Wiley Series in Probability and Statistics

DESIGN OF EXPERIMENTS FOR RELIABILITY ACHIEVEMENT

STEVEN E. RIGDON | RONG PAN
DOUGLAS C. MONTGOMERY | LAURA J. FREEMAN

WILEY

Design of Experiments for Reliability Achievement

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by Walter A. Shewhart and Samuel S. Wilks

The **Wiley Series in Probability and Statistics** is well established and authoritative. It covers many topics of current research interest in both pure and applied statistics and probability theory. Written by leading statisticians and institutions, the titles span both state-of-the-art developments in the field and classical methods.

Reflecting the wide range of current research in statistics, the series encompasses applied, methodological and theoretical statistics, ranging from applications and new techniques made possible by advances in computerized practice to rigorous treatment of theoretical approaches. This series provides essential and invaluable reading for all statisticians, whether in academia, industry, government, or research.

Design of Experiments for Reliability Achievement

Steven E. Rigdon
Saint Louis University
Epidemiology and Biostatistics
Saint Louis, MO

Rong Pan
Arizona State University
School of Computing and Augmented Intelligence
Tempe, AZ

Douglas C. Montgomery
Arizona State University
School of Computing and Augmented Intelligence
Phoenix, AZ

Laura J. Freeman
Statistics
Virginia Tech
Blacksburg, VA

WILEY

This edition first published 2022

Registered Office
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Editorial Office
111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data Applied for:

ISBN: 9781119237693

Cover image: Wiley
Cover design by © Eoneren/Getty Images

Set in 9.5/12.5pt STIXTwoText by Straive, Chennai, India

10 9 8 7 6 5 4 3 2 1

In memory of our friend and colleague, Connie Margaret Borror (1966–2016)

Contents

Preface

Techniques of design of experiments (DOE) have for decades been used in industry to achieve quality products and processes. These methods often involve analyzing models that assume, at least approximately, that the outcome is normally distributed. When the outcome is a lifetime, similar techniques can be applied, although the methods require more complicated models. Reliability experiments are special in two respects: (i) there is almost always *censoring*, i.e. the termination of the experiment before all units have failed, and (ii) lifetime distributions are usually not well approximated by the normal. This book is about designing experiments and analyzing data when the outcome is a lifetime.

Condra (2001b) suggests three aspects of reliability methods:

1. Methods for measuring and predicting failures
2. Methods for accommodating failures
3. Methods for preventing failures

The first, measuring and predicting failures, usually involves fitting models to lifetime data in order to assess the reliability of a system. The second, accommodating failures, involves concepts like parallel redundancy (where failure of a single component does not cause failure of the system), repairability (the ability to quickly fix a problem and return the system to working condition), maintainability (the ability to keep a system in working condition), and others. The last, methods for preventing failures, is potentially the most useful. DOE methods can be used to find characteristics of the product, or maybe the process used to make the product, that lead to the highest possible reliability. Of course, this involves methods for measuring and predicting failures (the first item earlier) and it could involve the second (accommodating failures), but the idea of designing experiments to improve reliability is a powerful one. DOE has been used successfully in a number of areas where a normally distributed response is reasonable, but applications in reliability are rather sparse.

We have divided the book into four parts:

I Reliability Here we cover the basic concepts and definitions of reliability. We present models for lifetimes, including the exponential, Weibull, gamma, and log-normal. In addition, we discuss log-location-scale distributions, such as the smallest extreme value (SEV) distribution, which is a general class of distributions that can be used to model the logarithm of lifetimes. Inference for lifetime distributions, or log lifetime distributions, is the topic of Chapter 3. There, we develop point and interval estimate of model parameters and ways we could test hypotheses regarding those parameters.

II **Design of Experiments** In the second part we present the basic ideas of experimental design and analysis. We cover the DOE for linear and generalized linear models. Chapter 4 covers factorial designs in general, the 2^k design, and fractional two-level designs. Chapter 5 covers designs for response surfaces.

III **Regression Models for Reliability Studies** This part consists of two chapters. Chapter 6 covers parametric regression models. This includes models on transformed data, exponential regression, and Weibull regression. Chapter 7 covers semi-parametric regression models including the Cox proportional hazards model.

IV **Experimental Design for Reliability Studies** The final part addresses experimental designs for reliability studies. Chapter 8 covers tests done under a single test condition. Chapter 9 covers multiple-factor experiments, including accelerated life tests.

The material in this book requires a one- or two-semester course in probability and statistics that uses some calculus. Readers with a background in reliability but not DOE can skip Part I and proceed to Part II on experimental design, and then to Parts III and IV. Readers with a background in experimental design but not reliability can begin with Part I, skip Part II, and proceed to Parts III and IV. Those who are well versed in both reliability and DOE can proceed directly to Parts III and IV.

The book's companion web site contains the data sets used in the book, along with the R and JMP code used to obtain the analyses. The web site also contains lists of known errors in the book.

Steven E. Rigdon
Saint Louis
26 November 2021

Rong Pan, Douglas C. Montgomery
Tempe
26 November 2021

Laura J. Freeman
Arlington
26 November 2021

About the Companion Website

This book is accompanied by a companion website:

www.wiley.com/go/rigdon/designexperiments

The website includes data sets and computer code.

Part I

Reliability

1

Reliability Concepts

1.1 Definitions of Reliability

It is difficult to define reliability precisely because this term evokes many different meanings in different contexts. In the field of reliability engineering, we primarily deal with engineered devices and systems. Single-word descriptions may depict one or two aspects of reliability in an engineering application context, but they are inadequate for a technical definition of *engineering reliability*. So, how do engineers and technical experts define *reliability*?

- Radio Electronics Television Manufacturers Association (1955) – "Reliability is the probability of a device performing its purpose adequately for the period of time intended under the operating conditions encountered."
- ASQ (2020) – "Reliability is defined as the probability that a product, system, or service will perform its intended function adequately for a specified period of time, or will operate in a defined environment without failure."
- Meeker and Escobar (1998a) – "Reliability is often defined as the probability that a system, vehicle, machine, device, and so on will perform its intended function under operating conditions, for a specified period of time."
- Condra (2001a) – "Reliability is quality over time."
- Yang (2007) - "Reliability is defined as the probability that a product performs its intended function without failure under specified conditions for a specified period of time."

There are some variations in the aforementioned definitions, but they all either explicitly or implicitly state the following characteristics of reliability:

- Reliability is a probabilistic measure – the probability of a functioning product, service, or system.
- Reliability is a function of time – the probability function of successfully performing tasks, as designed, over time.
- Reliability is defined under specified or intended operating conditions.

We define a function, $S(t)$, to be the survival, or reliability, function, which is the probability of the product, service, or system being successfully operated under its normal operating condition at time t; in other words, the unit survived past time t.

Design of Experiments for Reliability Achievement, First Edition.
Steven E. Rigdon, Rong Pan, Douglas C. Montgomery, and Laura J. Freeman.

Companion website: www.wiley.com/go/rigdon/designexperiments

1.2 Concepts for Lifetimes

When an item fails, the "fix" sometimes involves making a repair to bring it back to a working condition. Another possibility is to discard the item and replace it with a working item. In general, the more complex a system is, the more likely we are to repair it, and the simpler it is the more likely we are to scrap it and replace it with a new item. For example, if the starter on our automobile fails, we would probably take out the old starter and replace it with a new one. In a case like this, the automobile is a *repairable* system, but the starter is *nonrepairable* since our fix has been to replace it entirely.

Since complex systems, which are usually repairable, are made up of component parts that are non-repairable, we will focus in this book on nonrepairable items. If these nonrepairable items are designed and built to have high reliability, then the system should be reliable as well. For nonrepairable systems we are interested in studying the distribution of the time to the first (and only) failure, or more generally, the effect of predictor variables on this lifetime. This lifetime need not be measured in calendar time; it could be measured in operating time (for an item that is switched on and off periodically), miles driven (for a motor vehicle like a car or truck), copies made (for a copier or printer), or cycles (for an industrial machine). For nonrepairable systems, we study the occurrence of events in time, such as failures (and subsequent repairs) or recurrence of a disease or its symptoms. See Rigdon and Basu (2000) for a treatment of repairable systems.

The lifetime T of a unit is a random variable that necessarily takes on nonnegative values. Usually, but not always, we think of T as a continuous random variable taking on values in the interval $[0, \infty)$. There are various forms that the distribution may take, many of which, including the exponential, Weibull and gamma, are presented in detail Chapter 2. Here we present the fundamental ideas and terms for continuous random variables.

Definition 1.1 *The probability density function (PDF) of a continuous random variable T is a function $f(t)$ with the property that*

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

Thus, probabilities for a continuous random variable are found as areas under the PDF. (See Figure 1.1a.) Note that a and b can be $-\infty$ or ∞. Since $\int_a^a f(t)\ dt = 0$, the probability that $T = a$, that is, the probability that T *equals* a particular value a, is equal to zero. This also implies that

$$P(a < T < b) = P(a < T \le b) = P(a \le T < b) = P(a \le T \le b) = \int_a^b f(t)\ dt.$$

See Figure 1.1b.

Since the probability is 1 that T is between $-\infty$ and ∞, we have the property that

$$\int_{-\infty}^{\infty} f(t)\ dt = 1. \tag{1.1}$$

See Figure 1.1c. Also, since all probabilities must be nonnegative, the PDF $f(t)$ must satisfy

$$f(t) \ge 0, \quad \text{all } t. \tag{1.2}$$

The results in (1.1) and (1.2) are the fundamental properties for a PDF.

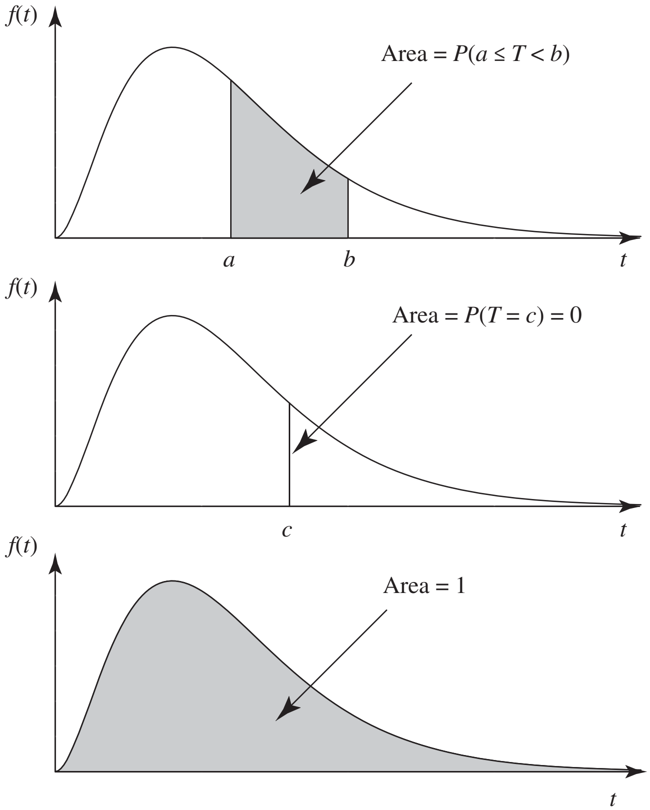

Figure 1.1 Properties of the PDF for a lifetime distribution.

The development earlier makes no assumption about the possible values that the random variable T can take on. For lifetimes, which must be nonnegative, we have $f(t) = 0$ for $t < 0$. Thus, for lifetimes, the PDF must satisfy

$$\begin{aligned} f(t) &\geq 0, \quad \text{all } t, \\ f(t) &= 0, \quad t < 0, \\ \int_0^\infty f(t)\ dt &= 1. \end{aligned}$$

The set of values for which the PDF of the random variable T is positive is called the *support* of T. The support for a lifetime distribution is $[0, \infty)$, although for some distributions we exclude the possibility of $t = 0$.

Note that the PDF does not give probabilities directly; for example, $f(4)$ does not give the probability that $T = 4$. Rather, as an approximation we can write

$$P(4 < T < 4 + \Delta t) = \int_4^{4+\Delta t} f(t)\ dt \approx \Delta t\, f(4).$$

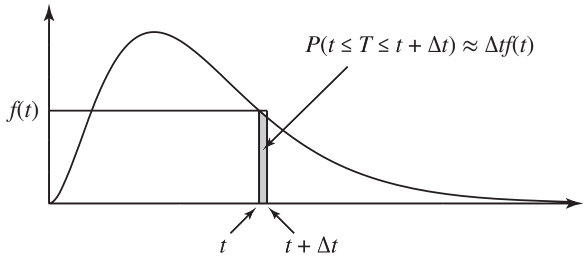

Figure 1.2 Approximation $P(t \le T \le t + \Delta t) \approx \Delta t f(t)$.

(See Figure 1.2.) Thus, the PDF can be interpreted as

$$\Delta t\, f(t) \approx P(t < T < t + \Delta t), \tag{1.3}$$

or equivalently

$$f(t) \approx \frac{P(t < T < t + \Delta t)}{\Delta t}.$$

To be precise, the PDF is equal to the limit of the right side earlier as $\Delta t \to 0^+$:

$$f(t) = \lim_{\Delta t \to 0^+} \frac{P(t < T < t + \Delta t)}{\Delta t}.$$

Definition 1.2 *The cumulative distribution function (CDF) of the random variable T is defined as*

$$F(t) = P(T \le t) = \int_{-\infty}^{t} f(x)\ dx, \tag{1.4}$$

where $f(x)$ is the PDF for T.

Note that we have changed the variable of integration from t to x, in order to avoid confusion with the upper limit on the integral. For a lifetime distribution with support $(0, \infty)$, we have the result

$$F(t) = P(T \le t) = P(0 \le T \le t) = \int_{0}^{t} f(x)\ dx. \tag{1.5}$$

As $t \to \infty$ on the right side earlier, the integral goes to $\int_0^\infty f(x)\ dx$, which equals 1. Also, since the probability of having a negative lifetime is 0, the CDF must be zero for all $t < 0$. Finally, since the CDF "accumulates" probability up to and including t, increasing t can only increase (or hold constant) the CDF. Thus, for a lifetime distribution, the CDF must satisfy

$$F(t) = 0, \quad \text{for } t \le 0,$$

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

$$F(t) \ \text{is nondecreasing.}$$

Equation (1.4) shows how to get the CDF given the PDF. A formula for the reverse (getting the PDF from the CDF) can be obtained by differentiating both sides of (1.4) with respect to t and applying the fundamental theorem of calculus:

$$F'(t) = \frac{d}{dt} \int_{-\infty}^{t} f(x)\ dx = f(t). \tag{1.6}$$

In other words, the PDF is the derivative of the CDF.

Definition 1.3 *The survival function, or reliability function, is defined to be*

$$S(t) = P(T > t) = \int_t^\infty f(x)\ dx.$$

In other words, $S(t)$ is the probability that an item survives past time t, while $F(t)$ is the probability that it fails at or before time t (that is, that it *doesn't* survive past time t). Thus, $S(t)$ and $F(t)$ are related by

$$S(t) = 1 - F(t).$$

One of the most important concepts in lifetime analysis is the hazard function.

Definition 1.4 *The hazard function is*

$$h(t) = \lim_{\Delta t \to 0^+} \frac{P(t \le T \le t + \Delta t | T > t)}{\Delta t} \tag{1.7}$$

As an approximation, we can write

$$\Delta t\ h(t) \approx P(t \le T \le t + \Delta t | T > t) \tag{1.8}$$

analogous to (1.3).

The probability in the definition of the hazard is a *conditional* probability; it is conditioned on survival to the beginning of the interval. This is a natural quantity to consider because it makes intuitive sense to talk about the failure probability of an item that is still working. It is conceptually more difficult to talk about the probability of an item failing if the item might or might not be working. If we replace the conditional probability in the definition of the hazard function with an unconditional probability, we get

$$\lim_{\Delta t \to 0^+} \frac{P(t \le T \le t + \Delta t)}{\Delta t} \tag{1.9}$$

which is equal to the PDF $f(t)$. Thus, the PDF is the (limit of) the probability of failing in a small interval when viewed *before* testing begins. The hazard is the (limit of) the probability of failing in a small interval for a unit that is known to be working.

The hazard function can be written as

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0^+} \frac{P(t < T \le t + \Delta t | T > t)}{\Delta t} \\ &= \lim_{\Delta t \to 0^+} \frac{P(t < T \le t + \Delta t \wedge T > t)}{\Delta t P(T > t)} \\ &= \lim_{\Delta t \to 0^+} \frac{P(t < T \le t + \Delta t) / \Delta t}{P(T > t)} \\ &= \frac{f(t)}{S(t)}. \end{aligned} \tag{1.10}$$

Indeed, many books *define* the hazard function in this way. We choose to define the hazard as the limit of a conditional probability because this intuitive concept is helpful for understanding the failure mechanism.

To illustrate the difference between hazard and density, consider a discrete case, say, where items are placed on test and are observed every 1000 hours. Let h_i denote the probability that an item fails in the ith interval $(1000(i-1), 1000i]$. Suppose first that $h_i = \frac{1}{10}$, so that there is a probability of $\frac{1}{10} = 0.1$ that a working unit will fail in any time interval. Thus, at the end of a time interval when we inspect those units still operating, we would expect that about one-tenth of them would fail. We could naturally ask the question "What is the probability that a unit fails in the ith interval?" This is different from the question "What is the probability that a unit that is currently operating fails in the ith interval?" The difference is that the latter is a conditional probability (conditioned on the unit still operating), whereas the former is an unconditional probability. The answer to the latter question is: the hazard h_i. To get at the answer to the latter question, we can observe that

$$\begin{aligned}
p_1 &= P\left(\text{failure in } (0, 1000]\right) = \frac{1}{10} \\
p_2 &= P\left(\text{failure in } (1000, 2000]\right) \\
&= P\left(\text{no failure in } (0, 1000]\right) P\left(\text{failure in } (1000, 2000] | \text{no failure in } (0, 1000]\right) \\
&= \frac{9}{10}\frac{1}{10} \\
p_3 &= P\left(\text{failure in } (2000, 3000]\right) \\
&= P\left(\text{no failure in } (0, 1000]\right) \\
&\quad \times P\left(\text{no failure in } (1000, 2000] | \text{no failure in } (0, 1000]\right) \\
&\quad \times P\left(\text{failure in } (2000, 3000] | \text{no failure in } (0, 2000]\right) \\
&= \frac{9}{10}\frac{9}{10}\frac{1}{10},
\end{aligned}$$

and in general,

$$\begin{aligned}
p_i &= P\left(\text{failure in } (1000\,(i-1)\,, 1000i]\right) \\
&= P\left(\text{no failure in } (0, 1000]\right) \\
&\quad \times P\left(\text{no failure in } (1000, 2000] | \text{no failure in } (0, 1000]\right) \\
&\quad \times \cdots \\
&\quad \times P\left(\text{failure in } (1000\,(i-1)\,, 1000i] | \text{no failure in } (0, 1000\,(i-1)]\right) \\
&= \left(\frac{9}{10}\right)^{i-1}\frac{1}{10}, \;\; i = 1, 2, \ldots
\end{aligned}$$

This is, of course, the geometric distribution. Plots of h_i and p_i for the case of a constant (discrete) hazard are shown in Figure 1.3.

Suppose now that the probability mass function (rather than the hazard function) is constant, with $p_i = 0.1$. Since $\sum p_i = 1$, we conclude that $p_i = 0.1$ for $i = 1, 2, \ldots, 10$. The hazard is then

$$\begin{aligned}
h_i &= P(\text{failure in interval } i | \text{survival to beginning of interval } i) \\
&= \frac{P\left(\text{failure in interval } i\right)}{P\left(\text{survival to beginning of interval } i\right)} \\
&= \frac{P\left(\text{failure in interval } i\right)}{1 - P\left(\text{death before beginning of interval } i\right)} \\
&= \frac{1/10}{1 - (i-1)/10} \\
&= \frac{1}{10 - i + 1}.
\end{aligned}$$

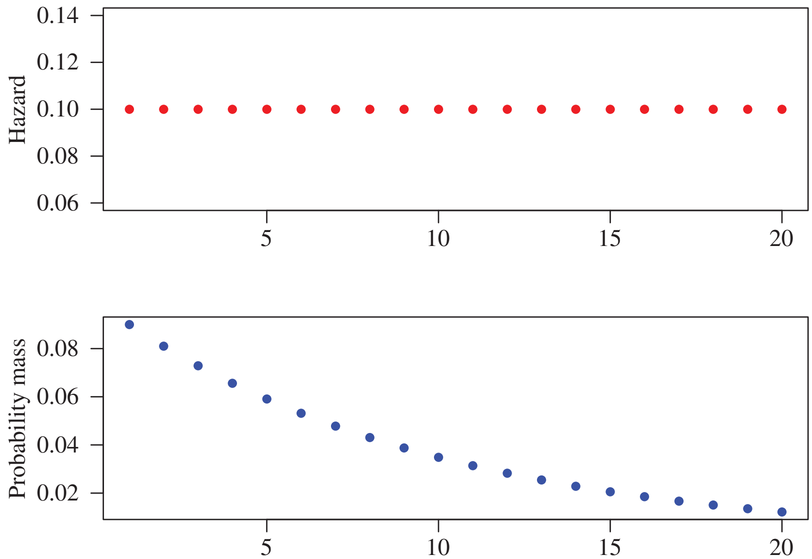

Figure 1.3 Hazard and probability mass function for the case of constant hazard.

Thus, $h_1 = \frac{1}{10}, h_2 = \frac{1}{9}, \dots, h_{10} = \frac{1}{1} = 1$. The last number, $h_{10} = 1$, may seem a little surprising, but if $p_i = 0$ for $i > 10$, and if an item hasn't failed up through interval $i = 9$, then it *must* fail at time $i = 10$. The hazard and probability mass for the constant probability case are shown in Figure 1.4.

The cumulative hazard is defined to be the accumulated area under the hazard function. To be precise, the cumulative hazard is defined to be

$$H(t) = \int_0^t h(u)\, du. \tag{1.11}$$

Any one of the PDF, CDF, survival function, hazard, or cumulative hazard function is enough to determine the lifetime distribution. In other words, knowing any one of these can get you all of the others. For example, Eq. (1.5) shows how you can get the CDF if you know the PDF. If we know the hazard function $h(t)$, we can use the relationship

$$h(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)}$$

to find $S(t)$. To see this, notice that this is a simple first-order linear differential equation with initial condition $S(0) = 1$, which can be solved by integrating both sides from $u = 0$ to $u = t$. This yields

$$\int_0^t h(u)\, du = \int_0^t -\frac{S'(u)}{S(u)}\, du = -\log S(t) + \log S(0) = -\log S(t) \tag{1.12}$$

from which we obtain the relationship

$$S(t) = \exp\left(-\int_0^t h(u)\, du\right). \tag{1.13}$$

We leave the other relationships as an exercise. Table 1.1 shows most of the relationships.

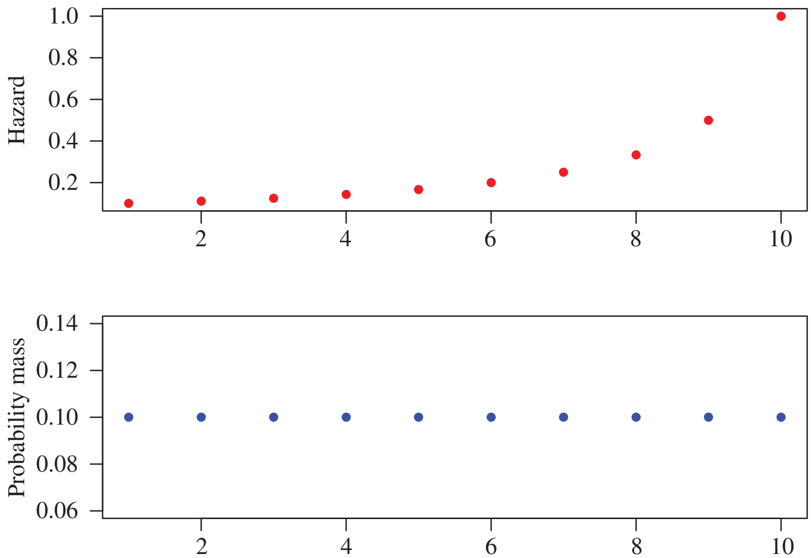

Figure 1.4 Hazard and probability mass function for the case of constant probability mass.

1.3 Censoring

A life testing experiment, where many units are operating simultaneously, may be terminated before all of the units have failed. This is a typical scenario, because many units will be highly reliable and will not fail during a test. This leads to the concept of censoring, where we cannot observe the lifetime of an item because of the design of the life testing experiment.

Definition 1.5 *When we cannot observe the exact failure time of an item, but rather we can only observe an* interval *in which the failure was observed, we say that the observation is* ***censored****. This interval could be an unbounded interval, such as* (c, ∞) *or a bounded interval such as* $(0, a)$.

The most common type of censoring occurs when the life test is stopped before all items have failed. Let τ denote the censoring time, that is, the time of termination of the test. In this case, we would know the exact failure times of all items that failed before time τ, but for those still operating at the end of the experiment, we know only that the failure would occur past time τ. For those item still operating, we would only know that the failure time was in the interval (τ, ∞). Actually, the censoring time need not be the same for all items. For example, if the items were placed into service at different times, then the censoring times would be different even if the test was terminated at the same (calendar) time. This type of lifetime censoring, where we observe a survival event (where we know that the failure must occur in the interval (τ, ∞)), is called **right censoring**.

Table 1.1 General relationships between PDF, CDF, hazard function, and cumulative hazard function.

	PDF $f(t)$	**CDF $F(t)$**	**Hazard $h(t)$**	**Cumulative hazard $H(t)$**
$f(t)$	—	$F(t) = \int_0^t f(u)\,du$	$h(t) = \frac{f(t)}{\int_t^\infty f(u)\,du}$	$H(t) = \exp\left(\frac{f(t)}{\int_t^\infty f(u)\,du}\right)$
$F(t)$	$f(t) = F'(t)$	—	$h(t) = \frac{F'(t)}{1 - F(t)}$	$H(t) = -\log(1 - F(t))$
$h(t)$	$f(t) = h(t)\exp\left(-\int_0^t h(u)\,du\right)$	$F(t) = 1 - \exp\left(-\int_0^t h(u)\,du\right)$	—	$H(t) = \int_0^t h(u)\,du$
$H(t)$	$f(t) = H'(t)\exp(-H(t))$	$F(t) = 1 - \exp(-H(t))$	$h(t) = H'(t)$	—

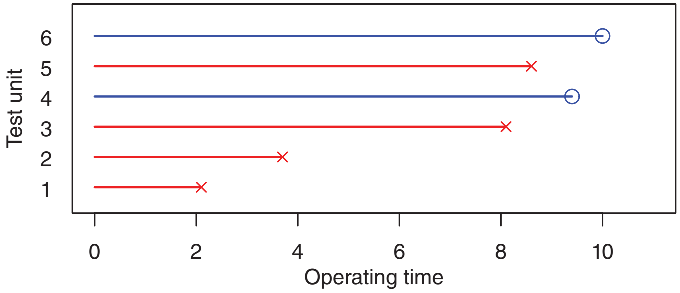

Figure 1.5 Times to failure for six items. The X indicates a failure time and the O indicates a censored time.

Figure 1.5 shows an illustration of right censoring. In this situation six items were placed on test at the same time. Items 1, 2, 3, and 5 failed during the test at times 2.1, 3.7, 8.1, and 8.6, respectively. The other two units were still operating at times 9.4 and 10.0. The failure times are denoted by an X and the censoring times are denoted by an O. The solid lines cover the times for which it is known that the items were operating.

It sometimes happens that observation of an item may begin well after the it was placed into service. This can occur when items are observed in the field. In some instances, it may be the case that the item is observed to have already failed when it is first inspected. For example, suppose an item is placed into service and then not inspected until an age of 100 days. If it was observed to be in a failed condition at that time, then we know only that it failed before time $t = 100$. In other words, the failure must have occurred sometime in the interval $(0, 100)$, but the exact failure time is unknown. This is called **left censoring**, because the failure is known to have occurred before time 100. Figure 1.6 illustrates the concept of left (and right) censoring. Here, items 1, 2, and 4 were placed on test and the failure times were observed to be 24, 37, and 77, respectively. Items 3 and 6 were observed at times 10 and 28, respectively, to be in a failed condition. A dashed line is used to indicate the possible times of actual failure. In general, we use a solid line to indicate times when the units were known to be operating. Finally, items 5 and 7 were still operating at time 99 when the test was terminated. Thus, items 3 and 6 were left censored and items 5 and 7 were right censored.

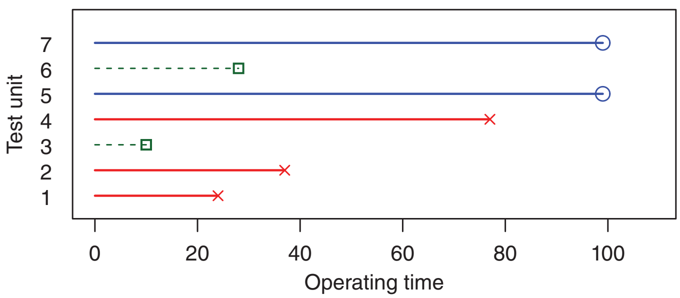

Figure 1.6 Times to failure for seven items. The X indicates a failure event and the O indicates a right censoring event, whereby the failure occurred to the right of the of the O. The dashed line that ends with a □ indicates a left censoring event, whereby the failure occurred to the left of the □.