

0 is the identity element for the union, i.e.,

$$0 + a_i = a_i \tag{3.18}$$

1 is the identity element for the product, i.e.,

$$1a_i = a_i \tag{3.19}$$

For any a_i its complement exists, i.e.,

$$a_i + \bar{a}_i = 1 \tag{3.20}$$

$$a_i\ \bar{a}_i = 0 \tag{3.21}$$

An important property is the duality principle: in any Boolean expression we can interchange unions and products and the elements 0 and 1 and thus produce another valid Boolean expression. This enables us to study either the failure or the success of a system.'

Suppose now that a function consists of unions, products and complements of n Boolean variables $x_1, x_2, \ldots x_n$ (e.g. x_i may be the indicator of a proposition and it can be 0 or 1). We call this function a Boolean function of the variables $x_1, x_2, \ldots x_n$ and we write $\phi(x_1, x_2, \ldots, x_n)$. When we consider the Boolean algebra of 0 and 1, it is clear that $\phi(x_1, \ldots, x_n)$ will also take the values of its variables. In this case ϕ is called a switching (or structure) function. It maps an n-dimensional vector $\underline{x} \equiv (x_1, \ldots x_n)$ of 0's and 1's onto 0 or 1 (see Refs. 104 and 105).

A theorem that is very useful in reliability studies concerns an expansion of a switching function as follows

$$\phi(x_1, x_2,\ldots,x_i,\ldots,x_n) = \phi(x_1, x_2,\ldots,1,\ldots,x_n)x_i + \\ + \phi(x_1, x_2,\ldots,0,\ldots,x_n)\bar{x}_i \qquad (3.22)$$

It states that the switching function is equal to the union of two products: the first is the product of one of the variables times the switching function with the variable assumed true and the second is the product of the complement of the same variable times the switching function with the variable assumed false.

A fundamental product of n variables is a product containing all of them complemented or not (but a variable cannot appear together with its complement in the product). For n variables there are 2^n such products, e.g., for n = 3 we have

$$x_1x_2x_3,\ x_1x_2\bar{x}_3,\ x_1\bar{x}_2x_3,\ x_1\bar{x}_2\bar{x}_3,\ \bar{x}_1x_2x_3,\ \bar{x}_1x_2\bar{x}_3,\ \bar{x}_1\bar{x}_2x_3,\ \bar{x}_1\bar{x}_2\bar{x}_3 \qquad (3.23)$$

that is, $2^3 = 8$ fundamental products. Clearly a fundamental product is 1 if and only if all its variables are 1. An important theorem is that a switching function can be written uniquely as the union of the fundamental products which correspond to the combinations of the variables which render the function true (i.e., ϕ takes the value 1). This is called the canonical expansion or disjunctive normal form of ϕ. For example, the switching function of a 2-out-of-3 system is expanded as (using 3.23).

$$\phi(x_1,x_2,x_3) = x_1x_2x_3 + x_1x_2\bar{x}_3 + x_1\bar{x}_2x_3 + \bar{x}_1x_2x_3 \qquad (3.24)$$

As another example, consider the gates of Fig. 3.3. The fundamental products are

$$x_1x_2,\ x_1\bar{x}_2,\ \bar{x}_1x_2,\ \bar{x}_1\bar{x}_2 \qquad (3.25)$$

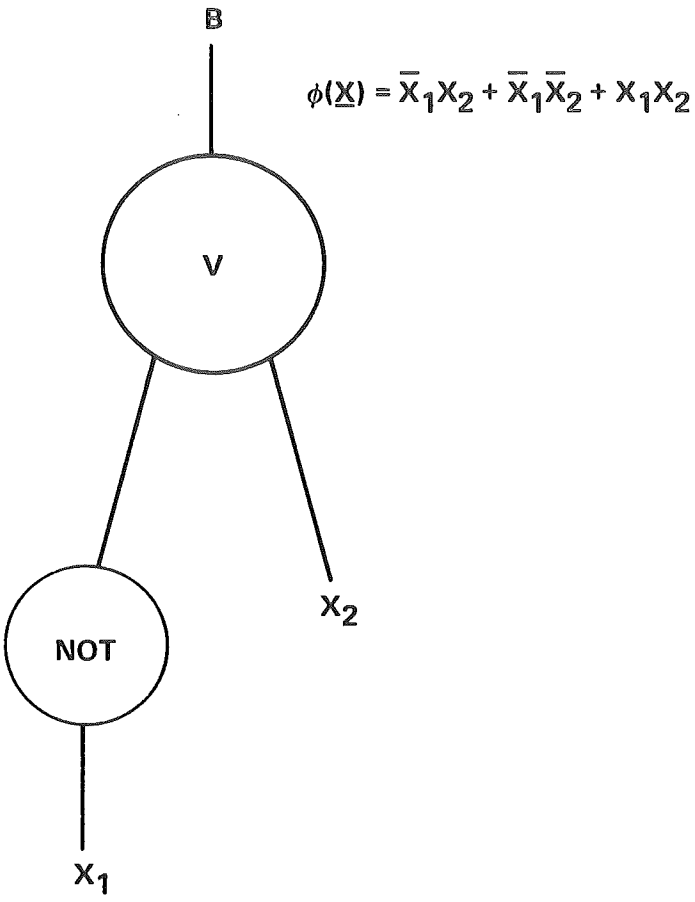


Figure 3.3. Switching Function of a Logic Diagram Involving a NOT Gate.

and the switching function is

$$\phi(\underline{x}) = x_1x_2 + \bar{x}_1x_2 + \bar{x}_1\bar{x}_2 \tag{3.26}$$

The expansions (3.24) and (3.26) can be simplified further as follows. We observe that in Eq. (3.24) the first two terms have the product x_1x_2 in common while the remaining part (x_3) of the first product appears complemented in the second. Using the distributive law, Eq. (3.12), we get

$$x_1x_2x_3 + x_1x_2\bar{x}_3 = x_1x_2(x_3 + \bar{x}_3) = x_1x_2$$

Repeating the above steps we finally reduce Eq. (3.24) to

$$\phi(x_1, x_2, x_3) = x_1x_2 + x_2x_3 + x_3x_1 \tag{3.27}$$

Similarly Eq. (3.26) can be written as

$$\phi(x_1, x_2) = \bar{x}_1 + x_2 \tag{3.28}$$

It is clear that no further simplification of the products (which, of course, are not fundamental any more) can be achieved. Each product represents the minimum number of propositions, which, if true, render ϕ true, e.g., in Eq. (3.27) if x_1 and x_2 are true then ϕ is true (1), and in Eq. (3.28) if $\bar{x}_1$ is true then ϕ is true. This discussion introduces the notion of a minimal path set which is the minimal set of the variables and/or their complements, which by being true cause the switching function to be true.

There is an important difference between the switching functions of Eqs. (3.27) and (3.28). In the first no complements of the variables appear, while in the second $\bar{x}_1$ alone is capable of yielding ϕ true. This is the consequence of the presence of a NOT gate in Fig. 3.3. Logical structures of the type of Eq. (3.27) are very common in applications and are called coherent or monotonic structures.[5,106,107] Their basic feature is that if a variable takes on the value 1 it can only contribute to the truthfulness of the switching

function (which, of course, is not the case with the variable x_1 of Fig. 3.3). More formally a coherent structure function has the following properties:

$$\phi(\underline{x}) = 1, \quad \text{if} \quad \underline{x} = (1, 1, \ldots, 1) \tag{3.29}$$

i.e., if all the variables are true, ϕ is true,

$$\phi(\underline{x}) = 0, \quad \text{if} \quad \underline{x} = (0, 0, \ldots, 0) \tag{3.30}$$

i.e., if all the variables are false, ϕ is false,

$$\phi(\underline{x}) \geq \phi(\underline{y}), \quad \text{if} \quad x_i \geq y_i \quad \text{for all } i \tag{3.31}$$

i.e., if a variable is false and it becomes true, this cannot cause the switching function to become 0, it either keeps its initial value, or it becomes 1.

For coherent structures it is easier to visualize the meaning of a path set: it is a set of the variables, which by being 1 (true) yield the structure function true. A minimal path set is a path set which does not have another path set as a subset. Thus, for a 2-out-of-3 system, the set $\{x_1,x_2,x_3\}$ is a path set but not a minimal path set, since $\{x_1,x_2\}$ is a path set also.

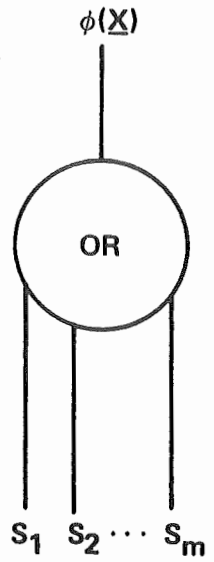
If the variables $x_{i_1}, x_{i_2} \ldots x_{i_k}$ form a minimal path set, it is convenient to define their product

$$s_i = x_{i_1} x_{i_2} \ldots x_{i_k} \tag{3.32}$$

and use these products to write the structure function as

$$\phi(\underline{x}) = 1 - \prod_{i=1}^{m} (1-s_i) \tag{3.33}$$

This equation simply expresses the fact that if any of the m minimal path sets is true, then the structure function is true (the symbol $\prod$ means the product of the Boolean variables). In this manner we have reduced the logical structure to an OR gate with inputs the minimal path sets (Fig. 3.4).



$$\phi = 1 - \prod_{i=1}^{m} (1-S_i)$$

Figure 3.4. Representation of a Structure Function as the Union of Minimal Path Sets.

Another simple representation of the structure function can be achieved with the use of minimal cut sets. A cut set is a set of variables which by being false cause $\phi(\underline{x})$ to be false. A minimal cut set does not contain another cut set. By taking one variable from each minimal path set, we can form a minimal cut set. Thus, for the 2-out-of-3 structure the minimal cut sets are $\{x_1,x_2\}$, $\{x_2,x_3\}$ and $\{x_3,x_1\}$. Since all the variables $x_{i_1}, x_{i_2},\ldots,x_{i_r}$ of a minimal cut set must be 0 for ϕ to be zero, we define the variable

$$C_i = 1 - \prod_{j=1}^{r}(1-x_{i_j}) \tag{3.34}$$

and we may represent the structure function as

$$\phi(\underline{x}) = \prod_{i=1}^{k} C_i \tag{3.35}$$

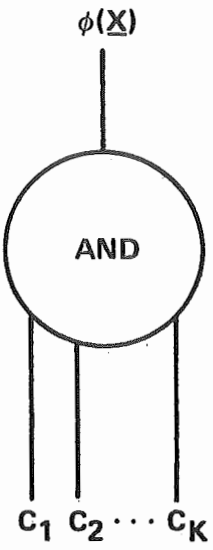
Eq. (3.35) states that if any of the minimal cut sets is zero, then ϕ is zero. The structure function, therefore, is reduced to an AND gate, as shown in Fig. (3.5).

The minimal path (cut) sets are extremely useful for the analysis of complex structures, since they deal with the critical combinations of the variables, that can yield the switching function true (false).

3.A.3 Reliability Diagrams and Fault Trees

Two of the most useful logic diagrams are the reliability block diagram and the fault tree.

The reliability diagram shows the functional relationships of the components of a system which is intended to accomplish a specified function between two points A (start) and B (finish). The series and parallel systems discussed in Sections 2.C.2 and 2.C.3 (Fig. 2.24 and 2.25) are obviously such diagrams. In practice most systems can be depicted as combinations of elements in series and parallel (coherent structures).



$$\phi(\underline{X}) = \prod_{i=1}^{K} C_i$$

Figure 3.5. Representation of a Structure Function as the Product of Minimal Cut Sets.

A simple example is the following: an experiment is conducted in a room in which the temperature must be kept within specified limits, otherwise the experiment should stop. Three sensors (thermocouples) monitor the room temperature and their outputs are connected to an indicator light. If the temperature is within the acceptable limits the light is off, otherwise the light is on and an operator proceeds to stop the experiment. To avoid false signals it is decided that at least two out of the three sensors must give an output for the light to be on (i.e., 2-out-of-3 logic is used). Denoting the sensors as S_1, S_2, S_3 and the light L, the reliability diagram is shown in Fig. 3.6. In 3.6.a and 3.6.b the diagram is drawn using series and parallel combinations. The use of the 2-out-of-3 gate in 3.6.c simplifies the diagram in that every component appears only once.

The switching function ϕ of this logical configuration is readily found to be

$$\phi = L(S_1S_2 + S_2S_3 + S_3S_1) \tag{3.36}$$

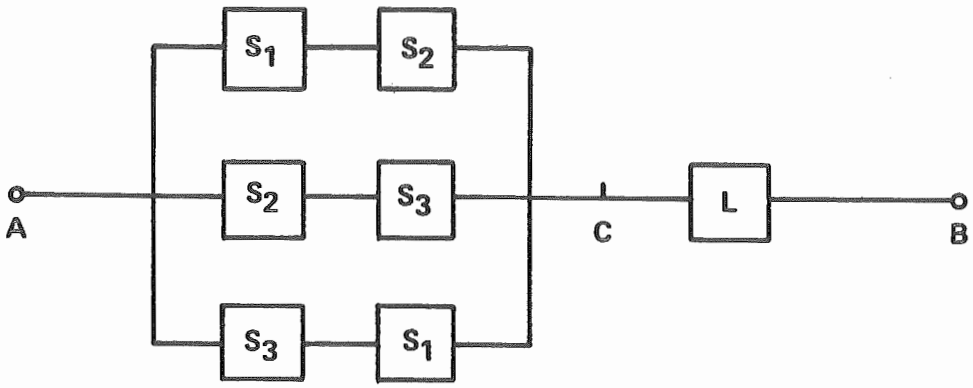
and the minimal path sets are

$$\{S_1,S_2,L\},\ \{S_2,S_3,L\},\ \{S_3,S_1,L\} \tag{3.37}$$

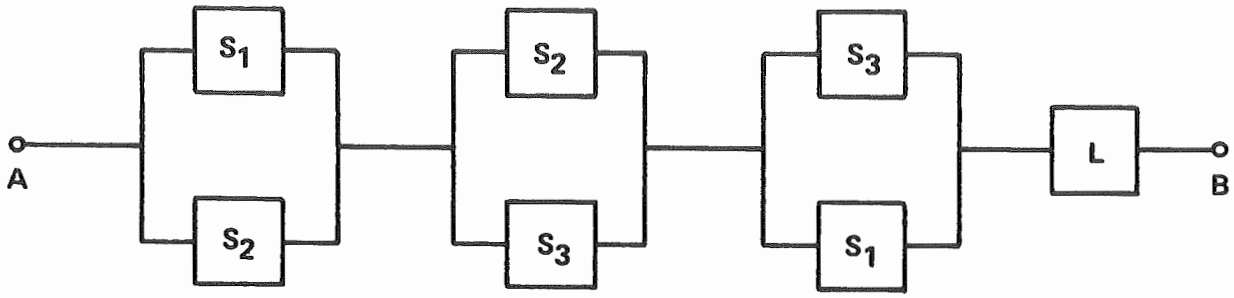
The minimal cut sets are

$$\{L\},\ \{S_1,S_2\},\ \{S_2,S_3\},\ \{S_3,S_1\} \tag{3.38}$$

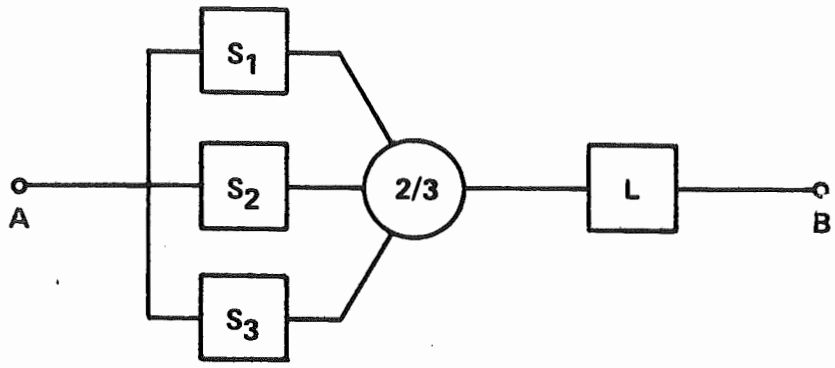
The logic diagram can be redrawn in the form of a tree if we define a top event (or proposition) as system success meaning that the temperature is correctly monitored by at least two of the sensors and the light is turned on. The tree is shown in Fig. 3.7 and it is equivalent to the block diagrams of Fig. 3.6. Using the duality principle we change the top event into SYSTEM FAILURE, interchange unions and intersections and replace the inputs by their complements. The resulting tree is shown in Fig. 3.8.



(a)



(b)



(c)

Figure 3.6. Different Forms of a Reliability Diagram.

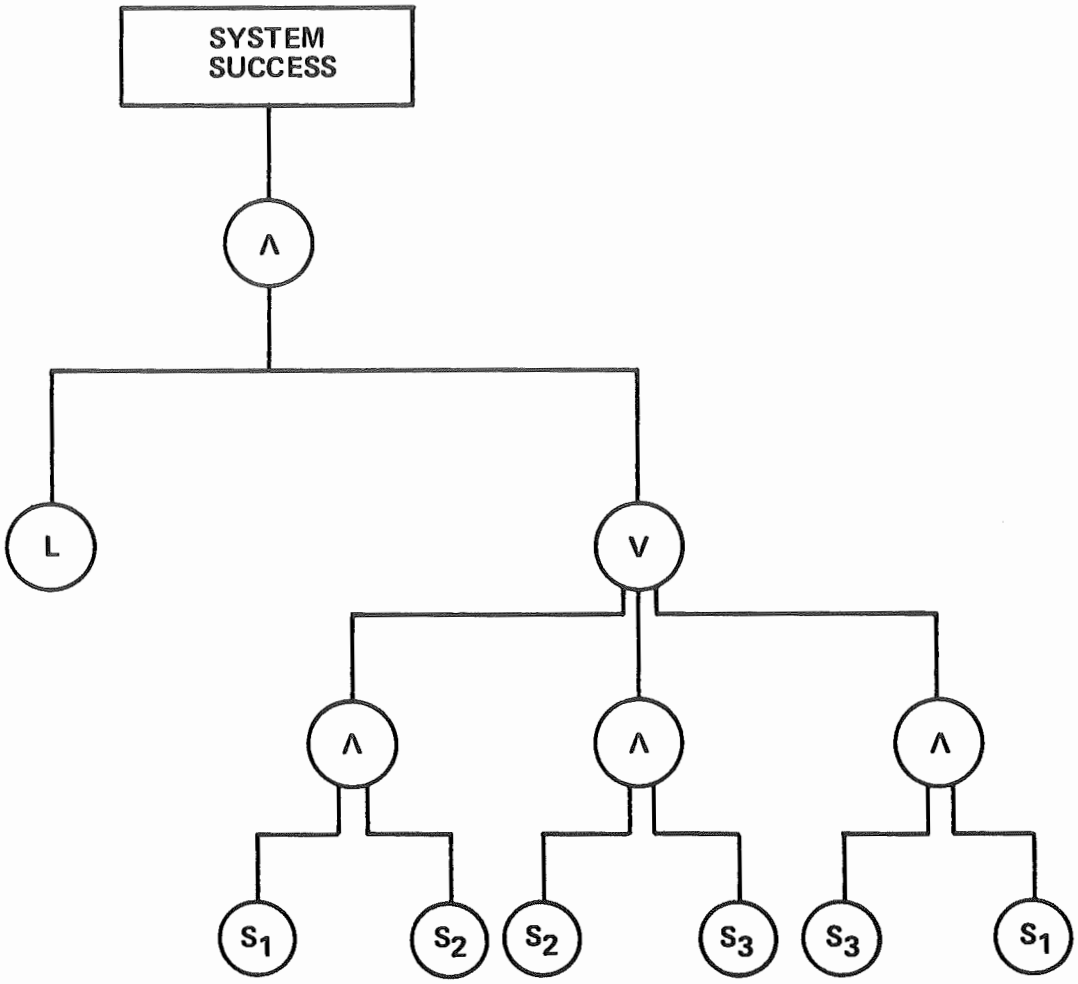


Figure 3.7. A Logic Diagram in a Tree-Form.

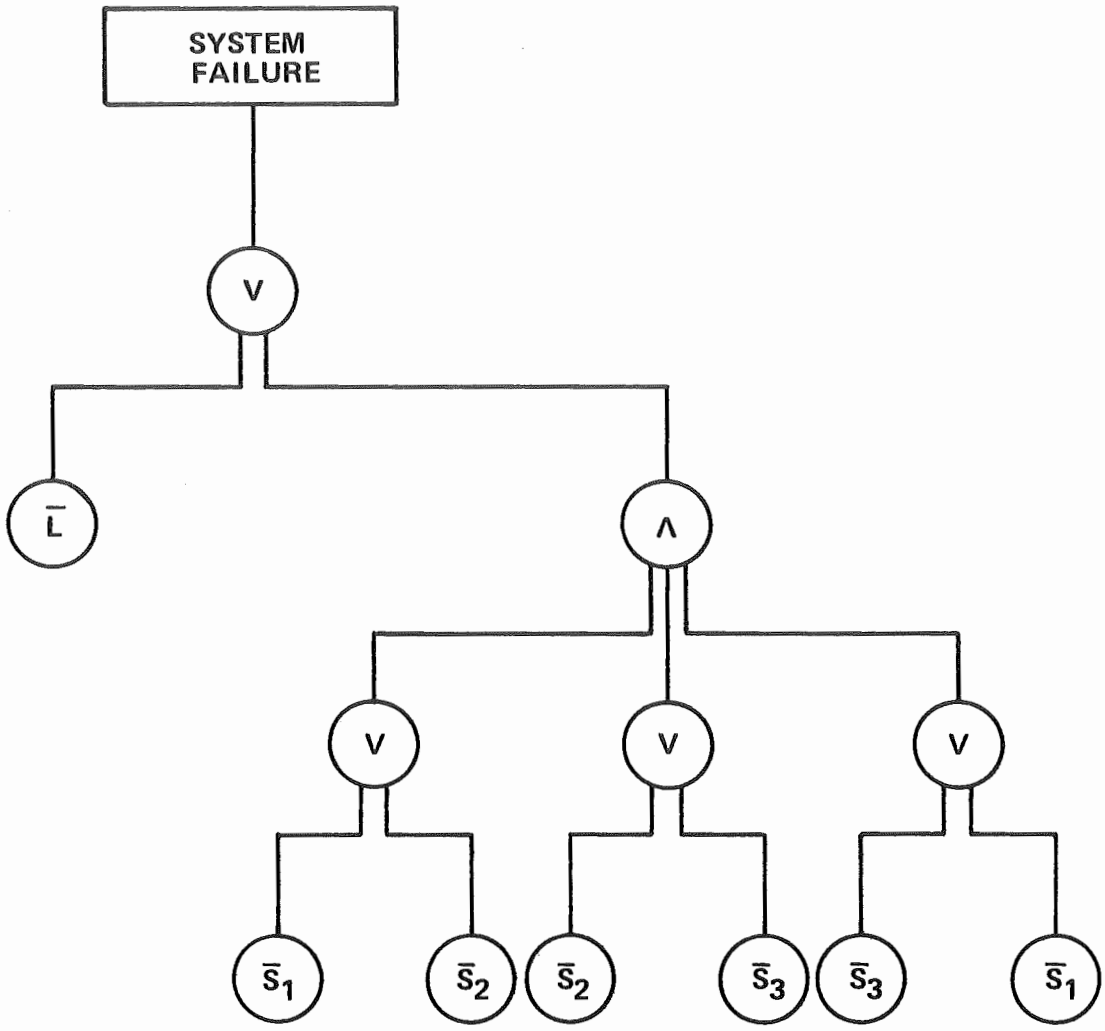


Figure 3.8. The Dual Form of the Tree of Figure 3.7.

The common feature of all the previous diagrams is that they depict the logical interconnection of the hardware of the system under study. In a safety study, however, there are additional factors which influence the performance of the system. These may be incorporated in the tree diagram and since it is more convenient to work with failures than successes we use the tree which has as top event a specified failure. Thus the fault tree approach results which has found wide applicability in safety analyses.

For a formal application of the method special symbols are used in addition to the logic gates (Fig. 3.9 and 3.10).

The unfavorable event which is put in the top is TEMPERATURE OUTSIDE LIMITS NOT DETECTED and it is analyzed as shown in Fig. 3.11. Notice that all possible causes of failure to monitor the unacceptable temperature are included. In fact the fault tree shows that failure of the operator alone to notice the light may cause the top event to occur, even though the hardware of the system function properly. Another detailed construction of a fault tree is given by Haasl (Ref. 108) in one of the early papers on the method.

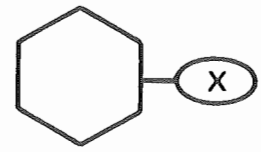
It is already apparent that the construction of the tree requires an intimate knowledge of the system and its environment. The fault tree of Fig. 3.11 is only one of the trees that might be drawn for the above example; in a real situation there may be other additional factors which would be deemed important enough to be incorporated in the tree. An attempt to develop a formal methodology for fault tree construction was made by Fussell (Ref. 109). It is limited to electrical systems and considers only hardware failures. The modes of failure of each component are stored in a library from which a program called DRAFT draws the appropriate primary inputs that can lead to the top event. To make the method systematic a whole set of definitions and classifications of events and other conditions is introduced, so that after a

THE EVENT X HAS BEEN ANALYZED TO ITS CAUSES AND IS STATED ONLY FOR CONVENIENCE IN READING THE FAULT TREE

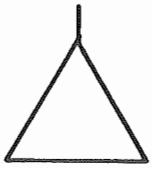
X HAS NOT BEEN ANALYZED TEMPORARILY BUT IT WILL WHEN THE FAULT TREE IS COMPLETED

X IS A PRIMARY INPUT EVENT WHICH NEEDS NO FURTHER ANALYSIS SINCE ITS PROBABILISTIC CHARACTERISTICS ARE KNOWN

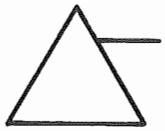
Figure 3.9. Fault Tree Symbols.



INHIBIT GATE: ITS OUTPUT IS PRODUCED BY THE (SINGLE) INPUT IF THE CONDITION X IS SATISFIED



TRANSFER-IN SYMBOL



TRANSFER-OUT SYMBOL. THE TRANSFER SYMBOLS CONNECT PARTS OF THE TREE DRAWN IN DIFFERENT POSITIONS.

Figure 3.10. Fault Tree Symbols.

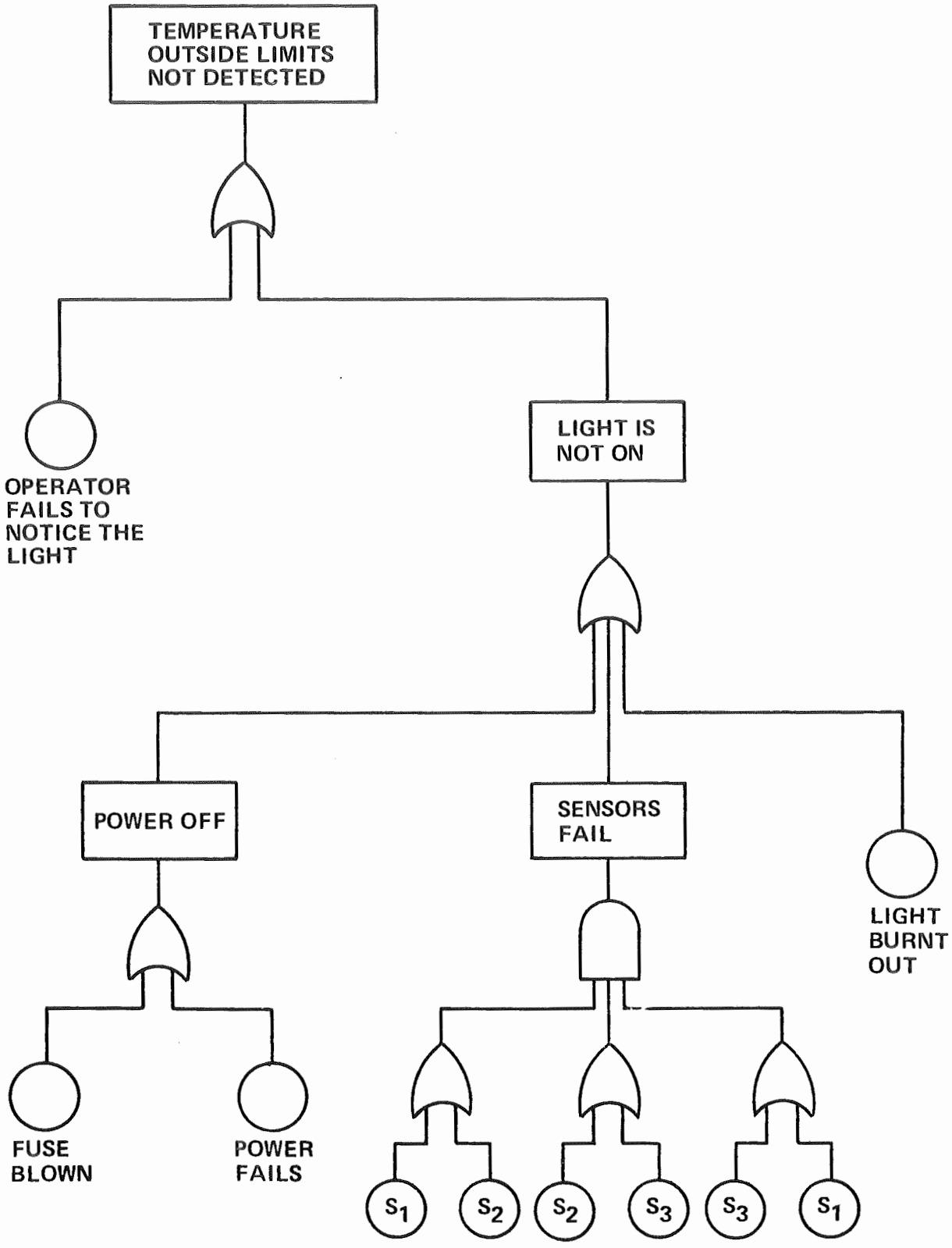


Figure 3.11. Fault Tree Example.

certain stage in the analysis of the top event has been reached, the program proceeds to complete the tree by piecing together the primary inputs.

3.A.4 Probability Relations

The logic diagrams form the basis upon which a probability analysis can be carried out. From the diagram, the laws of probability theory and the primary input information certain useful probabilistic quantities for the system and its subsystems can be calculated. A systematic approach is presented by Murchland in Ref. 110, while Refs. 111 and 112 contain the basic ideas and expressions.

The logic diagram (either a reliability block diagram or a fault tree) represents logical relations of propositions or events, as they are commonly called. We have agreed in Sec. 3.A.2 to assign the truth value (indicator variable) 1 to a true proposition and the truth value 0 to a false proposition The expected value of the indicator variable x_i of a proposition is the probability that the proposition is true, i.e.,

$$p_i = E[X_i] = p[x_i=1] \tag{3.39}$$

For a success diagram then the proposition will express successful performance and $p_i(t)$ will be the reliability or availability (if a maintenance scheme is employed) of the input. For example, in the block diagrams of Fig. 3.6 and in the tree diagram of Fig. 3.7 the input L corresponds to the statement "light is turned on". In a fault tree the propositions express failures and $p_i(t)$ will be the unreliability or the unavailability of the input. The inputs "fuse blown", "power fails" etc. of Fig. 3.11 serve as examples. To avoid confusion we will denote the availability as $p_i(t)$ and the unavailability as $q_i(t)$.

If the primary inputs are under a repair policy another quantity that is needed is the failure intensity $w_f(t)$ (which is a renewal density in the sense of Section 2.D.3). Its definition is

$w_f(t)dt$ = probability that a failure occurs in the interval $(t,t+dt)$

Similarly we define the repair intensity $w_r(t)$ (see Section 2.D.3).

An item or vertex is a generic term which refers to the whole system, a subsystem or a primary input. In Fig. 3.6.a the components between A and C can be referred to as an item and in the fault tree of Fig. 3.11 the top event and the events "Light is not on", "power off" etc. are items. For an item two fundamental relations are true. The first relates its unavailability $q(t)$, its expected number of failures $W_f(t)$ and its expected number of repairs $W_r(t)$ both evaluated over the interval $(0,t)$; this relation is

$$q(t) = W_f(t) - W_r(t) \tag{3.40}$$

given that initially

$$q(0) = W_f(0) - W_r(0)$$

The proof of Eq. (3.40) follows if we notice that

$$\text{(exact number of failures) - (exact number of repairs)} = \begin{cases} 1 & \text{if the item is failed at } t \\ 0 & \text{if the item is good at } t. \end{cases}$$

Taking the expectations of both sides of this equation we derive Eq. (3.40). For the availability the corresponding relation is, of course,

$$p(t) = 1 - [W_f(t) - W_r(t)] \tag{3.41}$$

which is Eq. (2.169) for a component.

The quantities $w_f(t)$ and $w_r(t)$ are evaluated by

$$W_f(t) = \int_0^t w_f(\tau)d\tau + q(0) \tag{3.42}$$

and

$$W_r(t) = \int_0^t w_r(\tau)\,d\tau \tag{3.43}$$

The second relation refers to the interval unreliability, which is defined as

$U(s,t) \equiv$ probability that the item is failed at s or, if it is good at s, it will fail before t.

For the unreliability upper and lower bounds are given as follows

$$\max_{s \le a \le t} q(\alpha) \le U(s,t) \le \min \{q(s) + W_f(t) - W_f(s),\ 1\} \tag{3.44}$$

For the interval (0,t) this relation becomes

$$\max_{0 \le a \le t} q(a) \le U(0,t) \le \min \{W_f(t),\ 1\} \tag{3.45}$$

For non-repairable items we have

$$U(s,t) = q(t) \tag{3.46}$$

The proof can be found in Ref. 110.

The above relations, though very useful, do not indicate how the quantities appearing are calculated from the primary input information. A fundamental assumption for such an analysis is that the performance of each primary input is statistically independent of that of the other inputs. This assumption forbids, for example, the presence of two components in parallel with one repairman. However in such special cases this subsystem can be analyzed with the methods of Sections 2.D.4, 2.D.5 or 2.D.6 and the resulting $q(t)$ and $w_f(t)$ of the subsystem are used as representative functions of a single input to the logic diagram.

A probability analysis is possible if the switching function of the item is available. Then we utilize Eq. (3.39) and we take the expectation value of $\phi(x)$, or we use elementary probability laws to find the unavailability (or

availability) of the item in terms of the q's (or p's) of its predecessors. We recall that for an AND gate (items in series) the structure function is

$$B = A_1 A_2 \ldots A_n$$

and the probability of this intersection of independent events is

$$q_B(t) = \prod_{i=1}^{n} q_{A_i}(t) \tag{3.47}$$

For an OR gate (items in parallel) the switching function is

$$B = A_1 + A_2 + \ldots + A_n$$

and the probability is found from Eq. (2.2) which we rewrite here

$$q_B(t) = \sum_{j=1}^{n} q_{A_j}(t) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} q_{A_i A_j}(t) +$$

$$+ \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=i+2}^{n} q_{A_i A_j A_k}(t) - \ldots + (-1)^{n+1} q_{A_1 A_2 \ldots A_n}(t) \tag{3.48}$$

For two events the formula gives

$$q_B(t) = q_{A_1}(t) + q_{A_2}(t) - q_{A_1 A_2}(t) \tag{3.49}$$

For more than two inputs Eq. (3.48) is quite complicated, but useful bounds can be estimated as

$$q_B(t) \leq \sum_{j=1}^{n} q_{A_j}(t) \tag{3.50}$$

$$q_B(t) \geq \sum_{j=1}^{n} q_{A_j}(t) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} q_{A_i A_j}(t) \tag{3.51}$$

etc.

Finally for a NOT gate (complement of an event) we have

$$B = \overline{A}$$

and $$q_B(t) = 1 - q_A(t) \tag{3.52}$$

Consider as an example an item which is a 2-out-of-3 system. Its switching function is given by Eq. (3.27) and applying the above rules we find for the probabilities

$$q(t) = q_1(t)q_2(t) + q_2(t)q_3(t) + q_3(t)q_1(t) - 2q_1(t)q_2(t)q_3(t) \qquad (3.53)$$

The same result is obtained if we start with the form of the switching function given in Eq. (3.24) and expand the complements as in Eq. (3.52).

The switching function can be found by the methods of Section 3.A.2. We start from the items of least complexity and proceed to build the system switching function (or directly the probability expression). Special care is required when one item appears more than once in the structure (or, equivalently, if an item has two or more successors in a tree diagram). In cases like a 2-out-of-3 system the subsystem is analyzed separately and its output is added as a single input to the whole system. Another approach is to use the expansion of Eq. (3.22). Logical structures in which all items have only one successor are called simple and obviously their hierarchical structure makes them easier to analyze. For detailed applications of this method of successive reduction of the diagram see Refs. 113 and 114. For large block diagrams the British program NOTED (Ref. 115) utilizes the method of reduction to estimate the unavailability. Various distributions for the inputs are available (exponential, normal, log-normal, Weibull) and repair and periodic inspection can be included.

For every item the unavailability q is a polynomial of the unavailabilities q_i of each predecessor; each q_i appears in the first power in the polynomial (this is clear from the way the item unavailability was constructed). Therefore, we can write

$$q = q(q_1, q_2, \ldots, q_n) \qquad (3.54)$$

and for the partial derivatives

$$\frac{\partial q}{\partial q_j} = \text{independent of } q_j.$$

For coherent structures the following is true (compare with Eqs. (3.29), (3.30) and (3.31))

$$q(0,0,\ldots,0) = 0 \tag{3.55}$$

(if all predecessors are good the item unavailability is zero)

$$q(1,1,\ldots,1) = 0 \tag{3.56}$$

(if all predecessors are failed the item is failed also)

$$\frac{\partial q}{\partial q_j} \geq 0 \tag{3.57}$$

(if the unavailability of a predecessor increases the item unavailability cannot decrease; as a result NOT gates cannot be present).

These introductory comments will now be used to derive an expression for the failure intensity of the item $w_f(t)$ in terms of the failure intensities of its predecessors $w_{f,i}(t)$. A further assumption is introduced here which states that in the interval (t, t + dt) only one component can fail, the probability of two or more failing being of second or higher order in dt. Under this assumption and for coherent structures it is proven in Ref. 110 that

$$w_f(t) = \sum_{\substack{\text{all} \\ \text{predecessors}}} w_{f,i}(t) \frac{\partial q(t)}{\partial q_i(t)} \tag{3.58}$$

Notice that $\frac{\partial q}{\partial q_i}$ can be interpreted as the probability that the item will be failed, if the predecessor i fails.

A similar relation exists for the repair intensity, i.e.,

$$w_r(t) = \sum_{\substack{\text{all} \\ \text{predecessors}}} w_{r,i}(t) \frac{\partial p(t)}{\partial p_i(t)} \tag{3.59}$$

The expected number of failures or repairs in (0,t) can be found by integrating Eqs. (3.58) and (3.59) as Eqs. (3.42) and (3.43) indicate.

Several simple examples will illustrate the use of Eq. (3.58):

AND gate (series system), using Eq. (3.47),

$$w_{f,B}(t) = \sum_i w_{f,A_i}(t) \prod_{\substack{j \\ j\neq i}} q_{A_j}(t) \tag{3.60}$$

OR gate (parallel system, using Eq. (3.48),

$$\frac{\partial q_B(t)}{\partial q_{A_i}} = 1 - \sum_{\substack{j=1 \\ j\neq i}}^{n} q_{A_j}(t) + \sum_{\substack{j=1 \\ j\neq i}}^{n-1} \sum_{\substack{k=j+1 \\ k\neq i}}^{n} q_{A_j A_k}(t) - \ldots (-1)^{n+1} q \prod_{i\neq j} A_i(t) \tag{3.61}$$

therefore for an OR gate with two inputs

$$w_{f,B}(t) = w_{f,A_1}(t)\, p_{A_2}(t) + w_{f,A_2}(t)\, p_{A_1}(t) \tag{3.62}$$

2-out-of-3 system, using Eq. (3.53),

$$w_f = (q_2+q_3-q_2q_3)w_{f,1} + (q_1+q_3-q_1q_3)w_{f,2} + (q_1+q_2-q_1q_2)w_{f,3} \tag{3.63}$$

The relations developed above assume that the probabilistic characteristics of the primary inputs are known exactly. As discussed in Section 2.B.5 though the parameters of the distributions are estimated from tests either as point estimates or in an interval with a certain confidence. In the latter case the question arises how the uncertainty about the primary input information is carried over to the probabilistic treatment of a complex logical structure. Rosenblatt (Ref. 125) gives a good introduction to this problem of confidence bounds for complex structures and the book by Mann, Schafer and Singpurwalla (Ref. 126) surveys the statistical techniques employed as well as an extensive list of references.

Murchland and Weber (Ref. 111) utilize simple statistical relations to estimate the mean and the variance of the unavailability of a coherent structure

given the corresponding quantities for the primary inputs. From the mean and variance a conservative confidence bound is calculated for the unavailability with the use of Tchebycheff's inequality, Eq. (2.9), which we reproduce here

$$P[m-k\sigma < X < m + k\sigma] \geq 1 - \frac{1}{k^2} \tag{3.64}$$

As it has been stated before, the unavailability of each item is a linear polynomial of the unavailabilities of its predecessors. Thus the mean of the item unavailability is found by inserting in the polynomial the average unavailabilities of the predecessors. For the variance we expand the polynomial about the mean values of the variables according to the multinomial theorem, square it and we take expectations. For example, the output of an AND gate with two inputs has the polynomial (Eq. (3.47))

$$q(t) = q_1(t)q_2(t) \tag{3.65}$$

The relation for the means is

$$\bar{q}(t) = \bar{q}_1(t)\bar{q}_2(t) \tag{3.66}$$

To find the variance we write

$$q(t) - \bar{q}(t) = \bar{q}_2(q_1-\bar{q}_1) + q_1(q_2-\bar{q}_2) + (q_1-\bar{q}_1)(q_2-\bar{q}_2) \tag{3.67}$$

and upon squaring and taking averages the cross products vanish the final result being

$$\sigma^2 = \sigma_1^2\,\sigma_2^2 + \bar{q}_2^2\,\sigma_1^2 + \bar{q}_1^2\,\sigma_2^2 \tag{3.68}$$

Similarly for an OR gate we find

$$\bar{q} = \bar{q}_1 + \bar{q}_2 - \bar{q}_1\bar{q}_2 \tag{3.69}$$

and

$$\sigma^2 = \sigma_1^2\,\sigma_2^2 + (1-\bar{q}_2)^2\,\sigma_1^2 + (1-\bar{q}_1)^2\,\sigma_2^2 \tag{3.70}$$

Therefore, starting from the bottom of a tree and proceeding to the top with the use of the above relations the mean and variance are calculated. The method can be applied to simple trees with independent inputs which are either nonmaintained or in the steady state. When the components are maintained the authors discuss various methods of approach and in the case of repairable inputs they indicate how the mean and variance of the failure intensity can be calculated. The method is essentially the same as above.

3.A.5 Solution Via Cut Sets.

The notion of minimal cut sets (mcs) introduced in Section 3.A.2 and the probability relations of Section 3.A.4 can be combined to provide with a systematic and economical method of analyzing the probabilistic behavior of a logic diagram. The method consists of identifying the mcs's, deriving the unavailability and failure intensity expressions for each mcs from those of the primary inputs and finally the corresponding quantities for the top event are determined from those of the mcs's. It should be noted that in practice we talk about cut sets whenever we study failures; strictly speaking, in a fault tree the combinations of components which can cause the top event to occur should be called path sets, since they render the top event true. We will follow the common usage. Furthermore, we will refer to logic diagrams in the form of trees, since any logic diagram can in general, be represented by a tree. All the structures are assumed coherent, i.e., only AND and OR gates are allowed.

Identification of Minimal Cut Sets.

For trees with relatively few inputs the mcs can be identified by inspection. Most often, however, such an approach is very inefficient, if possible at all, since the number of mcs increases very rapidly, as the complexity of the tree increases. There exist several approaches utilizing the computer as follows:

1. Deterministic and Monte Carlo methods (PREP code, Ref. 116)

The mcs's are identified either deterministically or by Monte Carlo simulation. In the deterministic method each input is failed individually and if the top event occurs that component is a mcs. The procedure is continued with the components failing in combinations of two, three etc. Each time the cut set is checked whether it contains another cut set in which case it is rejected. The method is reliable but it takes long computer times for large trees and usually it is stopped when the mcs's with one and two components have been found (if there are n primary inputs then the number of their combinations taken k at a time is $\binom{n}{k}$. For k = 3 and n = 500 and assuming that it takes the computer 10^{-5} min. to check each combination, the required time to check all of them is of the order of 10^{3} minutes).

In Monte Carlo simulation the components are failed randomly with times of failure chosen from their failure distribution. If, for example, the ith component has an exponential distribution, a random number r ($0 \leq r \leq 1$) is generated and the time of failure t_i is computed from

$$r = \frac{1-e^{-\lambda_i t_i}}{1-e^{-\lambda_i T}} \quad (T = \text{mission length}, \; t_i \leq T) \tag{3.74}$$

Thus a set of times-to-failure is obtained for the components which are ordered in increasing order, i.e.,

$$t_1 < t_2 < \ldots < t_n < T$$

The components are failed successively starting from the one with the smallest time to failure until the top event occurs. These components then form a cut set which is again tested against the already found cut sets, so that only the mcs's will be determined. This procedure identifies first the mcs's which are most important for the system, since the use of Eq. (3.71) insures that

components with high failure rate are failed more frequently. By selecting T to be very small the times-to-failure are chosen from the uniform distribution thus the failure rates are unimportant (to see this we expand $r = \frac{\lambda_i t_i}{\lambda_i T} = \frac{t_i}{T}$)

The PREP code can handle trees with up to 2000 inputs and up to 2000 gates.

2. Other Methods

In Ref. 117 a method is presented, which determines the mcs's utilizing the unique factorization theorem of prime numbers (that is, every natural number can be expressed as a unique product of prime numbers). Each primary input is assigned a prime number and each mcs is represented by a unique product of prime numbers which then identifies the primary inputs of the mcs. For example, in Fig. 3.12 the events x_1, x_2, x_3 are assigned the prime numbers 2, 3, 5 respectively. Working from the bottom we have

event $A = x_1 x_2 = 2 \times 3 = 6$

event $B = x_1 + x_2 = 3 + 5$ (the summation is not carried out)

$TOP = A + B = 6 + 3 + 5$

but 3 is a factor of 6, so the latter is eliminated and we have

$TOP = 3 + 5$

therefore there are two mcs's, $\{x_2\}$ and $\{x_3\}$.

This is the basis of the program ELRAFT, which also proceeds to calculate the unavailability of the TOP event.

A method which proceeds from the TOP to the primary inputs by successively eliminating the secondary events is presented in Ref. 118. Each AND gate encountered increases the size of a cut set while an OR increases the number of cut sets.

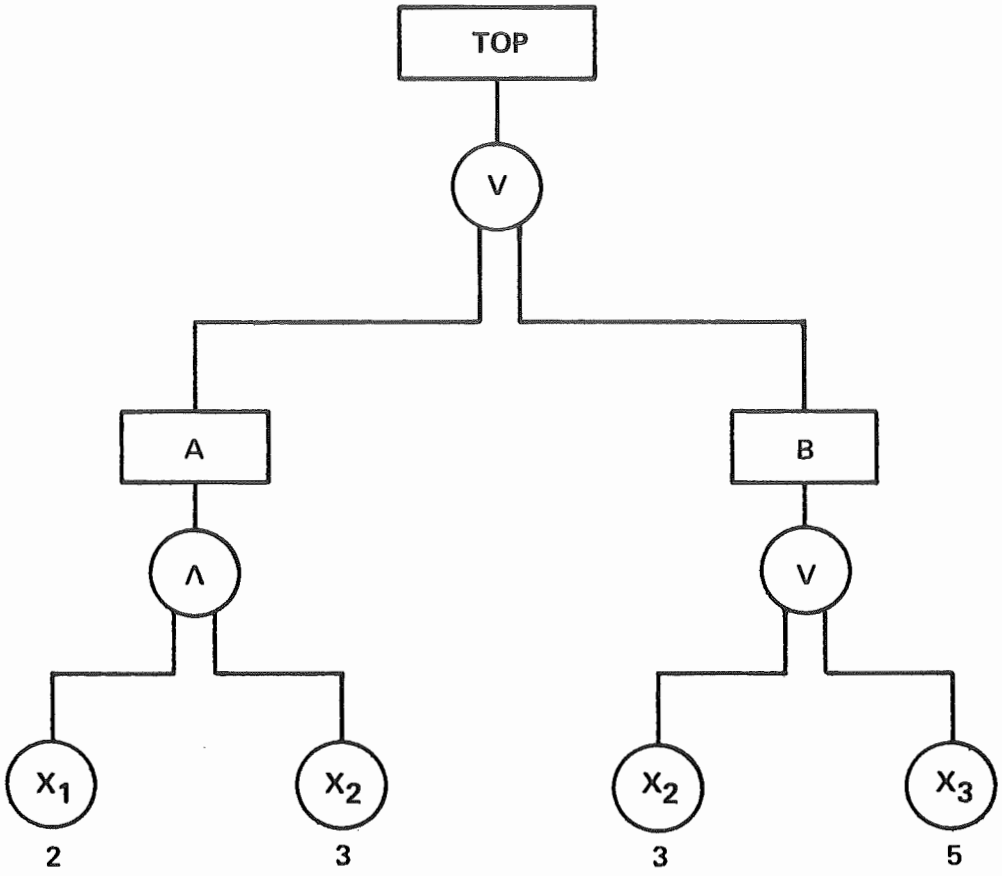


Figure 3.12. Sample Fault Tree.

Evaluation

Having identified the minimal cut sets we can proceed and evaluate the unavailability of each mcs and its failure intensity. The relations derived in Section 3.A.4 are immediately used for this task (see also Ref. 119).

Each mcs C_i, $i = 1,\ldots,k$, by its definition, is the product of all its inputs. Therefore, Eq. (3.47) for an AND gate yields the unavailability of C_i in terms of the unavailabilities of its inputs as

$$q_{C_i}(t) = \prod_{j \epsilon C_i} q_j(t) \tag{3.72}$$

(the notation $j \epsilon C_i$ indicates that the product is to be taken over all the components of C_i).

Similarly, for the failure intensity of C_i we have a result analogous to Eq. (3.60), i.e.,

$$w_{f,C_i}(t) = \sum_{j \epsilon C_j} w_{f,j}(t) \prod_{\substack{m \epsilon C_j \\ m \neq j}} q_m(t) \tag{3.73}$$

Since the top event is the union of the mcs's, its probabilistic quantities can be calculated. Eq. (3.48) is directly applicable for the unavailability, i.e.,

$$q_T(t) = \sum_{j=1}^{k} q_{C_j}(t) - \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} q_{C_i C_j}(t) + (-1)^k q_{C_1 \ldots C_k}(t) \tag{3.74}$$

The bracketing procedure of Eqs. (3.50) and (3.51) can also be applied here. Special care should be taken, when the unavailabilities of products of mcs's are calculated, since one or more components may appear in more than one mcs. In this case the theorem of conditional probabilities should be used, i.e.,

$$P(C_i C_j) = P(C_i) P(C_j / C_i) \tag{3.75}$$

An upper bound to $q_T(t)$ can be found by utilizing the expansion (3.33), which here becomes (recall that what we call here a cut set is a path set in the theory of Section 3.A.2)

$$TOP = 1 - \prod_{j=1}^{k} (1-C_j) \tag{3.76}$$

From Eq. (3.76) we get

$$q_T(t) \leq 1 - \prod_{j=1}^{k} (1-q_{C_j}(t) \tag{3.77}$$

and we reason as follows: if the mcs's did not have common components, (3.77) would be an exact equality, as it is readily seen, if the expectations of both sides of Eq. (3.76) are taken. The right-hand side of Eq. (3.77) assumes that no such common components exist, thus in the case of common components, it overestimates the unavailability of the structure, hence the inequality sign follows.

The failure intensity of the top event is given by expressions analogous to Eqs. (3.58) and (3.61). It is clear that the evaluation of the partial derivatives (Eq. (3.61)) of the unavailability polynomial (Eq. (3.74)) is quite complex. The general expression is

$$w_{f,T}(t) = \sum_{\text{all } C_i} w_{f,C_i}(t) \frac{\partial q_T(t)}{\partial q_{C_i}(t)} \tag{3.78}$$

where $w_{f,C_i}(t)$ is as given by Eq. (3.66) and

$$\frac{\partial q_T(t)}{\partial q_{C_i}(t)} = 1 - \sum_{\substack{j=1 \\ j\neq 1}}^{k} q_{C_j}(t) + \sum_{\substack{j=1 \\ j\neq i}}^{k-1} \sum_{\substack{m=j+1 \\ m\neq i}}^{k} q_{C_j C_m}(t) + (-1)^{k+1} q_{\prod_{j\neq i} C_j}(t) \tag{3.79}$$

Vesely (Ref. 119) proceeds to evaluate these terms as functions of the component unavailabilities. It is worth noticing that due to the smallness of

the unavailabilities upper and lower bounds can again be found for $w_{f,T}(t)$. An upper bound, which is usually satisfactory, is

$$w_{f,T}(t) \leq \sum_{j=1}^{k} w_{f,C_i}(t) \tag{3.80}$$

where the right-hand side is simply the sum of the failure intensities of the minimal cut sets (Eq. (3.66)).

A computer code based in the above analysis is the companion to the PREP package, KITT-1 and KITT-2 (Ref. 116).

KITT-1 calculates the unavailability and failure intensity of the minimal cut sets and the top event from those of the primary inputs. Primary failures are assumed exponential; the components may be non-repairable or repairable with either constant repair rate or fixed time-to-repair (see Sec. 2.D.3.).

KITT-2 does the same calculations but it is a multiphase code, i.e. the characteristics of the components may change (arbitrarily) at certain times. Up to 50 phases can be handled.

3.A.6 Simulation Techniques

One of first methods of estimating the unavailability of the top event in large fault trees was by simulation on a computer. The method has already been applied in Sect. 3.A.5 to find the minimal cut sets.

The idea of a Monte Carlo simulation is to generate random numbers and from these and the probabilities that the components are in a certain state (up or down) a set of random component states is generated. Then the tree logic is checked, the state of the top event is recorded and the process is repeated. Each such cycle is called a trial and their total number as well as the number of trials in which the top event occurs are used for quantitative analysis of the system. The principles of Monte Carlo simulation can be found in Refs. 8, 120 and 121.

The random numbers are generated by a computer algorithm. A frequently used method is the congruential multiplicative method. The random numbers are generated from the relation

$$x_j = \alpha\, x_{j-1} + c \quad \text{(modulo m)} \tag{3.81}$$

where m is an integer defining the period after which the numbers repeat themselves, α is a scale integer factor and c is an integer. The expression modulo m means that x_j is the remainder of the division of $(\alpha x_{j-1} + c)$ by m. The number m is chosen to be larger than the digit capacity of the computer and usually $m=2^n$, n=20 or 30. If we choose c odd and $\alpha=4k+1$ (k = integer) the period of the random numbers is m (and, as a consequence of the above choice of m, very long). Then the numbers x_j/m are uniformly distributed in the interval (0,1). As a simple example, suppose that $m = 2^3 = 8$, $c = 1$, $\alpha = 5$ and we start with $x_o = 2$. Then $5 \times 2 + 1 = 11$ and since $11 = 8 + 3$, the next number is $x_1 = 3$. Continuing this way we generate the sequence

$$x = 2,3,0,1,7,4,5,2,3, \text{ etc.},$$

with period 8. Dividing by m = 8 we scale the numbers in the interval (0,1), i.e., r=0.25, 0.375, 0, 0.125, 0.875, 0.50, 0.625, 0.25 etc.

Consider now a fault tree with non-maintained primary inputs. From the random numbers r and the cumulative failure distribution $F_j(t)$ of each input we generate a random time-to-failure for each by writing

$$r = F_j(T_j)$$

hence

$$T_j = F_j^{-1}(r) \quad . \tag{3.82}$$

If $F_j(t) = 1 - e^{-\lambda_j t}$ we have

$$T_j = - \frac{1}{\lambda_j} \ln(1-r)$$

but 1-r is also uniformly distributed in (0,1), thus we can generate T_j from

$$T_j = - \frac{1}{\lambda_j} \ln r \quad . \tag{3.83}$$

Ordering the T_j in increasing order the components are failed successively starting from the one with the smallest T and proceeding until top failure occurs or a specified time has elapsed. This ends one trial and the process is repeated from the beginning. This procedure is employed by the program SAFTE-2 (Ref. 105), which can handle up to 500 primary components with exponential failure distributions. The output yields the system MTTF and analyzes its distribution function.

If the components are repairable, in addition to random times-to-failure the program generates random repair times for each component from its repair distribution. The random times are again ordered and each component is failed and repaired according to its corresponding times, until failure of the top occurs or the mission time has elapsed. This technique is utilized in the program SAFTE-1 (Ref. 105). Failures are exponentially distributed and repairs are normally distributed. The program calculates the MTTF, MTTR and related statistical quantities for the system. Another program along the same lines was developed by Crosetti (Ref. 122). Exponential failure times are again assumed but the repair model can be either the normal distribution or the fixed-time-to-repair model. The output includes the estimate of the failure probability of the system and contributions to it from the failure of subsystems.

If the components have reached their steady-state unavailabilities $q_\infty(\lambda/\mu+\lambda$ or $\lambda\tau/1+\lambda\tau$, see Section 2.D.3) their states are determined by comparing q_∞ with r. Thus, if $r \leq q_\infty$ the component is assumed failed and repeating the comparison for all inputs a state of the system is generated and the occurrence or not of the top event is checked (SAFTE-3, Ref. 105).

When the random failure and repair times are calculated from Eq. (3.82) the process is called direct simulation. However problems regarding the number of trials arise due to the smallness of the unavailabilities related to a fault tree (Ref. 123). If the top failure has occurred n times out of N trials, its probability is estimated by

$$\hat{q} = \frac{n}{N} \quad . \tag{3.84}$$

The variance of $\hat{q}$ is

$$\hat{\sigma}^2_{\hat{q}} = \frac{\hat{q}(1-\hat{q})}{N} \tag{3.85}$$

(binomial distribution).

Suppose now that we wish to estimate $\hat{q}$ to within ±10%, meaning that the sampling will continue until the standard deviation of the estimate $\hat{q}$ is less than or equal to 10% of $\hat{q}$. This leads to

$$\sqrt{\frac{\hat{q}(1-\hat{q})}{N}} = 0.1\ \hat{q}$$

or using Eq. (3.84) and the approximation $1 - \hat{q} \cong 1$,

$$\sqrt{\frac{n}{N^2}} = 0.1\ \frac{n}{N}$$

or

$$n = 100. \tag{3.86}$$

Therefore the top failure must occur at least 100 times before the trials stop. If its probability is of the order 10^{-4} the number of trials should be $N = n/10^{-4} = 10^6$ which is prohibitive in terms of running computer time. To overcome this difficulty the technique of importance sampling (Ref. 123) is employed. The random times are not calculated from the time distributions, Eq. (3.82), but from an artificial distribution which predicts higher failure probabilities for the mission time interval. The use of weighting factors in the calculation of $\hat{q}$ describes the results and an estimate of $\hat{q}$ is obtained in shorter computer times. Some deficiencies of the method are argued in Ref. 124. The SAFTE programs and the one developed by Crosetti have utilized importance sampling.

The advantage of Monte Carlo methods over the analytical approach of Section 3.A.5 is the greater flexibility. Special logic symbols (like priority gates, NOT gates et al.) can be handled, as well as other special features of the system regarding its operation and maintenance. Furthermore uncertainties in the input data can be included and their effect on the top event probability can be analyzed (Ref. 125).

3.A.7 Applications

Block diagrams and fault trees have been used extensively in the literature of probabilistic safety and the purpose of this section is to give several references where detailed analyses are presented.

One of the early computer programs which handled block diagrams was ARMM (Automatic Reliability Mathematical Model). The program selects combinations of components which can cause system failure (i.e. cut sets) but those with a pre-specified number of components (e.g. 3 or 4) or less. Then it computes the reliability of the system (no repair is allowed) and the contribution of each failure mode and component to the system unreliability. Details and

applications to the reliability analysis of a reactor primary containment and safety injection system are given in Ref. 127.

In Ref. 128 block diagrams are employed to analyze a reactor automatic protective system. The system monitors a number of reactor parameters (e.g. temperature, pressure) and if they are within acceptable limits the sensors feed signals to the shutdown system which keeps the control rod actuators energized. In case the parameter limits are violated no signals are sent to the shutdown system and the control rods rapidly fall into the reactor. Two block diagrams are drawn for the two possible failures: the reactor is shut down while no parameter exceeds its limits (failed safely) and the reactor is not shut down while the monitored parameters are actually beyond the acceptable limits (failed dangerously). The diagrams are reduced to simpler forms by combining series and parallel elements successively and the unavailability is evaluated. In addition, critical components are identified. A study of a similar system is given in Ref. 129.

Snaith (Ref. 130) uses block diagrams to study the probability of failure per demand of an electrical supply system (Fig. 3.13). The 500 Mw(e) turbo-alternator is connected to the 400 kv metwork and it also feeds the 11 kv reactor unit board through the two unit transformers. The two bus-section switches (5 and 6) divide the board into three sections of which the two outer ones feed three primary coolant circulators (GC1, GC3, GC5 and GC2, GC4, GC6) each. If a loss of supply via the transformers occurs the underfrequency relays (1 and 2) detect it and the two gas turbines (4 and 7) start automatically feeding the two outer sections of the board with power and one gas circulator on each (GC1 and GC2) starts running again. These two circulators are sufficient for cooling of the reactor and all other connections with the two sections of the board are removed. The logic diagram for the successful performance of the

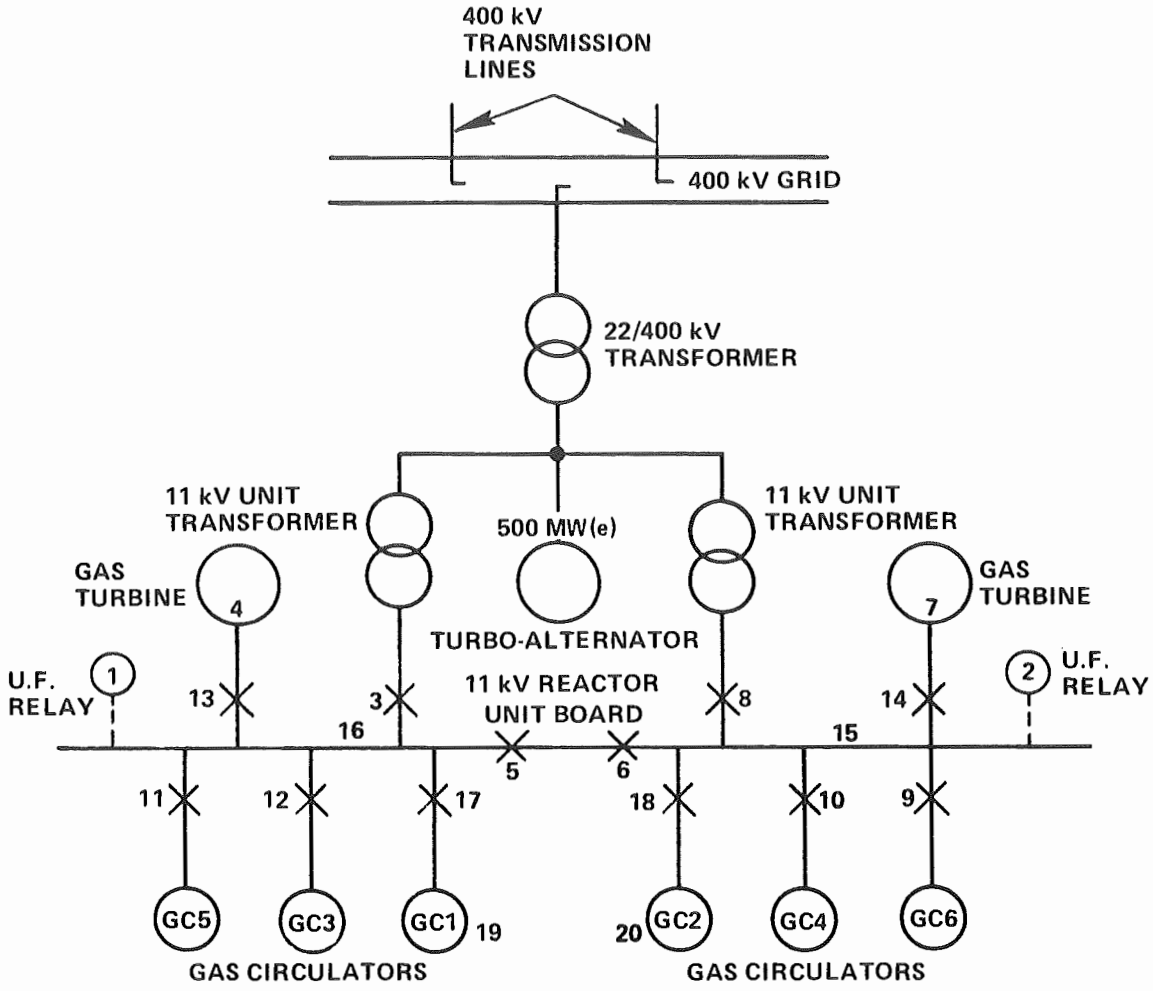


Figure 3.13. Diagram of Electrical Supply System. (Ref. 130)

system is shown in Fig. 3.14. The component data are presented in Table 3.3. Since the maintenance period is one year a calculation was carried out to estimate how the probability that the system will fail when needed (per demand) increases between two inspections. This was done with the use of the program NOTED and the results are shown in Table 3.4. A detailed study of an emergency core cooling system using the same approach as in the previous example is presented in Ref. 131.

The deductive logic of the fault tree has made it a useful and popular tool in various fields. The aircraft industry was one of the first to utilize fault trees in safety studies. Feutz and Waldeck (Ref. 132) present a detailed fault tree for the top event "aircraft destroyed." The tree was solved by simulation and the probability of the top event occurring as well as the critical paths leading to it were calculated. Based on this information the design was modified to increase safety and a new fault tree was drawn and solved. The authors discuss the problems which arise in the construction and solution of a fault tree.

Crosetti and Bruce (Ref. 133) show how the fault tree analysis can be proven useful in system studies. From the nuclear industry they summarize the results of Cole (Ref. 134) concerning the reliability of the fog spray system of Hanford's N-Reactor. Then they proceed to discuss the possibilities of using fault trees in reliability and optimization studies of communications systems, in the automobile industry, in freeway planning and in evaluating marketing alternatives.

Salvatori (Ref. 135) uses a fault tree to establish acceptable probabilities of occurrence for the events which may cause a dangerous situation in a nuclear power plant (Fig. 3.15). The top event is called "limit consequence status" and it represents the event which, when reached, can be classified as a

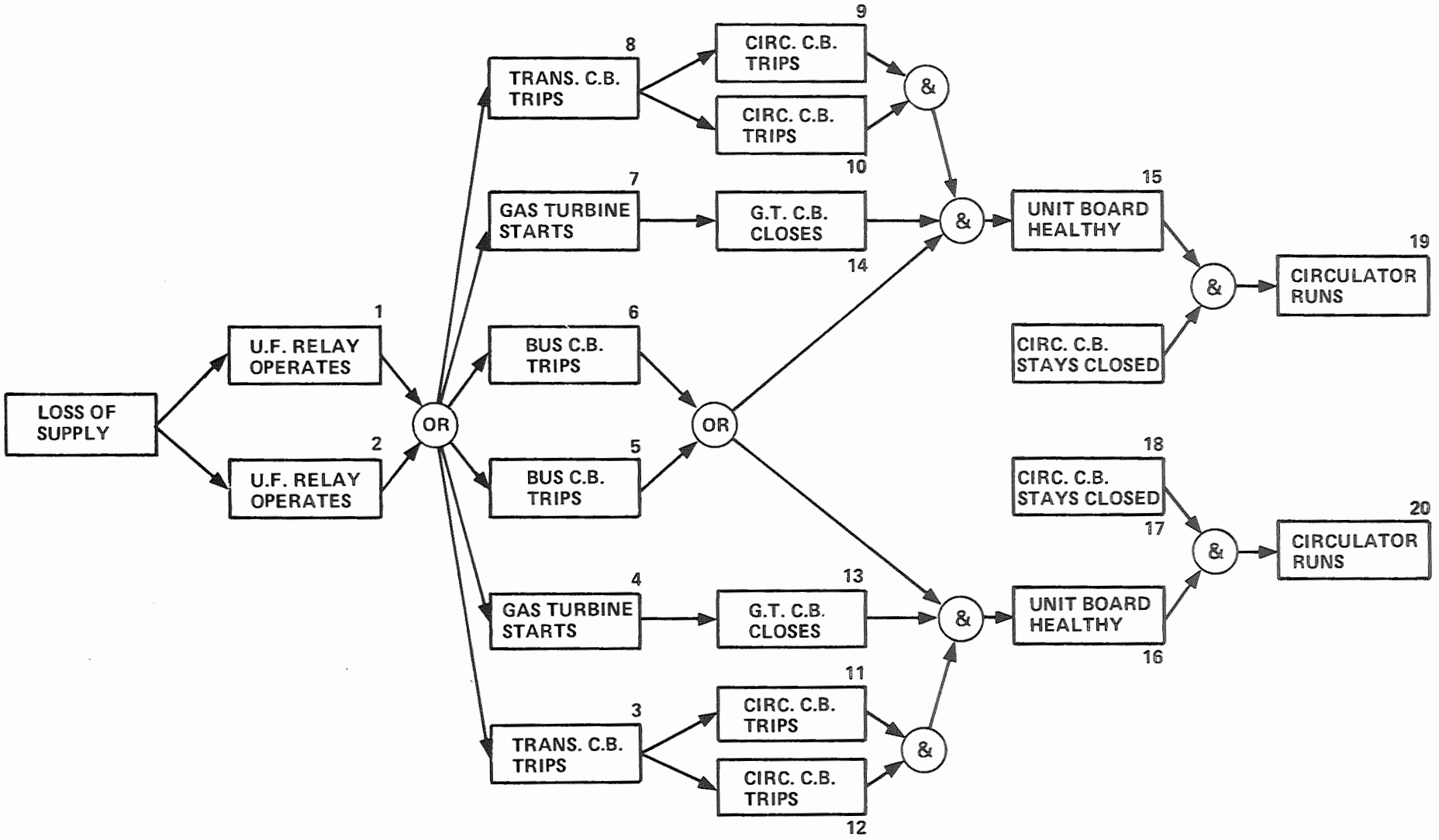


Figure 3.14. Reliability Diagram for the Electrical Supply System of Figure 3.13.

TABLE 3.3. COMPONENT DATA FOR THE LOGIC DIAGRAM OF FIG. 3.14. θ IS THE FAILURE RATE PER YEAR, τ_c IS THE INSPECTION INTERVAL AND τ_r IS THE MEAN REPAIR TIME

Item No.	Component	θ	τ_c (years)	τ_r (years)	Remarks
1, 2	Under frequency relay	0.01	1		
3, 8	Transformer circuit breaker (11 kV)	0.005	1	4×10^{-4}	Faults causing failure to trip on demand
5, 6	Bus-section circuit breaker (11 kV)	0.005	1	4×10^{-4}	Faults causing failure to trip on demand
4, 7	Gas turbine	—	—	—	0.023 probability of failure to start
13, 14	Gas turbine circuit breaker	0.02	1	4×10^{-4}	Faults causing failure to close on demand
9, 10, 11, 12	Circulator circuit breaker	0.005	1	4×10^{-4}	Faults causing failure to trip on demand
17, 18	Circulator circuit breaker (i.e., pre-selected circulators)	0.005	1	4×10^{-4}	Faults causing spurious opening
15, 16	Bus-bar	0.006	1	5×10^{-3}	Faults per bus-bar section
19, 20	Circulator				0.01 probability of failure to start

TABLE 3.4. PROBABILITY OF FAILURE PER DEMAND TO RESTORE TWO GAS CIRCULATORS

Time (years)	Probability of failure per demand
0.0	0.063
0.2	0.080
0.4	0.096
0.6	0.111
0.8	0.126
1.0	0.141

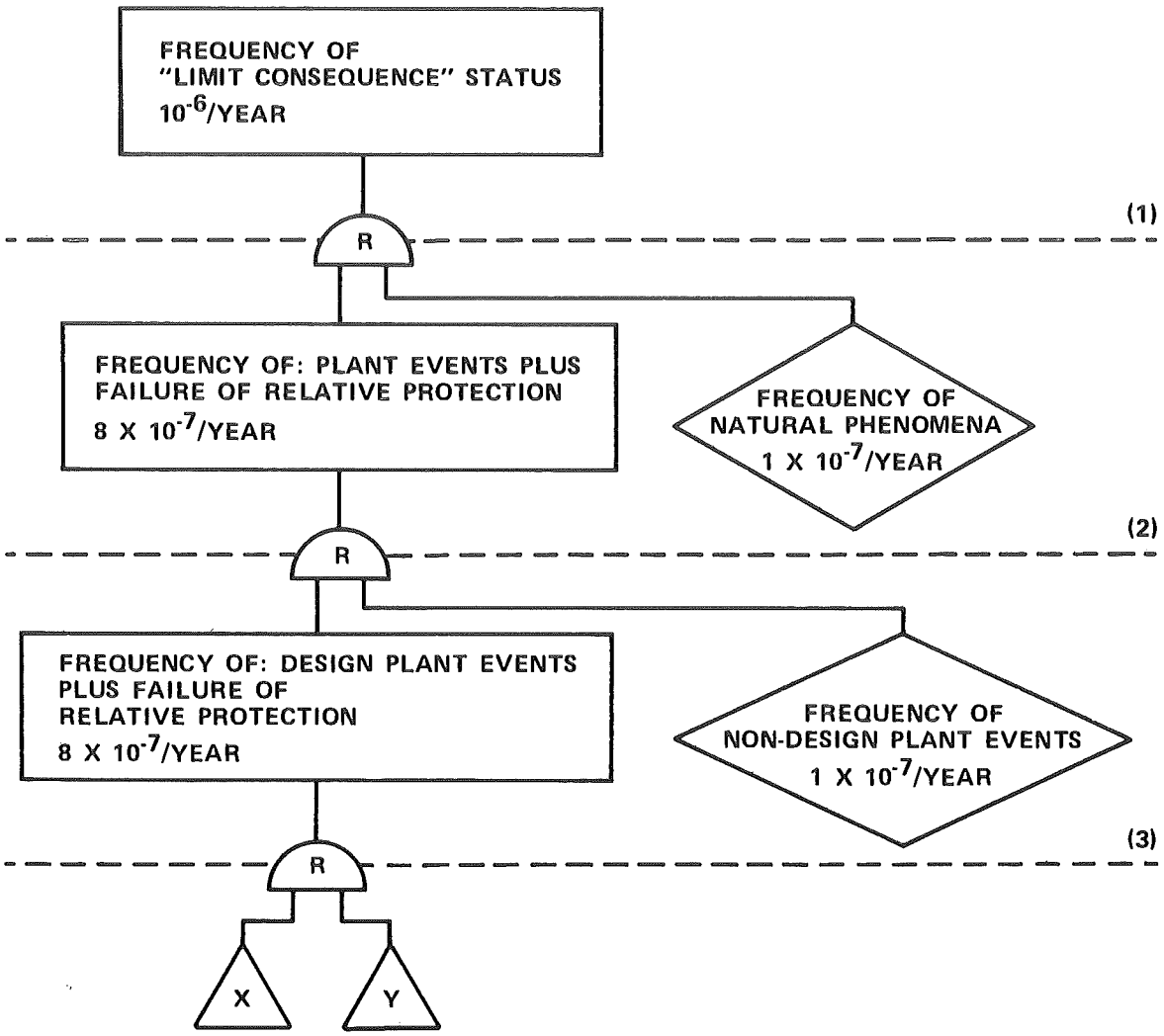


Figure 3.15. Fault Tree for the Assessment of Acceptable Probability Levels for Potentially Hazardous Events (Ref. 135).

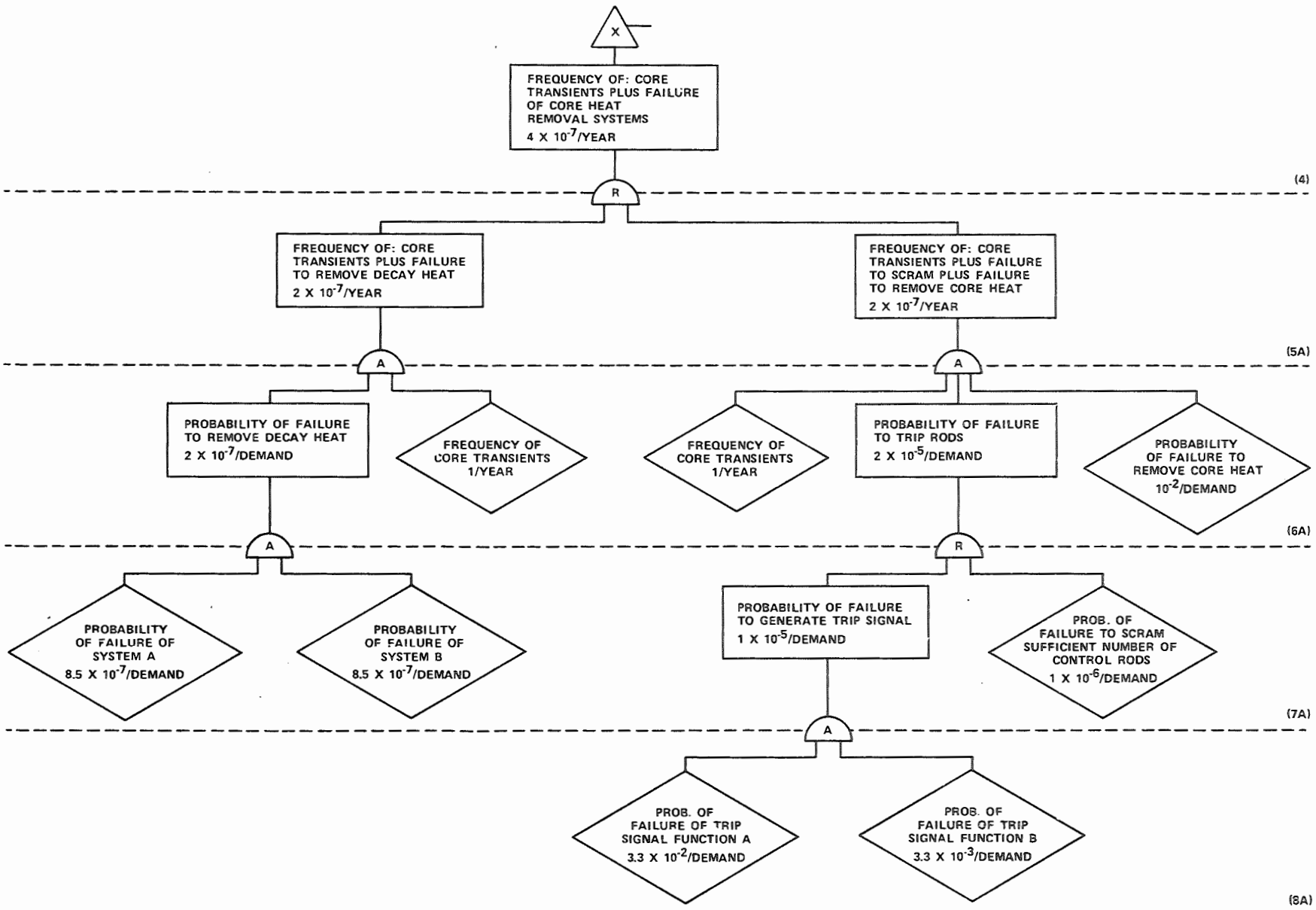


Figure 3.15. con't.

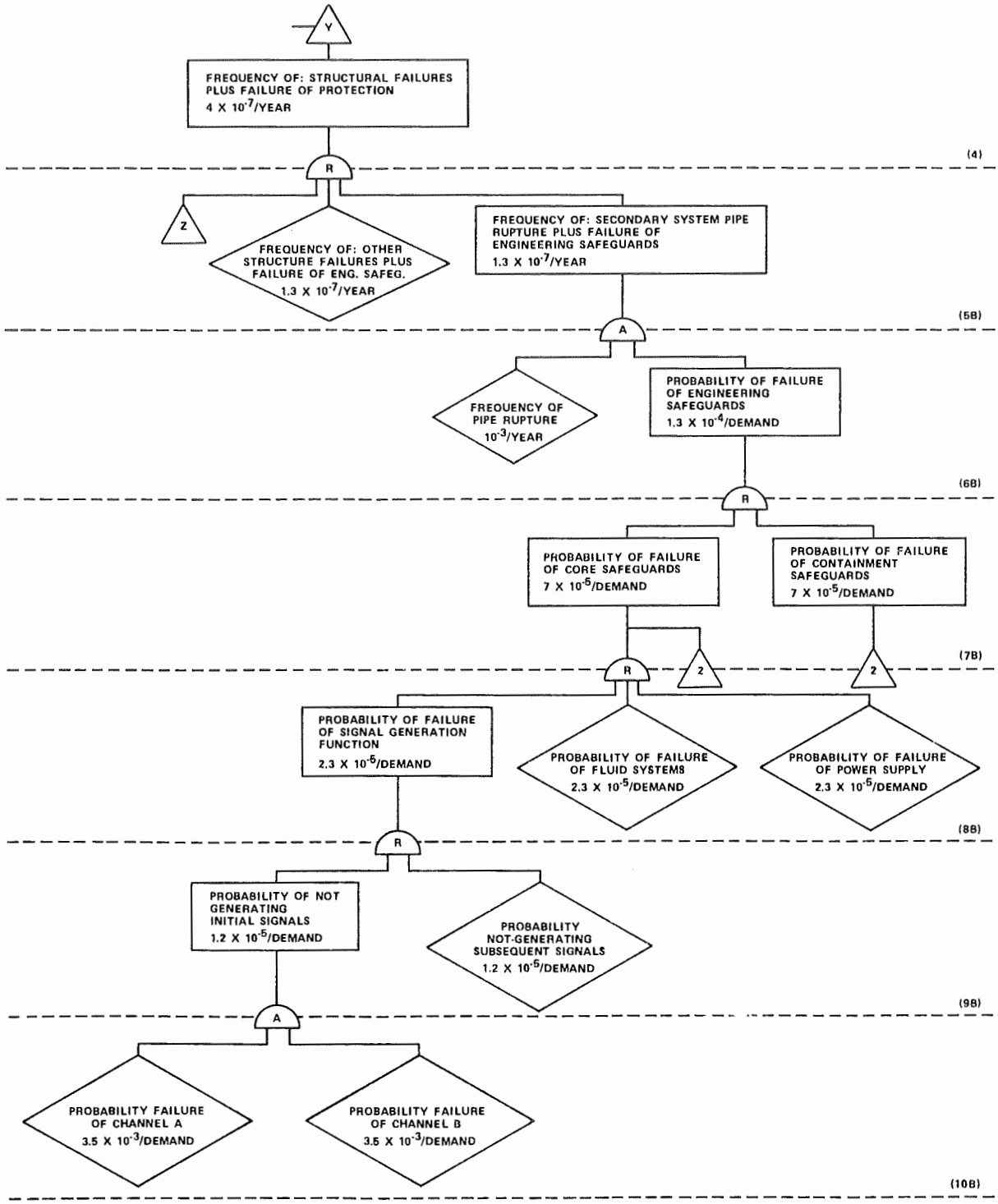


Figure 3.15. con't.

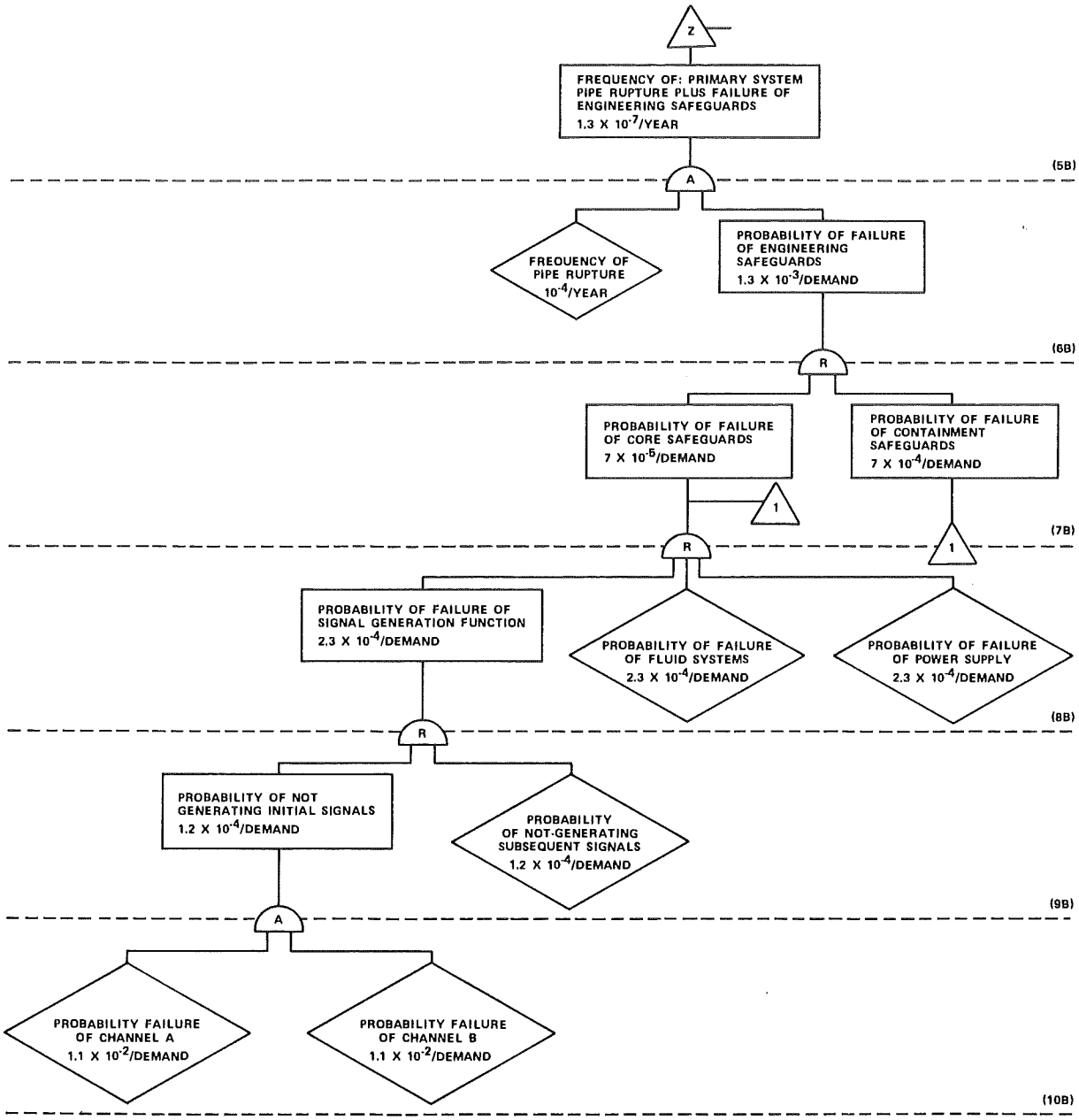


Figure 3.15. con't.

disaster. Its acceptable probability level is set at 10^{-6} per year, which is deemed to be of the same order as natural disasters. Using the deductive logic of the fault tree the chains of events that can lead the reactor to its limit consequence status are developed and judgment is applied to assign acceptable probability levels for each. If for a specific plant the probability of an event is greater than the level established above, special care should be taken to reduce it below the limit. As an example the author mentions the results of a study concerning the probability of a missile generated by a disk rupture hitting a critical plant component. At design speed this probability was estimated to be 10^{-12}/yr and at overspeed 10^{-10}/yr. The accident is classified as a non-design plant event, thus its probability of occurrence should not exceed the value 10^{-7}/yr (Fig. 3.15); since the estimated probabilities are much smaller than this limit, it is concluded that it is not necessary to make provisions in the plant design to contain the consequences of this accident.

Balfanz (Ref. 136) suggests the use of fault trees to estimate the failure rates of mechanical and electrical equipments. The method can be used as a supplement of the statistical analysis of the failures of identical items or as a method of estimation when statistical data are missing. Stewart and Hensley (Ref. 137) study a chemical plant in which oxygenated material is produced from oxygen and hydrocarbon. Figure 3.16 shows the physical process that takes place. The hydrocarbon and the oxygen are led into the reactors to produce the oxygenated material under high temperature and pressure. The rest of the cycle is self-explanatory. The objective of the study is to design an automatic protective system to prevent explosion. The fault tree of Fig. 3.17 was drawn to identify the events that could lead to explosion and which parameters should be monitored by the protective system. The numbers in the figure show the parameters selected.

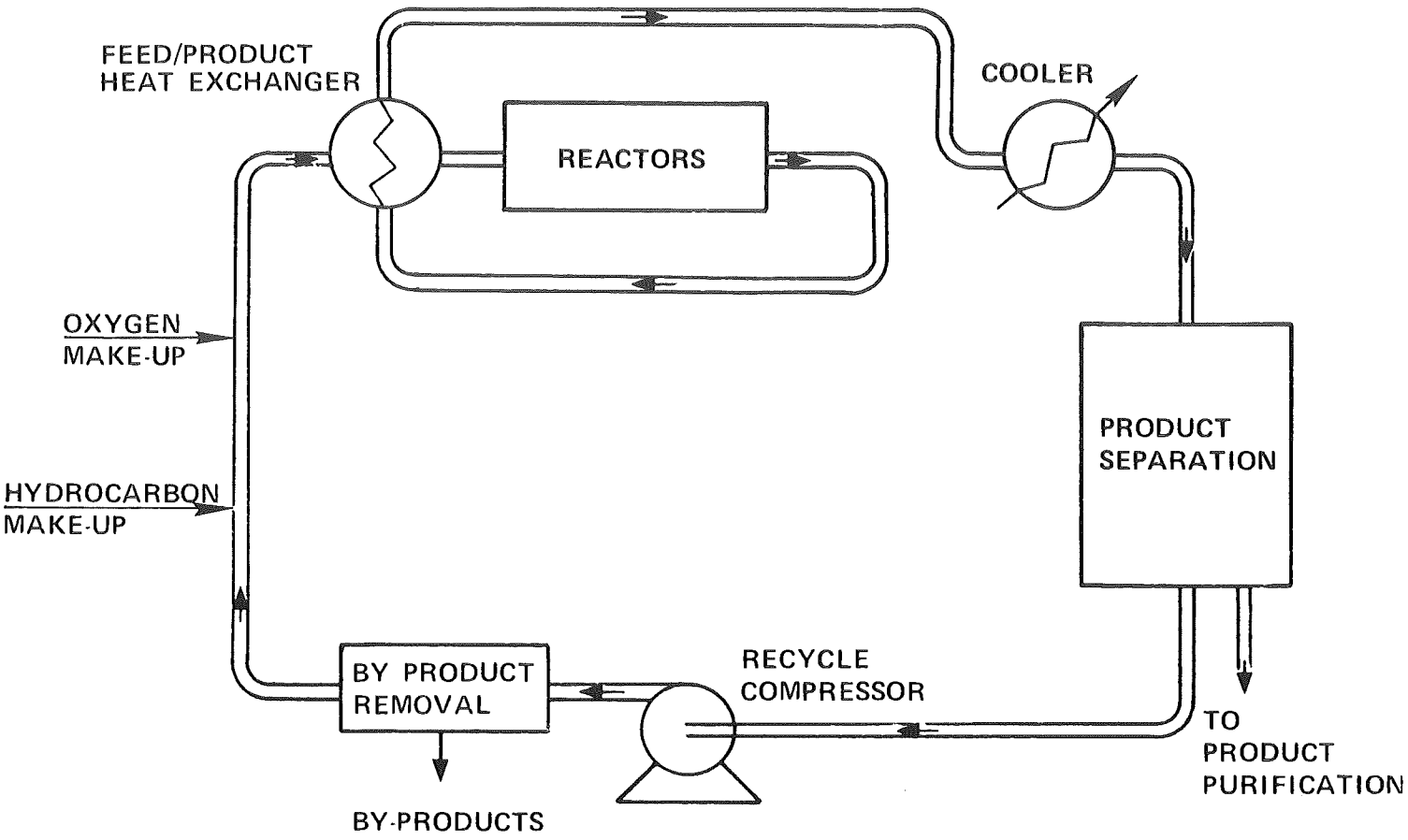


Figure 3.16. Process Diagram for a Chemical Plant (Ref 137).

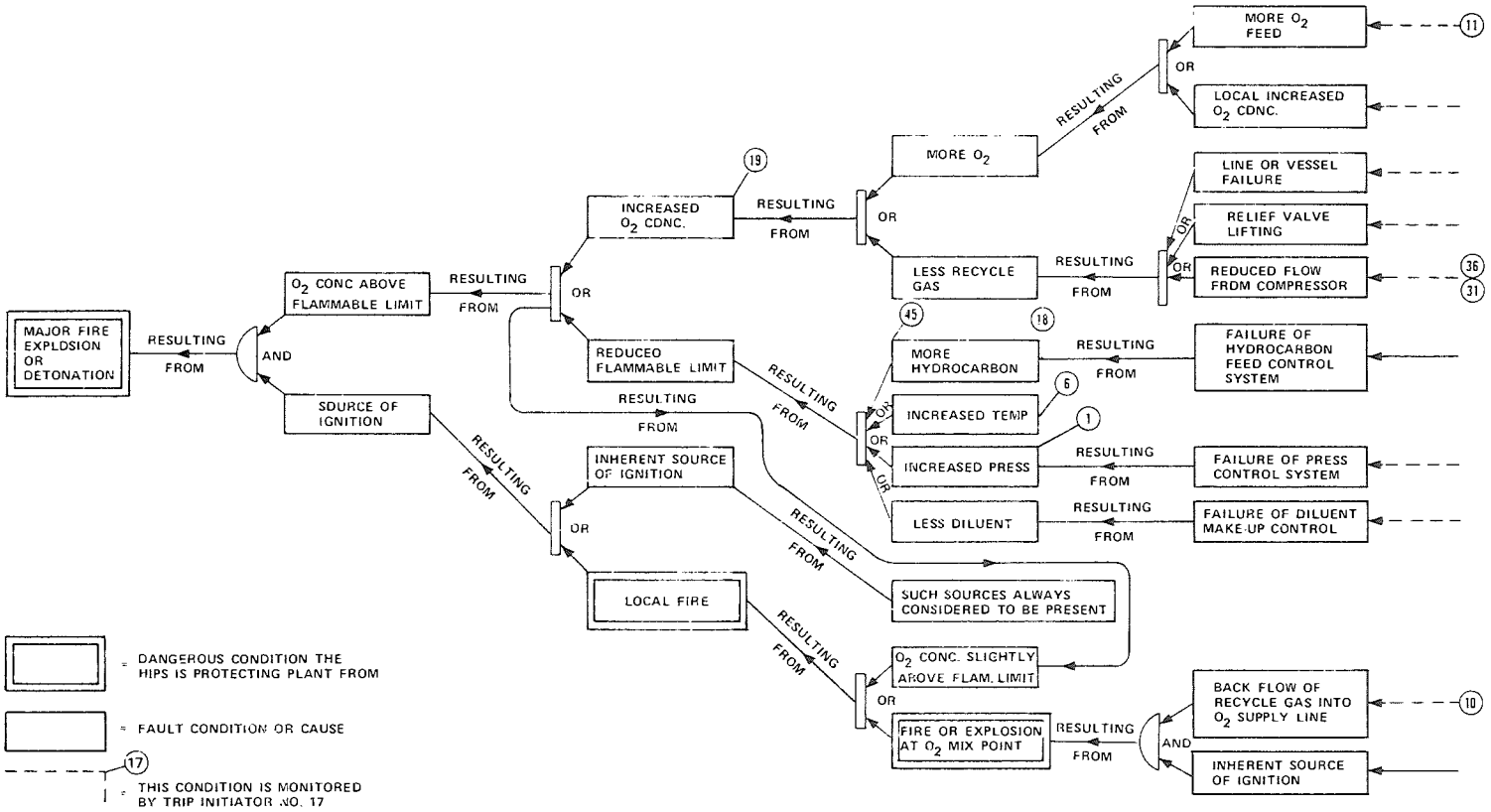


Figure 3.17. Fault Tree for the Chemical Plant of Figure 3.16 (Ref 137).

PART OF THE LOGIC DIAGRAM USED TO DETERMINE FAULT CONDITIONS WHICH COULD LEAD TO DANGEROUS CONDITIONS

In Ref. 138 Griffin utilizes the visibility that the method provides to compare the relative safety of two alternatives for the containment system of the Heavy Water Organic Cooled Reactor. His second application concerns the increase of the risk that the public is subjected to due to the existence of a sodium facility in a certain area. Finally, in the last application a fault tree is drawn to analyze the risk that the owner of a test facility complex assumes due to the potential of damage to the population, the environment etc. Risk is defined as the product of the probability of damage times its consequences. The tree is shown in Fig. 3.18.

The safety systems of a nuclear power plant have been popular subjects of investigation via fault trees. Hörtner et al. (Ref. 139) estimate the unavailability of all the systems (mechanical, protective and power supply) that are required to function in the case of rupture of the primary coolant line in a PWR. The results for the overall system and for each subsystem studied separately (as it is usually the case in the literature) are shown in Table 3.5. It is seen that the total unavailability is greater than the sum of the unavailabilities from the independent calculations. This is due to failures in the integrated system, which are missed in the separate calculations, i.e., failures due to weak points resulting from the interconnection of the subsystems.

Bustl (Ref. 140) gives an overall view of the problems which arise when reliability techniques are applied on nuclear power plants at the component, equipment and system level. Specific examples include the actuator command unit which controls a motor valve and the emergency power supply to the ECCS.

Gangloff et al. (Ref. 141) outline a study of the unavailability of the containment spray system of a PWR. Identification of the minimal cut sets

TABLE 3.5 UNAVAILABILITIES OF THE SYSTEM DEMANDED TO CONTROL A LOCA.

ASSUMED PROBABILITY OF LOSS OF STATION POWER SUPPLY 0.1 (Ref. 139)

System	Unavailability
Over-all Safety System	2.6×10^{-4}
Mechanical System	0.39×10^{-4}
Protection System	0.50×10^{-4}
Electrical Power Supply	0.39×10^{-4}

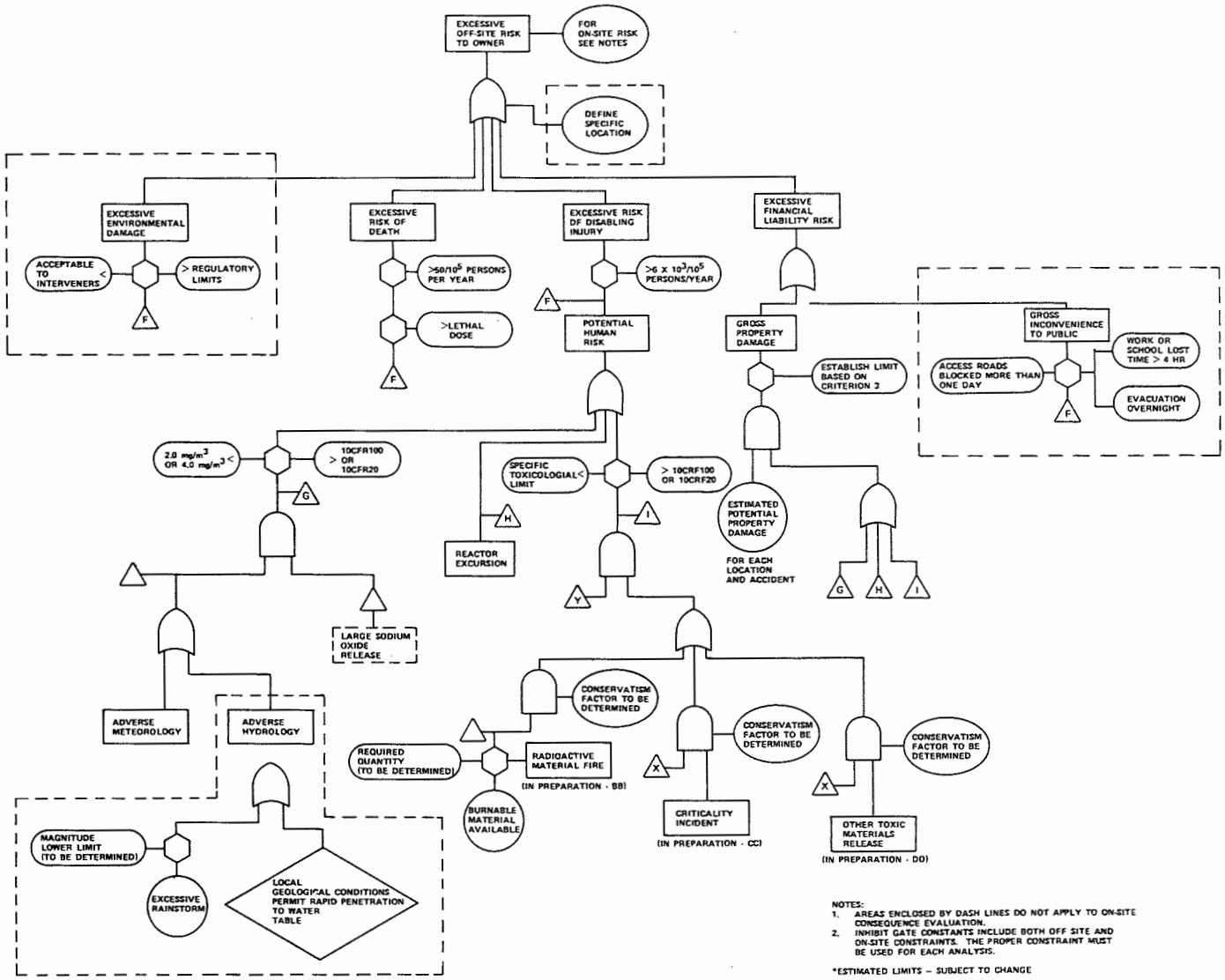


Figure 3.18. Risk Analysis for a Test Facility Complex (Ref 138).

reveals that no single failure could result to system failure. The probability of no spray is estimated to be 9.982×10^{-5} (per demand) and that of inadvertent spray 9.70×10^{-7}.

Erdmann, Okrent et al. (Ref. 142) study the loss of isolation between the high and low pressure portions of the residual heat removal system of a BWR. A simplified schematic of the system is shown in Fig. 3.19. The RHR system is a low pressure system (400 psi) directly connected to the primary system which is at higher pressure (1200 psi). Its objectives are to remove decay and residual heat from the reactor so that refueling and servicing can be performed, to supplement the spent fuel cooling system capacity when necessary to provide additional cooling, to condense reactor steam so that decay and residual heat may be removed if the main condenser is unavailable following a reactor scram and it forms an essential part of the low-pressure core flooding system which is part of the ECCS. In Fig. 3.19 the valves shown in black are normally closed. Failure of isolation can occur if the following groups of valves fail.

{F019, F022, F023}, {F050, F015},

{F009, F0018, F006B or F006A} .

A fault tree was built to analyze the event "Loss of isolation on leg 1" and it is shown in Fig. 3.20. Seven schemes were analyzed as follows:

Scheme 1: F019 and F022 are required to function properly while F023 is non-existent (i.e., open). Only failure data from nuclear experience are used.

Scheme 2: Valve F019 is not included and again nuclear data is used.

Scheme 3: The original design is considered with all available data.

Scheme 4: The pressure interlock on valve F022 is removed and nuclear data is used.

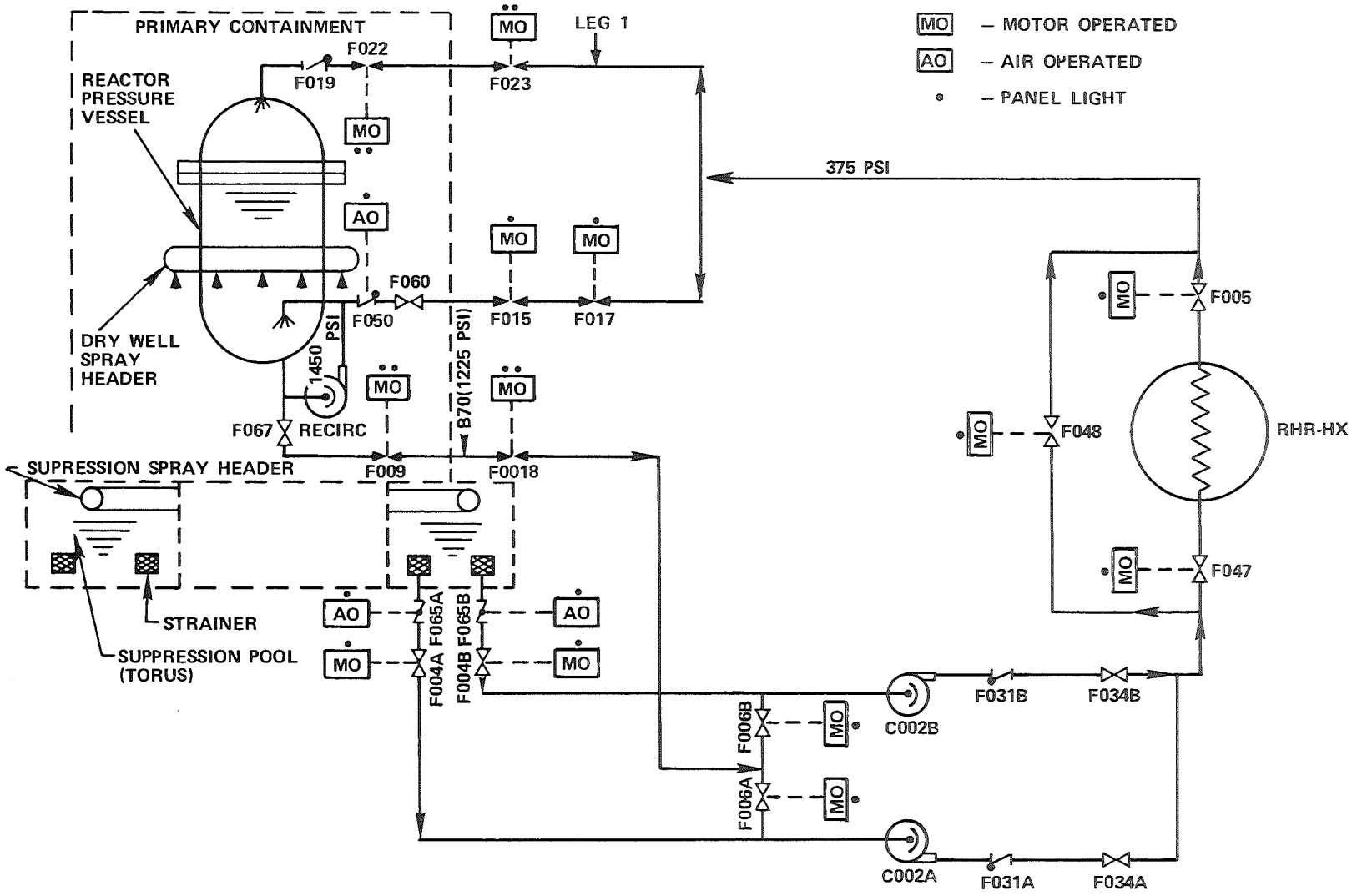


Figure 3.19. Schematic of RHR System for a BWR (Ref. 142).

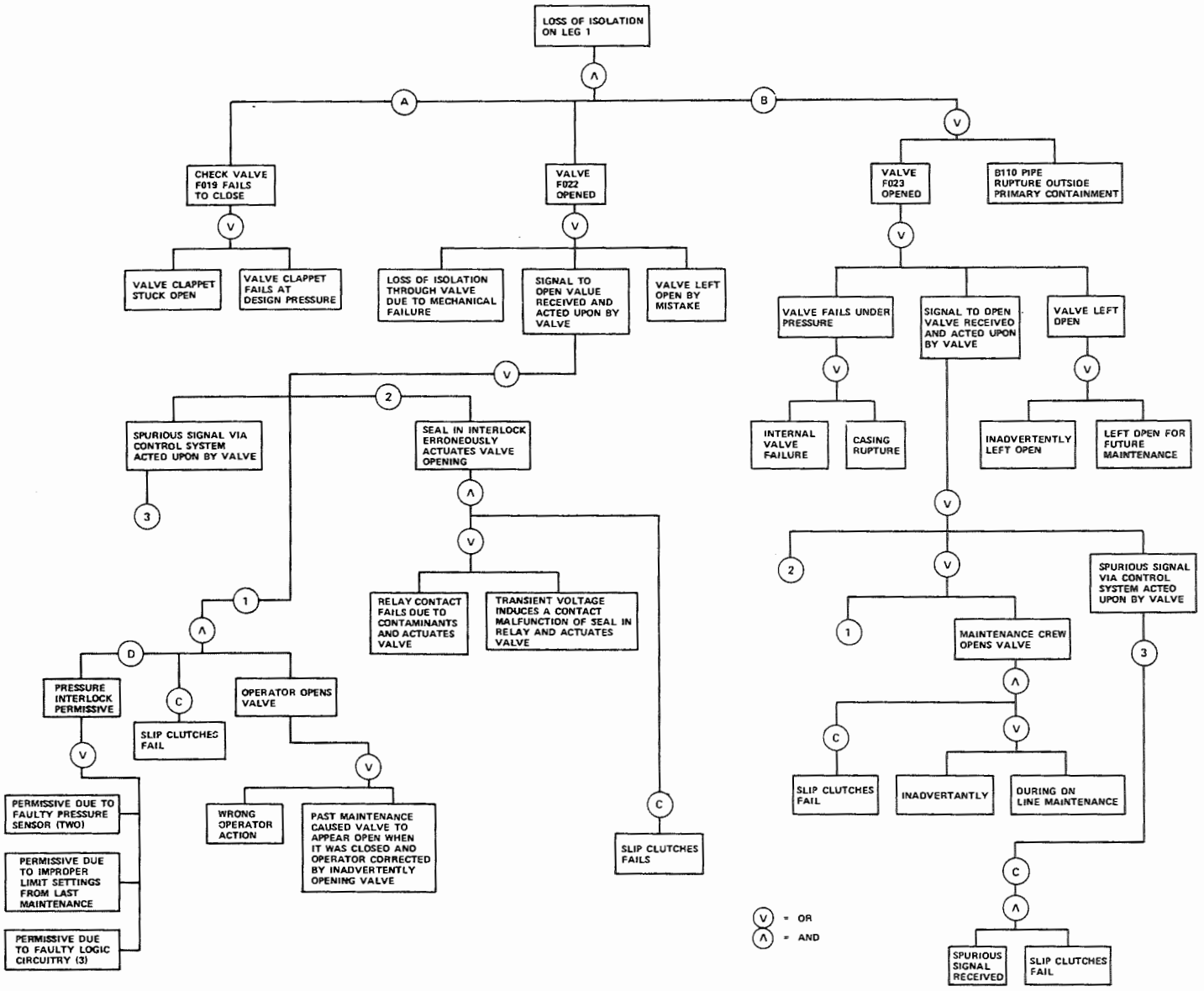


Figure 3.20. Fault Tree for the Loss of Isolation on Leg 1 of Figure 3.19 (Ref. 142).

Scheme 5: Identical to Scheme 3 but only nuclear data is used.

Scheme 6: Slip clutches are added to valves F022 and F023, nuclear data is used.

Scheme 7: Identical to Scheme 6, but the slip clutches are replaced by limited torque motors.

The failure rates are shown on Table 3.6. λ_{20} means that 20% of the data are smaller than the selected values. Similarly for λ_{80}. The logarithmic mean $\bar{\lambda}$ is calculated from

$$\ln \bar{\lambda} = \frac{1}{2} (\ln \lambda_{20} + \ln \lambda_{80}) \quad .$$

The fixed-time-to-repair model was employed with τ=3 months. Since the study is for one year, no repair is equivalent to setting τ = 12 months. The tree was solved with the aid of the code EKKFA (Ref. 143) which is a modification of the PREP and KITT codes.

The critical minimal cut sets for Scheme 3 were found to be:

1. random failure of valves F019, F022 and F023,
2. random failure of valve F019 and inadvertent opening by either the operator or the automatic control system of valves F022 and F023,
3. random failure of valves F019 and F022 and pipe failure outside the primary containment between the F022 and F023 valves.

The results for the various schemes are shown on Table 3.7.

3.A.8 Event Trees

The event (or flow, or accident-process) tree is a logic diagram similar to the fault tree with one fundamental difference: the logic is inductive, i.e. starting from an initiating event the tree proceeds to uncover its consequences. It is similar to the decision tree of decision analysis,[144] in safety studies, however, the event trees do not, as a rule, include any decisions.

TABLE 3.6 FAILURE RATES FOR THE FAULT TREE OF FIG. 3.20 (Ref. 142)

	Nuclear Data			All Data		
	λ_{20}	$\bar{\lambda}$	λ_{80}	λ_{20}	$\bar{\lambda}$	λ_{80}
Automatic Control System--Spurious or Improper Signal	6.7×10^{-8}	1.1×10^{-7}	2.1×10^{-7}	2.0×10^{-7}	5.5×10^{-7}	1.6×10^{-6}
Clutches	7.0×10^{-8}	2.5×10^{-7}	0.9×10^{-6}	7.0×10^{-8}	2.8×10^{-8}	1.1×10^{-6}
Contaminants--Failure Event Due to Contaminants	--	--	--	--	--	--
Circuit Breaker--High Current	1.0×10^{-6}	1.0×10^{-5}	1.0×10^{-4}	1.0×10^{-6}	4.0×10^{-5}	2.0×10^{-3}
Design Error	2.0×10^{-7}	2.4×10^{-7}	3.0×10^{-7}	2.0×10^{-7}	2.4×10^{-7}	2.5×10^{-7}
Instrumentation						
A) Detectors or Transducers	6.0×10^{-6}	1.6×10^{-5}	4.0×10^{-5}	8.0×10^{-6}	2.2×10^{-5}	5.6×10^{-5}
B) Recorders, Display Units, etc.	2.5×10^{-6}	9.3×10^{-6}	3.5×10^{-5}	4.5×10^{-6}	1.5×10^{-5}	5.2×10^{-5}
Interlocks--Relay or Switch	5.0×10^{-7}	3.0×10^{-6}	2.0×10^{-5}	5.0×10^{-7}	3.0×10^{-6}	2.0×10^{-5}
Maintenance Errors						
A) Installation	--	--	--	--	--	--
B) Modification	--	--	--	--	--	--
C) Operational or On-Line	--	--	--	--	--	--
D) General Non-Specific	2.0×10^{-8}	5.1×10^{-8}	1.3×10^{-7}	2.0×10^{-8}	1.4×10^{-7}	8.0×10^{-7}
Operator Errors	5.0×10^{-8}	1.7×10^{-7}	5.8×10^{-7}	5.0×10^{-8}	1.7×10^{-7}	5.8×10^{-7}
Pipe Rupture--Serious Leaks	2.0×10^{-11}	1.5×10^{-10}	1.0×10^{-9}	2.0×10^{-11}	1.5×10^{-10}	1.0×10^{-9}
Power Failures						
A) Primary	6.0×10^{-6}	8.2×10^{-6}	1.1×10^{-5}	6.0×10^{-6}	8.2×10^{-6}	1.1×10^{-5}
B) Standby	2.0×10^{-7}	2.0×10^{-6}	1.8×10^{-5}	2.0×10^{-7}	1.2×10^{-5}	7.0×10^{-4}
C) Power Supplies (H.V. or L.V.)	3.0×10^{-6}	9.2×10^{-6}	2.8×10^{-5}	7.0×10^{-7}	1.2×10^{-5}	2.0×10^{-4}
Pump Failures	--	--	--	--	--	--
Relay--Low Current	1.0×10^{-7}	3.0×10^{-7}	1.0×10^{-6}	1.9×10^{-7}	1.7×10^{-6}	1.6×10^{-5}
Valve--Isolation or Throttle						
A) Mechanical	5.0×10^{-7}	2.5×10^{-6}	1.1×10^{-5}	5.6×10^{-7}	3.0×10^{-6}	1.5×10^{-5}
B) Actuation	5.0×10^{-7}	2.5×10^{-6}	1.0×10^{-5}	5.8×10^{-7}	3.0×10^{-6}	1.2×10^{-5}
C) Non-Specific Causes	8.0×10^{-7}	3.3×10^{-6}	1.4×10^{-5}	1.0×10^{-6}	4.3×10^{-6}	1.9×10^{-5}
Valve--Check	3.0×10^{-7}	1.3×10^{-6}	6.0×10^{-6}	5.0×10^{-7}	2.5×10^{-6}	1.2×10^{-5}
Valve--Safety	--	--	--	--	--	--
Valve--Relief	--	--	--	--	--	--

TABLE 3.7 FAILURE PROBABILITIES FOR THE VARIOUS SCHEMES OF THE FAULT TREE OF FIG. 3.20 (Ref. 142)

Description of Schemes	Failure Probabilities per Year		
	Q_{80}	$\overline{Q}$	Q_{20}
1. Modified Design (One Check and One Isolation Valve)--Nuclear Statistical Data Used			
A) No Repair	2.0×10^{-3}	1.4×10^{-4}	1.0×10^{-5}
B) Repair	9.4×10^{-4}	6.7×10^{-5}	4.6×10^{-6}
2. Modified Design (No Check Valve)--Nuclear Statistical Data Used			
A) No Repair	1.2×10^{-3}	1.2×10^{-4}	1.3×10^{-5}
B) Repair	3.1×10^{-4}	2.3×10^{-5}	3.5×10^{-6}
3. Original Design--All Statistical Data Used			
A) No Repair	4.7×10^{-3}	5.8×10^{-5}	7.2×10^{-7}
B) Repair	7.8×10^{-4}	1.0×10^{-5}	1.5×10^{-7}
4. Modified Design (Pressure Interlock on One Valve Only)--Nuclear Statistical Data Used			
A) No Repair	8.1×10^{-5}	2.0×10^{-6}	4.9×10^{-8}
B) Repair	1.9×10^{-5}	4.6×10^{-7}	1.1×10^{-8}
5. Original Design--Nuclear Statistical Data Used			
A) No Repair	7.8×10^{-5}	1.8×10^{-6}	4.2×10^{-8}
B) Repair	1.4×10^{-5}	3.3×10^{-7}	8.1×10^{-9}
6. Modified Design (Slip Clutches)--Nuclear Statistical Data Used			
A) No Repair	1.3×10^{-5}	2.0×10^{-7}	4.0×10^{-7}
B) Repair	1.5×10^{-6}	3.0×10^{-8}	4.6×10^{-10}
7. Modified Design (Limited Torque Motors)--Nuclear Statistical Data Used			
A) No Repair	1.0×10^{-5}	1.6×10^{-7}	3.0×10^{-9}
B) Repair	1.0×10^{-6}	2.0×10^{-8}	4.0×10^{-10}

The merits of event trees are similar to those of fault trees: they are useful visual aids for understanding the consequences of an event, critical chains of events can be readily identified and simple probabilistic calculations can be performed.

Event trees are particularly useful in a probabilistic assessment of the risk from a power plant. In contrast to the design basis accident approach, where a maximum credible accident is postulated and the plant is designed to limit its consequences under the assumption that all the relevant parameters are unfavorable (worst case analysis), the current philosophy is not to differentiate between credible and incredible accidents, but to assign probabilities to all conceivable accidents and analyze their consequences via event trees. In this manner some measure of risk can be established which can be compared with the acceptable criteria. Details on this line of thought are given in Refs. 145, 146 and 147.

The graphical representation of an event tree is much simpler than that of fault trees, since no special symbols are used. Several examples will illustrate the methodology.

Pugh (Ref. 148) discusses the application of event trees on reactor systems (scientists of UKAEA usually call them fault diagrams or fault trees). A typical event tree in simplified form is shown in Fig. 3.21. The author proposes that for the initiating event in addition to its frequency of occurrence, the trip signals resulting from the detection of the fault should be identified for better coordination between the designer and the safety analyst. On the left of the diagram the sequence of protective systems that will be required is shown. The two branches of the tree which originate from each junction correspond to the two mutually exclusive events: "the system functions"

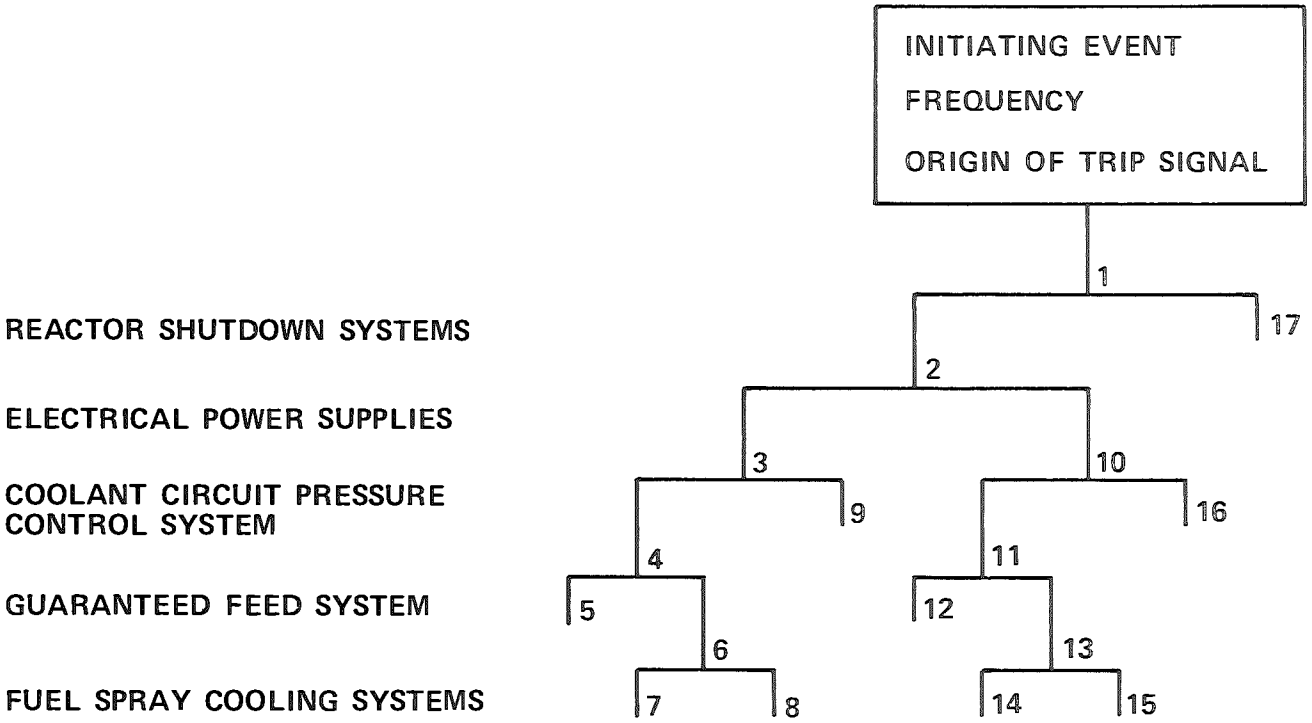


Figure 3.21. Typical Event Tree (Ref. 148).

and "the system fails." In an application the corresponding probabilities (usually per demand) should be known. An additional convention of the author is that branches on the left represent successful performance of the system. Thus the sequences of events leading to the points 5,7,12 and 14 are considered safe (the reactor is shut down and is sufficiently cooled). The points 8,9, 15,16 and 17 should be examined more carefully to assess the potential damage (e.g., if the system reaches point 16 the reactor is shut down but the coolant circuit pressure is not reduced thus a break of the circuit will occur). The probability of the system state represented by each end-point can be found by multiplication of the probabilities of occurrence of the events which lead to that point. The author proceeds to apply the method to initiating events that can occur in a heavy water reactor. Figure 3.22 shows two such applications (the negative numbers at each end-point are the powers of ten which give the probability of that state per year).

Doron and Albers (Ref. 149) use event trees to find the possible sequences of events that may occur after a loss of coolant accident in a PWR. The tree is shown on Fig. 3.23 with the annual frequency of each event and of each branch leading to the end-points 1-22. The authors then give the estimated activity released (in Curies) for each final state of the system along the lines that Farmer proposed in Ref. 145. The next column in the figure shows the product of the activity release times the frequency of the final state (Curies/year) and can be used as a measure of the risk from the plant. As it is shown on Figure 3.23 the most critical branches are 18 and 21.

The loss of coolant accident and the handling of related problems via event trees in a BWR is the subject of Ref. 150.

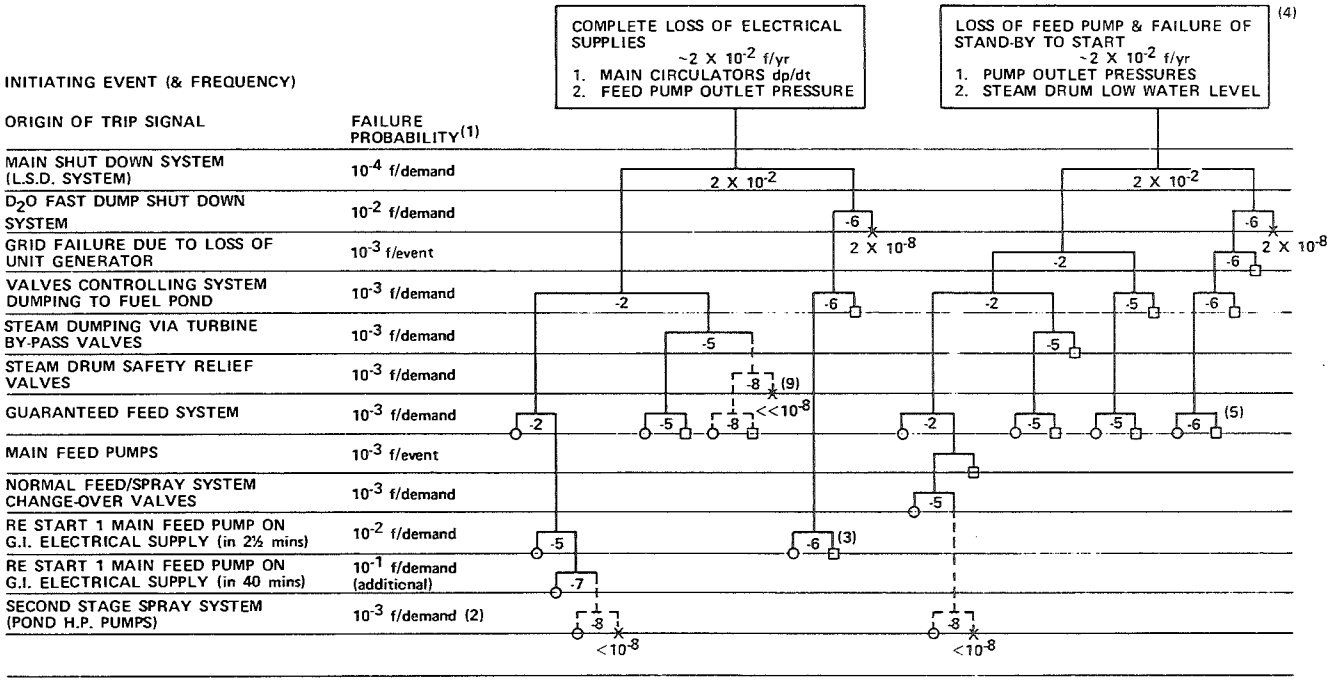


Figure 3.22. Applications of Event Trees (Ref. 148).

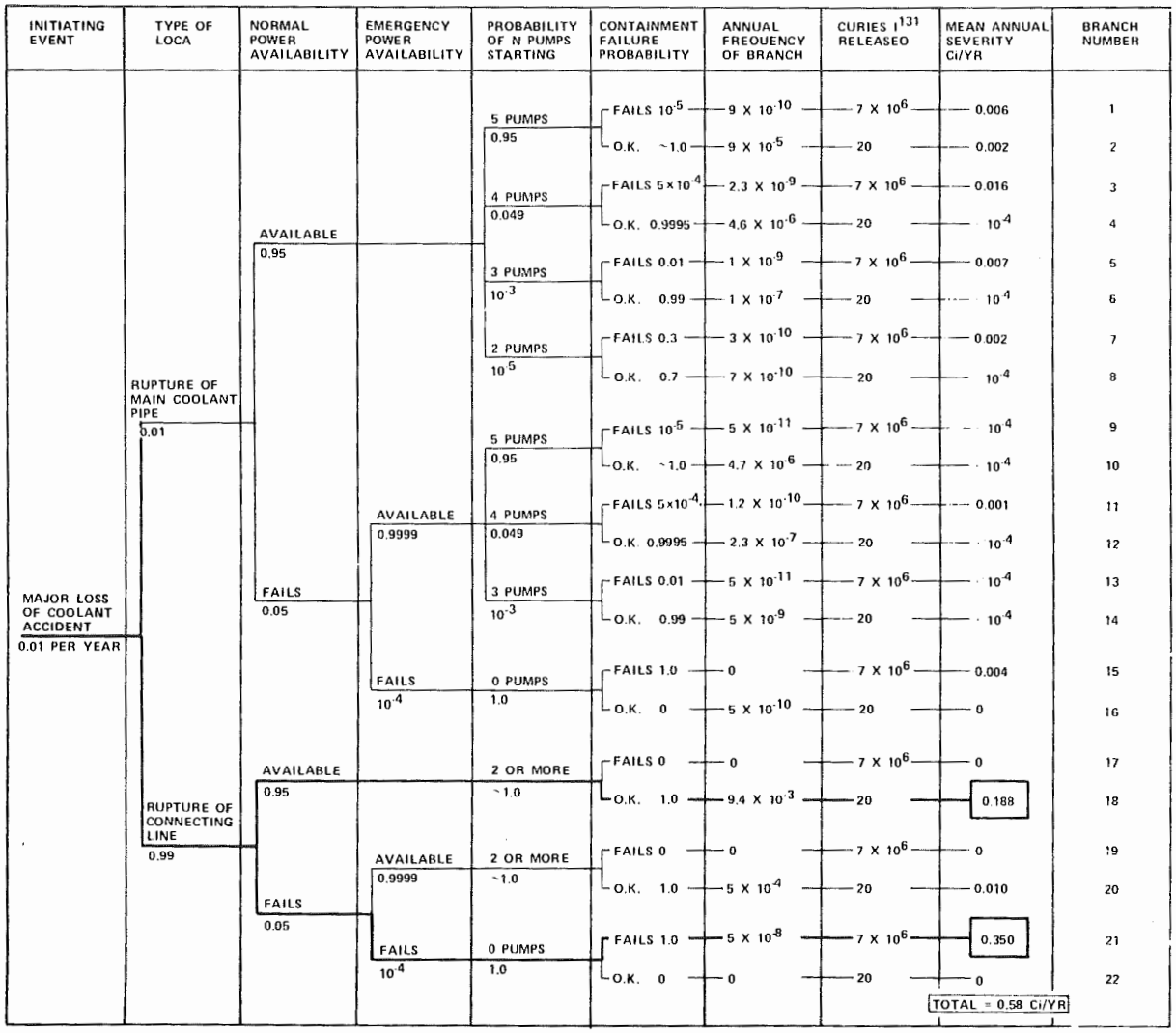


Figure 3.23. Event Tree for a Loss of Coolant Accident of a PWR (Ref. 149).

3.A.9 Qualitative Methods

The methods discussed here are qualitative and inductive in nature and they form an integral part of a safety analysis. They usually precede a fault tree study and their conclusions are of great help in the construction of the fault tree. A survey of the various methods is done by Balfanz in Ref. 151.

Although different names are given to these methods one could simply say that they are the natural approach an analyst with common sense would take in order to understand the system and its safety aspects. Thus the preliminary hazard analysis identifies the sources of energy which should be under control, like pressure tanks, fuels, etc.; the operational hazard analysis examines the functions and safety of the personnel employed in the plant, and so on.

Among the qualitative methods the one that finds extensive use is the failure modes and effects analysis (FMEA). As the name suggests, the analysis identifies the failure modes of the components of the system (all or the ones that are judged to be critical) and the effects that these may have upon the subsystem to which the component belongs and the system as a whole. For a systematic FMEA blank forms are provided on which as much information about the component is reported as it is deemed necessary and, of course, possible. As an example of FMEA Table 3.8 shows an application on a hand valve (Ref. 151). This is only one possible form and depending on the specific system under study additional columns can be used to enhance the amount of information given (like effect on personnel, corrective actions etc.). It is clear then how useful the FMEA can be, if done at the right time, in suggesting design modifications and in helping to build a fault tree for an unfavorable event.

TABLE 3.8 FAILURE MODES AND EFFECTS ANALYSIS OF A HAND VALVE (Ref. 151)

Name of Structural Part	Operating State	Failure Mode	Failure Cause	Failure rate (hr^{-1})	Repair Time (hr)	Effects on Input and Output of Structural Part	Effects on System
Valve 15	Closed	Opens incompletely unintentionally	Breakage of valve drive	0.5×10^{-6}	8	Complete failure of lock	No pressure in system
		Opens partially unintentionally	Leakage in valve seat	3×10^{-6}	24	Slight internal leakage	Pressure drop in system, follow up feed through regulation
		Does not open completely	Valve drive sluggish (rusted)	3×10^{-6}	24	Chocking effect due to valve	Delayed pressure relief in system
		Does not open	Blockage of drive or motor defective	5×10^{-6}	8	No preasure relief via valve	no pressure relief in system
	Open	Closes completely inintentionally	Sectional lock due to broken valve cone	0.8×10^{-6}	8	Blockage of pressure medium	Impermissible pressure build up in system

Some quantitative results can be determined from the FMEA showing the probability of occurrence of each failure mode and thus ranking them according to their importance. Details on this approach (criticality analysis) may be found in Ref. 152.

3.B. COMMON MODE FAILURES

A fundamental assumption of the fault tree analysis is that the failures of the primary components are random independent events which are described by exponential distributions. Some dependencies among redundant components may be predicted by changing the failure rates of the working components when some failures have occurred, but still the phenomenon is of random nature. A survey of the analyses performed on real systems and subsystems reveals that the degree of safety achieved against these random failures is acceptable. The reason behind this is that highly redundant systems are employed to perform a certain function.

However a system is not subject to random failures only but also to a different type of failures, called common mode failures, for which the methods we have presented are not applicable. Failure of many components due to a single cause is classified as a common mode failure. The problem is quite vague and general as stated in this definition and quantitative methods for handling it are lacking; the approach taken by safety analysts is qualitative and it is concerned with the classification of the various types of common mode failures into broad categories and based on this preventative measures are suggested (Refs. 153, 154 and 155). These categories are related to the cause of such a failure as follows:

1. Functional Deficiency.

The instrumentation used to monitor a certain variable is not appropriate for the intended use and it provides with wrong information. The conditions under which a system is supposed to operate are not well understood or they change unpredictably thus rendering the system inadequate. This type of a failure clearly has nothing to do with failure of the hardware itself and is

of systematic type in complete contrast to the random component failures which are usually encountered in reliability analyses.

2. Design Deficiency

Similar components have not been designed or manufactured properly. Equipments or subsystems thought to function independently actually have a common element failure of which can cause a common mode failure.

3. Maintainance Error

This category includes all errors that human operators may make regarding testing, repair, calibration and operation of the equipments of a system.

4. External Environment

Failures can be induced by fires, explosions, floods, earthquakes, tornados and other major external events. In addition other causes less dramatic but of importance may be unfavorable changes in the operating environment, such as accumulation of dirt and/or dust, high temperatures, humidity, vibration, etc.

The recommended measures against common mode failures are naturally based on different forms of diversity. The most important one which offers defense for all the previous categories is functional diversity by which more than one plant parameters are monitored to warn that an unfavorable situation is developing. Functional diversity should be combined with operational administrative diversity (more than one person should independently do and review personnel actions), equipment diversity (equipments of different types should be used to perform a certain function), physical diversity (redundant components should be physically separated) and design administrative diversity (reviews of the design and construction procedures). Although it is clear that such general recommendations are useful in reducing the probability of common mode failures, they are far from the quantitative methods of estimating

probabilities of failure due to random causes and the improvement of reliability due to redundancy, testing, etc.

A more systematic way of investigating common mode failures is described by Gangloff and Franke (Ref. 155) and Gangloff (Ref. 156). Fault trees (or other methods) are utilized to identify the combinations of events that can cause system failure (i.e., the minimal cut sets). From this information and the previous classification a table is built of the possible common mode failures of the system. Then the preventative measures are identified and each possible common mode failure is examined in detail to check whether it can be safely assumed that its likelihood of occurrence is very small. The method is of course, qualitative. Table 3.9 shows the format used to identify the common mode failures for a reactor protection system.

Epler (Ref. 157) uses the data on common mode failures that occurred at Oak Ridge National Laboratory to produce some numerical results and then make comparisons with random failures. The failures are described on Table 3.10. In addition to their causes the failures are also classified according to their rate of occurrence (instantaneous failures are those which occur so quickly that the operator does not have time to discover them and take the necessary steps to limit them). A further classification is with regard to whether the common mode failure was actually completed. The table shows that only 3 failures were completed while 7 were arrested in progress, result which verifies the usefulness of the preventative measures. Epler proceeds to estimate the rate of occurrence of common mode failures by estimating that the total number of subsystem years is 300 and with 3 failures having occurred the rate is 0.01 per subsystem year. To make comparisons with random failures it is assumed that the protection channels fail at a rate of 0.1 per year (random failures) and that the test interval is 0.1 year. The probability that a common mode

TABLE 3.9 POSSIBLE COMMON MODE FAILURES IN A REACTOR PROTECTION SYSTEM (Ref. 155)

REQUIRED FAILURE (EQUIPMENT AND MODE OF FAILURE)	EXTERNAL NORMAL ENVIRONMENT						DESIGN DEFICIENCY			OPERATION AND MAINTENANCE ERRORS						EXTERNAL PHENOMENA				FUNCTIONAL DEFICIENCY		
	DIRT	TEMPERATURE	MOISTURE	VIBRATION	WEAR	ELECTRICAL INTERFERENCE	UNRECOGNIZED INTERDEPENDENCE	COMMON ELEMENT	FAULT DEPENDENCE	MISCALIBRATION	IMPROPER TESTING	OUTDATED PRINTS	CARELESS MAINTENANCE	OPERATOR ERROR	INADEQUATE TRAINING	TORNADO	FIRE	FLOOD	EARTHQUAKE	PHENOMENOLOGICAL MISCONCEPTION	INADEQUATE PROTECTIVE ACTION	INADEQUATE INSTRUCTION
TRIP BREAKERS (DB-50 CIRCUIT BREAKER) FAIL TO OPEN CIRCUIT ON SIGNAL	X	X	X	X	X		X	X		X	X		X			X	X	X	X	X	X	X
TRIP RELAYS (BFD INDUSTRIAL CONTROL RELAY) FAIL TO OPEN CIRCUIT ON SIGNAL	X	X	X	X			X	X			X		X				X	X				X
LOGIC RELAYS (BF INDUSTRIAL CONTROL RELAYS) FAIL TO OPEN CIRCUIT ON SIGNAL	X	X	X	X			X	X			X		X				X	X				X
ANALOG CHANNELS (DIVERSE EQUIPMENT) FAIL TO REMOVE POWER TO RELAY COILS		X	X	X		X	X	X	X	X	X	X	X	X	X		X	X		X		X
PERMISSIVE FUNCTIONS (RELAYS AND SWITCHES) BYPASS ACTION OF LOGIC RELAYS	X	X	X	X		X	X	X	X	X	X	X	X	X	X		X	X		X		X
INTERCONNECTING WIRING SHORTS EQUIVALENT TO ABOVE FAULTS												X	X		X	X	X	X				
TEST CIRCUITRY BLOCK SIGNALS TO LOGIC OR TEST RELAYS											X	X	X				X	X				

NOTE: X IN BLOCK INDICATES POTENTIAL FOR COMMON-MODE FAILURE.

TABLE 3.10. COMMON MODE FAILURE EXPERIENCE AT ORNL (Ref. 157)

Item	Reactor	Class of failure	Environmental factor	Rate of Propagation	Description of failure; remarks
1	Tower Shielding Reactor (TSR)	Actual to completion	Disabled by accident	Instantaneous	The reactor was suspended between two towers by a cable from each in such a way that the severance of either cable would drop the reactor. Dropping the reactor or pulling a tower down by the cable hoist would be the maximum credible accident. Redundant switches were installed on a single bar to detect slack cable at the hoist. However, when one cable became slack, it struck away the bar and incapacitated the switches. Subsequently the cable was cut by the gears but was sufficiently fouled that it held. The hoist house was unshielded and hence unoccupied, so it was not until the following inspection that the condition was discovered.
2	Oak Ridge Graphite Reactor (X-10)	Actual to completion	Communication error	Instantaneous	In a test of the thermopile protection channels, all other protection channels were bypassed, and then all rods were withdrawn. The slowly responding thermopiles terminated the excursion at about 300% rated power.
3	Aberdeen Pulse Reactor (APR)*	Actual to completion	Disabled by accident	Instantaneous	When the reactor was pulsed during a test program at ORNL, the high ionization current destroyed the field-effect transistors in the flux amplifiers. These same amplifiers had been used in the High Flux Isotope Reactor and the Molten Salt Reactor Experiment, where protection response would have been fast enough to prevent the high current. A system with slower response would not have prevented the high current and would have allowed transistor failure. Protective diodes installed as a remedy must be tested periodically. In another installation, trouble developed from electrical-noise pickup which, it was found, could easily be remedied by removing the protective diodes. It now develops that the original field-effect transistors are no longer being manufactured and substitutes must be found for all existing amplifiers at all locations, with the overload problem being kept in mind.
4	Oak Ridge Research Reactor (ORR)	Arrested in progress	Change of characteristics	Instantaneous	Physicists flooded beam holes in accordance with procedures. This cut off the neutron beam and shielded the adjacent neutron detectors.
5	Low Intensity Test Reactor (LITR)	Arrested in progress	Unrecognized common element	Slow	The ionization chambers were purged continuously with gas from a common bottle. Contaminated gas caused most chambers to fail.
6	Low Intensity Test Reactor (LITR)	Arrested in progress	Change of characteristics	Slow	The temperatures at the ionization chambers were originally below 50°C. After 10 years of operation, the temperatures rose to 100°C and caused some of the chambers to fail.
7	Homogeneous Reactor Experiment No. 1 (HRE-1)	Arrested in progress	Unrecognized common element	Instantaneous	A steel housing was erected which enclosed a number of pneumatic devices. All affected instruments followed the pressure variations within the enclosure instead of the atmospheric pressure.
8	Molten Salt Reactor Experiment (MSRE)	Arrested in progress	Unrecognized common element	Instantaneous	Two d-c electric power supplies provided power of opposite polarities with respect to ground for a group of instruments used for protection and control. The positive voltage load was fail-safe, but it was discovered that loss of power to the negative voltage load was an unsafe failure.
9	Bulk Shielding Reactor (BSR-1)	Arrested in progress	Unrecognized common element	Instantaneous	All preamplifiers for protection and control were mounted on a single structural-steel member. During plant alterations it became necessary to unbolt the steel member and jerry-rig a support. The support sagged, strained the coaxial cables, and pulled loose the center conductors. Continuous monitoring detected the fault.
10	High Flux Isotope Reactor (HFIR)	Arrested in progress	Unrecognized common element	Instantaneous	It was necessary that the one rod with the greatest reactivity worth be stopped reliably. Redundant relays were provided to ensure reliability. Originally the two-phase motor stopped satisfactorily; however, when an improved power source was installed, the motor refused to stop when required. This failure was discovered on a periodic test.
11	Low Intensity Test Reactor (ORR)	Potential corrected	Unrecognized common element	Instantaneous	A single desiccant system served dry air to all coaxial cables for protection and control. These cables were under water, and a single leak could lower the signal level from all ionization chambers.
12	Low Intensity Test Reactor (LITR)	Potential, corrected	Change in characteristics	Indeterminate	The neoprene gasket between tank sections was embrittled by many years of radiation damage. A leak could wet the shielding material and cause neutron attenuation at the ionization chambers. The flux controller could raise the power, but the protection chambers would not indicate the power increase. The single-temperature channel would appear to be reading high incorrectly.
13	Oak Ridge Research Reactor (ORR)	Potential, corrected	Unrecognized common element	Instantaneous on loss of flow	A single electromechanical switch in the control system, which was never tested, has the capability, upon its failure, to defeat the low-flow protection system.
14	Oak Ridge Research Reactor (ORR)	Potential corrected	Communication error	Indeterminate	In an effort to improve maintenance procedures, instrument settings were typed and pasted near the related instruments. It was discovered that the typist had made an error and all identical instruments would have been incorrectly set.

*The accident occurred during a test at ORNL, not at the facility at Aberdeen.

failure will occur between two tests and for any number of channels is $0.01 \times 0.1 = 10^{-3}$. When random failures are considered the number of redundant channels is important, thus for one channel the probability of failure between two tests is $0.1 \times 0.1 = 10^{-2}$ and for a two-out-of-three logic scheme this probability is approximately 10^{-4}. It is seen that the probability of a common mode failure is 10 times larger and reliability calculations of random failures alone are not sufficient to demonstrate the safety of a system.

Similar rough calculations of the rate of occurrence of common mode failures in PWR's and BWR's were made by Williams (Ref. 153). Failures were reported and classified for the years 1969 and 1970 for both systems. It was found that in this period the common mode failures were 17% of the total number of failures for PWR systems and 25% for BWR systems. The main cause of such failures was identified to be human error for PWR's while for BWR's over half of the common mode failures were due to design deficiency. The rates of occurrence of these failures were estimated to be 1.67 per reactor year in PWR's and 2.24 per reactor year in BWR's, however, it should not be inferred that the rate of occurrence of common mode failures in BWR's is consistently higher for any year, since the data used for such calculations were not statistically significant.

3.C. HUMAN FACTORS AND SOFTWARE RELIABILITY

The effect that a human error can have on the safety of a system was discussed in the preceding section with regard to common mode failures. Operator errors were also included as events in the fault trees of Fig. 3.11 ("operator fails to notice the light") and Fig. 3.20 ("operator opens valve"); these events can be parts of minimal cut sets leading to the top event and the quantitative analysis of the trees requires that probability values be assigned to them.

The field of human factors has been studied extensively and its main task is the prevention of accidents; one of the qualitative methods of Sec. 3.A.9 dealt with personnel safety. However, the problem encountered here is of different nature since it is concerned with man-machine interactions and their quantitative description. The various aspects of the problem are discussed in Refs. 158, 159, and 160. The report by Garrick, Gekler, et al. (Ref. 161) is particularly relevant to the present discussion.

Human error can occur during testing, inspection, repair and operation. In most of the cases the system under study is automatic (e.g. reactor protection systems) and operational errors are not important. As it is to be expected, the probability of a human error increases as the number of functions that the operator is required to perform increases and as the time available decreases. Again the problem of good data arises; usually the available data come from laboratory experiments and not from actual situations but they do give a feeling of the order of magnitude of the probabilities. In Ref. 161, tables are provided which list several error rates for specific tasks (they have been compiled from other references listed in the report). In Table 3.11 we reproduce some of the data regarding the probability that a display will be read correctly or that a control device will be operated correctly. Table 3.12

TABLE 3.11 HUMAN RELIABILITY IN OPERATION OF CONTROLS AND DISPLAYS (Ref. 161)

Device and Parameter	Reliability
Counters	
Size (length), inches:	
1	0.9990
1 to 2	0.9998
3 and up	0.9995
Number of Drums or Digits:	
1 to 3	0.9997
4 to 5	0.9993
7 and up	0.9985
Lights	
Diameter, inches:	
Less than 1/4	0.9995
1/4 to 1/2	0.9997
1/2 to 1	0.9999
Number of lights on:	
1 or 2	0.9998
3 or 4	0.9975
5 to 7	0.9952
8 to 10	0.9945
Presentation	
Intermittent (blinking)	0.9998
continuous	0.9996
Push Buttons	
Size:	
miniature	0.9990
1/2 inch or more	0.9999
Number of push buttons in a group	
A. Single column or row	
1 to 5	0.9997
6 to 10	0.9995
11 to 25	0.9990
B. Double column or row or rows and column	
1 to 5	0.9997
6 to 10	0.9995
11 to 25	0.9990
C. Matrix	
6 to 10	0.9995
11 to 25	0.9995
25 or more	0.9985
Number of push buttons within group:	
2	0.9995
4	0.9991
8	0.9965
Distance between edges, inches:	
1/8 to 1/4	0.9985
3/8 to 1/2	0.9993
1/2 or more	0.9998
Detent:	
present	0.9998
absent (switch returns)	0.9993
Communicating	
Speaking	0.9998
Writing	0.9998
Recognition	0.9992
Decision Making	0.9992

TABLE 3.12. HUMAN RELIABILITY IN THE PERFORMANCE OF VARIOUS TASKS (Ref. 161)

Task Element	Estimated Reliability
Position multiple position electrical switch	0.9957
Install gasket	0.9962
Inspect for dents, cracks and scratches	0.9967
Tighten nuts, bolts and plugs	0.9970
Connect electrical cable (threaded)	0.9972
Inspect for air bubles (leak check)	0.9974
close hand valves	0.9983
Open hand valves	0.9985
Remove nuts, plugs and bolts	0.9988
Verify light illuminated or out	0.9996

shows the reliability per operation for certain tasks. If the failure rate λ is needed, we can calculate it by estimating the average time t required for the task and then divide the unreliability by t. In general, the error probability per operation lies in the interval 10^{-2} to 10^{-4}.

On a more theoretical level, we mention the work of Regulinski and Askren (Ref. 162). They conducted experiments involving continuous tasks and analyzed statistically the results. The conclusion was that the Weibull, gamma and log-normal distributions were reasonable models for the distributions of times-to-human failure, while the normal and exponential distributions were rejected.

All the models and methodologies presented thus far referred to the hardware of the system (human factors were also introduced in relation to their influence on the hardware). It has been recognized, however, that the reliability of the software should also be examined, since failures may emanate from it. Schick and Wolverton (Ref. 163) provide with the following definition: "Software reliability is defined to be the probability that the applications program, together with its operating system, data base, and computing environment, will perform its intended functions at the time when those functions are needed by the customer.".

Mathematical models dealing with the problem do not exist, although some attempts for quantification have been made. Investigations in the past have been mainly qualitative. In Ref. 164 the types of software errors are defined as deficiencies in fidelity, veracity and viability, where fidelity is the accuracy of mechanization of an algorithm for a given operating and hardware system, veracity is the adequacy of representation of a real problem by a given algorithm, and viability is the adequacy with which timing constraints are met by the mechanization of algorithm. This discussion concerns large programs which cannot be tested exhaustively so that all errors are discovered

and removed. The number of errors expected depends on the phase in which the program is: in the design and development phase (design of algorithm, preparation of flowcharts, subroutines, assembly of the program) large numbers of errors occur, but are not of interest to reliability analysts, since the program has not reached the user yet. When this happens, we can define a failure rate as usual which follows roughly the bath-tub curve (Ref. 165), that is, there is an initial phase (debugging time) where many errors are discovered and corrected and a subsequent period (useful life) where random errors occur, which are attributed by Williamson, et al. (Ref. 165) to inputs which cause out-of-tolerance outputs. The last part of the bath-tub curve (increasing failure rate) corresponds to the wear-out of components; computer programs, of course, do not wear-out, but we can say that by that time the algorithm is obsolete and must be replaced by a new one. If this modeling of the failure rate is accepted, standard reliability techniques can be applied to estimate software reliability, as it is done in Refs. 164 and 165.

The use of computers in reactor protection systems and the associated problems of hardware and software reliability is discussed by Hoermann in Ref. 166. Computerized functions include reactor scram, actuation of main steam, penetration and relief valves, turbine shut down and monitoring of temperature and coolant flow in the subassemblies of a breeder reactor (subject studied in detail in Ref. 112). An example of such a function involves a two-out-of-three protection system (Fig. 3.24). As shown in the figure there are two possibilities: in Fig. 3.24.b each computer receives only one signal and the majority voting is accomplished by interconnecting the computers, while in Fig. 3.24.c all three input signals are feeded into each computer, which operates independent of the others. The author proceeds to stress the need for both hardware and software reliability analyses of such systems. The discussion is qualitative and deals with the necessary actions that insure high software reliability (useful references in this context are 167 and 168).

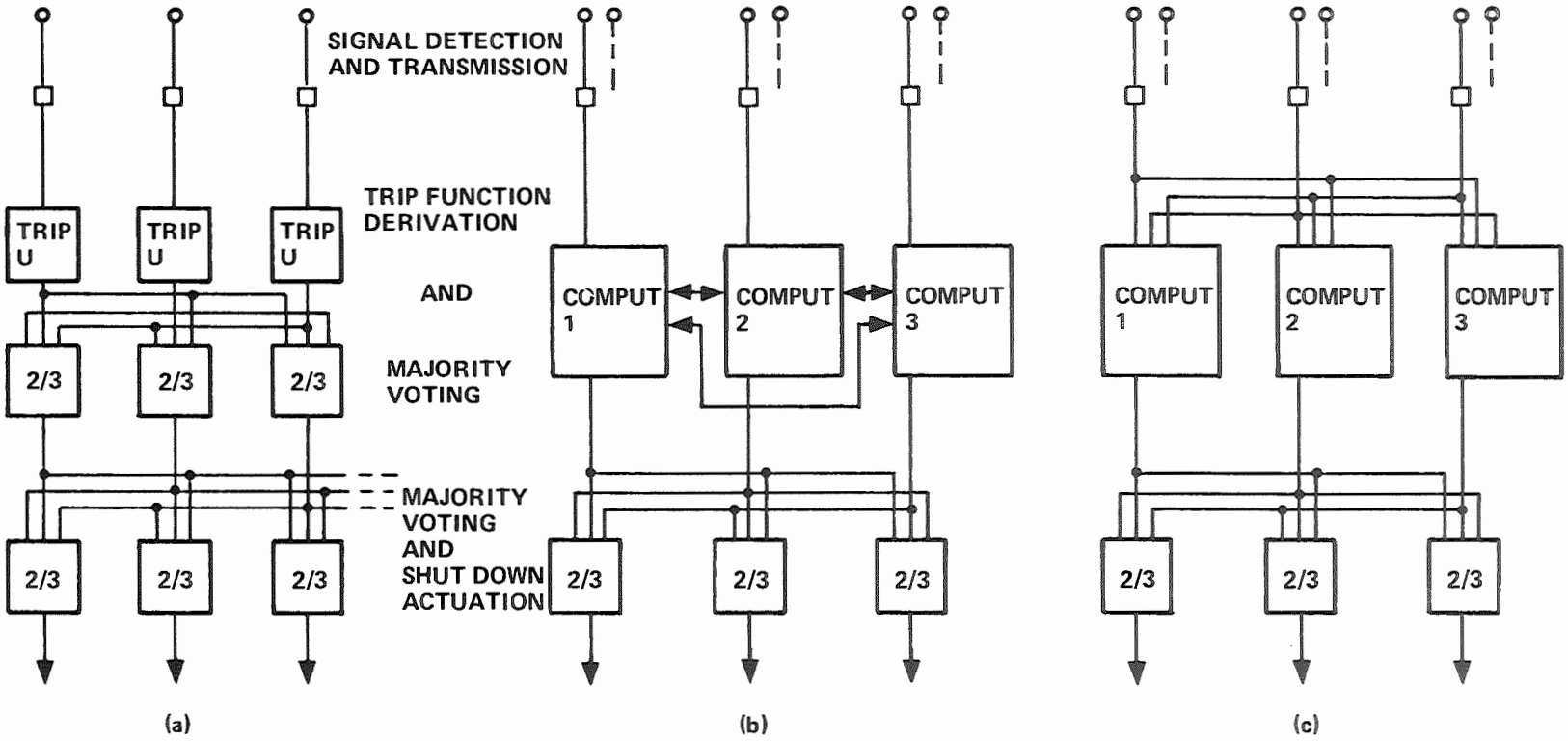


Figure 3.24. Two-out-of-three Protection System.
(a) Hard-Wired, (b) and (c) Computerized.
(Ref 166)

3.D. ANALYSES WITHOUT LOGIC DIAGRAMS

3.D.1 Introduction

The methods which have been presented in the preceding sections are the fundamental tools which enable one to study the safety of systems. The question as to which method is appropriate for a specific problem can be answered by looking at the available information. We recall that the general problem posed in quantitative safety studies is the evaluation of the probability that a specified event will occur in a period of time or that a specified function will be performed satisfactorily over a period of time or per demand. Then we can distinguish two cases:

i. The simplest case occurs when past experience from identical or basically similar situations as the one under investigation make it possible to calculate the required probabilities by statistical methods. The calculation may be non-parametric or a distribution function may be applicable, in which case its parameters are estimated from the data. Some simple probability laws may be used (e.g. the conditional probability theorem) in order to facilitate the calculations, but the procedure is essentially straightforward.

ii. This case is considerably more complex. It concerns studies involving multicomponent systems for which statistical data is lacking. Many factors influence the behavior of the system besides the failure properties of its parts, like inspection, maintenance, human operators, etc. In this case one exploits the hierarchical structure of the system to analyze it into simpler systems for which statistical failure data is available and the influence of maintenance can be incorporated in the calculation. Of course, logic diagrams are employed as well as qualitative methods for better understanding of the system.

In this section we focus our attention upon applications which do not require the use of logic diagrams. Of course, the methods themselves are not new.

3.D.2 Markov Models

The mathematics of Markov processes was presented in detail in Sections 2.D.4 and 2.D.5. It will be recalled that all the mutually exclusive states of the system must be enumerated and statistically significant data should be available for the estimation of the entries of the transition rate matrix. The number of states creates a problem, since it increases very rapidly with the complexity of the system; even in simple situations numerical methods are required to solve the Markov system of equations (Eq. 2.212).

Applications of Markov processes to power systems (transmission, distribution, bulk power supply systems) are discussed in Ref. 169.

Billinton and Lee (Ref. 170) present an application of the method to the reliability of the pumps of the heat transport system of a generating station. The station contains four 750 MW units and the heat transport system of each contains four 8,000 hp pumps. The pump configuration for each unit is assumed to operate independent of the configurations of the other units. This assumption enables one to study each unit separately with the advantage of fewer states in the Markov model; independence cannot be assumed if spare pumps are available for the whole station.

Each unit operates at full power when all four pumps are working and at 75% of power if one pump fails; failure of two or more pumps results to a shutdown of the unit. Fig. 3.25 shows the state space diagram for this case. Notice that in addition to the usual failure and repair rates (λ and μ) a third rate is also used, the rate of installation of a repaired unit γ.

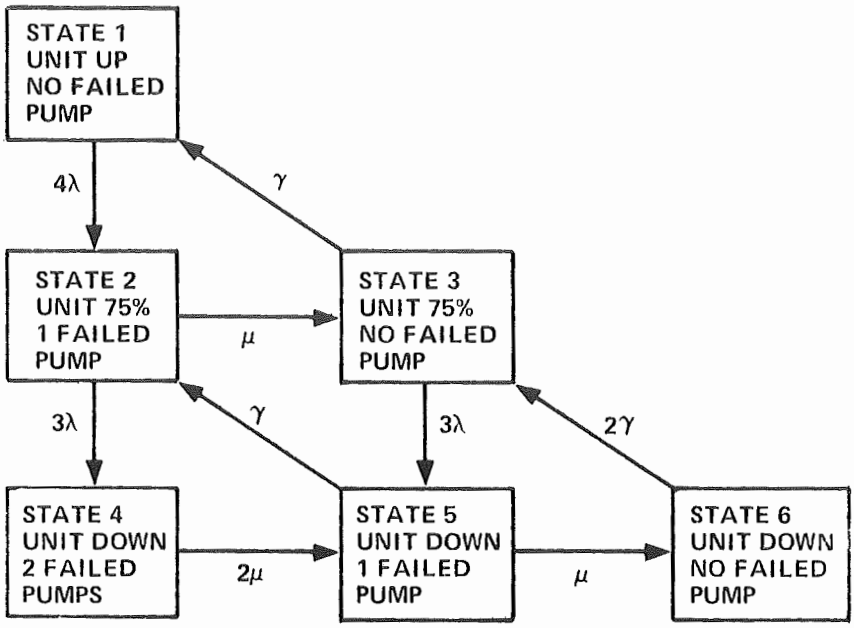


Figure 3.25. States and Transition Rates of One Generating Unit With Four Pumps. (Ref 170)

From the diagram we see that the transition rate matrix is

$$A \equiv \{a_{ij}\} = \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{pmatrix} -4\lambda & 4\lambda & 0 & 0 & 0 & 0 \\ 0 & -(3\lambda+\mu) & \mu & 3\lambda & 0 & 0 \\ \gamma & 0 & -(3\lambda+\gamma) & 0 & 3\lambda & 0 \\ 0 & 0 & 0 & -2\mu & 2\mu & 0 \\ 0 & \gamma & 0 & 0 & -(\mu+\gamma) & \mu \\ 0 & 0 & 2\gamma & 0 & 0 & -2\gamma \end{pmatrix} \end{array} \begin{array}{c} \text{States} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}$$

The steady-state probability vector $\underline{\Pi} = (\Pi_1,\ldots,\Pi_6)$ is calculated by solving the system

$$\begin{aligned} &\underline{\Pi}\, A = 0 \\ &\sum_{i=1}^{6} \Pi_i = 1 \end{aligned} \qquad (3.87)$$

Using the values $\lambda = 0.6\ yr^{-1}$, $\mu = 35.04\ yr^{-1}$ and $\gamma = 292\ yr^{-1}$ the numerical results on Table 3.13 are derived. The probability of the unit working at full power is Π_1 and at 75% of the power is $\Pi_2 + \Pi_3$. The frequency f_i of a state j per year is calculated from

$$f_i = \sum_j a_{ij}\, \Pi_i \qquad (3.88)$$

and the average fraction of the year spent in state i is calculated from

$$T_i = \frac{\Pi_i}{f_i} \qquad (3.89)$$

Since a pump itself is an item that is quite complex and can have a number of failure modes an improvement of the above model can be achieved by considering two types of failure: 1 — permanent failures which require actual removal of the pump and installation after repair, and 2 — temporary failures which can be repaired in a short time at the actual location of the pump. Assuming

TABLE 3.13. NUMERICAL RESULTS FOR THE MARKOV MODEL OF FIG. 3.25 (Ref. 170)

STATE VALUES				CAPACITY VALUES			
State	Probability	Frequency (per year)	Average duration (year)	Capacity	Probability	Frequency (per year)	Average duration (year)
1	0.9268535	2.2244485	0.4166667	100%	0.9268535	2.2244485	0.4166667
2	0.0634831	2.3387181	0.0271444	75%	0.0711011	2.3524304	0.0302245
3	0.0076180	2.2381608	0.0034037	0%	0.0020454	0.1279820	0.0159817
4	0.0016306	0.1142696	0.0142694				
5	0.0003913	0.1279820	0.0030577				
6	0.0000235	0.0137124	0.0017123				

that the temporary failure and repair rates are $\lambda' = 0.6\ yr^{-1}$ and $\mu' = 584\ yr^{-1}$ the state space diagram is shown on Fig. 3.26, where now there are ten possible states. The diagram has been drawn on three planes to facilitate reading

The authors proceed to study the effect of stand-by pumps on the availability of each unit and the problems that arise with regard to the number of states when all four generating units of the station must be investigated.'

Another application of Markov processes can be found in Ref. 171, where the reliability of the control rod drive system of a nuclear reactor (Otto Hahn nuclear ship) is investigated. Twenty system states are identified and the solution of the Markov system is carried out numerically with a program based on the Runge-Kutta method.

A fundamental assumption of Markov models is the constancy of the transition rates, which implies exponential distributions for the times spent by the system in each state. The field data, however, indicate in many situations that this assumption is not true. Di Marco (Ref. 90) and Patton (Ref. 172) study the reliability of generators and they point out that the available data suggest that the Weibull distribution is more appropriate to use than the exponential. In this case the results of Sec. 2.D.6 for non-Markovian systems apply.

3.D.3 Natural Phenomena

An important application of probabilistic models is in the study of natural phenomena (earthquakes, tornadoes etc.). The lack of real physical understanding of the natural processes which lead to such phenomena make it necessary to use the available statistical data in order to make predictions, even though many serious problems arise regarding the validity of the data and the models employed. Usually a distribution function is selected based on the knowledge we have about the nature of the phenomenon and its parameters are

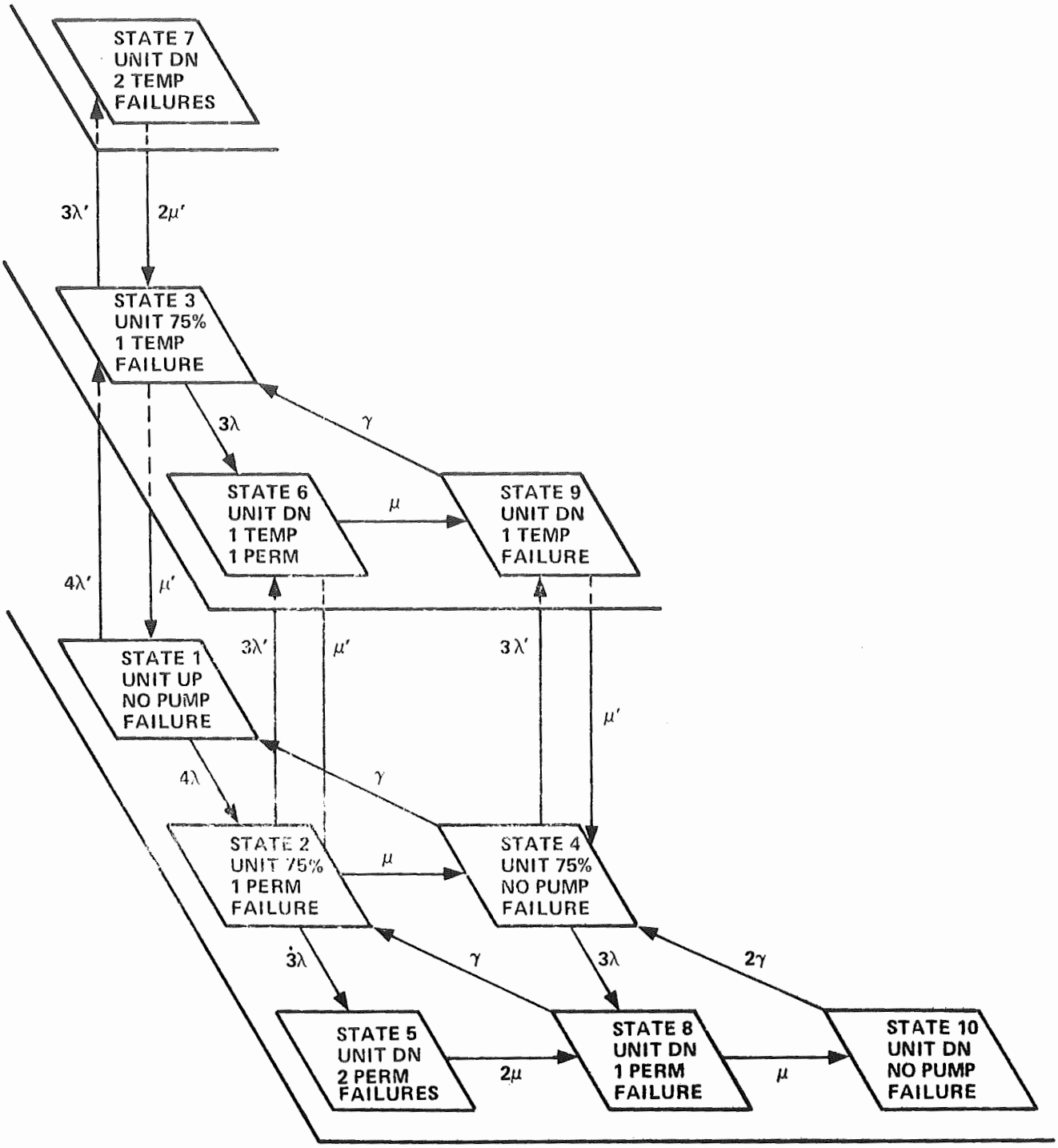


Figure 3.26. State Space Diagram for the Generating Unit With Two Failure Modes for Each Pump. (Ref 170)

estimated from the data. Table 3.14 is extracted from Ref. 173 and shows several applications of this approach and the corresponding references. Some theoretical aspects of probabilistic modeling of natural events are also examined by Hewitt in Ref. 173.

The first question is how the events occur in time; to answer it we use either a discrete distribution for the number of events occurring in a period of time or a continuous distribution for the time between successive events. Frequently a second random variable is associated with each event (e.g. the magnitude of an earthquake) and another continuous distribution is required to describe it. Two different approaches are possible in this case: either we use two distributions, as described above, one for the time occurrence and one for the magnitude, the latter being a conditional distribution in the sense that it gives the probability of the magnitude being in a certain interval given that an event has occurred, or we use one distribution which gives the probability of the largest or smallest magnitude over a period of time being in a certain interval.

The discrete distribution which is used widely in natural phenomena is the Poisson: if the average number of events per unit time is λ, then in the interval (0,t) the probability of k events is

$$p(k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \tag{3.90}$$

Basic assumptions for this distribution to be appropriate are: 1 — the characteristics of the phenomenon should be constant over the period of interest and 2 — non-overlapping time intervals are stochastically independent, that is, the number of events in an interval is not related with that in any other interval. Epstein and Lomnitz (Ref. 40) have used this distribution as a model of earthquake occurrence. The inter-arrival times are, of course, exponentially distributed, i.e.

TABLE 3.14 EXAMPLES OF NATURAL PHENOMENA DESCRIBED BY PROBABILITY DISTRIBUTIONS (Ref. 173)

Poisson Distribution	
1. Meteorite strikes on (potential) human targets	L. La Paz, Advances in Geophysics 4, Academic Press, 1958.
Negative Binomial	
1. Frequency of tornadoes	H.C.S. Thom, Monthly Weather Review, 1963
Gamma Distribution	
1. Sea waves: height	J.S. Longuet-Higgins, J. Marine Research, Vol. 11, 1952
2. River levels: recurrence of exceedances	C.A. McGilchrist, et al., Water Resources Research, Vol. 4, 1968, Vol. 5, 1969
3. Precipitation: drought occurrence	G.L. Barger and H.C.S. Thom, Agronomy Journ., Vol. 41, 1949.
Exponential Distribution	
1. River levels	McGilchrist et al., 1969.
Rayleigh and Weibull Distribution	
1. Wind Speed	A.G. Davenport, Wind Effects on Buildings and Structures, Univ. of Toronto Press, 1968.
2. Wave heights: trough-to-crest,	J.S. Longuet-Higgins, 1952.
Lognormal Distribution	
1. Tsunamis	W.G. Van Dorn, Advances in Hydroscience, Vol. II, Ven Te Chow, Ed., Academic Press, 1965.
2. Hydrologic Series (various examples)	Ven Te Chow, Proc. Amer. Soc. Div. Engr. Vol. 80, 1954
3. Tornadoes: dimension of damage Swath	H.C.S. Thom, Monthly Weather Review, 1963.
4. Flood damage magnitude: USA	Amer. Insur. Association, 1952-55
5. Earthquakes: magnitude and frequency	T. Asada, Journ. Phys. Earth (Tokyo), Vol. 5, 1957.

$$F(t) = 1 - e^{-\lambda t} \tag{3.91}$$

is the probability that one or more events occur in (0,t) and

$$R(t) = e^{-\lambda t} \tag{3.92}$$

is the probability of no events in (0,t). If we are interested in the distribution of the interval of time in which a specified number k of events occur, we must take the convolution of k exponentials, which leads to the gamma distribution (see Sects. 2.A.6 and 2.B.2), i.e.,

$$F(t) = 1 - \sum_{r=0}^{k-1} \frac{(\lambda t)^r}{r!} e^{-\lambda t} \quad , \qquad k = 1,2,... \tag{3.93}$$

A serious objection against the Poisson model is the requirement of independence of events. This is an idealization of the real world, since it is known that phenomena like earthquakes, tornadoes, floods tend to occur in clusters (in Ref. 26 this fact is pointed out for earthquakes). This problem of contagion (i.e. the occurrence of an event increases the probability of occurrence of another, a typical example being a contagious disease) has been studied and the usual approach is to modify the standard distributions to allow for the clustering (Refs. 173 and 174). For example, the Polya process is such a modification of a Poisson process (see also Feller, Ref. 1). Before using such models, however, it must be determined from the physical processes that occur, that indeed clustering is due to contagion and not to other reasons, such as uneven observations.

Having determined the distribution of the events in time the distribution of magnitudes should be found. Magnitude is treated as an independent random variable and it is, in most cases, continuous. Which distribution should be used is again determined from physical considerations and the available data. Thus, for earthquakes the distribution of magnitudes which is proposed in Ref. 40 is the exponential, i.e.

$$H(m) = 1 - e^{-\beta m} \tag{3.94}$$

which is interpreted as follows: given that an earthquake has occurred the probability that its magnitude is less than m is H(m).

From the assumed independence between frequency and magnitude it follows that, for example, the probability that n earthquakes will occur in (0,t) all having magnitude less than m is

$$P\left[n,\ M \leq m\right] = p(n)\left[H(m)\right]^n$$

$$= e^{-\lambda t}\frac{(\lambda t)^n}{n!}\left[1 - e^{-\beta m}\right]^n \tag{3.95}$$

The probability that all the earthquakes in (0,t) will have magnitude less than m is the sum of Eq. (3.95) over n, i.e.

$$G(M \leq m) = \sum_{n=0}^{\infty} p(n)\left[H(m)\right]^n$$

$$= \exp\left[-\lambda t e^{-\beta m}\right] \tag{3.96}$$

which is the extreme value distribution for the largest value, as expected.

A further implication of the independence of magnitude and frequency is that we can find the probability of exactly j earthquakes occurring in (0,t) with magnitude greater than m. If n earthquakes occur, then the probability that j of them have $M \geq m$ is given by the binomial distribution

$$P\left[j \text{ EQ's with } M \geq m/n \text{ EQ's occurred}\right] = \binom{n}{j}\left[1-H(m)\right]^j\left[H(m)\right]^{n-j}$$

The required probability is found by summing over n to allow for any number of earthquakes to have occurred, i.e.

$$P\left[j \text{ EQ's with } M \geq m\right] =$$

$$= \sum_{n=0}^{\infty} \binom{n}{j} \left[1-H(m)\right]^{j} \left[H(m)\right]^{n-j} e^{-\lambda t} \frac{(\lambda t)^{n}}{n!}$$

$$= e^{-\lambda t}\left[1-H(m)\right]^{j} \frac{(\lambda t)^{j}}{j!} \sum_{n=0}^{\infty} \frac{\left[\lambda t H(m)\right]^{n-j}}{(n-j)!}$$

$$= e^{-\lambda[1-H(m)]t} \frac{\{\lambda[1-H(m)]t\}^{j}}{j!} \tag{3.97}$$

This result shows that earthquakes with magnitude greater than m follow the Poisson distribution with rate

$$\lambda(m) = \lambda[1-H(m)] \tag{3.98}$$

These calculations show how the frequency and magnitude distributions can be combined to answer questions that may arise in applications. Of course, the exponential distribution is only one possible choice for the magnitude; other distributions that have been used incude the Rayleigh (wind speeds), the log-normal (tornadoes, floods), the Weibull (wind speeds), et al. (see Table 3.14 for references).

As stated earlier, the second approach to the magnitude-frequency problem involves extreme value distributions. We no longer assume independence of magnitude and frequency and their exact distributions need not be known. The most commonly applicable distribution is the Type I extreme value distribution of largest values

$$F(m) = \exp\left[-e^{-\alpha(m-\beta)}\right] \tag{3.99}$$

which gives the probability that the maximum magnitude in a specified period of time (usually a year) is less than m. Of course, Eqs. (3.96) and (3.99) are alternative expressions of the same distribution; there are however fundamental differences between the two, since Eq. (3.96) was derived from Eqs. (3.90) and (3.94), while for (3.99) no specific form of the initial distribution was

assumed (see also Sects. 2.A.6 and 2.B.2). This difference is reflected on the interpretation of the parameters and the method of their estimation from data.

The return period of the extremes of magnitude at least m is

$$T(m) = \frac{1}{1-F(m)} \quad \text{years} \tag{3.100}$$

where F(m) is given by Eq. (3.99). Thus it takes an average T(m) number of years for the annual maximum magnitude to be at least m once. If we wish to find the return period of a certain magnitude and not of the annual maximum we work with the parent population. From Eq. (3.98) we have that the number of events with magnitude greater than m is (per unit time) $\lambda(m)$, therefore the return period is

$$T'(m) = \frac{1}{\lambda[1-H(m)]} \tag{3.101}$$

or, using Eq. (3.94),

$$T'(m) = \frac{1}{\lambda} e^{\beta m} \quad \text{units of time} \tag{3.102}$$

T(m) and T'(m) are different for small values of m but for larger values the difference is insignificant (if m is large enough the population maximum and the annual maximum are the same). Shakal and Willis (Ref. 26) found that for earthquakes the two return periods are almost the same for values greater than 10 years. Besides earthquakes the extreme value distribution has been applied to the study of floods (Ref. 57), wind speeds (Ref. 175), etc. (see Gumbel, Refs. 16 and 17).

Finally, with regard to the general problem of modeling natural phenomena, Hewitt quotes Katti and Sly (Ref. 174) as follows:

> "1. No single theoretical distribution has been found to describe any large scale data.

2. For a number of data there could be two or more theoretical distributions that fit equally well and there is no way to choose between them based on fits alone.

3. Two or more physical models could lead to the same final statistical distribution and hence the estimation of the parameters of the distribution may not have unique meaning.

4. ... different methods of estimation lead to widely differing estimates when the methods are consistent ... there are a number of empirical frequencies to which the same theoretical frequency function has been fitted by different consistents methods..."

3.D.4 Various Probability Models

This section presents several examples of studies which show how simple probability relations, statistical data and empirical relations can be used to solve problems related to safety. Naturally, there is no general methodology which is universally applicable and each situation must be treated individually.

Four recent UCLA reports deal with probabilistic analysis of dam failures,[176] airplane crashes,[177] spills of toxic chemicals[178] and meteorite hazards.[179] In the last report the probability (per year) that a reactor will be directly hit or damaged by the heat generated by a near miss is calculated. If $A_{US} = 1.05 \times 10^{14}$ ft^2 is the area of the United States and A_{ET} is some area indicating the effective target, then the probability that N(W) meteorites with weight in some interval about W will damage the target is

$$P(W) = 1 - \left(1 - \frac{A_{ET}}{A_{US}}\right)^{N(W)} \tag{3.103}$$

The area A_{ET} for a given reactor consists of the area A_i of the plant itself plus the lethal area a(W) which is associated with each meteorite; this last area is included to account for the effect of a near miss. From Fig. 3.27 it follows that

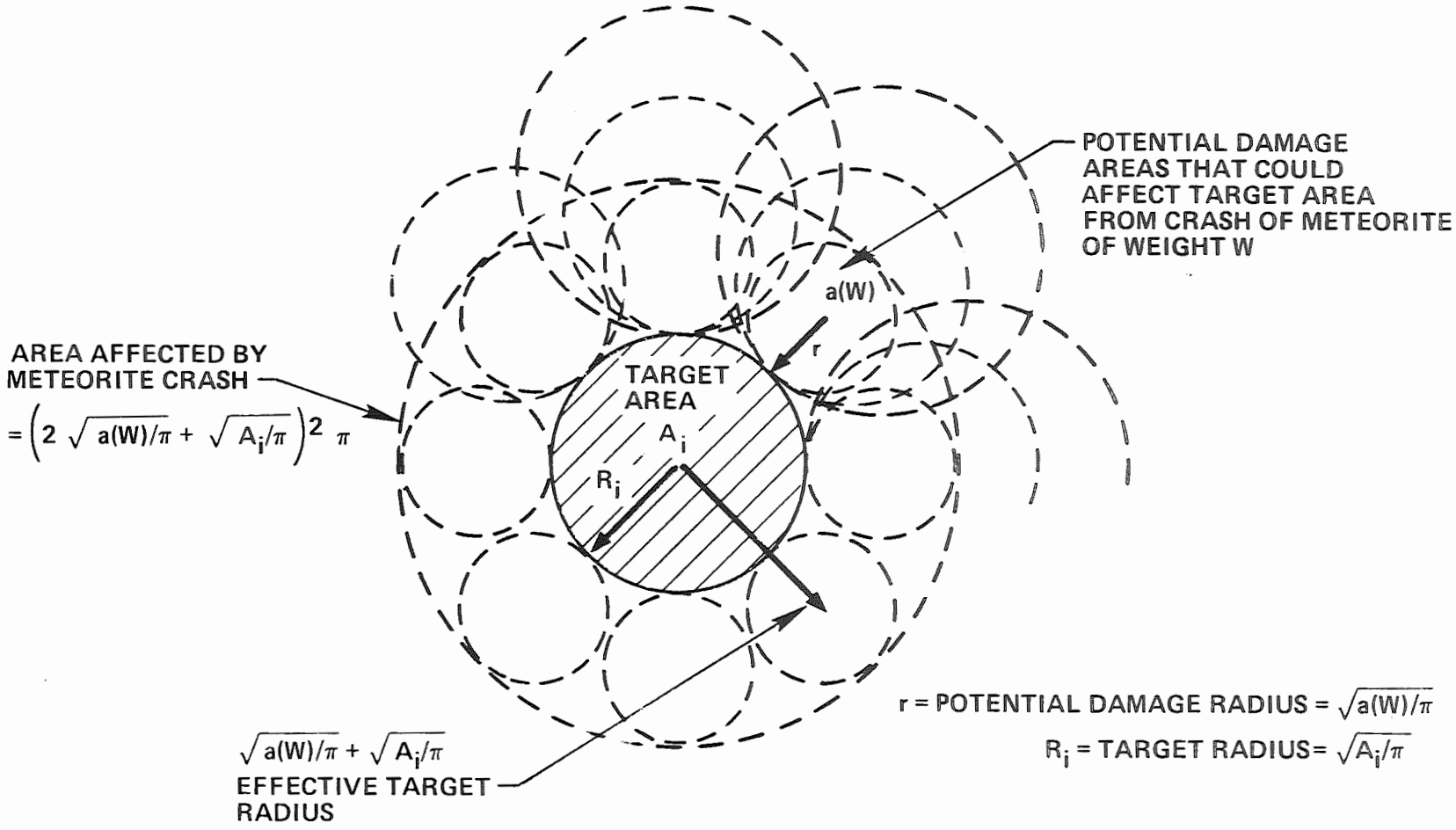


Figure 3.27. Comparison of Target Area and Affected Area for a Hypothetical Meteorite Crash; Assume Meteorite Can Only Crash on Land. (Ref 179)

$$A_{ET} = \left(\sqrt{A_i} + \sqrt{a(W)}\right)^2 \tag{3.104}$$

The number of meteorites in a certain weight range which hit the United States per year and the associated a(W) is shown on Table 3.15. The historical data are taken from Ref. 180. The vulnerable reactor area is estimated to be between 10^4 ft^2 and 10^5 ft^2. The table shows the probabilities as calculated from Eq. (3.103) and the potential damage. Since the first two weight ranges cannot damage the reactor, the probability of such a damage is calculated by summing the third through the sixteenth terms of the fourth column. For $A_i = 10^4$ ft. this probability is 7×10^{-10} per year and for $A_i = 10^5$ ft^2 it is 7×10^{-9} per year.

The important thing in such studies is to state clearly under what assumptions this use of the historical data is valid. The present problem is one of frequency-magnitude and it has been assumed that (a) the number of meteorites that fall is constant in one year and equal to 3500, (b) meteorites fall randomly throughout the surface of the earth; therefore, given that the area of the earth is 5.48×10^{15} ft^2 and that of the U.S. 1.05×10^{14} ft^2, we can find the number of meteorites per year falling on the United States with a simple calculation, i.e.

$$3500 \times \frac{1.05 \times 10^{14}}{5.48 \times 10^{15}} \cong 65$$

(c) the number of meteorites per year in a given weight range (as given on Table 3.15) does not change with time.

A different treatment of data is done by Bush in Ref. 181, where the probability of damage to critical plant components due to missiles from the turbine is calculated. This probability, P_4, is given by the product of three other probabilities, i.e.

TABLE 3.15. PROBABILITY OF A STONE OR IRON METEORITE HITTING AND DAMAGING A NUCLEAR REACTOR IN THE UNITED STATES (Ref. 179).

W Range of Meteorite Weight Hitting Earth (Tons)	N(W) Number of Meteorites In Weight Interval Per Year in United States	a(w) Potential Crash Area of Average Meteorite (ft^2)	P Probability of Hitting One Nuclear Reactor In United States $A_i=10^4$ ft^2	$A_i=10^5$ ft^2	Will Meteorite Cause Damage to Containment of a Nuclear Reactor? (Assume a Direct Hit)
$1/2 \times 10^{-3} - 10^{-3}$	45	2.5×10^{-1}	3×10^{-9}	3×10^{-8}	no
$10^{-3} - 10^{-2}$	12	2.1×10^{0}	8×10^{-10}	8×10^{-9}	very doubtful
$10^{-2} - 10^{-1}$	6	1.8×10^{1}	4×10^{-10}	4×10^{-9}	very possible
$10^{-1} - 10^{-0}$	2	1.1×10^{2}	2×10^{-10}	2×10^{-9}	certain rupture of containment
$10^{0} - 10^{1}$	14×10^{-2}	9.8×10^{2}	5×10^{-12}	5×10^{-11}	certain rupture of containment
$10^{1} - 10^{2}$	2×10^{-2}	1.9×10^{4}	5×10^{-12}	5×10^{-10}	serious damage
$10^{2} - 10^{3}$	6×10^{-3}	1.6×10^{5}	1×10^{-11}	3×10^{-11}	destroy nuclear reactor
$10^{3} - 10^{4}$	13×10^{-4}	1.1×10^{6}	5×10^{-12}	9×10^{-12}	destroy nuclear reactor
$10^{4} - 10^{5}$	2.1×10^{-4}	5.9×10^{6}	4×10^{-12}	5×10^{-12}	destroy nuclear reactor
$10^{5} - 10^{6}$	5.9×10^{-5}	5.9×10^{7}	2×10^{-12}	2×10^{-12}	destroy nuclear reactor and nearby area
$10^{6} - 10^{7}$	12×10^{8}	1.4×10^{8}	1×10^{-11}	1×10^{-11}	destroy nuclear reactor and surrounding areas
$10^{7} - 10^{8}$	2.6×10^{-6}	6.7×10^{8}	1×10^{11}	1×10^{-11}	destroy nuclear reactor and large surrounding area
$10^{8} - 10^{9}$	4.6×10^{-7}	3.0×10^{9}	8×10^{-12}	8×10^{-12}	" " "
$10^{10} - 10^{11}$	2.2×10^{-8}	6.7×10^{10}	7×10^{-12}	7×10^{-12}	" " "
$10^{11} - 10^{12}$	4.4×10^{-9}	3.0×10^{11}	1×10^{-11}	1×10^{-11}	" " "
$10^{12} - 10^{13}$	1.0×10^{-9}	1.4×10^{12}	1×10^{-11}	1×10^{-11}	" " "

$$P_4 = P_1 P_2 P_3 \tag{3.105}$$

where P_1 = probability of turbine failure and ejection of an energetic missile

P_2 = probability that such a missile strikes a critical component

P_3 = probability that the component suffers a significant damage.

P_1 is estimated from recorded turbine failures, which are shown on Table 3.16. The operating experience covers approximately 70,000 turbine years. The special feature of this historical record is that it cannot be assumed that the population of turbines, from which it was constructed, is homogeneous, because there has been an evolution in the design and manufacturing of the turbines over the years. Bush attributes the failures to three general causes: (a) metallurgical and/or design errors, (b) environmental effects, and (c) overspeed. As an example, the first category includes failures by brittle fracture due to retained oxygen and high nil-ductility temperatures; however, the processes of melting and heat treatment of the materials have been modified, so that such a failure is considered impossible now (the last one was recorded in 1956).

This continuous improvement of the quality of the turbines suggests that the failure rate is a decreasing function of time. This is the subject of a reliability-growth study (Refs. 25, 182); there are no standard rules indicating what model should be used and such a decision is largely a matter of judgment. The fact that the failure rate decreases with time does not imply that the bathtub curve is rejected. Referring to Fig. 3.28 we quote from Codier:[182] "... the bathtub curve is all right if it is understood that it describes the life-cycle behavior of a particular serial-numbered piece of hardware, but it has to be understood that the bathtub can be made to move up and down as the serial numbers change."

TABLE 3.16. CUMULATIVE TURBINE EXPERIENCE (Ref. 181)

Year	Plants New	Plants Total	Turbine Yr/Yr	Σ Turbine Yrs*	TOTAL FAILURES	TOTAL CUMULATIVE FAILURES	MISSILES FAILURES	MISSILES CUMULATIVE FAILURES
Pre-1950	-	1037	-	12,330	--	--		
1950	99	1136	1087	13,417	--	--		
1951	120	1256	1195	14,612	1	1	1	1
1952	108	1364	1310	15,922				
1953	149	1513	1385	17,360	1	2	0	1
1954	191	1704	1609	18,969	3	5	2	3
1955	146	1850	1777	20,746				
1956	127	1977	1914	22,660	1	6	0	3
1957	151	2128	2052	24,712				
1958	193	2321	2225	26,937				
1959	138	2459	2390	29,327	1	7	1+	4
1960	146	2605	2532	31,859	1	8	1+	5
1961	90	2695	2650	34,509				
1962	105	2800	2748	37,257				
1963	95	2895	2848	40,105				
1964	111	3006	2950	43,055				
1965	87	3093	3050	46,105				
1966	97	3190	3132	49,237				
1967	113	3303	3246	52,483				
1968	103	3406	3355	55,838				
1969	110	3516	3461	59,299	1	9	1+	6
1970	91	3607	3561	62,860				
1971	95	3702	3655	66,515				
1972	126	3828	3765	70,280	1	10	1+	7

* Values Σ Turbine Years Synthesized in Case of E E Data

\+ Overspeed, or out-of-phase, or generator failure

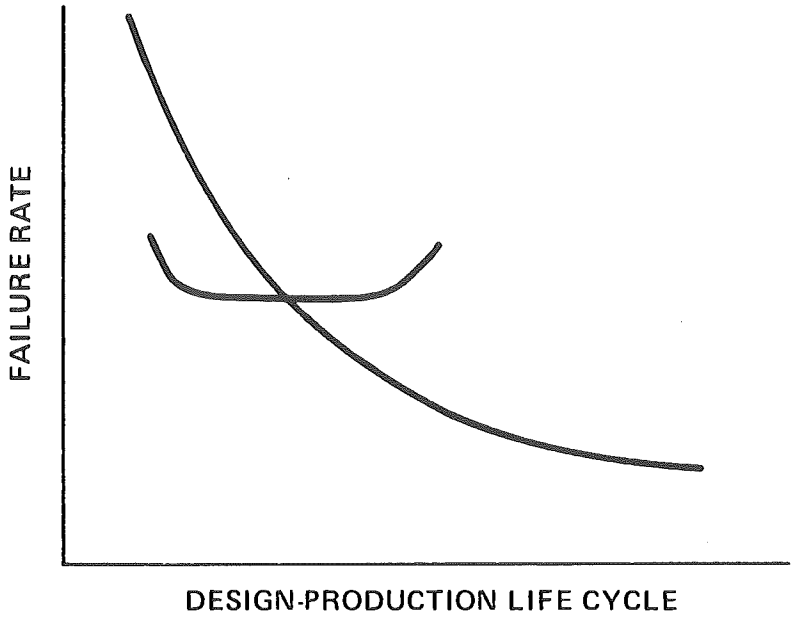


Figure 3.28. The Bathtub Curve and Reliability Growth. (Ref 182)

In the present problem the model used was the Duane growth model.[182] The cumulative failure rate λ_Σ is

$$\lambda_\Sigma \equiv \frac{F}{H} = KH^{-\alpha} \tag{3.106}$$

where H: total test time (years)

F: number of observed failures during H

k: a constant

α: growth rate usually in the range 0.3 to 0.5

If the data are plotted on a log-log paper and a straight line is fitted, the parameters k and α can be estimated. This is done in Fig. 3.29, where, in addition, the current value of the failure rate is shown. This current or instantaneous failure rate λ_i is interpreted as the failure rate of the equipments, if the reliability growth stopped at that time and it is calculated by differentiating Eq. (3.106) with respect to H, i.e.

$$\lambda_i = (1-\alpha)\lambda_\Sigma \tag{3.107}$$

The current failure rate is approximately 10^{-4} failures per year and it is projected to be $\sim 7 \times 10^{-5}$ within five years. The cumulative and instantaneous MTBF's are calculated by taking the inverse of the corresponding failure rates; the current MTBF is $\sim$ 10,000 turbine years and its projection after five years is 12,000 - 14,000 turbine years. From λ_i we estimate the value of P_1 as approximately equal to 10^{-4} per turbine year.

Having estimated P_1 there remains to find P_2 and P_3. The strike probability P_2 is estimated to be at most 10^{-3} for a target area of 1200 ft^2; the calculation takes into account the relative position and orientation of the turbine and the target area and it is described in detail by Bush. The damage probability P_3 is affected by the width of concrete protecting the critical components and the angle of incidence of the missile. Considering all the possibilities Bush concludes that the total probability P_4 is in the interval

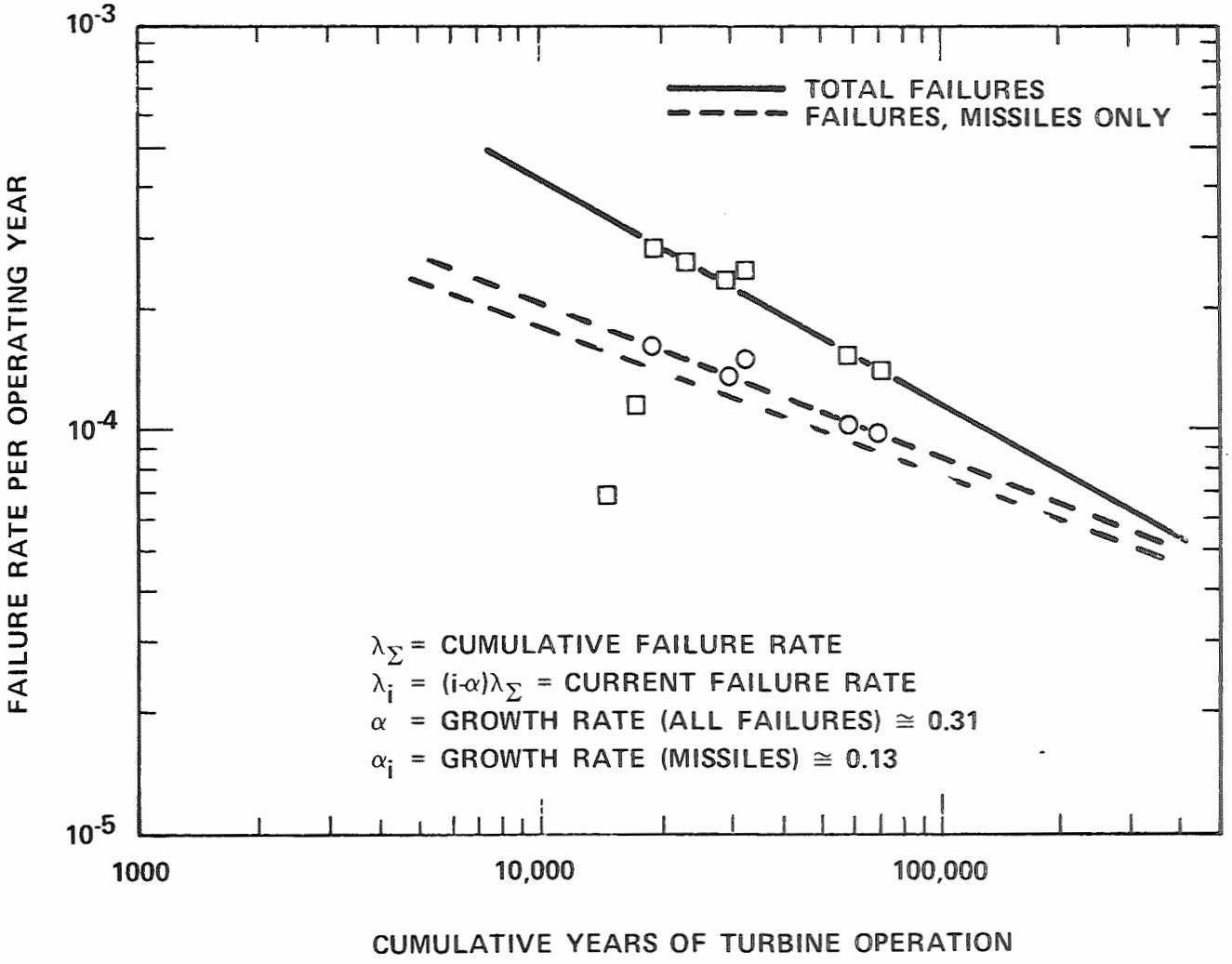


Figure 3.29. Cumulative and Instantaneous Failure Rates for Turbines as a Function of Cumulative Years of Turbine Operation From Table 3.16. (Ref 181)

10^{-6} to 10^{-8} per year. The problem of missile generation in a reactor plant is discussed in more general terms but without quantitative analysis by Gwaltney (Ref. 183) along with an extensive list of references.

Other probabilistic studies along the spirit of this section include the study of aircraft hazards for nuclear plants (Ref. 184) and a series of studies from UKAEA on the evaluation of the risk to a population resulting from the release of radioactivity (Refs. 185, 186 and 187).

REFERENCES

1. W. Feller, An Introduction to Probability Theory and its Applications, Vol. 1, 3rd Edition, John Wiley and Sons, New York, 1968.

2. A. Papoulis, Probability, Random Variables and Stochastic Processes, McGraw-Hill, New York, 1965.

3. E. Parzen, Modern Probability Theory and Its Applications, John Wiley and Sons, New York, 1960.

4. G.J. Hahn and S.S. Shapiro, Statistical Models in Engineering, John Wiley and Sons, New York, 1968.

5. R.E. Barlow and F. Proschan, Mathematical Theory of Reliability, John Wiley and Sons, New York, 1965.

6. ARINC Research Corporation, Reliability Engineering, Prentice-Hall, Englewood Cliffs, N.J., 1964.

7. I. Bazovsky, Reliability Theory and Practice, Prentice-Hall, Englewood Cliffs, N.J., 1961.

8. M.L. Shooman, Probabilistic Reliability: An Engineering Approach, McGraw-Hill, New York, 1968.

9. A.E. Green and A.J. Bourne, Reliability Technology, Wiley-Interscience, New York, 1972.

10. A. Papoulis, "The Abused Concept of Conditional Failure Rate," IEEE Trans. on Reliability, (Letters to the Editor), Vol. R-13, p. 62, June 1964.

11. D. Badenius, "Failure Rate/MTBF," IEEE Trans. on Reliability, Vol. R-19, No. 2, May 1970.

12. R.V. Hogg and A.T. Craig, Introduction to Mathematical Statistics, 2nd Edition, The Macmillan Company, New York, 1965.

13. H.L. Harter, New Tables of the Incomplete Gamma-Function Ratio and of Percentage Points of the Chi-square and Beta Distributions, Aerospace Research Laboratories, U.S. Air Force, 1964.

14. A. Hald, Statistical Tables and Formulas, John Wiley and Sons, New York, 1952.

15. K. Pearson, Tables of the Incomplete Beta Function, Biometrika Office, University College, London, 1948.

16. E.J. Gumbel, Statistical Theory of Extreme Values and Some Practical Applications, Nat. Bur. Std. Appl. Math. Ser. 33, 1954.

17. E.J. Gumbel, Statistics of Extremes, Columbia University Press, New York, 1958.

18. Probability Tables for the Analysis of Extreme Value Data, Nat. Bur. Std., Appl. Math. Ser. 22.

19. I.B. Gertsbakh and Kh. B. Kordonskiy, Models of Failure, Springer-Verlag, New York, 1969.

20. W.R. Buckland, Statistical Assessment of the Life Characteristic, Griffin's Statistical Monographs and Courses, No. 13, London, 1964.

21. N.D. Singpurwalla, "Statistical Fatigue Models: A Survey," IEEE Trans. on Rel., Vol. R-20, No. 3, Aug. 1971.

22. J.H.K. Kao, "A Graphical Estimation of Mixed Weibull Parameters in Life Testing of Electron Tubes," Technometrics, 1, 389, Nov. 1959.

23. J. Lieblein and M. Zelen, "Statistical Investigation of the Fatigue Life of Deep-Groove Ball Bearings," J. Res. Nat. Bur. Std., 57, 273, 1956.

24. A.M. Freudenthal, The Inelastic Behavior of Engineering Materials and Structures, John Wiley and Sons, New York, 1950.

25. D. Lloyd and M. Lipow, Reliability: Management, Methods and Mathematics, Prentice-Hall; Englewood Cliffs, N.J., 1964.

26. A.F. Shakal and P.E. Willis, "Estimated Earthquake Probabilities in the North Circum-Pacific Area," Bull. Seism. Soc. of Am., Vol. 62, 1397, 1972.

27. R.P. Haviland, Engineering Reliability and Long Life Design, D. Van Nostrand Co, Princeton, N.J., 1964.

28. J.H. Bompas-Smith, "The Determination of Distributions that Describe the Failures of Mechanical Components," Eighth Reliability and Maintainability Conference, 343, 1969.

29. D. Kececioglu, "Reliability Analysis of Mechanical Components and Systems," Nucl. Engin. and Design, 19, 259, 1972.

30. L. Caldarola, "A New Definition of Reliability, Continuous Lifetime Prediction and Learning Processes," Presented at the International NATO Conference on Reliability, Liverpool, England 16-17 July, 1973.

31. T. Yokobori, An Interdisciplinary Approach to Fracture and Strength of Solids, Wolters-Noordhoff, Groningen, Holland, 1968.

32. H.S. Endicott and T.M. Walsh, "Accelerated Testing of Component Parts," Annual Symposium on Reliability, 570, 1966.

33. D. Kececioglu, R.E. Smith and E.A. Felsted, "Distributions of Cycles-to-Failure in Simple Fatigue and the Associated Probabilities," Eighth Reliability and Maintainability Conference, 357, 1969.

34. E. Parzen, "On Models for the Probability of Fatigue Failure of a Structure," Time Series Analysis Papers, Holden-Day, San Francisco, 1967.

35. T.W. Calvin, "Modeling the Bathtub Curve," Annual Reliability and Maintainability Symposium, 577, 1973.

36. A.D. Soloviev, "Theory of Aging Elements," Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. III, University of California Press, 1967.

37. R.B. Schwartz, S.M. Seltzer and F.N. Stehle, "Failure Distribution Analysis," Fourth Reliability and Maintainability Conference, 817, 1965.

38. J.M. Nordquist, "Theory of Largest Values Applied to Earthquake Magnitudes," Trans. Am. Geophys. Union, 26, 29, 1945.

39. C. Lomnitz, "Statistical Prediction of Earthquakes," Reviews of Geophysics, 4, 377, 1966.

40. B. Epstein and C. Lomnitz, "A Model for the Occurrence of Large Earthquakes," Nature, 211, 954, 1966.

41. C.F. Richter, Elementary Seismology, 359, W.H. Freeman, San Francisco, 1958.

42. C.M. Ryerson, "Reliability Modeling and Microcircuit Prediction," Eighth Reliability and Maintainability Conference, 191, 1969.

43. B. Epstein, "Estimation from Life Test Data," Technometrics, 2, 447, Nov. 1960.

44. J.A. Greenwood and D. Durand, "Aids for Fitting the Gamma Distribution by Maximum Likelihood," Technometrics, 2, 55, Feb. 1960.

45. H.L. Harter and A.H. Moore, "Maximum Likelihood Estimation of the Parameters of Gamma and Weibull Populations from Complete and from Censored Samples," Technometrics, 4, 639, Nov. 1965.

46. A.H. Bowker and G.J. Lieberman, Engineering Statistics, Second Ed.; Prentice-Hall, Englewood Cliffs, N.J., 1972.

47. W.R. Mann, "Point and Interval Estimation Procedures for the Two-Parameter Weibull and Extreme-Value Distributions," Technometrics, 10, 231, May 1968.

48. A.C. Cohen, "Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and on Censored Samples," Technometrics, 7, 579, Nov. 1965.

49. M.V. Johns and G.J. Lieberman, "An Exact Asymptotically Efficient Confidence Bound for Reliability in the Case of the Weibull Distribution," Technometrics, 8, 135, Feb. 1966.

50. D.R. Thoman, L.J. Bain and C.E. Antle, "Inferences on the Parameters of the Weibull Distribution," Technometrics, 11, 445, August, 1969.

51. D.R. Thoman, L.J. Bain and C.E. Antle, "Maximum Likelihood Estimation, Exact Confidence Intervals for Reliability, and Tolerance Limits in the Weibull Distribution," Technometrics, 12, 363, May 1970.

52. N.R. Mann, "Tables for Obtaining the Best Linear Invariant Estimates of Parameters of the Weibull Distribution," Technometrics, 9, 629, Nov. 1967.

53. N.R. Mann and K.W. Fertig, "Tables for Obtaining Weibull Confidence Bounds and Tolerance Bounds Based on Best Linear Invariant Estimates of Parameters of the Extreme-Value Distribution," Technometrics, 15, 87, Feb. 1973.

54. B.F. Kimball, "Sufficient Statistical Estimation Functions for the Parameters of the Distribution of Maximum Values," Ann. of Math. Stat., 17, 299, 1946.

55. B.F. Kimball, "An Approximation to the Sampling Variance of an Estimated Maximum Value of Given Frequency Based on Fit of Doubly Exponential Distribution of Maximum Values," Ann. of Math. Stat., 20, 110, 1949.

56. J.T. De Oliveira, "Statistics for Gumbel and Fréchet Distributions," International Conference on Structural Safety and Reliability, A.M. Freudenthal, Ed., Pergamon Press, N.Y. 1972.

57. I.B. Wall, "Probabilistic Assessment of Flooding Hazard for Nuclear Power Plants," Trans. Am. Nucl. Soc., 16, 211, 1973.

58. T.J. Boardman and P.J. Kendell, "Estimation in Compound Exponential Failure Models," Technometrics, 12, 891, Nov. 1970.

59. T.J. Boardmann, "Estimation in Compound Exponential Failure Models-When the Data are Grouped," Technometrics, 15, 271, May 1973.

60. A.C. Cohen, "Estimation in Mixtures of Two Normal Distributions," Technometrics, 9, 15, Feb. 1967.

61. V. Hasselblad, "Estimation of Parameters for a Mixture of Normal Distributions," Technometrics, 8, 431, August 1966.

62. P.R. Rayment, "The Identification Problem for a Mixture of Observations from Two Normal Populations," Technometrics, 14, 911, Nov. 1972.

63. L.W. Falls, "Estimation of Parameters in Compound Weibull Distributions," Technometrics, 12, 399, May 1973.

64. W. Nelson, "Hazard Plotting for Incomplete Failure Data," Journal of Quality Technology, 1, 27, Jan. 1969.

65. W. Nelson, "Hazard Plotting Methods for Analysis of Life Data with Different Failure Modes," Journal of Quality Technology, 2, 126, July 1970.

66. W. Nelson, "Theory and Applications of Hazard Plotting for Censored Failure Data," Technometrics, 14, 945, Nov. 1972.

67. R.S. Pringle and P.M. Gresho, "A Comprehensive Reliability Analysis of Redundant Systems," J. Spacecraft and Rockets, 4, 631, May 1967.

68. D.R. Cox, Renewal Theory, Methuen, London, 1962.

69. G.H. Weiss, "A Survey of Some Mathematical Models in the Theory of Reliability," in Statistical Theory of Reliability, M. Zelen Ed., The University of Wisconsin Press, Madison, Wis., 1963.

70. E. Parzen, Stochastic Processes, Holden-Day, San Francisco, Calif., 1962.

71. W.L. Smith and M.R. Leadbetter, "On the Renewal Function for the Weibull Distribution," Technometrics, 5, 393, August 1963.

72. R.M. Soland, "A Renewal Theoretic Approach to the Estimation of Future Demand for Replacement Parts," Operations Research, 16, 36, Jan-Feb. 1968.

73. S.S. Gupta, "Order Statistics from the Gamma Distribution," Technometrics, 2, May 1960.

74. E.L. Peterson and H.B. Loo, "Maintainability Risk Analysis Using the Analytical Maintenance Model," Sixth Reliability and Maintainability Conference, 498, 1967.

75. E.J. Muth, "A Method for Predicting System Downtime," IEEE Trans. on Rel., Vol. R-17, No. 2, June 1968.

76. E.J. Muth, "On the Distribution and Prediction of Excess Time," Sixth Reliability and Maintainability Conf., p. 206, 1967.

77. E.J. Muth, "Excess Time, a Measure of System Repairability," IEEE Trans. on Rel., Vol. R-19, No. 1, Feb. 1970.

78. G.H. Sandler, System Reliability Engineering, Prentice-Hall, Englewood Cliffs, N.J., 1963.

79. L.A. Zadeh and C.A. Desoer, Linear System Theory, McGraw-Hill, New York, 1963.

80. J.A. Buzacott, "Markov Approach to Finding Failure Times of Repairable Systems," IEEE Trans on Rel., Vol. R-19, No. 4, Nov. 1970.

81. J. Kemeny and J. Snell, Finite Markov Chains, Van Nostrand, Princeton, N.J., 1960.

82. S. Karlin and J. McGregor, "The Classification of Birth and Death Processes," Trans. Am. Math. Soc., Vol. 86, p. 366, 1957.

83. D.R. Cox and W.L. Smith, Queues, Methuen, London, 1961.

84. S.L. Surana and A. Brameller, "Application of Reliability Theory to Computation of System Operation Security in a Power Industry," Generic Techniques in Systems Reliability Assessment, NATO Advanced Study Institute, Univ. of Liverpool, July 17-27, 1973.

85. R. Pyke, "Markov Renewal Processes: Definitions and Preliminary Properties," Ann. Math. Stat., 32, 1231, 1961.

86. R. Pyke, "Markov Renewal Processes with Finitely Many States," Ann. Math. Stat., 32, 1243, 1961.

87. S. Ross, Applied Probability Models with Optimization Applications, Holden-Day, San Francisco, 1970.

88. M.H. Branson and B. Shah, "Reliability Analysis of Systems Comprised of Units with Arbitrary Repair-Time Distributions," IEEE Trans. on Rel., Vol. R-20, No. 4, Nov. 1971.

89. S. Osaki, "On a Two-Unit Standby-Redundant System with Imperfect Switchover," IEEE Trans. on Rel., Vol. R-21, No. 1, Feb. 1972.

90. A. Di Marco, "A Semi-Markov Model of a Three-State Generating Unit," IEEE Trans. on Power Apparatus and Systems, Vol. PAS-91, No. 5, Sept.-Oct. 1972.

91. S.K. Srinivasan and M.N. Gopalan, "Probabilistic Analysis of a 2-Unit Cold-Standby System with a Single Repair Facility," IEEE Trans. on Rel., Vol. R-22, No. 5, Dec. 1973.

92. R.E. Barlow and F. Proschon, Availability Theory for Multicomponent Systems, Research Report ORC 72-8, University of California, Berkeley, April 1972.

93. S. Osaki, "Renewal Theoretic Aspects of Two-Unit Redundant Systems," IEEE Trans. on Rel., Vol. R-19, No. 3, August 1970.

94. D.K. Chow, "Reliability of Some Redundant Systems with Repair," IEEE Trans. on Rel., Vol. R-22, No. 4, Oct. 1973.

95. R.E. Barlow, "Maintenance and Replacement Policies," in Statistical Theory of Reliability, M. Zelen, Ed., The University of Wisconsin Press, Madison, Wis., 1963.

96. I.M. Jacobs, "Reliability of Engineered Safety Features as a Function of Testing Frequency," Nuclear Safety, Vol. 9, No. 4, July-Aug. 1968.

97. H.M. Hirsh, "Setting Test Intervals and Allowable Bypass Times as a Function of Protection System Goals," IEEE Trans. on Nucl. Sci., Vol. 18, No. 1, Feb. 1971.

98. H.M. Hirsh, Methods for Calculating Safe Test Intervals and Allowable Repair Times for Engineered Safeguard Systems, General Electric Report NEDO-10739, Jan. 1973.

99. IEEE Standard 352-1972 - Trial Use Guide: General Principles for Reliability Analysis of Nuclear Power Generating Station Protection Systems.

100. L.C. Hunter, Optimum Checking Procedures," in Statistical Theory of Reliability, M. Zelen Ed., The University of Wisconsin Press, Madison, 1963.

101. B.J. Flehinger, "A General Model for the Reliability Analysis of Systems Under Various Preventive Maintenance Policies," Ann. Math. Stat., Vol. 33, 137, 1962.

102. J. Coleman and J. Abrams, "Mathematical Model for Operational Readiness," J. Oper. Res. Soc., Vol. 10, 126, 1962.

103. B.J. Flehinger, "A Markovian Model for the Analysis of the Effects of Marginal Testing on System Reliability," Ann. Math. Stat., Vol. 33, 754, 1962.

104. F.E. Hohn, Applied Boolean Algebra, An Elementary Introduction, The Macmillan Co., N.Y., 1960.

105. B.J. Garrick, "Principles of Unified Systems Safety Analysis," Nucl. Eng. and Des., Vol. 13, 245, 1970.

106. Z.W. Birnbaum, J.D. Esary and S.C. Saunders, "Multi-component Systems and Structures and their Reliability," Technometrics, 3, 55, 1961.

107. J.D. Esary and F. Proschan, "Coherent Structures of Non-Identical Components," Technometrics, 5, 191, 1963.

108. D.F. Haasl, "Advanced Concepts in Fault Tree Analysis," System Safety Symposium, June 8-9, 1965, Seattle, Wash., The Boeing Company.

109. J.B. Fussell, "A Formal Methodology for Fault Tree Construction," Nucl. Sci. and Eng., 52, 421, 1973.

110. J.D. Murchland, "Fundamental Probability Relations for Repairable Items," NATO Advanced Study Institute on Generic Techniques in Systems Reliability Assessment, The University of Liverpool, July 17-27, 1973.

111. J.D. Murchland and G.G. Weber, "A Moment Method for the Calculation of a Confidence Interval for the Failure Probability of a System," Annual Reliability and Maintainability Symposium, 565, 1972.

112. S.K.W. Jacobi, R.O. Schneider and G.G. Weber, "Reliability and Availability of a Safety Shutdown System," Annual Reliability and Maintainability Symposium, 186, 1974.

113. J.A. Buzacott, "Finding the MTBF of Repairable Systems by Reduction of the Reliability Block Diagram," Microelectronics and Reliability, Vol. 6, 105-112, 1967.

114. J.A. Buzacott, "Network Approaches to Finding the Reliability of Repairable Systems," IEEE Trans. on Rel., Vol. R-19, No. 4, Nov. 1970.

115. E.R. Woodcock, "The Calculation of Reliability of Systems: the Program NOTED," UKAEA Report AHSB(S) R153, 1971.

116. W.E. Vesely and R.E. Narum, "PREP and KITT Computer Codes for the Automatic Evaluation of a Fault Tree," Idaho Nuclear Corporation, IN-1349.

117. S.N. Semanderes, "ELRAFT, a Computer Program for the Efficient Logic Reduction Analysis of Fault Trees," IEEE Trans. on Nucl. Sci., Vol. NS-18, No. 1, Feb. 1971.

118. J.B. Fussell and W.E. Vesely, "A New Methodology for Obtaining Cut Sets for Fault Trees," Trans. Am. Nucl. Soc., Vol. 15, 262, 1972.

119. W.E. Vesely, "A Time-Dependent Methodology for Fault-Tree Evaluation," Nucl. Eng. and Design, Vol. 13, 337, 1970.

120. J.M. Hammersley and D.C. Handscomb, Monte Carlo Methods, John Wiley and Sons, N.Y., 1964.

121. M. Clark, Jr. and K.F. Hansen, Numerical Methods of Reactor Analysis, Academic Press, N.Y., 1964.

122. P.A. Crosetti, "Computer Program for Fault Tree Analysis," DUN-5508, April 1969.

123. P. Nagel, "Importance Sampling in Systems Simulation," Fifth Reliability and Maintainability Conference, 330, 1966.

124. B. Schallopp and L. Kamarinopoylos, "Problems and Limitations of Reliability Calculations for Complex Systems," International Atomic Energy Agency, Symposium on Principles and Standards of Reactor Safety, 5-9 February, 1973, Paper IAEA/SM-169/21.

125. J.R. Rosenblatt, "Confidence Limits for the Reliability of Complex Systems," in Statistical Theory of Reliability, M. Zelen Ed., The University of Wisconsin Press, Madison, Wis., 1963.

126. N.R. Mann, R.E. Schafer and N.D. Singpurwalla, Methods for Statistical Analysis of Reliability and Life Data, John Wiley and Sons, New York, 1974.

127. B.J. Garrick, W.C. Gekler, L. Goldfisher, R.H. Karcher, B. Shimizu and J.H. Wilson, "Reliability Analysis of Nuclear Power Plant Protective Systems," Holmes and Narver, Inc., HN-190, May 1967.

128. J.A. Buzacott, A.H. Weaving and T.A. Wesolowski-Low, "Quantitative Safety," Trans. of the Soc. of Instr. Technol., Vol. 19, 60, 1967.

129. B.M. Tashjian, "Sensitivity Analysis of a Two-out-of-four Coincidence Logic Reactor Protective System," IEEE Trans. on Nucl. Sci., Vol. NS-18, No. 1, Feb. 1971.

130. E.R. Snaith, "Reliability Analysis of an Electrical Supply System for a Nuclear Power Station," Nucl. Eng. and Design, Vol. 13, 216, 1970.

131. A. Danielsen and E.R. Snaith, "The Use of Reliability Analysis Techniques Applied to Nuclear Power Station Emergency Core Cooling Systems," Specialist Meeting on the Development and Application of Reliability Techniques to Nuclear Plant, Liverpool, 8-10 April, 1974.

132. R.J. Feutz and T.A. Waldeck, "The Application of Fault Tree Analysis to Dynamic Systems," System Safety Symposium; June 8-9, 1965, Seattle, Wash, The Boeing Company.

133. P.A. Crosetti and R.A. Bruce, "Commercial Application of Fault Tree Analysis," Ninth Reliab. and Maint. Conf., 220, 1970.

134. D.E. Cole, "N-Reactor Confinement System Safety Analysis," Douglas United Nuclear, Inc., Report DUN-5890, 1969.

135. R. Salvatori, "Systematic Approach to Safety Design and Evaluation," IEEE Trans. on Nucl. Sci., Vol. NS-18, No. 1, 1971.

136. H.P. Balfanz, "Prediction of Failure Rates and Failure Modes of Mechanical and Electrical Equipment Components by the Fault Tree Method and Failure Effect Analysis," Kerntechnik, Vol. 13, No. 9, 1971.

137. R.M. Stewart and G. Hensley, "High Integrity Protective Systems on Hazardous Chemical Plants," CREST Specialist Meeting on Applicability of Quantitative Reliability Analysis of Complex Systems and Nuclear Plants in its Relation to Safety, Munich, May 26-28, 1971.

138. C.W. Griffin, "The Fault Tree as a Safety Optimization Design Tool," CONF-730304, Topical Meeting on Water-Reactor Safety, Salt Lake City, March 26-28, 1973.

139. H. Hörtner, E. Dressler, E. Nieckau and H. Spindler, "Reliability Investigation of the Reactor Safety System Used to Control a Loss-of-Coolant Accident," Specialist Meeting on the Development and Application of Reliability Techniques to Nuclear Plant, Liverpool, April 8-10, 1974.

140. W. Bastl, "Reliability of Nuclear Plant," NATO Advanced Study Institute on Generic Techniques in Systems Reliability Assessment, The University of Liverpool, July 17-27, 1973.

141. W.C. Gangloff, S.J. Sarver and T. Franke, "Evaluation of Safeguards Reliability-Containment Spray," Specialist Meeting on the Development and Application of Reliability Techniques to Nuclear Plants, Liverpool, April 8-10, 1974.

142. R.C. Erdmann, D. Okrent, P. Godbout and K.A. Solomon, "Fault Tree Analysis of Reactor Safety Systems with Application to the Residual Heat Removal System of a BWR," CONF-730414-P2, Topical Meeting on Mathematical Models and Computational Techniques for Analysis of Nuclear Systems, Ann Arbor, Mich., April 9-11, 1973.

143. P.J. Godbout, "A Probabilistic Approach for Determining Nuclear Reactor Control System Reliabilities and Related Accident Frequencies," Ph.D. Thesis, Energy and Kinetics Dept., UCLA, 1973.

144. H. Raiffa, Decision Analysis, Addison-Wesley, 1968.

145. F.R. Farmer, "Siting Criteria - A New Approach," paper SM-89/34, IAEA Symposium on the Containment and Siting of Nuclear Power Plants, Vienna, April, 1967.

146. H. Otway and R.C. Erdmann, "Reactor Siting from a Risk Viepoint," Nucl. Eng. and Design, Vol. 13, 365-376, 1970.

147. I.M. Jacobs, "Assessment of Risk in Reactor Safety," CONF-730304, Topical Meeting on Water-Reactor Safety, Salt Lake City, March 26-28, 1973.

148. M.C. Pugh, "Probability Approach to Safety Analysis," UKAEA, TRG report 1949(R), 1969.

149. Z.J. Doron and H. Albers, "Mean Annual Severity: An Extension of the Quantitative-Probabilistic Approach to Reactor Safety Analysis," Nucl. Eng. and Des., Vol. 9, 349-356, 1969.

150. P.W. Marriott, et al., "The Loss-of-Coolant Accident and the Environment - A Probabilistic View," ASME 72-WA/NE-9, Nov. 1972.

151. H.P. Balfanz, "Safety Analysis Plan — Use of Various Safety and Reliability Analyses at the Right Time and on Special Problems," German Scientific Reports, Institute for Reactor Safety, Report IRS-W-2, April, 1972 (Translation AEC-TR 7359).

152. W.E. Jordan, "Failure Modes, Effects and Criticality Analyses," Annual Reliability and Maintainability Symposium, 30, 1972.

153. D.W. Williams, "Common Mode Failures in U.S. Commercial Power Plants," M.S. Thesis, The University of Tennessee, Knoxville, Tennessee, June, 1972.

154. I.M. Jacobs, "The Common Mode Failure Study Discipline," IEEE Trans. on Nucl. Sci., Vol. NS-17, No. 1, Feb. 1970.

155. W.C. Gangloff and T. Franke, "An Engineering Approach to Common — Mode Failure Analysis," CSNI Specialist Meeting on the Development and Application of Reliability Techniques to Nuclear Plants, Liverpool, April 8-10, 1974.

156. W.C. Gangloff, "Probability Investigations into Anticipated Transients Without Reactor Trip," ASME 73-WA/NE-5, Nov., 1973.

157. E.P. Epler, "Common Mode Failure Considerations in the Design of Systems for Protection and Control," Nuclear Safety, Vol. 10, No. 1, Jan-Feb., 1969.

158. D. Meister, Human Factors: Theory and Practice, John Wiley and Sons, New York, 1971.

159. E.S. Brown, "System Safety and Human Factors: Some Necessary Relationships," Annual Reliability and Maintainability Symposium, 197, 1974.

160. B.P. Davis and C.N. Cordoni, "People Subsystem Measurement for Total Reliability," Ann. Rel. and Maint. Symp., 394, 1970.

161 B.J. Garrick, W.C. Gekler, et al, "The Effect of Human Error and Static Component Failure on Engineered Safety System Reliability," Holmes and Narver, Inc., HN-194, Nov. 1967.

162. T.L. Regulinski and W.B. Askren, "Mathematical Models of Human Performance Reliability," Annual Symposium on Reliability, 5, 1969.

163. G.J. Schick and R.W. Wolverton, "Achieving Reliability in Large Scale Software Systems," Ann. Symp. on Rel. and Maint., 302, 1974.

164. J.C. Dickson, J.L. Hesse, A.C. Kientz and M.L. Shooman, "Quantitative Analysis of Software Reliability," Ann. Symp. on Rel. and Maint., 148, 1972.

165. O.L. Williamson, G.G. Dorris, A.J. Ryberg and W.E. Straight, "A Software Reliability Program," Ann. Symp. on Rel. and Maint., 420, 1970.

166. H.A. Hoermann, "Principles of Reliability Assessment for Computerized Safety Systems," CSNI Specialist Meeting on the Development and Application of Reliability Techniques to Nuclear Plant, Liverpool, 8-10 April, 1974.

167. W.C. Hetzel, Principles of Computer Program Testing. Program Test Methods, Prentice-Hall, Englewood Cliffs, N.J., 1973.

168. J.S. Prokop, On Proving the Correctness of Computer Programs. Program Test Methods, Prentice-Hall, Englewood Cliffs, N.J., 1973.

169. R. Billinton, R.J. Ringlee and A.J. Wood, Power-System Reliability Calculations, The MIT Press, Cambridge, Mass., 1973.

170. R. Billinton and S.Y. Lee, "Availability Analysis of a Heat Transport Pump Configuration Using Markov Models," CSNI Specialist Meeting on the Development and Application of Reliability Techniques to Nuclear Plant, Liverpool, 8-10 April, 1974.

171. H. Zeibig and E. Valentin, "Application of Markov Processes to the Reliability Analysis of a Control Rod Drive System," Kerntechnik, Vol. 13, No. 9, 1971.

172. A.D. Patton, "A Probability Method for Bulk Power System Security Assessment. II A Development of Probability Models for Normally-Operating Components," IEEE Trans. on Power Apparatus and Systems, Vol. PAS-91, No. 6, Nov.-Dec. 1972.

173. K. Hewitt, "Probabilistic Approaches to Discrete Natural Events: A Review and Theoretical Discussion," Invitational Conference Commission on Quantitative Methods, Ann Arbor, Mich., 1969.

174. G.P. Patil, Classical and Contagious Discrete Distributions, Statist. Publishing Soc., Calcutta, 1965.

175. I.A. Singer and C.M. Nagle, "Variability of Wind Direction Within the United States," Nuclear Safety, Vol. 11, No. 1, Jan.-Feb. 1970.

176. P. Ayyaswamy, B. Hauss, T. Hsieh, A. Moscati, T.E. Hicks and D. Okrent, "Estimates of the Risks Associated with Dam Failures," UCLA-ENG-7423, March 1974.

177. K.A. Solomon, T.E. Hicks and D. Okrent, "Airplane Crash Risk to Ground Population," UCLA-ENG-7424, March 1974.

178. J.A. Simmons, R.C. Erdmann and B.N. Noft, "The Risk of Catastrophic Spills of Toxic Chemicals," UCLA-ENG-7425, March 1974.

179. K.A. Solomon, R.C. Erdmann, T.E. Hicks and D. Okrent, "Estimate of the Hazards to a Nuclear Reactor from the Random Impact of Meteorites," UCLA-ENG-7426, March 1974.

180. V.E. Blake, "A Prediction of the Hazards from the Random Impact of Meteorites in the Earth's Surface," Sandia Labs., Aerospace Nuclear Safety, SC-RR-68-838, Dec. 1968.

181. S.H. Bush, "Probability of Damage to Nuclear Components Due to Turbine Failure," CONF-730304, Topical Meeting on Water-Reactor Safety, Salt Lake City, March 26-28, 1973.

182. E.O. Codier, "Reliability Growth in Real Life," Annual Symposium on Reliability, 458, 1968.

183. R.C. Gwaltney, "Missile Generation and Protection in Light-Water-Cooled Reactors," Nuclear Safety, Vol. 10, No. 4, July-Aug. 1969.

184. C.V. Chelapati, R.P. Kennedy and I.B. Wall, "Probabilistic Assessment of Aircraft Hazard for Nuclear Power Plants," Nucl. Eng. and Design, Vol. 19, 333-364, 1972.

185. J.R. Beattie, G.D. Bell and J.E. Edwards, "Methods for the Evaluation of Risk," UKAEA Report, AHSB(S)R159, 1969.

186. G.D. Bell and J. Houghton, "Risk Evaluation for Stack Releases," UKAEA Report, AHSB(S)R173, 1969.

187. G.D. Bell, "Risk Evaluation for any Curie Release Spectrum and any Dose-Risk Relationship," UKAEA Report, AHSB(S)R 192, 1971.