A. BASICS OF PROBABILITY

A.1 Events

Any repeatable process for which the result is uncertain can be considered an experiment, such as counting failures over time or measuring time to failure of a specific item of interest. The result of one execution of the experiment is referred to as an **outcome**. Repetitions or trials of a defined experiment would not be expected to produce the same outcomes due to uncertainty associated with the process. The set of all possible outcomes of an experiment is defined as the **sample space**.

Sample spaces can contain discrete points (such as pass, fail) or points in a continuum (such as measurement of time to failure). An **event** E is a specified set of possible outcomes in a sample space S (denoted $E \subset S$, where $\subset$ denotes subset).

Most events of interest in practical situations are compound events, formed by some composition of two or more events. Composition of events can occur through the union, intersection, or complement of events, or through some combination of these.

For two events, E_1 and E_2, in a sample space S, the **union** of E_1 and E_2 is defined to be the event containing all sample points in E_1 or E_2 or both, and is denoted by the symbol $(E_1 \cup E_2)$. Thus, a union is simply the event that either E_1 or E_2 or both E_1 and E_2 occur.

For two events, E_1 and E_2, in a sample space S, the **intersection** of E_1 and E_2 is defined to be the event containing all sample points that are in both E_1 and E_2, denoted by the symbol $(E_1 \cap E_2)$. The intersection is the event that both E_1 and E_2 occur.

Figure A.1 shows a symbolic picture, called a **Venn diagram**, of some outcomes and events. In this example, the event E_1 contains three outcomes, event E_2 contains five outcomes, the union contains seven outcomes, and the intersection contains one outcome.

The **complement** of an event E is the collection of all sample points in S and not in E. The complement of E is denoted by the symbol $\overline{E}$ and is the event that all the outcomes in S that are not in E occur.

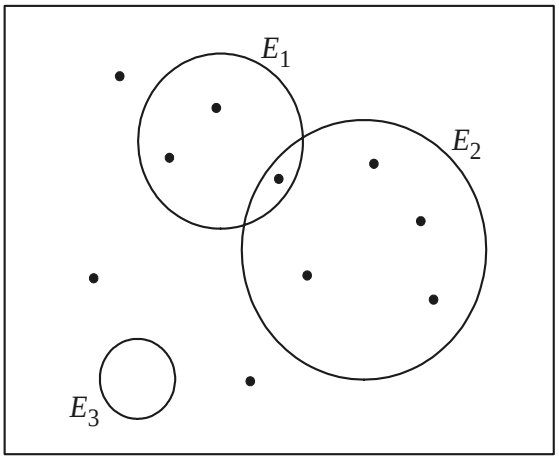

Figure A.1 Venn diagram, showing ten outcomes and three events.

It is sometimes useful to speak of the **empty** or **null** set, a set containing no outcomes. In Figure A.1, the event E_3 is empty. It cannot occur.

Two events, E_1 and E_2, are said to be **mutually exclusive** if the event $(E_1 \cap E_2)$ contains no outcomes in the sample space S. That is, the intersection of the two events is the null set. Mutually exclusive events are also referred to as **disjoint** events. Three or more events are called mutually exclusive, or disjoint, if each pair of events is mutually exclusive. In other words, no two events can happen together.

A.2 Basic Probability Concepts

Each of the outcomes in a sample space has a probability associated with it. **Probabilities of outcomes** are seldom known; they are usually estimated from relative frequencies with which the outcomes occur when the experiment is repeated many times. Once determined, the probabilities must satisfy two requirements:

1. The probability of each outcome must be a number ≥ 0 and ≤ 1.

2. The probabilities of all outcomes in a given sample space must sum to 1.

Associated with any event E of a sample space S is the **probability of the event**, Pr(E). Since an event represents a particular set of outcomes of an experiment, the values of Pr(E) are estimated from the outcomes of the experiment.

Probabilities are associated with each outcome in the sample space through a **probability model**. Probability models are often developed on the basis of information derived from outcomes obtained from an experiment. Probability models are also formulated in the context of mathematical functions.

The values of Pr(E) estimated from the experimental outcomes are often defined as being representative of the *long-run relative frequency* for event E. That is, the relative frequency of an outcome will tend toward some number between 0 and 1(inclusive) as the number of repetitions of the experiment increases. Thus, the probability of the outcome is the number about which the long-term relative frequency tends to stabilize.

This interpretation forms the basis of the **relative frequency definition of probability**, also referred to as the **frequentist view** of probability. In the frequentist view, a mathematical theory of probability is developed by deriving theorems based on the axioms of probability given in the next subsection. The probability of an event is considered to be a fixed quantity, either known or unknown, that is a property of the physical object involved and that can be estimated from data. A theorem derived from the three axioms describes the frequentist view:

If an experiment is repeated a large number of times, n, the observed relative frequency of occurrence, n_E/n, of the event E (where n_E = the number of repetitions when event E occurred) will tend to stabilize at a constant, Pr(E), referred to as the probability of E.

Another interpretation of probability leads to the so-called **classical definition of probability**, which can be stated as follows:

If an experiment can result in n equally likely and mutually exclusive outcomes and if n_E of these outcomes contain attribute E, then the probability of E is the ratio n_E / n.

For example, if each of the outcomes in Figure A.1 had equal probability, 0.1, then Pr(E_1) = 0.3, Pr(E_2) = 0.5, Pr($E_1 \cap E_2$) = 0.1, Pr($E_1 \cup E_2$) = 0.7, and Pr(E_3) = 0.

The classical definition also uses a set of axioms to precisely define probability and is more rigorous and logically consistent than the relative frequency definition. However, this axiomatic definition is less intuitive than the relative frequency definition. Since the true probabilities associated with the sample space are never known, the relative frequency definition is more useful than the classical definition. Both definitions, though, provide a mathematical framework for probability, an overview of which is addressed in Section A.3. Some texts, including parts of this handbook, use the terms *classical* and *frequentist* interchangeably.

Another interpretation of probability is as a **subjective probability**. Probabilities obtained from the opinions of people are examples of subjective probabilities. In this concept, probability can be thought of as a rational measure of belief. Any past information about the problem being considered can be used to help associate the various probabilities. In particular, information about the relative frequency of occurrence of an event could influence the assignment of probabilities.

The notion of subjective probability is the basis for Bayesian inference. In contrast to the relative frequency definition of probability that is based on properties of events, subjective probability can be extended to situations that cannot be repeated under identical conditions. However, the assignment of subjective probabilities can be done according to certain principles so that the frequency definition requirements of probability are satisfied. All the mathematical axioms and theorems developed for frequentist probability apply to subjective probability, but their interpretation is different.

Martz and Waller, 1991, present subjective probability as dealing not only with events but with propositions. A proposition is considered to be a collection of events that cannot be conceived as a series of repetitions, for example, a nuclear power plant meltdown. The degree of belief in proposition A, Pr(A), represents how strongly A is believed to be true. Thus, subjective probability refers to the degree of belief in a proposition. At the extremes, if A is believed to be true, Pr(A) = 1; if A is believed to be false, Pr(A) = 0. Points between 0 and 1 represent intermediate beliefs between false and true.

A.3 Basic Rules and Principles of Probability

The relative frequency, classical, and subjective probability definitions of probability satisfy the following axiomatic requirements of probability:

If $\Pr(E)$ is defined for a type of subset of the sample space S, and if

1. $\Pr(E) \geq 0$, for every event E,
2. $\Pr(E_1 \cup E_2 \cup \cdots) = \Pr(E_1) + \Pr(E_2) + \cdots$, where the events $E_1, E_2, \ldots$, are such that no two have a point in common, and
3. $\Pr(S) = 1$,

then $\Pr(E)$ is called a **probability function**.

A probability function specifies how the probability is distributed over various subsets E of a sample space S. From this definition, several rules of probability follow that provide additional properties of a probability function.

The probability of an impossible event (the empty or null set) is zero, written as:

$$\Pr(\varnothing) = 0,$$

where $\varnothing$ is the null set. The probability of the complement of E is given by:

$$\Pr(\overline{E}) = 1 - \Pr(E).$$

In general, the probability of the union of any two events is given by:

$$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2).$$

If E_1 and E_2 are mutually exclusive, then $\Pr(E_1 \cap E_2) = \Pr(\varnothing) = 0$, and

$$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2),$$

which is a special case of the second axiom of probability stated above and is sometimes referred to as the addition rule for probabilities.

For three events,

$$\begin{aligned}\Pr(E_1 \cup E_2 \cup E_3) = {} & \Pr(E_1) + \Pr(E_2) + \Pr(E_3) \\ & - \Pr(E_1 \cap E_2) - \Pr(E_1 \cap E_3) \\ & - \Pr(E_2 \cap E_3) + \Pr(E_1 \cap E_2 \cap E_3).\end{aligned}$$

This rule is also referred to as the **inclusion-exclusion principle** and can be generalized to n events. It is widely used in PRA to calculate the probability of an "or" gate (a union of events) in a fault tree (NRC 1994).

The inclusion-exclusion principle also provides useful upper and lower bounds on the probability of the union of n events that are not mutually exclusive. One such upper bound, referred to as the **rare event approximation**, is:

$$\Pr(E_1 \cup E_2 \cup \ldots \cup E_n) \leq \Pr(E_1) + \Pr(E_2) + \ldots + \Pr(E_n).$$

The rare event approximation should only be used when the probabilities of the n events are all very small (NRC 1994). If the n events are mutually exclusive, the error is zero. An approximation of the percent error is $n^2 \max [\Pr(E_i)]$, which is valid regardless of the independence of events (NRC 1994). The error in the approximation arises from the remaining terms in the full expansion of the left-hand side of the inequality. This approximation is frequently used in accident sequence quantification.

Many experimental situations arise in which outcomes are classified by two or more events occurring simultaneously. The simultaneous occurrence of two or more events (the intersection of events) is called a joint event, and its probability is called a **joint probability**. Thus, the joint probability of both events E_1 and E_2 occurring simultaneously is denoted by $\Pr(E_1 \cap E_2)$.

The probability associated with one event, irrespective of the outcomes for the other events, can be obtained by summing all the joint probabilities associated with all the outcomes for the other events, and is referred to as a **marginal probability**. A marginal probability is therefore the *unconditional* probability of an event, unconditioned on the occurrence of any other event.

Two events E_1 and E_2 are often related in such a way that the probability of occurrence of one depends on whether the other has or has not occurred. The **conditional probability** of one event, given that the other

has occurred, is equal to the joint probability of the two events divided by the marginal probability of the given event. Thus, the conditional probability of event E_2, given event E_1 has occurred, denoted $\Pr(E_2|E_1)$, is defined as:

$$\Pr(E_2|E_1) = \Pr(E_1 \cap E_2) / \Pr(E_1), \quad \text{(A.1)}$$

for $\Pr(E_1) > 0$. If $\Pr(E_1) = 0$, $\Pr(E_2|E_1)$ is undefined.

Rearranging this equation yields:

$$\begin{aligned}\Pr(E_1 \cap E_2) &= \Pr(E_1)\Pr(E_2|E_1) \\ &= \Pr(E_2)\Pr(E_1|E_2).\end{aligned} \quad \text{(A.2)}$$

Calculation of joint probability requires the concept of **statistical independence**. Two events E_1 and E_2 are statistically independent if the probability of one event does not change whenever the other event occurs or does not occur. Thus, E_2 is **independent** of E_1 if

$$\Pr(E_2|E_1) = \Pr(E_2).$$

If E_2 is independent of E_1, then E_1 is independent of E_2. It follows that events E_1 and E_2 are independent if their joint probability is equal to the product of the *unconditional*, or *marginal*, probabilities of the events:

$$\Pr(E_1 \cap E_2) = \Pr(E_1)\Pr(E_2),$$

which is sometimes referred to as the multiplication rule for probabilities. If $\Pr(E_1)$ varies depending on whether or not event E_2 has occurred, then events E_1 and E_2 are said to be **statistically dependent**.

If E_1, E_2, ... are mutually exclusive, and if the union of E_1, E_2, ... equals the entire sample space, then the events E_1, E_2, ... are said to form a **partition** of the sample space. Exactly one of the events must occur, not more than one but exactly one. In this case, the **law of total probability** says

$$\Pr(A) = \sum \Pr(A|E_i)\Pr(E_i) .$$

A special case can be written when there are only two sets. In this case, write E_1 simply as E and E_2 as $\overline{E}$.

Then the law of total probability simplifies to

$$\Pr(A) = \Pr(A|E)\Pr(E) + \Pr(A|\overline{E})\Pr(\overline{E})$$

for any event A. This formula is the basis for event trees, which are frequently used to diagram the possibilities in an accident sequence.

The concepts of mutually exclusive events and statistically independent events are often confused. If E_1 and E_2 are mutually exclusive events and $\Pr(E_1)$ and $\Pr(E_2)$ are nonzero, $\Pr(E_1 \cap E_2) = \Pr(\varnothing) = 0$. From Equation A.1, $\Pr(E_2|E_1) = 0$, which does not equal $\Pr(E_2)$. Thus, the two events are not independent. Mutually exclusive events cannot be independent and vice versa.

Equation A.2 can be used to calculate the probability of the intersection of a set of events (the probability that all the events occur simultaneously). For two events E_1 and E_2, the probability of simultaneous occurrence of the events is equal to the probability of E_1 times the probability of E_2 given that E_1 has already occurred. In general, the probability of the simultaneous occurrence of n events can be written as:

$$\Pr(E_1 \cap E_2 \cap \ldots \cap E_n) =$$

$$\Pr(E_1)\Pr(E_2|E_1)\Pr(E_3|E_2 \cap E_1) \ldots \Pr(E_n|E_{n-1} \cap \ldots \cap E_1),$$

which is referred to as the **chain rule**. This rule can be used to calculate the probability that a given accident sequence occurs, with E_1 denoting the initiating event and the remaining events corresponding to the failure or success of the systems that must function in order to mitigate such an accident.

The probability of occurrence of at least one of a set of statistically independent events yields a result that is important to PRA and fault tree applications. If E_1, E_2, ..., E_n are statistically independent events, the probability that at least one of the n events occurs is:

$$\Pr(E_1 \cup E_2 \cup \ldots \cup E_n) = \quad \text{(A.3)}$$

$$1 - [1 - (\Pr(E_1)][1 - (\Pr(E_2)] \ldots [1 - (\Pr(E_n)],$$

which is equivalent (with expansion) to using the inclusion-exclusion rule. For the simple case where $\Pr(E_1) = \Pr(E_2) = \ldots = \Pr(E_n) = p$, the right-hand side of this expression reduces to $1 - (1 - p)^n$.

The general result in Equation A.3 has application in PRA and fault tree analysis. For example, for a system in which system failure occurs if any one of n independent events occurs, the probability of system failure is given by Equation A.3. These events could be

failures of critical system components. In general, the events represent the modes by which system failure (the top event of the fault tree) can occur. These modes are referred to as the minimal cut sets of the fault tree and, if independent of each other (no minimal cut sets have common component failures), Equation A.3 applies. [See Vesely et al. (1981) for further discussion of fault trees and minimal cut sets.]

If the *n* events are not independent, the right-hand side of Equation A.3 may be greater than or less than the left-hand side. However, for an important situation that frequently arises in PRA, the right-hand side of Equation A.3 forms an upper bound for the left-hand side.

If the *n* events are cut sets that are positively associated (see Esary and Proschan 1970, 1963), then the right-hand side is an upper bound for $\Pr(E_1 \cup E_2 \cup ... \cup E_n)$ and is known as the **min cut upper bound** (NRC 1994). This name arises from common PRA applications where E_i is the i^{th} minimal cut set of a system or accident sequence of interest. In this case, the min cut upper bound is superior to the rare event approximation and can never exceed unity (as can happen with the rare event approximation). If the *n* events satisfy conditions similar to those of the rare event approximation, the min cut set upper bound is a useful approximation to the left hand side of Equation A.3. Note that the min cut upper bound is not applicable for mutually exclusive events.

A.4 Random Variables and Probability Distributions

A.4.1 Random Variables

A **random variable** is any rule that associates real numbers with the outcomes of an experiment. If the numbers associated with the outcomes of an experiment are all distinct and countable, the corresponding random variable is called a **discrete** random variable.

If the sample space contains an infinite number of outcomes (like those contained in any interval), the random variable is **continuous**. Time *T* is a common continuous random variable, for example, time to failure or time between failures, where the random variable *T* can assume any value over the range 0 to ∞.

A.4.2 Probability Distributions

A **probability function** (introduced at the beginning of section A.3) associates a probability with each possible value of a random variable and, thus, describes the distribution of probability for the random variable. For a discrete random variable, this function is referred to as a **discrete probability distribution function** (p.d.f.). A discrete p.d.f., commonly denoted by *f*, is also referred to as a **discrete distribution**, or **discrete probability mass function**.

If *x* denotes a value that the discrete random variable *X* can assume, the probability distribution function is often denoted $\Pr(x)$. The notation used here is that a random variable is denoted by an *upper case letter* and an observed value (a number or outcome) of the random variable is denoted by a *lower case letter*. The sum of the probabilities over all the possible values of *x* must be 1. Certain discrete random variables have wide application and have therefore been defined and given specific names. The two most commonly used discrete random variables in PRA applications are the **binomial** and **Poisson** random variables, which are presented in section A.6.

A continuously distributed random variable has a **density function**, a nonnegative integrable function, with the area between the graph of the function and the horizontal axis equal to 1. This density function is also referred to as the **continuous probability distribution function** (p.d.f.). If *x* denotes a value that the continuous random variable *X* can assume, the p.d.f. is often denoted as $f(x)$. The probability that *X* takes a value in a region *A* is the integral of $f(x)$ over *A*. In particular,

$$\Pr(a \le X \le b) = \int_a^b f(x)dx$$

and

$$\Pr(x \le X \le x + \Delta x) \approx f(x)\Delta x \qquad \text{(A.4)}$$

for small Δx.

The most commonly used continuous distributions in PRA are the **lognormal**, **exponential**, **gamma**, and **beta** distributions. Section A.7 summarizes the essential facts about these distributions, and also about less common but occasionally required distributions:

uniform, **normal**, **Weibull**, **chi-squared**, **inverted gamma**, **logistic-normal**, and **Student's *t***.

A.4.3 Cumulative Distribution Functions

Discrete probability distributions provide point probabilities for discrete random variables and continuous p.d.f.s provide point densities for continuous random variables. A related function useful in probability and PRA is the **cumulative distribution function** (c.d.f.). This function is defined as the probability that the random variable assumes values less than or equal to the specific value x, and is denoted $F(x)$.

For a discrete random variable X, with outcomes x_i, and the corresponding probabilities $\Pr(x_i)$, $F(x)$ is the sum of the probabilities of all $x_i \le x$. That is,

$$F(x) = \Pr(X \le x) = \sum_{x_i \le x} \Pr(x_i) .$$

For a continuous random variable X, $F(x)$ is the area beneath the p.d.f. $f(x)$ up to x. That is, $F(x)$ is the integral of $f(x)$:

$$F(x) = \Pr(X \le x) = \int_{-\infty}^{x} f(y)dy .$$

Thus, $f(x)$ is the derivative of $F(x)$. If X takes on only positive values, the limits of integration are 0 to x. Note that, because $F(x)$ is a probability, $0 \le F(x) \le 1$. If X ranges from $-\infty$ to $+\infty$, then

$F(-\infty) = 0$ and $F(+\infty) = 1$.

If X has a restricted range, with a and b being the lower and upper limits of X respectively, $a < X < b$, then

$F(a) = 0$ and $F(b) = 1$.

Also, $F(x)$ is a nondecreasing function of x, that is,

if $x_2 > x_1$, $F(x_2) \ge F(x_1)$.

Another important property of $F(x)$ is that

$\Pr(x_1 < X \le x_2) = F(x_2) - F(x_1)$

for discrete random variables and

$\Pr(x_1 \le X \le x_2) = F(x_2) - F(x_1)$

for continuous random variables.

An example of a p.d.f. and the associated c.d.f. for a continuous distribution is shown in Figure A.2.

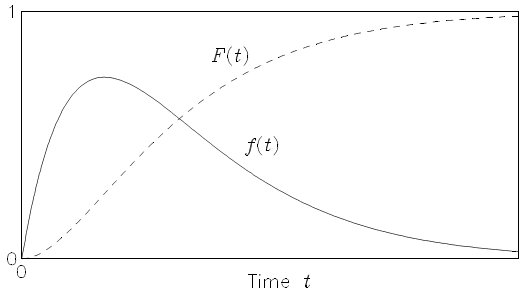

Figure A.2 Probability density function (p.d.f.) and cumulative distribution function (c.d.f.).

A.4.4 Reliability and Hazard Functions

A.4.4.1. Definitions

There are also characterizations that have special interpretations for time-to-failure distributions. Let T denote the random time to failure of a system. The **reliability function** of a system is defined as

$R(t) = \Pr(T > t)$.

Hence, $R(t)$, called the **reliability** at time t, is the probability that the system does not fail in the time interval $(0, t]$ or equivalently, the probability that the system is still operating at time t. (This discussion uses the notation $(a, b]$ to mean the set of times $> a$ and $\le b$, but the distinction between $<$ and $\le$ is a mathematical fine point, not important in practice.) The reliability function is also sometimes called the **survival function**. It is equal to $1 - F(t)$.

When used as a reliability criterion, it is common to state a time, say t_0, called the **mission time**, and require for a system that the reliability at mission time t_0 be at least some prescribed level, say R_0. For example, a pump might be required to operate successfully for at least 12 hours with probability at least 0.95. The requirement in this case is $R_0 = 0.95$ and $t_0 = 12$. In terms of the reliability function, this would mean $R(12) \ge 0.95$. One interpretation would be that such a pump would perform for the required mission time for 95% of the situations when it is called on to do so. Another

interpretation is that 95% of all such pumps would perform as required.

Consider a system that operates for a particular mission time, unless it fails. If it fails, no immediate repairs are attempted, so some authors call the system **nonrepairable**. A common way to characterize this system's reliability is in terms of the **hazard function**. Suppose that the system is still operating at time t, and consider the probability that it will fail in a small interval of time $(t, t + \Delta t]$. This is the conditional probability $\Pr(t < T \le t + \Delta t \mid T > t)$. The hazard function, h, is defined so that when Δt is small,

$$h(t)\Delta t \approx \Pr(t < T \le t + \Delta t \mid T > t) \ . \qquad \text{(A.5)}$$

This function is also encountered, under the name of λ, in some treatments of Poisson processes. Equation A.5 gives, approximately,

$$h(t)\Delta t \approx \frac{\Pr(t < T \le t + \Delta t)}{\Pr(T > t)} \approx \frac{f(t)\Delta t}{R(t)}$$

This is the basis for the formal definition of h:

$$h(t) = \frac{f(t)}{R(t)}$$

For details, see Bain and Engelhardt (1992, p. 541). Equation A.5 is analogous to Equation A.4, except that the probability in Equation A.5 is conditional on the system having survived until t, whereas Equation A.4 refers to all systems in the original population, either still surviving or not. Suppose a large number, say N, of identical systems are put into operation at time $t = 0$, and n is the number which fail in the interval $(t, t + \Delta t]$. It follows that $f(t)\Delta t \approx n/N$, the observed relative frequency of systems failed in the interval $(t, t + \Delta t]$. On the other hand, if N_t denotes the number of the original N systems which are still in operation at time t, then $h(t)\Delta t \approx n/N_t$, the observed relative frequency of *surviving* systems which fail in this same interval. Thus, $f(t)$ is a measure of the risk of failing at time t for any system in the original set, whereas $h(t)$ is a measure of the risk of failing at time t, but only for systems that have survived this long.

The hazard function is used as a measure of "aging" for systems in the population. If $h(t)$ is an increasing function, then systems are aging or wearing out with time. Of course, in general the hazard function can exhibit many types of behavior other than increasing with time, and other possible behaviors are discussed later in this handbook. In actuarial science the hazard function is called the **force of mortality**, and it is used as a measure of aging for individuals in a population. More generally, the hazard function gives an indication of "proneness to failure" of a system after time t has elapsed. Other terms which are also used instead of hazard function are **hazard rate** and **failure rate**. The term *failure rate* is often used in other ways in the literature of reliability (see Ascher and Feingold 1984, p. 19).

A.4.4.2 Relations among p.d.f., Reliability, and Hazard

Any one of the functions F, f, R, and h completely characterizes the distribution, and uniquely determines the other three functions. The definition

$$h(t) = \frac{f(t)}{R(t)}$$

was given above. The right hand side can be written as the derivative of $-\ln[R(t)]$, leading to

$$R(t) = \exp\left(-\int_0^t h(u)du\right) = \exp\left(-H(t)\right)$$

where the function $H(t)$ is called the **cumulative hazard function**. The reliability function, $R(t)$, and the c.d.f., $F(t) = 1 - R(t)$, are therefore uniquely determined by the hazard function, $h(t)$, and the p.d.f. can be expressed as

$$f(t) = h(t)\exp\left(-\int_0^t h(u)du\right).$$

Figure A.3 shows the reliability, hazard and the cumulative hazard function for the example of Figure A.2.

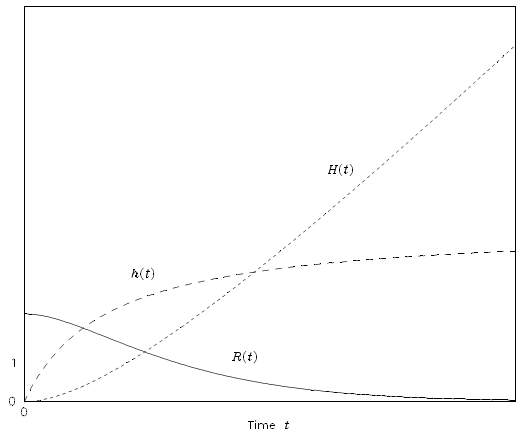

Figure A.3 The reliability function, hazard function and cumulative hazard function.

The hazard function in Figure A.3 is an increasing function of time. Therefore, it would be consistent with systems with a dominant wear-out effect for the entire life of the system. The lifetime of a system may be divided into three typical intervals: the **burn-in** or **infant** period, the **random** or **chance** failure period and the **wear-out** period. During the useful period, the dominant cause of failures is "random" failures. For example, systems might fail due to external causes such as power surges or other environmental factors rather than problems attributable to the defects or wear-out in the systems. This example is somewhat idealized because for many types of systems the hazard function will tend to increase slowly during the later stages of the chance failure period. This is particularly true of mechanical systems. On the other hand, for many electrical components such as transistors and other solid-state devices, the hazard function remains fairly flat once the burn-in failure period is over.

A.4.5 Joint, Marginal, and Conditional Distributions

Many statistical methods are based on selecting a sample of size n from a probability distribution $f(x)$. Such a sample is denoted by

$$(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = (x_1, x_2, ..., x_n),$$

where $x_1, x_2, ..., x_n$ are the actual values of the random variable X which has the distribution $f(x)$.

The concepts of simultaneous events and joint, marginal, and conditional probability, discussed in Section A.3, also pertain to random variables and probability distributions. Two random variables X_1 and X_2 (both continuous, both discrete, or one of each) can have a **joint distribution**, or joint p.d.f., denoted $f(x_1, x_2)$. The point (x_1, x_2) can be thought of as a point in two-dimensional Euclidean space. Similarly, n random variables have joint distribution $f(x_1, x_2, ..., x_n)$. Also, the n random variables have joint cumulative distribution $F(x_1, x_2, ..., x_n)$.

The **marginal distribution** of X_i is defined as the joint p.d.f. integrated (for continuous random variables) or summed (for discrete random variables) over the $n-1$ other corresponding dimensions, resulting in a function of x_i alone. Thus, the marginal distribution of X_i is the unconditional p.d.f. of X_i, $f_i(x_i)$.

The **conditional distribution** of X_1 given X_2, denoted $g(x_1 \mid x_2)$, is defined by

$$g(x_1 | x_2) = \frac{f(x_1, x_2)}{f_2(x_2)},$$

where $f_2(x_2) \neq 0$, and can be shown to satisfy the requirements of a probability function. Sampling from a conditional p.d.f. would produce only those values of X_1 that could occur for a given value of $X_2 = x_2$. The concept of a conditional distribution also extends to n random variables.

Two random variables X_1 and X_2 are independent if their joint p.d.f. is equal to the product of the two individual p.d.f.s. That is,

$$f(x_1, x_2) = f(x_1)\, f(x_2).$$

In general, $X_1, X_2, ..., X_n$ are **independent random variables** if

$$f(x_1, x_2, ..., x_n) = f(x_1) \cdot f(x_2) \cdot \ldots \cdot f(x_n).$$

A.4.6 Characterizing Random Variables and their Distributions

A.4.6.1 Distribution Characteristics

Probability distributions have many characteristics of interest, some of which are described by **distribution parameters**. The term parameter is used to refer to a

fixed characteristic. In contrast to a statistic, which changes from sample to sample, a parameter for a particular distribution is a constant and does not change. However, when a parameter's value is not known, sample statistics can be used to estimate the parameter value. Parameter estimation is discussed in Appendix B.

A very useful distribution characteristic is the parameter that serves as a measure of central tendency, which can be viewed as a measure of the middle of a distribution. When a change in the parameter slides the distribution sideways, as with the mean of a normal distribution, the parameter is referred to as the **location parameter**. It serves to locate the distribution along the horizontal axis. Sometimes, however, a change in the parameter squeezes or stretches the distribution toward or away from zero, as with the mean of the exponential distribution. In that case, the parameter is a **scale parameter**.

In any case, the most common measure of central tendency is the **mean**, μ, of the distribution, which is a weighted average of the outcomes, with the weights being probabilities of outcomes. For a discrete random variable X,

$$\mu_X = \sum_i x_i \Pr(x_i) .$$

For a continuous random variable X,

$$\mu_X = \int_{-\infty}^{\infty} x\, f(x) dx .$$

Another distribution characteristic commonly used as a measure of central tendency, or location, is the **median**, which is the point along the horizontal axis for which 50% of the area under the p.d.f. lies to its left and the other 50% to its right. The median of a random variable, X, is commonly designated med(X) or $x_{.50}$ and, for a continuous distribution, is the value for which $\Pr(X \le x_{.50}) = .50$ and $\Pr(X \ge x_{.50}) = .50$. In terms of the cumulative distribution, $F(x_{.50}) = .50$. The median is a specific case of the general 100αth **percentile**, x_α, for which $F(x_\alpha) = \alpha$. When the factor of 100 is dropped, x_α is called the α **quantile**. Along with the median as the 50th percentile (or equivalently, the 0.5 quantile), the 25th and 75th percentiles are referred to as **quartiles** of a distribution.

Figure A.4 shows the quartiles, $x_{0.25}$ and $x_{0.75}$, the median, $x_{0.50}$, and the mean. The quartiles and the median divide the area under the density curve into four pieces, each with the same area. Note that the mean is greater than the median in this example, which is the usual relation when the density has a long right tail, as this one does.

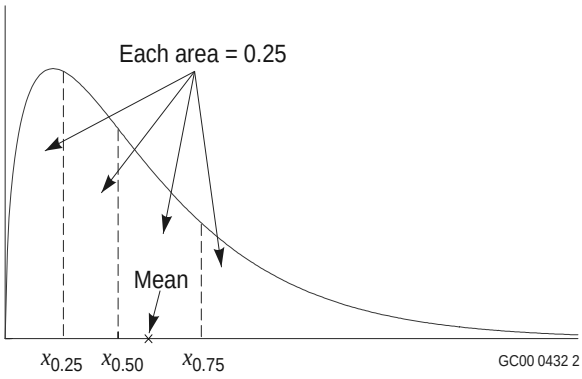

Figure A.4 Density, showing quartiles, median, and mean.

Figure A.5 shows the same quantities plotted with the c.d.f. By definition, the q quantile, x_q, satisfies $F(x_q) = q$.

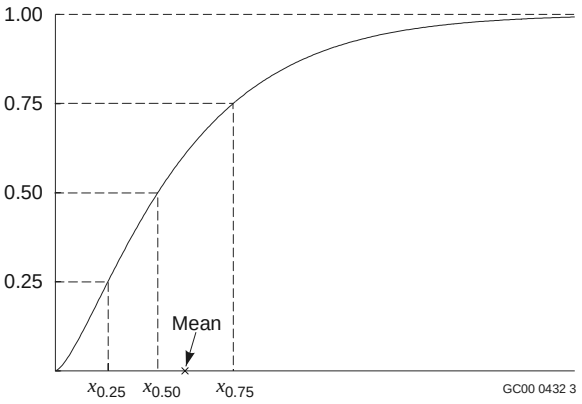

Figure A.5 Cumulative distribution function (c.d.f.) showing quartiles, median, and mean.

The mean and the median are used to measure the center or location of a distribution. Since the median is

less affected by tail-area probabilities, it can be viewed as a better measure of location than the mean for highly-skewed distributions. For symmetric distributions, the mean and median are equivalent.

A different measure of center or location of a distribution is the **mode**, which indicates the most probable outcome of a distribution. The mode is the point along the horizontal axis where the "peak" or maximum of the p.d.f. is located. Note that the mode does not necessarily have to be near the middle of the distribution. It simply indicates the most likely value of a distribution. Note also that a peak does not have to exist and, in some cases, more than one peak can exist.

Another important characteristic of a distribution is its **variance**, denoted by σ^2. The variance is the average of the squared deviations from the mean. The **standard deviation**, σ, of a distribution is the square root of its variance. Both the variance and standard deviation are measures of a distribution's spread or dispersion. For a discrete random variable X,

$$\sigma_X^2 = \sum_i (x_i - \mu)^2 \Pr(x_i) .$$

For a continuous random variable X,

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx .$$

Though less used than the mean and variance, the **skewness** is defined as

$E(X - \mu)^3/\sigma^3$.

It measures asymmetry. It is usually positive if the density has a longer right tail than left tail, and negative if the density has a longer left tail than right tail. For example, the density in Figure A.4 has positive skewness.

A.4.6.2 Mathematical Expectation

The definitions of distribution means and variances arise from **mathematical expectation** and **moments of a distribution**, which form an important method for calculating the parameters of a known p.d.f. In general, the **expectation** (**expected value** or mathematical expectation) of a function $g(X)$, denoted $E[g(X)]$, is

$$E[g(X)] = \sum_i g(x_i) \Pr(x_i) ,$$

when X is discrete, and

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)dx ,$$

when X is continuous.

Because of their wide use, several expectations have special names. For $g(X) = X$, the expectation $E(X)$ becomes the mean of X. Thus, the mean is also commonly referred to as the expected value (or expectation) of the random variable X. In addition, for $g(X) = X$, the expectation $E(X)$ is known as the **first moment about the origin**.

The variance, σ_X^2, also denoted by $\mathrm{Var}(X)$, of a distribution is defined by mathematical expectation with $g(X) = (X - \mu_X)^2$. Thus,

$\mathrm{Var}(X) = \sigma_X^2 = E[(X - \mu_X)^2] = E(X^2) - [E(X)]^2$,

which is known as the **second moment about the mean**.

Ordinary moments (moments about the origin) of a random variable X are defined as

$M_r = E(X^r)$,

for $r = 1, 2, \ldots$. Thus,

$\mathrm{Var}(X) = \sigma_X^2 = E(X^2) - [E(X)]^2 = M_2 - M_1^2$.

Central moments (moments about the mean) of a random variable X are defined as being equal to $E[(X - \mu)^r]$ for $r = 2, 3, \ldots$. The ordinary and central moments can be seen to define characteristics of distributions of random variables.

An important rule of expectation commonly used in PRA is that the expected value of a product of independent random variables is the product of their respective expected values. That is, $E(X_1 \cdot X_2 \cdot \ldots \cdot X_n) = E(X_1) \cdot E(X_2) \cdot \ldots \cdot E(X_n)$ when all X_i are independent. This rule also applies to conditionally independent random variables. If the random variables $X_2, X_3, \ldots, X_n$ are all conditionally independent given $X_1 = x_1$, then

$f(x_2, x_3, ..., x_n|x_1) = f(x_2|x_1)\cdot f(x_3|x_1)\cdot ... \cdot f(x_n|x_1)$.

It follows that

$E(X_2\cdot X_3\cdot ... \cdot X_n|x_1) = E(X_2|x_1)\cdot E(X_3|x_1)\cdot ... \cdot E(X_n|x_1)$.

Thus,

$E(X_1\cdot X_2\cdot ... \cdot X_n) = E[X_1\cdot E(X_2|x_1)\cdot E(X_3|x_1)\cdot ... \cdot E(X_n|x_1)]$.

A.4.6.3 Moment-Generating Functions

Another special mathematical expectation is the **moment-generating function** of a random variable. For a random variable X with p.d.f. $f(x)$, the moment-generating function of X (or of the distribution) is defined by $M(t) = E(e^{tX})$, if M exists for some interval $-h < t < h$. Therefore, if X is a continuous random variable,

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx .$$

If X is a discrete random variable,

$$M(t) = \sum_i e^{tx_i} f(x_i) .$$

Note that not every distribution has a moment-generating function.

The importance of the moment-generating function is that, when it does exist, it is unique and completely specifies the distribution of the random variable. If two random variables have the same moment-generating function, they have the same distribution.

It can be shown that the moments of a distribution can be found from the series expansion of $M(t)$. The moments of the distribution can also be determined from the moment-generating function by differentiating the moment-generating function with respect to t and setting $t = 0$. See Martz and Waller (1991) and any of several mathematical statistics texts, such as Hogg and Craig (1995), for further details on moment-generating functions.

A.4.6.4 Covariance and Correlation

For two random variables, X and Y, with means μ_x and μ_y, the expected value $E[(X - \mu_x)(Y - \mu_y)]$ is called the **covariance** of X and Y, denoted Cov(X, Y). The covariance of X and Y divided by the product of the standard deviations of X and Y is called the **correlation coefficient** (or correlation) between X and Y, denoted Cor(X, Y). That is,

$$\mathrm{Cor}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

$$= \frac{E(X-\mu_X)E(Y-\mu_Y)}{\sqrt{E[(X-\mu_X)^2]E[(Y-\mu_Y)^2]}}$$

.

The correlation coefficient measures the degree of association between X and Y, that is, the strength of a linear relationship between X and Y.

A.4.7 Distribution of a Transformed Random Variable

This section considers the distribution of $Y = h(X)$, when X has a known distribution and h is a known function. The problem is straightforward when X has a discrete distribution. When X is continuous and h is monotone, either increasing or decreasing, the c.d.f.s are also related in the natural way, as follows. Let F be the c.d.f. of X and let G be the c.d.f. of Y. Then we have
$G(y) = \Pr(Y \le y) = \Pr[h(X) \le y]$.
If h is monotone increasing, this equals
$\Pr[X \le h^{-1}(y)] = F(x)$,
where x and y are related by $y = h(x)$, $x = h^{-1}(y)$. In summary, $G(y) = F(x)$.

If, instead, h is monotone decreasing, then a similar argument gives
$G(y) = 1 - F(x)$.

The surprise comes with the densities. Differentiate both sides of either of the above equations with respect to y, to obtain the density of y. This involves using the chain rule for differentiation. The result is

$$g(y) = f(x)\left|\frac{dx}{dy}\right| .$$

That is, the density of Y is not simply equal to the density of X with a different argument. There is also a multiplier, the absolute value of the derivative. If $Y = \exp(X)$, then

$$g(y) = f[\ln(y)](1/y) .$$

If $Y = 1/X$, then

$$g(y) = f(1/y)(1/y^2) .$$

The formulas here are the basis for the densities of the lognormal distribution and the inverted gamma distribution.

A.5 Bayes' Theorem

It is frequently desired to calculate the probability of an event A given that another event B has occurred at some prior point in time. It can also be of interest to calculate the probability that a state of nature exists given that a certain sample is observed, for example, belonging to a certain population based on a sample measurement or observation. Conditional probability leads directly to **Bayes' Theorem**, which, along with subjective probability, forms the basis for Bayesian inference commonly used in PRA.

Bayes' Theorem states that: if A_1, A_2, ..., A_n are a sequence of disjoint events and if B is any other event such that $\Pr(B) > 0$, then

$$\Pr(A_i|B) = \frac{\Pr(B|A_i)\,\Pr(A_i)}{\Pr(B)} , \qquad \text{(A.6)}$$

where

$$\Pr(B) = \sum_{j=1}^{n} \Pr(B|A_j)\,\Pr(A_j) .$$

Equation A.6 follows from the definition of conditional probability in Equation A.2:

$$\Pr(A_i|B) = \frac{\Pr(B \cap A_i)}{\Pr(B)} = \frac{\Pr(B|A_i)\,\Pr(A_i)}{\Pr(B)} .$$

The $\Pr(A_i|B)$ is the **posterior** (or a posteriori) probability for the event A_i, meaning the probability of A_i once B is known. The $\Pr(A_i)$ is the **prior** (or a priori) probability of the event A_i before experimentation or observation. The event B is the observation. The $\Pr(B|A_i)$ is the probability of the observation given that A_i is true. The denominator serves as a normalizing constant.

Calculating the posterior probabilities $\Pr(A_i|B)$ requires knowledge of the probabilities $\Pr(A_i)$ and $\Pr(B|A_i)$, $i = 1, 2, ..., n$. The probability of an event can often be determined if the population is known, thus, the $\Pr(B|A_i)$ can be determined. However, the $\Pr(A_i)$, $i = 1, 2, ..., n$, are the probabilities that certain states of nature exist and are either unknown or difficult to ascertain. These probabilities, $\Pr(A_i)$, are called prior probabilities for the events A_i because they specify the distribution of the states of nature prior to conducting the experiment.

Application of Bayes' Theorem utilizes the fact that $\Pr(B|A_i)$ is easier to calculate than $\Pr(A_i|B)$. If probability is viewed as degree of belief, then the prior belief is changed, by the test evidence, to a posterior degree of belief. In many situations, some knowledge of the prior probabilities for the events A_1, A_2, ..., A_n exists. Using this prior information, inferring which of the set A_1, A_2, ..., A_n, is the true population can be achieved by calculating the $\Pr(A_i|B)$ and selecting the population that produces the highest probability.

Equation A.6 pertains to disjoint discrete events and discrete probability distributions. Bayes' Theorem has analogous results for continuous p.d.f.'s. Suppose X is a continuous random variable, with p.d.f. depending on parameter θ, and with conditional p.d.f. of X, given θ, specified by $f(x|\theta)$. Consider θ to be a possible value of the random variable Θ (using the convention of denoting random variables with uppercase letters). If the prior p.d.f. of Θ is denoted $g(\theta)$, then for every x such that $f(x) > 0$ exists, the posterior p.d.f. of Θ, given $X = x$, is

$$g(\theta|x) = \frac{f(x|\theta)g(\theta)}{f(x)} ,$$

where

$$f(x) = \int f(x|\theta)g(\theta)d\theta$$

is the marginal p.d.f. of X. Again, the prior and posterior p.d.f.'s can be used to represent the knowledge and beliefs about the likelihood of various values of a random variable Θ prior to and posterior to observing a value of another random variable X.

A.6 Discrete Random Variables

A.6.1 The Binomial Distribution

The **binomial** distribution describes the number of failures X in n independent trials. The random variable X has a binomial distribution if:

1. The number of random trials is one or more and is known in advance.
2. Each trial results in one of two outcomes, usually called success and failure (or could be pass-fail, hit-miss, defective-nondefective, etc.).
3. The outcomes for each trial are statistically independent.
4. The probability of failure, p, is constant across trials.

Equal to the number of failures in the n trials, a binomial random variable X can take on any integer value from 0 to n. The probability associated with each of these possible outcomes, x, is defined by the binomial(n, p) p.d.f. as

$$\Pr(X = x) = \binom{n}{x} p^x (1 - p)^{n-x},$$

$$x = 0, \ldots, n.$$

Here

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

is the **binomial coefficient** and

$n! = n(n-1)(n-2) \ldots (2)(1)$

denotes n factorial, with 0! defined to be equal to 1. This binomial coefficient provides the number of ways that exactly x failures can occur in n trials (number of combinations of n trials selected x at a time).

The binomial distribution has two parameters, n and p, of which n is known. (Although n may not always be known exactly, it is treated as known in this handbook.)

The mean and variance of a binomial(n, p) random variable X are

$E(X) = np$

and

$\mathrm{Var}(X) = np(1 - p)$.

Figure A.6 shows three binomial probability distribution functions, with parameter $p = 0.25$, and $n = 4$, 12, and 40. In each case, the mean is np. The means have been aligned in the three plots.

A.6.2 The Poisson Distribution

The **Poisson** distribution provides a discrete probability model that is appropriate for many random phenomena that involve counts. Examples are counts per fixed time interval of the number of items that fail, the number of customers arriving for service, and the number of telephone calls occurring. A common use of the Poisson distribution is to describe the behavior of many rare event occurrences. The Poisson distribution is also frequently used in applications to describe the occurrence of system or component failures under steady-state conditions.

The count phenomena that occur as Poisson random variables are not necessarily restricted to occurring over a time interval. They could also be counts of things occurring in some region, such as defects on a surface or within a certain material. A process that leads to a Poisson random variable is said to be a **Poisson process**.

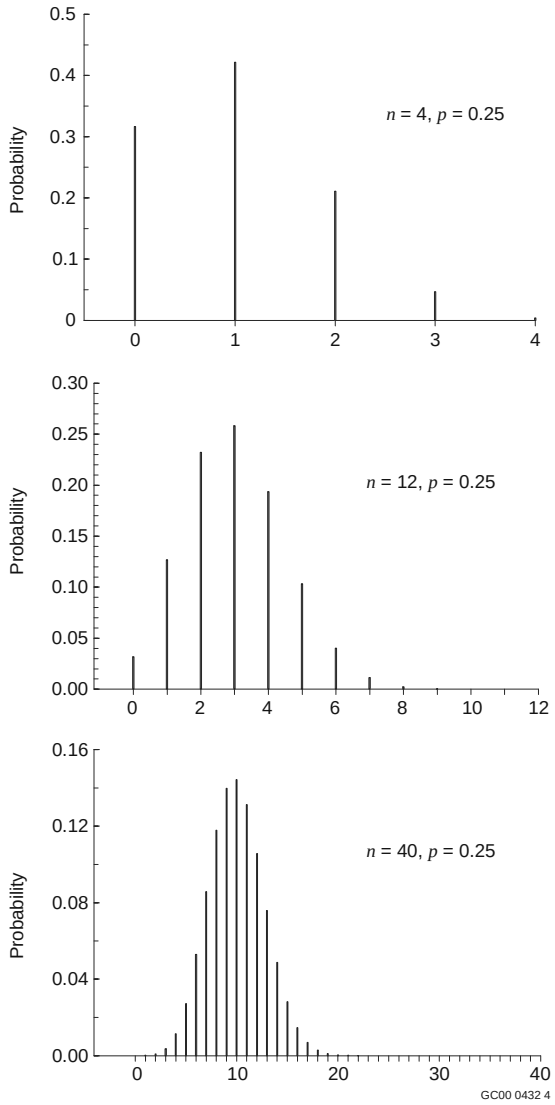

Figure A.6 Three binomial probability distribution functions.

The Poisson distribution describes the total number of events occurring in some interval of time t (or space). The p.d.f. of a Poisson random variable X, with parameter $\mu = \lambda t$, is

$$\Pr(X = x) = \frac{e^{-\mu}\mu^{x}}{x!} = \frac{e^{-\lambda t}(\lambda t)^{x}}{x!}, \qquad \text{(A.7)}$$

for $x = 0, 1, 2, \ldots$, and $x! = x(x-1)(x-2) \ldots (2)(1)$, as defined previously.

The Poisson distribution has a single parameter μ, denoted Poisson(μ). If X denotes the number of events that occur during some time period of length t, then X is often assumed to have a Poisson distribution with parameter $\mu = \lambda t$. In this case, X is considered to be a Poisson process with **intensity** $\lambda > 0$ (Martz and Waller 1991). The variable λ is also referred to as the **event rate** (or **failure rate** when the events are failures). Note that λ has units 1/time; thus, $\lambda t = \mu$ is dimensionless.

If only the total number of occurrences for a single time period t is of interest, the form of the p.d.f. in Equation A.7 using μ is simpler. If the event rate, λ, or various time periods, t, are of interest, the form of the p.d.f. in Equation A.7 using λt is more useful.

The expected number of events occurring in the interval 0 to t is $\mu = \lambda t$. Thus, the mean of the Poisson distribution is equal to the parameter of the distribution, which is why μ is often used to represent the parameter. The variance of the Poisson distribution is also equal to the parameter of the distribution. Therefore, for a Poisson(μ) random variable X,

$E(X) = \mathrm{Var}(X) = \mu = \lambda t$.

Figure A.7 shows three Poisson probability distribution functions, with means μ = 1.0, 3.0, and 10.0, respectively. The three means have been aligned in the graphs. Note the similarity between the Poisson distribution and the binomial distribution when $\mu = np$ and n is not too small.

Several conditions are assumed to hold for a Poisson process that produces a Poisson random variable:

1. For small intervals, the probability of exactly one occurrence is approximately proportional to the length of the interval (where λ, the event rate or intensity, is the constant of proportionality).
2. For small intervals, the probability of more than one occurrence is essentially equal to zero (see below).
3. The numbers of occurrences in two non-overlapping intervals are statistically independent.

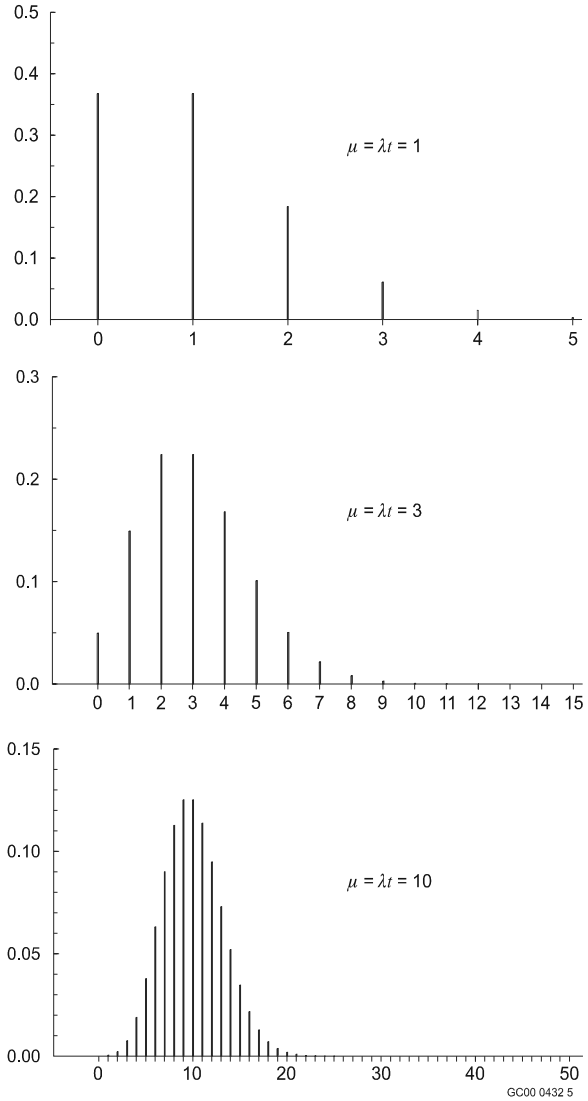

Figure A.7 Three Poisson probability distribution functions.

More precise versions of condition 2 are: (1) the probability of more than one event occurring in a very short time interval is negligible in comparison to the probability that only one event occurs (Meyer 1970), (2) the probability of more than one event occurring in a very short time interval goes to zero faster than the length of the interval (Pfeiffer and Schum 1973), and (3) simultaneous events occur only with probability zero (Çinlar 1975). All of these versions have the practical interpretation that common cause events do not occur. The phrase "do not occur" is used in this handbook, as it is in Thompson (1981).

The Poisson distribution also can serve as an **approximation to the binomial distribution**. Poisson random variables can be viewed as resulting from an experiment involving a large number of trials, n, that each have a small probability of occurrence, p, of an event. However, the rare occurrence is offset by the large number of trials. As stated above, the binomial distribution gives the probability that an occurrence will take place exactly x times in n trials. If $p = \mu/n$ (so that p is small for large n), and n is large, the binomial probability that the rare occurrence will take place exactly x times is closely approximated by the Poisson distribution with $\mu = np$. In general, the approximation is good for large n, small p, and moderate μ (say $\mu \leq 20$) (see Derman et al. 1973).

The Poisson distribution is important because it describes the behavior of many rare event occurrences, regardless of their underlying physical process. It also has many applications to describing the occurrences of system and component failures under steady-state conditions. These applications utilize the relationship between the Poisson and **exponential** (continuous random variable, see Section A.7.4) distributions: the times between successive events follow an exponential distribution.

A.7 Continuous Random Variables

A.7.1 The Uniform Distribution

A **uniform** distribution, also referred to as a rectangular distribution, represents the situation where any value in a specified interval, say $[a, b]$, is equally likely. For a uniform random variable, X, because the outcomes are equally likely, $f(x)$ is equal to a constant. The p.d.f. of a uniform distribution with parameters a and b, denoted uniform(a, b) is

$$f(x) = \frac{1}{b-a}$$

for $a \leq x \leq b$.

Figure A.8 shows the density of the uniform(a, b) distribution.

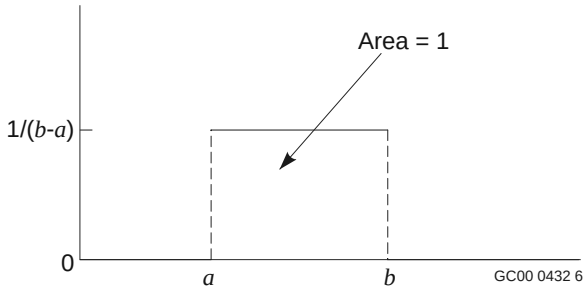

Figure A.8 Density of uniform(a, b) distribution.

The mean and variance of a uniform(a, b) distribution are

$$E(X) = \frac{b+a}{2}$$

and

$$\mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

A.7.2 The Normal Distribution

One of the most widely encountered continuous probability distributions is the **normal** distribution, which has the familiar bell shape and is symmetrical about its mean value. The importance of the normal distribution is due to: (1) its applicability in describing a very large number of random variables that occur in nature and (2) the fact that certain useful functions of nonnormal random variables are approximately normal. Details on the derivation of the normal distribution can be found in many basic mathematical statistics textbooks (e.g., Hogg and Craig 1995).

The normal distribution is characterized by two parameters, μ and σ. For a random variable, X, that is normally distributed with parameters μ and σ, the p.d.f. of X is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \qquad \text{(A.8)}$$

for $-\infty < x < \infty$, $-\infty < \mu < \infty$, and $\sigma > 0$. Increasing μ moves the density curve to the right and increasing σ spreads the density curve out to the right and left while lowering the peak of the curve. The units of μ and σ are the same as for X.

The mean and variance of a normal distribution with parameters μ and σ are

$E(X) = \mu$

and

$\mathrm{Var}(X) = \sigma^2$.

The normal distribution is denoted normal(μ, σ^2).

Figure A.9 shows two normal(μ, σ^2) densities. The distribution is largest at μ and is more concentrated around μ when σ is small than when σ is large.

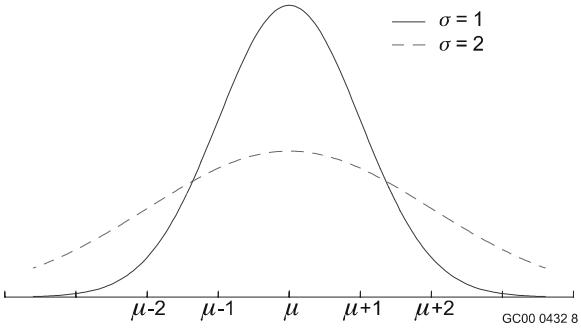

Figure A.9 Two normal densities.

Note the similarity of the normal density to a binomial p.d.f. with large np or a Poisson p.d.f. with large μ. This illustrates the fact that a normal distribution can sometimes be used to approximate those distributions.

The normal(0, 1) distribution is called the **standard normal** distribution, which, from Equation A.8, has p.d.f.

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \qquad \text{(A.9)}$$

for $-\infty < x < \infty$. The cumulative distribution of the standard normal distribution is denoted by Φ. Tables for the standard normal distribution are presented in Appendix C and in almost all books on statistics.

It can be shown that the transformed random variable $Z = (X - \mu)/\sigma$ is normal(0, 1). Thus, to calculate probabilities for a normal(μ, σ^2) random variable, X,

when $\mu \neq 0$ and/or $\sigma^2 \neq 1$, the tables for the standard normal can be used. Specifically, for any number a,

$$\Pr[X \leq a] = \Pr[(X - \mu)/\sigma \leq (a - \mu)/\sigma]$$

$$= \Pr[Z \leq (a - \mu)/\sigma]$$

$$= \Phi[(a - \mu)/\sigma] .$$

Part of the importance of the normal distribution is that it is the distribution that sample sums and sample means tend to possess as n becomes sufficiently large. This result is known as the **central limit theorem**, which states that, if X_1, X_2, ..., X_n, are independent random variables, each with mean μ and variance σ^2, the sum of these n random variables, $\Sigma_i X_i$, tends toward a normal($n\mu$, $n\sigma^2$) distribution for large enough n. Since the sample mean is a linear combination of this sum, the central limit theorem also applies. Thus, $\overline{X}$ = $\Sigma_i X_i/n$ tends to a normal(μ, σ^2/n) distribution. The importance of the central limit theorem is it can be used to provide approximate probability information for the sample sums and sample means of random variables whose distributions are unknown. Further, because many natural phenomena consist of a sum of several random contributors, the normal distribution is used in many broad applications.

Because a binomial random variable is a sum, it tends to the normal distribution as n gets large. Thus, the normal distribution can be used as an **approximation to the binomial distribution**. One rule of thumb is that the approximation is adequate for $np \geq 5$.

A Poisson random variable also represents a sum and, as presented previously, can also be used as an approximation to the binomial distribution. It follows that the normal distribution can serve as an **approximation to the Poisson distribution** when $\mu = \lambda t$ is large. One rule of thumb is that the approximation is adequate for $\mu \geq 5$.

A.7.3 The Lognormal Distribution

Use of the **lognormal distribution** has become increasingly widespread. It is commonly used as a distribution for failure time and in maintainability analysis (Martz and Waller 1991). It has also been widely used as a prior distribution for unknown positive parameters.

The lognormal distribution arises from the *product* of many independent random variables. If $Y = Y_1 \cdot Y_2 \cdot \ldots \cdot Y_n = \prod_i Y_i$ is the product of n independent positive random variables that are (nearly) identically distributed, then $\ln(Y) = \ln(\prod_i Y_i) = \Sigma_i \ln(Y_i)$ is a sum that tends toward a normal distribution.

The distribution of Y is defined to be lognormal when the distribution of $\ln(Y)$ is normal. That is, when Y is lognormal, $\ln(Y)$ is normal(μ, σ^2). The parameters of the lognormal distribution are μ and σ, the parameters from the underlying normal distribution. For a random variable, Y, that is lognormally distributed with parameters μ and σ, denoted lognormal(μ, σ^2), the p.d.f. of Y is

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

for $0 < y < \infty$, $-\infty < \mu < \infty$, and $\sigma > 0$. Note the y in the denominator, for reasons explained in Section A.4.7. The mean and variance of a lognormal(μ, σ^2) distribution are

$E(Y) = \exp(\mu + \sigma^2/2)$

and

$\mathrm{Var}(Y) = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1]$.

In addition, the median of a lognormal distribution is $\exp(\mu)$ and the mode is $\exp(\mu - \sigma^2)$. See Martz and Waller (1991) for more information on the lognormal distribution.

Sometimes the median of $Y = \exp(\mu)$ is used as a parameter. In addition, a parameter commonly used in PRA is the **error factor** (EF), where $EF = \exp(1.645\sigma)$, and is defined as

$\Pr[\mathrm{med}(Y)/EF \leq Y \leq \mathrm{med}(Y)*EF] = 0.90$.

Figure A.10 shows three lognormal densities. The value $\mu = -7$ corresponds to a median of about 1.E-3. [More exactly, it corresponds to $\exp(-7)$ = 9.E-4.] The value $\mu = -6.5$ corresponds to a median of about

1.5E−3. The value σ = 0.67 corresponds to an error factor EF = 3, and σ = 1.4 corresponds to an error factor EF = 10.

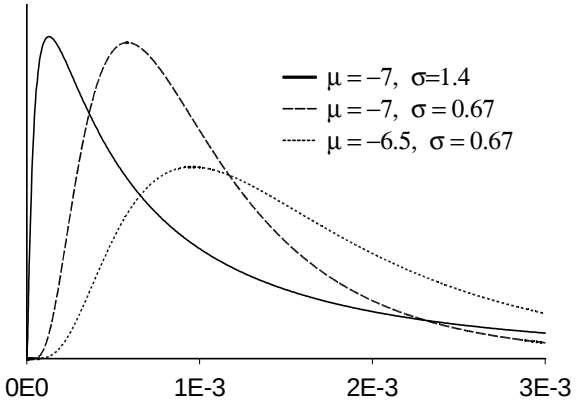

Figure A.10 Three lognormal densities.

The two distributions with σ = 0.67 and different values of μ have essentially the same shape, but with different scales. The larger μ corresponds to spreading the distribution out more from zero. The distribution with σ = 1.4, and therefore EF = 10, has a very skewed distribution.

To calculate probabilities for a lognormal(μ, σ^2) random variable, Y, the tables for the standard normal can be used. Specifically, for any number b,

$$\Pr[\, Y \le b \,] = \Pr[\, \ln(Y) \le \ln(b) \,]$$

$$= \Pr[\, X \le \ln(b) \,]$$

$$= \Phi[\, (\ln(b) - \mu)/\sigma \,] ,$$

where $X = \ln(Y)$ is normal(μ, σ^2).

A.7.4 The Exponential Distribution

The **exponential distribution** is widely used for modeling time to failure and is inherently associated with the Poisson process (see Martz and Waller 1991). For a Poisson random variable X defining the number of failures in a time interval t and for a random variable T defining the time to failure, it can be shown that T has the exponential p.d.f.

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

for $t > 0$. Thus, the time to first failure and the times between successive failures follow an exponential distribution and the number of failures in a fixed time interval follows a Poisson distribution.

Figure A.11 shows two exponential densities, for two values of λ. The intercept (height of the curve when t = 0) equals λ. Thus, the figure shows that the distribution is more concentrated near zero if λ is large. This agrees with the interpretation of λ as a frequency of failures and t as time to first failure.

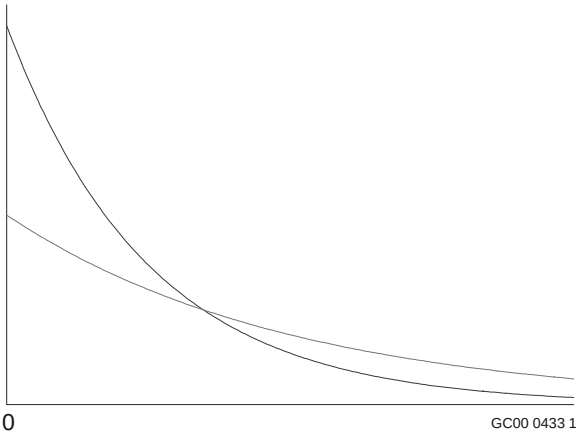

Figure A.11 Two exponential densities.

The exponential distribution parameter, λ, corresponds to the λt parameterization of the Poisson p.d.f. in Equation A.7. and is referred to as the **failure rate** if the component or system is repaired and restarted immediately after each failure. It is called the **hazard rate** if the component or system can only fail once and cannot be repaired. Section 4.6.2 discusses modeling duration times with different distributions and defines the hazard rate as $h(t) = f(t)/[1 - F(t)]$. For the exponential distribution, the hazard rate is constant, λ.

The c.d.f. of the exponential distribution is

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

The exponential distribution with parameter λ is denoted exponential(λ). The mean and variance of an exponential(λ) distribution are

$$E(T) = 1/\lambda$$

and

$$\mathrm{Var}(T) = 1/\lambda^2.$$

The relationship of the exponential distribution to the Poisson process can be seen by observing that the probability of no failures before time t can be viewed in two ways. First, the number of failures, X, can be counted. The probability that the count is equal to 0 is given by Equation A.7 as

$$\Pr(X = 0) = e^{-\lambda t}\frac{(\lambda t)^0}{0!} = e^{-\lambda t}.$$

Alternatively, the probability that first failure time, T, is greater than t is

$$\begin{aligned}\Pr(T > t) &= 1 - \Pr(T \le t)\\ &= 1 - F(t)\\ &= 1 - [1 - e^{-\lambda t}]\\ &= e^{-\lambda t}.\end{aligned}$$

Thus, the two approaches give the same expression for the probability of no failures before time t.

The assumptions of a Poisson process require a constant failure rate, λ, which can be interpreted to mean that the failure process has no memory (Martz and Waller 1991). Thus, if a device is still functioning at time t, it remains as good as new and its remaining life has the same exponential(λ) distribution. This constant failure rate corresponds to the flat part of the common "bathtub" failure curve (number of failures plotted against time) and does not pertain to initial (burn-in) failures and wear-out failures.

A different, sometimes useful, parameterization uses $\mu = 1/\lambda = E(T)$. For example, if T represents a time to failure, μ is called the mean time to failure. If T is the time to repair, or to fire suppression, or to some other event, the name for μ is the mean time to repair, or other appropriate name. The exponential(μ) distribution for T has density

$f(t) = (1/\mu)\exp(-t/\mu)$, for $t \ge 0$

and c.d.f.

$F(t) = 1 - \exp(-t/\mu)$, for $t \ge 0$.

The units of μ are the same as the units of t, minutes or hours or whatever the data have. The mean and variance are

$E(T) = \mu$
$\text{var}(T) = \mu^2$.

A.7.5 The Weibull Distribution

The **Weibull distribution** is widely used in reliability and PRA and generalizes the exponential distribution to include nonconstant failure or hazard rates (Martz and Waller 1991). Different Weibull distributions have been successfully used to describe initial failures and wear-out failures. The Weibull distribution is appropriate when a system is composed of a number of components, and system failure is due to any one of the components failing. It, therefore, is commonly referred to as a distribution corresponding to failure of the weakest link.

For a random variable, T, that has a Weibull distribution, the p.d.f. is

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

for $t \ge \theta \ge 0$ and parameters $\alpha > 0$ and $\beta > 0$. The parameter θ is a location parameter and corresponds to a period of guaranteed life that is not present in many applications (Martz and Waller 1991). Thus, θ is usually set to zero. The c.d.f. for T is

$$F(t) = 1 - \exp\left[-\left(\frac{t-\theta}{\alpha}\right)^{\beta}\right],$$

for $t \ge \theta$ and $\alpha > 0$ and $\beta > 0$.

The α parameter is a scale parameter that expands or contracts the density along the horizontal axis. The β parameter is a shape parameter that allows for a wide variety of distribution shapes [see Martz and Waller (1991) for further discussion and examples]. When $\beta = 1$, the distribution reduces to the exponential distribution. Therefore, the Weibull family of distributions includes the exponential family of distributions as a special case.

A Weibull distribution with parameters α, β, and θ is referred to as Weibull(α, β, θ) and, when $\theta = 0$, Weibull(α, β). The mean and variance of the Weibull distribution are given by Martz and Waller (1991) as

$\theta + \alpha\Gamma(1 + 1/\beta)$

and

$\alpha^2[\Gamma(1 + 2/\beta) - \Gamma^2(1 + 1/\beta)$.

Here, Γ is the gamma function, defined in Sec. A.7.6.

Figure A.12 shows four Weibull densities, all with the same scale parameter, α, and all with location parameter $\theta = 0$. The shape parameter, β, varies. When $\beta < 1$, the density becomes infinite at the origin. When $\beta = 1$, the distribution is identical to the exponential distribution. Surprisingly, the distribution is *not* asymptotically normal as β becomes large, although it is approximately normal when β is near 3.

A.7.6 The Gamma and Chi-Squared Distributions

The **gamma distribution** is an extension of the exponential distribution and is sometimes used as a failure time model (Martz and Waller, 1991). It is also often used as a prior distribution in Bayesian estimation (see Appendix B) of the failure rate parameter λ from Poisson(λt) or exponential(λ) data. The **chi-squared distribution** is a re-expression of a special case of the gamma distribution.

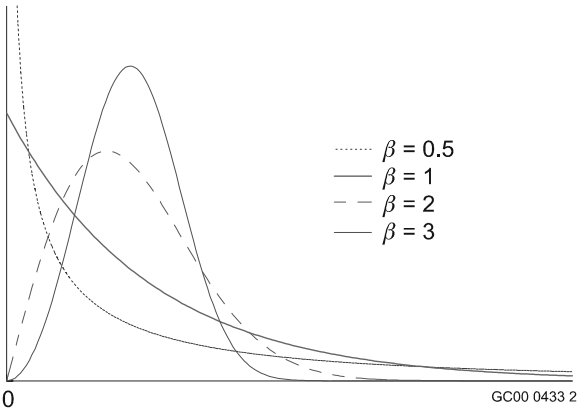

Figure A.12 Four Weibull densities, all having $\theta = 0$ and all having the same α.

The gamma distribution arises in many ways. The distribution of the sum of independent exponential(λ) random variables is gamma, which forms the basis for a confidence interval for λ from exponential(λ) data. Because the sum of n independent exponentially distributed random variables has a gamma distribution, the gamma distribution is often used as the distribution of the time, or waiting time, to the nth event in a Poisson process. In addition, the chi-squared distribution is the distribution for a sum of squares of independent, identically distributed normal random variables, which forms the basis for a confidence interval for the variance of a normal distribution. The gamma distribution is also often used as a distribution for a positive random variable, similar to the lognormal and Weibull distributions. In PRA work, it is often used as a Bayesian distribution for an uncertain positive parameter.

Two parameterizations of the gamma distribution are common, with various letters used for the parameters. The parameterization given here is most useful for Bayesian updates, the primary use of the gamma distribution in this handbook. For a random variable, T, that has a gamma distribution, the p.d.f. is

$$f(t) = \frac{\beta^a}{\Gamma(\alpha)} t^{\alpha - 1} \exp(-t\beta),$$

for t, α, and $\beta > 0$. Here

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

is the **gamma function** evaluated at α. If α is a positive integer, $\Gamma(\alpha) = (\alpha - 1)!$.

A gamma distribution with parameters α and β is referred to as gamma(α, β). The mean and variance of the gamma(α, β) random variable, T, are:

$E(T) = \alpha/\beta$

and

$\mathrm{Var}(T) = \alpha/\beta^2$.

The parameters α and β are referred to as the shape and scale parameters. The shape parameter α allows the density to have many forms. If α is near zero, the distribution is highly skewed. For $\alpha = 1$, the gamma distribution reduces to an exponential(β^{-1}) distribution. Also, the gamma($\alpha = n/2$, $\beta = ½$) distribution is known as the **chi-squared distribution** with n degrees of freedom, denoted $\chi^2(n)$. The p.d.f. for the $\chi^2(n)$ distribution is found by substituting these values into the above formula for the gamma p.d.f. It also can be found in many statistics texts (e.g., Hogg and Craig 1995, Chapter 4).

In addition, if T has a gamma(α, β) distribution, then $2\beta T$ has a $\chi^2(2\alpha)$ distribution, which forms the defining relationship between the two distributions. The gamma and chi-squared distributions can, therefore, be viewed as two ways of expressing one distribution. Since the chi-squared distribution usually is only allowed to have integer degrees of freedom, the gamma distribution can be thought of as an interpolation of the chi-squared distribution.

Percentiles of the chi-squared distribution are tabulated in Appendix C. These tables can be used as follows to find the percentiles of any gamma distribution. The $100\times p$ percentile of a gamma(α, β) distribution is $\chi^2_p(2\alpha)/(2\beta)$, where $\chi^2_p(2\alpha)$ denotes the $100\times p$ percentile of the chi-squared distribution with 2α degrees of freedom.

Figure A.13 shows gamma densities with four shape parameters, α. When $\alpha < 1$, the density becomes infinite at 0. When $\alpha = 1$, the density is identical to an exponential density. When α is large, the distribution is approximately a normal distribution.

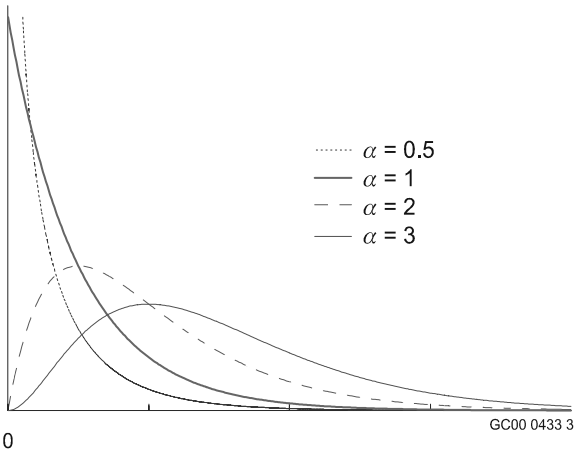

Figure A.13 Gamma densities with four shape parameters.

As stated previously, the sum of exponential lifetimes or waiting times has a gamma distribution, with the shape parameter α equal to the number of exponential lifetimes. Thus, when α is large, the **gamma distribution is approximately normal**.

An alternative parameterization of the gamma distribution uses the scale parameter, say $\tau = \beta^{-1}$. If T has a gamma(α, τ) distribution, its p.d.f. is

$$f(t) = \frac{1}{\tau^{\alpha}\Gamma(\alpha)} t^{\alpha-1}\exp(-t/\tau)$$

for t, α, and $\tau > 0$. The mean and variance of the gamma(α, τ) random variable, T, are:

$E(T) = \alpha\tau$

and

$\mathrm{Var}(T) = \alpha\tau^2$.

This alternative parameterization is useful in a very small portion of this handbook.

A.7.7 The Inverted Gamma and Inverted Chi-Squared Distributions

The **inverted gamma distribution** is often used as a prior distribution for Bayesian estimation of the time to failure of an exponential(λ) distribution (Martz and Waller 1991). It is also used as a prior and posterior distribution for σ^2 when the data have a normal distribution with variance σ^2 (Box and Tiao 1973, Lee 1997).

For a gamma(α, β) random variable, T, $W = 1/T$ has an inverted gamma distribution with p.d.f.

$$f(w) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\left(\frac{1}{w}\right)^{\alpha+1}\exp\left(-\frac{\beta}{w}\right),$$

for w, α, and $\beta > 0$. The parameters here are the same as for the gamma distribution. For example, if T has units of time then w and β both have units 1/time. A comparison of this density with the gamma density shows that this density has an extra w^2 in the denominator, for reasons explained in Section A.4.7.

The parameters of the inverted gamma distribution are α and β and this distribution is denoted inverted gamma(α, β). Similar to the gamma(α, β) distribution, α is the shape parameter and β is the scale parameter.

The distribution can also be parameterized in terms of $\tau = \beta^{-1}$.

The mean and variance of an inverted gamma(α, β) random variable, *W*, are

$$E(W) = \frac{\beta}{\alpha - 1}, \qquad \alpha > 1,$$

and

$$\mathrm{Var}(W) = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)}, \qquad \alpha > 2.$$

Note that, for $\alpha \le 1$, the mean and higher moments do not exist and, for $1 < \alpha \le 2$, the mean exists but the variance does not exist (Martz and Waller, 1991).

Figure A.14 shows four inverted gamma distributions, all having the same scale parameter, β, and having various shape parameters, α.

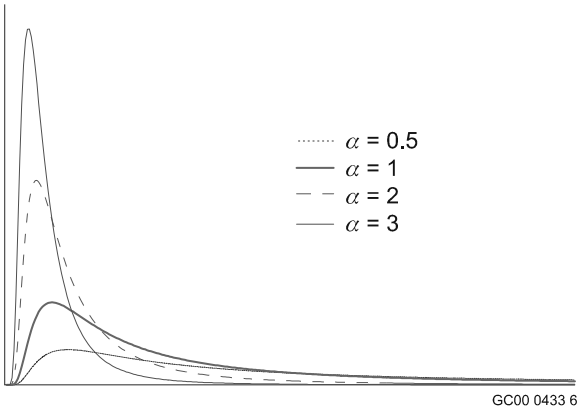

Figure A.14 Four inverted gamma densities, having the same scale parameter, β, and various shape parameters, α.

In the special case with $\alpha = n/2$ and $\beta = ½$, the distribution is called the **inverted chi-squared** distribution with *n* degrees of freedom. Values from this distribution are sometimes denoted $\chi^{-2}(n)$. This form of the distribution is often used in connection with a prior for σ^2 when the data are normally distributed.

A.7.8 The Beta Distribution

Many continuous quantitative phenomena take on values that are bounded by known numbers *a* and *b*. Examples are percentages, proportions, ratios, and distance to failure points on items under stress. The **beta distribution** is a versatile family of distributions that is useful for modeling phenomena that can range from 0 to 1 and, through a transformation, from *a* to *b*.

The beta distribution family includes the uniform distribution and density shapes that range from decreasing to uni-modal right-skewed to symmetric to U-shaped to uni-modal left-skewed to increasing (Martz and Waller 1991). It can serve as a model for a reliability variable that represents the probability that a system or component lasts at least *t* units of time. The beta distribution is also widely used in Bayesian estimation and reliability analysis as a prior distribution for the binomial distribution parameter *p* that represents a reliability or failure probability.

The p.d.f. of a beta random variable, *Y*, is

$$f(y) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha - 1}(1 - y)^{\beta - 1},$$

for $0 \le y \le 1$, with the parameters α, $\beta > 0$, and is denoted beta(α, β). The gamma functions at the front of the p.d.f. form a normalizing constant so that the density integrates to 1.

The mean and variance of the beta(α, β) random variable, *Y*, are

$$E(Y) = \frac{\alpha}{\alpha + \beta}$$

and

$$\mathrm{Var}(Y) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Various beta distributions are shown in Figures A.15 and A.16. Figure A.15 shows beta densities with $\alpha = \beta$, and therefore with mean 0.5. When $\alpha < 1$, the density becomes infinite at 0.0, and when $\beta < 1$, the density becomes infinite at 1.0. When $\alpha = \beta = 1$, the density is uniform. When α and β are large, the density is approximately normal.

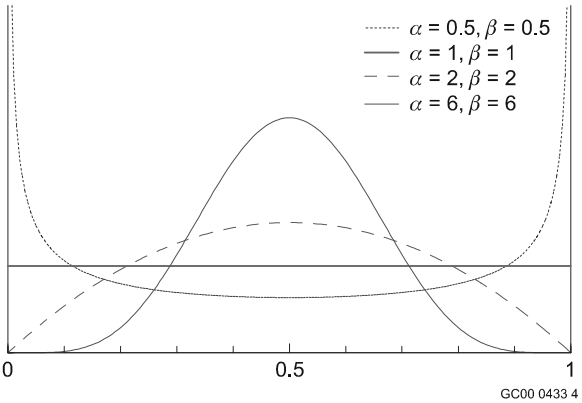

Figure A.15 Beta distributions with mean = 0.5.

Figure A.16 shows densities with mean 0.1. Again, when $\alpha < 1$, the density becomes infinite at 0.0, and when $\alpha > 1$, the density is zero at 0.0. As the parameters α and β become large, the density approaches a normal distribution.

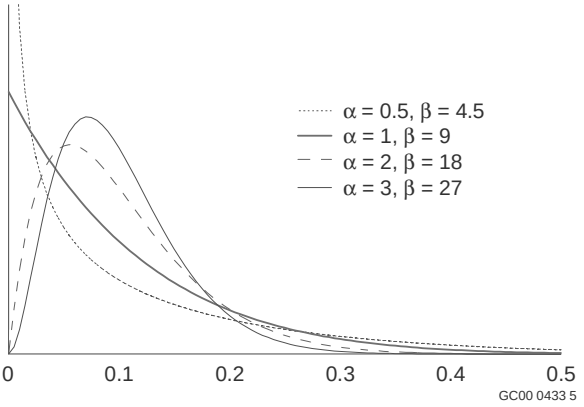

Figure A.16 Four beta distributions with mean 0.1.

Another parameterization of the beta distribution uses the parameters $x_0 = \alpha$ and $n_0 = \alpha + \beta$. This parameterization is used by Martz and Waller (1991) because it simplifies Bayes formulas and Bayesian estimation. The p.d.f. of a beta(x_0, n_0) is

$$f(y) = \frac{\Gamma(n_0)}{\Gamma(x_0)\Gamma(n_0 - x_0)} y^{x_0 - 1}(1-y)^{n_0 - x_0 - 1},$$

for $0 \le y \le 1$, with the parameters x_0 and n_0 satisfying

$n_0 > x_0 > 0$.

The mean and variance of the beta(x_0, n_0) random variable, Y, are

$$E(Y) = \frac{x_0}{n_0}$$

and

$$\mathrm{Var}(Y) = \frac{x_0(n_0 - x_0)}{n_0^2(n_0 + 1)}.$$

Percentiles of the beta distribution occur in the formula for a confidence interval for p, and in the formula for a Bayes credible interval for p when a conjugate prior is used. Some percentiles are tabulated in Appendix C. In addition, many software packages, including some commonly used spreadsheets, can calculate these percentiles. If none of these work, Martz and Waller (1991) give a method for finding the beta percentiles from the corresponding percentiles of an F distribution. The F distribution is tabulated in most statistics books, and can be interpolated if necessary with good accuracy. The relation is

$$\mathrm{beta}_q(\alpha, \beta) = \alpha / [\alpha + \beta F_{1-q}(2\beta, 2\alpha)]$$

for small q, and

$$\mathrm{beta}_q(\alpha, \beta) = \alpha F_q(2\alpha, 2\beta) / [\beta + \alpha F_q(2\alpha, 2\beta)]$$

for large q. Here $\mathrm{beta}_q(\alpha, \beta)$ denotes the q quantile, or the $100 \times q$ percentile, of the beta(α, β) distribution, and $F_q(d_1, d_2)$ denotes the q quantile of an F distribution with d_1 and d_2 degrees of freedom. So if all else fails, and a statistics book with F tables is nearby, the first formula can be used to find the lower percentile of the beta distribution and the second formula can be used to find the upper percentile. This method is not discussed further here, because it is not expected to be needed often.

A.7.9 The Logistic-Normal Distribution

While not widely used in PRA, this distribution is commonly used for Bayesian inference in other fields of application, especially as a prior for the binomial parameter p when p could plausibly be fairly large. X has a logistic-normal distribution if $\ln[X/(1 - X)]$ is normally distributed with some mean μ and variance σ^2. The function $\ln[X/(1 - X)]$ may appear strange, but it is common enough in some areas of application to have a

name, the **logit** function. Therefore, the above statements could be rewritten to say that X has a logistic-normal distribution if logit(X) is normally distributed.

Properties of the logistic-normal distribution are summarized here.

- Let $y = \ln[x/(1 - x)]$. Then $x = e^y / (1 + e^y)$. This implies that x must be between 0 and 1.
- As x increases from 0 to 1, $y = \ln[x/(1 - x)]$ increases monotonically from $-\infty$ to $+\infty$. Thus, y can be generated from a normal distribution with no problem of going outside the possible range.
- The monotonic relation between x and y means that the percentiles match. For example, the 95th percentile of Y is $\mu + 1.645\sigma$. Denote this by $y_{0.95}$. Therefore, the 95th percentile of X is
  $x_{0.95} = \exp(x_{0.95}) / [1 + \exp(x_{0.95})]$.
  Alternatively, this can be written as
  $y_{0.95} = \ln[x_{0.95} / (1 - x_{0.95})]$.
- If X is close to 0 with high probability, so that $X/(1 - X)$ is close to X with high probability, then the logistic-normal and lognormal distributions are nearly the same.

The third bullet shows how to find the percentiles of a logistic-normal distribution. Unfortunately there is no equally easy way to find the moments, such as the mean or variance. Moments must be found using numerical integration.

Figure A.17 shows several logistic normal distributions that all have median 0.5. These correspond to a normally distributed y with mean $\mu = 0$ and with various values of σ. Figure A.18 shows several logistic normal distributions that all have median 0.1. These correspond to a normally distributed y with mean $\mu = -2.2 = \ln[0.1/(1 - 0.1)]$.

Note the general similarities to the beta distributions in Figures A.15 and A.16. Note also the differences: Logistic-normal distributions are characterized most easily by percentiles, whereas beta distributions are characterized most easily by moments. Also, the beta densities can be J-shaped or U-shaped, but the logistic-normal densities always drop to zero at the ends of the range.

If Z has a standard normal distribution, X has a chi-squared distribution with d degrees of freedom, and Z

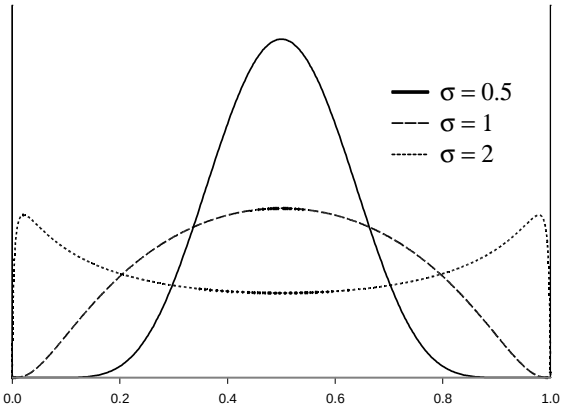

Figure A.17 Three logistic-normal densities with median = 0.5.

A.7.10 Student's *t* Distribution

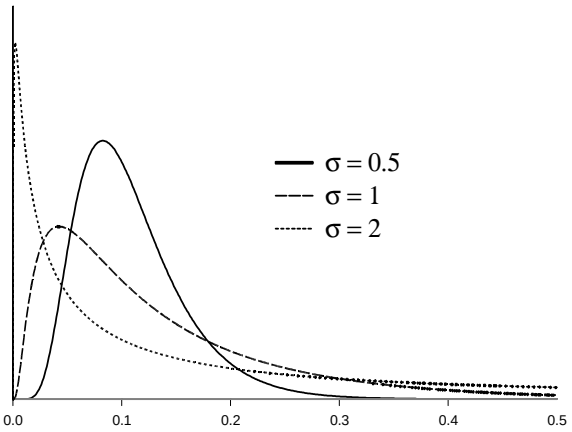

Figure A.18 Three logistic-normal densities with median = 0.1.

The **Student's *t*** distribution is not used in a central way in PRA. However, it appears in a peripheral way in places in this handbook, when dealing with the parameters of a normal or lognormal distribution, or in large-sample situations when a distribution is approximated as normal or lognormal. Therefore, the basic facts are summarized here.

and X are statistically independent, then

$$T = \frac{Z}{\sqrt{X / d}}$$

has a Student's t distribution with d degrees of freedom. Therefore, T has a distribution that is symmetrical about 0, and it can take values in the entire real line. If d is large, the denominator is close to 1 with high probability, and T has approximately a standard normal distribution. If d is smaller, the denominator adds extra variability, and the extreme percentiles of T are farther out than are the corresponding normal percentiles. Tables of the distribution are given in Appendix C.

Although not needed for ordinary work, the p.d.f. and first two moments of T are given here. (See many standard texts, such DeGroot 1975.) The p.d.f. is

$$f(t) = \frac{\Gamma[(d+1)/2]}{(d\pi)^{1/2}\Gamma(d/2)}\left[1+(t^2/2)\right]^{-(d+1)/2} .$$

If $d > 1$ the mean is 0. If $d > 2$ the variance is $d/(d-2)$. If $d \le 2$ the variance does not exist. If $d = 1$, even the mean does not exist; in this case the distribution is called a **Cauchy distribution**.

REFERENCES FOR APPENDIX A

Ascher, H. and H. Feingold, 1984, *Repairable Systems Reliability – Modeling, Inference, Misconceptions and Their Causes*, Marcel Dekker, Inc., New York.

Bain, L. J., and M. Engelhardt, 1992, *Introduction to Probability and Mathematical Statistics, 2nd Ed.*, PWS-Kent Pub., Boston.

Çinlar, E., 1975, *Introduction to Stochastic Processes*. Prentice-Hall, Englewood Cliffs, NJ.

DeGroot, Morris H., 1975, *Probability and Statistics*, Addison-Wesley, Reading, MA.

Derman, C., L. J. Gleser and I. Olkin, 1973, *A Guide to Probability Theory and Application*. Holt, Rinehart and Winston, NY.

Esary, J. D. and F. Proschan, 1970, A Reliability Bound for Systems of Maintained, Interdependent Components in *Journal of the American Statistical Association*, Vol. 65, pp. 329-338.

Esary, J. D. and F. Proschan, 1963, Coherent Structures of Non-Identical Components, in *Technometrics*, Vol 6, pp. 191-209.

Hogg, R. V. and A. T. Craig, 1995, *Introduction to Mathematical Statistics*. Macmillan, NY.

Lee, Peter M., 1997, *Bayesian Statistics: An Introduction, Second Edition*, Arnold, a member of the Hodder Headline Group, London.

Martz, H. F. and R. A. Waller, 1991, *Bayesian Reliability Analysis*. R. E. Krieger Publishing Co., Malabar, FL.

Meyer, P. L., 1970, *Introductory Probability and Statistical Applications*. Addison-Wesley Pub. Co., Reading, MA.

NRC, 1994, *A Review of NRC Staff Uses of Probabilistic Risk Assessment*, NUREG-1489. U. S. Nuclear Regulatory Commission, Washington, D.C.

Pfeiffer, P. E. and D. A. Schum, 1973, *Introduction to Applied Probability*. Academic Press, NY.

Snedecor, G. W. and W. G. Cochran, 1989, *Statistical Methods.* Iowa State University Press, Ames, IA.

Thompson, W. A., Jr., 1981, On the Foundations of Reliability, in *Technometrics*, Vol. 23, No. 1, pp. 1-13.

Vesely, W. E., F. F. Goldberg, N. H. Roberts and D. F. Haasl, 1981, *Fault Tree Handbook*. NUREG-0492. U.S. Nuclear Regulatory Commission, Washington, D.C.