# 1. Basic Statistical Concepts

This brief chapter presents some fundamental elements of engineering probability and statistics with which some readers are probably already familiar, but others may not be. Statistics is the study of how best one can describe and analyze the data and then draw conclusions or inferences based on the data available. The first section of this chapter begins with some basic definitions, including probability axioms, basic statistics and reliability measures.

The second section describes the most common distribution functions such as the binomial, Poisson, geometric, exponential, normal, log normal, Student's *t*, gamma, Pareto, Beta, Rayleigh, Cauchy, Weibull and Vtub-shaped hazard rate distributions, their applications and their use in engineering and applied statistics.

The third section describes statistical inference, including parameter estimation and confidence intervals. Statistical inference is the process by which information from sample data is used to draw conclusions about the population from which the sample was selected that hopefully represents the whole population. This discussion also introduces the maximum likelihood estimation (MLE) method, the method of moments, MLE with censored data, the statistical change-point estimation method, nonparametic tolerance limits, sequential sampling and Bayesian methods.

The fourth section briefly discusses stochastic processes, including Markov processes, Poisson processes, renewal processes, quasi-renewal processes, and nonhomogeneous Poisson processes.



Finally, the last section provides a short list of books for readers who are interested in advanced engineering and applied statistics.

## 1.1 Basic Probability Measures

We start off this chapter by defining several useful terms:

1. Outcome: A result or observation from an experiment, which cannot be predicted with certainty.
2. Event: Subset of a set of all possible outcomes.
3. Probability: The relative frequency at which an event occurs in a large number of identical experiments.
4. Random variable: A function which assigns real numbers to the outcomes of an experiment.

5. Statistics: A function (itself a random variable) of one or more random variables, that does not depend upon any unknown parameters.

### 1.1.1 Probability Axioms

Now let $C$ be a subset of the sample space ($C \subset \mathbb{Z}$). A probability set function, denoted by P($C$), has the following properties:

1. $P(\mathbb{Z}) = 1, \quad P(C) \geq 0$
2. $P(C_1 \cup C_2 \cup \ldots) = P(C_1) + P(C_2) + \ldots$

where the subsets $C_i$ have no elements in common (i. e., they are mutually exclusive).

Let $C_1$ and $C_2$ be two subsets of the sample space $\mathbb{Z}$. The conditional probability of getting an outcome in $C_2$ given that an outcome from $C_1$ is given by

$$P(C_2/C_1) = \frac{P(C_2 \cap C_1)}{P(C_1)}.$$

Let $C_1, C_2, \ldots, C_n$ be $n$ mutually disjoint subsets of the sample space $\mathbb{Z}$. Let $C$ be a subset of the union of the $C_i$s; that is

$$C \subset \bigcup_{i=1}^{n} C_i.$$

Then

$$P(C) = \sum_{i=1}^{n} P(C/C_i)P(C_i) \tag{1.1}$$

and

$$P(C_i/C) = \frac{P(C/C_i)P(C_i)}{\sum_{i=1}^{n} P(C/C_i)P(C_i)}.$$

Equation (1.1) is known as the law of total probability.

### 1.1.2 Basic Statistics

The cumulative distribution function (cdf) $F$ is a unique function which gives the probability that a random variable $X$ takes on values less than or equal to some value $x$. In other word, $F(x) = P(X \leq x)$.

The probability density function (pdf) $f$ is the probability that $X$ takes on the value $x$; that is, $f(x) = P(X = x)$.

The pdf satisfies the following two relations for discrete and continuous random variables, respectively,

$$\sum_{\text{all } x} f(x) = 1$$

and

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1.$$

In the continuous case, the pdf is the derivative of the cdf:

$$f(x) = \frac{\partial F(x)}{\partial x}.$$

The expected value of a random variable $X$ is given by

$$E(X) = \sum_{\text{all } x} x f(x)$$

in the discrete case, and by

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x$$

in the continuous case. Similarly, the variance of a random variable $X$, denoted by $\sigma^2$, is a measure of how the values of $X$ are spread about the mean value. It is defined as

$$\sigma^2 = E\,(X - \mu)^2\,.$$

It is calculated for discrete and continuous random variables, respectively, by

$$\sigma^2 = \sum_{\text{all } x} (x - \mu)^2 f(x)$$

and

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\,\mathrm{d}x.$$

The standard deviation of $X$, denoted by $\sigma$, is the square root of the variance.

The skewness coefficient of a random variable $X$ is a measure of the symmetry of the distribution of $X$ about its mean value $\mu$, and is defined as

$$S_\mathrm{c} = \frac{E(X - \mu)^3}{\sigma^3}.$$

The skewness is zero for a symmetric distribution, negative for a left-tailed distribution, and positive for a right-tailed distribution.

Similarly, the kurtosis coefficient of a random variable $X$ is a measure of how much of the mass of the distribution is contained in the tails, and is defined as

$$K_\mathrm{c} = \frac{E(X - \mu)^4}{\sigma^4}.$$

Obviously, kurtosis is always positive; however, larger values represent distributions with heavier tails.

Assume there are $n$ random variables $X_1, X_2, \ldots, X_n$ which may or may not be mutually independent. The joint cdf, if it exists, is given by

$$P(X_1 \le x_1, X_2 \le x_2, \ldots X_n \le x_n) = \int_{-\infty}^{x_n} \int_{-\infty}^{x_{n-1}} .. \int_{-\infty}^{x_1} f(t_1, t_2, .., t_n)\,\mathrm{d}t_1\,\mathrm{d}t_2..\mathrm{d}t_n$$

If the $n$ random variables are mutually statistically independent, then the joint pdf can be rewritten as

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i).$$

The conditional distribution of a random variable $Y$ given that another random variable $X$ takes on a value $x$ is given by:

$$f(y/X = x) = \frac{f(x, y)}{f_1(x)},$$

where

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}y.$$

Given a random sample of size $n$ from a distribution, the sample mean and sample variance are defined as, respectively,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2.$$

### 1.1.3 Reliability Measures

Definitions of reliability given in the literature vary according to the practitioner or researcher. The generally accepted definition is as follows.

*Definition 1.1*

*Reliability* is the probability of success or the probability that the system will perform its intended function under specified design limits.

More specifically, reliability is the probability that a product or system will operate properly for a specified period of time (design life) under the design operating conditions (such as temperature, voltage, etc.) without failure. In other words, reliability can be used as a measure of the system's success at providing its function properly. Reliability is one of the quality characteristics that consumers require from manufacturers.

Mathematically, reliability $R(t)$ is the probability that a system will be successful in the interval from time 0 to time $t$:

$$R(t) = P(T > t), \quad t \ge 0, \tag{1.2}$$

where $T$ is a random variable denoting the time-to-failure or failure time.

*Unreliability*, or the cdf $F(t)$, a measure of failure, is defined as the probability that the system will fail by time $t$.

$$F(t) = P(T \le t), \quad t \ge 0.$$

In other words, $F(t)$ is the failure distribution function. If the time-to-failure random variable $T$ has a density function $f(t)$, then

$$R(t) = \int_{t}^{\infty} f(s)\,\mathrm{d}s$$

or, equivalently,

$$f(t) = -\frac{\mathrm{d}}{\mathrm{d}t}[R(t)].$$

The density function can be mathematically described in terms of $T$:

$$\lim_{\Delta t \to 0} P(t < T \le t + \Delta t).$$

This can be interpreted as the probability that the failure time $T$ will occur between the operating time $t$ and the next interval of operation $t + \Delta t$.

Consider a new and successfully tested system that operates well when put into service at time $t = 0$. The system becomes less likely to remain successful as the time interval increases. The probability of success for an infinite time interval is, of course, zero. Thus, the system starts to function at a probability of one and eventually decreases to a probability of zero. Clearly, reliability is a function of mission time. For example, one can say that the reliability of the system is 0.995 for a mission time of 24 h.

*Example 1.1:* A computer system has an exponential failure time density function

$$f(t) = \frac{1}{9000} e^{-\frac{t}{9000}}, \quad t \geq 0.$$

The probability that the system will fail after the warranty (six months or 4380 h) and before the end of the first year (one year or 8760 h) is given by

$$P(4380 < T \leq 8760) = \int_{4380}^{8760} \frac{1}{9000} e^{-\frac{t}{9000}} dt$$
$$= 0.237.$$

This indicates that the probability of failure during the interval from six months to one year is 23.7%.

Consider the Weibull distribution, where the failure time density function is given by

$$f(t) = \frac{\beta t^{\beta-1}}{\theta^{\beta}} e^{-(\frac{t}{\theta})^{\beta}}, \quad t \geq 0,\ \theta > 0,\ \beta > 0.$$

Then the reliability function is

$$R(t) = e^{-(\frac{t}{\theta})^{\beta}}, \quad t \geq 0.$$

Thus, given a particular failure time density function or failure time distribution function, the reliability function can be obtained directly. Section 1.2 provides further insight for specific distributions.

### System Mean Time to Failure

Suppose that the reliability function for a system is given by $R(t)$. The expected failure time during which a component is expected to perform successfully, or the system mean time to failure (MTTF), is given by

$$\text{MTTF} = \int_0^{\infty} t f(t) dt \tag{1.3}$$

or, equivalently, that

$$\text{MTTF} = \int_0^{\infty} R(t) dt. \tag{1.4}$$

Thus, MTTF is the definite integral evaluation of the reliability function. In general, if $\lambda(t)$ is defined as the failure rate function, then, by definition, MTTF is not equal to $1/\lambda(t)$.

The MTTF should be used when the failure time distribution function is specified because the reliability level implied by the MTTF depends on the underlying failure time distribution. Although the MTTF measure is one of the most widely used reliability calculations, it is also one of the most misused calculations. It has been misinterpreted as a "guaranteed minimum lifetime". Consider the results given in Table 1.1 for a twelve-component life duration test.

**Table 1.1** Results from a twelve-component life duration test

| Component | Time to failure (h) |
|---|---|
| 1 | 4510 |
| 2 | 3690 |
| 3 | 3550 |
| 4 | 5280 |
| 5 | 2595 |
| 6 | 3690 |
| 7 | 920 |
| 8 | 3890 |
| 9 | 4320 |
| 10 | 4770 |
| 11 | 3955 |
| 12 | 2750 |

A component MTTF of 3660 h was estimated using a basic averaging technique. However, one of the components failed after 920 h. Therefore, it is important to note that the system MTTF denotes the average time to failure. It is neither the failure time that could be expected 50% of the time nor is it the guaranteed minimum time of system failure, but mostly depends on the failure distribution.

A careful examination of (1.4) will show that two failure distributions can have the same MTTF and yet produce different reliability levels.

### Failure Rate Function

The probability of a system failure in a given time interval $[t_1, t_2]$ can be expressed in terms of the reliability function as

$$\int_{t_1}^{t_2} f(t) dt = \int_{t_1}^{\infty} f(t) dt - \int_{t_2}^{\infty} f(t) dt$$
$$= R(t_1) - R(t_2)$$

or in terms of the failure distribution function (or the unreliability function) as

$$\int_{t_1}^{t_2} f(t) dt = \int_{-\infty}^{t_2} f(t) dt - \int_{-\infty}^{t_1} f(t) dt$$
$$= F(t_2) - F(t_1).$$

The rate at which failures occur in a certain time interval $[t_1, t_2]$ is called the *failure rate*. It is defined as the probability that a failure per unit time occurs in the interval, given that a failure has not occurred prior to $t_1$, the beginning of the interval. Thus, the failure rate is

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

Note that the failure rate is a function of time. If we redefine the interval as $[t, t + \Delta t]$, the above expression becomes

$$\frac{R(t) - R(t + \Delta t)}{\Delta t R(t)}.$$

The rate in the above definition is expressed in failures per unit time, but in reality the time units might instead correspond to miles, hours, trials, etc. The *hazard function* is defined as the limit of the failure rate as the interval approaches zero. Thus, the hazard function $h(t)$ is the instantaneous failure rate, and is defined by

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{R(t) - R(t + \Delta t)}{\Delta t R(t)} \\ &= \frac{1}{R(t)} \left[ -\frac{\mathrm{d}}{\mathrm{d}t} R(t) \right] \\ &= \frac{f(t)}{R(t)}. \end{aligned} \tag{1.5}$$

The quantity $h(t)\,\mathrm{d}t$ represents the probability that a device of age $t$ will fail in the small interval of time $t$ to $(t + \mathrm{d}t)$. The importance of the hazard function is that it indicates the change in the failure rate over the life of a population of components by plotting their hazard functions on a single axis. For example, two designs may provide the same reliability at a specific point in time, but the failure rates up to this point in time can differ.

The death rate, in statistical theory, is analogous to the failure rate, as the nature of mortality is analogous to the hazard function. Therefore, the hazard function, hazard rate or failure rate function is the ratio of the pdf to the reliability function.

## 1.2 Common Probability Distribution Functions

This section presents some of the most common distribution functions and several hazard models that are applied in engineering statistics [1.1].

### 1.2.1 Discrete Random Variable Distributions

#### Binomial Distribution

The binomial distribution is one of the most widely used discrete random variable distributions in reliability and quality inspection. It has applications in reliability engineering, for example when one is dealing with a situation in which an event is either a success or a failure.

The binomial distribution can be used to model a random variable $X$ which represents the number of successes (or failures) in $n$ independent trials (these are referred to as Bernoulli trials), with the probability of success (or failure) being $p$ in each trial. The pdf of the distribution is given by

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, 2, \ldots, n,$$

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

where $n =$ number of trials, $x =$ number of successes, $p =$ single trial probability of success.

The mean of the binomial distribution is $np$ and the variance is $np(1 - p)$. The coefficient of skewness is given by

$$S_c = \frac{1 - 2p}{\sqrt{np(1 - p)}}$$

and the coefficient of kurtosis is

$$K_c = 3 - \frac{6}{n} + \frac{1}{np(1 - p)}.$$

The reliability function $R(k)$ (i. e., at least $k$ out of $n$ items are good) is given by

$$R(k) = \sum_{x=k}^{n} \binom{n}{x} p^x (1 - p)^{n-x}.$$

*Example 1.2:* Suppose that, during the production of lightbulbs, 90% are found to be good. In a random sample of 20 lightbulbs, the probability of obtaining at least 18 good lightbulbs is given by

$$\begin{aligned} R(18) &= \sum_{x=18}^{20} \binom{20}{18} (0.9)^x (0.1)^{20-x} \\ &= 0.667. \end{aligned}$$

### Poisson Distribution

Although the Poisson distribution can be used in a manner similar to the binomial distribution, it is used to deal with events in which the sample size is unknown. A Poisson random variable is a discrete random variable distribution with a probability density function given by

$$P(X = x) = \frac{\lambda^x \mathrm{e}^{-\lambda}}{x!} \quad \text{for } x = 0, 1, 2, \ldots \tag{1.6}$$

where $\lambda =$ constant failure rate; $x =$ is the number of events. In other words, $P(X = x)$ is the probability that exactly $x$ failures occur.

A Poisson distribution is used to model a Poisson process. A Poisson random variable has a mean and a variance both equal to $\lambda$ where $\lambda$ is called the parameter of the distribution. The skewness coefficient is

$$S_\mathrm{c} = \frac{1}{\sqrt{\lambda}}$$

and the kurtosis coefficient is

$$K_\mathrm{c} = 3 + \frac{1}{\lambda}.$$

The Poisson distribution reliability up to time $t$, $R(k)$ (the probability of $k$ or fewer failures), can be defined as follows

$$R(k) = \sum_{x=0}^{k} \frac{(\lambda t)^x \mathrm{e}^{-\lambda t}}{x!}.$$

This distribution can be used to determine the number of spares required for a system during a given mission.

*Example 1.3:* A nuclear plant is located in an area susceptible to both high winds and earthquakes. From historical data, the mean frequency of large earthquakes capable of damaging important plant structures is one every 50 y. The corresponding frequency of damaging high winds is once in 25 y. During a strong earthquake, the probability of structure damage is 0.1. During high winds, the damage probability is 0.05. Assume that earthquakes and high winds can be described by independent Poisson random variables and that the damage caused by these events are independent. Let us answer the following questions:

1. What is the probability of having strong winds but not large earthquakes during a 10y period?
2. What is the probability of having strong winds and large earthquakes in the 10y period?
3. What is the probability of building damage during the 10y period?

Considering the first question, let the random variables $X$ and $Y$ represent the number of earthquakes and the number of occurrences of high winds, respectively. We assume that the two random variables are statistically independent. The means of $X$ and $Y$ are, respectively, given by

$$\lambda_X = \frac{1}{50\,\mathrm{y}}\,(10\,\mathrm{y}) = 0.2$$

and

$$\lambda_Y = \frac{1}{25\,\mathrm{y}}\,(10\,\mathrm{y}) = 0.4\,.$$

The conditional damage probabilities are given as follows:

$$P(\text{damage/earthquake}) = 0.1$$

and

$$P(\text{damage/wind}) = 0.05.$$

Let event

$A =$ {strong winds and no earthquakes},
$B =$ {strong winds and large earthquakes},
$C =$ {building damage}.

Assuming that the winds and earthquakes are independent of each other, the probability of having strong winds but not earthquakes during the 10 y period can be written as

$$\begin{aligned} P(A) &= P(\text{winds})P(\text{no earthquakes}) \\ &= [1 - P(\text{no winds})]P(\text{no earthquakes}) \end{aligned}$$

Therefore, we obtain

$$P(A) = (1 - \mathrm{e}^{-0.4})(\mathrm{e}^{-0.2}) = 0.27$$

For the second question, the probability of having strong winds and earthquakes during the 10 y period can be obtained from

$$\begin{aligned} P(B) &= P(\text{winds})P(\text{earthquakes}) \\ &= [1 - P(\text{no winds})][1 - P(\text{no earthquakes})] \\ &= (1 - \mathrm{e}^{-0.4})(1 - \mathrm{e}^{-0.2}) = 0.06\,. \end{aligned}$$

Finally, for the third question, we assume that multiple occurrences of earthquakes and high winds do not occur during the 10 y period. Therefore, the probability of

building damage can be written as

$$\begin{aligned}P(C) &= P(\text{damage/earthquakes})P(\text{earthquakes})\\ &\quad + P(\text{damage/wind})P(\text{wind})\\ &\quad - P(\text{damage/earthquakes and wind})\\ &\quad P(\text{earthquake and wind})\\ &= P(\text{damage/earthquakes})P(\text{earthquakes})\\ &\quad + P(\text{damage/wind})P(\text{wind})\\ &\quad - P(\text{damage/earthquakes})P(\text{damage/wind})\\ &\quad P(\text{earthquake and wind})\\ &= (1-\mathrm{e}^{-0.2})(0.1)+(1-\mathrm{e}^{-0.4})(0.05)\\ &\quad -(0.05)(0.1)(0.06)\\ &= 0.0343\,.\end{aligned}$$

#### Geometric Distribution

Consider a sequence of independent trials where each trial has the same probability of success, $p$. Let $N$ be a random variable representing the number of trials until the first success. This distribution is called the geometric distribution. It has a pdf given by

$$P(N=n) = p\,(1-p)^{n-1}\,, \quad n = 1, 2, \ldots\,.$$

The corresponding cdf is

$$F(n) = 1-(1-p)^{n}\,, \quad n = 1, 2, \ldots\,.$$

The expected value and the variance are, respectively,

$$E(N) = \frac{1}{p}$$

and

$$V(N) = \frac{1-p}{p^2}.$$

#### Hypergeometric Distribution

The hypergeometric distribution is a discrete distribution that arises in sampling, for example. It has a pdf given by

$$f(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \quad x = 0, 1, 2, \ldots, n. \tag{1.7}$$

Typically, $N$ will be the number of units in a finite population; $n$ will be the number of samples drawn without replacement from $N$; $k$ will be the number of failures in the population; and $x$ will be the number of failures in the sample.

The expected value and variance of the hypergeometric random variable $X$ are, respectively

$$E(X) = \frac{nk}{N}$$

and

$$V(X) = \frac{k(N-k)n(N-n)}{N^2(N-1)}.$$

### 1.2.2 Continuous Distributions

#### Exponential Distribution

The exponential distribution plays an essential role in reliability engineering because it has a constant failure rate. It has been used to model the lifetimes of electronic and electrical components and systems. This distribution is applicable to the case where a used component that has not failed is as good as a new component – a rather restrictive assumption. It should therefore be used carefully, since there are numerous situations where this assumption (known as the "memoryless property" of the distribution) is not valid.

If the time to failure is described by an exponential failure time density function, then

$$f(t) = \frac{1}{\theta}\,\mathrm{e}^{-\frac{t}{\theta}}, \quad t \geq 0,\ \theta > 0 \tag{1.8}$$

and this will lead to the reliability function

$$R(t) = \int_t^\infty \frac{1}{\theta}\,\mathrm{e}^{-\frac{s}{\theta}}\,\mathrm{d}s = \mathrm{e}^{-\frac{t}{\theta}}, \quad t \geq 0,$$

where $\theta = 1/\lambda > 0$ is an MTTF's parameter and $\lambda \geq 0$ is a constant failure rate.

The hazard function or failure rate for the exponential density function is constant, i. e.,

$$h(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{\theta}\,\mathrm{e}^{-\frac{1}{\theta}}}{\mathrm{e}^{-\frac{1}{\theta}}} = \frac{1}{\theta} = \lambda.$$

The failure rate for this distribution is $\lambda$, a constant, which is the main reason for this widely used distribution. Because of its constant failure rate, the exponential is an excellent model for the long flat "intrinsic failure" portion of the bathtub curve. Since most parts and systems spend most of their lifetimes in this portion of the bathtub curve, this justifies frequent use of the exponential distribution (when early failures or wearout is not a concern). The exponential model works well for interarrival times. When these events trigger failures, the exponential lifetime model can be used.

We will now discuss some properties of the exponential distribution that can be used to understand its characteristics and when and where it can be applied.

*Property 1.1*

(*Memoryless property*) The exponential distribution is the only continuous distribution that satisfies

$$P\{T \geq t\} = P\{T \geq t+s | T \geq s\} \quad \text{for } t > 0,\ s > 0. \tag{1.9}$$

This result indicates that the conditional reliability function for the lifetime of a component that has survived to time $s$ is identical to that of a new component. This term is the so-called "used as good as new" assumption.

*Property 1.2*

If $T_1, T_2, \ldots, T_n$, are independently and identically distributed exponential random variables (r.v.'s) with a constant failure rate $\lambda$, then

$$2\lambda \sum_{i=1}^{n} T_i \sim \chi^2(2n), \tag{1.10}$$

where $\chi^2(2n)$ is a chi-squared distribution with $2n$ degrees of freedom. This result is useful for establishing a confidence interval for $\lambda$.

### Uniform Distribution

Let $X$ be a random variable with a uniform distribution over the interval $(a, b)$ where $a < b$. The pdf is given by

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}.$$

The expected value and variance are, respectively,

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

and

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

### Normal Distribution

The normal distribution plays an important role in classical statistics due to the *Central Limit Theorem*. In production engineering, the normal distribution primarily applies to measurements of product susceptibility and external stress. This two-parameter distribution is used to describe mechanical systems in which a failure results from some wearout effect. The normal distribution takes the well-known bell shape. This distribution is symmetrical about the mean and the spread is measured by the variance. The larger the value, the flatter the distribution. The pdf is given by

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}, \quad -\infty < t < \infty,$$

where $\mu$ is the mean value and $\sigma$ is the standard deviation. The cumulative distribution function (cdf) is

$$F(t) = \int_{-\infty}^{t} \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{s-\mu}{\sigma}\right)^2} \mathrm{d}s.$$

The reliability function is

$$R(t) = \int_{t}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{s-\mu}{\sigma}\right)^2} \mathrm{d}s.$$

There is no closed-form solution for the above equation. However, tables for the standard normal density function are readily available (see Table 1.6 in Sect. 1.A) and can be used to find probabilities for any normal distribution. If

$$Z = \frac{T-\mu}{\sigma}$$

is substituted into the normal pdf, we obtain

$$f(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}}, \quad -\infty < Z < \infty.$$

This is a so-called standard normal pdf, with a mean value of 0 and a standard deviation of 1. The standardized cdf is given by

$$\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}s^2} \mathrm{d}s, \tag{1.11}$$

where $\Phi$ is a standard normal distribution function. Thus, for a normal random variable $T$, with mean $\mu$ and standard deviation $\sigma$,

$$P(T \leq t) = P\left(Z \leq \frac{t-\mu}{\sigma}\right) = \Phi\left(\frac{t-\mu}{\sigma}\right),$$

where $\Phi$ yields the relationship required if standard normal tables are to be used.

It should be noted that the coefficent of kurtosis in the normal distribution is 3. The hazard function for a normal distribution is a monotonically increasing function

of $t$. This is easily shown by proving that $h'(t) \geq 0$ for all $t$. Since

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

then

$$h'(t) = \frac{R(t)f'(t) + f^2(t)}{R^2(t)} \geq 0.$$

One can attempt this proof by using the basic definition of a normal density function $f$.

*Example 1.4:* A component has a normal distribution of failure times with $\mu = 2000\,\text{h}$ and $\sigma = 100\,\text{h}$. The reliability of the component at 1900 h is required.

Note that the reliability function is related to the standard normal deviate $z$ by

$$R(t) = P\left(Z > \frac{t-\mu}{\sigma}\right),$$

where the distribution function for $Z$ is given by (1.11). For this particular application,

$$\begin{aligned} R(1900) &= P\left(Z > \frac{1900-2000}{100}\right) \\ &= P(z > -1). \end{aligned}$$

From the standard normal table in Table 1.6 in Sect. 1.A, we obtain

$$R(1,900) = 1 - \Phi(-1) = 0.8413.$$

The value of the hazard function is found from the relationship

$$h(t) = \frac{f(t)}{R(t)} = \frac{\Phi\left(\frac{t-\mu}{\sigma}\right)}{\sigma R(t)},$$

where $\Phi$ is the pdf of the standard normal density. Here

$$\begin{aligned} h(1900) &= \frac{\Phi(-1.0)}{\sigma R(t)} = \frac{0.1587}{100(0.8413)} \\ &= 0.0019 \text{ failures/cycle}. \end{aligned}$$

The normal distribution is flexible enough to make it a very useful empirical model. It can be theoretical derived under assumptions matching many failure mechanisms. Some of these are: corrosion, migration, crack growth, and failures resulting from chemical reactions or processes in general. That does not mean that the normal distribution is always the correct model for these mechanisms, but it does perhaps explain why it has been empirically successful in so many of these cases.

### Log Normal Distribution

The log normal lifetime distribution is a very flexible model that can empirically fit many types of failure data. This distribution, when applied in mechanical reliability engineering, is able to model failure probabilities of repairable systems, the compressive strength of concrete cubes, the tensile strength of fibers, and the uncertainty in failure rate information. The log normal density function is given by

$$f(t) = \frac{1}{\sigma t\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln t-\mu}{\sigma}\right)^2}, \quad t \geq 0, \tag{1.12}$$

where $\mu$ and $\sigma$ are parameters such that $-\infty < \mu < \infty$, and $\sigma > 0$. Note that $\mu$ and $\sigma$ are not the mean and standard deviations of the distribution.

Its relationship to the normal (just take natural logarithms of all of the data and time points and you have "normal" data) makes it easy to work with many good software analysis programs used to treat normal data.

Mathematically, if a random variable $X$ is defined as $X = \ln T$, then $X$ is normally distributed with a mean of $\mu$ and a variance of $\sigma^2$. That is,

$$E(X) = E(\ln T) = \mu$$

and

$$V(X) = V(\ln T) = \sigma^2.$$

Since $T = e^X$, the mean of the log normal distribution can be found via the normal distribution. Consider that

$$E(T) = E(e^X) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\left[x - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]} dx.$$

By rearranging the exponent, this integral becomes

$$E(T) = e^{\mu+\frac{\sigma^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\left[x-(\mu+\sigma^2)\right]^2} dx.$$

Thus, the mean of the log normal distribution is

$$E(T) = e^{\mu+\frac{\sigma^2}{2}}.$$

Proceeding in a similar manner,

$$E(T^2) = E(e^{2X}) = e^{2(\mu+\sigma^2)}$$

so the variance for the log normal is

$$V(T) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1).$$

The coefficient of skewness of this distribution is

$$S_c = \frac{e^{3\sigma^2} - 3e^{\sigma^2} + 2}{\left(e^{\sigma^2} - 1\right)^{\frac{3}{2}}}.$$

It is interesting that the skewness coefficient does not depend on $\mu$ and grows rapidly as the variance $\sigma^2$ increases.

The cumulative distribution function for the log normal is

$$F(t) = \int_0^t \frac{1}{\sigma s\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln s - \mu}{\sigma}\right)^2} ds$$

and this can be related to the standard normal deviate $Z$ by

$$F(t) = P(T \le t) = P(\ln T \le \ln t) = P\left(Z \le \frac{\ln t - \mu}{\sigma}\right).$$

Therefore, the reliability function is given by

$$R(t) = P\left(Z > \frac{\ln t - \mu}{\sigma}\right) \tag{1.13}$$

and the hazard function would be

$$h(t) = \frac{f(t)}{R(t)} = \frac{\Phi\left(\frac{\ln t - \mu}{\sigma}\right)}{\sigma t R(t)}$$

where $\Phi$ is the cdf of standard normal density.

The log normal lifetime model, like the normal, is flexible enough to make it a very useful empirical model. It can be theoretically derived under assumptions matching many failure mechanisms, including corrosion, migration, crack growth and failures resulting from chemical reactions or processes in general. As with the normal distribution, this does not mean that the log normal is always the correct model for these mechanisms, but it suggests why it has been empirically successful in so many of these cases.

### Student's *t* Distribution

Student's $t$ probability density function of a random variable $T$ is given by:

$$f(t) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{r}{2}\right)\left(1 + \frac{t^2}{r}\right)^{\frac{r+1}{2}}} \quad \text{for } -\infty < t < \infty. \tag{1.14}$$

In other words, if a random variable $T$ is defined as

$$T = \frac{W}{\sqrt{\frac{V}{r}}},$$

where $W$ is a standard normal random variable and $V$ has a chi-square distribution with $r$ degrees of freedom, and $W$ and $V$ are statistically independent, then $T$ is Student's $t$-distributed, and parameter $r$ is referred to as the degrees of freedom (see Table 1.7 in Sect. 1.A).

### The *F* Distribution

Let us define the random variable $F$ is as follows

$$F = \frac{U/r_1}{V/r_2},$$

where $U$ has a chi-square distribution with $r_1$ degrees of freedom, $V$ is also chi-square-distributed, with $r_2$ degrees of freedom, and $U$ and $V$ are statistically independent, then the probability density function of $F$ is given by

$$f(t) = \frac{\Gamma\left(\frac{r_1+r_2}{2}\right)\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}}(t)^{\frac{r_1}{2}-1}}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)\left(1 + \frac{r_1 t}{r_2}\right)^{\frac{r_1+r_2}{2}}} \quad \text{for } t > 0. \tag{1.15}$$

The $F$ distribution is a two-parameter – $r_1$ and $r_2$ – distribution where the parameters are the degrees of freedom of the underlying chi-square random variables (see Table 1.8 in Sect. 1.A).

It is worth noting that if $T$ is a random variable with a $t$ distribution and $r$ degrees of freedom, then the random variable $T^2$ has an $F$ distribution with parameters $r_1 = 1$ and $r_2 = r$. Similarly, if $F$ is $F$-distributed with $r_1$ and $r_2$ degrees of freedom, then the random variable $Y$, defined as

$$Y = \frac{r_1 F}{r_2 + r_1 F}$$

has a beta distribution with parameters $r_1/2$ and $r_2/2$.

### Weibull Distribution

The exponential distribution is often limited in applicability owing to its memoryless property. The Weibull distribution [1.2] is a generalization of the exponential distribution and is commonly used to represent fatigue life, ball-bearing life and vacuum tube life. The Weibull distribution is extremely flexible and appropriate for modeling component lifetimes with fluctuating hazard rate functions and is used to represent various

types of engineering applications. The three-parameter probability density function is

$$f(t) = \frac{\beta(t-\gamma)^{\beta-1}}{\theta^{\beta}} \mathrm{e}^{-\left(\frac{t-\gamma}{\theta}\right)^{\beta}}, \quad t \geq \gamma \geq 0, \tag{1.16}$$

where $\theta$ and $\beta$ are known as the scale and shape parameters, respectively, and $\gamma$ is known as the location parameter. These parameters are always positive. By using different parameters, this distribution can follow the exponential distribution, the normal distribution, etc. It is clear that, for $t \geq \gamma$, the reliability function $R(t)$ is

$$R(t) = \mathrm{e}^{-\left(\frac{t-\gamma}{\theta}\right)^{\beta}} \quad \text{for } t > \gamma > 0,\ \beta > 0,\ \theta > 0 \tag{1.17}$$

hence,

$$h(t) = \frac{\beta(t-\gamma)^{\beta-1}}{\theta^{\beta}}, \quad t > \gamma > 0,\ \beta > 0,\ \theta > 0. \tag{1.18}$$

It can be shown that the hazard function decreases for $\beta < 1$, increases for $\beta > 1$, and is constant when $\beta = 1$.

Note that the Rayleigh and exponential distributions are special cases of the Weibull distribution at $\beta = 2$, $\gamma = 0$, and $\beta = 1$, $\gamma = 0$, respectively. For example, when $\beta = 1$ and $\gamma = 0$, the reliability of the Weibull distribution function in (1.17) reduces to

$$R(t) = \mathrm{e}^{-\frac{t}{\theta}}$$

and the hazard function given in (1.18) reduces to $1/\theta$, a constant. Thus, the exponential is a special case of the Weibull distribution. Similarly, when $\gamma = 0$ and $\beta = 2$, the Weibull probability density function becomes the Rayleigh density function. That is,

$$f(t) = \frac{2}{\theta} t \mathrm{e}^{-\frac{t^2}{\theta}} \quad \text{for } \theta > 0,\ t \geq 0.$$

### Other Forms of Weibull Distributions

The Weibull distribution is widely used in engineering applications. It was originally proposed in order to represent breaking strength distributions of materials. The Weibull model is very flexible and has also been applied in many applications as a purely empirical model, with theoretical justification. Other forms of Weibull probability density function include, for example,

$$f(x) = \lambda\gamma x^{\gamma-1} \mathrm{e}^{-\lambda t^{\gamma}}. \tag{1.19}$$

When $\gamma = 2$, the density function becomes a Rayleigh distribution.

It is easy to show that the mean, variance and reliability of the above Weibull distribution are, respectively:

$$\begin{aligned} \text{Mean} &= \lambda^{\frac{1}{\gamma}} \Gamma\left(1+\frac{1}{\gamma}\right); \\ \text{Variance} &= \lambda^{\frac{2}{\gamma}} \left\{ \Gamma\left(1+\frac{2}{\gamma}\right) - \left[\Gamma\left(1+\frac{1}{\gamma}\right)\right]^2 \right\}; \\ \text{Reliability} &= \mathrm{e}^{-\lambda t^{\gamma}}. \end{aligned} \tag{1.20}$$

*Example 1.5:* The time to failure of a part has a Weibull distribution with $\frac{1}{\lambda} = 250$ (measured in $10^5$ cycles) and $\gamma = 2$. The part reliability at $10^6$ cycles is given by:

$$R(10^6) = \mathrm{e}^{-(10)^2/250} = 0.6703.$$

The resulting reliability function is shown in Fig. 1.1.

### Gamma Distribution

The gamma distribution can be used as a failure probability function for components whose distribution is skewed. The failure density function for a gamma distribution is

$$f(t) = \frac{t^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} \mathrm{e}^{-\frac{t}{\beta}}, \quad t \geq 0,\ \alpha,\ \beta > 0, \tag{1.21}$$

where $\alpha$ is the shape parameter and $\beta$ is the scale parameter. In this expression, $\Gamma(\alpha)$ is the gamma function, which is defined as

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} \mathrm{e}^{-t} \mathrm{d}t \quad \text{for } \alpha > 0.$$

Hence,

$$R(t) = \int_t^{\infty} \frac{1}{\beta^{\alpha}\Gamma(\alpha)} s^{\alpha-1} \mathrm{e}^{-\frac{s}{\beta}} \mathrm{d}s.$$

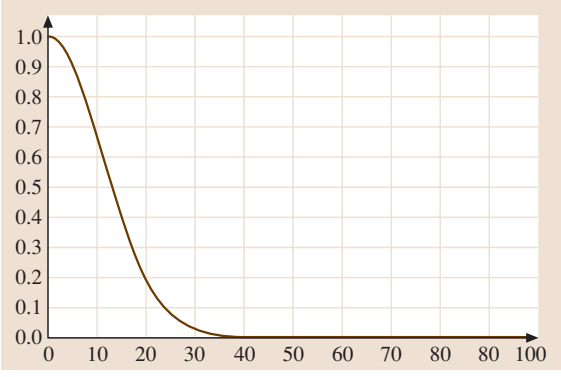


**Fig. 1.1** Weibull reliability function versus time

If $\alpha$ is an integer, it can be shown by successive integration by parts that

$$R(t) = \mathrm{e}^{-\frac{t}{\beta}} \sum_{i=0}^{\alpha-1} \frac{\left(\frac{t}{\beta}\right)^i}{i!} \tag{1.22}$$

and

$$h(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{\beta^\alpha \Gamma(\alpha)} t^{\alpha-1} \mathrm{e}^{-\frac{t}{\beta}}}{\mathrm{e}^{-\frac{t}{\beta}} \sum_{i=0}^{\alpha-1} \frac{\left(\frac{t}{\beta}\right)^i}{i!}}. \tag{1.23}$$

The gamma density function has shapes that are very similar to the Weibull distribution. At $\alpha = 1$, the gamma distribution becomes the exponential distribution with a constant failure rate $1/\beta$. The gamma distribution can also be used to model the time to the $n$th failure of a system if the underlying failure distribution is exponential. Thus, if $X_i$ is exponentially distributed with parameter $\theta = 1/\beta$, then $T = X_1 + X_2 + \cdots + X_n$ is gamma-distributed with parameters $\beta$ and $n$.

*Example 1.6:* The time to failure of a component has a gamma distribution with $\alpha = 3$ and $\beta = 5$. Obtain the reliability of the component and the hazard rate at 10 time units.

Using (1.22), we compute

$$R(10) = \mathrm{e}^{-\frac{10}{5}} \sum_{i=0}^{2} \frac{\left(\frac{10}{5}\right)^i}{i!} = 0.6767\,.$$

The hazard rate is given by

$$h(10) = \frac{f(10)}{R(10)} = \frac{0.054}{0.6767} = 0.798 \text{ failures/unit time.}$$

The other form of the gamma probability density function can be written as follows:

$$f(x) = \frac{\beta^\alpha t^{\alpha-1}}{\Gamma(\alpha)} \mathrm{e}^{-t\beta}, \quad t > 0. \tag{1.24}$$

This pdf is characterized by two parameters: the shape parameter $\alpha$ and the scale parameter $\beta$. When $0 < \alpha < 1$, the failure rate monotonically decreases; when $\alpha > 1$, the failure rate monotonically increases; when $\alpha = 1$ the failure rate is constant.

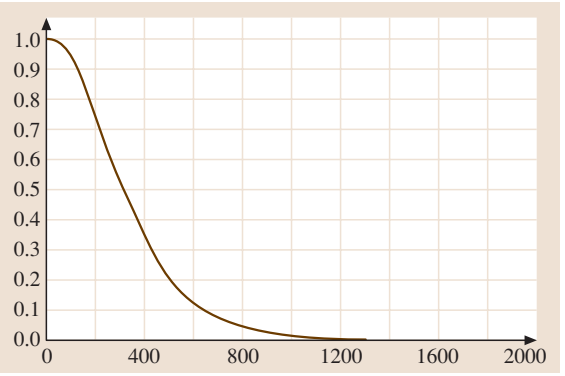


**Fig. 1.2** Gamma reliability function versus time

The mean, variance and reliability of the gamma random variable are:

$$\text{Mean (MTTF)} = \frac{\alpha}{\beta};$$

$$\text{Variance} = \frac{\alpha}{\beta^2};$$

$$\text{Reliability} = \int_t^\infty \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} \mathrm{e}^{-x\beta}\,\mathrm{d}x.$$

*Example 1.7:* A mechanical system time to failure is gamma-distributed with $\alpha = 3$ and $1/\beta = 120$. The system reliability at 280 h is given by

$$R(280) = \mathrm{e}^{\frac{-280}{120}} \sum_{k=0}^{2} \frac{\left(\frac{280}{120}\right)^2}{k!} = 0.851\,19$$

and the resulting reliability plot is shown in Fig. 1.2.

The gamma model is a flexible lifetime model that may offer a good fit to some sets of failure data. Although it is not widely used as a lifetime distribution model for common failure mechanisms, the gamma lifetime model is commonly used in Bayesian reliability applications.

### Beta Distribution

The two-parameter beta density function $f(t)$ is given by

$$f(t) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha-1}(1-t)^{\beta-1}\,,$$
$$0 < t < 1,\ \alpha > 0,\ \beta > 0\,, \tag{1.25}$$

where $\alpha$ and $\beta$ are the distribution parameters. This two-parameter beta distribution is commonly used in many reliability engineering applications and also

plays an important role in the theory of statistics. Note that the beta-distributed random variable takes on values in the interval $(0, 1)$, so the beta distribution is a natural model when the random variable represents a probability. Likewise, when $\alpha = \beta = 1$, the beta distribution reduces to a uniform distribution.

The mean and variance of the beta random variable are, respectively,

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

and

$$V(T) = \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2}.$$

### Pareto Distribution

The Pareto distribution was originally developed to model income in a population. Phenomena such as city population size, stock price fluctuations and personal incomes have distributions with very long right tails. The probability density function of the Pareto distribution is given by

$$f(t) = \frac{\alpha k^\alpha}{t^{\alpha+1}}, \quad k \le t \le \infty. \tag{1.26}$$

The mean, variance and reliability of the Pareto distribution are:

$$\text{Mean} = k/(\alpha - 1) \quad \text{for } > 1;$$
$$\text{Variance} = \alpha k^2/[(\alpha - 1)^2(\alpha - 2)] \quad \text{for } \alpha > 2;$$
$$\text{Reliability} = \left(\frac{k}{t}\right)^\alpha.$$

The Pareto and log normal distributions are commonly used to model population size and economical incomes. The Pareto is used to fit the tail of the distribution, and the log normal is used to fit the rest of the distribution.

### Rayleigh Distribution

The Rayleigh model is a flexible lifetime model that can apply to many degradation process failure modes. The Rayleigh probability density function is

$$f(t) = \frac{t}{\sigma^2} \exp\left(\frac{-t^2}{2\sigma^2}\right). \tag{1.27}$$

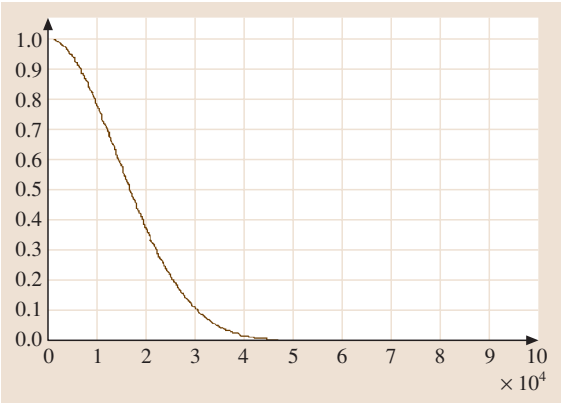


**Fig. 1.3** Rayleigh reliability function versus time

The mean, variance and reliability of the Rayleigh function are:

$$\text{Mean} = \sigma\left(\frac{\pi}{2}\right)^{\frac{1}{2}};$$
$$\text{Variance} = \left(2 - \frac{\pi}{2}\right)\sigma^2;$$
$$\text{Reliability} = \mathrm{e}^{\frac{-\sigma t^2}{2}}.$$

*Example 1.8:* Rolling resistance is a measure of the energy lost by a tire under load when it resists the force opposing its direction of travel. In a typical car traveling at sixty miles per hour, about 20% of the engine power is used to overcome the rolling resistance of the tires. A tire manufacturer introduces a new material that, when added to the tire rubber compound, significantly improves the tire rolling resistance but increases the wear rate of the tire tread. Analysis of a laboratory test of 150 tires shows that the failure rate of the new tire increases linearly with time ($h$). This is expressed as

$$h(t) = 0.5 \times 10^{-8} t.$$

The reliability of the tire after one year (8760 h) of use is

$$R(1\text{ y}) = \mathrm{e}^{-\frac{0.5}{2} \times 10^{-8} \times (8760)^2} = 0.8254.$$

Figure 1.3 shows the resulting reliability function.

### Vtub-Shaped Hazard Rate Distribution

*Pham* recently developed a two-parameter lifetime distribution with a Vtub-shaped hazard rate, known as

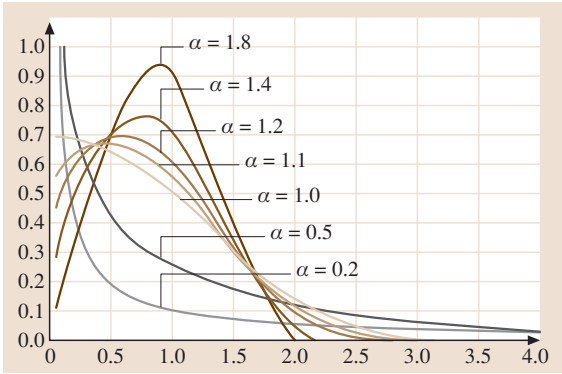


**Fig. 1.4** Probability density function for various values of $\alpha$ with $a = 2$

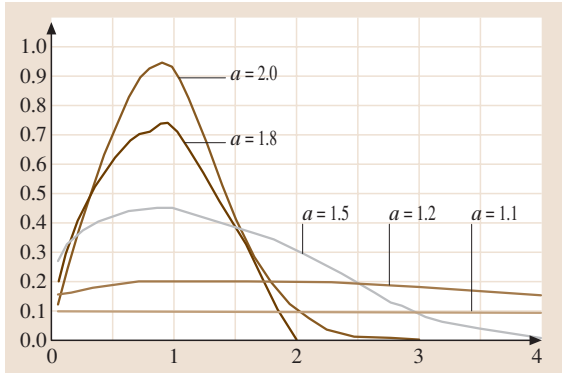


**Fig. 1.5** Probability density function for various values of $a$ with $\alpha = 1.5$

a *loglog distribution with a Vtub-shaped hazard rate* or a *Pham distribution* for short [1.3].

Note that the loglog distribution with a Vtub-shaped hazard rate and the Weibull distribution with bathtub-shaped failure rates are not the same. For the bathtub-shaped failure rate, after an initial "infant mortality period", the useful life of the system begins. During its useful life, the system fails at a constant rate. This period is then followed by a wearout period during which the system failure rate slowly increases with the onset of wearout. For the Vtub-shaped, after the infant mortality period, the system experiences a relatively low but increasing failure rate. The failure rate increases due to aging.

The Pham probability density function is given as follows [1.3]:

$$f(t) = \alpha \ln a t^{\alpha-1} a^{t^\alpha} \mathrm{e}^{1-a^{t^\alpha}}$$
$$\text{for } t > 0,\ a > 0,\ \alpha > 0. \tag{1.28}$$

The Pham distribution and reliability functions are

$$F(t) = \int_0^t f(x)\,\mathrm{d}x = 1 - \mathrm{e}^{1-a^{t^\alpha}}$$

and

$$R(t) = \mathrm{e}^{1-a^{t^\alpha}}, \tag{1.29}$$

respectively. The corresponding failure rate of the Pham distribution is given by

$$h(t) = \alpha \ln a t^{\alpha-1} a^{t^\alpha}. \tag{1.30}$$

Figures 1.4 and 1.5 describe the density functions and failure rate functions for various values of $a$ and $\alpha$.

### Two-Parameter Hazard Rate Function

This is a two-parameter function that can have increasing and decreasing hazard rates. The hazard rate $h(t)$, the reliability function $R(t)$ and the pdf are, respectively, given as follows

$$h(t) = \frac{n\lambda t^{n-1}}{\lambda t^n + 1} \quad \text{for } n \geq 1,\ \lambda > 0,\ t \geq 0, \tag{1.31}$$
$$R(t) = \mathrm{e}^{-\ln(\lambda t^N + 1)}$$

and

$$f(t) = \frac{n\lambda t^{n-1}}{\lambda t^n + 1} \mathrm{e}^{-\ln(\lambda t^n + 1)}, \quad n \geq 1,\ \lambda > 0,\ t \geq 0,$$

where $n$ = shape parameter; $\lambda$ = scale parameter.

### Three-Parameter Hazard Rate Function

This is a three-parameter distribution that can have increasing and decreasing hazard rates. The hazard rate $h(t)$ is given as

$$h(t) = \frac{\lambda(b+1)[\ln(\lambda t + \alpha)]^b}{(\lambda t + \alpha)},$$
$$b \geq 0,\ \lambda > 0,\ \alpha \geq 0,\ t \geq 0. \tag{1.32}$$

The reliability function $R(t)$ for $\alpha = 1$ is

$$R(t) = \mathrm{e}^{-[\ln(\lambda t + \alpha)]^{b+1}}.$$

The probability density function $f(t)$ is

$$f(t) = \mathrm{e}^{-[\ln(\lambda t + \alpha)]^{b+1}} \frac{\lambda(b+1)[\ln(\lambda t + \alpha)]^b}{(\lambda t + \alpha)},$$

where $b =$ shape parameter; $\lambda =$ scale parameter, and $\alpha =$ location parameter.

### Extreme-Value Distribution

The extreme-value distribution can be used to model external events such as floods, tornadoes, hurricanes and high winds in risk applications. The cdf of this distribution is given by

$$F(t) = \mathrm{e}^{-\mathrm{e}^{y}} \quad \text{for } -\infty < t < \infty. \tag{1.33}$$

### Cauchy Distribution

The Cauchy distribution can be applied when analyzing communication systems where two signals are received and one is interested in modeling the ratio of the two signals. The Cauchy probability density function is given by

$$f(t) = \frac{1}{\pi(1+t^2)} \quad \text{for } -\infty < t < \infty. \tag{1.34}$$

It is worth noting that the ratio of two standard normal random variables is a random variable with a Cauchy distribution.

## 1.3 Statistical Inference and Estimation

The problem of "point estimation" is that of estimating the parameters of a population, such as $\lambda$ or $\theta$ from an exponential, $\mu$ and $\sigma^2$ from a normal, etc. It is assumed that the type of population distribution involved is known, but the distribution parameters are unknown and they must be estimated using collected failure data. This section is devoted to the theory of estimation and discusses several common estimation techniques, such as maximum likelihood, method of moments, least squared estimation, and Bayesian methods. We also discuss confidence interval and tolerance limit estimation. For example, assume that $n$ independent samples are drawn from the exponential density function $f(x;\lambda) = \lambda\,\mathrm{e}^{-\lambda x}$ for $x > 0$ and $\lambda > 0$. Then the joint probability density function (pdf) or sample density (for short) is given by

$$f(x_1,\lambda)\cdot f(x_1,\lambda)\cdots f(x_1,\lambda) = \lambda^n\,\mathrm{e}^{-\lambda\sum_{i-1}^{n} x_i}. \tag{1.35}$$

The problem here is to find a "good" point estimate of $\lambda$, which is denoted by $\hat{\lambda}$. In other words, we want to find a function $h(X_1, X_2, \ldots, X_n)$ such that, if $x_1, x_2, \ldots, x_n$ are the observed experimental values of $X_1, X_2, \ldots, X_n$, the value $h(x_1, x_2, \ldots, x_n)$ will be a good point estimate of $\lambda$. By "good", we mean that it possesses the following properties:

- unbiasedness,
- consistency,
- efficiency (minimum variance),
- sufficiency.

In other words, if $\hat{\lambda}$ is a good point estimate of $\lambda$, then one can select a function $h(X_1, X_2, \ldots, X_n)$ where $h(X_1, X_2, \ldots, X_n)$ is an unbiased estimator of $\lambda$ and the variance of $h(X_1, X_2, \ldots, X_n)$ is a minimum. We will now present the following definitions.

*Unbiasedness.* For a given positive integer $n$, the statistic $Y = h(X_1, X_2, \ldots, X_n)$ is called an unbiased estimator of the parameter $\theta$ if the expectation of $Y$ is equal to a parameter $\theta$; that is,

$$E(Y) = \theta.$$

*Consistency.* The statistic $Y$ is called a consistent estimator of the parameter $\theta$ if $Y$ converges stochastically to a parameter $\theta$ as $n$ approaches infinity. If $\varepsilon$ is an arbitrarily small positive number when $Y$ is consistent, then

$$\lim_{n\to\infty} P(|Y-\theta| \le \varepsilon) = 1.$$

*Minimum Variance.* The statistic $Y$ is called the minimum variance unbiased estimator of the parameter $\theta$ if $Y$ is unbiased and the variance of $Y$ is less than or equal to the variance of every other unbiased estimator of $\theta$. An estimator that has the property of minimum variance in large samples is said to be efficient.

*Sufficiency.* The statistic $Y$ is said to be sufficient for $\theta$ if the conditional distribution of $X$, given that $Y = y$, is independent of $\theta$.

This is useful when determining a lower bound on the variance of all unbiased estimators. We now establish a lower bound inequality known as the *Cramér–Rao inequality*.

*Cramér–Rao Inequality.* Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution with pdf $f(x;\theta)$ for $\theta_1 < \theta < \theta_2$, where $\theta_1$ and $\theta_2$ are known. Let $Y =$

$h(X_1, X_2, \ldots, X_n)$ be an unbiased estimator of $\theta$. The lower bound inequality on the variance of $Y$, $\mathrm{Var}(Y)$, is given by

$$\mathrm{Var}(Y) \geq \frac{1}{nE\left[\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2\right]} = \frac{1}{-nE\left(\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2}\right)}. \tag{1.36}$$

*Theorem 1.1*

An estimator $\hat{\theta}$ is said to be asymptotically efficient if $\sqrt{n}\hat{\theta}$ has a variance that approaches the Cramér–Rao lower bound for large $n$; that is,

$$\lim_{n\to\infty} \mathrm{Var}(\sqrt{n}\hat{\theta}) = \frac{1}{-nE\left(\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2}\right)}. \tag{1.37}$$

### 1.3.1 Parameter Estimation

We now discuss some basic methods of parameter estimation, including the method of maximum likelihood estimation (MLE) and the method of moments. The assumption that the sample is representative of the population will be made both in the example and in later discussions.

#### Maximum Likelihood Estimation Method

In general, one deals with a sample density

$$f(x_1, x_2, \ldots, x_n) = f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta),$$

where $x_1, x_2, \ldots, x_n$ are random, independent observations of a population with density function $f(x)$.

For the general case, we would like to find an estimate or estimates, $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$ (if such exist), where

$$f(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots, \theta_m) > f(x_1, x_2, \ldots, x_n; \theta_1', \theta_2', \ldots, \theta_m').$$

The notation $\theta_1', \theta_2', \ldots, \theta_n'$ refers to any other estimates different to $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$.

Consider a random sample $X_1, X_2, \ldots, X_n$ from a distribution with a pdf $f(x;\theta)$. This distribution has a vector $\theta = (\theta_1, \theta_2, \ldots, \theta_m)'$ of unknown parameters associated with it, where $m$ is the number of unknown parameters. Assuming that the random variables are independent, then the likelihood function, $L(X;\theta)$, is the product of the probability density function evaluated at each sample point

$$L(X,\theta) = \prod_{i=1}^{n} f(X_i;\theta), \tag{1.38}$$

where $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$. The maximum likelihood estimator $\hat{\theta}$ is found by maximizing $L(\boldsymbol{X};\theta)$ with respect to $\theta$. In practice, it is often easier to maximize $\ln[L(\boldsymbol{X};\theta)]$ in order to find the vector of MLEs, which is valid because the logarithmic function is monotonic. The log likelihood function is given by

$$\ln L(X,\theta) = \sum_{i=1}^{n} \ln f(X_i;\theta) \tag{1.39}$$

and is asymptotically normally distributed since it consists of the sum of $n$ independent variables and the central limit theorem is implied. Since $L(X;\theta)$ is a joint probability density function for $X_1, X_2, \ldots, X_n$, its integral must be 1; that is,

$$\int_0^\infty \int_0^\infty \cdots \int_0^\infty L(X;\theta)\,\mathrm{d}X = 1.$$

Assuming that the likelihood is continuous, the partial derivative of the left-hand side with respect to one of the parameters, $\theta_i$, yields

$$\begin{aligned}
&\frac{\partial}{\partial\theta_i}\int_0^\infty \int_0^\infty \cdots \int_0^\infty L(X;\theta)\,\mathrm{d}X \\
&= \int_0^\infty \int_0^\infty \cdots \int_0^\infty \frac{\partial}{\partial\theta_i} L(X;\theta)\,\mathrm{d}X \\
&= \int_0^\infty \int_0^\infty \cdots \int_0^\infty \frac{\partial \log L\,(X;\theta)}{\partial\theta_i} L(X;\theta)\,\mathrm{d}X \\
&= E\left(\frac{\partial \log L\,(X;\theta)}{\partial\theta_i}\right) \\
&= E[U_i(\theta)] \quad \text{for } i = 1, 2, \ldots, m,
\end{aligned}$$

where $\boldsymbol{U}(\theta) = [U_1(\theta), U_2(\theta), \ldots U_n(\theta)]'$ is often called the score vector, and the vector $\boldsymbol{U}(\theta)$ has components

$$U_i(\theta) = \frac{\partial[\log L\,(X;\theta)]}{\partial\theta_i} \quad \text{for } i = 1, 2, \ldots, m \tag{1.40}$$

which, when equated to zero and solved, yields the MLE vector $\boldsymbol{\theta}$.

Suppose that we can obtain a nontrivial function of $X_1, X_2, \ldots, X_n$, say $h(X_1, X_2, \ldots, X_n)$, such that, when $\theta$ is replaced by $h(X_1, X_2, \ldots, X_n)$, the likelihood function $L$ will achieve a maximum. In other words,

$$L[X, h(X)] \geq L(X,\theta)$$

for every $\theta$. The statistic $h(X_1, X_2, \ldots, X_n)$ is called a maximum likelihood estimator of $\theta$ and will be denoted as

$$\hat{\theta} = h(x_1, x_2, \ldots, x_n). \tag{1.41}$$

The observed value of $\hat{\theta}$ is called the MLE of $\theta$. In general, the mechanics for obtaining the MLE can be obtained as follows:

*Step 1.* Find the joint density function $L(X, \theta)$
*Step 2.* Take the natural log of the density $\ln L$
*Step 3.* Find the partial derivatives of $\ln L$ with respect to each parameter
*Step 4.* Set these partial derivatives to "zero"
*Step 5.* Solve for parameter(s).

*Example 1.9:* Let $X_1, X_2, \ldots, X_n$, denote a random sample from the normal distribution $N(\mu, \sigma^2)$. Then the likelihood function is given by

$$L(X, \mu, \sigma^2) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \frac{1}{\sigma^n} \mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

and

$$\ln L = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Thus, we have

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0,$$
$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} - \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0.$$

Solving the two equations simultaneously, we obtain

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n},$$
$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

Note that the MLEs, if they exist, are both sufficient and efficient estimates. They also have an additional property called invariance – in other words, for an MLE of $\theta$, $\mu(\theta)$ is the MLE of $\mu(\theta)$. However, they are not necessarily unbiased (i. e., $E(\hat{\theta}) = \theta$). In fact, the point is that

$$E(\hat{\sigma}^2) = \left(\frac{n-1}{n}\right)\sigma^2 \neq \sigma^2.$$

Therefore, for small $n$, $\sigma^2$ is usually adjusted to account for this bias, and the best estimate of $\sigma^2$ is

$$\hat{\sigma}^2 = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

Sometimes it is difficult, if not impossible, to obtain maximum likelihood estimators in a closed form, and therefore numerical methods must be used to maximize the likelihood function.

*Example 1.10:* Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from the Weibull distribution with pdf

$$f(x, \alpha, \lambda) = \alpha\lambda x^{\alpha-1}\mathrm{e}^{-\lambda x^{\alpha}}. \tag{1.42}$$

The likelihood function is

$$L(X, \alpha, \lambda) = \alpha^n \lambda^n \prod_{i=1}^{n} x_i^{\alpha-1} \mathrm{e}^{-\lambda\sum_{i=1}^{n} x_i^{\alpha}}.$$

Then

$$\ln L = n\log\alpha + n\log\lambda + (\alpha-1)\sum_{i=1}^{n}\log x_i - \lambda\sum_{i=1}^{n} x_i^{\alpha},$$
$$\frac{\partial \ln L}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n}\log x_i - \lambda\sum_{i=1}^{n} x_i^{\alpha}\log x_i = 0,$$
$$\frac{\partial \ln L}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i^{\alpha} = 0.$$

As noted, solutions of the above two equations for $\alpha$ and $\lambda$ are extremely difficult to obtain and require the application of either graphical or numerical methods. It is sometimes desirable to use a quick method of estimation, which leads to a discussion of the method of moments.

### Method of Moments

Here one simply sets the sample moments equal to the corresponding population moments. For example, for the gamma distribution, the mean and the variance of

the distribution are, respectively, $\frac{\alpha}{\beta}$ and $\frac{\alpha}{\beta^2}$. Therefore, one has the following two equations in two unknowns:

$$\bar{X} = \frac{\alpha}{\beta},$$
$$S^2 = \frac{\alpha}{\beta^2}.$$

Solving these two equations simultaneously, we obtain

$$\alpha = \frac{\bar{X}^2}{S^2},$$
$$\beta = \frac{\bar{X}}{S^2}.$$

### 1.3.2 Maximum Likelihood Estimation with Censored Data

Censored data arises when we monitor for a random variable of interest – unit failure, for example – but the monitoring is stopped before measurements are complete (i. e. before the unit fails). In other words, censored observation contains only partial information about the random variable of interest. In this section, we consider two types of censoring. The first type of censoring is called Type I censoring, where the event is only observed if it occurs prior to some prespecified time. The second type of censoring is Type II censoring, in which the study continues until the failure of the first $r$ units (or components), where $r$ is some predetermined integer ($r < n$).

Examples of Type II censoring are often used when testing equipment life. Here our items are tested at the same time, and the test is terminated when $r$ of the $n$ items have failed. These approaches may save time and resources because it may take a very long time for all of the items to fail. Both Type I and Type II censoring arise in many reliability applications.

For example, let's say that we have a batch of transistors or tubes. We begin to test them all at $t = 0$, and record their times to failure. Some transistors may take a long time to burn out, and we will not want to wait that long to end the experiment. We might stop the experiment at a prespecified time $t_c$, in which case we have Type I censoring. On the other hand, we might not know what fixed value to use for the censoring time beforehand, so we decide to wait until a prespecified number of units have failed, $r$, in which case we have Type II censoring.

Censoring times may vary from individual to individual or from application to application. We now discuss a general case known as multiple-censored data.

#### Parameter Estimate with Multiple-Censored Data

The likelihood function for multiple-censored data is given by

$$\begin{aligned} L &= f(t_{1,\mathrm{f}}, \ldots, t_{r,\mathrm{f}}, t_{1,\mathrm{s}}, \ldots, t_{m,\mathrm{s}}) \\ &= C \prod_{i=1}^{r} f(t_{i,\mathrm{f}}) \prod_{j=1}^{m} [1 - F(t_{j,\mathrm{s}})], \end{aligned} \tag{1.43}$$

where $C$ is a constant, $f(.)$ is the density function and $F(.)$ is the distribution function. There are $r$ failures at times $t_{1,\mathrm{f}}, \ldots, t_{r,\mathrm{f}}$ and $m$ units with censoring times $t_{1,\mathrm{s}}, \ldots, t_{m,\mathrm{s}}$.

Note that we obtain Type-I censoring by simply setting $t_{i,\mathrm{f}} = t_{i,n}$ and $t_{j,\mathrm{s}} = t_0$ in the likelihood function in (1.43). The likelihood function for Type II censoring is similar to Type I censoring except $t_{j,\mathrm{s}} = t_r$ in (1.43). In other words, the likelihood function for the first $r$ observations from a sample of size $n$ drawn from the model in both Type I and Type II censoring is given by

$$L = f(t_{1,n}, \ldots, t_{r,n}) = C \prod_{i=1}^{r} f(t_{i,n}) [1 - F(t_*)]^{n-r}, \tag{1.44}$$

where $t_* = t_0$, the time of cessation of the test for Type I censoring and $t_* = t_r$, the time of the $r$th failure for Type II censoring.

*Example 1.11:* Consider a two-parameter probability density distribution with multiple-censored data and a distribution function with bathtub shaped failure rate, as given by [1.4]:

$$f(t) = \lambda\beta t^{\beta-1} \exp[t^{\beta} + \lambda(1 - \mathrm{e}^{t^{\beta}})], \; t, \lambda, \beta > 0 \tag{1.45}$$

and

$$F(t) = 1 - \exp[\lambda(1 - \mathrm{e}^{t^{\beta}})], \; t, \lambda, \beta > 0, \tag{1.46}$$

respectively.

Substituting the functions $f(t)$ and $F(t)$ into (1.45) and (1.46) into (1.44), we obtain the logarithm of the likelihood function:

$$\begin{aligned} \ln L = {} & \ln C + r \ln \lambda + r \ln \beta + \sum_{i=1}^{r} (\beta - 1) \ln t_i \\ & + (m + r)\lambda + \sum_{i=1}^{r} t_i^{\beta} - \left[ \sum_{i=1}^{r} \lambda \mathrm{e}^{t_i^{\beta}} + \sum_{j=1}^{m} \lambda \mathrm{e}^{t_j^{\beta}} \right]. \end{aligned}$$

The function $\ln L$ can be maximized by setting the partial derivative of $\ln L$ with respect to $\lambda$ and $\beta$ equal to zero,

and solving the resulting equations simultaneously for $\lambda$ and $\beta$. Therefore, we obtain

$$\frac{\partial \ln L}{\partial \lambda} = \frac{r}{\lambda} + (m+r) - \sum_{i=1}^{r} e^{t_i^{\beta}} - \sum_{j=1}^{m} e^{t_j^{\beta}} \equiv 0,$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{r}{\beta} + \sum_{i=1}^{r} \ln t_i + \sum_{i=1}^{r} t_i^{\beta} \ln t_i - \lambda \left( \sum_{i=1}^{r} e^{t_i^{\beta}} t_i^{\beta} \ln t_i + \sum_{j=1}^{m} e^{t_j^{\beta}} t_j^{\beta} \ln t_j \right) \equiv 0.$$

This implies that

$$\hat{\lambda} = \frac{r}{\left( \sum_{i=1}^{r} e^{t_i^{\hat{\beta}}} + \sum_{j=1}^{m} e^{t_j^{\hat{\beta}}} \right) - m - r} \tag{1.47}$$

and that $\hat{\beta}$ is the solution of

$$\frac{r}{\hat{\beta}} + \sum_{i=1}^{r} \ln t_i + \sum_{i=1}^{r} t_i^{\hat{\beta}} \ln t_i = \frac{r}{\left( \sum_{i=1}^{r} e^{t_i^{\hat{\beta}}} + \sum_{j=1}^{m} e^{t_j^{\hat{\beta}}} \right) - m - r} \left( \sum_{i=1}^{r} e^{t_i^{\hat{\beta}}} t_i^{\hat{\beta}} \ln t_i + \sum_{j=1}^{m} e^{t_j^{\hat{\beta}}} t_j^{\hat{\beta}} \ln t_j \right). \tag{1.48}$$

We now discuss two special cases.

*Case 1: Type I or Type II censored data*

From (1.44), the likelihood function for the first $r$ observations from a sample of size $n$ drawn from the model in both Type I and Type II censoring is

$$L = f(t_{1,n}, \ldots, t_{r,n}) = C \prod_{i=1}^{r} f(t_{i,n}) [1 - F(t_*)]^{n-r},$$

where $t_* = t_0$, the test cessation time for Type I censoring, and $t_* = t_r$, the time of the $r$th failure for Type II censoring. Equations (1.47) and (1.48) become

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{r} e^{t_i^{\hat{\beta}}} + (n-r) e^{t_*^{\hat{\beta}}} - n},$$

$$\frac{r}{\hat{\beta}} + \sum_{i=1}^{r} \ln t_i + \sum_{i=1}^{r} t_i^{\hat{\beta}} \ln t_i = \frac{r}{\sum_{i=1}^{r} e^{t_i^{\hat{\beta}}} + (n-r) e^{t_*^{\hat{\beta}}} - n} \times \left( \sum_{i=1}^{r} e^{t_i^{\hat{\beta}}} t_i^{\hat{\beta}} \ln t_i + \sum_{j=1}^{m} e^{t_j^{\hat{\beta}}} t_j^{\hat{\beta}} \ln t_j \right)$$

*Case 2: Complete censored data*

Simply replace $r$ with $n$ in (1.47) and (1.48) and ignore the $t_j$ portions. The maximum likelihood equations for the $\lambda$ and $\beta$ are given by

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} e^{t_i^{\hat{\beta}}} - n},$$

$$\frac{n}{\hat{\beta}} + \sum_{i=1}^{n} \ln t_i + \sum_{i=1}^{n} t_i^{\hat{\beta}} \ln t_i = \frac{n}{\sum_{i=1}^{n} e^{t_i^{\hat{\beta}}} - n} \times \sum_{i=1}^{n} e^{t_i^{\hat{\beta}}} t_i^{\hat{\beta}} \ln t_i .$$

## Confidence Intervals of Estimates

The asymptotic variance–covariance matrix for the parameters ($\lambda$ and $\beta$) is obtained by inverting the Fisher information matrix

$$I_{ij} = E\left( -\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} \right), \quad i, j = 1, 2, \tag{1.49}$$

where $\theta_1, \theta_2 = \lambda$ or $\beta$ [1.5]. This leads to

$$\begin{pmatrix} \mathrm{Var}(\hat{\lambda}) & \mathrm{Cov}(\hat{\lambda}, \hat{\beta}) \\ \mathrm{Cov}(\hat{\lambda}, \hat{\beta}) & \mathrm{Var}(\hat{\beta}) \end{pmatrix} = \begin{pmatrix} E\left( -\frac{\partial^2 \ln L}{\partial^2 \lambda} |_{\hat{\lambda}, \hat{\beta}} \right) & E\left( -\frac{\partial^2 \ln L}{\partial \lambda \partial \beta} |_{\hat{\lambda}, \hat{\beta}} \right) \\ E\left( -\frac{\partial^2 \ln L}{\partial \beta \partial \lambda} |_{\hat{\lambda}, \hat{\beta}} \right) & E\left( -\frac{\partial^2 \ln L}{\partial^2 \beta} |_{\hat{\lambda}, \hat{\beta}} \right) \end{pmatrix}^{-1}. \tag{1.50}$$

We can obtain approximate $(1-\alpha)100\%$ confidence intervals for the parameters $\lambda$ and $\beta$ based on the asymptotic normality of the MLEs [1.5] as:

$$\hat{\lambda} \pm Z_{\alpha/2} \sqrt{\mathrm{Var}(\hat{\lambda})} \quad \text{and} \quad \hat{\beta} \pm Z_{\alpha/2} \sqrt{\mathrm{Var}(\hat{\beta})}, \tag{1.51}$$

where $Z_{\alpha/2}$ is the upper percentile of the standard normal distribution.

*Application 1.* Consider the lifetime of a part from a helicopter's main rotor blade. Data on lifetime of the part taken a system database collected from October 1995 to September 1999 [1.3] are shown in Table 1.2. In this application, we consider several distribution functions for this data, including Weibull, log normal, normal, and loglog distribution functions.

The Pham pdf with parameters $a$ and $\alpha$ is

$$f(t) = \alpha(\ln a)t^{\alpha-1}a^{t^\alpha}\,\mathrm{e}^{1-a^{t^\alpha}} \quad \text{for } t>0,\ \alpha>0,\ a>1$$

and its corresponding log likelihood function (1.39) is

$$\begin{aligned}\log L(a,\alpha) &= n\log\alpha + n\ln(\ln a)\\ &\quad + (\alpha-1)\left(\sum_{i=1}^{n}\ln t_i\right)\\ &\quad + \ln a\cdot\sum_{i=1}^{n}t_i^{\alpha} + n - \sum_{i=1}^{n}a^{t_i^{\alpha}}\,.\end{aligned}$$

We then determine the confidence intervals for parameter estimates $a$ and $\alpha$. From the above log likelihood function, we can obtain the Fisher information matrix $\mathbf{H}$ as $\mathbf{H} = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22}\end{pmatrix}$, where

$$h_{11} = E\left(-\frac{\partial^2 \log L}{\partial a^2}\right),$$

$$h_{12} = h_{21} = E\left(-\frac{\partial^2 \log L}{\partial a\partial\alpha}\right),$$

$$h_{22} = E\left(-\frac{\partial^2 \log L}{\partial \alpha^2}\right).$$

The variance matrix $\mathbf{V}$ can be obtained as follows:

$$\mathbf{V} = (\mathbf{H})^{-1} = \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22}\end{pmatrix}. \tag{1.52}$$

The variances of $a$ and $\alpha$ are

$$\mathrm{Var}(a) = v_{11} \quad \mathrm{Var}(\alpha) = v_{22}.$$

**Table 1.2** Main rotor blade data

| Part code | Time to failure (h) | Part code | Time to failure (h) |
|---|---|---|---|
| xxx-015-001-107 | 1634.3 | xxx-015-001-107 | 403.2 |
| xxx-015-001-107 | 1100.5 | xxx-015-001-107 | 2898.5 |
| xxx-015-001-107 | 1100.5 | xxx-015-001-107 | 2869.1 |
| xxx-015-001-107 | 819.9 | xxx-015-001-107 | 26.5 |
| xxx-015-001-105 | 1398.3 | xxx-015-001-107 | 26.5 |
| xxx-015-001-107 | 1181 | xxx-015-001-107 | 3180.6 |
| xxx-015-001-107 | 128.7 | xxx-015-001-107 | 644.1 |
| xxx-015-001-107 | 1193.6 | xxx-015-001-107 | 1898.5 |
| xxx-015-001-107 | 254.1 | xxx-015-001-107 | 3318.2 |
| xxx-015-001-107 | 3078.5 | xxx-015-001-107 | 1940.1 |
| xxx-015-001-107 | 3078.5 | xxx-015-001-107 | 3318.2 |
| xxx-015-001-107 | 3078.5 | xxx-015-001-107 | 2317.3 |
| xxx-015-001-107 | 26.5 | xxx-015-001-107 | 1081.3 |
| xxx-015-001-107 | 26.5 | xxx-015-001-107 | 1953.5 |
| xxx-015-001-107 | 3265.9 | xxx-015-001-107 | 2418.5 |
| xxx-015-001-107 | 254.1 | xxx-015-001-107 | 1485.1 |
| xxx-015-001-107 | 2888.3 | xxx-015-001-107 | 2663.7 |
| xxx-015-001-107 | 2080.2 | xxx-015-001-107 | 1778.3 |
| xxx-015-001-107 | 2094.3 | xxx-015-001-107 | 1778.3 |
| xxx-015-001-107 | 2166.2 | xxx-015-001-107 | 2943.6 |
| xxx-015-001-107 | 2956.2 | xxx-015-001-107 | 2260 |
| xxx-015-001-107 | 795.5 | xxx-015-001-107 | 2299.2 |
| xxx-015-001-107 | 795.5 | xxx-015-001-107 | 1655 |
| xxx-015-001-107 | 204.5 | xxx-015-001-107 | 1683.1 |
| xxx-015-001-107 | 204.5 | xxx-015-001-107 | 1683.1 |
| xxx-015-001-107 | 1723.2 | xxx-015-001-107 | 2751.4 |

One can approximately obtain the $(1-\beta)100\%$ confidence intervals for $a$ and $\alpha$ based on the normal distribution as $\left[\hat{a}-z_{\frac{\beta}{2}}\sqrt{v_{11}}, \hat{a}+z_{\frac{\beta}{2}}\sqrt{v_{11}}\right]$ and $\left[\hat{\alpha}-z_{\frac{\beta}{2}}\sqrt{v_{22}}, \hat{\alpha}+z_{\frac{\beta}{2}}\sqrt{v_{22}}\right]$, respectively, where $v_{ij}$ is given in (1.52) and $z_\beta$ is $(1-\frac{\beta}{2})100\%$ of the standard normal distribution. Having obtained $\hat{a}$ and $\hat{\alpha}$, the MLE of the reliability function can be computed as

$$\hat{R}(t) = \mathrm{e}^{1-\hat{a}t^{\hat{\alpha}}}. \tag{1.53}$$

Let us define a partial derivative vector for reliability $R(t)$ as:

$$v[R(t)] = \left(\frac{\partial R(t)}{\partial a}\ \frac{\partial R(t)}{\partial \alpha}\right)$$

Then the variance of $R(t)$ can be obtained as

$$\mathrm{Var}\,[R(t)] = v[R(t)]\mathrm{V}\,(v[R(t)])^{\mathrm{T}},$$

where $V$ is given in (1.52).

One can approximately obtain the $(1-\beta)100\%$ confidence interval for $R(t)$ is

$$\left[\hat{R}(t) - z\beta\sqrt{\mathrm{Var}\,[R(t)]},\ \hat{R}(t) + z\beta\sqrt{\mathrm{Var}\,[R(t)]}\right].$$

The MLE parameter estimations for the loglog distribution and its corresponding parameters, based on the data set shown in Table 1.2, are:

$$\hat{\alpha} = 1.1075, \quad \mathrm{Var}(\hat{\alpha}) = 0.0162,$$
$$95\%\ \text{CI for}\ \hat{\alpha}: [0.8577, 1.3573];$$
$$\hat{a} = 1.0002, \quad \mathrm{Var}(\hat{a}) = 2.782\mathrm{e}^{-08},$$
$$95\%\ \text{CI for}\ a: [0.9998, 1.0005].$$

Similarly, the C.I. for $R(t)$ can be obtained directly using (1.53).

### 1.3.3 Statistical Change-Point Estimation Methods

The change-point problem has been widely studied in reliability applications in areas such as biological sciences, survival analysis and environmental statistics.

Assume that there is a sequence of random variables $X_1, X_2, \ldots, X_n$, that represents the inter-failure times, and that an index change-point $\tau$ exists, such that $X_1, X_2, \ldots, X_\tau$ have a common distribution $F$ with a density function $f(t)$ and $X_{\tau+1}, X_{\tau+2}, \ldots, X_n$ have a distribution $G$ with a density function $g(t)$, where $F \neq G$. Consider the following assumptions:

1. There is a finite but unknown number of units $N$ to be tested.
2. At the beginning, all of the units have the same lifetime distribution $F$. After $\tau$ failures are observed, the remaining $(N-\tau)$ items have the distribution $G$. The change-point $\tau$ is assumed unknown.
3. The sequence $\{X_1, X_2, \ldots, X_\tau\}$ is statistically independent of the sequence $\{X_{\tau+1}, X_{\tau+2}, \ldots, X_n\}$.
4. The lifetime test is performed according to the Type II censoring approach, in which the number of failures $n$ is predetermined.

Note that the total number of units to put up for testing $N$ can be determined in advance in hardware reliability testing. However, in software reliability testing, the parameter $N$ can be defined as the initial number of faults in the software, and it can be considered to be an unknown parameter. Let $T_1, T_2, \ldots, T_n$ be the arrival times for sequential failures. Then

$$\begin{aligned} T_1 &= X_1, \\ T_2 &= X_1 + X_2, \\ &\vdots \\ T_n &= X_1 + X_2 + \cdots X_n. \end{aligned} \tag{1.54}$$

The failure times $T_1, T_2, \ldots, T_\tau$ are the first $\tau$ order statistics of a sample of size $N$ from the distribution $F$. The failure times $T_{\tau+1}, T_{\tau+2}, \ldots, T_n$ are the first $(n-\tau)$ order statistics of a sample of size $(N-\tau)$ from the distribution $G$.

*Example 1.12:* The Weibull change-point model of the lifetime distributions $F$ and $G$ with parameters $(\lambda_1, \beta_1)$ and $(\lambda_2, \beta_2)$, respectively, can be expressed as

$$F(t) = 1 - \exp\left(-\lambda_1 t^{\beta_1}\right), \tag{1.55}$$

$$G(t) = 1 - \exp\left(-\lambda_2 t^{\beta_2}\right). \tag{1.56}$$

Assume that the distributions belong to parametric families $\{F(t \mid \theta_1), \theta_1 \in \Theta_1\}$ and $\{G(t \mid \theta_2), \theta_2 \in \Theta_2\}$. Assume that $T_1, T_2, \ldots, T_\tau$ are the first $\tau$ order statistics of a sample of size $N$ from the distribution $\{F(t \mid \theta_1), \theta_1 \in \Theta_1\}$ and that $T_{\tau+1}, T_{\tau+2}, \ldots, T_n$ are the first $(n-\tau)$ order statistics of a sample of size $(N-\tau)$ from the distribution $\{G(t \mid \theta_2), \theta_2 \in \Theta_2\}$, where $N$ is unknown. The log likelihood function can be expressed

as follows [1.6]:

$$
\begin{aligned}
&L(\tau, N, \theta_1, \theta_2 \mid T_1, T_2, \ldots, T_n) \\
&= \sum_{i=1}^{n} (N-i+1) + \sum_{i=1}^{\tau} f(T_i \mid \theta_1) \\
&\quad + \sum_{i=\tau+1}^{n} g(T_i \mid \theta_2) + (N-\tau)\log\left[1 - F(T_\tau \mid \theta_1)\right] \\
&\quad + (N-n)\log\left[1 - G(T_n \mid \theta_2)\right]. \qquad (1.57)
\end{aligned}
$$

If the parameter $N$ is known in which where hardware reliability is commonly considered for example, then the likelihood function is given by

$$
\begin{aligned}
&L(\tau, \theta_1, \theta_2 \mid T_1, T_2, \ldots, T_n) \\
&= \sum_{i=1}^{\tau} f(T_i \mid \theta_1) + \sum_{i=\tau+1}^{n} g(T_i \mid \theta_2) \\
&\quad + (N-\tau)\log\left[1 - F(T_\tau \mid \theta_1)\right] + (N-n) \\
&\quad \log\left[1 - G(T_n \mid \theta_2)\right].
\end{aligned}
$$

The maximum likelihood estimator (MLE) of the change-point value $\hat{\tau}$ and $(\hat{N}, \hat{\theta}_1, \hat{\theta}_2)$ can be obtained by taking partial derivatives of the log likelihood function in (1.57) with respect to the unknown parameters that maximize the function. It should be noted that there is no closed form for $\hat{\tau}$, but it can be obtained by calculating the log likelihood for each possible value of $\tau$, $1 \le \tau \le (n-1)$, and selecting the value that maximizes the log likelihood function.

### Application 2: A Software Model with a Change Point

In this application, we examine the case where the sample size $N$ is unknown. Consider a software reliability model developed by *Jelinski* and *Moranda* in 1972, often called the Jelinski–Moranda model. The assumptions of the model are as follows:

1. There are $N$ initial faults in the program.
2. A detected fault is removed instantaneously and no new fault is introduced.
3. Each failure caused by a fault occurs independently and randomly in time according to an exponential distribution.
4. The functions $F$ and $G$ are exponential distributions with failure rate parameters $\lambda_1$ and $\lambda_2$, respectively.

Based on these assumptions, the inter-failure times $X_1, X_2, \ldots, X_n$ are independently exponentially distributed. Specifically, $X_i = T_i - T_{i-1}, i = 1, 2, \ldots \tau$, are exponentially distributed with parameter $\lambda_1 (N-i+1)$, where $\lambda_1$ is the initial fault detection rate of the first $\tau$ failures, and $X_j = T_j - T_{j-1}$, $j = \tau+1, \tau+2, \ldots n$ are exponentially distributed with parameter $\lambda_2 (N-\tau-j+1)$, where $\lambda_2$ is the fault detection rate of the first $n-\tau$ failures. If $\lambda_1 = \lambda_2$, it means that each fault removal is the same and that the change-point model becomes the Jelinski–Moranda software reliability model [1.7].

The MLEs of the parameters $(\tau, N, \lambda_1, \lambda_2)$ can be obtained by solving the following equations simultaneously:

$$
\hat{\lambda}_1 = \frac{\tau}{\sum_{i=1}^{\tau} (\hat{N}-i+1) x_i}, \qquad (1.58)
$$

$$
\hat{\lambda}_2 = \frac{(n-\tau)}{\sum_{i=\tau+1}^{n} (\hat{N}-i+1) x_i}, \qquad (1.59)
$$

$$
\sum_{i=1}^{n} \frac{1}{(\hat{N}-i+1)} = \hat{\lambda}_1 \sum_{i=1}^{\tau} x_i + \hat{\lambda}_2 \sum_{i=\tau+1}^{n} x_i. \qquad (1.60)
$$

To illustrate the model, we use the data set shown in Table 1.3 to obtain the unknown parameters $(\tau, N, \lambda_1, \lambda_2)$ using (1.58)–(1.60). The data in Table 1.3 [1.8] shows the successive inter-failure times for a real-time command and control system. The table reads from left to right in rows, and the recorded times are execution times, in seconds. There are 136 failures in total. Figure 1.6 plots the log-likelihood function versus the number of failures. The MLEs of the parameters $(\tau, N, \lambda_1, \lambda_2)$ with one change point are given by

$$
\hat{\tau} = 16,\ \hat{N} = 145,\ \hat{\lambda}_1 = 1.1 \times 10^{-4},\ \hat{\lambda}_2 = 0.31 \times 10^{-4}.
$$

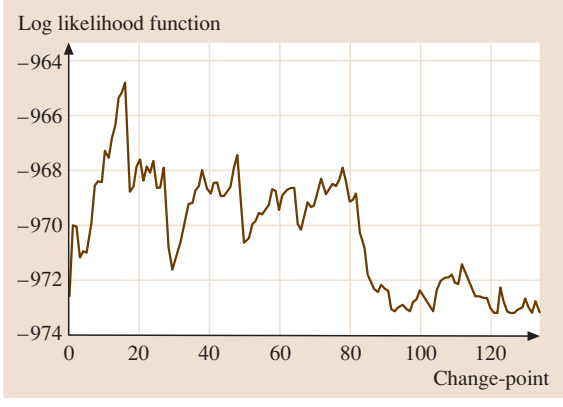


**Fig. 1.6** The log likelihood function versus the number of failures

**Table 1.3** Successive inter-failure times (in s) for a real-time command system

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 30 | 113 | 81 | 115 | 9 | 2 | 91 | 112 | 15 |
| 138 | 50 | 77 | 24 | 108 | 88 | 670 | 120 | 26 | 114 |
| 325 | 55 | 242 | 68 | 422 | 180 | 10 | 1146 | 600 | 15 |
| 36 | 4 | 0 | 8 | 227 | 65 | 176 | 58 | 457 | 300 |
| 97 | 263 | 452 | 255 | 197 | 193 | 6 | 79 | 816 | 1351 |
| 148 | 21 | 233 | 134 | 357 | 193 | 236 | 31 | 369 | 748 |
| 0 | 232 | 330 | 365 | 1222 | 543 | 10 | 16 | 529 | 379 |
| 44 | 129 | 810 | 290 | 300 | 529 | 281 | 160 | 828 | 1011 |
| 445 | 296 | 1755 | 1064 | 1783 | 860 | 983 | 707 | 33 | 868 |
| 724 | 2323 | 2930 | 1461 | 843 | 12 | 261 | 1800 | 865 | 1435 |
| 30 | 143 | 108 | 0 | 3110 | 1247 | 943 | 700 | 875 | 245 |
| 729 | 1897 | 447 | 386 | 446 | 122 | 990 | 948 | 1082 | 22 |
| 75 | 482 | 5509 | 100 | 10 | 1071 | 371 | 790 | 6150 | 3321 |
| 1045 | | 648 | 5485 | 1160 | 1864 | 4116 | | | |

If we do not consider a change point in the model, the MLEs of the parameters $N$ and $\lambda$, can be given as

$$\hat{N} = 142,\ \hat{\lambda} = 0.35 \times 10^{-4}.$$

From Fig. 1.6,

it is clear that it is worth considering change points in reliability functions.

### 1.3.4 Goodness of Fit Techniques

The problem discussed here is one of comparing an observed sample distribution with a theoretical distribution. Two common techniques that will be discussed are the $\chi^2$ goodness-of-fit test and the Kolmogorov–Smirnov "$d$" test.

#### Chi-Squared Test

The following statistic

$$\chi^2 = \sum_{i=1}^{k} \left( \frac{x_i - \mu_i}{\sigma_i} \right)^2 \tag{1.61}$$

has a chi-squared ($\chi^2$) distribution with $k$ degrees of freedom. The procedure used for the chi-squared test is:

1. Divide the sample data into mutually exclusive cells (normally 8–12) such that the range of the random variable is covered.
2. Determine the frequency, $f_i$, of sample observations in each cell.
3. Determine the theoretical frequency, $F_i$, for each cell (the area under density function between cell boundaries $X_n$ – total sample size). Note that the theoretical frequency for each cell should be greater than 1. This step normally requires estimates for the population parameters, which can be obtained from the sample data.
4. Form the statistic

$$A = \sum_{i=1}^{k} \frac{(f_i - F_i)^2}{F_i}. \tag{1.62}$$

5. From the $\chi^2$ tables, choose a value of $\chi^2$ with the desired significance level and degrees of freedom ($= k - 1 - r$, where $r$ is the number of population parameters estimated).
6. Reject the hypothesis that the sample distribution is the same as the theoretical distribution if

$$A > \chi^2_{1-\alpha,k-1-r},$$

where $\alpha$ is called the significance level.

*Example 1.13:* Given the data in Table 1.4, can the data be represented by the exponential distribution with a significance level of $\alpha$?

From the above calculation, $\hat{\lambda} = 0.002\,63$, $R_i = \mathrm{e}^{-\lambda t_i}$ and $Q_i = 1 - R_i$. Given that the significance level $\alpha$ is 0.1, from (1.62), we obtain

$$A = \sum_{i=1}^{11} \frac{(f_i - F_i)^2}{F_i} = 6.165\,.$$

From Table 1.9 in Sect. 1.A, the value of $\chi^2$ with nine degrees of freedom and $\alpha = 0.1$ is 14.68; that is,

$$\chi^2_{9,0.1} = 14.68\,.$$

Since $S = 6.165 < 14.68$, we would not reject the hypothesis of an exponential with $\lambda = 0.002\,63$.

**Table 1.4** Sample observations for each cell boundary

| Cell boundaries | $f_i$ | $Q_i = (1 - R_i)\,60$ | $F_i = Q_i - Q_{i-1}$ |
|---|---|---|---|
| 0 – 100 | 10 | 13.86 | 13.86 |
| 100 – 200 | 9 | 24.52 | 10.66 |
| 200 – 300 | 8 | 32.71 | 8.19 |
| 300 – 400 | 8 | 39.01 | 6.30 |
| 400 – 500 | 7 | 43.86 | 4.85 |
| 500 – 600 | 6 | 47.59 | 3.73 |
| 600 – 700 | 4 | 50.45 | 2.86 |
| 700 – 800 | 4 | 52.66 | 2.21 |
| 800 – 900 | 2 | 54.35 | 1.69 |
| 900 – 1000 | 1 | 55.66 | 1.31 |
| > 1000 | 1 | 58.83 | 2.17 |

If in the statistic

$$A = \sum_{i=1}^{k} \left( \frac{f_i - F_i}{\sqrt{F_i}} \right)^2, \quad \left( \frac{f_i - F_i}{\sqrt{F_i}} \right)$$

is approximately normal for large samples, then $A$ also has a $\chi^2$ distribution. This is the basis for the goodness of fit test.

### Kolmogorov–Smirnov $d$ Test

Both the $\chi^2$ and "$d$" tests are nonparametric tests. However, the $\chi^2$ test largely assumes sample normality of the observed frequency about its mean, while "$d$" assumes only a continuous distribution. Let $X1 \le X_2 \le X_3 \le \ldots \le X_n$ denote the ordered sample values. Define the observed distribution function, $F_n(x)$, as:

$$F_n(X) = \begin{cases} 0 & \text{for } x \le x_1 \\ \frac{i}{n} & \text{for } x_i < x \le x_{i+1} \\ 1 & \text{for } x > x_n \end{cases} .$$

Assume the testing hypothesis

$$H_0 : F(x) = F_0(x),$$

where $F_0(x)$ is a given continuous distribution and $F(x)$ is an unknown distribution. Let

$$d_n = \sup_{-\infty < x < \infty} |F_n(x) - F_0(x)|.$$

Since $F_0(x)$ is a continuous increasing function, we can evaluate $|F_n(x) - F_0(x)|$ for each $n$. If $d_n \le d_{n,\alpha}$, then we will not reject the hypothesis $H_0$; otherwise, we will reject it when $d_n > d_{n,\alpha}$. The value of $d_{n,\alpha}$ can be found in Table 1.10 in Sect. 1.A, where $n$ is the sample size and $a$ is the level of significance.

## 1.3.5 Least Squared Estimation

One common approach to curve fitting, which is unrelated to normal regression theory and MLE estimates of coefficients but uses identical formulae, is called the method of least squares. This method is based on minimizing the sum of the squared distances from the best fit line to the actual data points. It just so happens that finding the MLEs for the coefficients of the regression line also involves this sum of squared distances.

### Normal Linear Regression

Regression considers the distribution of one variable as a function of another when the other variable is fixed at each of several levels. In the normal bivariate case, consider the distribution of $X$ as a function of given values of $Z$ where $X = \alpha + \beta Z$. Consider a sample of $n$ observations $(x_i, z_i)$. We can obtain the likelihood and its natural log for the normal distribution as follows:

$$f(x_1, x_2, \ldots, x_n) = \frac{1}{2\pi^{\frac{n}{2}}} \left( \frac{1}{\sigma^2} \right)^{\frac{n}{2}} \mathrm{e}^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \alpha - \beta z_i)^2},$$

$$\ln L = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \alpha - \beta z_i)^2.$$

Taking the partial derivatives of $\ln L$ with respect to $\alpha$ and $\beta$, we have

$$\frac{\partial \ln L}{\partial \alpha} = \sum_{i=1}^{n} (x_i - \alpha - \beta z_i)^2 = 0,$$

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{n} z_i (x_i - \alpha - \beta z_i) = 0.$$

The solutions to the simultaneous equations are

$$\hat{\alpha} = \bar{X} - \beta \bar{Z},$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Z_i - \bar{Z})}{\sum_{i=1}^{n} (Z_i - \bar{Z})^2}. \tag{1.63}$$

### Least Squared Straight Line Fit

Assume that there is a linear relationship between $X$ and $E(Y|x)$; that is, that $E(Y|x) = a + bx$. Given a set of data, we want to estimate the coefficients $a$ and $b$

that minimize the sum of the squares. Suppose that the desired polynomial, $p(x)$, is written as

$$\sum_{i=0}^{m} a_i x^i,$$

where $a_0, a_1, \ldots, a_m$ are to be determined. The method of least squares chooses as "solutions" those coefficients that minimize the sum of the squares of the vertical ($y$) distances from the data points to the presumed polynomial. This means that the "best" polynomial is the one whose coefficients minimize the function $L$, where

$$L = \sum_{i=1}^{n} [y_i - p(x_i)]^2.$$

Here, we will only treat the linear case, where $X = \alpha + \beta Z$. The procedure for higher order polynomials is identical, although the computations become much more tedious. Assume a straight line of the form $X = \alpha + \beta Z$. For each observation $(x_i, z_i)$ : $X_i = \alpha + \beta Z_i$, let

$$Q = \sum_{i=1}^{n} (x_i - \alpha - \beta z_i)^2.$$

We wish to find estimates for $\alpha$ and $\beta$ that minimize $Q$. Taking partial differentials, we obtain

$$\frac{\partial Q}{\partial \alpha} = -2 \sum_{i=1}^{n} (x_i - \alpha - \beta z_i) = 0,$$

$$\frac{\partial Q}{\partial \beta} = -2 \sum_{i=1}^{n} z_i (x_i - \alpha - \beta z_i) = 0.$$

Note that the above are the same as the MLE equations for normal linear regression. Therefore, we obtain the following results:

$$\hat{\alpha} = \bar{x} - \beta \bar{z},$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(z_i - \bar{z})}{\sum_{i=1}^{n} (z_i - \bar{z})^2}. \qquad (1.64)$$

The above gives an example of least squares applied to a linear case. The same pattern applies for higher-order curves with 3, 4 and so on solutions.

### 1.3.6 Interval Estimation

A point estimate is sometimes inadequate at providing an estimate for an unknown parameter, since it rarely coincides with the true value of the parameter. An alternative way is to obtain a confidence interval estimation of the form $[\theta_L, \theta_U]$ where $\theta_L$ is the lower bound and $\theta_U$ is the upper bound.

Point estimates can become more useful if some measure of their error is given; in other words, if some kind of tolerance for their high and low values is given. Thus, if an interval estimator is $[\theta_L, \theta_U]$ with a given probability $(1-\alpha)$, then $\theta_L$ and $\theta_U$ are called the $100\,(1-\alpha)\,\%$ confidence limits for the given parameter $\theta$, and the interval between them is a $100\,(1-\alpha)\,\%$ confidence interval, while $(1-\alpha)$ is called the confidence coefficient.

#### Confidence Intervals for Normal Parameters

The one-dimensional normal distribution has two parameters: mean $\mu$ and variance $\sigma^2$. The simultaneous employment of both parameters in a confidence statement concerning percentages of the population will be discussed in the next section on tolerance limits. Hence, individual confidence statements about $\mu$ and $\sigma^2$ will be discussed here.

*Confidence Limits for a Mean $\mu$ with a Known $\sigma^2$.* It is easy to show that the statistic

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

is a standard normal distribution, where

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Hence, a $100\,(1-\alpha)\,\%$ confidence interval for the mean $\mu$ is given by

$$P\left[\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha. \qquad (1.65)$$

In other words,

$$\mu_L = \bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad \text{and} \quad \mu_U = \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}.$$

*Example 1.14:* Draw a sample of size 4 from a normal distribution with a known variance = 9, say $x_1 = 2$, $x_2 = 3$, $x_3 = 5$, $x_4 = 2$. Determine the location of the true mean ($\mu$). The sample mean can be calculated as

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2+3+5+2}{4} = 3.$$

Assuming that $\alpha = 0.05$ and, from the standard normal distribution (Table 1.6 in Sect. 1.A), $Z_{0.025} = 1.96$, then we obtain

$$P\left[3 - 1.96\frac{3}{\sqrt{4}} < \mu < 3 + 1.96\frac{3}{\sqrt{4}}\right] = 0.95,$$
$$P[0.06 < \mu < 5.94] = 0.95.$$

This example shows that there is a 95% probability that the true mean is somewhere between 0.06 and 5.94. Now, $\mu$ is a fixed parameter and does not vary, so how do we interpret the probability? If samples of size 4 were to be repeatedly drawn, a different set of limits would be constructed each time. If this is the case, the interval becomes the random variable and the interpretation is that, for 95% of the time, the interval constructed in this way will contain the true (fixed) parameter.

*Confidence Limits for a Mean $\mu$ with an Unknown $\sigma^2$.* Let

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}. \tag{1.66}$$

It can be shown that the statistic

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

has a $t$ distribution with $(n-1)$ degrees of freedom (see Table 1.7 in Appendix A). Thus, for a given sample mean and sample standard deviation, we obtain

$$P\left[|T| < t_{\frac{\alpha}{2},n-1}\right] = 1 - \alpha.$$

Hence, a $100\,(1-\alpha)\,\%$ confidence interval for the mean $\mu$ is given by

$$p\left[\bar{X} - t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}\right]$$
$$= 1 - \alpha. \tag{1.67}$$

*Example 1.15:* The variability of a new product was investigated. An experiment was run using a sample of size $n = 25$; the sample mean was found to be $\bar{X} = 50$ and the variance $\sigma^2 = 16$. From Table 1.7 in Appendix A, $t_{\frac{\alpha}{2},n-1} = t_{0.025,24} = 2.064$. The 95% confidence limit for $\mu$ is given by

$$P\left[50 - 2.064\sqrt{\frac{16}{25}} < \mu\right.$$
$$\left. < 50 + 2.064\sqrt{\frac{16}{25}}\right] = 0.95\,,$$
$$P[48.349 < \mu < 51.651]x = 0.95\,.$$

Note that, for one-sided limits, one should choose $t_{\alpha}$, or $t_{1-\alpha}$.

*Confidence Limits on $\sigma^2$.* Note that $n\hat{\sigma}^2/\sigma^2$ has a $\chi^2$ distribution with $(n-1)$ degrees of freedom. Correcting for the bias in $\hat{\sigma}^2$, it is clear that $(n-1)\hat{\sigma}^2/\sigma^2$ has this same distribution. Hence,

$$P\left[\chi^2_{\frac{\alpha}{2},n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{1-\frac{\alpha}{2},n-1}\right] = 1 - \alpha$$

or

$$P\left[\frac{\sum(x_i - \bar{x})^2}{\chi^2_{1-\frac{\alpha}{2},n-1}} < \sigma^2 < \frac{\sum(x_i - \bar{x})^2}{\chi^2_{\frac{\alpha}{2},n-1}}\right] = 1 - \alpha. \tag{1.68}$$

Again, for one-sided limits, one should choose $\chi^2(\alpha)$ or $\chi^2(1-\alpha)$.

### Confidence Intervals for Exponential Parameters

The pdf and cdf for the exponential distribution are

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

and

$$F(x) = 1 - \mathrm{e}^{-\lambda x},$$

respectively. It was shown that the distribution of a function of the estimate

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{n} x_i + (n-r)x_r} \tag{1.69}$$

derived from a test of $n$ identical components with common exponential failure density (failure rate $\lambda$), whose testing was stopped after the $r$th failure, was chi-squared ($\chi^2$), i. e.,

$$2r\frac{\lambda}{\hat{\lambda}} = 2\lambda T$$
($\chi^2$ distribution with $2r$ degrees of freedom),

where $T$ is the total time accrued by all units. Knowing the distribution of $2\lambda T$ allows us to obtain the confidence limits on the parameter as:

$$P\left[\chi^2_{1-\frac{\alpha}{2},2r} < 2\lambda T < \chi^2_{\frac{\alpha}{2},2r}\right] = 1 - \alpha$$

or, equivalently, that

$$P\left[\frac{\chi^2_{1-\frac{\alpha}{2},2r}}{2T} < \lambda < \frac{\chi^2_{\frac{\alpha}{2},2r}}{2T}\right] = 1 - \alpha.$$

**Table 1.5** Confidence limits for $\theta$

| Confidence limits | Fixed number of failures | Fixed time |
|---|---|---|
| One-sided lower limit | $\frac{2T}{\chi^2_{\alpha,2r}}$ | $\frac{2T}{\chi^2_{\alpha,2r+2}}$ |
| One-sided upper limit | $\frac{2T}{\chi^2_{1-\alpha,2r}}$ | $\frac{2T}{\chi^2_{1-\alpha,2r}}$ |
| Two-sided limits | $\frac{2T}{\chi^2_{\alpha/2,2r}}, \frac{2T}{\chi^2_{1-\alpha/2,2r}}$ | $\frac{2T}{\chi^2_{\alpha/2,2r+2}}, \frac{2T}{\chi^2_{1-\alpha/2,2r}}$ |

This means that in $(1-\alpha)\%$ of the samples of a given size $n$, the random interval

$$\left(\frac{\chi^2_{1-\frac{\alpha}{2},2r}}{2T}, \frac{\chi^2_{\frac{\alpha}{2},2r}}{2T}\right)$$

will contain a population of constant failure rate. For $\theta = 1/\lambda$ or the MTBF, the above confidence limits change to

$$P\left[\frac{2T}{\chi^2_{\frac{\alpha}{2},2r}} < \theta < \frac{2T}{\chi^2_{1-\frac{\alpha}{2},2r}}\right] = 1-\alpha.$$

If testing is stopped at a fixed time rather than a fixed number of failures, the number of degrees of freedom in the lower limit increases by two. Table 1.5 shows the confidence limits for $\theta$, the mean of the exponential density.

### Confidence Intervals for Binomial Parameters

Consider a sequence of $n$ Bernoulli trials with $k$ successes and $(n-k)$ failures. We now determine one-sided upper and lower and two-sided limits on the parameter $p$, the probability of success. For the lower limit, the binomial sum is set up such that the probability of $k$ or more successes with a true $p$ as low as $p_L$ is only $\alpha/2$. This means that the probability of $k$ or more successes with a true $p$ higher than $p_L$ is $\left(1-\frac{\alpha}{2}\right)$.

$$\sum_{i=k}^{n}\binom{n}{i}p_L^i(1-p_L)^{n-i} = \frac{\alpha}{2}.$$

Similarly, the binomial sum for the upper limit is

$$\sum_{i=k}^{n}\binom{n}{i}p_U^i(1-p_U)^{n-i} = 1-\frac{\alpha}{2}$$

or, equivalently,

$$\sum_{i=0}^{k-1}\binom{n}{i}p_U^i(1-p_U)^{n-i} = \frac{\alpha}{2}.$$

Solving for $p_L$ and $p_U$ in the above equations,

$$P[p_L < p < p_U] = 1-\alpha.$$

For one-sided limits, merely change $\alpha/2$ to $\alpha$.

*Example 1.16:* Given $n = 100$ with 25 successes, and 75 failures, an 80% two-sided confidence limit on $p$ can be obtained as follows:

$$\sum_{i=25}^{100}\binom{100}{i}p_L^i(1-p_L)^{100-i} = 0.10,$$

$$\sum_{i=0}^{24}\binom{100}{i}p_U^i(1-p_U)^{100-i} = 0.10.$$

Solving the above two equations simultaneously, we obtain

$$p_L \approx 0.194 \quad \text{and} \quad p_U \approx 0.313,$$

$$P[0.194 < p < 0.313] = 0.80.$$

Continuing with Example 1.16 above, we now find an 80% one-sided confidence limit on $p$.

We start by setting the top equation to 0.20 and solving for $p_L$. It is then easy to obtain $p_L = 0.211$ and $P[p > 0.211] = 0.80$.

Let us now define $\bar{p} = k/n$, the number of successes divided by the number of trials. For large values of $n$ and if $np > 5$ and $n(1-p) > 5$, and from the central limit theorem [1.9], the statistic

$$Z = \frac{(\bar{p}-p)}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}}$$

approximates to the standard normal distribution. Hence

$$P[-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}] = 1-\alpha.$$

Then

$$P\left[\bar{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} < p < \bar{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right] = 1-\alpha.$$

*Example 1.17:* Find the two-sided confidence limit for $n = 900$, $k = 180$, and $\alpha = 0.05$. Then we obtain $p = 180/900 = 0.2$ and

$$P\left[0.2 - 1.96\sqrt{\frac{0.2(0.8)}{900}} < p < 0.2 + 1.96\sqrt{\frac{0.2(0.8)}{900}}\right] = 0.95,$$

$$P[0.174 < p < 0.226] = 0.95.$$

#### Confidence Intervals for Poisson Parameters

Limits for the Poisson parameters are completely analogous to those for the binomial distribution except that the sample space is denumerable instead of finite. The lower and upper limits can be solved simultaneously in the following equations:

$$\sum_{i=k}^{\infty} \frac{\lambda_L^i e^{-\lambda_L}}{i!} = \frac{\alpha}{2},$$

$$\sum_{i=k}^{\infty} \frac{\lambda_U^i e^{-\lambda_U}}{i!} = 1 - \frac{\alpha}{2},$$

or, equivalently,

$$\sum_{i=k}^{\infty} \frac{\lambda_L^i e^{-\lambda_L}}{i!} = \frac{\alpha}{2},$$

$$\sum_{i=0}^{k-1} \frac{\lambda_U^i e^{-\lambda_U}}{i!} = \frac{\alpha}{2}.$$

The one-sided limits are constructed in the same way as for binomial limits.

### 1.3.7 Nonparametric Tolerance Limits

Nonparametric tolerance limits are based on the smallest and largest observation in the sample, designated $X_S$ and $X_L$, respectively. Due to their nonparametric nature, these limits are quite insensitive, and obtaining precisions similar to the parametric methods necessitates much larger samples. An interesting question here is to determine the sample size required to include at least $100\,(1-\alpha)\,\%$ of the population between $X_S$ and $X_L$ with a given probability $y$.

For two-sided tolerance limits, if $(1-\alpha)$ is the minimum proportion of the population contained between the largest observation $X_L$ and the smallest observation $X_S$ with a confidence $(1-\gamma)$, then it can be shown that

$$n(1-\alpha)^{n-1} - (n-1)(1-\alpha)^n = \gamma.$$

Therefore, the number of observations required is given by

$$n = \left( \frac{(2-\alpha)}{4\alpha} \chi^2_{1-\gamma,4} + \frac{1}{2} \right) + 1,$$

where a value of $\chi^2_{1-\gamma,4}$ is given in Table 1.8 of Sect. 1.A.

*Example 1.18:* Determine the tolerance limits that include at least 90% of the population with probability 0.95. Here,

$$\alpha = 0.1,\ \gamma = 0.95 \quad \text{and} \quad \chi^2_{0.05,4} = 9.488.$$

Therefore, a sample of size

$$n = \left[ \frac{(2-0.1)}{4(0.1)} (9.488) + \frac{1}{2} \right] + 1 = 46$$

is required. For a one-sided tolerance limit, the number of observations required is given by

$$n = \left( \frac{\log(1-\gamma)}{\log(1-\alpha)} \right) + 1.$$

*Example 1.19:* As in Example 1.18, we wish to find a lower tolerance limit; that is, the number of observations required such that the probability is 0.95 that at least 90% of the population will exceed $X_S$. This is given by

$$n = \left( \frac{\log(1-0.95)}{\log(1-0.1)} \right) + 1 = 30.$$

One can easily generate a table containing the sample size required to include a given percentage of the population between $X_S$ and $X_L$ with a given confidence, or the sample size required to include a given percentage of the population above or below $X_S$ or $X_L$, respectively.

### 1.3.8 Sequential Sampling

A sequential sampling scheme is one in which items are drawn one at a time and the results at any stage determine whether sampling or testing should stop. Thus, any sampling procedure for which the number of observations is a random variable can be regarded as sequential sampling. Sequential tests derive their name from the fact that the sample size is not determined in advance, but allowed to "float" with a decision (accept, reject, or continue test) after each trial or data point.

In general, let us consider the hypothesis

$$H_0 : f(x) = f_0(x) \text{ versus } H_1 : f(x) = f_1(x).$$

For an observational test, say $X_1$, if $X_1 \le A$, then we will accept the testing hypothesis $[H_0 : f(x) = f_0(x)]$; if $X_1 \ge A$ we will reject $H_0$ and accept $H_I : f(x) = f_1(x)$. Otherwise, we will continue to perform at least one more test. The interval $X_1 \le A$ is called the acceptance region. The interval $X_1 \ge A$ is called the rejection or critical region.

A "good" test is one that makes the $\alpha$ and $\beta$ errors as small as possible. However, there is not much freedom to do this without increasing the sample size. A common procedure is to fix the $\beta$ error and then choose a critical region to minimize the error or to maximize the "power"

(power $= 1-\beta$) of the test, or to choose the critical region so as to equalize the $\alpha$ and $\beta$ errors to reasonable levels.

One criterion (similar to the MLE) used to construct tests is called the "probability ratio", which is the ratio of the sample densities under $H_1/H_0$. Consider the ratio of probabilities

$$\lambda_m = \frac{\prod_{i=1}^{n} f_1(x_i)}{\prod_{i=1}^{n} f_0(x_i)} > k.$$

Here, $x_1, x_2, \ldots, x_n$ are $n$ independent random observations and $k$ is chosen to give the desired a error.

Recall from the MLE discussion in Sect. 1.3.1 that $f_1(x_1), f_1(x_2), \ldots, f_1(x_n)$ are maximized under $H_1$ when the parameter(s), e.g. $\theta = \theta_1$ and similarly $f_0(x_1), f_0(x_2), \ldots, f_0(x_n)$, are maximized when $\theta = \theta_0$. Thus, the ratio will become large if the the sample favors $H_1$ and will become small if the sample favors $H_0$. Therefore, the test will be called a sequential probability ratio test if we

1. stop sampling and reject $H_0$ as soon as $\lambda_\mathrm{m} \geq A$;
2. stop sampling and accept $H_0$ as soon as $\lambda_\mathrm{m} \leq B$;
3. continue sampling as long as $B < \lambda_\mathrm{m} < A$, where $A > B$.

The selection of $A$ and $B$ using the above test, as suggested by *Wald* (see [1.9]), can be determined as follows:

$$B = \frac{\beta}{1-\alpha} \quad \text{and} \quad A = \frac{1-\beta}{\alpha}$$

The bases for $\alpha$ and $\beta$ are therefore:

$$P[\lambda_\mathrm{m} > A|H_0] = \alpha$$
$$P[\lambda_\mathrm{m} < A|H_1] = \beta$$

### 1.3.9 Bayesian Methods

The Bayesian approach to statistical inference is based on a theorem first presented by Thomas Bayes. To demonstrate the approach, let $X$ have a pdf $f(x)$, which is dependent on $\theta$. In the traditional statistical inference approach, $\theta$ is an unknown parameter, and hence is a constant. We now describe our prior supposition for the value of $\theta$ by a pdf of $h(\theta)$. This amounts to quantitatively assessing subjective judgment and should not be confused with the so-called objective probability assessment derived from the long-term frequency approach. Thus, $\theta$ will now essentially be treated as a random variable $\theta$ with a pdf of $h(\theta)$.

Consider a random sample $X_1, X_2, \ldots, X_n$ from $f(x)$ and define a statistic $Y$ as a function of this random sample. Then there exists a conditional pdf $g(y|\theta)$ of $Y$ for a given $\theta$. The joint pdf for $y$ and $\theta$ is

$$f(\theta, y) = h(\theta)g(y|\theta).$$

If $\theta$ is continuous, then

$$f_1(y) = \int_\theta h(\theta)g(y|\theta)\,\mathrm{d}\theta$$

is the marginal pdf for the statistic $y$. Given the information $y$, the conditional pdf for $\theta$ is

$$k(\theta|y) = \frac{h(\theta)g(y|\theta)}{f_1(y)} \quad \text{for } f_1(y) > 0$$
$$= \frac{h(\theta)g(y|\theta)}{\int_\theta h(\theta)g(y|\theta)\,\mathrm{d}\theta}$$

If $\theta$ is discrete, then

$$f_1(y) = \sum_k P(\theta_k)P(y|\theta_k)$$

and

$$P(\theta_i|y_i) = \frac{P(\theta_k)P(y_i|\theta_i)}{\sum_k P(\theta_k)P(y_j|\theta_k)}$$

where $P(\theta_j)$ is the prior probability of event $\theta_i$ and $P(\theta_j|y_j)$ is the posterior probability of event $y_j$ given $\theta_i$. This is simply a form of Bayes' theorem. Here, $h(\theta)$ is the prior pdf that expresses our belief about the value of $\theta$ before the data $(Y = y)$ became available. Then $k(\theta|y)$ is the posterior pdf, given the data $(Y = y)$.

Note that the difference in the shape of the prior pdf $h(\theta)$ compared to the posterior pdf $k(\theta|y)$ due to the information is a result of the product of $g(y|\theta)$ and $h(\theta)$, because $f_l(y)$ is simply a normalization constant for a fixed $y$. The idea of reliability is to take "prior" data and combine it with current data to gain a better estimate or confidence interval or test than would be possible with either on their own. As more current data is acquired, the prior data is "washed out" [1.1].

## 1.4 Stochastic Processes

Stochastic processes are used to describe the operation of a system over time. There are two main types of stochastic processes: continuous and discrete. A complex continuous process is a process describing a system transition from state to state. The simplest process that will be discussed here is a Markov process. In this case, the future behavior of the process does not depend on its past or present behavior. In many systems that arise in practice, however, past and present states of the system influence the future states, even if they do not uniquely determine them.

### 1.4.1 Markov Processes

In this section, we will discuss discrete stochastic processes. As an introduction to the Markov process, let us examine the following example.

*Example 1.20:* Consider a parallel system consisting of two components. From a reliability point of view, the states of the system can be described by

*State 1*: Full operation (both components operating);
*State 2*: One component is operating and one component has failed;
*State 3*: Both components have failed.

Define

$$P_i(t) = P[X(t) = i] = P[\text{system is in state } i \text{ at time } t]$$

and

$$P_i(t + \mathrm{d}t) = P[X(t + \mathrm{d}t) = i] = P[\text{system is in state } i \text{ at time } t + \mathrm{d}t].$$

Define a random variable $X(t)$ which can assume the values 1, 2, or 3 corresponding to the states mentioned above. Since $X(t)$ is a random variable, one can discuss $P[X(t) = 1]$, $P[X(t) = 2]$ or the conditional probability $P[X(t_1) = 2|X(t_0) = 1]$. Again, $X(t)$ is defined as a function of time $t$, while the conditional probability $P[X(t_1) = 2|X(t_0) = 1]$ can be interpreted as the probability of being in state 2 at time $t_1$, given that the system was in state 1 at time $t_0$. In this example, the "stage space" is discrete, i. e., 1, 2, 3, etc., and the parameter space (time) is continuous. The simple process described above is called a stochastic process: a process that develops over time (or space) in accordance with some probabilistic (stochastic) laws. There are many types of stochastic processes.

Here we emphasize the Markov process, which is a special type of stochastic process. Let the system be observed at discrete moments of time $t_n$, where $n = 0, 1, 2, \ldots$, and let $X(t_n)$ denote the state of the system at time $t_n$.

*Definition 1.2*
Let $t_0 < t_1 < \ldots < t_n$. If

$$P[X(t_n) = x_n|X(t_{n-1}) = x_{n-1}, X(t_{n-2}) = x_{n-2}, \ldots, X(t_0) = x_0] = P[X(t_n) = x_n|X(t_{n-1}) = x_{n-1}]$$

then the process is called a *Markov process*.

From the definition of a Markov process, given the present state of the process, its behavior in the future does not depend on its behavior in the past. Many systems have this property, which is called the *Markov property*, and systems that have this property are called *Markov chains*. The Markov property is precisely defined by the following requirement:

$$P[X(t_n) = x_n|X(t_{n-1}) = x_{n-1}, X(t_{n-2}) = x_{n-2}, \ldots, X(t_0) = x_0] = P[X(t_n) = x_n|X(t_{n-1}) = x_{n-1}].$$

The essential characteristic of a Markov process is that it is a process that has no memory; its future is determined by the present and not the past. If, in addition to having no memory, the process is such that it depends only on the difference $(t + \mathrm{d}t) - t = \mathrm{d}t$ and not the value of $t$ – in other words $P[X(t + \mathrm{d}t) = j|X(t) = i]$ is independent of $t$ – then the process is Markov with stationary transition probabilities or is homogeneous in time. This is the same property noted in exponential event times; in fact, referring back to the graphical representation of $X(t)$, the times between state changes are exponential if the process has stationary transition probabilities.

Thus, a Markov process which is homogeneous in time can describe processes with exponential event occurrence times. The random variable of the process is $X(t)$, the state variable rather than the time to failure used in the exponential failure density. To illustrate the types of processes that can be described, we now review the exponential distribution and its properties. Recall that, if $X_1, X_2, \ldots, X_n$, are independent random variables, each with exponential density and a mean of $1/\lambda_i$, then $\min\{X_1, X_2, \ldots, X_n\}$ has an exponential density with a mean of $\left(\sum \lambda_i\right)^{-1}$.

The significance of this property is as follows:

1. The failure behavior of components operated simultaneously can be characterized by an exponential density with a mean equal to the reciprocal of the sum of the failure rates.
2. The joint failure/repair behavior of a system where components are operating and/or undergoing repair can be characterized by an exponential density with a mean equal to the reciprocal of the sum of the failure and repair rates.
3. The failure/repair behavior of a system similar to that described in (2) above but further complicated by active and dormant operating states and sensing and switching can be characterized by an exponential density.

The above property means that almost all reliability and availability models can be characterized by a time-homogeneous Markov process if the various failure times and repair times are exponential. The notation for the Markov process is $\{X(t), t > 0\}$, where $X(t)$ is discrete (state space) and $t$ is continuous (parameter space). By convention, this type of Markov process is called a continuous-parameter Markov chain.

From a reliability/availability viewpoint, there are two types of Markov processes. These are defined as follows:

1. *Absorbing process:* Contains an "absorbing state", which is a state that, once entered, the system can never leave (e.g. a failure which aborts a flight or a mission).
2. *Ergodic process:* Contains no absorbing states, meaning that $X(t)$ can move around indefinitely (e.g. the operation of a ground power plant where failure only temporarily disrupts the operation).

*Pham* ([1.1], p. 265) presents a summary of Markov processes broken down into absorbing and ergodic categories. Both the reliability and the availability can be described in terms of the probability of the process or system being in defined "up" states, e.g. states 1 and 2 in the initial example. Likewise, the MTBF can be described as the total time spent in the "up" states before proceeding to the absorbing state or failure state.

Define the incremental transition probability as

$$P_{ij}(\mathrm{d}t) = P[X(t+\mathrm{d}t) = j | X(t) = i].$$

This is the probability that the process [random variable $X(t)$] will move to state $j$ during the increment $t$ to $(t+\mathrm{d}t)$, given that it was in state $i$ at time $t$. Since we are dealing with time-homogeneous Markov processes (exponential failure and repair times), the incremental transition probabilities can be derived from an analysis of the exponential hazard function. It was shown that the hazard function for an exponential with a mean of $1/\lambda$ was just $\lambda$. This means that the limiting (as $\mathrm{d}t \to 0$) conditional probability of an event occurring between $t$ and $t+\mathrm{d}t$, given that an event had not occurred at time $t$, is simply $\lambda$, in other words:

$$h(t) = \lim_{\mathrm{d}t \to 0} \frac{P[t < X < t+\mathrm{d}t | X > t]}{\mathrm{d}t} = \lambda.$$

The equivalent statement for the random variable $X(t)$ is

$$h(t)\,\mathrm{d}t = P[X(t+\mathrm{d}t) = j | X(t) = i] = \lambda\,\mathrm{d}t.$$

Now, $h(t)\,\mathrm{d}t$ is in fact the incremental transition probability, so $P_{ij}(\mathrm{d}t)$ can be stated in terms of the basic failure and/or repair rates.

Returning to Example 1.20, it is easy to construct a state transition showing the incremental transition probabilities between all possible states for the process:

*State 1*: Both components operating;
*State 2*: One component up and one component down;
*State 3*: Both components down (absorbing state).

The loops in *Pham* ([1.1], p. 265) indicate the probability of remaining in the present state during the $\mathrm{d}t$ increment

$$P_{11}(\mathrm{d}t) = 1 - 2\lambda\,\mathrm{d}t \qquad P_{12}(\mathrm{d}t) = 2\lambda\,\mathrm{d}t$$
$$P_{21}(\mathrm{d}t) = 0 \qquad P_{22}(\mathrm{d}t) = 1 - \lambda\,\mathrm{d}t$$
$$P_{31}(\mathrm{d}t) = 0 \qquad P_{32}(\mathrm{d}t) = 0$$

$$P_{13}(\mathrm{d}t) = 0$$
$$P_{23}(\mathrm{d}t) = \lambda\,\mathrm{d}t$$
$$P_{33}(\mathrm{d}t) = 1$$

The zeros for $P_{ij}$, $i > j$ show that the process cannot go backwards: this is not a repair process. The zero on $P_{13}$ shows that, for a process of this type, the probability of more than one event (e.g. failure, repair, etc.) occurring in the incremental time period $\mathrm{d}t$ approaches zero as $\mathrm{d}t$ approaches zero.

Except for the initial conditions of the process (the state in which the process starts), the process is completely specified by incremental transition probabilities. The reason that this is useful is that assuming exponential event (failure or repair) times allows us to characterize the process at any time $t$, since the process depends only on what happens between $t$ and $(t+\mathrm{d}t)$. The incremental transition probabilities can be arranged

into a matrix in a way that depicts all possible statewide movements. Thus, for parallel configurations,

$$[p_{ij}(\mathrm{d}t)] = \begin{pmatrix} 1 & 2 & 3 \\ 1-2\lambda\,\mathrm{d}t & 2\lambda\,\mathrm{d}t & 0 \\ 0 & 1-\lambda\,\mathrm{d}t & \lambda\,\mathrm{d}t \\ 0 & 0 & 1 \end{pmatrix}$$

for $i, j = 1, 2$, or 3. The matrix $[P_{ij}(\mathrm{d}t)]$ is called the incremental, one-step transition matrix. It is a stochastic matrix (the rows sum to 1.0). As mentioned earlier, this matrix, along with the initial conditions, completely describes the process.

Now, $[P_{ij}(\mathrm{d}t)]$ gives the probabilities of remaining or moving to all of the various states during the interval $t$ to $t+\mathrm{d}t$; hence,

$$\begin{aligned} P_1(t+\mathrm{d}t) &= (1-2\lambda\,\mathrm{d}t)P_1(t) \\ P_2(t+\mathrm{d}t) &= 2\lambda\,\mathrm{d}tP_1(t)(1-\lambda\,\mathrm{d}t)P_2(t) \\ P_3(t+\mathrm{d}t) &= \lambda\,\mathrm{d}tP_2(t)+P_3(t) \end{aligned}$$

By algebraic manipulation, we have

$$\begin{aligned} \frac{[P_1(t+\mathrm{d}t)-P_1(t)]}{\mathrm{d}t} &= -2\lambda P_1(t), \\ \frac{[P_2(t+\mathrm{d}t)-P_2(t)]}{\mathrm{d}t} &= 2\lambda P_1(t)-\lambda P_2(t), \\ \frac{[P_3(t+\mathrm{d}t)-P_3(t)]}{\mathrm{d}t} &= \lambda P_2(t). \end{aligned}$$

Taking limits of both sides as $\mathrm{d}t \to 0$, we obtain

$$\begin{aligned} P_1'(t) &= -2\lambda P_1(t), \\ P_2'(t) &= 2\lambda P_1(t)-\lambda P_2(t), \\ P_3'(t) &= \lambda P_2(t). \end{aligned}$$

The above system of linear first-order differential equations can be easily solved for $P_1(t)$ and $P_2(t)$, meaning that the reliability of the configuration can be obtained:

$$R(t) = \sum_{i=1}^{2} P_i(t).$$

Actually, there is no need to solve all three equations, only the first two, because $P_3(t)$ does not appear and also $P_3(t) = [1-P_1(t)]-P_2(t)$. The system of linear, first-order differential equations can be solved by various means, including both manual and machine methods. We use manual methods employing the Laplace transform (Appendix B) here.

$$\begin{aligned} L[P_i(t)] &= \int_0^\infty \mathrm{e}^{-st}P_i(t)\,\mathrm{d}t = f_i(s), \\ L[P_i'(t)] &= \int_0^\infty \mathrm{e}^{-st}P_i'(t)\,\mathrm{d}t = sf_i(s)-P_i(0). \end{aligned}$$

Application of the Laplace transform will allow us to transform the system of linear, first-order differential equations into a system of linear algebraic equations that can easily be solved, and solutions of $P_i(t)$ can be determined via the inverse transforms.

Returning to the example, the initial condition of a parallel configuration is assumed to be "fully up", such that

$$P_1(t=0) = 1,\ P_2(t=0) = 0,\ P_3(t=0) = 0.$$

Transforming the equations for $P_1'(t)$ and $P_2'(t)$ gives

$$\begin{aligned} sf_1(s)-P_1(t)|_{t=0} &= -2\lambda f_1(s), \\ sf_2(s)-P_2(t)|_{t=0} &= 2\lambda f_1(s)-\lambda f_2(s). \end{aligned}$$

Evaluating $P_1(t)$ and $P_2(t)$ at $t=0$ gives

$$\begin{aligned} sf_1(s)-1 &= -2\lambda f_1(s), \\ sf_2(s)-0 &= 2\lambda f_1(s)-\lambda f_2(s), \end{aligned}$$

from which we obtain

$$\begin{aligned} (s+2\lambda)f_1(s) &= 1, \\ -2\lambda f_1(s)+(s+\lambda)f_2(s) &= 0. \end{aligned}$$

Solving the above equations for $f_1(s)$ and $f_2(s)$, we have

$$\begin{aligned} f_1(s) &= \frac{1}{(s+2\lambda)}, \\ f_2(s) &= \frac{2\lambda}{[(s+2\lambda)(s+\lambda)]}. \end{aligned}$$

From the inverse Laplace transforms in Appendix B,

$$\begin{aligned} P_1(t) &= \mathrm{e}^{-2\lambda t}, \\ P_2(t) &= 2\mathrm{e}^{-\lambda t}-2\mathrm{e}^{-2\lambda t}, \\ R(t) &= P_1(t)+P_2(t) = 2\mathrm{e}^{-\lambda t}-\mathrm{e}^{-2\lambda t}. \end{aligned}$$

The example given above is that of a simple absorbing process where we are concerned about reliability. If a repair capability were added to the model in the form of a repair rate $\mu$, the methodology would remain the same, with only the final result changing. With

a repair rate $\mu$ added to the parallel configuration, the incremental transition matrix would be

$$[P_{ij}(\mathrm{d}t)] = \begin{pmatrix} 1-2\lambda\,\mathrm{d}t & 2\lambda\,\mathrm{d}t & 0 \\ \mu\,\mathrm{d}t & 1-(\lambda+\mu)\,\mathrm{d}t & \lambda\,\mathrm{d}t \\ 0 & 0 & 1 \end{pmatrix}.$$

The differential equations would become

$$P_1'(t) = -2\lambda P_1(t) + \mu P_2(t),$$
$$P_2'(t) = 2\lambda P_1(t) - (\lambda+\mu) P_2(t),$$

and the transformed equations would become

$$(s+2\lambda) f_1(s) - \mu f_2(s) = 1,$$
$$-2\lambda f_1(s) + (s+\lambda+\mu) f_2(s) = 0.$$

Hence, we obtain

$$f_1(s) = \frac{(s+\lambda+\mu)}{(s-s_1)(s-s_2)},$$
$$f_2(s) = \frac{2\lambda}{(s-s_1)(s-s_2)},$$

where

$$s_1 = \frac{-(3\lambda+\mu)+\sqrt{(3\lambda+\mu)2-8\lambda^2}}{2},$$
$$s_2 = \frac{-(3\lambda+\mu)-\sqrt{(3\lambda+\mu)2-8\lambda^2}}{2}. \qquad (1.70)$$

Using the Laplace transform, we obtain

$$P_1(t) = \frac{(s_1+\lambda+\mu)\mathrm{e}^{-s_1 t}}{(s_1-s_2)} + \frac{(s_2+\lambda+\mu)\mathrm{e}^{-s_2 t}}{(s_2-s_1)},$$
$$P_2(t) = \frac{2\lambda\,\mathrm{e}^{-s_1 t}}{(s_1-s_2)} + \frac{2\lambda\,\mathrm{e}^{-s_2 t}}{(s_2-s_1)},$$

where $s_1$ and $s_2$ are given in (1.70).

Thus, the reliability of two components in a parallel system is given by

$$R(t) = P_1(t) + P_2(t) = \frac{(s_1+3\lambda+\mu)\mathrm{e}^{-s_1 t} - (s_2+3\lambda+\mu)\mathrm{e}^{-s_2 t}}{(s_1-s_2)} \qquad (1.71)$$

### System Mean Time Between Failures

Another parameter of interest for absorbing Markov processes is the MTBF. Recalling the previous example of a parallel configuration with repair, the differential equations $P_1'(t)$ and $P_2'(t)$ describing the process were

$$P_1'(t) = -2\lambda P_1(t) + \mu P_2(t),$$
$$P_2'(t) = 2\lambda P_1(t) - (\lambda+\mu) P_2(t).$$

Integrating both sides of the above equations yields

$$\int_0^\infty P_1'(t)\,\mathrm{d}t = -2\lambda \int_0^\infty P_1(t)\,\mathrm{d}t + \mu \int_0^\infty P_2(t)\,\mathrm{d}t,$$
$$\int_0^\infty P_2'(t)\,\mathrm{d}t = 2\lambda \int_0^\infty P_1(t)\,\mathrm{d}t - (\lambda+\mu) \int_0^\infty P_2(t)\,\mathrm{d}t.$$

For the repairable system and from (1.4),

$$\int_0^\infty R(t)\,\mathrm{d}t = \text{MTBF}.$$

Similarly,

$$\int_0^\infty P_1(t)\,\mathrm{d}t = \text{mean time spent in state 1, and}$$
$$\int_0^\infty P_2(t)\,\mathrm{d}t = \text{mean time spent in state 2.}$$

Designating these mean times as $T_1$ and $T_2$, respectively, we have

$$P_1(t)\,\mathrm{d}t|_0^\infty = -2\lambda T_1 + \mu T_2,$$
$$P_2(t)\,\mathrm{d}t|_0^\infty = 2\lambda T_1 - (\lambda+\mu) T_2.$$

But $P_1(t) = 0$ as $t \to \infty$ and $P_1(t) = 1$ for $t = 0$. Likewise, $P_2(t) = 0$ as $t \to \infty$ and $P_2(t) = 0$ for $t = 0$. Thus,

$$-1 = -2\lambda T_1 + \mu T_2,$$
$$0 = 2\lambda T_1 - (\lambda+\mu) T_2,$$

or, equivalently,

$$\begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} -2\lambda & \mu \\ 2\lambda & -(\lambda+\mu) \end{pmatrix} \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}.$$

Therefore,

$$T_1 = \frac{(\lambda+\mu)}{2\lambda^2}, \quad T_2 = \frac{1}{\lambda},$$
$$\text{MTBF} = T_1 + T_2 = \frac{(\lambda+\mu)}{2\lambda^2} + \frac{1}{\lambda} = \frac{(3\lambda+\mu)}{2\lambda^2}.$$

The MTBF for unmaintained processes is developed in exactly the same way as just shown.

The last case to consider for absorbing processes is that of the availability of a maintained system. The difference between reliability and availability is somewhat subtle for absorbing processes. A good example is that of a communications system where the mission would

continue if such a system failed temporarily, but if it failed permanently the mission would be aborted. Consider a cold-standby system consisting of two units: one main unit and one spare unit [1.1]:

*State 1*: Main unit operating and the spare is OK;
*State 2*: Main unit out and restoration underway;
*State 3*: Spare unit is installed and operating;
*State 4*: Permanent failure (no spare available).

The incremental transition matrix is given by

$$[P_{ij}(\mathrm{d}t)] = \begin{pmatrix} 1-\lambda\,\mathrm{d}t & \lambda\,\mathrm{d}t & 0 & 0 \\ 0 & 1-\mu\,\mathrm{d}t & \mu\,\mathrm{d}t & 0 \\ 0 & 0 & 1-\lambda\,\mathrm{d}t & \lambda\,\mathrm{d}t \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

We obtain

$$P_1'(t) = -\lambda P_1(t),$$
$$P_2'(t) = \lambda P_1(t) - \mu P_2(t),$$
$$P_3'(t) = \mu P_2(t) - \lambda P_3(t).$$

Using the Laplace transform, we obtain the following results.

The probability of full-up performance $P_1(t)$ is given by

$$P_1(t) = \mathrm{e}^{-\lambda t}.$$

The probability of a down system that is under repair $P_2(t)$ is

$$P_2(t) = \left(\frac{\lambda}{(\lambda-\mu)}\right)\left(\mathrm{e}^{-\mu t} - \mathrm{e}^{-\lambda t}\right).$$

Similarly, the probability of a fully up system with no spare available $P_3(t)$ is

$$P_3(t) = \left(\frac{\lambda\mu}{(\lambda-\mu)^2}\right)[\mathrm{e}^{-\mu t} - \mathrm{e}^{-\lambda t} - (\lambda-\mu)t\,\mathrm{e}^{-\lambda t}].$$

Hence, the point availability $A(t)$ is given by

$$A(t) = P_1(t) + P_3(t).$$

If average or interval availability is required, this is achieved by

$$\left(\frac{1}{t}\right)\int_0^T A(t)\,\mathrm{d}t = \left(\frac{1}{t}\right)\int_0^T [P_1(t) + P_3(t)]\,\mathrm{d}t,$$

where $T$ is the interval of concern.

Ergodic processes, as opposed to absorbing processes, do not have any absorbing states, and hence movement between states can go on indefinitely. For the latter reason, availability (point, steady-state, or interval) is the only meaningful measure. As an example of an ergodic process, we will use a ground-based power unit configured in parallel.

The parallel units are identical, each with exponential failure and repair times with means $1/\lambda$ and $1/\mu$, respectively. Assume a two-repairmen capability if required (both units down), then

*State 1*: Fully up (both units operating);
*State 2*: One unit down and under repair (other unit up);
*State 3*: Both units down and under repair.

It should be noted that, as in the case of failure events, two or more repairs cannot be made in the $\mathrm{d}t$ interval.

$$[P_{ij}(\mathrm{d}t)] = \begin{pmatrix} 1-2\lambda\,\mathrm{d}t & 2\lambda\,\mathrm{d}t & 0 \\ \mu\,\mathrm{d}t & 1-(\lambda+\mu)\,\mathrm{d}t & \lambda\,\mathrm{d}t \\ 0 & 2\mu\,\mathrm{d}t & 1-2\mu\,\mathrm{d}t \end{pmatrix}.$$

Case I: *Point Availability – Ergodic Process.* For an ergodic process, as $t \to \infty$ the availability settles down to a constant level. Point availability allows us to study the process before this "settling down", and it reflects the initial conditions in the process. We can obtain a solution for the point availability in a similar way to that for absorbing processes, except that the last row and column of the transition matrix must be retained and entered into the system of equations. For example, the system of differential equations becomes

$$\begin{pmatrix} P_1'(t) \\ P_2'(t) \\ P_3'(t) \end{pmatrix} = \begin{pmatrix} -2\lambda & \mu & 0 \\ 2\lambda & -(\lambda+\mu) & 2\mu \\ 0 & \lambda & -2\mu \end{pmatrix} \begin{pmatrix} P_1(t) \\ P_2(t) \\ P_3(t) \end{pmatrix}.$$

Similar to the absorbing case, the Laplace transform can be used to solve for $P_1(t)$, $P_2(t)$ and $P_3(t)$; the point availability $A(t)$ is given by

$$A(t) = P_1(t) + P_2(t).$$

Case II: *Interval Availability – Ergodic Process.* This is the same as the absorbing case, with integration over the time period $T$ of interest. The interval availability, $A(T)$, is

$$A(T) = \frac{1}{T}\int_0^T A(t)\,\mathrm{d}t.$$

Case III: *Steady State Availability – Ergodic Process.* Here, the process is examined as $t \to \infty$, with complete "washout" of the initial conditions. By letting $t \to \infty$,

the system of differential equations can be transformed into linear algebraic equations. Thus,

$$\lim_{t\to\infty}\begin{pmatrix}P_1'(t)\\P_2'(t)\\P_3'(t)\end{pmatrix} = \lim_{t\to\infty}\begin{pmatrix}-2\lambda & \mu & 0\\ 2\lambda & -(\lambda+\mu) & 2\mu\\ 0 & \lambda & -2\mu\end{pmatrix}\begin{pmatrix}P_1(t)\\P_2(t)\\P_3(t)\end{pmatrix}.$$

As $t\to\infty$, $P_i(t)\to$ constant and $P_i'(t)\to 0$. This leads to an unsolvable system, namely,

$$\begin{pmatrix}0\\0\\0\end{pmatrix} = \begin{pmatrix}-2\lambda & \mu & 0\\ 2\lambda & -(\lambda+\mu) & 2\mu\\ 0 & \lambda & -2\mu\end{pmatrix}\begin{pmatrix}P_1(t)\\P_2(t)\\P_3(t)\end{pmatrix}.$$

To avoid the above difficulty, an additional equation is introduced:

$$\sum_{i=1}^{3} P_i(t) = 1.$$

With the introduction of the new equation, one of the original equations is deleted and a new system is formed:

$$\begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}1 & 1 & 1\\ -2\lambda & \mu & 0\\ 2\lambda & -(\lambda+\mu) & 2\mu\end{pmatrix}\begin{pmatrix}P_1(t)\\P_2(t)\\P_3(t)\end{pmatrix}$$

or, equivalently,

$$\begin{pmatrix}P_1(t)\\P_2(t)\\P_3(t)\end{pmatrix} = \begin{pmatrix}1 & 1 & 1\\ -2\lambda & \mu & 0\\ 2\lambda & -(\lambda+\mu) & 2\mu\end{pmatrix}^{-1}\begin{pmatrix}1\\0\\0\end{pmatrix}.$$

We now obtain the following results:

$$P_1(t) = \frac{\mu^2}{(\mu+\lambda)^2},$$

$$P_2(t) = \frac{2\lambda\mu}{(\mu+\lambda)^2},$$

and

$$P_3(t) = 1 - P_1(t) - P_2(t),$$

$$= \frac{\lambda^2}{(\mu+\lambda)^2}.$$

Therefore, the steady state availability $A(\infty)$ is given by

$$A_3(\infty) = P_1(t) + P_2(t)$$

$$= \frac{\mu(\mu+2\lambda)}{(\mu+\lambda)^2}.$$

Note that Markov methods can also be employed when failure or repair times are not exponential but can be represented as the sum of exponential times with identical means (an Erlang distribution or gamma distribution with integer-valued shape parameters). Basically, the method involves introducing "dummy" states which, although being of no particular interest in themselves, change the hazard function from constant to increasing.

### 1.4.2 Counting Processes

Among various discrete stochastic processes, counting processes are widely used in engineering statistics to describe the appearance of events in time, such as failures, the number of perfect repairs, etc. The simplest counting process is a Poisson process. The Poisson process plays a special role in many applications related to reliability [1.1]. A classic example of such an application is the decay of uranium. Here, radioactive particles from nuclear material strike a certain target in accordance with a Poisson process of some fixed intensity. One well-known counting process is the so-called renewal process. This process is described as a sequence of events where the intervals between the events are independent and identically distributed random variables. In reliability theory, this type of mathematical model is used to describe the number of occurrences of an event over a time interval. In this section, we also discuss the quasi-renewal process and the nonhomogeneous Poisson process.

A non-negative, integer-valued stochastic process $N(t)$ is called a counting process if $N(t)$ represents the total number of occurrences of an event in the time interval $[0, t]$ and satisfies these two properties:

1. if $t_1 < t_2$, then $N(t_1) \le N(t_2)$,
2. if $t_1 < t_2$, then $N(t_2) - N(t_1)$ is the number of occurrences of the event in the interval $[t_1, t_2]$.

For example, if $N(t)$ equals the number of persons who have entered a restaurant at or prior to time $t$, then $N(t)$ is a counting process in which an event occurs whenever a person enters the restaurant.

#### Poisson Processes

One of the most important counting processes is the Poisson process.

*Definition 1.3*

A counting process $N(t)$ is said to be a Poisson process with intensity $\lambda$ if

1. the failure process $N(t)$ has stationary independent increments;
2. the number of failures in any time interval of length $s$ has a Poisson distribution with a mean of $\lambda s$; in other words

$$P\{N(t+s)-N(t)=n\}=\frac{e^{-\lambda s}(\lambda s)^n}{n!} \qquad n=0,1,2,\ldots; \qquad (1.72)$$

3. the initial condition is $N(0)=0$.

This model is also called a homogeneous Poisson process, indicating that the failure rate $\lambda$ does not depend on time $t$. In other words, the number of failures that occur during the time interval $(t, t+s]$ does not depend on the current time $t$, only the length of the time interval $s$. A counting process is said to possess independent increments if the number of events in disjoint time intervals are independent.

For a stochastic process with independent increments, the autocovariance function is

$$\mathrm{Cov}[X(t_1), X(t_2)] = \begin{cases} \mathrm{Var}[N(t_1+s)-N(t_2)] & \text{for } 0<t_2-t_1<s \\ 0 & \text{otherwise} \end{cases},$$

where

$$X(t)=N(t+s)-N(t).$$

If $X(t)$ is Poisson-distributed, then the variance of the Poisson distribution is

$$\mathrm{Cov}[X(t_1), X(t_2)] = \begin{cases} \lambda[s-(t_2-t_1)] & \text{for } 0<t_2-t_1<s \\ 0 & \text{otherwise} \end{cases}.$$

This result shows that the Poisson increment process is covariance stationary. We now present several properties of the Poisson process.

*Property 1.3*

The sum of independent Poisson processes $N_1(t)$,$N_2(t)$, $\ldots$, $N_k(t)$ with mean values $\lambda_1 t, \lambda_2 t, \ldots, \lambda_k t$, respectively, is also a Poisson process with mean $\left(\sum_{i=1}^{k}\lambda_i\right)t$. In other words, the sum of the independent Poisson processes is also a Poisson process with a mean that is equal to the sum of the means of the individual Poisson processes.

*Property 1.4*

The difference between two independent Poisson processes, $N_1(t)$, and $N_2(t)$, with mean $\lambda_1 t$ and $\lambda_2 t$, respectively, is not a Poisson process. Instead, it has a probability mass function of

$$P[N_1(t)-N_2(t)=k] = e^{-(\lambda_1+\lambda_2)t}\left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{k}{2}} I_k\left(2\sqrt{\lambda_1\lambda_2}t\right), \qquad (1.73)$$

where $I_k(.)$ is a modified Bessel function of order $k$.

*Property 1.5*

If the Poisson process $N(t)$ with mean $\lambda t$ is filtered such that not every occurrence of the event is counted, then the process has a constant probability $p$ of being counted. The result of this process is a Poisson process with mean $\lambda pt$[].

*Property 1.6*

Let $N(t)$ be a Poisson process and $Y_n$ a family of independent and identically distributed random variables which are also independent of $N(t)$. A stochastic process $X(t)$ is said to be a compound Poisson process if it can be represented as

$$X(t)=\sum_{i=1}^{N(t)} Y_i.$$

### Renewal Processes

A renewal process is a more general case of the Poisson process in which the inter-arrival times of the process or the times between failures do not necessarily follow the exponential distribution. For convenience, we will call the occurrence of an event a renewal, the inter-arrival time the renewal period, and the waiting time the renewal time.

*Definition 1.4*

A counting process $N(t)$ that represents the total number of occurrences of an event in the time interval $(0, t]$ is called a renewal process if the times between the failures are independent and identically distributed random variables.

The probability that exactly $n$ failures occur by time $t$ can be written as

$$P[N(t)=n] = P[N(t)\geq n] - P[N(t)>n]. \quad (1.74)$$

Note that the times between the failures are $T_1, T_2, \ldots, T_n$, so the failures occurring at time $W_k$ are

$$W_k = \sum_{i=1}^{k} T_i$$

and

$$T_k = W_k - W_{k-1}.$$

Thus,

$$\begin{aligned} P[N(t)=n] &= P[N(t)\geq n] - P[N(t)>n] \\ &= P[W_n \leq t] - P[W_{n+1}\leq t] \\ &= F_n(t) - F_{n+1}(t), \end{aligned}$$

where $F_n(t)$ is the cumulative distribution function for the time of the $n$th failure and $n = 0, 1, 2, \ldots$.

*Example 1.21:* Consider a software testing model for which the time at which an error is found during the testing phase has an exponential distribution with a failure rate of $X$. It can be shown that the time of the $n$th failure follows the gamma distribution with parameters $k$ and $n$. From (1.74), we obtain

$$\begin{aligned} P[N(t)=n] &= P[N(t)\leq n] - P[N(t)\leq n-1] \\ &= \sum_{k=0}^{n} \frac{(\lambda t)^k}{k!} e^{-\lambda t} - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \\ &= \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad \text{for } n = 0, 1, 2, \ldots. \end{aligned}$$

Several important properties of the renewal function are given below.

*Property 1.7*

The mean value function of the renewal process, denoted by $m(t)$, is equal to the sum of the distribution functions for all renewal times, that is,

$$\begin{aligned} m(t) &= E[N(t)] \\ &= \sum_{n=1}^{\infty} F_n(t). \end{aligned}$$

*Property 1.8*

The renewal function $m(t)$ satisfies the following equation:

$$m(t) = F_a(t) + \int_0^t m(t-s)\,dF_a(s), \quad (1.75)$$

where $F_a(t)$ is the distribution function of the inter-arrival time or the renewal period.

In general, let $y(t)$ be an unknown function to be evaluated and $x(t)$ be any non-negative and integrable function associated with the renewal process. Assume that $F_a(t)$ is the distribution function of the renewal period. We can then obtain the following result.

*Property 1.9*

Let the renewal equation be

$$y(t) = x(t) + \int_0^t y(t-s)\,dF_a(s). \quad (1.76)$$

Then its solution is given by

$$y(t) = x(t) + \int_0^t x(t-s)\,dm(s),$$

where $m(t)$ is the mean value function of the renewal process.

The proof of the above property can be easily derived using the Laplace transform. Let $x(t) = a$. Thus, in Property 1.9, the solution $y(t)$ is given by

$$\begin{aligned} y(t) &= x(t) + \int_0^t x(t-s)\,dm(s) \\ &= a + \int_0^t a\,dm(s) \\ &= a\{1 + E[N(t)]\}. \end{aligned}$$

### Quasi-Renewal Processes

In this section we discuss a general renewal process: the quasi-renewal process. Let $\{N(t), t>0\}$ be a counting process and let $X_n$ be the time between the $(n-1)$th and the $n$th event of this process, $n \geq 1$.

*Definition 1.5*

[1.10]: If the sequence of non-negative random variables $\{X_1, X_2, \dots\}$ is independent and

$$X_i = \alpha X_{i-1} \tag{1.77}$$

for $i \geq 2$ where $\alpha > 0$ is a constant, then the counting process $\{N(t), t \geq 0\}$ is said to be a quasi-renewal process with parameter $\alpha$ and the first inter-arrival time $X_1$.

When $\alpha = 1$, this process becomes the ordinary renewal process. This quasi-renewal process can be used to model reliability growth processes in software testing phases and hardware burn-in stages for $\alpha > 1$, and in hardware maintenance processes when $\alpha \leq 1$.

Assume that the probability density function, cumulative distribution function, survival function and failure rate of random variable $X_1$ are $f_1(x)$, $F_1(x)$, $s_1(x)$ and $r_1(x)$, respectively. Then the pfd, cdf, survival function, and failure rate of $X_n$ for $n = 1, 2, 3, \dots$ are, respectively, given below [1.10]:

$$f_n(x) = \frac{1}{\alpha^{n-1}} f_1\left(\frac{1}{\alpha^{n-1}}x\right),$$
$$F_n(x) = F_1\left(\frac{1}{\alpha^{n-1}}x\right),$$
$$s_n(x) = s_1\left(\frac{1}{\alpha^{n-1}}x\right),$$
$$f_n(x) = \frac{1}{\alpha^{n-1}} r_1\left(\frac{1}{\alpha^{n-1}}x\right).$$

Similarly, the mean and variance of $X_n$ is given as

$$E(X_n) = \alpha^{n-1} E(X_1),$$
$$\mathrm{Var}(X_n) = \alpha^{2n-2}\mathrm{Var}(X_1).$$

Because of the non-negativity of $X_1$, and the fact that $X_1$ is not identically 0, we obtain

$$E(X_1) = \mu_1 \neq 0.$$

It is worth noting that the shape parameters for $X_n$ are the same for $n = 1, 2, 3, \dots$ for a quasi-renewal process if $X_1$ follows the gamma, Weibull, or log normal distribution.

This means that the shape parameters of the inter-arrival time will not change after "renewal". In software reliability, the assumption that the software debugging process does not change the error-free distribution seems reasonable. Thus, if a quasi-renewal process model is used, the error-free times that occur during software debugging will have the same shape parameters. In this sense, a quasi-renewal process is suitable for modeling the increase in software reliability. It is worth noting that

$$\lim_{n\to\infty} \frac{E(X_1 + X_2 + \cdots + X_n)}{n} = \lim_{n\to\infty} \frac{\mu_1(1-\alpha^n)}{(1-\alpha)n},$$
$$= 0 \quad \text{if } \alpha < 1,$$
$$= \infty \text{ if } \alpha > 1.$$

Therefore, if the inter-arrival time represents the error-free time of a software system, then the average error-free time approaches infinity when its debugging process has been operating for a long debugging time.

*Distribution of N(t).* Consider a quasi-renewal process with parameter $\alpha$ and a first inter-arrival time $X_1$. Clearly, the total number of renewals $N(t)$ that occur up to time $t$ has the following relationship to the arrival time of the $n$th renewal $SS_n$:

$$N(t) \geq n \quad \text{if and only if } SS_n \leq t.$$

In other words, $N(t)$ is at least $n$ if and only if the $n$th renewal occurs prior to time $t$. It is easily seen that

$$SS_n = \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \alpha^{i-1} X_1 \quad \text{for } n \geq 1. \tag{1.78}$$

Here, $SS_0 = 0$. Thus, we have

$$P\{N(t) = n\} = P\{N(t) = n\} - P\{N(t) \geq n+1\}$$
$$= P\{SS_n \leq t\} - P\{SS_{n+1} \leq t\}$$
$$= G_n(t) - G_{n+1}(t),$$

where $G_n(t)$ is the convolution of the inter-arrival times $F_1, F_2, F_3, \dots, F_n$. In other words,

$$G_n(t) = P\{F_1 + F_2 + \cdots + F_n \leq t\}.$$

If the mean value of $N(t)$ is defined as the renewal function $m(t)$, then

$$m(t) = E[N(t)]$$
$$= \sum_{n=1}^{\infty} P\{N(t) \geq n\}$$
$$= \sum_{n=1}^{\infty} P\{SS_n \leq t\}$$
$$= \sum_{n=1}^{\infty} G_n(t).$$

The derivative of $m(t)$ is known as the renewal density

$$\lambda(t) = m'(t).$$

In renewal theory, random variables representing inter-arrival distributions assume only non-negative values, and the Laplace transform of its distribution $F_1(t)$ is defined by

$$\mathcal{L}\{F_1(s)\} = \int_0^\infty \mathrm{e}^{-sx}\,\mathrm{d}F_1(x).$$

Therefore,

$$\mathcal{L}F_n(s) = \int_0^\infty \mathrm{e}^{-a^{n-1}st}\,\mathrm{d}F_1(t) = \mathcal{L}F_1(\alpha^{n-1}s)$$

and

$$\begin{aligned}\mathcal{L}m_n(s) &= \sum_{n=1}^{\infty} \mathcal{L}G_n(s)\\ &= \sum_{n=1}^{\infty} \mathcal{L}F_1(s)\mathcal{L}F_1(\alpha s)\cdots\mathcal{L}F_1(\alpha^{n-1}s).\end{aligned}$$

Since there is a one-to-one correspondence between distribution functions and its Laplace transform, it follows that the first inter-arrival distribution of a quasi-renewal process uniquely determines its renewal function.

If the inter-arrival time represents the error-free time (time to first failure), a quasi-renewal process can be used to model reliability growth in both software and hardware.

Suppose that all software faults have the same chance of being detected. If the inter-arrival time of a quasi-renewal process represents the error-free time of a software system, then the expected number of software faults in the time interval $[0, t]$ can be defined by the renewal function, $m(t)$, with parameter $\alpha > 1$. Denoted by $m_\mathrm{r}(t)$, the number of remaining software faults at time $t$, it follows that

$$m_\mathrm{r}(t) = m(T_\mathrm{c}) - m(t)$$

where $m(T_\mathrm{c})$ is the number of faults that will eventually be detected through a software lifecycle $T_\mathrm{c}$.

### Nonhomogeneous Poisson Processes

The nonhomogeneous Poisson process model (NHPP), which represents the number of failures experienced up to time $t$, is a nonhomogeneous Poisson process $\{N(t)$ with $t \geq 0\}$. The main issue with the NHPP model is to determine an appropriate mean value function to denote the expected number of failures experienced up to a certain time.

Different assumptions mean that the model will end up with different functional forms of the mean value function. Note that the exponential assumption for the inter-arrival time between failures is relaxed in a renewal process, and the stationary assumption is relaxed in the NHPP.

The NHPP model is based on the following assumptions:

- The failure process has an independent increment; in other words, the number of failures during the time interval $(t, t+s)$ depends on the current time $t$ and the length of the time interval $s$, and does not depend on the past history of the process.
- The failure rate of the process is given by

$$\begin{aligned}&P\{\text{exactly one failure in } (t, t+\Delta t)\}\\ &\quad = P\{N(t+\Delta t) - N(t) = 1\}\\ &\quad = \lambda(t)\Delta t + o(\Delta t),\end{aligned}$$

  where $\lambda(t)$ is the intensity function.
- During a small interval $\Delta t$, the probability of more than one failure is negligible; that is,

$$P\{\text{two or more failures in } (t, t+\Delta t)\} = o(\Delta t),$$

- The initial condition is $N(0) = 0$.

Based on these assumptions, the probability that exactly $n$ failures occur during the time interval $(0, t)$ for the NHPP is given by

$$\Pr\{N(t) = n\} = \frac{[m(t)]^n}{n!}\,\mathrm{e}^{-m(t)} \quad n = 0, 1, 2, \ldots, \tag{1.79}$$

where $m(t) = E[N(t)] = \int_0^t \lambda(s)\mathrm{d}s$ and $\lambda(t)$ is the intensity function. It is easily shown that the mean value function $m(t)$ is nondecreasing.

The reliability $R(t)$, defined as the probability that there are no failures in the time interval $(0, t)$, is given by

$$\begin{aligned}R(t) &= P\{N(t) = 0\}\\ &= \mathrm{e}^{-m(t)}.\end{aligned}$$

In general, the reliability $R(x|t)$ – the probability that there are no failures in the interval $(t, t+x)$ – is given by

$$\begin{aligned}R(x|t) &= P\{N(t+x) - N(t) = 0\}\\ &= \mathrm{e}^{-[m(t+x)-m(t)]}\end{aligned}$$

and its density is given by

$$f(x) = \lambda(t+x)\,\mathrm{e}^{-[m(t+x)-m(t)]},$$

where

$$\lambda(x) = \frac{\partial}{\partial x}[m(x)].$$

The variance of the NHPP can be obtained as follows:

$$\mathrm{Var}[N(t)] = \int_0^t \lambda(s)\,\mathrm{d}s$$

and the autocorrelation function is given by

$$\begin{aligned}\mathrm{Cor}[s] &= E[N(t)]E[N(t+s)-N(t)] + E[N^2(t)] \\ &= \int_0^t \lambda(s)\,\mathrm{d}s \int_0^{t+s} \lambda(s)\,\mathrm{d}s + \int_0^t \lambda(s)\,\mathrm{d}s \\ &= \int_0^t \lambda(s)\,\mathrm{d}s \left[1 + \int_0^{t+s} \lambda(s)\,\mathrm{d}s\right].\end{aligned}$$

*Example 1.22:* Assume that the intensity $\lambda$ is a random variable with pdf $f(\lambda)$. Then the probability that exactly $n$ failures occur during the time interval $(0, t)$ is given by

$$P\{N(t) = n\} = \int_0^\infty \mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!} f(\lambda)\,\mathrm{d}\lambda.$$

If the pdf $f(\lambda)$ is given as the following gamma density function with parameters $k$ and $m$:

$$f(\lambda) = \frac{1}{\Gamma(m)} k^m \lambda^{m-1}\,\mathrm{e}^{-k\lambda} \quad \text{for } \lambda \geq 0$$

then it can be shown that

$$P\{N(t) = n\} = \binom{n+m-1}{n} p^m q^n \quad n = 0, 1, 2, \ldots$$

(this is also called a negative binomial density function), where

$$p = \frac{k}{t+k} \quad \text{and} \quad q = \frac{t}{t+k} = 1 - p.$$

## 1.5 Further Reading

The reader interested in a deeper understanding of advanced probability theory and stochastic processes should note the following citations, which refer to highly recommended books: [1.9, 11–13]

### References


1.1 H. Pham: *Software Reliability* (Springer, Berlin, Heidelberg 2000)

1.2 W. Weibull: A statistical distribution function of wide applicability, J. Appl. Mech. **18**, 293–297 (1951)

1.3 H. Pham: A Vtub-shaped hazard rate function with applications to system safety, Int. J. Reliab. Appl. **3**(1), 1–16 (2002)

1.4 Z. Chen: Exact confidence interval for the shape parameter of a loglogistic distribution, J. Stat. Comput. Sim. **56**, 193–211 (1997)

1.5 W. Nelson: *Applied Life Data Analysis* (Wiley, New York 1982)

1.6 M. Zhao: Statistical reliability change-point estimation models. In: *Handbook of Reliability Engineering*, ed. by H. Pham (Springer, Berlin, Heidelberg 2003) pp. 157–163

1.7 Z. Jelinski, P. B. Moranda: Software reliability research. In: *Statistical Computer Performance Evaluation*, ed. by W. Freiberger (Academic, New York 1972)

1.8 J. D. Musa, A. Lannino, K. Okumoto: *Software Reliability: Measurement, Prediction, and Application* (McGraw-Hill, New York 1987)

1.9 W. Feller: *An Introduction to Probability Theory and Its Applications*, 3rd edn. (Wiley, New York 1994)

1.10 H. Wang, H. Pham: A quasi renewal process and its applications in imperfect maintenance, Int. J. Syst. Sci. **27**(10), 1055–1062 (1996)

1.11 J. L. Devore: *Probability and Statistics for Engineering and the Sciences*, 3rd edn. (Brooks Cole, Pacific Grove 1991)

1.12 B. V. Gnedenko, I. A. Ushakov: *Probabilistic Reliability Engineering* (Wiley, New York 1995)

1.13 J. G. Hahn, W. Q. Meeker: *Statistical Intervals: A Guide for Practitioners* (Wiley, New York 1991)

# 1.A Appendix: Distribution Tables

**Table 1.6** Cumulative areas under the standard normal distribution

| Z | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| −3.0 | 0.0013 | 0.0010 | 0.0007 | 0.0005 | 0.0003 | 0.0002 | 0.0002 | 0.0001 | 0.0001 | 0.0000 |
| −2.9 | 0.0019 | 0.0018 | 0.0017 | 0.0017 | 0.0016 | 0.0016 | 0.0015 | 0.0015 | 0.0014 | 0.0014 |
| −2.8 | 0.0026 | 0.0025 | 0.0024 | 0.0023 | 0.0023 | 0.0022 | 0.0021 | 0.0021 | 0.0020 | 0.0019 |
| −2.7 | 0.0035 | 0.0034 | 0.0033 | 0.0032 | 0.0031 | 0.0030 | 0.0029 | 0.0028 | 0.0027 | 0.0026 |
| −2.6 | 0.0047 | 0.0045 | 0.0044 | 0.0043 | 0.0041 | 0.0040 | 0.0039 | 0.0038 | 0.0037 | 0.0036 |
| −2.5 | 0.0062 | 0.0060 | 0.0059 | 0.0057 | 0.0055 | 0.0054 | 0.0052 | 0.0051 | 0.0049 | 0.0048 |
| −2.4 | 0.0082 | 0.0080 | 0.0078 | 0.0075 | 0.0073 | 0.0071 | 0.0069 | 0.0068 | 0.0066 | 0.0064 |
| −2.3 | 0.0107 | 0.0104 | 0.0102 | 0.0099 | 0.0096 | 0.0094 | 0.0091 | 0.0089 | 0.0087 | 0.0084 |
| −2.2 | 0.0139 | 0.0136 | 0.0132 | 0.0129 | 0.0126 | 0.0122 | 0.0119 | 0.0116 | 0.0113 | 0.0110 |
| −2.1 | 0.0179 | 0.0174 | 0.0170 | 0.0166 | 0.0162 | 0.0158 | 0.0154 | 0.0150 | 0.0146 | 0.0143 |
| −2.0 | 0.0228 | 0.0222 | 0.0217 | 0.0212 | 0.0207 | 0.0202 | 0.0197 | 0.0192 | 0.0188 | 0.0183 |
| −1.9 | 0.0287 | 0.0281 | 0.0274 | 0.0268 | 0.0262 | 0.0256 | 0.0250 | 0.0244 | 0.0238 | 0.0233 |
| −1.8 | 0.0359 | 0.0352 | 0.0344 | 0.0336 | 0.0329 | 0.0322 | 0.0314 | 0.0307 | 0.0300 | 0.0294 |
| −1.7 | 0.0446 | 0.0436 | 0.0427 | 0.0418 | 0.0409 | 0.0401 | 0.0392 | 0.0384 | 0.0375 | 0.0367 |
| −1.6 | 0.0548 | 0.0537 | 0.0526 | 0.0516 | 0.0505 | 0.0495 | 0.0485 | 0.0475 | 0.0465 | 0.0455 |
| −1.5 | 0.0668 | 0.0655 | 0.0643 | 0.0630 | 0.0618 | 0.0606 | 0.0594 | 0.0582 | 0.0570 | 0.0559 |
| −1.4 | 0.0808 | 0.0793 | 0.0778 | 0.0764 | 0.0749 | 0.0735 | 0.0722 | 0.0708 | 0.0694 | 0.0681 |
| −1.3 | 0.0968 | 0.0951 | 0.0934 | 0.0918 | 0.0901 | 0.0885 | 0.0869 | 0.0853 | 0.0838 | 0.0823 |
| −1.2 | 0.1151 | 0.1131 | 0.1112 | 0.1093 | 0.1075 | 0.1056 | 0.1038 | 0.1020 | 0.1003 | 0.0985 |
| −1.1 | 0.1357 | 0.1335 | 0.1314 | 0.1292 | 0.1271 | 0.1251 | 0.1230 | 0.1210 | 0.1190 | 0.1170 |
| −1.0 | 0.1587 | 0.1562 | 0.1539 | 0.1515 | 0.1492 | 0.1469 | 0.1446 | 0.1423 | 0.1401 | 0.1379 |
| −0.9 | 0.1841 | 0.1814 | 0.1788 | 0.1762 | 0.1736 | 0.1711 | 0.1685 | 0.1660 | 0.1635 | 0.1611 |
| −0.8 | 0.2119 | 0.2090 | 0.2061 | 0.2033 | 0.2005 | 0.1977 | 0.1949 | 0.1922 | 0.1894 | 0.1867 |
| −0.7 | 0.2420 | 0.2389 | 0.2358 | 0.2327 | 0.2297 | 0.2266 | 0.2236 | 0.2206 | 0.2177 | 0.2148 |
| −0.6 | 0.2743 | 0.2709 | 0.2676 | 0.2643 | 0.2611 | 0.2578 | 0.2546 | 0.2514 | 0.2483 | 0.2451 |
| −0.5 | 0.3085 | 0.3050 | 0.3015 | 0.2981 | 0.2946 | 0.2912 | 0.2877 | 0.2843 | 0.2810 | 0.2776 |
| −0.4 | 0.3446 | 0.3409 | 0.3372 | 0.3336 | 0.3300 | 0.3264 | 0.3228 | 0.3192 | 0.3156 | 0.3121 |
| −0.3 | 0.3821 | 0.3783 | 0.3745 | 0.3707 | 0.3669 | 0.3632 | 0.3594 | 0.3557 | 0.3520 | 0.3483 |
| −0.2 | 0.4207 | 0.4168 | 0.4129 | 0.4090 | 0.4052 | 0.4013 | 0.3974 | 0.3936 | 0.3897 | 0.3859 |
| −0.1 | 0.4602 | 0.4562 | 0.4522 | 0.4483 | 0.4443 | 0.4404 | 0.4364 | 0.4325 | 0.4286 | 0.4247 |
| −0.0 | 0.5000 | 0.4960 | 0.4920 | 0.4880 | 0.4840 | 0.4801 | 0.4761 | 0.4721 | 0.4681 | 0.4641 |
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7703 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |

**Table 1.6** (cont.)

| Z | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9278 | 0.9292 | 0.9306 | 0.9319 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9430 | 0.9441 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |
| 1.8 | 0.9641 | 0.9648 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9686 | 0.9693 | 0.9700 | 0.9706 |
| 1.9 | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9762 | 0.9767 |
| 2.0 | 0.9772 | 0.9778 | 0.9783 | 0.9788 | 0.9793 | 0.9798 | 0.9803 | 0.9808 | 0.9812 | 0.9817 |
| 2.1 | 0.9821 | 0.9826 | 0.9830 | 0.9834 | 0.9838 | 0.9842 | 0.9846 | 0.9850 | 0.9854 | 0.9857 |
| 2.2 | 0.9861 | 0.9864 | 0.9868 | 0.9871 | 0.9874 | 0.9878 | 0.9881 | 0.9884 | 0.9887 | 0.9890 |
| 2.3 | 0.9893 | 0.9896 | 0.9898 | 0.9901 | 0.9904 | 0.9906 | 0.9909 | 0.9911 | 0.9913 | 0.9916 |
| 2.4 | 0.9918 | 0.9920 | 0.9922 | 0.9925 | 0.9927 | 0.9929 | 0.9931 | 0.9932 | 0.9934 | 0.9936 |
| 2.5 | 0.9938 | 0.9940 | 0.9941 | 0.9943 | 0.9945 | 0.9946 | 0.9948 | 0.9949 | 0.9951 | 0.9952 |
| 2.6 | 0.9953 | 0.9955 | 0.9956 | 0.9957 | 0.9959 | 0.9960 | 0.9961 | 0.9962 | 0.9963 | 0.9964 |
| 2.7 | 0.9965 | 0.9966 | 0.9967 | 0.9968 | 0.9969 | 0.9970 | 0.9971 | 0.9972 | 0.9973 | 0.9974 |
| 2.8 | 0.9974 | 0.9975 | 0.9976 | 0.9977 | 0.9977 | 0.9978 | 0.9979 | 0.9979 | 0.9980 | 0.9981 |
| 2.9 | 0.9981 | 0.9982 | 0.9982 | 0.9983 | 0.9984 | 0.9984 | 0.9985 | 0.9985 | 0.9986 | 0.9986 |
| 3.0 | 0.9987 | 0.9990 | 0.9993 | 0.9995 | 0.9997 | 0.9998 | 0.9998 | 0.9999 | 0.9999 | 1.000 |

**Table 1.7** Percentage points for the $t$-distribution ($t_{\alpha,r}$)

| r / α | 0.100 | 0.050 | 0.025 | 0.01 | 0.005 | 0.0025 | 0.001 |
|---|---|---|---|---|---|---|---|
| 1 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 127.32 | 318.310 |
| 2 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 14.089 | 23.326 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 7.453 | 10.213 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 5.598 | 7.173 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 4.773 | 5.893 |
| 6 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 4.317 | 5.208 |
| 7 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 4.029 | 4.785 |
| 8 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 3.833 | 4.501 |
| 9 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 3.690 | 4.297 |
| 10 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 3.581 | 4.144 |
| 11 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 3.497 | 4.025 |
| 12 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 3.428 | 3.930 |
| 13 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 3.372 | 3.852 |
| 14 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 3.326 | 3.787 |
| 15 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 3.286 | 3.733 |
| 16 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 3.252 | 3.686 |
| 17 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.222 | 3.646 |
| 18 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.197 | 3.610 |
| 19 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.174 | 3.579 |
| 20 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.153 | 3.552 |
| 21 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 3.135 | 3.527 |
| 22 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 3.119 | 3.505 |
| 23 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 3.104 | 3.485 |
| 24 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 3.091 | 3.467 |

**Table 1.7** (cont.)

| r / α | 0.100 | 0.050 | 0.025 | 0.01 | 0.005 | 0.0025 | 0.001 |
|---|---|---|---|---|---|---|---|
| 25 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 3.078 | 3.450 |
| 26 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.067 | 3.435 |
| 27 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.057 | 3.421 |
| 28 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.047 | 3.408 |
| 29 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 3.038 | 3.396 |
| 30 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | 3.030 | 3.385 |
| 40 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | 2.971 | 3.307 |
| 60 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | 2.915 | 3.232 |
| 120 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 | 2.860 | 3.160 |
| ∞ | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 2.807 | 3.090 |

**Table 1.8** Percentage points for the $F$-distribution $F_{0.05}, \nu_2/\nu_1$

| $\nu_2$ / $\nu_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 161.40 | 199.50 | 215.70 | 224.60 | 230.20 | 234.00 | 236.80 | 238.90 | 240.50 | 241.90 |
| 2 | 18.51 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.35 | 19.37 | 19.38 | 19.40 |
| 3 | 10.13 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 | 8.85 | 8.81 | 8.79 |
| 4 | 7.71 | 6.94 | 6.59 | 6.39 | 9.26 | 9.16 | 6.09 | 6.04 | 6.00 | 5.96 |
| 5 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 | 4.82 | 4.77 | 4.74 |
| 6 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 | 4.15 | 4.10 | 4.06 |
| 7 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 | 3.73 | 3.68 | 3.64 |
| 8 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 | 3.44 | 3.39 | 3.35 |
| 9 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 | 3.23 | 3.18 | 3.14 |
| 10 | 4.95 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 | 3.07 | 3.02 | 2.98 |
| 11 | 4.84 | 3.98 | 3.59 | 3.36 | 3.20 | 3.09 | 3.01 | 2.95 | 2.90 | 2.85 |
| 12 | 4.75 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.91 | 2.85 | 2.80 | 2.75 |
| 13 | 4.67 | 3.81 | 3.41 | 3.18 | 3.03 | 2.92 | 2.83 | 2.77 | 2.71 | 2.67 |
| 14 | 4.60 | 3.74 | 3.34 | 3.11 | 2.96 | 2.85 | 2.76 | 2.70 | 2.65 | 2.60 |
| 15 | 4.54 | 3.68 | 3.29 | 3.06 | 2.90 | 2.79 | 2.71 | 2.64 | 2.59 | 2.54 |
| 16 | 4.49 | 3.63 | 3.24 | 3.01 | 2.85 | 2.74 | 2.66 | 2.59 | 2.54 | 2.49 |
| 17 | 4.45 | 3.59 | 3.20 | 2.96 | 2.81 | 2.70 | 2.61 | 2.55 | 2.49 | 2.45 |
| 18 | 4.41 | 3.55 | 3.16 | 2.93 | 2.77 | 2.66 | 2.58 | 2.51 | 2.46 | 2.41 |
| 19 | 4.38 | 3.52 | 3.13 | 2.90 | 2.74 | 2.63 | 2.54 | 2.48 | 2.42 | 2.38 |
| 20 | 4.35 | 3.49 | 3.10 | 2.87 | 2.71 | 2.60 | 2.51 | 2.45 | 2.39 | 2.35 |
| 21 | 4.32 | 3.47 | 3.07 | 2.84 | 2.68 | 2.57 | 2.49 | 2.42 | 2.37 | 2.32 |
| 22 | 4.30 | 3.44 | 3.05 | 2.82 | 2.66 | 2.55 | 2.46 | 2.40 | 2.34 | 2.30 |
| 23 | 4.28 | 3.42 | 3.03 | 2.80 | 2.64 | 2.53 | 2.44 | 2.37 | 2.32 | 2.27 |
| 24 | 4.26 | 3.40 | 3.01 | 2.78 | 2.62 | 2.51 | 2.42 | 2.36 | 2.30 | 2.25 |
| 25 | 4.24 | 3.39 | 2.99 | 2.76 | 2.60 | 2.49 | 2.40 | 2.34 | 2.28 | 2.24 |
| 26 | 4.23 | 3.37 | 2.98 | 2.74 | 2.59 | 2.47 | 2.39 | 2.32 | 2.27 | 2.22 |
| 27 | 4.21 | 3.35 | 2.96 | 2.73 | 2.57 | 2.46 | 2.37 | 2.31 | 2.25 | 2.20 |
| 28 | 4.20 | 3.34 | 2.95 | 2.71 | 2.56 | 2.45 | 2.36 | 2.29 | 2.24 | 2.19 |
| 29 | 4.18 | 3.33 | 2.93 | 2.70 | 2.55 | 2.43 | 2.35 | 2.28 | 2.22 | 2.18 |
| 30 | 4.17 | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.33 | 2.27 | 2.21 | 2.16 |
| 40 | 4.08 | 3.23 | 2.84 | 2.61 | 2.45 | 2.34 | 2.25 | 2.18 | 2.12 | 2.08 |
| 60 | 4.00 | 3.15 | 2.76 | 2.53 | 2.37 | 2.25 | 2.17 | 2.10 | 2.04 | 1.99 |
| 120 | 3.92 | 3.07 | 2.68 | 2.45 | 2.29 | 2.17 | 2.09 | 2.02 | 1.96 | 1.91 |
| ∞ | 3.84 | 3.00 | 2.60 | 2.37 | 2.21 | 2.10 | 2.01 | 1.94 | 1.88 | 1.83 |

**Table 1.9** Percentage points for the $\chi^2$ distribution

| $\nu$ / $\chi^2_\alpha$ | $\chi^2_{0.99}$ | $\chi^2_{0.975}$ | $\chi^2_{0.95}$ | $\chi^2_{0.90}$ | $\chi^2_{0.10}$ | $\chi^2_{0.05}$ | $\chi^2_{0.025}$ | $\chi^2_{0.01}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0.00 | 0.00 | 0.02 | 2.71 | 3.84 | 5.02 | 6.64 |
| 2 | 0.02 | 0.05 | 0.10 | 0.21 | 4.61 | 5.99 | 7.38 | 9.21 |
| 3 | 0.12 | 0.22 | 0.35 | 0.58 | 6.25 | 7.82 | 9.35 | 11.35 |
| 4 | 0.30 | 0.48 | 0.71 | 1.06 | 7.78 | 9.49 | 11.14 | 13.28 |
| 5 | 0.55 | 0.83 | 1.15 | 1.61 | 9.24 | 11.07 | 12.83 | 15.09 |
| 6 | 0.87 | 1.24 | 1.64 | 2.20 | 10.65 | 12.59 | 14.45 | 16.81 |
| 7 | 1.24 | 1.69 | 2.17 | 2.83 | 12.02 | 14.07 | 16.01 | 18.48 |
| 8 | 1.65 | 2.18 | 2.73 | 3.49 | 13.36 | 15.51 | 17.54 | 20.09 |
| 9 | 2.09 | 2.70 | 3.33 | 4.17 | 14.68 | 16.92 | 19.02 | 21.67 |
| 10 | 2.56 | 3.25 | 3.94 | 4.87 | 15.99 | 18.31 | 20.48 | 23.21 |
| 11 | 3.05 | 3.82 | 4.58 | 5.58 | 17.28 | 19.68 | 21.92 | 24.73 |
| 12 | 3.57 | 4.40 | 5.23 | 6.30 | 18.55 | 21.92 | 23.34 | 26.22 |
| 13 | 4.11 | 5.01 | 5.89 | 7.04 | 19.81 | 22.36 | 24.74 | 27.69 |
| 14 | 4.66 | 5.63 | 6.57 | 7.79 | 21.06 | 23.69 | 26.12 | 29.14 |
| 15 | 5.23 | 6.26 | 7.26 | 8.57 | 22.31 | 25.00 | 27.49 | 30.58 |
| 16 | 5.81 | 6.91 | 7.96 | 9.31 | 23.54 | 26.30 | 28.85 | 32.00 |
| 17 | 6.41 | 7.56 | 8.67 | 10.09 | 24.77 | 27.59 | 30.19 | 33.41 |
| 18 | 7.02 | 8.23 | 9.39 | 10.87 | 25.99 | 28.87 | 31.53 | 34.81 |
| 19 | 7.63 | 8.91 | 10.12 | 11.65 | 27.20 | 30.14 | 32.85 | 36.19 |
| 20 | 8.26 | 9.59 | 10.85 | 12.44 | 28.41 | 31.41 | 34.17 | 37.57 |
| 21 | 8.90 | 10.28 | 11.59 | 13.24 | 29.62 | 32.67 | 35.48 | 38.93 |
| 22 | 9.54 | 10.98 | 12.34 | 14.04 | 30.81 | 33.92 | 36.78 | 40.29 |
| 23 | 10.20 | 11.69 | 13.09 | 14.85 | 32.01 | 35.17 | 38.08 | 41.64 |
| 24 | 10.86 | 12.40 | 13.85 | 15.66 | 33.20 | 36.42 | 39.36 | 42.98 |
| 25 | 11.52 | 13.12 | 14.61 | 16.47 | 34.38 | 37.65 | 40.65 | 44.31 |
| 26 | 12.20 | 13.84 | 15.38 | 17.29 | 35.56 | 38.89 | 41.92 | 45.64 |
| 27 | 12.88 | 14.57 | 16.15 | 18.11 | 36.74 | 40.11 | 43.19 | 46.96 |
| 28 | 13.57 | 15.31 | 16.93 | 18.94 | 37.92 | 41.34 | 44.46 | 48.28 |
| 29 | 14.26 | 16.05 | 17.71 | 19.77 | 39.09 | 42.56 | 45.72 | 49.59 |
| 30 | 14.95 | 16.79 | 18.49 | 20.60 | 40.26 | 43.77 | 46.98 | 50.89 |
| 35 | 18.48 | 20.56 | 22.46 | 24.81 | 46.03 | 49.80 | 53.21 | 57.36 |
| 40 | 22.14 | 24.42 | 26.51 | 29.07 | 51.78 | 55.76 | 59.35 | 63.71 |
| 50 | 29.69 | 32.35 | 34.76 | 37.71 | 63.14 | 67.50 | 71.42 | 76.17 |
| 60 | 37.47 | 40.47 | 43.19 | 46.48 | 74.37 | 79.08 | 83.30 | 88.39 |
| 70 | 45.43 | 48.75 | 51.74 | 55.35 | 85.50 | 90.53 | 95.03 | 100.44 |
| 80 | 53.53 | 57.15 | 60.39 | 64.30 | 96.55 | 101.88 | 106.63 | 112.34 |
| 90 | 61.74 | 65.64 | 69.12 | 73.31 | 107.54 | 113.15 | 118.14 | 124.13 |
| 100 | 70.05 | 74.22 | 77.93 | 82.38 | 118.47 | 124.34 | 129.57 | 135.81 |
| 110 | 78.45 | 82.86 | 86.79 | 91.50 | 129.36 | 135.48 | 140.92 | 147.42 |
| 120 | 86.91 | 91.57 | 95.70 | 100.65 | 140.20 | 146.57 | 152.22 | 158.96 |

**Table 1.10** Critical values $d_{n,\alpha}$ for the Kolmogorov–Smirnov test

| $n$ / $\alpha$ | 0.2 | 0.1 | 0.05 | 0.02 | 0.01 | $n$ / $\alpha$ | 0.2 | 0.1 | 0.05 | 0.02 | 0.01 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.900 | 0.950 | 0.975 | 0.990 | 0.995 | 16 | 0.258 | 0.295 | 0.327 | 0.366 | 0.392 |
| 2 | 0.684 | 0.776 | 0.842 | 0.900 | 0.929 | 17 | 0.250 | 0.286 | 0.318 | 0.355 | 0.381 |
| 3 | 0.565 | 0.636 | 0.708 | 0.785 | 0.829 | 18 | 0.244 | 0.279 | 0.309 | 0.346 | 0.371 |
| 4 | 0.493 | 0.565 | 0.624 | 0.689 | 0.734 | 19 | 0.237 | 0.271 | 0.301 | 0.337 | 0.361 |
| 5 | 0.447 | 0.509 | 0.563 | 0.627 | 0.669 | 20 | 0.232 | 0.265 | 0.294 | 0.329 | 0.352 |
| 6 | 0.410 | 0.468 | 0.519 | 0.577 | 0.617 | 21 | 0.226 | 0.259 | 0.287 | 0.321 | 0.344 |
| 7 | 0.381 | 0.436 | 0.483 | 0.538 | 0.576 | 22 | 0.221 | 0.253 | 0.281 | 0.314 | 0.337 |
| 8 | 0.358 | 0.410 | 0.454 | 0.507 | 0.542 | 23 | 0.216 | 0.247 | 0.275 | 0.307 | 0.330 |
| 9 | 0.339 | 0.387 | 0.430 | 0.480 | 0.513 | 24 | 0.212 | 0.242 | 0.264 | 0.301 | 0.323 |
| 10 | 0.323 | 0.369 | 0.409 | 0.457 | 0.489 | 25 | 0.208 | 0.238 | 0.264 | 0.295 | 0.317 |
| 11 | 0.308 | 0.352 | 0.391 | 0.437 | 0.468 | 26 | 0.204 | 0.233 | 0.259 | 0.290 | 0.311 |
| 12 | 0.296 | 0.338 | 0.375 | 0.419 | 0.449 | 27 | 0.200 | 0.229 | 0.254 | 0.284 | 0.305 |
| 13 | 0.285 | 0.325 | 0.361 | 0.404 | 0.432 | 28 | 0.197 | 0.225 | 0.250 | 0.279 | 0.300 |
| 14 | 0.275 | 0.314 | 0.349 | 0.390 | 0.418 | 29 | 0.193 | 0.221 | 0.246 | 0.275 | 0.295 |
| 15 | 0.266 | 0.304 | 0.338 | 0.377 | 0.404 | 30 | 0.190 | 0.218 | 0.242 | 0.270 | 0.281 |

## 1.B Appendix: Laplace Transform

If a function $h(x)$ can be obtained from some prescribed operation on a function $f(x)$, then $h(x)$ is often called a transform of $f(x)$. For example,

$$h(x) = \sqrt{2 + f(x)},$$

$$h(x) = \frac{\partial}{\partial x} f(x).$$

The Laplace transform of $f(t)$ is the function $f^*(s)$, where

$$f^*(s) = \int_0^\infty \mathrm{e}^{-st} f(t)\,\mathrm{d}t.$$

The Laplace transform is often denoted by $f^*(s)$ or $\mathcal{L}(f(t))$ or $\mathcal{L}(f)$. The results of the Laplace transform for a few simple functions are presented below.

### Results

1.
$$\mathcal{L}(1) = \int_0^\infty \mathrm{e}^{-st}\,\mathrm{d}t = \frac{1}{s};$$

2.
$$\mathcal{L}(\mathrm{e}^{-at}) = \int_0^\infty \mathrm{e}^{-st}\,\mathrm{e}^{-at}\,\mathrm{d}t = \int_0^\infty \mathrm{e}^{-(s+a)t}\,\mathrm{d}t = \frac{1}{s+a}.$$

3. If $f(t) = \frac{1}{a}\mathrm{e}^{-\frac{t}{a}}$, then
$$\mathcal{L}[f(t)] = \int_0^\infty \mathrm{e}^{-st}\frac{1}{a}\mathrm{e}^{-\frac{t}{a}}\,\mathrm{d}t = \frac{1}{1+sa}.$$

4. If $f(t) = t\,\mathrm{e}^{at}$, then
$$\mathcal{L}[f(t)] = \int_0^\infty \mathrm{e}^{-st}t\,\mathrm{e}^{at}\,\mathrm{d}t = \frac{1}{(s-a)^2}.$$

5. If $f(t) = \frac{1}{a}(\mathrm{e}^{at} - 1)$, then
$$\mathcal{L}[f(t)] = \int_0^\infty \mathrm{e}^{-st}\frac{1}{a}(\mathrm{e}^{at} - 1)\,\mathrm{d}t = \frac{1}{s(s-a)}.$$

6. If $f(t) = (1+at)\mathrm{e}^{at}$, then
$$\mathcal{L}[f(t)] = \int_0^\infty \mathrm{e}^{-st}(1+at)\mathrm{e}^{at}\,\mathrm{d}t = \frac{s}{(s-a)^2}.$$

Similarly, we can obtain the following results:

7. If $f(t)=\frac{a\mathrm{e}^{at}-b\mathrm{e}^{bt}}{a-b}$, then
$$\mathcal{L}[f(t)]=\frac{s}{(s-a)(s-b)} \quad \text{for } a\neq b.$$
8. If $f(t)=\frac{\alpha^k t^{k-1}\mathrm{e}^{-at}}{\Gamma(k)}$ then
$$\mathcal{L}[f(t)]=\left(\frac{\alpha}{\alpha+s}\right)^k.$$
9. If $f(t)=\frac{\mathrm{e}^{at}-\mathrm{e}^{bt}}{a-b}$, for $a\neq b$, then
$$\mathcal{L}[f(t)]=\frac{1}{(s-a)(s-b)}.$$
10. If $f(t)=\lambda\mathrm{e}^{-\lambda t}$, then
$$\mathcal{L}[f(t)]=\frac{\lambda}{\lambda+s}.$$
11.
$$\begin{aligned}&\mathcal{L}[c_1f_1(t)+c_2f_2(t)]\\&=\int_0^\infty \mathrm{e}^{-st}[c_1f_1(t)+c_2f_2(t)]\,\mathrm{d}t\\&=c_1\mathcal{L}[f_1(t)]+c_2\mathcal{L}[f_2(t)]\,.\end{aligned}$$
12. If $f_i(t)=\lambda_i\mathrm{e}^{-\lambda_i t}$, then
$$\mathcal{L}\left[\sum_{i=1}^n f_i(t)\right]=\sum_{i=1}^n\frac{\lambda_i}{\lambda_i+s}.$$
13.
$$\mathcal{L}\left[\sum_{i=1}^n f_i(t)\right]=\sum_{i=1}^n\mathcal{L}\,[f_i(t)].$$
14.
$$\begin{aligned}\mathcal{L}[f'(t)]&=\int_0^\infty \mathrm{e}^{-st}f'(t)\,\mathrm{d}t\\&=f(t)\mathrm{e}^{-st}\Big|_0^\infty+s\int_0^\infty f(t)\mathrm{e}^{-st}\,\mathrm{d}t\\&=-f(0^+)+sf^*(s)\\&=-f(0^+)+s\mathcal{L}[f(t)].\end{aligned}$$