Chapter 4

Probability Models in Survival Analysis

One of the central aspects of survival analysis is the investigation of the probability distribution of a random variable T which has nonnegative support. In some settings, there are covariates that influence the probability distribution of T. In addition, the data collected on the random variable T is often right censored, which means that only a lower bound is available on the value of T. So there is a bit of a mental adjustment that needs to be made from probability theory, where X is usually used to denote a random variable, to survival analysis, where T is used to denote a random variable that can only assume nonnegative values. The choice of T is made because the random variable of interest is typically *time*. This chapter introduces probability models for T.

Four lifetime distribution representations that are commonly used to define the probability distribution of a random variable T are introduced in this chapter: the *survivor function*, the *probability density function*, the *hazard function*, and the *cumulative hazard function*. These four representations apply to both continuous (for example, the lifetime of a light bulb) and discrete (for example, the lifetime of the landing gear on an airplane) lifetimes. The survival time distribution of a drill bit, an automobile, a cat, and a recession are vastly different. One would certainly not want to use the same failure time distribution with identical parameters to model these diverse lifetimes. This chapter surveys two probability distributions (the *exponential* distribution as an example of a one-parameter distribution and the *Weibull* distribution as an example of a two-parameter distribution) that are commonly used to model lifetimes. The exponential distribution is central to survival analysis just as the normal distribution is central to classical statistics. After sections that survey other lifetime distributions and moment ratio diagrams, the *Cox proportional hazards model* is introduced. The proportional hazards model is appropriate for incorporating a vector of covariates that influence survival (for example, the turning speed and feed rate for a drill bit) into a lifetime model.

4.1 Lifetime Distribution Representations

The application areas associated with the probability distribution of the nonnegative random variable T are quite wide.

- In reliability engineering, T is typically the lifetime of a component or a system of components. Examples include the lifetime of a light bulb or the lifetime of a tennis racket.

- In biostatistics, T is typically the survival time of a patient. To be more specific, this might be the survival time of a patient after a particular type of surgery. More generally, the lifetime T could be the time between the end of radiation treatment for a particular cancer and the time the cancer recurs. In other words, T is the remission time.

- In actuarial science, T is often the lifetime of an insured individual in the life insurance industry. On the casualty and property side of actuarial science, T is often the lifetime of a structure or a vehicle.

- In sociology, T can model the duration of a strike, the duration of a marriage, or the duration of a business partnership. More generally, T might model the social distance between two strangers having a conversation.

- In economics, T can be the time between recessions or the absolute change in a stock market index from one year to the next.

- In systems engineering, T could be the length of time that it takes to screen a passenger at an airport. The time that a customer spends in a slow-moving queue before exiting the queue is another nonnegative random variable that might be of interest.

- In public policy, T could be the response time by emergency vehicles to a reported building fire. Alternatively, T could be the time for a released inmate to return to prison in a recidivism application within the criminal justice system.

- In library science, T could be the time that a book is checked out. The time between an interlibrary loan request and its fulfillment is another nonnegative random variable of interest to librarians.

- In meteorology, T could be the time between the formation of a tropical storm and the time it makes landfall. The time that a severe hurricane spends as a Category 5 hurricane is another nonnegative random variable of interest to meteorologists.

- In chemistry, T could be the length of time required to complete a chemical reaction. A chemist could also use T to denote the bond length between two atoms.

The long list given above is intended to highlight that survival analysis is a field that has a very wide range of applications. Although the letter T has been selected because it most often represents time, there are many applications in which it represents something other than time (for example, social distance or bond length).

When T represents time, T can be thought of as the time between two events. For this reason, this part of survival analysis is often referred to as *time-to-event* modeling. The time of purchase and the time of failure, for example, might be the two events for a manufactured product. Since the applications of survival analysis are wide, we will use the generic terms "failure" of an "item" when referring to the second of the two events.

This section introduces four functions that define the probability distribution of a continuous, nonnegative random variable T, the lifetime of an item. The four representations presented in this chapter are not the only ways to define the distribution of T. Other methods include the moment generating function $E\left[e^{sT}\right]$, the characteristic function $E\left[e^{isT}\right]$, the Mellin transform $E\left[T^s\right]$, the mean residual life function $E[T - t \,|\, T \geq t]$, and the reversed failure rate $f(t)/F(t)$. The four representations used here have been chosen because of their intuitive appeal, usefulness in problem solving, and popularity in the literature.

4.1.1 Survivor Function

The first lifetime distribution representation is the *survivor function* $S(t)$. The survivor function is the probability that an item is functioning at any time t.

Definition 4.1 The *survivor function* for a nonnegative random variable T is

$$S(t) = P(T \geq t) \qquad t \geq 0,$$

where $S(t) = 1$ for all $t < 0$.

A survivor function is also known as the reliability function [because $S(t)$ is the reliability of an item at time t] and the complementary cumulative distribution function [because $S(t) = 1 - F(t)$ for continuous random variables, where $F(t) = P(T \leq t)$ is the cumulative distribution function]. All survivor functions must satisfy three conditions:

$$S(0) = 1 \qquad \lim_{t \to \infty} S(t) = 0 \qquad S(t) \text{ is nonincreasing.}$$

There are two interpretations of the survivor function. First, $S(t)$ is the probability that an individual item is functioning at time t. Second, if there is a large population of items with identically distributed lifetimes, $S(t)$ is the expected fraction of the population that is functioning at time t.

The survivor function is useful for comparing the survival patterns of several populations of items. The graph in Figure 4.1 shows survivor functions $S_1(t)$ and $S_2(t)$, where $S_1(t)$ corresponds to population 1 and $S_2(t)$ corresponds to population 2. Since $S_1(t) \geq S_2(t)$ for all t values, it can be concluded that the items in population 1 are superior to those in population 2 with regard to survival.

The conditional survivor function, $S_{T \mid T \geq a}(t)$, is the survivor function of an item that is functioning at time a:

$$S_{T \mid T \geq a}(t) = \frac{P(T \geq t \text{ and } T \geq a)}{P(T \geq a)} = \frac{P(T \geq t)}{P(T \geq a)} = \frac{S(t)}{S(a)} \qquad t \geq a.$$
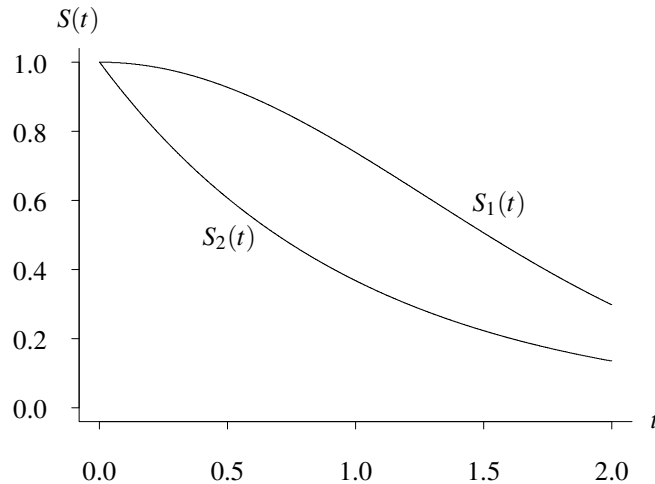
Figure 4.1: Two survivor functions.

Figure 4.2 shows the original survivor function $S(t)$ and the conditional survivor function $S_{T|T \geq a}(t)$ when $a = 0.5$. Since the conditional survivor function is rescaled by the factor $S(a)$, it has the same shape as the remaining portion of the original survivor function. The conditional survivor function is useful for comparing the survival experience of a group of items that has survived to time a. Examples include manufactured items surviving a burn-in test and cancer patients surviving 5 years after diagnosis and treatment. The conditional survivor function is of particular interest to actuaries. If a 37-year-old woman, for example, is purchasing a one-year term life insurance policy, an estimate of $S_{T|T \geq 37}(38)$ is required to determine an appropriate premium for the policy.
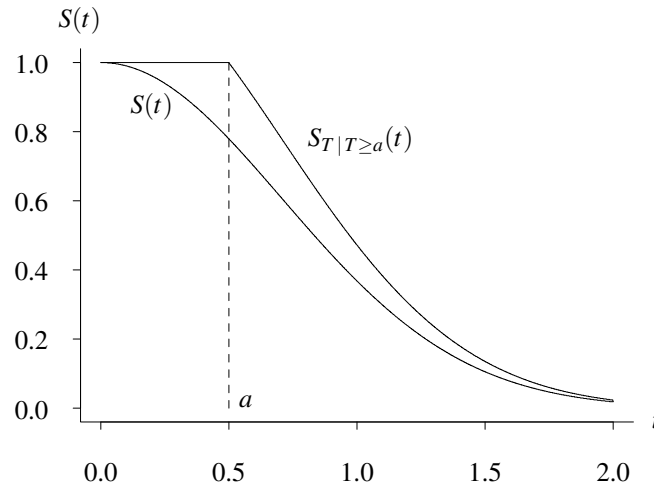

Figure 4.2: Conditional survivor function.

4.1.2 Probability Density Function

The second lifetime distribution representation is the familiar *probability density function*, which is defined as the negative of the derivative of the survivor function.

Definition 4.2 The *probability density function* of the nonnegative random variable T is

$$f(t) = -S'(t) \qquad t \geq 0,$$

where $S(t)$ is the survivor function and its derivative exists.

The probability density function has the probabilistic interpretation

$$f(t)\Delta t \cong P(t \leq T \leq t + \Delta t)$$

for small Δt values. Although the probability density function is not as effective as the survivor function for comparing the survival patterns of two populations, a graph of $f(t)$ indicates the likelihood of failure for any t. The probability of failure between times a and b is calculated by an integral:

$$P(a \leq T \leq b) = \int_a^b f(t)\,dt.$$

All probability density functions for lifetimes must satisfy two conditions:

$$\int_0^\infty f(t)\,dt = 1 \qquad\qquad f(t) \geq 0 \text{ for all } t \geq 0.$$

It is assumed that $f(t) = 0$ for all $t < 0$, which is consistent with our assumption that the random variable T is nonnegative. This assumption excludes distributions with negative support, such as the normal distribution. The probability density function shown in Figure 4.3 illustrates the relationship between the cumulative distribution function $F(t)$ and the survivor function $S(t)$ for a continuous lifetime. The area under $f(t)$ to the left of the arbitrary time t_0 is $F(t_0)$; the area under $f(t)$ to the right of t_0 is $S(t_0)$.
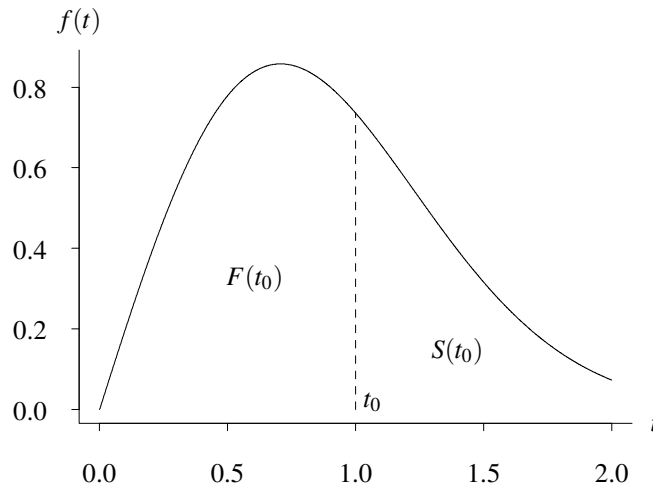

Figure 4.3: Relationship between the survivor and cumulative distribution functions.

4.1.3   Hazard Function

The *hazard function*, $h(t)$, is perhaps the most popular of the four representations for survival analysis due to its intuitive interpretation as the amount of *risk* associated with an item at time t. A second reason for its popularity is its usefulness in comparing the way risks change over time for several populations of items by plotting their hazard functions on a single axis. A third reason is that the hazard function is a special case of the intensity function for a nonhomogeneous Poisson process, which will be introduced in a subsequent chapter. A hazard function models the occurrence of one event, a failure, whereas the intensity function models the occurrence of a sequence of events over time. The hazard function goes by several aliases: in reliability it is also known as the hazard rate or failure rate; in actuarial science it is known as the force of mortality or force of decrement; in point process and extreme value theory it is known as the rate or intensity function; in vital statistics it is known as the age-specific death rate; and in economics its reciprocal is known as Mill's ratio.

The hazard function can be derived using conditional probability. First, consider the probability of failure between t and $t + \Delta t$:

$$P(t \leq T \leq t + \Delta t) = \int_t^{t+\Delta t} f(\tau)d\tau = S(t) - S(t + \Delta t).$$

Conditioning on the event that the item is working at time t yields

$$P(t \leq T \leq t + \Delta t \mid T \geq t) = \frac{P(t \leq T \leq t + \Delta t)}{P(T \geq t)} = \frac{S(t) - S(t + \Delta t)}{S(t)}.$$

If this conditional probability is averaged over the interval $[t, \, t + \Delta t]$ by dividing by Δt, an average rate of failure is obtained:

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

As $\Delta t \to 0$, this becomes the instantaneous failure rate, which is the hazard function

$$\begin{aligned}
h(t) &= \lim_{\Delta t \to 0} \frac{S(t) - S(t + \Delta t)}{S(t) \Delta t} \\
&= -\frac{S'(t)}{S(t)} \\
&= \frac{f(t)}{S(t)} \qquad t \geq 0
\end{aligned}$$

using the definition of the derivative from calculus. This forms the basis for the following definition.

Definition 4.3 The *hazard function* for a nonnegative random variable T is

$$h(t) = \frac{f(t)}{S(t)} \qquad t \geq 0,$$

where $f(t)$ is the probability density function and $S(t)$ is the survivor function.

Thus, the hazard function is the ratio of the probability density function to the survivor function. Using the previous derivation, a probabilistic interpretation of the hazard function is

$$h(t) \Delta t \cong P(t \leq T \leq t + \Delta t \mid T \geq t)$$

for small Δt values, which is a conditional version of the interpretation for the probability density function. All hazard functions must satisfy two conditions:

$$\int_0^\infty h(t) \, dt = \infty \qquad h(t) \geq 0 \text{ for all } t \geq 0.$$

Example 4.1 Consider the Weibull distribution defined by the survivor function

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

with positive scale parameter λ and positive shape parameter κ. Find the hazard function.

By differentiating the survivor function with respect to t and negating, the probability density function is

$$f(t) = \lambda \kappa (\lambda t)^{\kappa - 1} e^{-(\lambda t)^\kappa} \qquad t \geq 0,$$

so the hazard function is

$$h(t) = \frac{f(t)}{S(t)} = \lambda\kappa(\lambda t)^{\kappa-1} \qquad\qquad t \geq 0.$$

Figure 4.4 illustrates the shape of the hazard function for the Weibull distribution with $\lambda = 1$ and three κ values. The hazard function is constant when $\kappa = 1$, increasing when $\kappa > 1$, and decreasing when $\kappa < 1$.
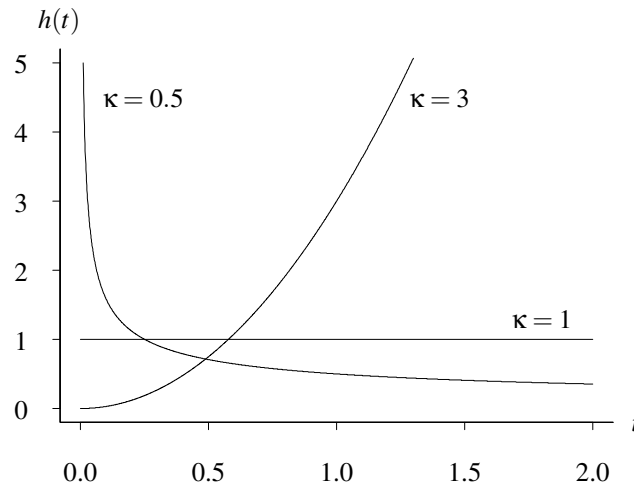

Figure 4.4: Hazard functions for the Weibull distribution.

The *units* on a hazard function are typically given in failures per unit time. In Example 4.1, if $\lambda = 0.01$, $\kappa = 1$, and time is measured in hours, then $h(t) = 0.01$ *failures per hour*. Manufactured items are often so reliable that to avoid hazard functions such as $h(t) = 0.00000128$ failures per hour the units are changed so that the hazard function may be expressed as $h(t) = 1.28$ failures per 10^6 hours. Another way to avoid writing too many leading zeroes is to change the units to years, where one year equals 8760 hours.

The shape of the hazard function indicates how an item ages. The intuitive interpretation of $h(t)$ as the amount of *risk* an item is subject to at time t implies that when the hazard function is larger the item is under greater risk of failure, and when the hazard function is smaller the item is under less risk of failure. The three hazard functions plotted in Figure 4.5 correspond to an increasing hazard function (labeled IFR for increasing failure rate), a decreasing hazard function (labeled DFR for decreasing failure rate), and a bathtub-shaped hazard function (labeled BT for bathtub-shaped failure rate).

The increasing hazard function is probably the most common situation of the three depicted in Figure 4.5. In this case, items are more likely to fail as time passes. In other words, items wear out or degrade with time. This is almost certainly the case with mechanical items that undergo wear or fatigue. It can also be the case in certain biomedical experiments. Let T, for example, be the time until a tumor appears after the injection of a substance into a laboratory animal. If the substance makes the tumor more likely to appear as time passes, then the hazard function associated with T is increasing. This leads to the formal definition of the IFR class. Notice the loose use of the term
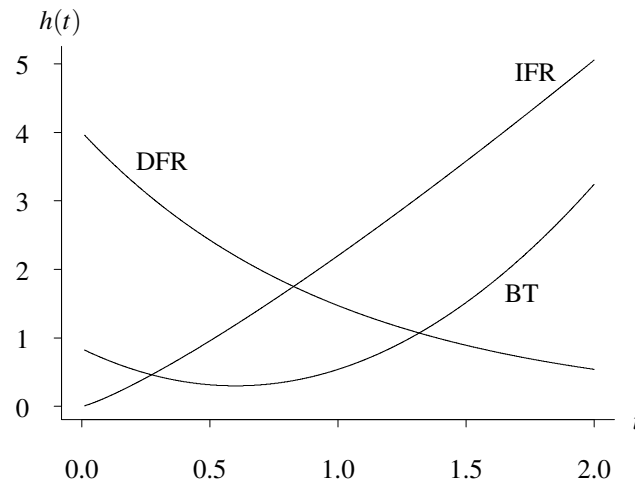
Figure 4.5: Common hazard function shapes.

increasing in the definition of the IFR class (because IFR distributions have *nondecreasing* hazard functions).

Definition 4.4 The distribution of a nonnegative random variable T is in the IFR (increasing failure rate) class if $h(t)$ is a nondecreasing function of t.

The second situation depicted in Figure 4.5, the decreasing hazard function, is less common. In this case, the item is less likely to fail as time passes. Items with this type of hazard function improve with time. Some metals work-harden through use and thus have increased strength as time passes. Another situation for which a decreasing hazard function might be appropriate for modeling is in working bugs out of computer programs. Bugs are more likely to appear initially, but the likelihood of them appearing decreases as time passes. This leads to the formal definition of the DFR class.

Definition 4.5 The distribution of a nonnegative random variable T is in the DFR (decreasing failure rate) class if $h(t)$ is a nonincreasing function of t.

The loose use of the term *increasing* in the definition of the IFR class and the term *decreasing* in the definition of the DFR class allows a distribution with a constant hazard function, the exponential distribution, to serve as a boundary between the two classes. The exponential distribution's hazard function $h(t) = \lambda$ for $t \geq 0$, is both nondecreasing and nonincreasing, so it belongs to both the IFR and DFR classes. As shown in the Venn diagram in Figure 4.6, this definition of IFR and DFR classifies all lifetime distributions into one of four sets: a constant hazard function (that is, the exponential distribution, which is the intersection of the IFR and DFR classes), strictly increasing hazard functions, strictly decreasing hazard functions, and other hazard functions (such as bathtub-shaped hazard functions).

The third situation depicted in Figure 4.5, a bathtub-shaped hazard function, occurs when the hazard function decreases initially and then increases as items age. Items improve initially and then degrade as time passes. One situation in which the bathtub-shaped hazard function arises is in the
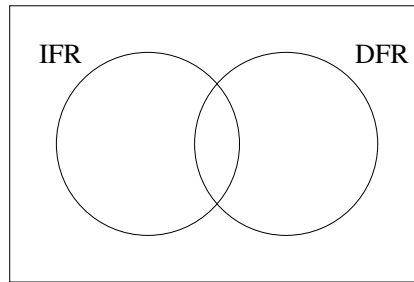
Figure 4.6: Venn diagram for IFR and DFR distribution classes.

lifetimes of manufactured items. Often manufacturing, design, or component defects cause early failures. The period in which these failures occur is sometimes called the *burn-in* period. If failure is particularly catastrophic, this part of the lifetime will often be consumed by the manufacturer in a controlled environment. The time value during which early failures have been eliminated may be valuable to a producer who is determining an appropriate warranty period. Once items pass through this early part of their lifetime, they have a fairly constant hazard function, and failures are equally likely to occur at any point in time. Finally, as items continue to age, the hazard function increases, resulting in *wear-out* failures. The three paragraphs that follow give examples of applications of the bathtub-shaped hazard function.

The bathtub-shaped hazard function can be envisioned for calculators; the burn-in period corresponds to the first few weeks of use when manufacturing, design, or component defects arise. Wear-out failures occur after a few years of use when the buttons are about ready to fall off. Failures due to calculators being dropped occur throughout the life of a calculator. If these failures are equally likely at any time, the hazard function will be increased by a constant that reflects the probability of dropping the calculator for all time values.

The bathtub-shaped hazard function also arises in the lifetimes of people. In this case, the early failures are known as *infant mortality* deaths and occur during the first few years of life. After this time, the hazard function has a very gentle increase through the teenage years and into adulthood. Finally, *old age* deaths occur during the later years of life. The magnitude of the hazard function depends on factors such as the standard of living and medical services available. Also, occupation (for example, flower arranger versus stunt man) and lifestyle (for example, eating habits, sleeping habits, smoking habits, stress level) affect the lifetime distribution of a person. The hazard function is used in actuarial science; the appropriate premium for a life insurance policy is based on probabilities associated with the lifetime distribution. The lowest life insurance premiums are usually for children who have survived the infant mortality part of their lifetimes.

There are dozens of other lifetime distribution classes beyond just the IFR and DFR classes. These include IFRA (increasing failure rate on average), DFRA (decreasing failure rate on average), IMRL (increasing mean residual life), and DMRL (decreasing mean residual life).

Care must be taken to differentiate between the hazard function for a *population* and the hazard function for an individual *item* under consideration. To use human lifetimes as an illustration, consider the following question: do two healthy 11-year-old boys living in the same town necessarily have the same hazard function? The answer is no. The reason is that all people are born with genetic predispositions that will influence their risk as they age. So, although a hazard function exists for all 11-year-old boys living in that particular town, it is an aggregate hazard function representing the population, and individual boys may be at increased or decreased risk. This is why life insurance

companies typically require a medical exam to determine whether an individual is at higher risk than the rest of the population. The common assumption in most probabilistic models and statistical analyses is that of mutually independent and identically distributed random variables, which in this case are lifetimes. This assumption is not always valid in survival analysis applications because items are often manufactured in diverse conditions (for example, different temperatures or raw materials).

4.1.4 Cumulative Hazard Function

The fourth lifetime distribution representation, the *cumulative hazard function*, is defined as the integral of the hazard function.

Definition 4.6 The *hazard function* for a nonnegative random variable T is

$$H(t) = \int_0^t h(\tau)\, d\tau \qquad t \geq 0,$$

where $h(t)$ is the hazard function.

Whereas the hazard function reflects the risk pattern associated with an item over time, the cumulative hazard function gives the accumulated risk at time t. Similar to the way a cumulative distribution function accumulates probability, the cumulative hazard function $H(t)$ accumulates the risk from time 0 to time t. All cumulative hazard functions must satisfy three conditions:

$$H(0) = 0 \qquad \lim_{t \to \infty} H(t) = \infty \qquad H(t) \text{ is nondecreasing.}$$

The cumulative hazard function is valuable for random variate generation in Monte Carlo simulation, implementing certain procedures in statistical inference, and defining certain distribution classes (for example, the IFRA class).

The four lifetime distribution representations presented here are equivalent in the sense that each completely specifies a lifetime distribution. In addition, any one lifetime distribution representation implies the other three. Algebra and calculus can be used to find one lifetime distribution representation given that another is known. For example, if the survivor function is known, the cumulative hazard function can be determined by

$$H(t) = \int_0^t h(\tau)\, d\tau = \int_0^t \frac{f(\tau)}{S(\tau)}\, d\tau = -\ln S(t),$$

where ln is the natural logarithm (log base e). The from–to matrix in Table 4.1 shows that any of the three other lifetime distribution representations (given by the columns) can be found if one of the representations (given by the rows) is known. It is assumed that the support of the lifetime T is $[0, \infty)$ in Table 4.1.

Example 4.2 Given $h(t) = 18t$ for $t \geq 0$, find $f(t)$.

Using the $\big(h(t),\, f(t)\big)$ element of the from–to matrix in Table 4.1,

$$f(t) = h(t)e^{-\int_0^t h(\tau)\, d\tau}$$
$$= 18t\, e^{-\int_0^t 18\tau\, d\tau}$$
$$= 18t\, e^{-9t^2} \qquad t \geq 0,$$

which is a special case of the Weibull distribution with $\lambda = 3$ and $\kappa = 2$.

	$f(t)$	$S(t)$	$h(t)$	$H(t)$
$f(t)$	$\bullet$	$\displaystyle\int_t^\infty f(\tau)d\tau$	$\dfrac{f(t)}{\displaystyle\int_t^\infty f(\tau)d\tau}$	$-\ln\left[\displaystyle\int_t^\infty f(\tau)d\tau\right]$
$S(t)$	$-S'(t)$	$\bullet$	$\dfrac{-S'(t)}{S(t)}$	$-\ln S(t)$
$h(t)$	$h(t)\,e^{-\int_0^t h(\tau)d\tau}$	$e^{-\int_0^t h(\tau)d\tau}$	$\bullet$	$\displaystyle\int_0^t h(\tau)d\tau$
$H(t)$	$H'(t)\,e^{-H(t)}$	$e^{-H(t)}$	$H'(t)$	$\bullet$

Table 4.1: Lifetime distribution representation relationships.

4.2   Exponential Distribution

Just as the normal distribution plays a pivotal role in classical statistics because of the central limit theorem, the exponential distribution plays a pivotal role in survival analysis because it is the only continuous distribution with a constant hazard function. The exponential distribution has a single positive scale parameter λ, often called the *failure rate* by reliability engineers.

Definition 4.7 The four lifetime distribution representations associated with a random variable T having the exponential distribution with positive rate parameter λ are

$$S(t) = e^{-\lambda t} \qquad f(t) = \lambda e^{-\lambda t} \qquad h(t) = \lambda \qquad H(t) = \lambda t$$

for $t \geq 0$. Symbolically, this is written as $T \sim \text{exponential}(\lambda)$.

The four lifetime distribution representations are plotted in Figure 4.7 for $\lambda = 1$ and $\lambda = 2$. Two-parameter distributions, which are more complex but can model a wider variety of situations, are presented in subsequent sections.

The centrality, tractability, and importance of the exponential distribution make it a key probability distribution to know well. In that light, this section surveys several probabilistic properties
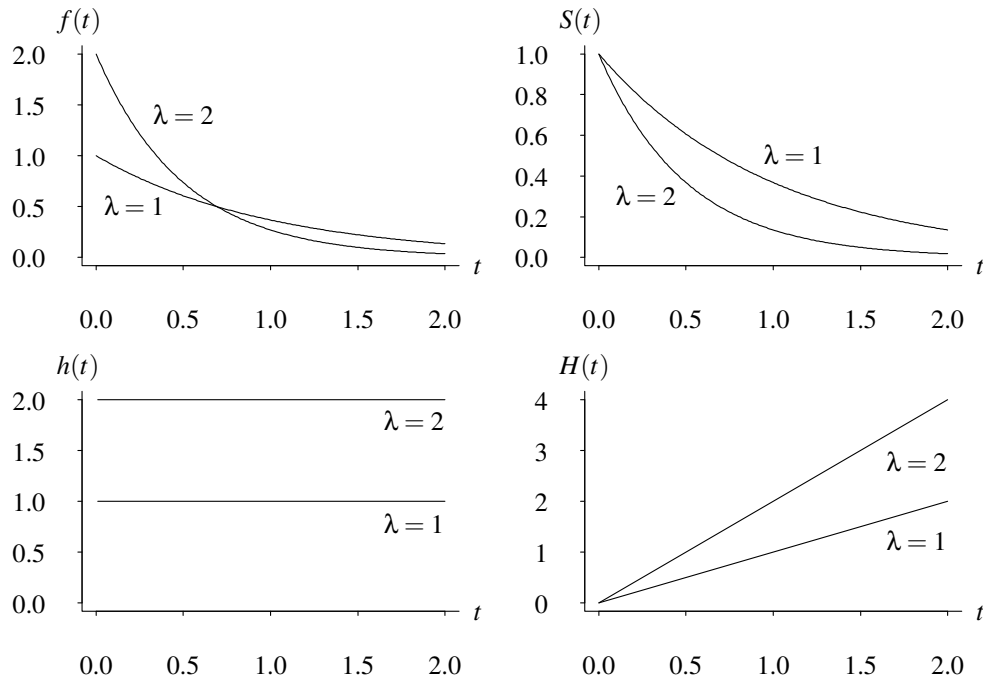
Figure 4.7: Lifetime distribution representations for the exponential distribution.

of the exponential distribution that are useful in understanding how it is unique and when it should be applied. In all the properties, it is assumed that the nonnegative lifetime T has the exponential distribution with parameter λ.

Theorem 4.1 (memoryless property) If $T \sim \text{exponential}(\lambda)$, then

$$P(T \geq t) = P(T \geq t + s \,|\, T \geq s) \qquad t \geq 0; \; s \geq 0.$$

Proof The probability of surviving to time $t + s$ conditioned on survival to time s is

$$
\begin{aligned}
P(T \geq t + s \,|\, T \geq s) &= \frac{P(T \geq t + s)}{P(T \geq s)} \\
&= \frac{S(t + s)}{S(s)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} \\
&= e^{-\lambda t} \\
&= S(t) \\
&= P(T \geq t)
\end{aligned}
$$

for all $t \geq 0$ and $s \geq 0$. $\qquad\qquad\square$

As shown in Figure 4.8 for $\lambda = 1$ and $s = 0.5$, this result indicates that the conditional survivor function for the lifetime of an item that has survived to time s is identical to the survivor function for the lifetime of a brand new item. This used-as-good-as-new assumption is very strong. Consider, for example, whether the exponential distribution should be used to model the lifetime of a candle with an expected burning time of 5 hours. If several candles are sampled and burned, we could imagine a bell-shaped histogram for candle lifetimes, centered around 5 hours. The exponential lifetime model is certainly *not* appropriate in this case, because a candle that has burned for 4 hours does not have the same remaining lifetime distribution as that of a brand new candle. The exponential distribution would only be appropriate for candle lifetimes if the remaining lifetime of a used candle is identical to the lifetime of a new candle. An electrical component for which the exponential lifetime assumption might be justified is a fuse. A fuse is designed to fail when there is a power surge that causes the fuse to fail, resulting in a blown fuse which must be replaced. Assuming that the fuse does not undergo any weakening or degradation over time and that power surges that cause failure occur at a constant rate over time, the exponential lifetime assumption is appropriate, and a used fuse that has not failed is as good as a new one in terms of longevity.
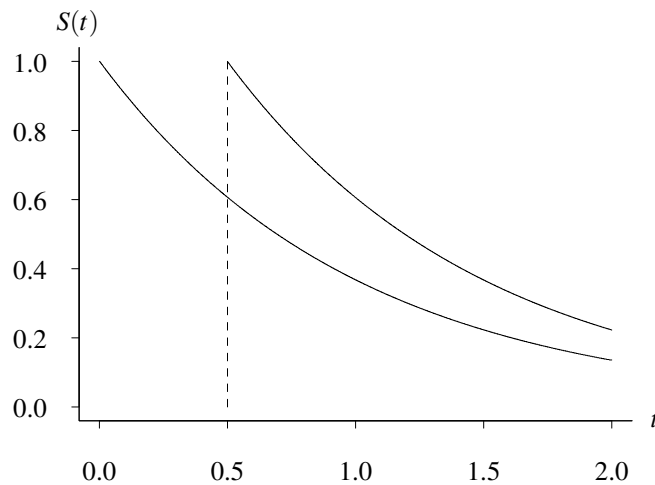
Figure 4.8: The memoryless property of the exponential distribution.

The exponential distribution should be applied judiciously because the memoryless property restricts its applicability. It is often misapplied for the sake of simplicity because the statistical techniques for the exponential distribution are particularly tractable, or because small sample sizes do not support more than a one-parameter distribution.

The exponential distribution is the only continuous distribution with the memoryless property. The exponential distribution is the only continuous lifetime distribution for which the conditional lifetime distribution of a used item is identical to the original lifetime distribution. The only discrete distribution with the memoryless property is the geometric distribution.

Theorem 4.2 If T is a continuous nonnegative random variable with cumulative hazard function $H(t)$, then $H(T) \sim \text{exponential}(1)$.

Proof The survivor function for λT is

$$
\begin{aligned}
P(\lambda T \geq t) &= P(T \geq t/\lambda) \\
&= e^{-\lambda(t/\lambda)} \\
&= e^{-t} \qquad\qquad t \geq 0,
\end{aligned}
$$

so λT has survivor function e^{-t}, which is exponential(1). $\qquad\square$

This property is mathematically equivalent to the probability integral transformation, which states that $F(T) \sim U(0,1)$, resulting in the inverse-cdf technique for generating random variates for Monte Carlo simulation: $T \leftarrow F^{-1}(U)$, where $U \sim U(0,1)$. Using Theorem 4.2, random lifetime variates are generated by

$$
T \leftarrow H^{-1}\big(-\ln(1-U)\big)
$$

because $-\ln(1-U)$ is a unit exponential random variate. Random lifetimes generated in this fashion are generated by the *cumulative hazard function technique*.

Example 4.3 Assuming that the failure time of an item has the Weibull distribution with survivor function

$$
S(t) = e^{-(\lambda t)^\kappa} \qquad\qquad t \geq 0
$$

for positive scale parameter λ and positive shape parameter κ, find an equation to convert $U(0,1)$ random numbers to Weibull random variates.

The cumulative hazard function for the Weibull distribution is

$$
H(t) = -\ln S(t) = (\lambda t)^\kappa \qquad\qquad t \geq 0,
$$

which has inverse

$$
H^{-1}(y) = \frac{y^{1/\kappa}}{\lambda} \qquad\qquad y \geq 0.
$$

Weibull random variates can be generated by

$$
T \leftarrow \frac{1}{\lambda}\left[-\ln(1-U)\right]^{1/\kappa},
$$

where U is uniformly distributed between 0 and 1.

Figure 4.9 illustrates the geometry associated with generating a variate from the cumulative hazard function. The value of $-\ln(1-U)$, the unit exponential random variate, is indicated on the vertical axis, and the corresponding random variate T is indicated on the horizontal axis.

The next result gives a general expression for the sth moment of an exponential random variable.

Theorem 4.3 If $T \sim$ exponential(λ), then

$$
E[T^s] = \frac{\Gamma(s+1)}{\lambda^s} \qquad\qquad s > -1,
$$

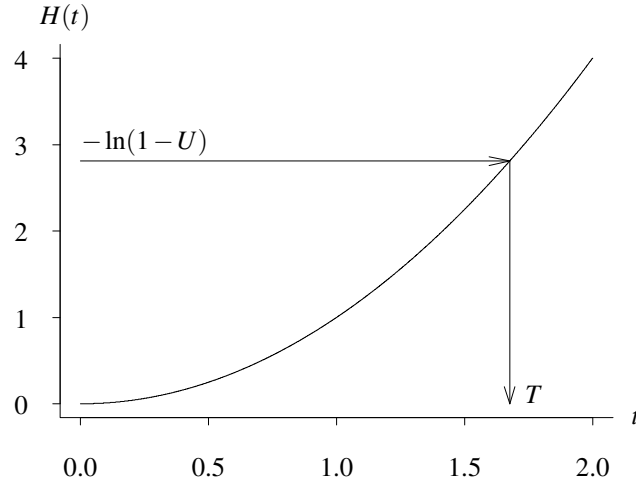where $\Gamma(\alpha) = \displaystyle\int_0^\infty x^{\alpha-1} e^{-x} \, dx.$

Figure 4.9: Generating a random variate by the inverse cumulative hazard function technique.

Proof Using the substitution $x = \lambda t$, the expected value of T^s is

$$E[T^s] = \int_0^\infty t^s \lambda e^{-\lambda t}\, dt$$
$$= \lambda^{-s} \int_0^\infty x^s e^{-x}\, dx$$
$$= \lambda^{-s} \Gamma(s+1) \qquad s > -1. \qquad \square$$

When s is a nonnegative integer, this expression reduces to $E[T^s] = s!/\lambda^s$. By setting $s = 1, 2, 3,$ and 4, the population mean, variance, coefficient of variation, skewness, and kurtosis can be obtained:

$$E[T] = \frac{1}{\lambda} \qquad V[T] = \frac{1}{\lambda^2} \qquad \gamma = 1 \qquad \gamma_3 = 2 \qquad \gamma_4 = 9.$$

Since the coefficient of variation of an exponential random variable is 1, a quick check for exponentiality for a data set is to see if the ratio of the sample standard deviation to the sample mean is approximately 1. The histogram of the sample values should also have the appropriate shape.

Theorem 4.4 (self-reproducing)   If $T_1, T_2, \ldots, T_n$ are mutually independent random variables, $T_i \sim \text{exponential}(\lambda_i)$, for $i = 1, 2, \ldots, n$, and $T = \min\{T_1, T_2, \ldots, T_n\}$, then

$$T \sim \text{exponential}\left(\sum_{i=1}^n \lambda_i\right).$$

Proof The survivor function for T is

$$S_T(t) = P(T \geq t)$$

$$
\begin{aligned}
&= P(\min\{T_1, T_2, \ldots, T_n\} \geq t) \\
&= P(T_1 \geq t, T_2 \geq t, \ldots, T_n \geq t) \\
&= P(T_1 \geq t) P(T_2 \geq t) \ldots P(T_n \geq t) \\
&= e^{-\lambda_1 t} e^{-\lambda_2 t} \ldots e^{-\lambda_n t} \\
&= e^{-\sum_{i=1}^{n} \lambda_i t} \qquad\qquad t \geq 0.
\end{aligned}
$$

Therefore, $T \sim$ exponential $\left(\sum_{i=1}^{n} \lambda_i\right)$. $\qquad\qquad\qquad\qquad\qquad\qquad$ □

This result indicates that the minimum of n exponential random lifetimes also has the exponential distribution. This is important in two applications. First, if n components, each with mutually independent exponential times to failure, are arranged in a series system, then the distribution of the system failure time is also exponential with a failure rate equal to the sum of the component failure rates. Second, when there are several mutually independent, exponentially distributed *causes* of failure competing for the lifetime of an item (for example, failing by open or short circuit for an electronic item or death by various diseases for a human being), then the lifetime can be modeled as the minimum of the individual lifetimes from each cause of failure.

Theorem 4.5 If $T_1, T_2, \ldots, T_n$ are mutually independent and identically distributed exponential(λ) random variables, then

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

where $\chi^2(2n)$ denotes the chi-square distribution with $2n$ degrees of freedom.

Proof Since $T_1, T_2, \ldots, T_n$ are mutually independent and identically distributed exponential(λ) random variables,

$$
\sum_{i=1}^{n} T_i \sim \text{Erlang}(\lambda, n).
$$

Furthermore

$$
\lambda \sum_{i=1}^{n} T_i \sim \text{Erlang}(1, n),
$$

which implies that

$$
2\lambda \sum_{i=1}^{n} T_i \sim \chi^2(2n). \qquad\qquad\qquad\qquad\qquad □
$$

This property is useful for determining a confidence interval for λ based on a data set of n mutually independent exponential(λ) lifetimes. With probability $1 - \alpha$,

$$
\chi^2_{2n,\, 1-\alpha/2} < 2\lambda \sum_{i=1}^{n} T_i < \chi^2_{2n,\, \alpha/2},
$$

where the left- and right-hand sides of this inequality are the $\alpha/2$ and $1 - \alpha/2$ fractiles of the chi-square distribution with $2n$ degrees of freedom. This notation is illustrated in Figure 4.10, with the three areas under the probability density function of the chi-square random variable plotted on the
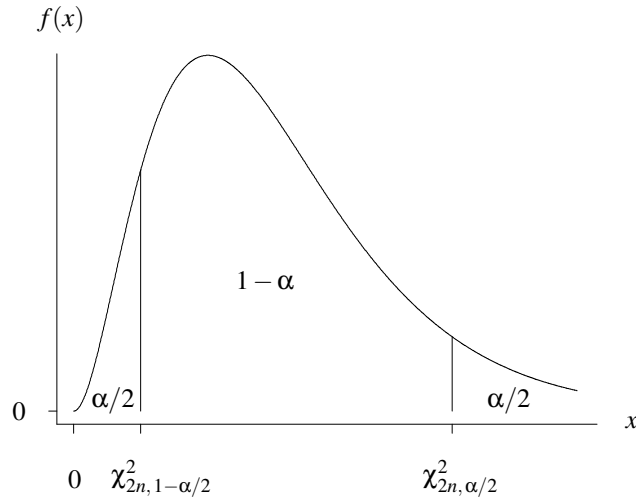
Figure 4.10: Fractiles of the chi-square distribution with $2n$ degrees of freedom.

graph. Rearranging this expression yields an exact $100(1-\alpha)\%$ two-sided confidence interval for λ:

$$\frac{\chi^2_{2n,\,1-\alpha/2}}{2\displaystyle\sum_{i=1}^{n} T_i} < \lambda < \frac{\chi^2_{2n,\,\alpha/2}}{2\displaystyle\sum_{i=1}^{n} T_i}.$$

Theorem 4.6 If $T \sim \text{exponential}(\lambda)$, then $\lceil T \rceil \sim \text{Geometric}\left(1 - e^{-\lambda}\right)$.

Proof Let $N = \lceil T \rceil$. The ceiling function applied to the continuous random variable T means that the random variable N is discrete. Since the support of T is $t \geq 0$, the support of N is $n = 1, 2, \ldots$. The probability mass function of N is

$$f_N(n) = P(N = n) = P(n-1 \leq T < n) = \int_{n-1}^{n} \lambda e^{-\lambda t}\, dt = \left[-e^{-\lambda t}\right]_{n-1}^{n} = e^{-(n-1)\lambda} - e^{-n\lambda}$$

for $n = 1, 2, \ldots$. Equivalently,

$$f_N(n) = e^{-(n-1)\lambda}\left(1 - e^{-\lambda}\right) \qquad n = 1, 2, \ldots,$$

which can be recognized as a $\text{Geometric}\left(1 - e^{-\lambda}\right)$ random variable with support beginning at 1.   □

This property involves the only two probability distributions with the memoryless property. The ceiling function returns the next highest integer associated with the continuous failure time T. A modeling situation in which this property might be of interest occurs when an item with an exponential(λ) failure time distribution is placed on test at noon on a particular day. (The item could be a manufactured item such as a light bulb in the reliability setting or a subject such as a laboratory animal with cancer in a biostatistics setting.) Rather than continuously monitoring the item in order

to observe its failure time T, you instead check the item for failure each subsequent day at noon. If time is measured in days, then the day number in which you observe failure is $\lceil T \rceil$, which the property indicates has a geometric distribution. Data collected in this fashion is known as *current status data*. A time to failure is known as *interval censored* when only a lower bound and upper bound are known on a failure time, as is the case in the scenario presented here.

The exponential distribution, for which the item under study does not age in a probabilistic sense, is the simplest of the lifetime models. Three are many other important properties of the exponential distribution in addition to those presented in this section. The two-parameter Weibull distribution, which includes the exponential distribution as a special case, is presented next. It is more flexible for modeling, although more complex mathematically.

4.3   Weibull Distribution

The exponential distribution is limited in applicability because of the memoryless property. The assumption that a lifetime has a constant failure rate is often too restrictive or inappropriate. Mechanical items, for instance, typically degrade over time and hence their lifetimes are more likely to follow a probability distribution with a strictly increasing hazard function. The Weibull distribution, named after Swedish mathematician Waloddi Weibull, is a generalization of the exponential distribution that is appropriate for modeling lifetimes having constant, strictly increasing, or strictly decreasing hazard functions.

Definition 4.8 The four lifetime distribution representations associated with a random variable T having the Weibull distribution with positive scale parameter λ and positive shape parameter κ are

$$S(t) = e^{-(\lambda t)^{\kappa}} \qquad f(t) = \kappa \lambda^{\kappa} t^{\kappa-1} e^{-(\lambda t)^{\kappa}} \qquad h(t) = \kappa \lambda^{\kappa} t^{\kappa-1} \qquad H(t) = (\lambda t)^{\kappa}$$

for $t \geq 0$. Symbolically, this is written as $T \sim \text{Weibull}(\lambda, \kappa)$.

The first four lifetime distribution representations for the Weibull(λ, κ) distribution are for $t \geq 0$, where $\lambda > 0$ and $\kappa > 0$ are the scale and shape parameters of the distribution. The hazard function approaches zero from infinity for $\kappa < 1$, is constant for $\kappa = 1$, the exponential case, and increases from zero for $\kappa > 1$. One other special case occurs when $\kappa = 2$, commonly known as the *Rayleigh distribution*, which has a linear hazard function with slope $2\lambda^2$. When $3 < \kappa < 4$, the probability density function resembles that of a normal probability density function, and the mode and median of the distribution are equal when $\kappa \cong 3.26$. The R code for plotting these lifetime distribution representations for $\lambda = 1$ and $\kappa = 0.5, 1, 2, 3$ is given below. The `by` argument in the call to the `seq` function controls the spacing between the t values plotted. The `matplot` function plots several functions on a single plot simultaneously.

```
par(mfrow = c(2, 2))
kappa = c(0.5, 1, 2, 3)
t     = seq(0, 1.5, by = 0.05)
f     = cbind(dweibull(t, kappa[1]), dweibull(t, kappa[2]),
              dweibull(t, kappa[3]), dweibull(t, kappa[4]))
matplot(t, f, type = "l")
S     = cbind(1 - pweibull(t, kappa[1]), 1 - pweibull(t, kappa[2]),
              1 - pweibull(t, kappa[3]), 1 - pweibull(t, kappa[4]))
matplot(t, S, type = "l")
```

```
h      = f / S
matplot(t, h, type = "l")
H      = -log(S)
matplot(t, H, type = "l")
```

These four functions are plotted in Figure 4.11 for $\lambda = 1$ and $\kappa = 0.5, 1, 2, 3$.

The *characteristic life* of the Weibull distribution is a special fractile defined by $t_c = 1/\lambda$. All Weibull survivor functions pass through the point $(1/\lambda, 1/e)$, regardless of the value of κ, as shown in Figure 4.11 for $\lambda = 1$. Also, since $H(t) = -\ln S(t)$, all Weibull cumulative hazard functions pass through the point $(1/\lambda, 1)$, regardless of the value of κ.

There are several ways to *parameterize* the Weibull distribution. The previous two paragraphs introduced one such parameterization with a scale parameter λ and a shape parameter κ. Another common way to parameterize the Weibull distribution is with the survivor function

$$S(t) = e^{-(t/\eta)^\beta} \qquad t \geq 0,$$

where η is a positive scale parameter and β is a positive shape parameter. This is the parameterization used in R. Comparing the two survivor functions, it is clear that the two shape parameters κ and β play identical roles, and the two scale parameters λ and η are reciprocals. Both parameterizations correspond to the Weibull distribution, but some careful bookkeeping is necessary to account for the different roles of the various parameters. The version of the Weibull distribution with parameters λ and κ will be used consistently throughout this book.
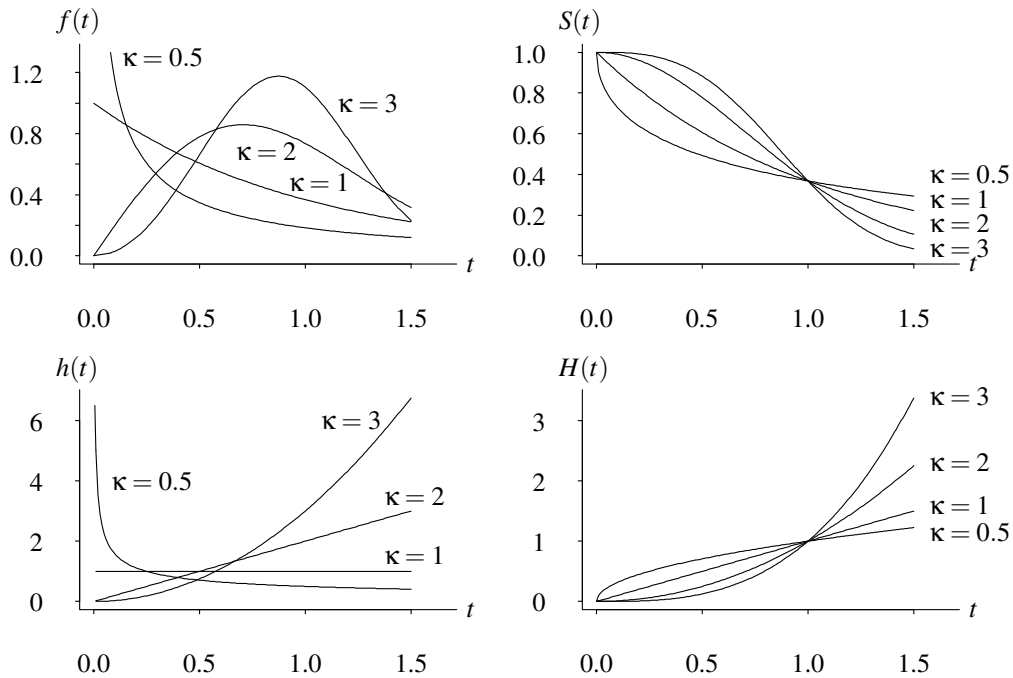

Figure 4.11: Lifetime distribution representations for the Weibull distribution for $\lambda = 1$.

Moments for the Weibull distribution are not as mathematically tractable as those for the exponential distribution. Using the substitution $u = (\lambda t)^\kappa$, the rth central moment about the origin is

$$
\begin{aligned}
E\left[T^r\right] &= \int_0^\infty t^r f(t)\, dt \\
&= \int_0^\infty \kappa \lambda^\kappa t^{\kappa+r-1} e^{-(\lambda t)^\kappa}\, dt \\
&= \frac{1}{\lambda^r} \int_0^\infty u^{r/\kappa} e^{-u}\, du \\
&= \frac{1}{\lambda^r} \Gamma\left(1 + \frac{r}{\kappa}\right) \\
&= \frac{r}{\kappa \lambda^r} \Gamma\left(\frac{r}{\kappa}\right)
\end{aligned}
$$

for $r = 1, 2, \ldots$. Using this expression for $E\left[T^r\right]$ with $r = 1$ and $r = 2$ and the shortcut formula for the population variance, the population mean and variance of a Weibull(λ, κ) random variable are

$$
E[T] = \frac{1}{\lambda} \Gamma\left(1 + \frac{1}{\kappa}\right) = \frac{1}{\lambda \kappa} \Gamma\left(\frac{1}{\kappa}\right)
$$

and

$$
V[T] = \frac{1}{\lambda^2}\left\{ \Gamma\left(1 + \frac{2}{\kappa}\right) - \left[\Gamma\left(1 + \frac{1}{\kappa}\right)\right]^2 \right\} = \frac{1}{\lambda^2}\left\{ \frac{2}{\kappa}\Gamma\left(\frac{2}{\kappa}\right) - \left[\frac{1}{\kappa}\Gamma\left(\frac{1}{\kappa}\right)\right]^2 \right\}.
$$

The associated coefficient of variation is

$$
\gamma = \frac{\sigma}{\mu} = \frac{\left\{ \frac{2}{\kappa}\Gamma\left(\frac{2}{\kappa}\right) - \left[\frac{1}{\kappa}\Gamma\left(\frac{1}{\kappa}\right)\right]^2 \right\}^{1/2}}{\frac{1}{\kappa}\Gamma\left(\frac{1}{\kappa}\right)}.
$$

Using this expression for $E\left[T^r\right]$ with $r = 3$ and $r = 4$ yields the population skewness and kurtosis:

$$
\gamma_3 = \left\{ \frac{2}{\kappa}\Gamma\left(\frac{2}{\kappa}\right) - \left[\frac{1}{\kappa}\Gamma\left(\frac{1}{\kappa}\right)\right]^2 \right\}^{-3/2} \left\{ \frac{3}{\kappa}\Gamma\left(\frac{3}{\kappa}\right) - \frac{6}{\kappa^2}\Gamma\left(\frac{1}{\kappa}\right)\Gamma\left(\frac{2}{\kappa}\right) + 2\left[\frac{1}{\kappa}\Gamma\left(\frac{1}{\kappa}\right)\right]^3 \right\},
$$

$$
\begin{aligned}
\gamma_4 = \left\{ \frac{2}{\kappa}\Gamma\left(\frac{2}{\kappa}\right) - \left[\frac{1}{\kappa}\Gamma\left(\frac{1}{\kappa}\right)\right]^2 \right\}^{-2} &\left\{ \frac{4}{\kappa}\Gamma\left(\frac{4}{\kappa}\right) - \frac{12}{\kappa^2}\Gamma\left(\frac{1}{\kappa}\right)\Gamma\left(\frac{3}{\kappa}\right) \right. \\
&\left. + \frac{12}{\kappa^3}\left[\Gamma\left(\frac{1}{\kappa}\right)\right]^2\Gamma\left(\frac{2}{\kappa}\right) - \frac{3}{\kappa^4}\left[\Gamma\left(\frac{1}{\kappa}\right)\right]^4 \right\}.
\end{aligned}
$$

The next example applies the formulas developed thus far for the Weibull distribution to the lifetime of a spring.

Example 4.4 The lifetime of a certain type of spring used continuously under known operating conditions has the Weibull distribution with $\lambda = 0.0014$ and $\kappa = 1.28$, where time is measured in hours. (Estimating the parameters for the Weibull distribution from a data set is introduced in the next chapter; the parameters are assumed to be known constants in this example.)

- Find the population mean time to failure.

- Find the probability that a new spring will operate for 400 hours.

- Find the probability that a spring that has operated for 200 hours without failure will operate another 400 hours.

The population mean time to failure is

$$\mu = E[T] = \frac{1}{(0.0014)(1.28)} \Gamma\left(\frac{1}{1.28}\right) \cong 661.8 \text{ hours.}$$

The probability that a new spring will operate for 400 hours is

$$S(400) = e^{-[(0.0014)(400)]^{1.28}} \cong 0.6222.$$

To calculate the conditional probability that a used spring lasts another 400 hours requires a conditional survivor function. The conditional survivor function for a spring that has operated for 200 hours is

$$S_{T\,|\,T\geq 200}(t) = \frac{S(t)}{S(200)} = \frac{e^{-(0.0014t)^{1.28}}}{e^{-[(0.0014)(200)]^{1.28}}} \qquad t \geq 200.$$

So the conditional probability that a spring that has operated for 200 hours lasts another 400 hours is $S_{T\,|\,T\geq 200}(600) \cong 0.5469$, as illustrated in Figure 4.12. It is not surprising that this conditional survival probability is slightly lower than the probability that a new spring survives 400 hours. Since the shape parameter $\kappa = 1.28$ is greater than 1, the spring's lifetime is in the IFR class, which means that the spring degrades over time.
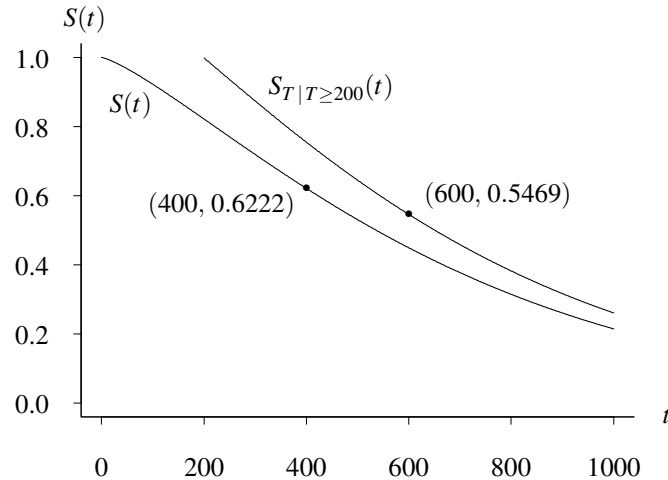

Figure 4.12: The original and conditional survivor functions.

Finding fractiles of the Weibull distribution requires only a few steps of algebra. The pth fractile of a Weibull(λ, κ) random variable, denoted by t_p, can be found by equating the cumulative

distribution of T to p and solving for t_p:

$$p = 1 - e^{-(\lambda t_p)^\kappa} \qquad \Rightarrow \qquad t_p = \frac{1}{\lambda}\left[-\ln(1-p)\right]^{1/\kappa}.$$

These fractiles can be useful for establishing warranty periods or setting burn-in periods for manufactured items.

The Weibull distribution has the self-reproducing property. If $T_1, T_2, \ldots, T_n$ are mutually independent component lifetimes having the Weibull distribution with the same shape parameters, then the minimum of these values has the Weibull distribution. More specifically, if $T_i \sim$ Weibull(λ_i, κ) for $i = 1, 2, \ldots, n$, then $\min\{T_1, T_2, \ldots, T_n\} \sim$ Weibull$\left((\sum_{i=1}^n \lambda_i^\kappa)^{1/\kappa}, \kappa\right)$.

4.4 Other Lifetime Distributions

Although the exponential and Weibull distributions are popular lifetime models, they are limited in their modeling capability. For example, if it were determined that an item had a bathtub-shaped hazard function, none of these three models would be appropriate unless a piecewise model over time segments of the lifetime were used. Several other models that may be used to describe the distribution of a continuous lifetime T are surveyed in this section.

The distributions for the nonnegative random variable T described here have three types of parameters: location parameters, denoted by a, b, and μ; scale parameters, denoted by λ and σ; and shape parameters, denoted by κ, γ, and δ. All distributions have support on $[0, \infty)$ except for the uniform and Pareto distributions.

The one-parameter lifetime models that are summarized in this section are the exponential(λ) and Muth(κ) distributions. The two-parameter lifetime models summarized here are the Weibull(λ, κ), gamma(λ, κ), uniform(a, b), log normal(μ, σ), log logistic(λ, κ), inverse Gaussian(λ, μ), exponential power(λ, κ), Pareto(λ, κ), and Gompertz(δ, κ) distributions. The three-parameter lifetime models summarized here are the Makeham(δ, κ, γ), IDB(δ, κ, γ), and generalized Pareto(δ, κ, γ) distributions. The n-parameter lifetime models summarized here are the hypoexponential$(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and hyperexponential$(\lambda_1, \lambda_2, \ldots, \lambda_n)$ distributions.

The shapes of the lifetime distribution representations, particularly the hazard function, are useful in determining the appropriate distribution to use to model a lifetime. One-, two-, three-, and n-parameter lifetime distributions are described consecutively in the following paragraphs.

4.4.1 Some One-Parameter Lifetime Models

The one-parameter lifetime distributions defined here have their $f(t)$, $S(t)$, $h(t)$, and $H(t)$ functions given in Table 4.2. The simplest lifetime distribution is the *exponential distribution*, with a positive scale parameter λ. As indicated in Section 4.2, it is the only continuous distribution with a constant failure rate.

Muth developed a distribution with a single shape parameter κ $(0 < \kappa \le 1)$. The *Muth distribution* is asymptotically equivalent to the unit exponential distribution as $\kappa \to 0$ and has a hazard function that increases from $h(0) = 1 - \kappa$, for all κ.

4.4.2 Some Two-Parameter Lifetime Models

The two-parameter lifetime distributions defined here have their lifetime distribution representations given in Table 4.3. As outlined in Section 4.3, the *Weibull distribution*, having positive scale parameter λ and positive shape parameter κ, is one of the most popular two-parameter lifetime models

Distribution	$f(t)$	$S(t)$	$h(t)$	$H(t)$	Parameters
Exponential	$\lambda e^{-\lambda t}$	$e^{-\lambda t}$	λ	λt	$\lambda > 0$
Muth	$(e^{\kappa t} - \kappa)\, e^{\left[-\frac{1}{\kappa}e^{\kappa t}+\kappa t+\frac{1}{\kappa}\right]}$	$e^{\left[-\frac{1}{\kappa}e^{\kappa t}+\kappa t+\frac{1}{\kappa}\right]}$	$e^{\kappa t} - \kappa$	$\frac{1}{\kappa}e^{\kappa t} - \kappa t - \frac{1}{\kappa}$	$0 < \kappa \le 1$

Table 4.2: One-parameter univariate lifetime distributions.

used in survival analysis. The Weibull distribution includes the exponential distribution as a special case when $\kappa = 1$, and the hazard function increases from zero to infinity when $\kappa > 1$ and decreases from infinity to zero when $\kappa < 1$.

The *gamma distribution*, has positive scale parameter λ and positive shape parameter κ. As with

Distribution	$f(t)$	$S(t)$	$h(t)$	$H(t)$	Parameters
Weibull	$\kappa\lambda^{\kappa}t^{\kappa-1}e^{-(\lambda t)^{\kappa}}$	$e^{-(\lambda t)^{\kappa}}$	$\kappa\lambda^{\kappa}t^{\kappa-1}$	$(\lambda t)^{\kappa}$	$\lambda > 0; \ \kappa > 0$
Gamma	$\dfrac{\lambda(\lambda t)^{\kappa-1}e^{-\lambda t}}{\Gamma(\kappa)}$	$1 - I(\kappa, \lambda t)$	$\dfrac{\lambda(\lambda t)^{\kappa-1}e^{-\lambda t}}{\Gamma(\kappa)[1-I(\kappa, \lambda t)]}$	$-\ln[1-I(\kappa, \lambda t)]$	$\lambda > 0; \ \kappa > 0$
Uniform	$\dfrac{1}{b-a}$	$\dfrac{b-t}{b-a}$	$\dfrac{1}{b-t}$	$-\ln\left(\dfrac{b-t}{b-a}\right)$	$a \le t \le b;$ $0 \le a < b$
Log normal	$\dfrac{1}{\sigma t\sqrt{2\pi}}e^{-(\ln t-\mu)^2/2\sigma^2}$	$\displaystyle\int_t^{\infty} f(\tau)d\tau$	$\dfrac{f(t)}{S(t)}$	$-\ln S(t)$	$-\infty < \mu < \infty;$ $\sigma > 0$
Log logistic	$\dfrac{\lambda\kappa(\lambda t)^{\kappa-1}}{[1+(\lambda t)^{\kappa}]^2}$	$\dfrac{1}{1+(\lambda t)^{\kappa}}$	$\dfrac{\lambda\kappa(\lambda t)^{\kappa-1}}{1+(\lambda t)^{\kappa}}$	$\ln\left[1+(\lambda t)^{\kappa}\right]$	$\lambda > 0; \ \kappa > 0$
Inverse Gaussian	$\sqrt{\dfrac{\lambda}{2\pi t^3}}e^{-\lambda(t-\mu)^2/2\mu^2 t}$	$\displaystyle\int_t^{\infty} f(\tau)d\tau$	$\dfrac{f(t)}{S(t)}$	$-\ln S(t)$	$\lambda > 0; \ \mu > 0$
Exponential Power	$\lambda\kappa t^{\kappa-1}e^{1-e^{\lambda t^{\kappa}}+\lambda t^{\kappa}}$	$e^{1-e^{\lambda t^{\kappa}}}$	$e^{\lambda t^{\kappa}}\lambda\kappa t^{\kappa-1}$	$e^{\lambda t^{\kappa}} - 1$	$\lambda > 0; \ \kappa > 0$
Pareto	$\dfrac{\kappa\lambda^{\kappa}}{t^{\kappa+1}}$	$\left(\dfrac{\lambda}{t}\right)^{\kappa}$	$\dfrac{\kappa}{t}$	$\kappa\ln\left(\dfrac{t}{\lambda}\right)$	$t \ge \lambda;$ $\lambda > 0; \ \kappa > 0$
Gompertz	$\delta\kappa^t e^{-\delta(\kappa^t-1)/\ln\kappa}$	$e^{-\delta(\kappa^t-1)/\ln\kappa}$	$\delta\kappa^t$	$\dfrac{\delta(\kappa^t-1)}{\ln\kappa}$	$\kappa > 1; \ \delta > 0$

Table 4.3: Two-parameter univariate lifetime distributions.

the Weibull distribution, the gamma distribution includes the exponential distribution as a special case when $\kappa = 1$. The hazard function increases from zero to λ when $\kappa > 1$, decreases from infinity to λ when $\kappa < 1$.

The *uniform distribution* is a simple two-parameter model. The main application of the uniform distribution in survival analysis is to approximate lifetime distributions over relatively small intervals. The uniform distribution has support on $[a, b]$ with location parameters a and b, where $0 \leq a < b$. The hazard function increases from $h(a) = 1/(b-a)$ to infinity. When $a = 0$ and $b = 1$, the uniform distribution can be used to generate random variates for Monte Carlo simulation by inversion of the cumulative distribution function based on the probability integral transformation.

The *log normal distribution* has a hazard function shape that places it in the UBT [upside-down bathtub-shaped, or hump-shaped, where $h(t)$ increases initially and then decreases] class. It is parameterized by μ and σ because the logarithm of a log normal random variable is a normal random variable with population mean μ and standard deviation σ. One historical reason that the log normal distribution has been less popular than the Weibull distribution is that its survivor function is not closed form. This is important for estimating parameters for right-censored data sets, although widespread algorithms and computer routines can overcome this issue. The survivor function for a log normal random variable is

$$S(t) = 1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right) \qquad t \geq 0,$$

where Φ is the cumulative distribution function of a standard normal random variable.

The *log logistic distribution* has positive scale parameter λ and positive shape parameter κ. The hazard function is decreasing when $\kappa \leq 1$ and is UBT for $\kappa > 1$. As with the exponential and Weibull distributions, its survivor function can be inverted in closed form, so log logistic variates can easily be generated by inversion for Monte Carlo simulation. The log logistic distribution is widely used in biomedical applications.

The *inverse Gaussian distribution* has a positive parameter μ and positive scale parameter λ. Similar to the log normal distribution, the inverse Gaussian distribution is also in the UBT class. The survivor function is not closed form, but can be written in terms of the cumulative distribution function of a standard normal random variable. The population mean of the inverse Gaussian distribution is μ and the population variance is μ^3/λ, so the parameter μ is not a true location parameter because it does more than just shift the location of the distribution.

The *exponential power distribution* has a positive scale parameter λ and a positive shape parameter κ. The exponential power distribution has two properties that make it unique. First, the hazard function increases exponentially in t, whereas the Weibull hazard function increases in a polynomial fashion. Second, the exponential power distribution is one of the few two-parameter distributions that has a hazard function that can assume a bathtub shape. The hazard function achieves a minimum at $t = [(1 - \kappa)/(\lambda\kappa)]^{1/\kappa}$ when $\kappa < 1$. For $\kappa > 1$, the hazard function increases from zero to infinity, and for $\kappa = 1$ the hazard function increases from λ. The distribution has a characteristic life of $(1/\lambda)^{1/\kappa}$. The exponential power distribution's survivor function, which is

$$S(t) = e^{1 - e^{\lambda t^\kappa}} \qquad t \geq 0,$$

can be inverted in closed form, so random variates can easily be generated by inversion.

Pareto devised a probability distribution with support on $t \geq \lambda$, where κ is a positive shape parameter and λ is a positive scale parameter. The hazard function for the *Pareto distribution* decreases to zero from $h(\lambda) = \kappa/\lambda$.

The *Gompertz distribution* is a lifetime model that has been used to model adult lifetimes in actuarial applications. This distribution has positive shape parameters δ and κ. Gompertz assumed that Mill's ratio, the reciprocal of the hazard function, measures human resistance to death. He assumed this resistance decreases over time at a rate proportional to itself; that is,

$$\frac{d}{dt}\left[\frac{1}{h(t)}\right] = \kappa\left[\frac{1}{h(t)}\right],$$

where κ is a constant. The solution to this separable differential equation is $h(t) = \delta e^{ct}$, where $e^c = \kappa$. The hazard function increases from $h(0) = \delta$.

4.4.3   Some Three-Parameter Lifetime Models

The three-parameter lifetime distributions defined here have their $f(t)$, $S(t)$, $h(t)$, and $H(t)$ functions given in Table 4.4. The *Makeham distribution* has three positive shape parameters and is a generalization of the Gompertz distribution with γ included in the hazard function. Whereas the Gompertz distribution has been used to model lifetimes in terms of death from natural causes, the Makeham distribution takes into account the possibility of accidental deaths by including the extra parameter. The hazard function increases from $\delta + \gamma$.

The *IDB* (increasing, decreasing, bathtub) *distribution* is a three-parameter model with a hazard function that can exhibit increasing ($\delta \geq \gamma\kappa$), decreasing ($\delta = 0$), and bathtub shapes ($0 < \delta < \gamma\kappa$). The distribution has shape parameters $\delta \geq 0$, $\kappa \geq 0$, and $\gamma \geq 0$. Special cases of the IDB distribution are the Rayleigh distribution when $\gamma = 0$ and the exponential distribution when $\delta = \kappa = 0$ and $\gamma > 0$.

The *generalized Pareto distribution* is another three-parameter distribution with shape parameters δ, κ, and γ. It is able to achieve an increasing hazard function when $\kappa < 0$, a decreasing hazard function when $\kappa > 0$, and a constant hazard function when $\kappa = 0$. For all parameter values, $h(0) = \gamma + \kappa/\delta$ and $\lim_{t \to \infty} h(t) = \gamma$. The special cases of $\gamma = 0$ and $\kappa = -\delta\gamma$ result in the hazard functions

$$h(t) = \frac{\kappa}{t+\delta} \qquad \text{and} \qquad h(t) = \frac{\gamma t}{t+\delta}$$

for $t \geq 0$.

Distribution	$f(t)$	$S(t)$	$h(t)$	$H(t)$	Parameters
Makeham	$(\gamma+\delta\kappa^t)e^{-\gamma t-\delta(\kappa^t-1)/\ln\kappa}$	$e^{-\gamma t-\delta(\kappa^t-1)/\ln\kappa}$	$\gamma+\delta\kappa^t$	$\gamma t+\dfrac{\delta(\kappa^t-1)}{\ln\kappa}$	$\delta \geq 0;\ \kappa > 1;$ $\gamma > 0$
IDB	$\dfrac{(1+\kappa t)\delta t+\gamma}{(1+\kappa t)^{\gamma/\kappa+1}}e^{-\delta t^2/2}$	$(1+\kappa t)^{-\gamma/\kappa}e^{-\delta t^2/2}$	$\delta t+\dfrac{\gamma}{1+\kappa t}$	$\dfrac{\delta}{2}t^2+\dfrac{\gamma}{\kappa}\ln(1+\kappa t)$	$\delta \geq 0;\ \kappa \geq 0;$ $\gamma \geq 0$
Generalized Pareto	$\left(\gamma+\dfrac{\kappa}{t+\delta}\right)\left(1+\dfrac{t}{\delta}\right)^{-\kappa}e^{-\gamma t}$	$\left(1+\dfrac{t}{\delta}\right)^{-\kappa}e^{-\gamma t}$	$\gamma+\dfrac{\kappa}{t+\delta}$	$\gamma t+\kappa\ln\left(1+\dfrac{t}{\delta}\right)$	$\delta > 0;\ \gamma \geq 0;$ $\kappa \geq -\delta\gamma$

Table 4.4: Three-parameter univariate lifetime distributions.

4.4.4 Some *n*-Parameter Lifetime Models

Two *n*-parameter distributions are related to the exponential distribution. The first is the *hypoexponential distribution*. If $T_i \sim$ exponential(λ_i) for $i = 1, 2, \ldots, n$, then $T = T_1 + T_2 + \cdots + T_n$ has the hypoexponential distribution. The hypoexponential distribution collapses to the Erlang distribution with parameters λ and n when $\lambda = \lambda_1 = \lambda_2 = \cdots = \lambda_n$. The hypoexponential distribution is in the IFR class for all values of its parameters.

A second *n*-parameter distribution is the *hyperexponential distribution*. If $T_i \sim$ exponential(λ_i) for $i = 1, 2, \ldots, n$, and T has probability density function

$$f_T(t) = p_1 f_{T_1}(t) + p_2 f_{T_2}(t) + \cdots + p_n f_{T_n}(t),$$

where $p_1 + p_2 + \cdots + p_n = 1$ and $p_i > 0$ for $i = 1, 2, \ldots, n$, then T has the hyperexponential distribution. This lifetime distribution is a mixture of exponential distributions. The hyperexponential distribution collapses to the exponential distribution with failure rate λ when $\lambda = \lambda_1 = \lambda_2 = \cdots = \lambda_n$. The hyperexponential distribution is in the DFR class for all values of its parameters.

4.4.5 Summary

Figure 4.13 shows how these univariate lifetime distributions are related to one another. Each oval represents one lifetime distribution, listing its name, parameter(s), and support. Solid arrows connecting the distributions denote special cases and transformations. An example of a special case is the arrow pointing from the Weibull distribution to the exponential distribution with the label $\kappa = 1$. An example of a transformation is the arrow pointing from the exponential distribution to the chi-square distribution with the label $2\lambda \sum_{i=1}^{n} T_i$ (iid). This result is given in Theorem 4.5. Another example of a transformation is the self-loop on the exponential distribution, where the minimum of independent exponential random variables is also exponential. This result is given in Theorem 4.4. Dashed arrows denote limiting distributions, which typically arise as one of the parameters approaches 0 or infinity. An example of a limiting distribution is the arrow pointing from the gamma distribution to the normal distribution with the label $\kappa \to \infty$. The limiting distribution of a gamma random variable converges to the normal distribution as its shape parameter increases.

Table 4.5 contains a summary of the distribution classes to which the distributions belong. Double lines are used to separate the distributions by the number of parameters. For each class to which a distribution belongs, the corresponding set of parameter values is specified. The distribution classes that are considered are IFR, DFR, BT, and UBT.

4.5 Moment Ratio Diagrams

The lifetime distributions introduced in this chapter have been presented in a serial fashion without much attention being directed toward looking at all of them simultaneously. Isolating their presentation in this fashion is unfortunate; it would be of benefit to view all of these distributions simultaneously. One way to view these probability distributions simultaneously is to place them on a graph of their moments. These graphs are often known as *moment-ratio diagrams*. A moment-ratio diagram is the locus of pairs of standardized moments for a particular probability distribution plotted on a single set of axes. Moment-ratio diagrams are useful for (1) quantifying the "distance" or "proximity" between univariate probability distributions based on their second, third, and fourth moments, (2) illustrating the limiting behavior of probability distributions, (3) highlighting the versatility of a particular probability distribution based on the range of values that the moments can assume, and (4) generating a list of potential probability models based on a data set.
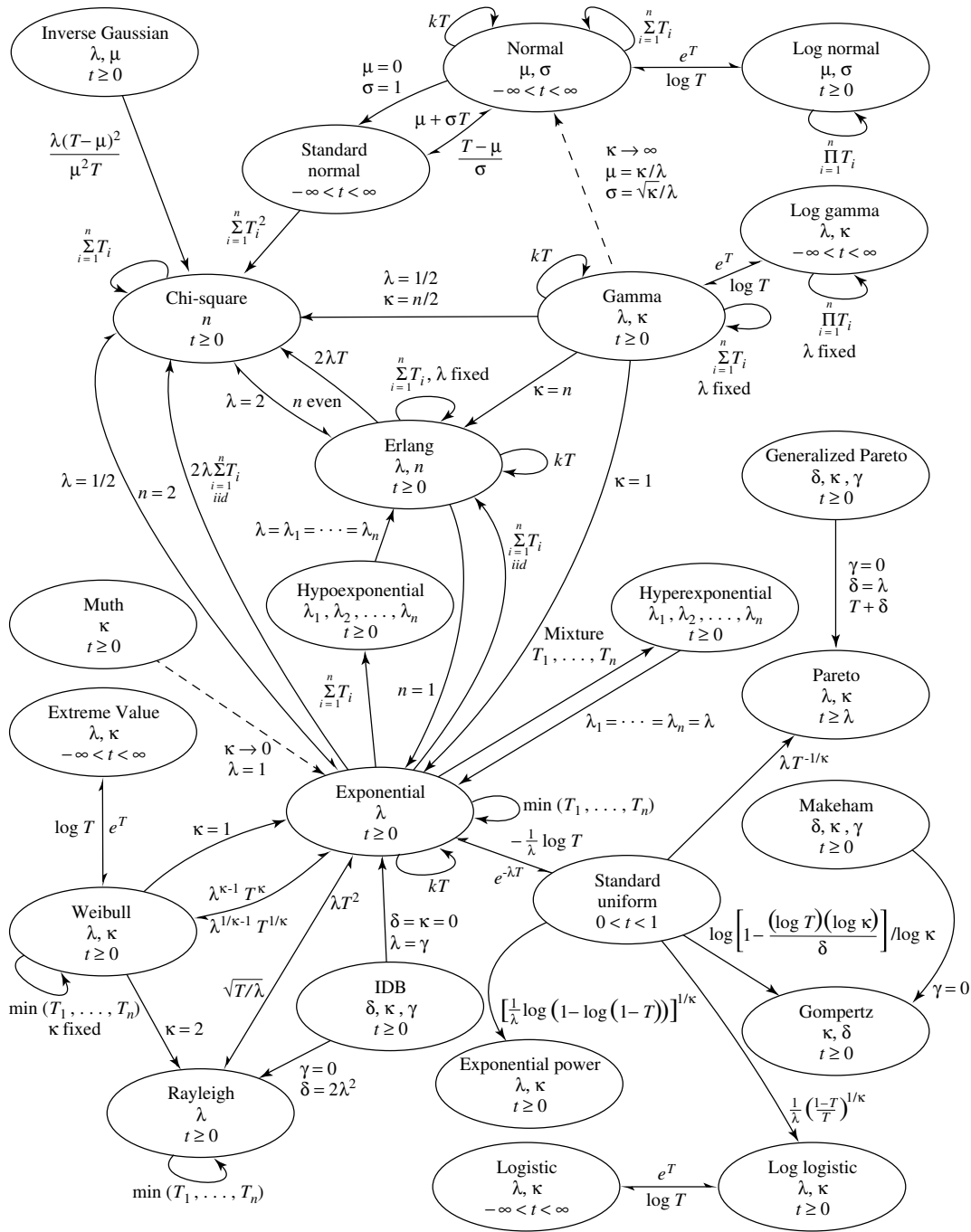
Inverse Gaussian
λ, μ
$t \geq 0$

kT

Normal
μ, σ
$-\infty < t < \infty$

$\sum\limits_{i=1}^{n} T_i$

e^T

$\log T$

Log normal
μ, σ
$t \geq 0$

$\mu = 0$
$\sigma = 1$

$\mu + \sigma T$

$\dfrac{T - \mu}{\sigma}$

$\kappa \to \infty$
$\mu = \kappa/\lambda$
$\sigma = \sqrt{\kappa}/\lambda$

$\dfrac{\lambda(T - \mu)^2}{\mu^2 T}$

Standard
normal
$-\infty < t < \infty$

$\sum\limits_{i=1}^{n} T_i^2$

$\prod\limits_{i=1}^{n} T_i$

Log gamma
λ, κ
$-\infty < t < \infty$

$\sum\limits_{i=1}^{n} T_i$

Chi-square
n
$t \geq 0$

$\lambda = 1/2$
$\kappa = n/2$

kT

e^T

$\log T$

Gamma
λ, κ
$t \geq 0$

$\prod\limits_{i=1}^{n} T_i$
λ fixed

$2\lambda T$

$\sum\limits_{i=1}^{n} T_i$, λ fixed

$\kappa = n$

$\sum\limits_{i=1}^{n} T_i$
λ fixed

$\lambda = 2$ n even

$2\lambda \sum\limits_{\substack{i=1 \\ iid}}^{n} T_i$

Erlang
λ, n
$t \geq 0$

kT

$\kappa = 1$

$\lambda = 1/2$ $n = 2$

$\lambda = \lambda_1 = \cdots = \lambda_n$

$\sum\limits_{\substack{i=1 \\ iid}}^{n} T_i$

Generalized Pareto
δ, κ, γ
$t \geq 0$

Hypoexponential
$\lambda_1, \lambda_2, \ldots, \lambda_n$
$t \geq 0$

$\sum\limits_{i=1}^{n} T_i$

Hyperexponential
$\lambda_1, \lambda_2, \ldots, \lambda_n$
$t \geq 0$

$\gamma = 0$
$\delta = \lambda$
$T + \delta$

Muth
κ
$t \geq 0$

Mixture
$T_1, \ldots, T_n$

$n = 1$

Pareto
λ, κ
$t \geq \lambda$

Extreme Value
λ, κ
$-\infty < t < \infty$

$\kappa \to 0$
$\lambda = 1$

$\lambda_1 = \cdots = \lambda_n = \lambda$

$\lambda T^{-1/\kappa}$

$\log T$ e^T

$\kappa = 1$

Exponential
λ
$t \geq 0$

$\min(T_1, \ldots, T_n)$

$-\dfrac{1}{\lambda} \log T$

Makeham
δ, κ, γ
$t \geq 0$

$e^{-\lambda T}$

$\lambda^{\kappa-1} T^{\kappa}$

λT^2

kT

$\log\left[1 - \dfrac{(\log T)(\log \kappa)}{\delta}\right] / \log \kappa$

Weibull
λ, κ
$t \geq 0$

$\lambda^{1/\kappa-1} T^{1/\kappa}$

$\delta = \kappa = 0$
$\lambda = \gamma$

Standard
uniform
$0 < t < 1$

$\gamma = 0$

$\min(T_1, \ldots, T_n)$
κ fixed

$\kappa = 2$

$\sqrt{T/\lambda}$

IDB
δ, κ, γ
$t \geq 0$

$\left[\frac{1}{\lambda} \log(1 - \log(1 - T))\right]^{1/\kappa}$

Gompertz
κ, δ
$t \geq 0$

$\gamma = 0$
$\delta = 2\lambda^2$

Exponential power
λ, κ
$t \geq 0$

$\frac{1}{\lambda}\left(\frac{1-T}{T}\right)^{1/\kappa}$

Rayleigh
λ
$t \geq 0$

$\min(T_1, \ldots, T_n)$

Logistic
λ, κ
$-\infty < t < \infty$

e^T

$\log T$

Log logistic
λ, κ
$t \geq 0$

Figure 4.13: Relationships among continuous univariate lifetime distributions.

Distribution	IFR	DFR	BT	UBT
Exponential	$\text{YES}_{\text{all }\lambda}$	$\text{YES}_{\text{all }\lambda}$	NO	NO
Muth	$\text{YES}_{\text{all }\kappa}$	NO	NO	NO
Weibull	$\text{YES}_{\kappa \geq 1}$	$\text{YES}_{\kappa \leq 1}$	NO	NO
Gamma	$\text{YES}_{\kappa \geq 1}$	$\text{YES}_{\kappa \leq 1}$	NO	NO
Uniform	$\text{YES}_{\text{all }a\text{ and }b}$	NO	NO	NO
Log normal	NO	NO	NO	$\text{YES}_{\text{all }\mu\text{ and }\sigma}$
Log logistic	NO	$\text{YES}_{\kappa \leq 1}$	NO	$\text{YES}_{\kappa > 1}$
Inverse Gaussian	NO	NO	NO	$\text{YES}_{\text{all }\lambda\text{ and }\mu}$
Exponential Power	$\text{YES}_{\kappa \geq 1}$	NO	$\text{YES}_{\kappa < 1}$	NO
Pareto	NO	$\text{YES}_{\text{all }\kappa}$	NO	NO
Gompertz	$\text{YES}_{\text{all }\delta\text{ and }\kappa}$	NO	NO	NO
Makeham	$\text{YES}_{\text{all }\delta\text{ and }\kappa}$	NO	NO	NO
IDB	$\text{YES}_{\delta \geq \gamma\kappa}$	$\text{YES}_{\delta = 0}$	$\text{YES}_{0 < \delta < \gamma\kappa}$	NO
Generalized Pareto	$\text{YES}_{\kappa \leq 0}$	$\text{YES}_{\kappa \geq 0}$	NO	NO
Hypoexponential	$\text{YES}_{\text{all }\lambda_1, \lambda_2, \ldots, \lambda_n}$	$\text{YES}_{n=1}$	NO	NO
Hyperexponential	$\text{YES}_{\lambda_1 = \lambda_2 = \cdots = \lambda_n}$	$\text{YES}_{\text{all }\lambda_1, \lambda_2, \ldots, \lambda_n}$	NO	NO

Table 4.5: Distribution classes.
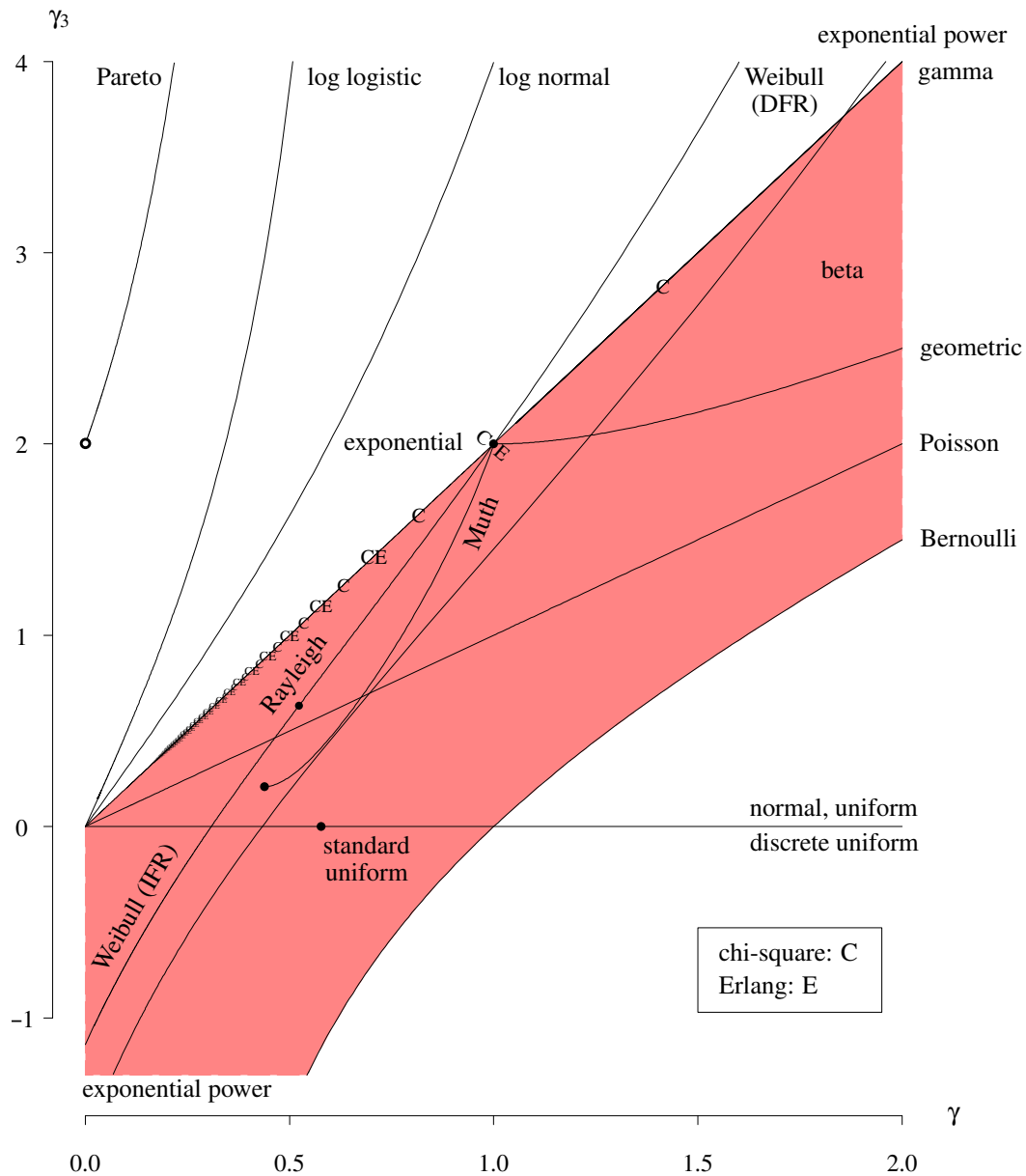
4.5.1 Skewness vs. Coefficient of Variation

As one illustration of a moment-ratio diagram, Figure 4.14 contains a plot of the population skewness

$$\gamma_3 = E\left[\left(\frac{T - \mu}{\sigma}\right)^3\right]$$

on the vertical axis, versus the population coefficient of variation

$$\gamma = \frac{\sigma}{\mu}$$

on the horizontal axis for several of the lifetime distributions introduced in this chapter, where μ and σ are the population mean and standard deviation of the random variable T. Some features of this moment-ratio diagram are listed below.

Figure 4.14: Skewness γ_3 versus coefficient of variation γ.

- The locus of points associated with the various probability distribution consist of either a single point (for example, the Rayleigh distribution), a curve (for example, the log logistic distribution), or a region (for example, the beta distribution).

- There are two gathering points: the exponential distribution at $(\gamma, \gamma_3) = (1, 2)$ and a degenerate

distribution at $(\gamma, \gamma_3) = (0, 0)$.

- The Poisson distribution, with $\gamma_3 = \gamma$, and the gamma distribution, with $\gamma_3 = 2\gamma$, have linear relationships between γ and γ_3.

- The limiting values of the beta distribution region are the line associated with the gamma distribution $\gamma_3 = 2\gamma$ and the curve associated with the Bernoulli distribution $\gamma_3 = \gamma - 1/\gamma$.

- Symmetric distributions, such as the $N\left(\mu, \sigma^2\right)$, $U(a, b)$, and discrete uniform distributions, all have population skewness $\gamma_3 = 0$.

- The curves associated with the gamma and Weibull distributions intersect at the exponential distribution, which is associated with shape parameter $\kappa = 1$.

- The open point associated with the Pareto distribution gives the limiting distribution as $\kappa \to \infty$. The values of γ and γ_3 are defined for $\kappa > 3$.

- The chi-square distribution, indicated by a C for various values of its degrees of freedom, and the Erlang distribution, indicated by an E for various values of its integer shape parameter, coincide when the degrees of freedom for the chi-square distribution are even. This accounts for the alternating pattern of C and CE labels along the line for the gamma distribution.

4.5.2 Kurtosis vs. Skewness

A second moment-ratio diagram, which is given in Figure 4.15 is a plot of the population kurtosis

$$\gamma_4 = E\left[\left(\frac{T - \mu}{\sigma}\right)^4\right]$$

on the vertical axis, versus the population skewness

$$\gamma_3 = E\left[\left(\frac{T - \mu}{\sigma}\right)^3\right]$$

on the horizontal axis for several lifetime distributions introduced in this chapter. (Some authors prefer to work with the *excess population kurtosis* $\gamma_4 - 3$.) Although it uses higher-order moments, it is considered the more classic moment-ratio diagram because the distributions plotted consist of points, curves, and regions that are independent of location and scale parameters. The population skewness scale can sometimes be replaced by the squared skewness, resulting in what is known as a *Cullen and Frey graph*. Figure 4.15 contains a moment-ratio diagram for the population skewness versus the population kurtosis, plotted upside down per tradition.

The locus of (γ_3, γ_4) values that a distribution occupies in Figure 4.15 typically depends on the number of shape parameters. The Rayleigh(λ) distribution, for example, with just a scale parameter, occupies just the single point because it has no shape parameters. The gamma distribution, on the other hand, occupies the curve $\gamma_3 = 3\gamma_2^2 + 3$ because it has one shape parameter. Finally, the beta distribution occupies a region because it has two shape parameters. Some further features of this moment-ratio diagram are listed below.
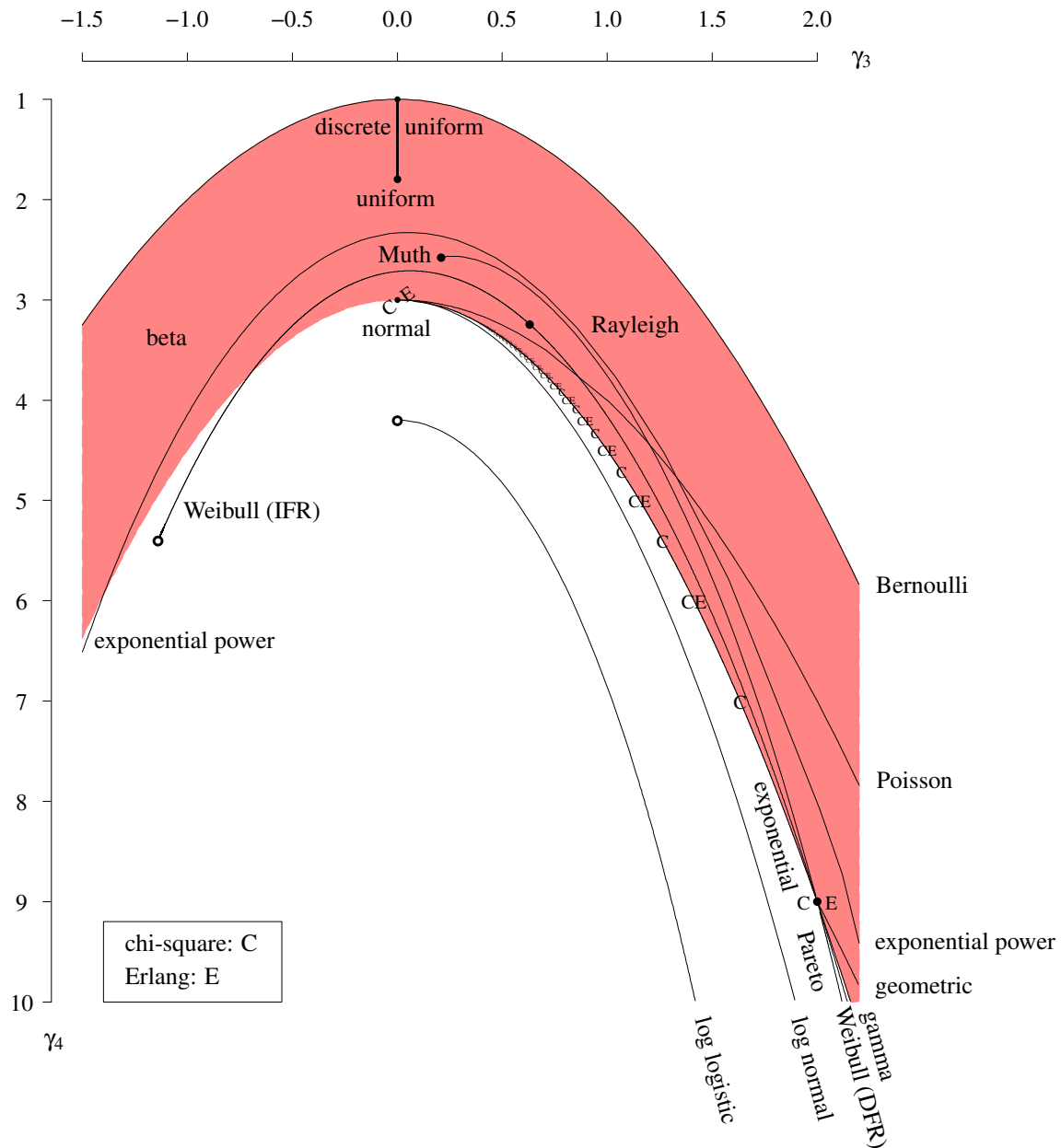
Figure 4.15: Population kurtosis γ_4 versus population skewness γ_3.

- There are two gathering points: the exponential distribution at $(\gamma_3, \gamma_4) = (2, 9)$ and the normal distribution at $(\gamma_3, \gamma_4) = (0, 3)$. This is further evidence of the centrality of these two distributions in probability theory: the exponential distribution plays a pivotal role in stochastic processes (for example, queueing theory and survival analysis) and the normal distribution

plays a pivotal role in classical statistics. Stochastic processes can be thought of as probability over time; statistics can be thought of as probability applied to data.

- Once again, the beta distribution covers the largest amount of territory in Figure 4.15, high-lighting its versatility as a probability model.

- The smallest possible value of the population kurtosis is $\gamma_4 = 1$, which is achieved by the Bernoulli distribution with $p = 1/2$. This distribution is equivalent to the discrete uniform distribution with parameters 0 and 1.

- The curves associated with the gamma and Weibull distributions again intersect at the expo-nential distribution at $(\gamma_3, \gamma_4) = (2, 9)$, which corresponds to the shape parameter $\kappa = 1$.

- The discrete uniform distribution is plotted as a solid line between its limits as a two-mass value at $(\gamma_3, \gamma_4) = (0, 1)$ and its limiting distribution (as the number of mass values increases) at $(\gamma_3, \gamma_4) = (0, 1.8)$. The locus of points is actually a series of points along this line.

A statistician can plot the *sample skewness* and the *sample kurtosis*

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{T_i - \bar{T}}{S}\right)^3 \qquad \text{and} \qquad \frac{1}{n}\sum_{i=1}^{n}\left(\frac{T_i - \bar{T}}{S}\right)^4$$

for a set of data values $T_1, T_2, \ldots, T_n$, sample mean $\bar{T}$, and sample standard deviation S on Fig-ure 4.15 for a particular data set. Doing so gives a sense for which of the distributions might be candidate probability models for the implied population distribution. The moment-ratio diagram allows a statistician to compare several candidate distributions simultaneously in terms of their mo-ments.

4.6 Proportional Hazards Model

The *proportional hazards* model is appropriate for including a vector of covariates (for example, the turning speed and feed rate for a drill bit) in a lifetime model. It is often known as the *Cox proportional hazards model* because it was devised by British statistician Sir David Cox in 1972. A *covariate*—often called an *explanatory variable*—is a variable that influences the survival time of the item under consideration. Covariates might account for the fact that the population is not truly homogeneous, or they might account for treatments imposed on the population.

The $q \times 1$ vector $\mathbf{z} = (z_1, z_2, \ldots, z_q)'$ contains q covariates associated with a particular item. These covariates might be treatments, stresses, intrinsic properties of items, or exogenous (environ-mental) variables. The simplest case is the two-population situation modeled by a single ($q = 1$) binary covariate z, where $z = 0$ typically corresponds to the control group and $z = 1$ typically corre-sponds to the treatment group. A second, slightly more complicated example arises when a single covariate assumes a continuous value (for example, dosage in a medical setting or turning speed in a manufacturing setting). The objective in an analysis of this type might be to find the dosage or turning speed that minimizes risks or costs, respectively. Other possibilities for the elements of $\mathbf{z}$ include cumulative load applied, time-varying stresses, and environmental factors.

The covariates increase or decrease the hazard function in the proportional hazards model. This model was originally developed for medical settings in which covariates are usually patient char-acteristics such as age, gender, cholesterol level, or blood pressure. The models are often used to determine which covariate has the most significant impact on survival or to compare the survival

patterns for different treatments (for example, chemotherapy versus surgery for cancer) by factoring out the impact of the covariates.

One issue of immediate interest is how to link the covariates to a lifetime distribution. One approach is to define one lifetime model when $z = 0$ (often called the *baseline* distribution) and other models when $z \neq 0$. One problem that arises with this approach is that there might be dozens or even thousands of possible values associated with $z \neq 0$, and a separate lifetime model would need to be defined for each of these vectors. The more practical approach is to define a single lifetime model which is appropriate for all values of z in order to simplify the modeling.

The baseline distribution corresponds to having all the covariates equal to zero. In a reliability setting, this is typically the normal operating conditions for the item. Other covariate vectors are often used for accelerated environmental conditions. In a biomedical setting, the baseline is typically the control group that receives either no treatment or the standard treatment for a particular disease. The covariates are linked to the lifetime by the link function $\psi(z)$, which typically satisfies $\psi(0) = 1$ and $\psi(z) > 0$ for all z. When a link function satisfies these conditions, then $z = 0$ implies that $S_0(t) \equiv S(t)$. The most general case is to let $\psi(z)$ be any function of the covariates.

Definition 4.9 Let the $q \times 1$ vector $z = (z_1, z_2, \ldots, z_q)'$ denote q covariates associated with the lifetime of an item. The *proportional hazards* model can be defined by

$$h(t) = \psi(z)h_0(t) \qquad t \geq 0,$$

where $h_0(t)$ is a baseline hazard function and $\psi(z) > 0$ is a link function.

The covariates increase the hazard function when $\psi(z) > 1$ or decrease the hazard function when $\psi(z) < 1$. A popular choice is the log-linear link function $\psi(z) = e^{\beta' z}$, where $\beta = (\beta_1, \beta_2, \ldots, \beta_q)'$ is a $q \times 1$ vector of regression coefficients corresponding to the q covariates. The log-linear link function satisfies $\psi(z) > 0$ for all vectors z and β. Other, less popular choices for the link function are $\psi(z) = \beta' z$ and $\psi(z) = (\beta' z)^{-1}$. Both alternative choices suffer from the limitation that $\psi(z) < 0$ for some values of β and z, resulting in a constrained optimization problem when the models are fitted to data. The left-hand side of this model is often written as $h(t; z)$ because survival is now a function of both time and the covariate vector z.

Regression modeling tools, such as indicator variables, modeling of interaction terms, modeling of nonlinear relationships between variables, and stepwise selection of significant covariates, can all be used here in the same fashion as in regression modeling covered earlier in the text. Estimation of the regression coefficients $\beta_1, \beta_2, \ldots, \beta_q$ and the baseline distribution parameters from a data set consisting of times to failure and associated covariates is introduced in the next chapter. The proportional hazards model has a unique feature that allows estimation of the regression parameters (the β vector) without knowledge of the baseline distribution.

Reliability engineers often use accelerated conditions to induce failures. These conditions include voltage, current, pressure, impact, and humidity. The results from the fitted proportional hazards model can then be extrapolated back to the standard operating conditions by adjusting the values of the covariates. The accelerated levels of the covariates must be chosen carefully based on sound engineering judgment and previous experience in order to assure that failure modes that would not occur in the standard operating conditions are not induced by the accelerated testing environment.

The other lifetime distribution representations can be determined for the proportional hazards

model. For example, the cumulative hazard function for a random variable T with covariates z is

$$
\begin{aligned}
H(t) &= \int_0^t h(\tau)\,d\tau \\
&= \int_0^t \psi(z)h_0(\tau)\,d\tau \\
&= \psi(z)\int_0^t h_0(\tau)\,d\tau \\
&= \psi(z)H_0(t) \qquad t \geq 0.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
S(t) &= e^{-H(t)} \\
&= e^{-\psi(z)H_0(t)} \\
&= \left(e^{-H_0(t)}\right)^{\psi(z)} \\
&= \left[S_0(t)\right]^{\psi(z)} \qquad t \geq 0.
\end{aligned}
$$

Finally,

$$
\begin{aligned}
f(t) &= S(t)h(t) \\
&= \left[S_0(t)\right]^{\psi(z)}\psi(z)h_0(t) \\
&= \left[S_0(t)\right]^{\psi(z)-1}\psi(z)S_0(t)h_0(t) \\
&= \left[S_0(t)\right]^{\psi(z)-1}\psi(z)f_0(t) \qquad t \geq 0.
\end{aligned}
$$

The notation has been simplified in the three expressions above; these functions are more accurately expressed as $H(t, z)$, $S(t, z)$, and $f(t, z)$. Table 4.6 summarizes the various lifetime distribution representations for the proportional hazards models. This table allows a modeler to determine any of the four lifetime distribution representations for either model once the baseline distribution and link function are specified, as illustrated in the next example.

Representation	Proportional Hazards
$S(t)$	$\left[S_0(t)\right]^{\psi(z)}$
$f(t)$	$f_0(t)\psi(z)\left[S_0(t)\right]^{\psi(z)-1}$
$h(t)$	$\psi(z)h_0(t)$
$H(t)$	$\psi(z)H_0(t)$

Table 4.6: Lifetime distribution representations for the proportional hazards model.

Example 4.5 Consider the case of a Weibull baseline function in a proportional hazards model. Find the hazard function, survivor function, and the mean time to failure for an item having covariate vector $\mathbf{z}$.

The baseline hazard function is Weibull with parameters λ and κ:

$$h_0(t) = \kappa \lambda^\kappa t^{\kappa-1} \qquad\qquad t \geq 0.$$

So the hazard function for an item with covariates $\mathbf{z}$ is

$$h(t) = \psi(\mathbf{z}) h_0(t) = \psi(\mathbf{z}) \kappa \lambda^\kappa t^{\kappa-1} \qquad\qquad t \geq 0.$$

Using Table 4.6, the appropriate formula for determining the survivor function is

$$S(t) = [S_0(t)]^{\psi(\mathbf{z})} \qquad\qquad t \geq 0.$$

Using the usual baseline survivor function for the Weibull distribution,

$$S(t) = \left[e^{-(\lambda t)^\kappa} \right]^{\psi(\mathbf{z})} = e^{-(\lambda t)^\kappa \psi(\mathbf{z})} \qquad\qquad t \geq 0.$$

This survivor function can be recognized as that of a Weibull lifetime with scale parameter $\lambda \psi(\mathbf{z})^{1/\kappa}$ and shape parameter κ. The population mean time to failure for an item with covariate vector $\mathbf{z}$ is

$$E[T] = \frac{1}{\lambda \psi(\mathbf{z})^{1/\kappa} \kappa} \Gamma\left(\frac{1}{\kappa} \right).$$

As before, the notation has been simplified. It is certainly more accurate to write this as $E[T \,|\, \mathbf{z}]$.

This chapter has contained a brief introduction to probability models for univariate lifetime distributions, both without and with associated covariates. These models are appropriate for a nonnegative random variable T with applications in reliability, biostatistics, actuarial science, economics, sociology, etc. The distribution of T can be defined by one of five lifetime distribution representations: the survivor function, the probability density function, the hazard function, or the cumulative hazard function. The exponential distribution is a key central lifetime distribution because it is the only continuous distribution having both a constant hazard function and the memoryless property. The Weibull distribution is a two-parameter lifetime distribution that includes the exponential distribution as a special case when its shape parameter κ is equal to 1. The Cox proportional hazards model provides one way to incorporate a vector of covariates $\mathbf{z}$ into a lifetime model. This model contains a link function $\psi(\mathbf{z})$ which links the values of the covariates to the failure time distribution. The next chapter introduces statistical methods that can be applied to lifetime data.

4.7   Exercises

4.1   Let $t^* > 0$ be the mode value for a continuous lifetime T. Show that $h'(t^*) = [h(t^*)]^2$.

4.2   The probability that an item will survive a 1000-hour mission is 0.4. If the item is operating 800 hours into the mission, the probability of surviving the remaining 200 hours of the mission is 0.85. What is the probability that the item survives the initial 800 hours of the mission?

4.3 The hazard function shown below is for a continuous random variable measured in hours.

(a) Find $S(4)$.

(b) Find $S(10)$.

(c) Find $f(10)$.
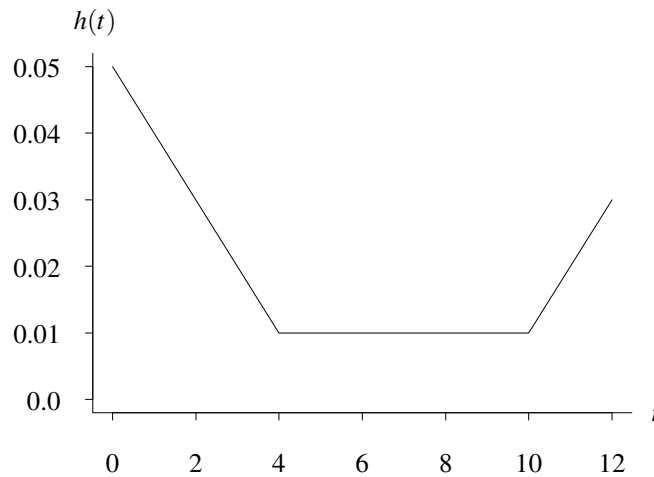
4.4 Draw the survivor function corresponding to the probability density function illustrated below. Use a straight edge whenever the function is linear. The rectangles and triangle on the probability density function all have area $1/3$.
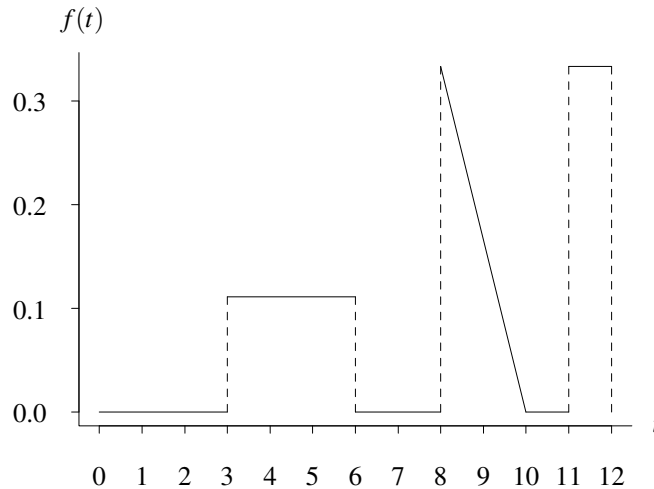
4.5 Consider the hazard function

$$h(t) = \alpha + e^{\beta t} \qquad t \geq 0.$$

What conditions must the parameters α and β meet for $h(t)$ to be a legitimate hazard function for a random lifetime T?

4.6   Jordan has designed a new screwdriver. Its lifetime, measured in years, follows a distribution with survivor function

$$S(t) = \frac{1}{1 + (e^{\lambda t} - 1)^\kappa} \qquad t \geq 0,$$

where λ and κ are positive parameters. If $\kappa = 1/2$ and $\lambda = 1/7$, find the probability that a screwdriver that is still functioning after 5 years of use will last another 3 years.

4.7   Let the time to failure of a bar code reader have survivor function

$$S(t) = \frac{2}{1 + e^{\theta t}} \qquad t \geq 0,$$

where $\theta > 0$. If Ellen places n bar code readers on test simultaneously at time $t = 0$, find the expected number that fail by time t_0.

4.8   The lifetime of a motor, in years, is a continuous random variable with probability density function

$$f(t) = 3(t+3)^{-2} \qquad t \geq 0.$$

Find the warranty period so that 5% of the motors will fail during the warranty period on average.

4.9   Carrie purchases a hammer whose lifetime T is a random variable with survivor function

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

where λ is a positive scale parameter and κ is a positive shape parameter. Assuming that $\lambda = 0.001$ and $\kappa = 2$, find $P(T > 80 \,|\, T > 50)$.

4.10   Let the lifetime T have hazard function

$$h(t) = \begin{cases} \lambda & 0 < t < 1 \\ \lambda t & t \geq 1 \end{cases}$$

for positive parameter λ. Find the associated survivor function $S(t)$.

4.11   Consider a random lifetime T with survivor function

$$S(t) = \begin{cases} 1 & t \leq 3 \\ 1 - (t-3)/(t-2) & t > 3. \end{cases}$$

Give a variate generation algorithm for this probability distribution.

4.12   Show that

$$E[T] = \int_0^\infty t f(t) \, dt$$

can also be found by

$$E[T] = \int_0^\infty S(t) \, dt$$

for any continuous random variable T with nonnegative support and a finite population mean whenever $\lim_{t \to \infty} t\,S(t) = 0$.

4.13 Consider the random variable T with hazard function

$$h(t) = \begin{cases} 1 & 0 \le t \le 1 \\ t & t > 1. \end{cases}$$

(a) Find $S(t)$.

(c) Find the median of T.

(d) Find the 95th percentile of T.

4.14 Let T be a random variable with hazard function

$$h(t) = \kappa + t \qquad t \ge 0,$$

where κ is a positive parameter.

(a) Use numerical methods to determine a value of κ such that $E[T] = 1/5$.

(b) Conduct a Monte Carlo simulation experiment that supports the value of κ determined in part (a).

4.15 The random variable T has hazard function

$$h(t) = 1 + 2t \qquad t \ge 0.$$

Find $E[T]$.

4.16 An insurance company issues a 30-month warranty on an automobile transmission. Lisa has purchased a 30-month warranty on her transmission and has not made a claim during the 30 months. She would like to purchase a 12-month extension to the warranty. This insurance company will pay a fixed $5000 for a new transmission on the extended warranty if her transmission fails within the next 12 months. Let the continuous random variable T denote the lifetime of Lisa's transmission, measured in months, from the date of the purchase of the automobile. The survivor function of T is $S(t)$. Find an expression for the expected payment that the insurance company will make on Lisa's extended warranty.

4.17 Find the population skewness and kurtosis for an exponential(λ) random variable.

4.18 Which of the following lifetimes is the best candidate for being well approximated by an exponential distribution?

(a) The burning time for a particular type of candle.

(b) The duration of a woman's pregnancy associated with her newborn baby.

(c) The duration of a strike.

(d) The duration of a men's haircut by John at John's barbershop.

(e) The breaking strength of a particular type of yarn.

4.19 Tami purchases a battery whose lifetime T is exponential(λ), for some fixed, positive failure rate λ. The battery is placed in storage on the interval $[0, t_0]$, where t_0 is a fixed, positive constant time value, then monitored continuously for failure thereafter. Thus, the time to detect failure is $X = \max\{T, t_0\}$. Find $E[X]$.

4.20 Consider the continuous random variables T_1 and T_2, each with positive support. Let $S_1(t)$ denote the survivor function of T_1 and $S_2(t)$ denote the survivor function of T_2. When $S_1(t) \geq S_2(t)$ for all values of $t \geq 0$, T_1 is said to "stochastically dominate" T_2. This is one way of showing the superiority of the probability distribution of T_1 over the probability distribution of T_2. One measure of the difference between T_1 and T_2 is the Kolmogorov metric

$$D = \sup_t |S_1(t) - S_2(t)|.$$

Calculate the value of D for the independent random variables $T_1 \sim \text{exponential}(\lambda_1)$ and $T_2 \sim \text{exponential}(\lambda_2)$, where $\lambda_1 < \lambda_2$.

4.21 Rosie purchases a light bulb whose lifetime follows an exponential distribution. If the mean lifetime is one year longer than the median lifetime, find the value of the mean lifetime.

4.22 Marian purchases 30 sixty-watt light bulbs, each having a lifetime which is exponentially distribution with a mean of 1000 hours. If she places the 30 bulbs on a life test without replacement upon failure, find the probability that 10 or fewer of these light bulbs survive to 1200 hours.

4.23 An automobile insurance policy is structured as follows. For claims below $1000, the policyholder bears the entire cost of the claim. For claims between $1000 and $2000, the policyholder bears the first $1000 of the claim and the policy pays any additional amount. For claims over $2000, the policyholder bears the first $1000 of the claim and the policy pays $1000 plus half of the amount that exceeds $2000. The distribution of a claim T has the exponential distribution with mean $3000. Find the cumulative distribution function of the reimbursement amount that the insurance company pays on a claim.

4.24 Let $T_1, T_2, \ldots, T_5$ be independent and identically distributed exponential(λ) random variables.

(a) Find the probability density function of the second order statistic $T_{(2)}$.

(b) Find $P\left(T_{(2)} \leq 1/\lambda\right)$.

4.25 Let T_1, T_2, T_3, be independent and identically distributed exponential(λ) random variables. Find the 96th percentile of the random variables:

(a) $3\min\{T_1, T_2, T_3\}$,

(b) $T_1 + T_2 - T_3$.

4.26 Troy is a bicyclist. His bike has a front wheel with eight spokes. The wheel is in the failed state when two consecutive spokes have failed. The initial lifetime distributions of the eight spokes are independent and identically distributed exponential random variables with failure rates of one failure per year. When a spoke fails, the failure rate of the two adjacent operating spokes doubles. Conduct a Monte Carlo simulation experiment to estimate the population mean time to wheel failure (considering only the spokes) to two-digit accuracy.

4.27 For the Weibull random lifetime T, show that

$$P\left(T < \frac{1}{\lambda}\right) = 1 - \frac{1}{e},$$

regardless of the value of κ.

4.28   Find the population skewness and kurtosis of a Weibull(λ, κ) random variable.

4.29   Find the value of the shape parameter in the Weibull distribution associated with a population skewness of zero.

4.30   Drea is using a graphics design software package that has a Weibull time to failure. Find the mode of the time to failure distribution. For which parameter values is this value the mode? Find the probability that the software package is still functioning at the mode value.

4.31   Let $T \sim$ Weibull(λ, κ).

   (a) Find expressions for the mean, median, and mode of T. (*Hint*: they might not all be closed-form.)

   (b) Find parameter values associated with the following three cases: the median and mode of the distribution are equal; the mean and median of the distribution are equal; the mean and mode of the distribution are equal.

4.32   Katherine designs a scanner and desires a one-month reliability of 0.8. She finds that the failure time of the scanner has a Weibull distribution with parameters $\lambda = 8.33$ and $\kappa = 0.334$, with time measured in months. Unfortunately, she finds that the one-month reliability is

$$S(1) = e^{-8.33^{0.334}} = 0.13,$$

which is clearly unacceptable. Fortunately, this Weibull distribution has a decreasing failure rate, so she knows that if she burns in the scanners, she can increase their one-month reliability. How long should she burn in the scanners to achieve a one-month reliability of 0.8 for scanners that survive the test? What fraction of the scanners placed on the burn-in test will fail during the test?

4.33   Statistical applications involving the Weibull distribution can benefit from reparameterizing the distribution. One such reparameterization replaces the scale parameter λ with a particular fractile of the distribution. More specifically, let p_0 be a prescribed constant satisfying $0 < p_0 < 1$. Denote the associated fractile of the Weibull distribution as t_{p_0}. Perform the necessary algebra to write the survivor function of the reparameterized Weibull distribution in terms of the parameters t_{p_0} and κ.

4.34   Alex purchases a laptop computer with a lifetime T, in years, which has a Weibull distribution with $\lambda = 0.2$ and $\kappa = 2$. The laptop computer can be purchased for $600. The manufacturer of the laptop provides a full refund if the laptop fails within the first year after purchase, a one-third refund if the laptop fails during the second year after purchase, and no refund if the laptop fails thereafter. What is the expected refund on a laptop?

4.35   Steve takes three generators to a work site. He will use the three generators in a cold standby system to provide electrical power. The lifetimes of the generators are exponentially distributed with mean 1000 hours. Find the variance of the total amount of time that electrical power can be supplied by the generators.

4.36   Find $E[T^r]$ for $r = 1, 2, \ldots$ for a log logistic random variable.

4.37   Find the population skewness and kurtosis of a log logistic random variable.

4.38 Lindsay purchases a jack hammer. Let the lifetime of an item be defined by a special case of the log logistic distribution with survivor function

$$S(t) = \frac{1}{1 + \lambda t} \qquad t \geq 0,$$

where λ is a positive scale parameter. If the item has been operating for a time units, find

(a) the probability it will last another r time units,

(b) the expected remaining time to failure.

4.39 Consider the random variable X having the *logistic* distribution with location parameter η, positive scale parameter $\rho > 0$, and probability density function

$$f(x) = \frac{e^{(x-\eta)/\rho}}{\rho \left(1 + e^{(x-\eta)/\rho}\right)^2} \qquad -\infty < x < \infty.$$

Show that e^X has the log logistic distribution.

4.40 Derive $\lim_{t \to \infty} h(t)$ for the log normal distribution.

4.41 Many life insurance companies offer a "last-to-die" policy for couples or business partners that pays out when the second of the two individuals dies. These policies are often purchased to pay tax liabilities on small businesses. Assume that David, age 40, and his wife Laura, age 35, celebrate their mutual birthday by purchasing a one-year, $100,000 term last-to-die policy. Find, to the nearest penny, the revenue-neutral premium (that is, where the premium equals the expected payout). For simplicity, assume that

- all new-born baby boys have Weibull random lifetimes with $\lambda = 1/65$ and $\kappa = 3/2$,

- all new-born baby girls have exponential power random lifetimes with $\lambda = 1/12$ and $\kappa = 1/2$,

- health care, lifestyle, environmental factors, etc. remain constant throughout David and Laura's lifetimes,

- there is no overhead or profit associated with the premium,

- the prevailing interest rate during the next year is 0%, and

- their two lifetimes are independent.

4.42 Meghan purchases a book stand for a rare book, which has lifetime T. If T has the log logistic distribution, $S(1) = 1/5$ and $S(3) = 1/37$, find $S(2)$.

4.43 Joanna purchases a food truck whose lifetime is a continuous random variable T with the *power distribution*, having probability density function

$$f(t) = \frac{\beta t^{\beta-1}}{\alpha^\beta} \qquad 0 < t < \alpha,$$

where α is a positive scale parameter and β is a positive shape parameter. Find the median of T.

4.44 Summer and Brigid are conducting a study concerning the random time T required to reshelve a book after it has been returned to a library. The time between the return of a book and the time it is reshelved has a special case of the *extreme value distribution* with survivor function

$$S(t) = 1 - e^{-e^{-t}} \qquad -\infty < t < \infty.$$

For real constants a and b satisfying $a < b$, find $P(a < T < b)$.

4.45 Daneen is modeling the lifetimes of light bulbs (in years) with the proportional hazards model with $q = 2$ covariates: wattage (z_1) and operating temperature in degrees Fahrenheit (z_2). The baseline distribution is exponential with a failure rate of 1.1 failures per year and the log-linear form of the link function $\psi(\mathbf{z})$ is used. If previous data has shown that the associated regression coefficients are $\beta_1 = 0.003$ and $\beta_2 = 0.004$, what is the expected time to failure of a 60-watt bulb operating in a constant $72°$F environment?

4.46 In a *log logistic regression model* with a single covariate z, the lifetime T can be expressed as

$$T = e^{\beta_0 + \beta_1 z + \theta Y},$$

where β_0 and β_1 are regression parameters, $\theta > 0$ is a parameter of the model, and Y has probability density function

$$f_Y(y) = \frac{e^y}{(1 + e^y)^2} \qquad -\infty < y < \infty.$$

(a) Find the survivor function of T for one particular value of the covariate z; that is, find $S_{T|Z=z}(t \,|\, Z = z)$.

(b) The *odds ratio*

$$\frac{1 - S_{T|Z=z}(t \,|\, Z = z)}{S_{T|Z=z}(t \,|\, Z = z)}$$

gives the odds that an item fails by time t for one particular value of the covariate z. Calculate the odds ratio for the log logistic regression model.

(c) Consider two different items with covariates z_1 and z_2. Prove that the quotient of their odds ratios is independent of t for any time $t > 0$.

4.47 Consider the baseline hazard function

$$h_0(t) = \begin{cases} 1 & 0 \le t < 1 \\ t & t \ge 1. \end{cases}$$

In a proportional hazards model, find the probability that an item with covariates $\mathbf{z}$ and link function $\psi(\mathbf{z})$ survives to time t.

4.48 A proportional hazards model is applied to a lifetime that has a single binary covariate z with regression coefficient β, link function $\psi(z) = e^{\beta z}$, and Weibull baseline hazard function. Find

(a) the survivor function for the time to failure,

(b) the mean time to failure when $z = 0$,

(c) the mean time to failure when $z = 1$.

4.49 Ali purchases a freezer with a lifetime that is well approximated by the proportional hazards model with $q = 2$ covariates: external temperature z_1 (measured in degrees Celsius) and humidity z_2. Assume that the Weibull baseline distribution and a log-linear link function are used.

  (a) What would you expect the sign (positive or negative) of $\hat{\beta}_1$ to be if a large sample of failure times and associated covariates was collected? Explain your reasoning.

  (b) Find the probability that such a component survives to time t for any covariate vector z and regression coefficients $\boldsymbol{\beta}$.

4.50 Write a few sentences describing the suitability of the link functions

$$\psi(z) = 1 + \beta_1 z_1 + \beta_2 z_2 \qquad \text{and} \qquad \psi(z) = (\beta_1 z_1 + \beta_2 z_2)^2$$

for a proportional hazards model with $q = 2$ covariates.

4.51 Consider the Cox proportional hazards model

$$h(t) = \psi(z)h_0(t) \qquad t \geq 0,$$

with $q = 2$ covariates, z_1 and z_2, which includes an interaction term between the covariates. The link function assumes the log linear form:

$$\psi(z) = e^{\beta_1 z_1 + \beta_2 z_2 + \beta_3 z_1 z_2}.$$

Find the ratio of the hazard function for covariates z_1 and z_2 to the hazard function for covariates z_1 and $z_2 + 5$.