

This chapter reviews basic notions of probability (or “stochastic variability”) which is the formal study of the laws of chance, i.e., where the ambiguity in outcome is inherent in the nature of the process itself. Both the primary views of probability, namely the frequentist (or classical) and the Bayesian, are covered, and some of the important probability distributions are presented. Finally, an effort is made to explain how probability is different from statistics, and to present different views of probability concepts such as absolute, relative and subjective probabilities.

2.1 Introduction

2.1.1 Outcomes and Simple Events

A *random variable* is a numerical description of the outcome of an experiment whose value depends on chance, i.e., whose outcome is not entirely predictable. Tossing a dice is a random experiment. There are two types of random variables:

- (i) *discrete* random variable is one that can take on only a finite or countable number of values,
- (ii) *continuous* random variable is one that may take on any value in an interval.

The following basic notions relevant to the study of probability apply primarily to *discrete* random variables.

- *Outcome* is the result of a single trial of a random experiment. It cannot be decomposed into anything simpler. For example, getting a {2} when a dice is rolled.
- *Sample space* (some refer to it as “universe”) is the set of all possible outcomes of a single trial. For the rolling of a dice, the sample space is $S = \{1, 2, 3, 4, 5, 6\}$.
- *Event* is the combined outcomes (or a collection) of one or more random experiments defined in a specific manner. For example, getting a pre-selected number (say, 4) from adding the outcomes of two dices would constitute a simple event: $A = \{4\}$.
- *Complement of a event* is the set of outcomes in the sample not contained in A. $\bar{A} = \{2, 3, 5, 6, 7, 8, 9, 10, 11, 12\}$ is the complement of the event stated above.

2.1.2 Classical Concept of Probability

Random data by its very nature is indeterminate. So how can a scientific theory attempt to deal with indeterminacy? Probability theory does just that, and is based on the fact that though the result of any particular result of an experiment cannot be predicted, a long sequence of performances taken together reveals a stability that can serve as the basis for fairly precise predictions.

Consider the case when an experiment was carried out a number of times and the anticipated event E occurred in some of them. *Relative frequency* is the ratio denoting the fraction of events when success has occurred. It is usually estimated empirically *after* the event from the following proportion:

$$p(E) = \frac{\text{number of times E occurred}}{\text{number of times the experiment was carried out}} \quad (2.1)$$

For certain simpler events, one can determine this proportion without actually carrying out the experiment; this is referred to as “*wise before the event*”. For example, the relative frequency of getting heads (selected as a “success” event) when tossing a fair coin is 0.5. In any case, this *a priori* proportion is interpreted as the long run relative frequency, and is referred to as *probability*. This is the classical, or frequentist or traditionalist definition, and has some theoretical basis. This interpretation arises from the *strong law of large numbers* (a well-known result in probability theory) which states that the average of a sequence of independent random variables having the same distribution will converge to the mean of that distribution. If a dice is rolled, the probability of getting a pre-selected number between 1 and 6 (say, 4) will vary from event to event, but on an average will tend to be close to 1/6.

2.1.3 Bayesian Viewpoint of Probability

The classical or traditional probability concepts are associated with the frequentist view of probability, i.e., interpreting

probability as the long run frequency. This has a nice intuitive interpretation, hence its appeal. However, people have argued that most processes are unique events and do not occur repeatedly, thereby questioning the validity of the frequentist or objective probability viewpoint. Even when one may have some basic preliminary idea of the probability associated with a certain event, the frequentist view excludes such subjective insights in the determination of probability. The Bayesian approach, however, recognizes such issues by allowing one to update assessments of probability that *integrate prior knowledge with observed events*, thereby allowing better conclusions to be reached. Both the classical and the Bayesian approaches converge to the same results as increasingly more data (or information) is gathered. It is when the data sets are small that the additional benefit of the Bayesian approach becomes advantageous. Thus, the Bayesian view is not an approach which is at odds with the frequentist approach, but rather adds (or allows the addition of) refinement to it. This can be a great benefit in many types of analysis, and therein lies its appeal. The Bayes' theorem and its application to discrete and continuous probability variables are discussed in Sect. 2.5, while Sect. 4.6 (of Chap. 4) presents its application to estimation and hypothesis problems.

2.2 Classical Probability

2.2.1 Permutations and Combinations

The very first concept needed for the study of probability is a sound knowledge of *combinatorial mathematics* which is concerned with developing rules for situations involving permutations and combinations.

(a) Permutation $P(n, k)$ is the number of ways that k objects can be selected from n objects *with the order being important*. It is given by:

$$P(n, k) = \frac{n!}{(n - k)!} \quad (2.2a)$$

A special case is the number of permutations of n objects taken n at a time:

$$P(n, n) = n! = n(n - 1)(n - 2) \dots (2)(1) \quad (2.2b)$$

(b) Combinations $C(n, k)$ is the number of ways that k objects can be selected from n objects *with the order not being important*. It is given by:

$$C(n, k) = \frac{n!}{(n - k)!k!} \equiv \binom{n}{k} \quad (2.3)$$

Note that the same equation also defines the *binomial* coefficients since the expansion of $(a+b)^n$ according to the Binomial theorem is

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k. \quad (2.4)$$

Example 2.2.1: (a) Calculate the number of ways in which three people from a group of seven people can be seated in a row.

This is a case of permutation since the order is important. The number of possible ways is:

$$P(7, 3) = \frac{7!}{(7 - 3)!} = \frac{(7) \cdot (6) \cdot (5)}{1} = 2110$$

(b) Calculate the number of combinations in which three people can be selected from a group of seven.

Here the order is not important and the combination formula can be used. Thus:

$$C(7, 3) = \frac{7!}{(7 - 3)!3!} = \frac{(7) \cdot (6) \cdot (5)}{(3) \cdot (2)} = 35 \quad \blacksquare$$

Another type of combinatorial problem is the *factorial problem* to be discussed in Chap. 6 while dealing with design of experiments. Consider a specific example involving *equipment scheduling* at a physical plant of a large campus which includes primemovers (diesel engines or turbines which produce electricity), boilers and chillers (vapor compression and absorption machines). Such equipment need a certain amount of time to come online and so operators typically keep some of them “idling” so that they can start supplying electricity/heating/cooling at a moment's notice. Their operating states can be designated by a binary variable; say “1” for on-status and “0” for off-status. Extensions of this concept include cases where, instead of two states, one could have m states. An example of 3 states is when say two identical boilers are to be scheduled. One could have three states altogether: (i) when both are off (0–0), (ii) when both are on (1–1), and (iii) when only one is on (1–0). Since the boilers are identical, state (iii) is identical to 0–1. In case, the two boilers are of different size, there would be four possible states. The number of combinations possible for “ n ” such equipment where each one can assume “ m ” states is given by m^n . Some simple cases for scheduling four different types of energy equipment in a physical plant are shown in Table 2.1.

2.2.2 Compound Events and Probability Trees

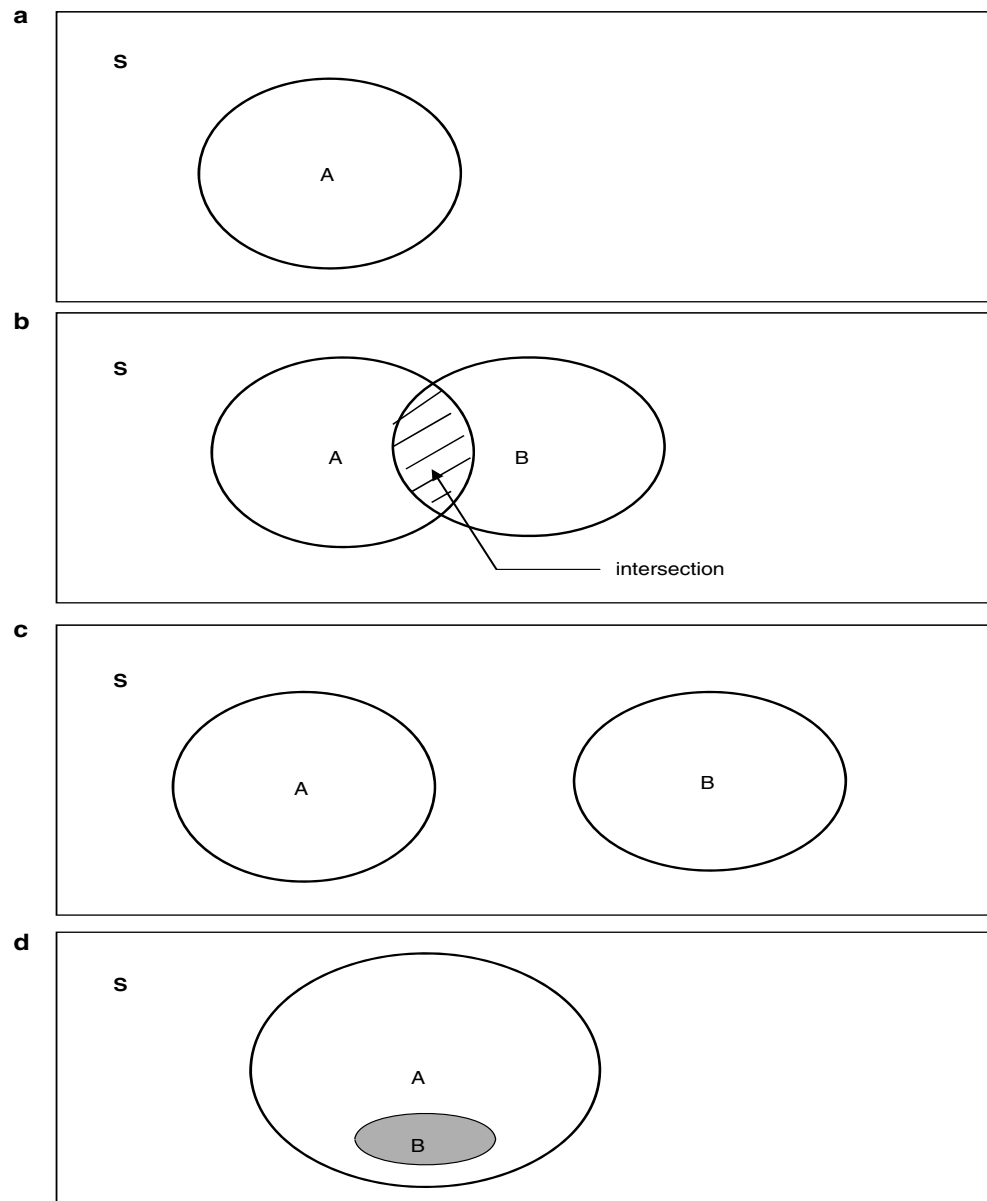
A compound or joint or composite event is one which arises from operations involving two or more events. The use of Venn's diagram is a very convenient manner of illustrating and understanding compound events and their probabilities (see Fig. 2.1).

Table 2.1 Number of combinations for equipment scheduling in a large facility

	Status (0- off, 1- on)				Number of Combinations
	Primemovers	Boilers	Chillers-Vapor compression	Chillers-Absorption	
One of each	0-1	0-1	0-1	0-1	$2^4=16$
Two of each-assumed identical	0-0, 0-1, 1-1	0-0, 0-1, 1-1	0-0, 0-1, 1-1	0-0, 0-1, 1-1	$3^4=81$
Two of each-non-identical except for boilers	0-0, 0-1, 1-0, 1-1	0-0, 0-1, 1-0	0-0, 0-1, 1-0, 1-1	0-0, 0-1, 1-0, 1-1	$4^3 \times 3^1=192$

- The universe of outcomes or sample space is denoted by a rectangle, while the probability of a particular event (say, event A) is denoted by a region (see Fig. 2.1a);
- *union* of two events A and B (see Fig. 2.1b) is represented by the set of outcomes in *either* A or B or both, and is denoted by $A \cup B$ (where the symbol \cup is conveniently remembered as “u” of “union”). An example is the number of cards in a pack which are either hearts or spades (26 nos.);
- *intersection* of two events A and B is represented by the set of outcomes in *both* A and B simultaneously, and is denoted by $A \cap B$. It is represented by the hatched area in Fig. 2.1b. An example is the number of red cards which are jacks (2 nos.);
- *mutually exclusive events* or disjoint events are those which have no outcomes in common (Fig. 2.1c). An example is the number of red cards with spades seven (nil);

Fig. 2.1 Venn diagrams for a few simple cases. **a** Event A is denoted as a region in space S. Probability of event A is represented by the area inside the circle to that inside the rectangle. **b** Events A and B are intersecting, i.e., have a common overlapping area (shown hatched). **c** Events A and B are mutually exclusive or are disjoint events. **d** Event B is a subset of event A



- event B is inclusive in event A when all outcomes of B are contained in those of A, i.e., B is a *sub-set* of A (Fig. 2.1d). An example is the number of cards less than six (event B) which are red cards (event A).

2.2.3 Axioms of Probability

Let the sample space S consist of two events A and B with probabilities $p(A)$ and $p(B)$ respectively. Then:

- (i) probability of any event, say A, cannot be negative. This is expressed as:

$$p(A) \geq 0 \quad (2.5)$$

- (ii) probabilities of all events must be unity (i.e., normalized):

$$p(S) \equiv p(A) + p(B) = 1 \quad (2.6)$$

- (iii) probabilities of mutually exclusive events add up:

$$p(A \cup B) = p(A) + p(B) \quad (2.7)$$

if A and B are mutually exclusive

If a dice is rolled, the outcomes are mutually exclusive. If event A is the occurrence of 2 and event B that of 3, then $p(A \text{ or } B) = 1/6 + 1/6 = 1/3$. Mutually exclusive events and independent events are not to be confused. While the former is a property of the events themselves, the latter is a property that arises from the event probabilities and their intersections (this is elaborated further below).

Some other inferred relations are:

- (iv) probability of the complement of event A:

$$p(\bar{A}) = 1 - p(A) \quad (2.8)$$

- (v) probability for *either* A or B (when they are not mutually exclusive) to occur is equal to:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) \quad (2.9)$$

This is intuitively obvious from the Venn diagram (see Fig. 2.1b) since the hatched area (representing $p(A \cap B)$) gets counted twice in the sum and, so needs to be deducted once. This equation can also be deduced from the axioms of probability. Note that if events A and B are mutually exclusive, then Eq. 2.9 reduces to Eq. 2.7.

2.2.4 Joint, Marginal and Conditional Probabilities

- (a) *Joint probability* of two independent events represents the case when both events occur together, i.e. $p(A \text{ and } B) = p(A \cap B)$. It is equal to:

$$p(A \cap B) = p(A) \cdot p(B) \quad (2.10)$$

if A and B are independent

These are called *product models*. Consider a dice tossing experiment. If event A is the occurrence of an even number, then $p(A) = 1/2$. If event B is that the number is less than or equal to 4, then $p(B) = 2/3$. The probability that both events occur when a dice is rolled is $p(A \text{ and } B) = 1/2 \times 2/3 = 1/3$. This is consistent with our intuition since events {2,4} would satisfy both the events.

(b) *Marginal probability* of an event A refers to the probability of A in a joint probability setting. For example, consider a space containing two events, A and B. Since S can be taken to be the sum of event space B and its complement \bar{B} , the probability of A can be expressed in terms of the sum of the disjoint parts of B:

$$p(A) = p(A \cap B) + p(A \cap \bar{B}) \quad (2.11)$$

This notion can be extended to the case of more than two joint events.

Example 2.2.2: Consider an experiment involving drawing two cards from a deck with replacement. Let event A = {first card is a red one} and event B = {card is between 2 and 8 inclusive}. How Eq. 2.11 applies to this situation is easily shown. Possible events A: hearts (13 cards) plus diamonds (13 cards)

Possible events B: 4 suites of 2, 3, 4, 5, 6, 7, 8.

$$\text{Also, } p(A \cap B) = \frac{1}{2} \cdot \frac{(7) \cdot (4)}{52} = \frac{14}{52} \text{ and}$$

$$p(A \cap \bar{B}) = \frac{1}{2} \cdot \frac{(13 - 7) \cdot (4)}{52} = \frac{12}{52}$$

$$\text{Consequently, from Eq. 2.11: } p(A) = \frac{14}{52} + \frac{12}{52} = \frac{1}{2}.$$

This result of $p(A) = 1/2$ is obvious in this simple experiment, and could have been deduced intuitively. However, intuition may mislead in more complex cases, and hence, the usefulness of this approach. ■

(c) *Conditional probability*: There are several situations involving compound outcomes that are sequential or successive in nature. The chance result of the first stage determines the conditions under which the next stage occurs. Such events, called *two-stage* (or multi-stage) events, involve step-by-step outcomes which can be represented as a *probability tree*. This allows better visualization of how the probabilities progress from one stage to the next. If A and B are events, then the probability that event B occurs given that A has already occurred is given by:

$$p(B/A) = \frac{p(A \cap B)}{p(A)} \quad (2.12)$$

A special but important case is when $p(B/A)=p(B)$. In this case, B is said to be *independent* of A because the fact that event A has occurred does not affect the probability of B occurring. Thus, two events A and B are mutually exclusive if $p(B/A)=p(B)$. In this case, one gets back Eq. 2.10.

An example of a conditional probability event is the drawing of a spade from a pack of cards from which a first card was already drawn. If it is known that the first card *was not* a spade, then the probability of drawing a spade the second time is $12/51=4/17$. On the other hand, if the first card drawn *was* a spade, then the probability of getting a spade on the second draw is $11/51$.

Example 2.2.3: A single fair dice is rolled. Let event A= {even outcome} and event B={outcome is divisible by 3}.

- List the various events in the sample space: {1 2 3 4 5 6}
- List the outcomes in A and find $p(A)$: {2 4 6}, $p(A)=1/2$
- List the outcomes of B and find $p(B)$: {3 6}, $p(B)=1/3$
- List the outcomes in $A \cap B$ and find $p(A \cap B)$: {6}, $p(A \cap B)=1/6$
- Are the events A and B independent? Yes, since Eq. 2.10 holds ■

Example 2.2.4: Two defective bulbs have been mixed with 10 good ones. Let event A= {first bulb is good}, and event B= {second bulb is good}.

- If two bulbs are chosen at random with replacement, what is the probability that both are good?
 $p(A)=8/10$ and $p(B)=8/10$. Then:

$$p(A \cap B) = \frac{8}{10} \cdot \frac{8}{10} = \frac{64}{100} = 0.64$$

- What is the probability that two bulbs drawn in sequence (i.e., not replaced) are good where the status of the bulb can be checked after the first draw?

From Eq. 2.12, $p(\text{both bulbs drawn are good})$:

$$p(A \cap B) = p(A) \cdot p(B/A) = \frac{8}{10} \cdot \frac{7}{9} = \frac{28}{45} = 0.622$$

Example 2.2.5: Two events A and B have the following probabilities: $p(A) = 0.3$, $p(B) = 0.4$ and $p(\bar{A} \cap B) = 0.28$.

- Determine whether the events A and B are independent or not?

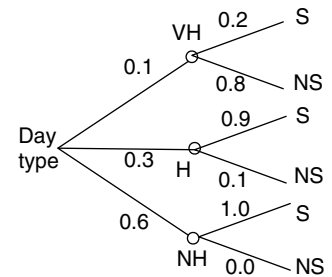
From Eq. 2.8, $P(\bar{A}) = 1 - p(A) = 0.7$. Next, one will verify whether Eq. 2.10 holds or not. In this case, one needs to verify whether: $p(\bar{A} \cap B) = p(\bar{A}) \cdot p(B)$ or whether 0.28 is equal to (0.7×0.4) . Since this is correct, one can state that events A and B are independent.

- Find $p(A \cup B)$

From Eqs. 2.9 and 2.10:

$$\begin{aligned} p(A \cup B) &= p(A) + p(B) - p(A \cap B) \\ &= p(A) + p(B) - p(A) \cdot p(B) \\ &= 0.3 + 0.4 - (0.3)(0.4) = 0.58 \end{aligned}$$

Fig. 2.2 The forward probability tree for the residential air-conditioner when two outcomes are possible (S satisfactory or NS not satisfactory) for each of three day-types (VH very hot, H hot and NH not hot)



Example 2.2.6: Generating a probability tree for a residential air-conditioning (AC) system.

Assume that the AC is slightly under-sized for the house it serves. There are two possible outcomes (S- satisfactory and NS- not satisfactory) depending on whether the AC is able to maintain the desired indoor temperature. The outcomes depend on the outdoor temperature, and for simplicity, its annual variability is grouped into three categories: very hot (VH), hot (H) and not hot (NH). The probabilities for outcomes S and NS to occur in each of the three day-type categories are shown in the probability tree diagram (Fig. 2.2) while the joint probabilities computed following Eq. 2.10 are assembled in Table 2.2.

Note that the relative probabilities of the three branches in both the first stage as well as in each of the two branches of each outcome add to unity (for example, in the Very Hot, the S and NS outcomes add to 1.0, and so on). Further, note that the joint probabilities shown in the table also have to sum to unity (it is advisable to perform such verification checks). The probability of the indoor conditions being satisfactory is determined as: $p(S)=0.02+0.27+0.6=0.89$ while $p(NS)=0.08+0.03+0=0.11$. It is wise to verify that $p(S)+p(NS)=1.0$. ■

Example 2.2.7: Consider a problem where there are two boxes with marbles as specified:

Box 1: 1 red and 1 white and Box 2: 4 red and 1 green

A box is chosen at random and a marble drawn from it. What is the probability of getting a red marble?

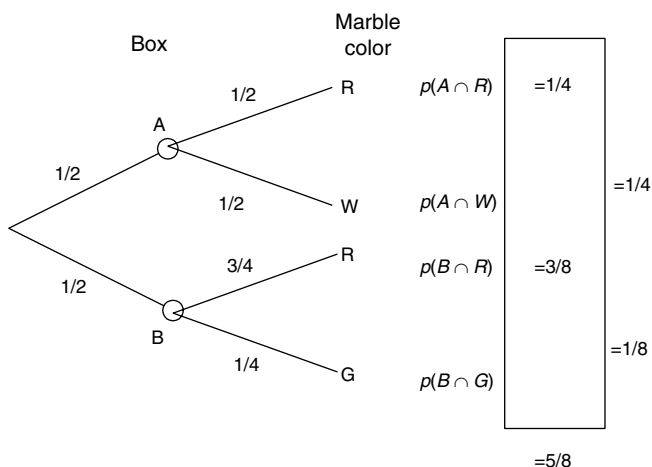
One is tempted to say that since there are 4 red marbles in total out of 6 marbles, the probability is $2/3$. However, this is incorrect, and the proper analysis approach requires that one frame this problem as a two-stage experiment. The first stage is the selection of the box, and the second the drawing

Table 2.2 Joint probabilities of various outcomes

$p(VH \cap S) = 0.1 \times 0.2 = 0.02$
$p(VH \cap NS) = 0.1 \times 0.8 = 0.08$
$p(H \cap S) = 0.3 \times 0.9 = 0.27$
$p(H \cap NS) = 0.3 \times 0.1 = 0.03$
$p(NH \cap S) = 0.6 \times 1.0 = 0.6$
$p(NH \cap NS) = 0.6 \times 0 = 0$

Table 2.3 Probabilities of various outcomes

$p(A \cap R) = 1/2 \times 1/2 = 1/4$	$p(B \cap R) = 1/2 \times 3/4 = 3/8$
$p(A \cap W) = 1/2 \times 1/2 = 1/4$	$p(B \cap W) = 1/2 \times 0 = 0$
$p(A \cap G) = 1/2 \times 0 = 0$	$p(B \cap G) = 1/2 \times 1/4 = 1/8$

**Fig. 2.3** The first stage of the forward probability tree diagram involves selecting a box (either A or B) while the second stage involves drawing a marble which can be red (R), white (W) or green (G) in color. The total probability of drawing a red marble is 5/8

of the marble. Let event A (or event B) denote choosing Box 1 (or Box 2). Let R, W and G represent red, white and green marbles. The resulting probabilities are shown in Table 2.3.

Thus, the probability of getting a red marble = $1/4 + 3/8 = 5/8$. The above example is depicted in Fig. 2.3 where the reader can visually note how the probabilities propagate through the probability tree. This is called

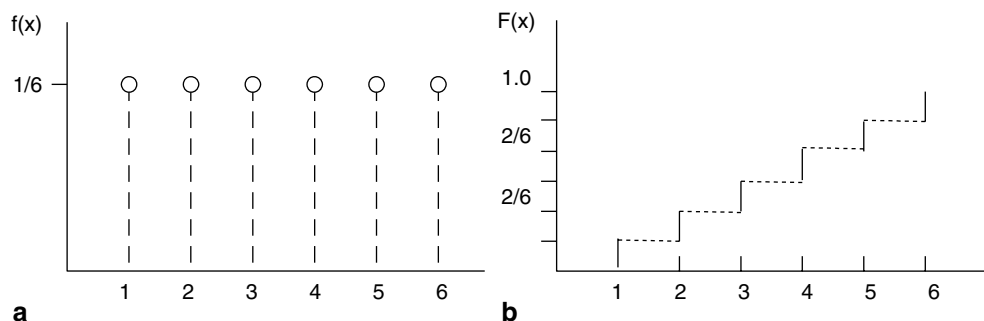
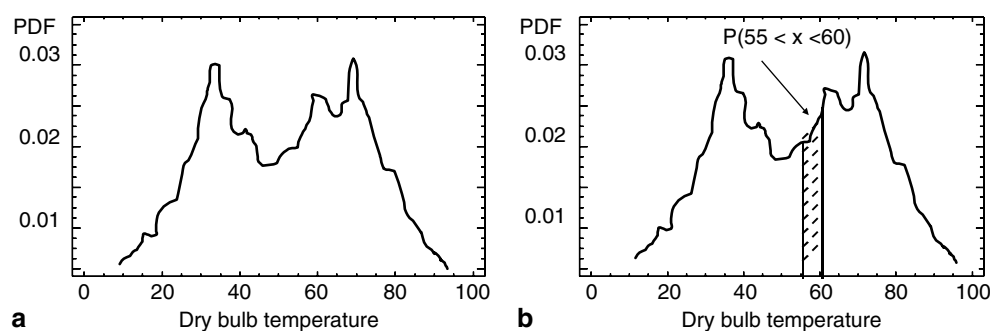
the “*forward tree*” to differentiate it from the “*reverse*” tree discussed in Sect. 2.5.

The above example illustrates how a two-stage experiment has to be approached. First, one selects a box which by itself does not tell us whether the marble is red (since one has yet to pick a marble). Only after a box is selected, can one use the prior probabilities regarding the color of the marbles inside the box in question to determine the probability of picking a red marble. These prior probabilities can be viewed as conditional probabilities; i.e., for example, $p(A \cap R) = p(R/A) \cdot p(A)$ ■

2.3 Probability Distribution Functions

2.3.1 Density Functions

The notions of discrete and continuous random variables were introduced in Sect. 2.1.1. The distribution of a random variable represents the probability of it taking its various possible values. For example, if the y-axis in Fig. 1.1 of the dice rolling experiment were to be changed into a relative frequency (= 1/6), the resulting histogram would graphically represent the corresponding *probability density function (PDF)* (Fig. 2.4a). Thus, the probability of getting a 2 in the rolling of a dice is 1/6th. Since, this is a discrete random variable, the function takes on specific values at discrete points of the x-axis (which represents the outcomes). The same type of y-axis normalization done to the data shown in Fig. 1.2 would result in the PDF for the case of continuous random data. This is shown in Fig. 2.5a for the random variable taken to be the hourly outdoor dry bulb temperature over the year at Phila-

Fig. 2.4 Probability functions for a discrete random variable involving the outcome of rolling a dice. **a** Probability density function. **b** Cumulative distribution function**Fig. 2.5** Probability density function and its association with probability for a continuous random variable involving the outcomes of hourly outdoor temperatures at Philadelphia, PA during a year. The probability that the temperature will be between 55° and 60°F is given by the shaded area. **a** Density function. **b** Probability interpreted as an area

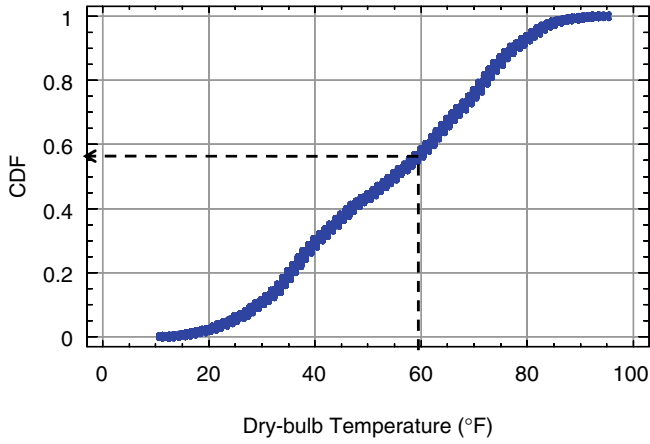


Fig. 2.6 The cumulative distribution function (CDF) for the PDF shown in Fig. 2.5. Such a plot allows one to easily determine the probability that the temperature is less than 60°F

delphia, PA. Notice that this is the envelope of the histogram of Fig. 1.2. Since the variable is continuous, it is implausible to try to determine the probability of, say temperature outcome of 57.5°F. One would be interested in the probability of outcomes within a range, say 55°–60°F. The probability can then be determined as the area under the PDF as shown in Fig. 2.5b. It is for such continuous random variables that the *cumulative distribution function (CDF)* is useful. It is simply the cumulative area under the curve starting from the lowest value of the random variable to the current value (Fig. 2.6). The vertical scale directly gives the probability (or, in this case, the fractional time) that X is less than or greater than a certain value. Thus, the probability ($x \leq 60$) is about 0.58. The concept of CDF also applies to discrete variables as illustrated in Fig. 2.4b for the dice rolling example.

To restate, depending on whether the random variable is discrete or continuous, one gets discrete or continuous probability distributions. Though most experimentally gathered data is discrete, the underlying probability theory is based on the data being continuous. Replacing the integration sign by the summation sign in the equations that follow allows extending the following definitions to discrete distributions. Let $f(x)$ be the probability distribution function associated with a random variable X . This is a function which provides the probability that a discrete random variable X takes on some specific value x among its various possible values. The axioms of probability (Eqs. 2.5 and 2.6) for the discrete case are expressed for the case of continuous random variables as:

- PDF cannot be negative:

$$f(x) \geq 0 \quad -\infty < x < \infty \quad (2.13)$$

- Probability of the sum of all outcomes must be unity

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad (2.14)$$

The cumulative distribution function (CDF) or $F(a)$ represents the area under $f(x)$ enclosed in the range $-\infty < x < a$:

$$F(a) = p\{X \leq a\} = \int_{-\infty}^a f(x) dx \quad (2.15)$$

The inverse relationship between $f(x)$ and $F(a)$, provided a derivative exists, is:

$$f(x) = \frac{dF(x)}{dx} \quad (2.16)$$

This leads to the probability of an outcome $a \leq X \leq b$ given by:

$$\begin{aligned} p\{a \leq X \leq b\} &= \int_a^b f(x) dx \\ &= \int_{-\infty}^b f(x) dx - \int_{-\infty}^a f(x) dx \\ &= F(b) - F(a) \end{aligned} \quad (2.17)$$

Notice that the CDF for discrete variables will be a step function (as in Fig. 2.4b) since the PDF is defined at discrete values only. Also, the CDF for continuous variables is a function which increases monotonically with increasing x . For example, the probability of the outdoor temperature being between 55° and 60°F is given by $p\{55 \leq X \leq 60\} = F(b) - F(a) = 0.58 - 0.50 = 0.08$ (see Fig. 2.6).

The concept of probability distribution functions can be extended to the treatment of simultaneous outcomes of multiple random variables. For example, one would like to study how temperature of quenching of a particular item made of steel affects its hardness. Let X and Y be the two random variables. The probability that they occur together can be represented by a function $f(x, y)$ for any pair of values (x, y) within the range of variability of the random variables X and Y . This function is referred to as the *joint probability density function* of X and Y which has to satisfy the following properties for continuous variables:

$$f(x, y) \geq 0 \quad \text{for all } (x, y) \quad (2.18)$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1 \quad (2.19)$$

$$p[(X, Y) \in A] = \int \int_A f(x, y) dx dy \quad (2.20)$$

where A is any region in the xy plane.

If X and Y are two *independent* random variables, their joint PDF will be the product of their marginal ones:

$$f(x, y) = f(x) \cdot f(y) \quad (2.21)$$

Note that this is the continuous variable counterpart of Eq. 2.10 which gives the joint probability of two discrete events.

The *marginal distribution* of X given two jointly distributed random variables X and Y is simply the probability distribution of X ignoring that of Y. This is determined for X as:

$$g(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad (2.22)$$

Finally, the *conditional probability distribution* of X given that $X=x$ for two jointly distributed random variables X and Y is:

$$f(y/x) = \frac{f(x, y)}{g(x)} \quad g(x) > 0 \quad (2.23)$$

Example 2.3.1: Determine the value of c so that each of the following functions can serve as probability distributions of the discrete random variable X:

$$(a) \quad f(x) = c(x^2 + 4) \quad \text{for } x = 0, 1, 2, 3$$

$$(b) \quad f(x) = ax^2 \quad \text{for } -1 < x < 2$$

(a) One uses the discrete version of Eq. 2.14, i.e.,

$$\sum_{i=0}^3 f(x_i) = 1 \text{ leads to } 4c + 5c + 8c + 13c = 1 \text{ from which}$$

$$c = 1/30$$

(b) One uses Eq. 2.14 modified for the limiting range in x:

$$\int_{-1}^2 ax^2 dx = 1 \text{ from which } \left[\frac{ax^3}{3} \right]_{-1}^2 = 1 \text{ resulting in}$$

$$a = 1/3 \quad \blacksquare$$

Example 2.3.2: The operating life in weeks of a high efficiency air filter in an industrial plant is a random variable X having the PDF:

$$f(x) = \frac{20}{(x + 100)^3} \quad \text{for } x > 0$$

Find the probability that the filter will have an operating life of:

(a) at least 20 weeks

(b) anywhere between 80 and 120 weeks

First, determine the expression for the CDF from Eq. 2.14. Since the operating life would decrease with time, one needs to be careful about the limits of integration applicable to this case. Thus,

$$\text{CDF} = \int_x^0 \frac{20}{(x' + 100)^3} dx' = \left[-\frac{10}{(x' + 100)^2} \right]_x^0$$

(a) with $x=20$, the probability that the life is at least 20 weeks:

$$p(20 < X < \infty) = \left[-\frac{10}{(x + 100)^2} \right]_{20}^{\infty} = 0.000694$$

(b) for this case, the limits of integration are simply modified as follows:

$$p(80 < X < 120) = \left[-\frac{10}{(x + 100)^2} \right]_{80}^{120} = 0.000102 \quad \blacksquare$$

Example 2.3.3: Consider two random variables X and Y with the following joint density function:

$$f(x, y) = \frac{2}{5}(2x + 3y)$$

for $0 \leq x \leq 1, 0 \leq y \leq 1$

(a) Verify whether the normalization criterion is satisfied. This is easily verified from Eq. 2.19:

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy &= \int_0^1 \int_0^1 \frac{2}{5}(2x + 3y) dx dy \\ &= \int_0^1 \left[\frac{2x^2}{5} + \frac{6xy}{5} \right]_{x=0}^{x=1} dy \\ &= \int_0^1 \left(\frac{2}{5} + \frac{6y}{5} \right) dy = \frac{2}{5} + \frac{3}{5} = 1 \end{aligned}$$

(b) Determine the joint probability in the region $(0 < x < 1/2, 1/4 < y < 1/2)$. In this case, one uses Eq. 2.20 as follows:

$$\begin{aligned} p(0 < X < 1/2, 1/4 < Y < 1/2) &= \int_{1/4}^{1/2} \int_0^{1/2} \frac{2}{5}(2x + 3y) dx dy \\ &= \frac{13}{160} \end{aligned}$$

(c) Determine the marginal distribution $g(x)$. From Eq. 2.22:

$$g(x) = \int_0^1 \frac{2}{5}(2x + 3y) dy = \left[\frac{4xy}{5} + \frac{6y^2}{10} \right]_{y=0}^{y=1} = \frac{4x + 3}{5} \quad \blacksquare$$

Table 2.4 Computing marginal probabilities from a probability table

Age (Y)	Income (X)			Marginal probability of Y
	>\$ 40,000	40,000–90,000	<90,000	
Under 25	0.15	0.09	0.05	0.29
Between 25–40	0.10	0.16	0.12	0.38
Above 40	0.08	0.20	0.05	0.33
Marginal probability of X	0.33	0.45	0.22	Should sum to 1.00 both ways

Example 2.3.4: The percentage data of annual income versus age has been gathered from a large population living in a certain region— see Table 2.4. Let X be the income and Y the age. The marginal probability of X for each class is simply the sum of the probabilities under each column and that of Y the sum of those for each row. Thus, $p(X \geq 40,000) = 0.15 + 0.10 + 0.08 = 0.33$, and so on. Also, verify that the sum of the marginal probabilities of X and Y sum to 1.00 (so as to satisfy the normalization condition). ■

2.3.2 Expectation and Moments

This section deals with ways by which one can summarize the characteristics of a probability function using a few important measures. Commonly, the mean or the expected value $E[X]$ is used as a measure of the central tendency of the distribution, and the variance $\text{var}[X]$ as a measure of dispersion of the distribution about its mean. These are very similar to the notions of arithmetic mean and variance of a set of data. As before, the equations which apply to continuous random variables are shown below; in case of discrete variables, the integrals have to be replaced with summations.

- *expected value of the first moment or mean*

$$E[X] \equiv \mu = \int_{-\infty}^{\infty} xf(x)dx. \quad (2.24)$$

The mean is exactly analogous to the physical concept of center of gravity of a mass distribution. This is the reason why PDF are also referred to as the “mass distribution function”. The concept of symmetry of a PDF is an important one implying that the distribution is symmetric about the mean. A distribution is symmetric if: $p(\mu - x) = p(\mu + x)$ for every value of x .

- *variance*

$$\text{var}[X] \equiv \sigma^2 = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx \quad (2.25a)$$

Alternatively, it can be shown that for any discrete distribution:

$$\text{var}[X] = E[X^2] - \mu^2 \quad (2.25b)$$

Notice the appearance of the expected value of the second moment $E[X^2]$ in the above equation. The variance is analogous to the physical concept of the moment of inertia of a mass distribution about its center of gravity.

In order to express the variance which is a measure of dispersion in the same units as the random variable itself, the square root of the variance, namely the *standard deviation* σ is used. Finally, errors have to be viewed, or evaluated, in terms of the magnitude of the random variable. Thus, the *relative error* is often of more importance than the actual error. This has led to the widespread use of a dimensionless quantity called the *Coefficient of Variation* (CV) defined as the percentage ratio of the standard deviation to the mean:

$$CV = 100 \cdot \left(\frac{\sigma}{\mu}\right) \quad (2.26)$$

2.3.3 Function of Random Variables

The above definitions can be extended to the case when the random variable X is a function of several random variables; for example:

$$X = a_0 + a_1 X_1 + a_2 X_2 \dots \quad (2.27)$$

where the a_i coefficients are constants and X_i are random variables.

Some important relations regarding the mean:

$$\begin{aligned} E[a_0] &= a_0 \\ E[a_1 X_1] &= a_1 E[X_1] \\ E[a_0 + a_1 X_1 + a_2 X_2] &= a_0 + a_1 E[X_1] + a_2 E[X_2] \end{aligned} \quad (2.28)$$

Similarly there are a few important relations that apply to the variance:

$$\begin{aligned} \text{var}[a_0] &= 0 \\ \text{var}[a_1 X_1] &= a_1^2 \text{var}[X_1] \end{aligned} \quad (2.29)$$

Again, if the two random variables are independent,

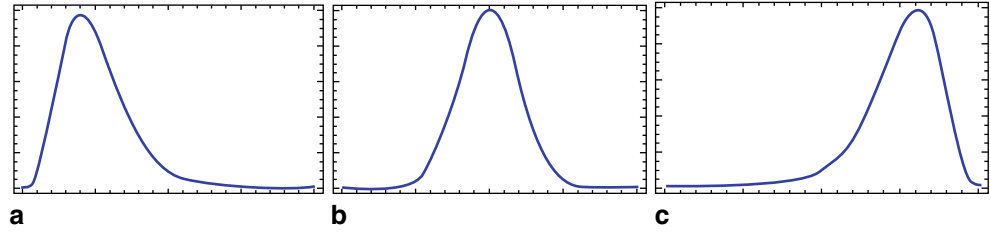
$$\text{var}[a_0 + a_1 X_1 + a_2 X_2] = a_1^2 \text{var}[X_1] + a_2^2 \text{var}[X_2] \quad (2.30)$$

The notion of *covariance* of two random variables is an important one since it is a measure of the tendency of two random variables to vary together. The covariance is defined as:

$$\text{cov}[X_1, X_2] = E[(X_1 - \mu_1) \cdot (X_2 - \mu_2)] \quad (2.31)$$

where μ_1 and μ_2 are the mean values of the random variables X_1 and X_2 respectively. Thus, for the case of two random

Fig. 2.7 Skewed and symmetric distributions. **a** Skewed to the right. **b** Symmetric. **c** Skewed to the left



variables which are not independent, Eq. 2.30 needs to be modified into:

$$\text{var}[a_0 + a_1X_1 + a_2X_2] = a_1^2\text{var}[X_1] + a_2^2\text{var}[X_2] + 2a_1a_2\text{cov}[X_1, X_2] \quad (2.32)$$

Moments higher than the second are sometimes used. For example, the third moment yields the *skewness* which is a measure of the symmetry of the PDF. Figure 2.7 shows three distributions: one skewed to the right, a symmetric distribution, and one skewed to the left. The fourth moment yields the *coefficient of kurtosis* which is a measure of the peakiness of the PDF.

Two commonly encountered terms are the median and the mode. The value of the random variable at which the PDF has a peak is the *mode*, while the *median* divides the PDF into two equal parts (each part representing a probability of 0.5).

Finally, distributions can also be described by the number of “humps” they display. Figure 2.8 depicts the case of uni-modal and bi-modal distributions, while Fig. 2.5 is the case of a distribution with three humps.

Example 2.3.5: Let X be a random variable representing the number of students who fail a class. Its PDF is given in Table 2.5.

The discrete event form of Eqs. 2.24 and 2.25 is used to compute the mean and the variance:

$$\begin{aligned} \mu &= (0)(0.51) + (1)(0.38) + (2)(0.10) + (3)(0.01) \\ &= 0.61 \end{aligned}$$

Further:

$$\begin{aligned} E(X^2) &= (0)(0.51) + (1^2)(0.38) + (2^2)(0.10) + (3^2)(0.01) \\ &= 0.87 \end{aligned}$$

Hence:

$$\sigma^2 = 0.87 - (0.61)^2 = 0.4979 \quad \blacksquare$$

Table 2.5 PDF of number of students failing a class

X	0	1	2	3
$f(x)$	0.51	0.38	0.10	0.01

Example 2.3.6: Consider Example 2.3.2 where a PDF of X is defined. Let $g(x)$ be a function of this PDF such that $g(x) = 4x + 3$.

One wishes to determine the expected value of $g(X)$. From Eq. 2.24,

$$E[f(x)] = \int_{-\infty}^{\infty} \frac{20x}{(x+100)^3} dx = 0.1$$

Then from Eq. 2.28

$$E[g(X)] = 3 + 4 \cdot E[f(X)] = 3.4 \quad \blacksquare$$

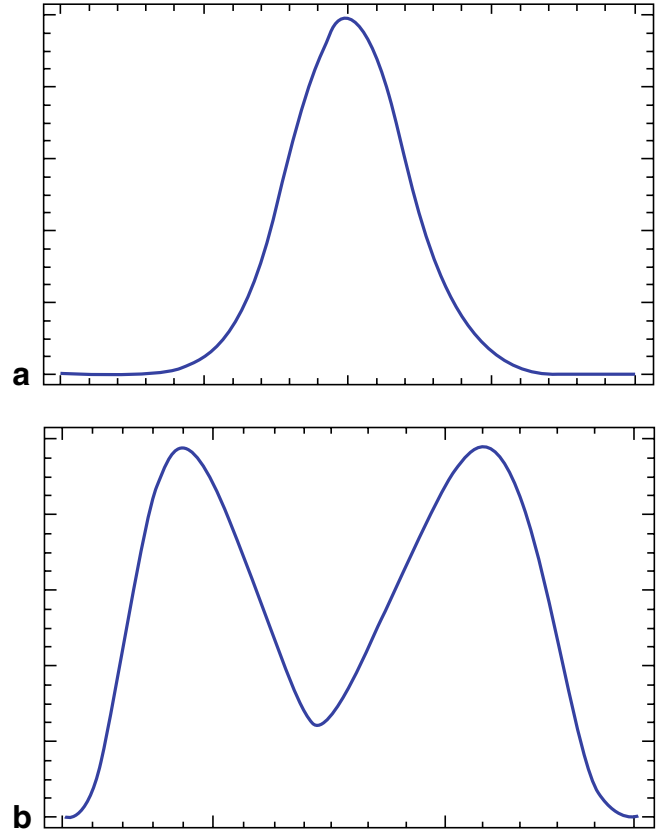


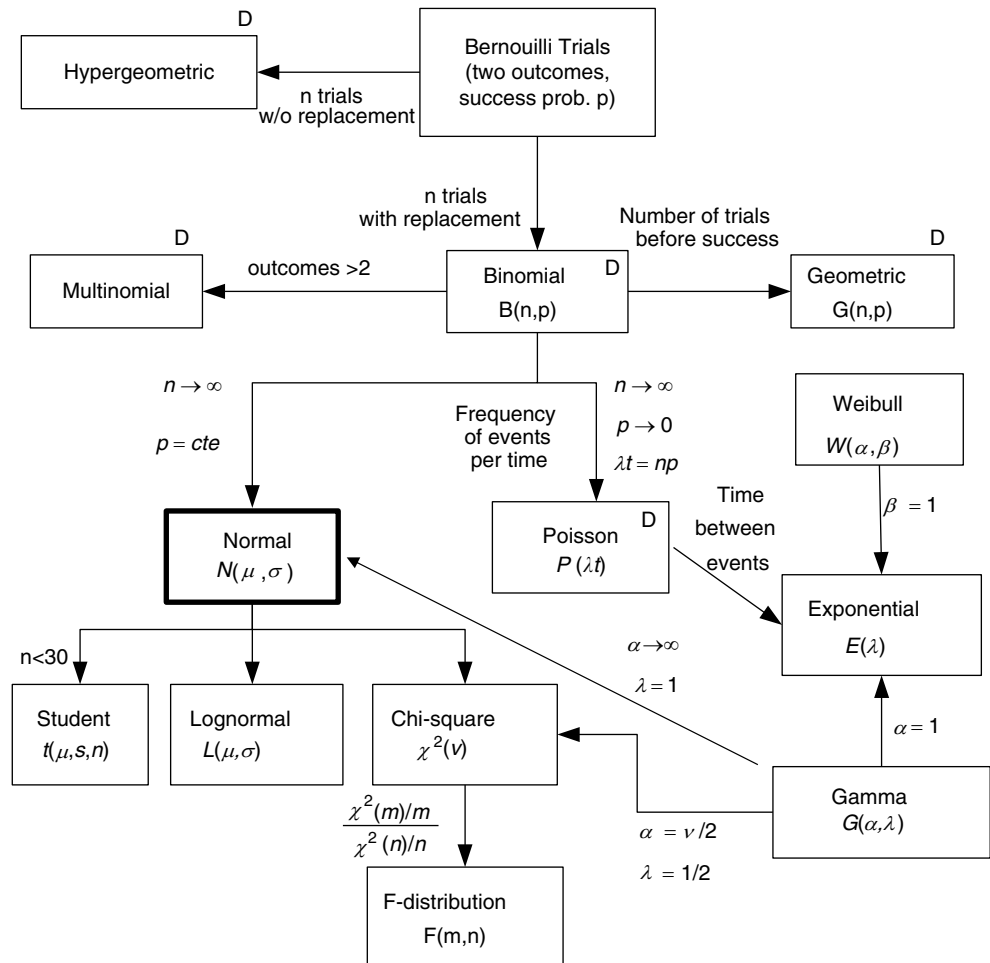
Fig. 2.8 Unimodal and bi-modal distributions. **a** Unimodal, **b** Bi-modal

2.4 Important Probability Distributions

2.4.1 Background

Data arising from an occurrence or phenomenon or descriptive of a class or a group can be viewed as a distribution of a random variable with a PDF associated with it. A majority of data sets encountered in practice can be described by one (or two) among a relatively few PDFs. The ability to characterize data in this manner provides distinct advantages to the analysts in terms of: understanding the basic dynamics of the phenomenon, in prediction and confidence interval specification, in classification, and in hypothesis testing (discussed in Chap. 4). Such insights eventually allow better decision making or sounder structural model identification since they provide a means of quantifying the random uncertainties inherent in the data. Surprisingly, most of the commonly encountered or important distributions have a common genealogy, shown in Fig. 2.9 which is a useful mnemonic for the reader.

Fig. 2.9 Genealogy between different important probability distribution functions. Those that are discrete functions are represented by “D” while the rest are continuous functions. (Adapted with modification from R. E. Lave, Jr. of Stanford University)



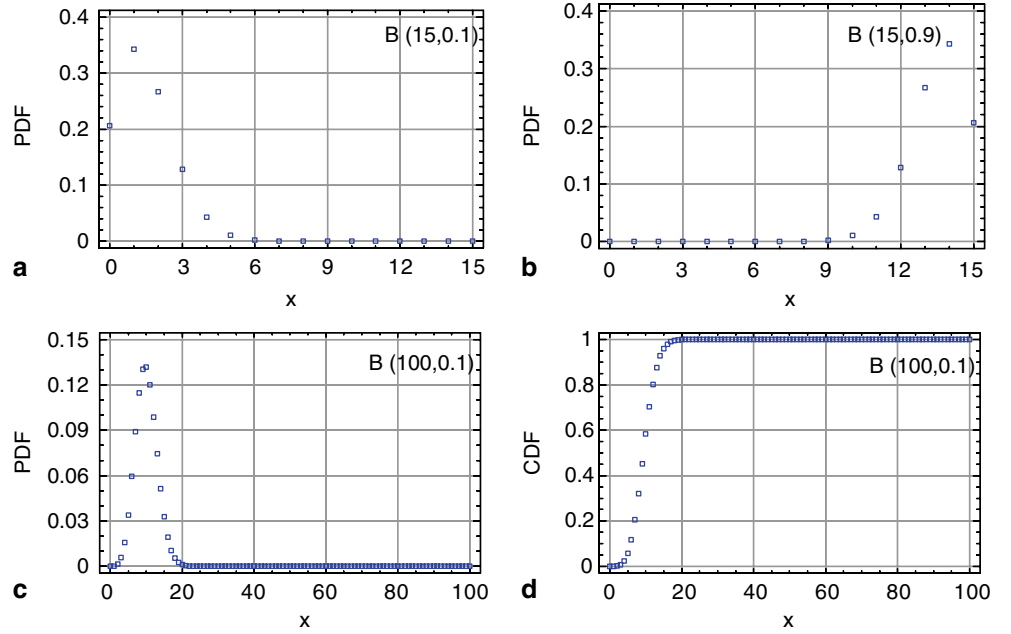
2.4.2 Distributions for Discrete Variables

(a) Bernoulli Process. Consider an experiment involving repeated trials where only two complementary outcomes are possible which can be labeled either as a “success” or a “failure”. Such a process is called a Bernoulli process: (i) if the successive trials are independent, and (ii) if the probability of success p remains constant from one trial to the next. Note that the number of partitions or combinations of n outcomes into two groups with x in one group and $(n-x)$ in the other is equal to

$$C(n, x) = \binom{n}{x}$$

(b) Binomial Distribution. The number of successes in n Bernoulli trials is called a binomial random variable. Its PDF is called a Binomial distribution (so named because of its association with the terms of the binomial expansion). It is a unimodal distribution which gives the probability of x successes in n independent trials, if the probability of success in any one trial is p . Note that the outcomes must be Bernoulli trials. This distribution is given by:

Fig. 2.10 Plots for the Binomial distribution illustrating the effect of probability of success p with X being the probability of the number of “successes” in a total number of n trials. Note how the skewness in the PDF is affected by p (frames a and b), and how the number of trials affects the shape of the PDF (frame a and c). Instead of vertical bars at discrete values of X as is often done for discrete distributions such as the Binomial, the distributions are shown as contour points so as to be consistent with how continuous distributions are represented. **a** $n=15$ and $p=0.1$, **b** $n=15$ and $p=0.9$, **c** $n=100$ and $p=0.1$, **d** $n=100$ and $p=0.1$



$$B(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x} \quad (2.33a)$$

with mean: $\mu = (n \cdot p)$ and variance

$$\sigma^2 = np(1-p) \quad (2.33b)$$

When n is small, it is easy to compute Binomial probabilities using Eq. 2.33a. For large values of n , it is more convenient to refer to tables which apply not to the PDF but to the corresponding cumulative distribution functions, referred to here as *Binomial probability sums*, defined as:

$$B(r; n, p) = \sum_{x=0}^r B(x; n, p) \quad (2.33c)$$

There can be numerous combinations of n , p and r , which is a drawback to such tabular determinations. Table A1 in Appendix A illustrates the concept only for $n=15$ and $n=20$ and for different values of p and r . Figure 2.10 illustrates how the skewness of the Binomial distribution is affected by p , and by the total number of trials n .

Example 2.4.1: Let k be the number of heads in $n=4$ independent tosses of a coin. Then the mean of the distribution $= (4) \cdot (1/2) = 2$, and the variance $\sigma^2 = (4) \cdot (1/2) \cdot (1-1/2) = 1$. From Eq. 2.33a, the probability of two successes in four tosses =

$$\begin{aligned} B(2; 4, 0.5) &= \binom{4}{2} \left(\frac{1}{2}\right)^2 \left(1 - \frac{1}{2}\right)^{4-2} \\ &= \frac{4 \times 3}{2} \times \frac{1}{4} \times \frac{1}{4} = \frac{3}{8} \end{aligned}$$

Example 2.4.2: The probability that a patient recovers from a type of cancer is 0.6. If 15 people are known to have contracted this disease, then one can determine probabilities of various types of cases using Table A1. Let X be the number of people who survive.

(a) The probability that at least 5 survive is:

$$\begin{aligned} p(X \geq 5) &= 1 - p(X < 5) = 1 - \sum_{x=0}^4 B(x; 15, 0.6) \\ &= 1 - 0.0094 = 0.9906 \end{aligned}$$

(b) The probability that there will be 5 to 8 survivors is:

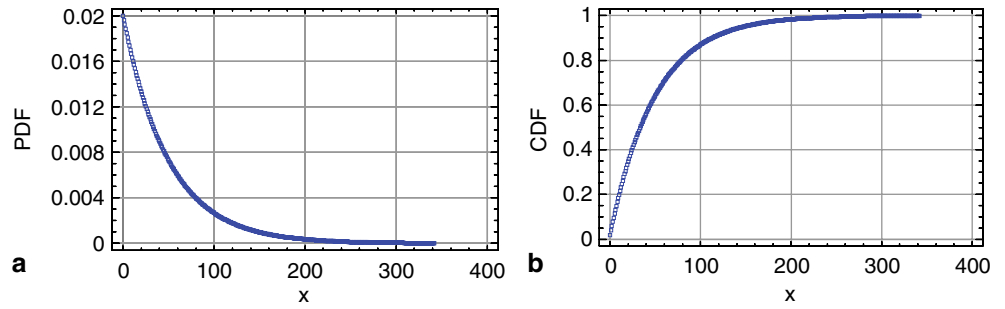
$$\begin{aligned} P(5 \leq X \leq 8) &= \sum_{x=5}^8 B(x; 15, 0.6) - \sum_{x=0}^4 B(x; 15, 0.6) \\ &= 0.3902 - 0.0094 = 0.3808 \end{aligned}$$

(c) The probability that exactly 5 survive:

$$\begin{aligned} p(X = 5) &= \sum_{x=0}^5 B(x; 15, 0.6) - \sum_{x=0}^4 B(x; 15, 0.6) \\ &= 0.0338 - 0.0094 = 0.0244 \end{aligned}$$

(c) Geometric Distribution. Rather than considering the number of successful outcomes, there are several physical instances where one would like to ascertain the time interval for a certain probability event to occur the first time (which could very well destroy the physical system). This probability (p) is given by the geometric distribution which can be derived from the Binomial distribution. Consider N to be the random variable representing the number of trials until the event does occur. Note that if an event occurs the first

Fig. 2.11 Geometric distribution $G(x; 0.02)$ for Example 2.4.3 where the random variable is the number of trials until the event occurs, namely the “50 year design wind” at the coastal location in question. **a** PDF. **b** CDF



time during the n^{th} trial then it did not occur during the previous $(n-1)$ trials. Then, the geometric distribution is given by:

$$G(n; p) = p \cdot (1 - p)^{n-1} \quad n = 1, 2, 3, \dots \quad (2.34a)$$

An extension of the above concept relates to the time between two successive occurrences of the same event, called the recurrence time. Since the events are assumed independent, the mean recurrence time denoted by random variable T between two consecutive events is simply the expected value of the Bernoulli distribution:

$$\begin{aligned} \bar{T} = E(T) &= \sum_{t=1}^{\infty} t \cdot p(1-p)^{t-1} \\ &= p[1 + 2(1-p) + 3(1-p)^2] \approx \frac{1}{p} \end{aligned} \quad (2.34b)$$

Example 2.4.3:¹ Using geometric PDF for 50 year design wind problems

The design code for buildings in a certain coastal region specifies the 50-year wind as the “design wind”, i.e., a wind velocity with a return period of 50 years, or one which may be expected to occur once every 50 years. What are the probabilities that:

- the design wind is encountered in any given year. From Eq. 2.34b, $p = \frac{1}{\bar{T}} = \frac{1}{50} = 0.02$
- the design wind is encountered during the fifth year of a newly constructed building (from Eq. 2.34a): $G(5; 0.02) = (0.02) \cdot (1 - 0.02)^4 = 0.018$
- the design wind is encountered within the first 5 years:

$$\begin{aligned} G(n \leq 5; p) &= \sum_{t=1}^5 (0.02) \cdot (1 - 0.02)^{t-1} = 0.02 \\ &\quad + 0.0196 + 0.0192 + 0.0188 + 0.0184 = 0.096 \end{aligned}$$

Figure 2.11 depicts the PDF and the CDF for the geometric function corresponding to this example. ■

(d) Hypergeometric Distribution. The Binomial distribution applies in the case of independent trials or when sampling from a batch of items is done with replacement. Another type of dependence arises when sampling is done *without* replacement. This case occurs frequently in areas such as acceptance sampling, electronic testing and quality assurance where the item is destroyed during the process of testing. If n items are to be selected without replacement from a set of N items which contain k items that pass a success criterion, the PDF of the number X of successful items is given by the hypergeometric distribution:

$$\begin{aligned} H(x; N, n, k) &= \frac{C(k, x) \cdot C(N - k, n - x)}{C(N, n)} \\ &= \frac{\binom{k}{x} \binom{N - k}{n - x}}{\binom{N}{n}} \end{aligned} \quad (2.35a)$$

$x = 0, 1, 2, 3, \dots$

$$\begin{aligned} \text{with mean } \mu &= \frac{nk}{N} \text{ and} \\ \text{variance } \sigma^2 &= \frac{N - n}{N - 1} \cdot n \cdot \frac{k}{N} \cdot \left(1 - \frac{k}{N}\right) \end{aligned} \quad (2.35b)$$

Note that $C(k, x)$ is the number of ways x items can be chosen from the k “successful” set, while $C(N - k, n - x)$ is the number of ways that the remainder $(n - x)$ items can be chosen from the “unsuccessful” set of $(N - k)$ items. Their product divided by the total number of combinations of selecting equally likely samples of size n from N items is represented by Eq. 2.35a.

Example 2.4.4: Lots of 10 computers each are called acceptable if they contain no fewer than 2 defectives. The procedure for sampling the lot is to select 5 computers at random and test for defectives. What is the probability that exactly one defective is found in the sample if there are 2 defectives in the entire lot?

Using the hypergeometric distribution given by Eq. 2.35a with $n=5$, $N=10$, $k=2$ and $x=1$:

¹ From Ang and Tang (2007) by permission of John Wiley and Sons.

$$H(1; 10, 5, 2) = \frac{\binom{2}{1} \binom{10-2}{5-1}}{\binom{10}{5}} = 0.444 \quad \blacksquare$$

(e) Multinomial Distribution. A logical extension to Bernoulli experiments where the result is a two-way outcome, either success/good or failure/defective, is the multinomial experiment where k possible outcomes are possible. An example of $k=5$ is when the grade of a student is either A, B, C, D or F. The issue here is to find the number of combinations of n items which can be partitioned into k independent groups (a student can only get a single grade for the same class) with x_1 being in the first group, x_2 in the second, ... This is represented by:

$$\binom{n}{x_1, x_2, \dots, x_k} = \frac{n!}{x_1! x_2! \dots x_k!} \quad (2.36a)$$

with the conditions that $(x_1 + x_2 + \dots + x_k) = n$ and that all partitions are mutually exclusive and occur with equal probability from one trial to the next. It is intuitively obvious that when n is large and k is small, the hypergeometric distribution will tend to closely approximate the Binomial.

Just like Bernoulli trials lead to the Binomial distribution, the multinomial experiment leads to the multinomial distribution which gives the probability distribution of k random variables x_1, x_2, \dots, x_k in n independent trials occurring with probabilities p_1, p_2, \dots, p_k :

$$f(x_1, x_2, \dots, x_k) = \binom{n}{x_1, x_2, \dots, x_k} p_1^{x_1} \cdot p_2^{x_2} \dots p_k^{x_k} \quad (2.36b)$$

$$\text{with } \sum_{i=1}^k x_i = n \quad \text{and} \quad \sum_{i=1}^k p_i = 1$$

Example 2.4.5: Consider an examination given to 10 students. The instructor, based on previous years' experience, expects the distribution given in Table 2.6.

On grading the exam, he finds that 5 students got an A, 3 got a B and 2 got a C, and no one got either D or F. What is the probability that such an event could have occurred purely by chance?

This answer is directly provided by Eq. 2.36b which yields the corresponding probability of the above event taking place:

Table 2.6 PDF of student grades for a class

X	A	B	C	D	F
p(X)	0.2	0.3	0.3	0.1	0.1

$$f(A, B, C, D, F) = \binom{10}{5, 3, 2, 0, 0} (0.2^5) \cdot (0.3^3) \cdot (0.3^2) \cdot (0.1^0) \cdot (0.1^0) \approx 0.00196$$

This is very low, and hence this occurrence is unlikely to have occurred purely by chance. \blacksquare

(f) Poisson Distribution. Poisson experiments are those that involve the number of outcomes of a random variable X which occur per unit time (or space); in other words, as describing the occurrence of isolated events in a continuum. A Poisson experiment is characterized by: (i) independent outcomes (also referred to as memoryless), (ii) probability that a single outcome will occur during a very short time is proportional to the length of the time interval, and (iii) probability that more than one outcome occurs during a very short time is negligible. These conditions lead to the Poisson distribution which is the limit of the Binomial distribution when $n \rightarrow \infty$ and $p \rightarrow 0$ in such a way that the product $(n \cdot p) = \lambda t$ remains constant. It is given by:

$$p(x; \lambda t) = \frac{(\lambda t)^x \exp(-\lambda t)}{x!} \quad x = 0, 1, 2, 3 \dots \quad (2.37a)$$

where λ is called the “mean occurrence rate”, i.e., the average number of occurrences of the event per unit time (or space) interval t . A special feature of this distribution is that its mean or average number of outcomes μ per time t and its variance σ^2 are such that

$$\mu(X) = \sigma^2(X) = \lambda t = n \cdot p \quad (2.37b)$$

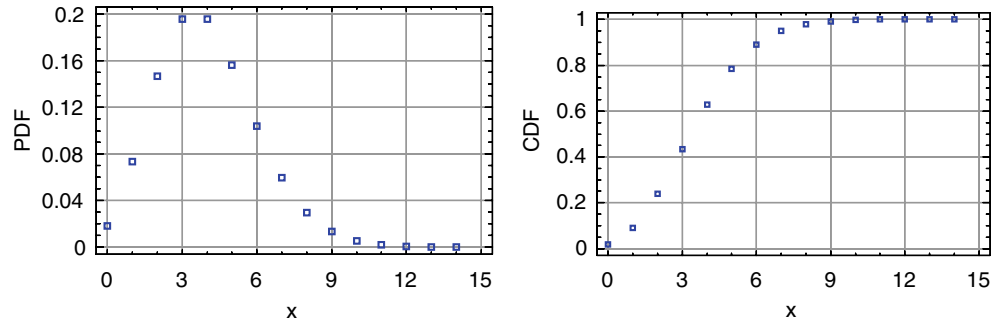
Akin to the Binomial distribution, tables for certain combinations of the two parameters allow the cumulative Poisson distribution to be read off directly (see Table A2) with the latter being defined as:

$$P(r; \lambda t) = \sum_{x=0}^r P(x; \lambda t) \quad (2.37c)$$

Applications of the Poisson distribution are widespread: the number of faults in a length of cable, number of suspended particles in a volume of gas, number of cars in a fixed length of roadway or number of cars passing a point in a fixed time interval (traffic flow), counts of α -particles in radio-active decay, number of arrivals in an interval of time (queuing theory), the number of noticeable surface defects found by quality inspectors on a new automobile, ...

Example 2.4.6: During a laboratory experiment, the average number of radioactive particles passing through a counter in 1 millisecond is 4. What is the probability that 6 particles enter the counter in any given millisecond?

Fig. 2.12 Poisson distribution for the number of storms per year where $\lambda t = 4$



Using the Poisson distribution function (Eq. 2.37a) with $x=6$ and $\lambda t=4$:

$$P(6; 4) = \frac{4^6 \cdot e^{-4}}{6!} = 0.1042$$

Example 2.4.7: The average number of planes landing at an airport each hour is 10 while the maximum number it can handle is 15. What is the probability that on a given hour some planes will have to be put on a holding pattern?

In this case, Eq. 2.37c is used. From Table A2, with $\lambda t = 10$

$$P(X > 15) = 1 - P(X \leq 15) = 1 - \sum_{x=0}^{15} P(x; 10) = 1 - 0.9513 = 0.0487$$

Example 2.4.8: Using Poisson PDF for assessing storm frequency

Historical records at Phoenix, AZ indicate that on an average there are 4 dust storms per year. Assuming a Poisson distribution, compute the probabilities of the following events using Eq. 2.37a:

(a) that there would not be any storms at all during a year:

$$p(X = 0) = \frac{(4)^0 \cdot e^{-4}}{0!} = 0.018$$

(b) the probability that there will be four storms during a year:

$$p(X = 4) = \frac{(4)^4 \cdot e^{-4}}{4!} = 0.195$$

Note that though the average is four, the probability of actually encountering four storms in a year is less than 20%. Figure 2.12 represents the PDF and CDF for different number of X values for this example.

distributions. It is a special case of the Binomial distribution with the same values of mean and variance but applicable when n is sufficiently large ($n > 30$). It is a two-parameter distribution given by:

$$N(x; \mu, \sigma) = \frac{1}{\sigma(2\pi)^{1/2}} \exp \left[-\frac{(x - \mu)^2}{\sigma^2} \right] \quad (2.38a)$$

where μ and σ are the mean and standard deviation respectively of the random variable X . Its name stems from an erroneous earlier perception that it was the natural pattern followed by distributions and that any deviation from it required investigation. Nevertheless, it has numerous applications in practice and is the most important of all distributions studied in statistics. Further, it is the parent distribution for several important continuous distributions as can be seen from Fig. 2.9. It is used to model events which occur by chance such as variation of dimensions of mass-produced items during manufacturing, experimental errors, variability in measurable biological characteristics such as people's height or weight,... Of great practical import is that normal distributions apply in situations where the random variable is the result of a sum of several other variable quantities acting independently on the system.

The shape of the normal distribution is unimodal and symmetrical about the mean, and has its maximum value at $x = \mu$ with points of inflexion at $x = \mu \pm \sigma$. Figure 2.13 illustrates its shape for two different cases of μ and σ . Further, the normal distribution given by Eq. 2.38a provides a convenient approximation for computing binomial probabilities for large number of values (which is tedious), provided $[n \cdot p \cdot (1-p)] > 10$.

In problems where the normal distribution is used, it is more convenient to standardize the random variable into a new random variable $z \equiv \frac{x - \mu}{\sigma}$ with mean zero and variance of unity. This results in the *standard normal curve* or *z-curve*:

$$N(z; 0, 1) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2). \quad (2.38b)$$

2.4.3 Distributions for Continuous Variables

(a) Gaussian Distribution. The Gaussian distribution or normal error function is the best known of all continuous

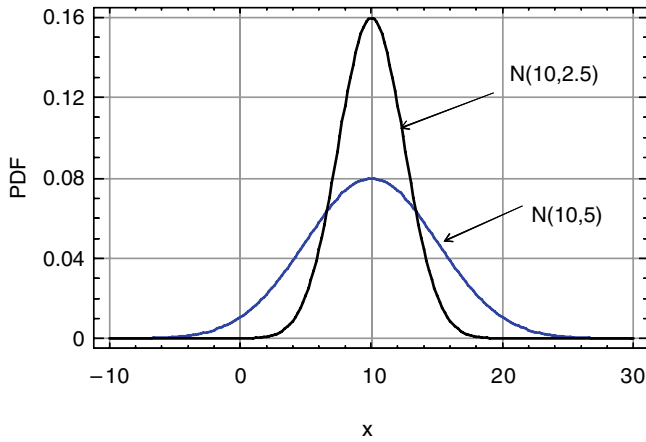


Fig. 2.13 Normal or Gaussian distributions with same mean of 10 but different standard deviations. The distribution flattens out as the standard deviation increases

In actual problems, the standard normal distribution is used to determine the probability of the variate having a value within a certain interval, say z between z_1 and z_2 . Then Eq. 2.38a can be modified into:

$$N(z_1 \leq z \leq z_2) = \int_{z_1}^{z_2} \frac{1}{\sqrt{2\pi}} \exp(-z^2/2) dz \quad (2.38c)$$

The shaded area in Table A3 permits evaluating the above integral, i.e., determining the associated probability assuming $z_1 = -\infty$. Note that for $z=0$, the probability given by the shaded area is equal to 0.5. Since not all texts adopt the same format in which to present these tables, the user is urged to use caution in interpreting the values shown in such tables.

Example 2.4.9: *Graphical interpretation of probability using the standard normal table*

Resistors made by a certain manufacturer have a nominal value of 100 ohms but their actual values are normally distributed with a mean of $\mu = 100.6$ ohms and standard deviation $\sigma = 3$ ohms. Find the percentage of resistors that will have values:

- (i) higher than the nominal rating. The standard normal variable $z(x=100) = (100 - 100.6)/3 = -0.2$. From Table A3, this corresponds to a probability of $(1 - 0.4207) = 0.5793$ or 57.93%.
- (ii) within 3 ohms of the nominal rating (i.e., between 97 and 103 ohms). The lower limit $z_1 = (97 - 100.6)/3 = -1.2$, and the tabulated probability from Table A3 is $p(z = -1.2) = 0.1151$ (as illustrated in Fig. 2.14a). The upper limit is: $z_2 = (103 - 100.6)/3 = 0.8$. However, care should be taken in properly reading the corresponding value from Table A3 which only gives probability values of $z < 0$. One first determines the probability about the negative value symmetric about 0, i.e., $p(z = -0.8) = 0.2119$ (shown in Fig. 2.14b). Since the total area under the curve is 1.0, $p(z = 0.8) = 1.0 - 0.2119 = 0.7881$. Finally, the required probability $p(-1.2 < z < 0.8) = (0.7881 - 0.1151) = 0.6730$ or 67.3%. ■

Inspection of Table A3 allows the following statements which are important in statistics:

The interval $\mu \pm \sigma$ contains approximately $[1 - 2(0.1587)] = 0.683$ or 68.3% of the observations,

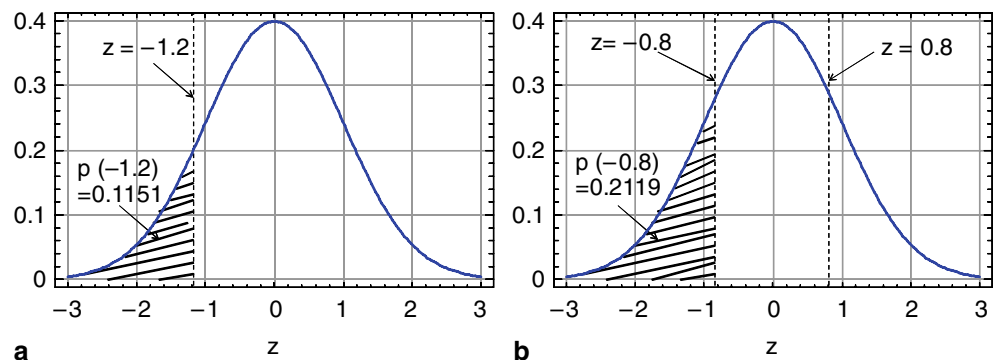
The interval $\mu \pm 2\sigma$ contains approximately 95.4% of the observations,

The interval $\mu \pm 3\sigma$ contains approximately 99.7% of the observations,

Another manner of using the standard normal table is for the “backward” problem. Instead of being specified the z value and having to deduce the probability, such a problem arises when the probability is specified and the z value is to be deduced.

Example 2.4.10: Reinforced and pre-stressed concrete structures are designed so that the compressive stresses are carried mostly by the concrete itself. For this and other reasons the main criterion by which the quality of concrete is assessed is its compressive strength. Specifications for concrete used in civil engineering jobs may require specimens of specified size and shape (usually cubes) to be cast and tested on site. One can assume the normal distribution to apply. If the mean and standard deviation of this distribution are μ and σ , the civil engineer wishes to determine the “statistical minimum strength” x specified as the strength below which

Fig. 2.14 Figures meant to illustrate that the shaded areas are the physical representations of the tabulated standardized probability values in Table A3. **a** Lower limit. **b** Upper limit



only say 5% of the cubes are expected to fail. One searches Table A3 and determines the value of z for which the probability is 0.05, i.e., $p(z = -1.645) = 0.05$. Hence, one infers that this would correspond to $x = \mu - 1.645\sigma$.

(b) Student t Distribution. One important application of the normal distribution is that it allows making statistical inferences about population means from random samples (see Sect. 4.2). In case the random samples are small ($n < 30$), then the t-student distribution, rather than the normal distribution, should be used. If one assumes that the sampled population is approximately normally distributed, then the random variable $t = \frac{x - \mu}{s/\sqrt{n}}$ has the Student t-distribution $t(\mu, s, v)$ where s is the sample standard deviation and v is the degrees of freedom ($= n - 1$). Thus, the number of degrees of freedom (d.f.) equals the number of data points minus the number of constraints or restrictions placed on the data. Table A4 (which is set up differently from the standard normal table) provides numerical values of the t-distribution for different degrees of freedom at different confidence levels. How to use these tables will be discussed in Sect. 4.2. Unlike the z curve, one has a family of t-distributions for different values of v . Qualitatively, the t-distributions are similar to the standard normal distribution in that they are symmetric about a zero mean, while they are but slightly wider than the corresponding normal distribution as indicated in Fig. 2.15. However, in terms of probability values represented by areas under the curves as in Example 2.4.9, the differences between the normal and the student-t distributions are large enough to warrant retaining this distinction.

(c) Lognormal Distribution. This distribution is appropriate for non-negative outcomes which are the product of a number of quantities. In such cases, the data are skewed and the symmetrical normal distribution is no longer appropriate. If a variate X is such that $\log(X)$ is normally distributed, then

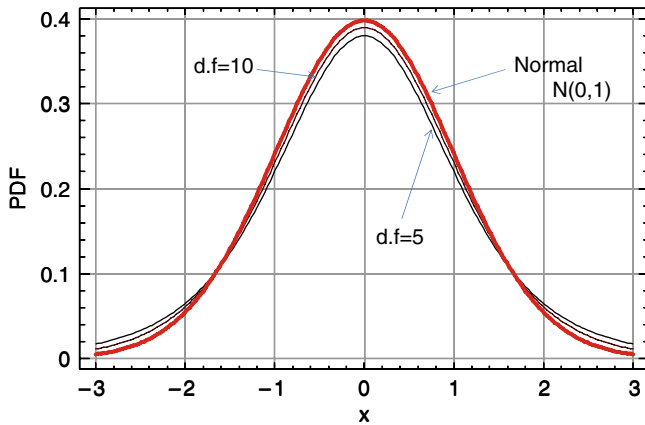


Fig. 2.15 Comparison of the normal (or Gaussian) z curve to two Student-t curves with different degrees of freedom (d.f.). As the d.f. increase, the PDF for the Student-t distribution flattens out and deviates increasingly from the normal distribution

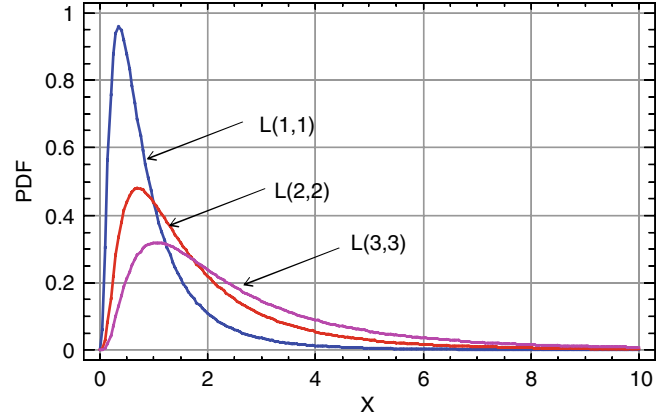


Fig. 2.16 Lognormal distributions for different mean and standard deviation values

the distribution of X is said to be lognormal. With X ranging from $-\infty$ to $+\infty$, $\log(X)$ would range from 0 to $+\infty$. Not only does the lognormal model accommodate skewness, but it also captures the non-negative nature of many variables which occur in practice. It is characterized by two parameters, the mean and variance (μ, σ), as follows:

$$L(x; \mu, \sigma) = \frac{1}{\sigma \cdot x (\sqrt{2\pi})} \exp \left[-\frac{(\ln x - \mu)^2}{2\sigma^2} \right] \quad \text{when } x \geq 0$$

$$= 0 \quad \text{elsewhere} \quad (2.39)$$

The lognormal curves are a family of skewed curves as illustrated in Fig. 2.16. Lognormal failure laws apply when the degradation in lifetime is proportional to the previous amount of degradation. Typical applications in civil engineering involve flood frequency, in mechanical engineering with crack growth and mechanical wear, and in environmental engineering with pollutants produced by chemical plants and threshold values for drug dosage.

Example 2.4.11: Using lognormal distributions for pollutant concentrations

Concentration of pollutants produced by chemical plants is known to resemble lognormal distributions and is used to evaluate issues regarding compliance of government regulations. The concentration of a certain pollutant, in parts per million (ppm), is assumed lognormal with parameters $\mu = 4.6$ and $\sigma = 1.5$. What is the probability that the concentration exceeds 10 ppm?

One can use Eq. 2.39, or simpler still, use the z tables (Table A3) by suitable transformations of the random variable.

$$L(X > 10) = N[\ln(10), 4.6, 1.5] = N \left[\frac{\ln(10) - 4.6}{1.5} \right]$$

$$= N(-1.531) = 0.0630 \quad \blacksquare$$

(d) Gamma Distribution. There are several processes where distributions other than the normal distribution are warranted. A distribution which is useful since it is versatile in the shapes it can generate is the gamma distribution (also called the *Erlang distribution*). It is a good candidate for modeling random phenomena which can only be positive and are unimodal. The gamma distribution is derived from the gamma function for positive values of α , which one may recall from mathematics, is defined by the integral:

$$\Gamma_x(\alpha) = \int_0^x x^{\alpha-1} \cdot e^{-x} dx \quad (2.40a)$$

Recall that for non-negative integers k :

$$\Gamma(k+1) = k! \quad (2.40b)$$

The continuous random variable X has a gamma distribution with positive parameters α and λ if its density function is given by:

$$G(x; \alpha, \lambda) = \lambda^\alpha e^{-\lambda x} \frac{x^{\alpha-1}}{(\alpha-1)!} \quad x > 0 \quad (2.40c)$$

$$= 0 \quad \text{elsewhere}$$

The mean and variance of the gamma distribution are:

$$\mu = \alpha/\lambda \quad \text{and} \quad \sigma^2 = \alpha/\lambda^2 \quad (2.40d)$$

Variation of the parameter α (called the shape factor) and λ (called the scale parameter) allows a wide variety of shapes to be generated (see Fig. 2.17). From Fig. 2.9, one notes that the Gamma distribution is the parent distribution of many other distributions discussed below. If $\alpha \rightarrow \infty$ and $\lambda = 1$, the gamma distribution approaches the normal (see Fig. 2.9). When $\alpha = 1$, one gets the exponential distribution. When $\alpha = \nu/2$ and $\lambda = 1/2$, one gets the chi-square distribution (discussed below).

(e) Exponential Distribution. A special case of the gamma distribution for $\alpha=1$ is the exponential distribution. It is the continuous distribution analogue to the geometric distribution

which applied to the discrete case. It is used to model the interval between two occurrences, e.g. the distance between consecutive faults in a cable, or the time between chance failures of a component (such as a fuse) or a system, or the time between consecutive emissions of α -particles, or the time between successive arrivals at a service facility. Its PDF is given by

$$E(x; \lambda) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (2.41a)$$

where λ is the mean value per unit time or distance. The mean and variance of the exponential distribution are:

$$\mu = 1/\lambda \quad \text{and} \quad \sigma^2 = 1/\lambda^2 \quad (2.41b)$$

The distribution is represented by a family of curves for different values of λ (see Fig. 2.18). Exponential failure laws apply to products whose current age does not have much effect on their remaining lifetimes. Hence, this distribution is said to be “memoryless”. Notice the relationship between the exponential and the Poisson distributions. While the latter represents the *number* of failures per unit time, the exponential represents the *time* between successive failures. Its CDF is given by:

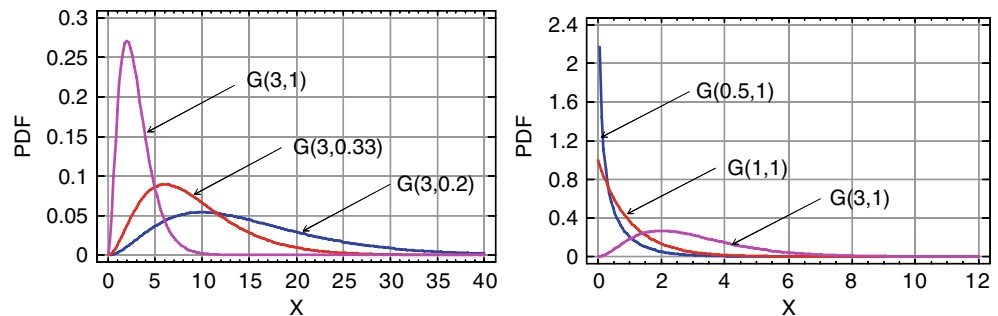
$$CDF[E(a, \lambda)] = \int_0^a \lambda \cdot e^{-\lambda x} dx = 1 - e^{-\lambda a} \quad (2.41c)$$

Example 2.4.12: Temporary disruptions to the power grid can occur due to random events such as lightning, transformer failures, forest fires,... The Poisson distribution has been known to be a good function to model such failures. If these occur, on average, say, once every 2.5 years, then $\lambda = 1/2.5 = 0.40$ per year.

- (a) What is the probability that there will be at least one disruption next year?

$$CDF[E(X \leq 1; \lambda)] = 1 - e^{-0.4(1)} = 1 - 0.6703 = 0.3297$$

Fig. 2.17 Gamma distributions for different combinations of the shape parameter α and the scale parameter $\beta = 1/\lambda$



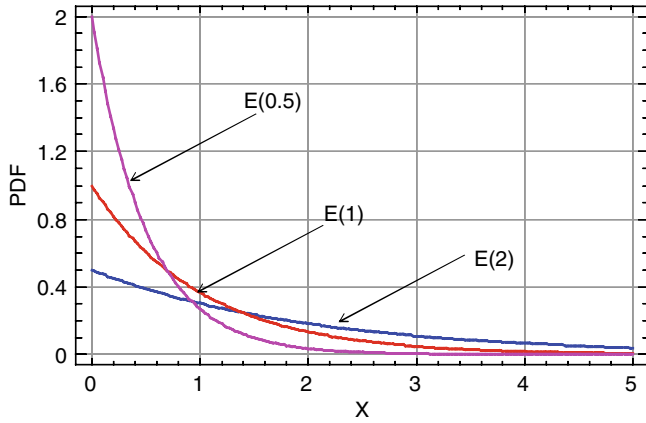


Fig. 2.18 Exponential distributions for three different values of the parameter λ

- (b) What is the probability that there will be no more than two disruptions next year?

This is the complement of at least two disruptions.

$$\begin{aligned} \text{Probability} &= 1 - CDF[E(X \leq 2; \lambda)] \\ &= 1 - [1 - e^{-0.4(2)}] = 0.4493 \end{aligned}$$

(f) Weibull Distribution. Another versatile and widely used distribution is the Weibull distribution which is used in applications involving reliability and life testing; for example, to model the time of failure or life of a component. The continuous random variable X has a Weibull distribution with parameters α and β (shape and scale factors respectively) if its density function is given by:

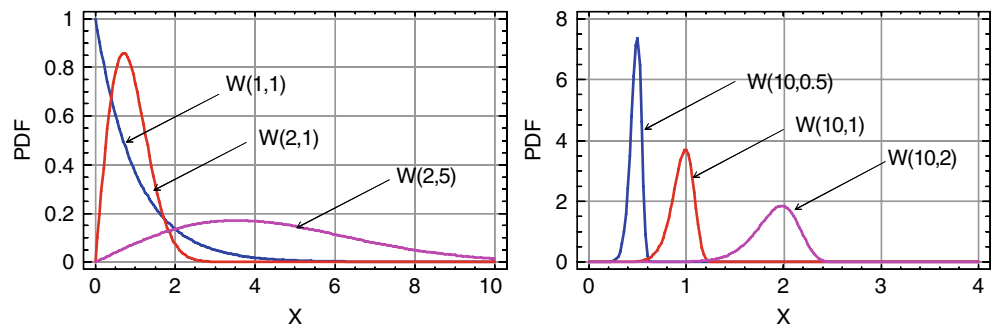
$$\begin{aligned} W(x; \alpha, \beta) &= \frac{\alpha}{\beta^\alpha} \cdot x^{\alpha-1} \cdot \exp[-(x/\beta)^\alpha] \quad \text{for } x > 0 \\ &= 0 \quad \text{elsewhere} \end{aligned} \quad (2.42a)$$

with mean

$$\mu = \beta \cdot \Gamma\left(1 + \frac{1}{\alpha}\right) \quad (2.42b)$$

Figure 2.19 shows the versatility of this distribution for different sets of α and β values. Also shown is the special case

Fig. 2.19 Weibull distributions for different values of the two parameters α and β (the shape and scale factors respectively)



of $W(1,1)$ which is the exponential distribution. For $\beta > 1$, the curves become close to bell-shaped and somewhat resemble the normal distribution. The Weibull distribution has been found to be very appropriate to model reliability of a system i.e., the failure time of the weakest component of a system (bearing, pipe joint failure,...).

Example 2.4.13: Modeling wind distributions using the Weibull distribution

The Weibull distribution is also widely used to model the hourly variability of wind velocity in numerous locations worldwide. The mean wind speed and its distribution on an annual basis, which are affected by local climate conditions, terrain and height of the tower, are important in order to determine annual power output from a wind turbine of a certain design whose efficiency changes with wind speed. It has been found that the shape factor α varies between 1 and 3 (when $\alpha=2$, the distribution is called the *Rayleigh distribution*). The probability distribution shown in Fig. 2.20 has a mean wind speed of 7 m/s. Determine:

- (a) the numerical value of the parameter β assuming the shape factor $\alpha=2$

One calculates the gamma function $\Gamma(1 + \frac{1}{2}) = 0.8862$ from which $\beta = \frac{\mu}{0.8862} = 7.9$

- (b) using the PDF given by Eq. 2.42, it is left to the reader to compute the probability of the wind speed being equal to 10 m/s (and verify the solution against the figure which indicates a value of 0.064).

(g) Chi-square Distribution. A third special case of the gamma distribution is when $\alpha = \nu/2$ and $\lambda = 1/2$ where ν is a positive integer, and is called the degrees of freedom. This distribution called the chi-square (χ^2) distribution plays an important role in inferential statistics where it is used as a test of significance for hypothesis testing and analysis of variance type of problems. Just like the t-statistic, there is a family of distributions for different values of ν (Fig. 2.21). Note that the distribution cannot assume negative values, and that it is positively skewed. Table A5 assembles critical values of the Chi-square distribution for different values of the degrees of freedom parameter ν and for different signifi-

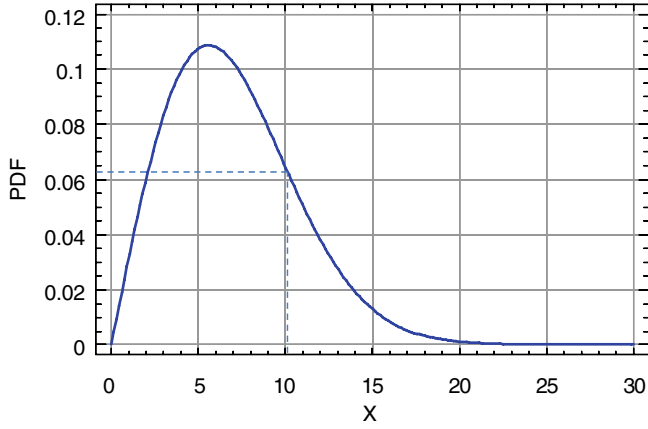


Fig. 2.20 PDF of the Weibull distribution $W(2, 7.9)$

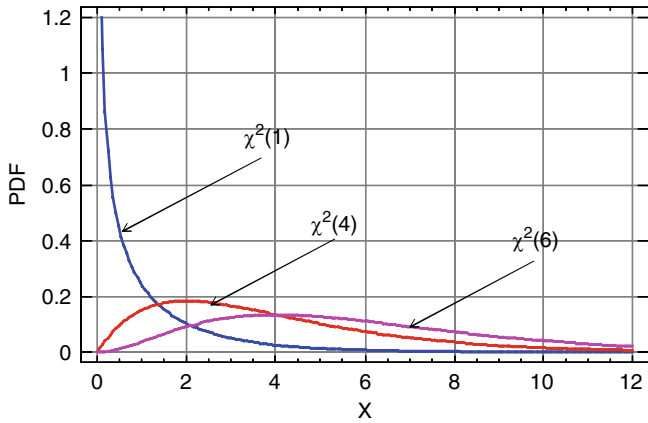


Fig. 2.21 Chi-square distributions for different values of the variable v denoting the degrees of freedom

cance levels. The usefulness of these tables will be discussed in Sect. 4.2.

The PDF of the chi-square distribution is:

$$\chi^2(x; v) = \frac{1}{2^{v/2} \Gamma(v/2)} \cdot x^{v/2-1} \cdot e^{-x/2} \quad x > 0 \quad (2.43a)$$

$$= 0 \quad \text{elsewhere}$$

while the mean and variance values are :

$$\mu = v \quad \text{and} \quad \sigma^2 = 2v \quad (2.43b)$$

(h) F-Distribution. While the t-distribution allows comparison between two sample means, the F distribution allows comparison between two or more sample variances. It is defined as the ratio of two independent chi-square random variables, each divided by its degrees of freedom. The F distribution is also represented by a family of plots (see Fig. 2.22) where each plot is specific to a set of numbers representing the degrees of freedom of the two random variables (v_1, v_2). Table A6 assembles critical values of the F-distribution.

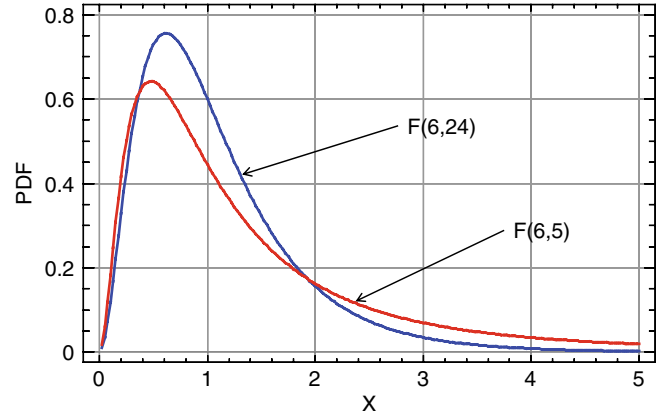


Fig. 2.22 Typical F distributions for two different combinations of the random variables (v_1 and v_2)

butions for different combinations of these two parameters, and its use will be discussed in Sect. 4.2.

(i) Uniform Distribution. The uniform probability distribution is the simplest of all PDFs and applies to both continuous and discrete data whose outcomes are all equally likely, i.e. have equal probabilities. Flipping a coin for heads/tails or rolling a dice for getting numbers between 1 and 6 are examples which come readily to mind. The probability density function for the discrete case where X can assume values x_1, x_2, \dots, x_k is given by:

$$U(x; k) = \frac{1}{k} \quad (2.44a)$$

with mean $\mu = \frac{\sum_{i=1}^k x_i}{k}$ and

$$\text{variance } \sigma^2 = \frac{\sum_{i=1}^k (x_i - \mu)^2}{k} \quad (2.44b)$$

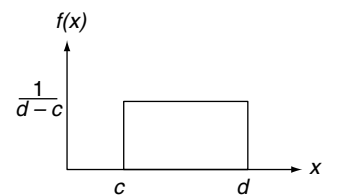
For random variables that are continuous over an interval (c, d) as shown in Fig. 2.23, the PDF is given by:

$$U(x) = \frac{1}{d-c} \quad \text{when } c < x < d$$

$$= 0 \quad \text{otherwise} \quad (2.44c)$$

The mean and variance of the uniform distribution (using notation shown in Fig. 2.23) are given by:

Fig. 2.23 The uniform distribution assumed continuous over the interval $[c, d]$



$$\mu = \frac{c+d}{2} \quad \text{and} \quad \sigma^2 = \frac{(d-c)^2}{12} \quad (2.44d)$$

The probability of random variable X being between say x_1 and x_2 is:

$$U(x_1 \leq X \leq x_2) = \frac{x_2 - x_1}{d - c} \quad (2.44e)$$

Example 2.4.14: A random variable X has a uniform distribution with $c=-5$ and $d=10$ (see Fig. 2.23). Determine:

- On an average, what proportion will have a negative value? (Answer: $1/3$)
- On an average, what proportion will fall between -2 and 2 ? (Answer: $4/15$)

(j) Beta Distribution. A very versatile distribution is the Beta distribution which is appropriate for discrete random variables between 0 and 1 such as representing proportions. It is a two parameter model which is given by:

$$\text{Beta}(x; p, q) = \frac{(p+q+1)!}{(p-1)!(q-1)!} x^{p-1} (1-x)^{q-1} \quad (2.45a)$$

Depending on the values of p and q , one can model a wide variety of curves from u shaped ones to skewed distributions (see Fig. 2.24). The distributions are symmetrical when p and q are equal, with the curves becoming peakier as the numerical values of the two parameters increase. Skewed distributions are obtained when the parameters are unequal.

$$\begin{aligned} \text{The mean of the Beta distribution } \mu &= \frac{p}{p+q} \quad \text{and} \\ \text{variance } \sigma^2 &= \frac{pq}{(p+q)^2(p+q+1)} \end{aligned} \quad (2.45b)$$

This distribution originates from the Binomial distribution, and one can detect the obvious similarity of a two-outcome affair with specified probabilities. The usefulness of this distribution will become apparent in Sect. 2.5.3 dealing with the Bayesian approach to probability problems.

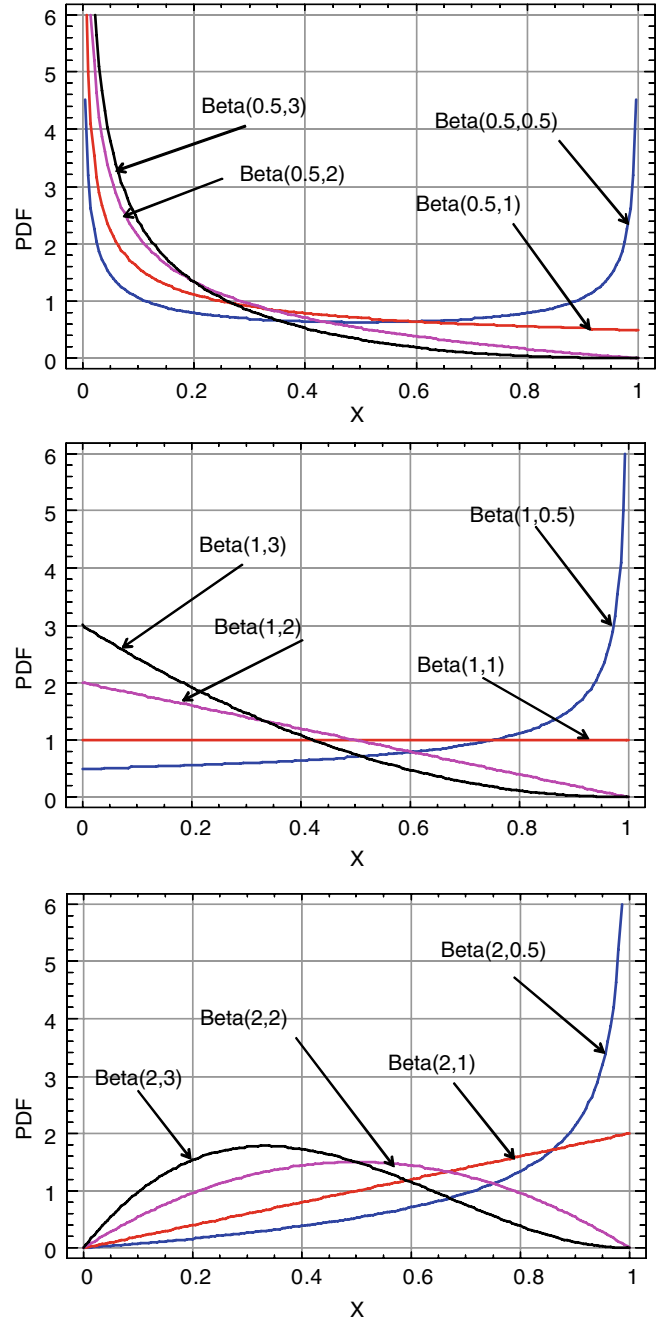


Fig. 2.24 Various shapes assumed by the Beta distribution depending on the values of the two model parameters

2.5 Bayesian Probability

2.5.1 Bayes' Theorem

It was stated in Sect. 2.1.4 that the Bayesian viewpoint can enhance the usefulness of the classical frequentist notion of probability². Its strength lies in the fact that it provides a framework to include prior information in a two-stage (or

multi-stage) experiment. If one substitutes the term $p(A)$ in Eq. 2.12 by that given by Eq. 2.11, one gets :

$$p(B/A) = \frac{p(A \cap B)}{p(A \cap B) + p(A \cap \bar{B})} \quad (2.46)$$

Also, one can re-arrange Eq. 2.12 into: $p(A \cap B) = p(A) \cdot p(B/A)$ or $= p(B) \cdot p(A/B)$. This allows expressing

² There are several texts which deal with Bayesian statistics; for example, Bolstad (2004).

Eq. 2.46 into the following expression referred to as the *law of total probability* or *Bayes' theorem*:

$$p(B/A) = \frac{p(A/B) \cdot p(B)}{p(A/B) \cdot p(B) + p(A/\bar{B}) \cdot p(\bar{B})} \quad (2.47)$$

Bayes theorem, superficially, appears to be simply a restatement of the conditional probability equation given by Eq. 2.12. The question is why is this reformulation so insightful or advantageous? First, the probability is now re-expressed in terms of its disjoint parts $\{B, \bar{B}\}$, and second the probabilities have been “flipped”, i.e., $p(B/A)$ is now expressed in terms of $p(A/B)$. Consider the two events A and B. If event A is observed while event B is not, this expression allows one to infer the “flip” probability, i.e. probability of occurrence of B from that of the observed event A. In Bayesian terminology, Eq. 2.47 can be written as:

$$\begin{aligned} &\text{Posterior probability of event B given event A} \\ &= \frac{(\text{Likelihood of A given B}) \cdot (\text{Prior probability of B})}{\text{Prior probability of A}} \end{aligned} \quad (2.48)$$

Thus, the probability $p(B)$ is called the *prior probability* (or unconditional probability) since it represents opinion *before* any data was collected, while $p(B/A)$ is said to be the *posterior probability* which is reflective of the opinion revised in light of new data. The likelihood is identical to the conditional probability of A given B i.e., $p(A/B)$.

Equation 2.47 applies to the case when only one of two events is possible. It can be extended to the case of more than two events which partition the space S. Consider the case where one has n events, $B_1 \dots B_n$ which are disjoint and make up the entire sample space. Figure 2.25 shows a sample space

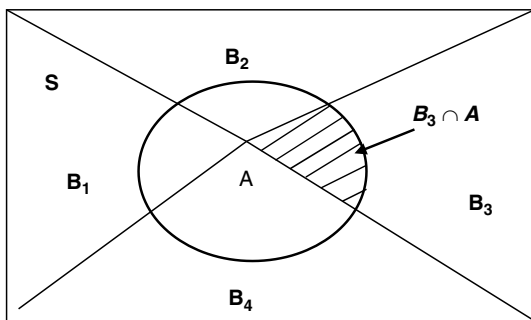


Fig. 2.25 Bayes theorem for multiple events depicted on a Venn diagram. In this case, the sample space is assumed to be partitioned into four discrete events $B_1 \dots B_4$. If an observable event A has already occurred, the conditional probability of B_3 : $p(B_3/A) = \frac{p(B_3 \cap A)}{p(A)}$. This is the ratio of the hatched area to the total area inside the ellipse

ce of 4 events. Then, the law of total probability states that the probability of an event A is the sum of its disjoint parts:

$$p(A) = \sum_{j=1}^n p(A \cap B_j) = \sum_{j=1}^n p(A/B_j) \cdot p(B_j) \quad (2.49)$$

$$\text{Then } \underbrace{p(B_i/A)}_{\text{posterior probability}} = \frac{p(A \cap B_i)}{p(A)} = \frac{p(A/B_i) \cdot p(B_i)}{\sum_{j=1}^n \underbrace{p(A/B_j)}_{\text{likelihood}} \cdot \underbrace{p(B_j)}_{\text{prior}}} \quad (2.50)$$

which is known as *Bayes' theorem for multiple events*. As before, the marginal or prior probabilities $p(B_i)$ for $i = 1, \dots, n$ are assumed to be known in advance, and the intention is to update or revise our “belief” on the basis of the observed evidence of event A having occurred. This is captured by the probability $p(B_i/A)$ for $i = 1, \dots, n$ called the posterior probability or the weight one can attach to each event B_i after event A is known to have occurred.

Example 2.5.1: Consider the two-stage experiment of Example 2.2.7. Assume that the experiment has been performed and that a red marble has been obtained. One can use the information known beforehand i.e., the prior probabilities R, W and G to determine from which box the marble came from. Note that the probability of the red marble having come from box A represented by $p(A/R)$ is now the conditional probability of the “flip” problem. This is called

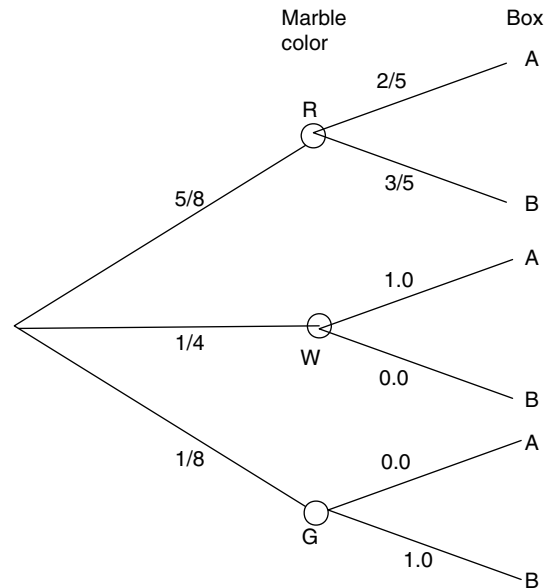
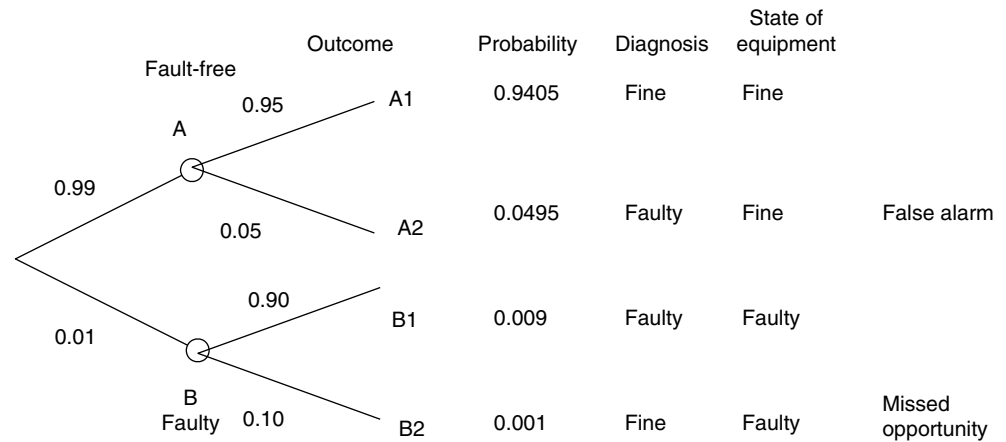


Fig. 2.26 The probabilities of the reverse tree diagram at each stage are indicated. If a red marble (R) is picked, the probabilities that it came from either Box A or Box B are 2/5 and 3/5 respectively

Fig. 2.27 The forward tree diagram showing the four events which may result when monitoring the performance of a piece of equipment



the *posterior probabilities* of event A with R having occurred, i.e., they are relevant after the experiment has been performed. Thus, from the law of total probability (Eq. 2.47):

$$p(B/R) = \frac{\frac{1}{2} \cdot \frac{3}{4}}{\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{3}{4}} = \frac{3}{5}$$

and

$$p(A/R) = \frac{\frac{1}{2} \cdot \frac{1}{2}}{\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{3}{4}} = \frac{2}{5}$$

The reverse probability tree for this experiment is shown in Fig. 2.26. The reader is urged to compare this with the forward tree diagram of Example 2.2.7. The probabilities of 1.0 for both W and G outcomes imply that there is no uncertainty at all in predicting where the marble came from. This is obvious since only Box A contains W, and only Box B contains G. However, for the red marble, one cannot be sure of its origin, and this is where a probability measure has to be determined. ■

Example 2.5.2: Forward and reverse probability trees for fault detection of equipment

A large piece of equipment is being continuously monitored by an add-on fault detection system developed by another vendor in order to detect faulty operation. The vendor of the fault detection system states that their product correctly identifies faulty operation when indeed it is faulty (this is referred to as *sensitivity*) 90% of the time. This implies that there is a probability $p=0.10$ of a false negative occurring (i.e., a missed opportunity of signaling a fault). Also, the vendor quoted that the correct status prediction rate or *specificity* of the detection system (i.e., system identified as healthy when indeed it is so) is 0.95, implying that the false positive or

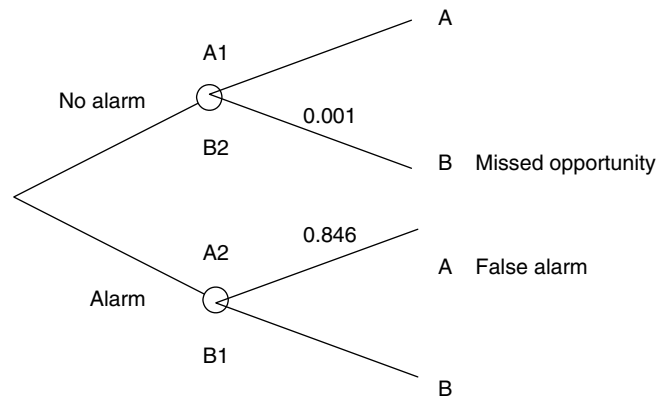


Fig. 2.28 Reverse tree diagram depicting two possibilities. If an alarm sounds, it could be either an erroneous one (outcome A from A2) or a valid one (B from B1). Further, if no alarm sounds, there is still the possibility of missed opportunity (outcome B from B2). The probability that it is a false alarm is 0.846 which is too high to be acceptable in practice. How to decrease this is discussed in the text

false alarm rate is 0.05. Finally, historic data seem to indicate that the large piece of equipment tends to develop faults only 1% of the time.

Figure 2.27 shows how this problem can be systematically represented by a forward tree diagram. State A is the fault-free state and state B is represented by the faulty state. Further, each of these states can have two outcomes as shown. While outcomes A1 and B1 represent correctly identified fault-free and faulty operation, the other two outcomes are errors arising from an imperfect fault detection system. Outcome A2 is the “*false negative*” (or false alarm or error type II which will be discussed at length in Sect. 4.2 of Chap. 4), while outcome B2 is the *false positive* rate (or missed opportunity or error type I). The figure clearly illustrates that the probabilities of A and B occurring along with the conditional probabilities $p(A1/A)=0.95$ and $p(B1/B)=0.90$, result in the probabilities of each the four states as shown in the figure.

The reverse tree situation, shown in Fig. 2.28, corresponds to the following situation. A fault has been signaled. What is the probability that this is a false alarm? Using Eq. 2.47:

$$\begin{aligned}
 p(A/A2) &= \frac{(0.99).(0.05)}{(0.99).(0.05) + (0.01).(0.90)} \\
 &= \frac{0.0495}{0.0495 + 0.009} \\
 &= 0.846
 \end{aligned}$$

This is very high for practical situations and could well result in the operator disabling the fault detection system altogether. One way of reducing this false alarm rate, and thereby enhance robustness, is to increase the sensitivity of the detection device from its current 90% to something higher by altering the detection threshold. This would result in a higher missed opportunity rate, which one has to accept for the price of reduced false alarms. For example, the current missed opportunity rate is:

$$\begin{aligned}
 p(B/B1) &= \frac{(0.01) \cdot (0.10)}{(0.01) \cdot (0.10) + (0.99) \cdot (0.95)} \\
 &= \frac{0.001}{0.001 + 0.9405} = 0.001
 \end{aligned}$$

This is probably lower than what is needed, and so the above suggested remedy is one which can be considered. Note that as the piece of machinery degrades, the percent of time when faults are likely to develop will increase from the current 1% to something higher. This will have the effect of lowering the false alarm rate (left to the reader to convince himself why). ■

Bayesian statistics provide the formal manner by which prior opinion expressed as probabilities can be revised in the light of new information (from additional data collected) to yield posterior probabilities. When combined with the relative consequences or costs of being right or wrong, it allows one to address decision-making problems as pointed out in the example above (and discussed at more length in Sect. 12.2.9). It has had some success in engineering (as well as in social sciences) where subjective judgment, often referred to as intuition or experience gained in the field, is relied upon heavily.

The Bayes' theorem is a consequence of the probability laws and is accepted by all statisticians. It is the interpretation of probability which is controversial. Both approaches differ in how probability is defined:

- classical viewpoint: long run relative frequency of an event
- Bayesian viewpoint: degree of belief held by a person about some hypothesis, event or uncertain quantity (Phillips 1973).

Advocates of the classical approach argue that human judgment is fallible while dealing with complex situations, and this was the reason why formal statistical procedures were developed in the first place. Introducing the vagueness of human judgment as done in Bayesian statistics would dilute the “purity” of the entire mathematical approach. Ad-

vocates of the Bayesian approach, on the other hand, argue that the “personalist” definition of probability should not be interpreted as the “subjective” view. Granted that the prior probability is subjective and varies from one individual to the other, but with additional data collection all these views get progressively closer. Thus, with enough data, the initial divergent opinions would become indistinguishable. Hence, they argue, the Bayesian method brings consistency to informal thinking when complemented with collected data, and should, thus, be viewed as a mathematically valid approach.

2.5.2 Application to Discrete Probability Variables

The following example illustrates how the Bayesian approach can be applied to discrete data.

Example 2.5.3:³ *Using the Bayesian approach to enhance value of concrete piles testing*

Concrete piles driven in the ground are used to provide bearing strength to the foundation of a structure (building, bridge,...). Hundreds of such piles could be used in large construction projects. These piles could develop defects such as cracks or voids in the concrete which would lower compressive strength. Tests are performed by engineers on piles selected at random during the concrete pour process in order to assess overall foundation strength. Let the random discrete variable be the proportion of defective piles out of the entire lot which is taken to assume five discrete values as shown in the first column of Table 2.7. Consider the case where the prior experience of an engineer as to the proportion of defective piles from similar sites is given in the second column of the table below.

Before any testing is done, the *expected value* of the probability of finding one pile to be defective is: $p = (0.20)(0.30) + (0.4)(0.40) + (0.6)(0.15) + (0.8)(0.10) + (1.0)$

Table 2.7 Illustration of how a prior PDF is revised with new data

Proportion of defectives (x)	Probability of being defective			
	Prior PDF of defectives	After one pile tested is found defective	After two piles tested are found defective	Limiting case of infinite defectives
0.2	0.30	0.136	0.049	...
0.4	0.40	0.364	0.262	0.0
0.6	0.15	0.204	0.221	0.0
0.8	0.10	0.182	0.262	0.0
1.0	0.05	0.114	0.205	...
Expected probability of defective pile	0.44	0.55	0.66	1.0

³ From Ang and Tang (2007) by permission of John Wiley and Sons.

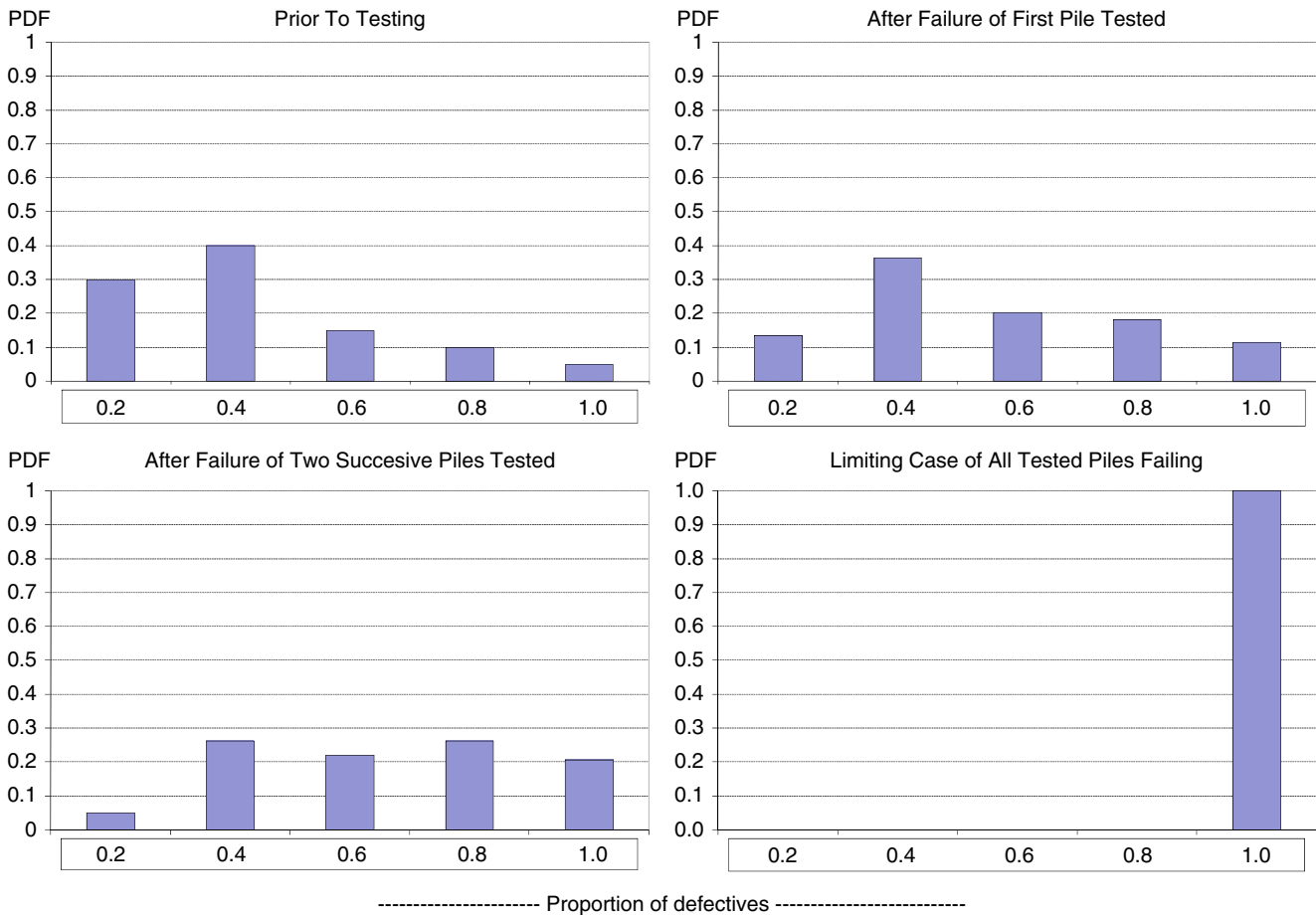


Fig. 2.29 Illustration of how the prior discrete PDF is affected by data collection following Bayes' theorem

(0.05)=0.44 (as shown in the last row under the second column). This is the prior probability.

Suppose the first pile tested is found to be defective. How should the engineer revise his prior probability of the proportion of piles likely to be defective? This is given by Bayes' theorem (Eq. 2.50). For proportion $x=0.2$, the posterior probability is:

$$\begin{aligned}
 p(x = 0.2) &= \frac{(0.2)(0.3)}{(0.2)(0.3) + (0.4)(0.4) + (0.6)(0.15) + (0.8)(0.10) + (1.0)(0.05)} \\
 &= \frac{0.06}{0.44} \\
 &= 0.136
 \end{aligned}$$

This is the value which appears in the first row under the third column. Similarly the posterior probabilities for different va-

lues of x can be determined as well as the expected value $E(x=1)$ which is 0.55. Hence, a single inspection has led to the engineer revising his prior opinion upward from 0.44 to 0.55. Had he drawn a conclusion on just this single test without using his prior judgment, he would have concluded that all the piles were defective; clearly, an over-statement. The engineer would probably get a second pile tested, and if it also turns

out to be defective, the associated probabilities are shown in the fourth column of Table 2.7. For example, for $x=0.2$:

$$p(x = 0.2) = \frac{(0.2)(0.136)}{(0.2)(0.136) + (0.4)(0.364) + (0.6)(0.204) + (0.8)(0.182) + (1.0)(0.114)} = 0.049$$

Table 2.8 Prior pdf of defective proportion

X	0.1	0.2
f(x)	0.6	0.4

The expected value in this case increases to 0.66. In the limit, if each successive pile tested turns out to be defective, one gets back the classical distribution, listed in the last column of the table. The progression of the PDF from the prior to the infinite case is illustrated in Fig. 2.29. Note that as more piles tested turn out to be defective, the evidence from the data gradually overwhelms the prior judgment of the engineer. However, it is only when collecting data is so expensive or time consuming that decisions have to be made from limited data that the power of the Bayesian approach becomes evident. Of course, if one engineer's prior judgment is worse than that of another engineer, then his conclusion from the same data would be poorer than the other engineer. It is this type of subjective disparity which antagonists of the Bayesian approach are uncomfortable with. On the other hand, proponents of the Bayesian approach would argue that experience (even if intangible) gained in the field is a critical asset in engineering applications and that discarding this type of knowledge entirely is naïve, and a severe handicap. ■

There are instances when no previous knowledge or information is available about the behavior of the random variable; this is sometime referred to as *prior of pure ignorance*. It can be shown that this assumption of the prior leads to results identical to those of the traditional probability approach (see Examples 2.5.5 and 2.5.6).

Example 2.5.4:⁴ Consider a machine whose prior pdf of the proportion x of defectives is given by Table 2.8.

If a random sample of size 2 is selected, and one defective is found, the Bayes estimate of the proportion of defectives produced by the machine is determined as follows.

Let y be the number of defectives in the sample. The probability that the random sample of size 2 yields one defective is given by the Binomial distribution since this is a two-outcome situation:

$$f(y/x) = B(y; n, x) = \binom{2}{y} x^y (1-x)^{2-y}; \quad y = 0, 1, 2$$

If $x = 0.1$, then

$$f(1/0.1) = B(1; 2, 0.1) = \binom{2}{1} (0.1)^1 (0.9)^{2-1} = 0.18$$

Similarly, for $x = 0.2$, $f(1/0.2) = 0.32$.

Thus, the total probability of finding one defective in a sample size of 2 is:

$$\begin{aligned} f(y = 1) &= (0.18)(0.6) + (0.32)(0.40) \\ &= (0.108) + (0.128) \\ &= 0.236 \end{aligned}$$

The posterior probability $f(x/y = 1)$ is then given:

- for $x = 0.1$: $0.108/0.236 = 0.458$
- for $x = 0.2$: $0.128/0.236 = 0.542$

Finally, the Bayes' estimate of the proportion of defectives x is:

$$x = (0.1)(0.458) + (0.2)(0.542) = 0.1542$$

which is quite different from the value of 0.5 given by the classical method. ■

2.5.3 Application to Continuous Probability Variables

The Bayes' theorem can also be extended to the case of continuous random variables (Ang and Tang 2007). Let x be the random variable with a prior PDF denoted by $p(x)$. Though any appropriate distribution can be chosen, the Beta distribution (given by Eq. 2.45) is particularly convenient⁵, and is widely used to characterize prior PDF. Another commonly used prior is the uniform distribution called a *diffuse prior*.

For consistency with convention, a slightly different nomenclature than that of Eq. 2.50 is adopted. Assuming the Beta distribution, Eq. 2.45a can be rewritten to yield the prior:

$$p(x) \propto x^a (1-x)^b \quad (2.51)$$

Recall that higher the values of the exponents a and b , the peakier the distribution indicative of the prior distribution being relatively well defined.

Let $L(x)$ represent the conditional probability or likelihood function of observing y "successes" out of n observations. Then, the posterior probability is given by:

$$f(x/y) \propto L(x) \cdot p(x) \quad (2.52)$$

In the context of Fig. 2.25, the *likelihood* of the unobservable events B_1, \dots, B_n is the conditional probability that A has occurred given B_i for $i = 1, \dots, n$, or by $p(A/B_i)$. The likelihood function can be gleaned from probability considerations in many cases. Consider Example 2.5.3 involving testing the foundation piles of buildings. The Binomial distribution gives the probability of x failures in n independent Bernoulli

⁴ From Walpole et al. (2007) by permission of Pearson Education.

⁵ Because of the corresponding mathematical simplicity which it provides as well as the ability to capture a wide variety of PDF shapes

trials, provided the trials are independent and the probability of failure in any one trial is p . This applies to the case when one holds p constant and studies the behavior of the pdf of defectives x . If instead, one holds x constant and lets $p(x)$ vary over its possible values, one gets the *likelihood function*. Suppose n piles are tested and y piles are found to be defective or sub-par. In this case, the likelihood function is written as follows for the Binomial PDF:

$$L(x) = \binom{n}{y} x^y (1-x)^{n-y} \quad 0 \leq x \leq 1 \quad (2.53)$$

Notice that the Beta distribution is the same form as the likelihood function. Consequently, the posterior distribution given by Eq. 2.53 assumes the form:

$$f(x/y) = k \cdot x^{a+y} (1-x)^{b+n-y} \quad (2.54)$$

where k is independent of x and is a normalization constant. Note that $(1/k)$ is the denominator term in Eq. 2.54 and is essentially a constant introduced to satisfy the probability law that the area under the PDF is unity. What is interesting is that the information contained in the prior has the net result of “artificially” augmenting the number of observations taken. While the classical approach would use the likelihood function with exponents y and $(n-y)$ (see Eq. 2.51), these are inflated to $(a+y)$ and $(b+n-y)$ in Eq. 2.54 for the posterior distribution. This is akin to having taken more observations, and supports the previous statement that the Bayesian approach is particularly advantageous when the number of observations is low.

Three examples illustrating the use of Eq. 2.54 are given below.

Example 2.5.5: Repeat Example 2.5.4 assuming that no information is known about the prior. In this case, assume a uniform distribution.

The prior pdf can be found from the Binomial distribution:

$$\begin{aligned} f(y/x) &= B(1; 2, x) = \binom{2}{1} x^1 (1-x)^{2-1} \\ &= 2x(1-x) \end{aligned}$$

The total probability of one defective is now given by:

$$f(y=1) = \int_0^1 2x(1-x)dx = \frac{1}{3}$$

The posterior probability is then found by dividing the above two expressions (Eq. 2.54):

$$f(x/y=1) = \frac{2x(1-x)}{1/3} = 6x(1-x)$$

Finally, the Bayes’ estimate of the proportion of defectives x is:

$$x = 6 \int_0^1 x^2(1-x)dx = 0.5$$

which can be compared to the value of 0.5 given by the classical method. ■

Example 2.5.6: Let us consider the same situation as that treated in Example 2.5.3. However, the proportion of defectives x is now a continuous random variable for which no prior distribution can be assigned. This implies that the engineer has no prior information, and in such cases, a uniform distribution is assumed:

$$p(x) = 1.0 \quad \text{for} \quad 0 \leq x \leq 1$$

The likelihood function for the case of the single tested pile turning out to be defective is x , i.e. $L(x)=x$. The posterior distribution is then:

$$f(x/y) = k \cdot x(1.0)$$

The normalizing constant

$$k = \left[\int_0^1 x dx \right]^{-1} = 2$$

Hence, the posterior probability distribution is:

$$f(x/y) = 2x \quad \text{for} \quad 0 \leq x \leq 1$$

The Bayesian estimate of the proportion of defectives is:

$$p = E(x/y) = \int_0^1 x \cdot 2x dx = 0.667 \quad \blacksquare$$

Example 2.5.7:⁶ *Enhancing historical records of wind velocity using the Bayesian approach*

Buildings are designed to withstand a maximum wind speed which depends on the location. The probability x that the wind speed will not exceed 120 km/h more than once in 5 years is to be determined. Past records of wind speeds of a nearby location indicated that the following beta distribution would be an acceptable prior for the probability distribution (Eq. 2.45):

$$p(x) = 20x^3(1-x) \quad \text{for} \quad 0 \leq x \leq 1$$

In this case, the likelihood that the annual maximum wind speed will exceed 120 km/h in 1 out of 5 years is given by:

$$L(x) = \binom{5}{4} x^4(1-x) = 5x^4(1-x)$$

⁶ From Ang and Tang (2007) by permission of John Wiley and Sons.

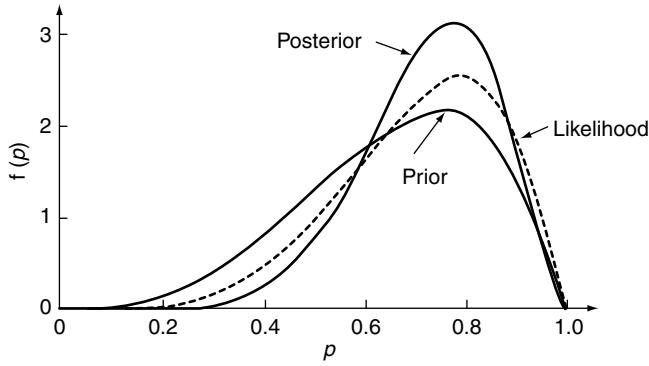


Fig. 2.30 Probability distributions of the prior, likelihood function and the posterior. (From Ang and Tang 2007 by permission of John Wiley and Sons)

Hence, the posterior probability is deduced following Eq. 2.54:

$$\begin{aligned} f(x/y) &= k \cdot [5x^4(1-x)] \cdot [20x^3(1-x)] \\ &= 100k \cdot x^7 \cdot (1-x)^2 \end{aligned}$$

where the constant k can be found from the normalization criterion:

$$k = \left[\int_0^1 100x^7(1-x)^2 dx \right]^{-1} = 3.6$$

Finally, the posterior PDF is given by

$$f(x/y) = 360x^7(1-x)^2 \quad \text{for } 0 \leq x \leq 1$$

Plots of the prior, likelihood and the posterior functions are shown in Fig. 2.30. Notice how the posterior distribution has become more peaked reflective of the fact that the single test data has provided the analyst with more information than that contained in either the prior or the likelihood function. ■

2.6 Probability Concepts and Statistics

The distinction between probability and statistics is often not clear cut, and sometimes, the terminology adds to the confusion⁷. In its simplest sense, probability generally allows one to predict the behavior of the system “before” the event under the stipulated assumptions, while statistics refers to a body of knowledge whose application allows one to make sense out of the data collected. Thus, probability concepts provide the theoretical underpinnings of those aspects of statistical analysis which involve random behavior or noise in the actual data being analyzed. Recall that in Sect. 1.5, a

⁷ For example, “statistical mechanics” in physics has nothing to do with statistics at all but is a type of problem studied under probability.

distinction had been made between four types of uncertainty or unexpected variability in the data. The first was due to the stochastic or inherently random nature of the process itself which no amount of experiment, even if done perfectly, can overcome. The study of probability theory is mainly mathematical, and applies to this type, i.e., to situations/processes/systems whose random nature is known to be of a certain type or can be modeled so that its behavior (i.e., certain events being produced by the system) can be predicted in the form of probability distributions. Thus, probability deals with the idealized behavior of a system under a *known type of randomness*. Unfortunately, most natural or engineered systems do not fit neatly into any one of these groups, and so when performance data is available of a system, the objective may be:

- (i) to try to understand the overall nature of the system from its measured performance, i.e., to explain what caused the system to behave in the manner it did, and
- (ii) to try to make inferences about the general behavior of the system from a limited amount of data.

Consequently, some authors have suggested that probability be viewed as a “deductive” science where the conclusion is drawn without any uncertainty, while statistics is an “inductive” science where only an imperfect conclusion can be reached, with the added problem that this conclusion hinges on the types of assumptions one makes about the random nature of the underlying drivers! Here is a simple example to illustrate the difference. Consider the flipping of a coin supposed to be fair. The probability of getting “heads” is $\frac{1}{2}$. If, however, “heads” come up 8 times out of the last 10 trials, what is the probability the coin is not fair? Statistics allows an answer to this type of enquiry, while probability is the approach for the “forward” type of questioning.

The previous sections in this chapter presented basic notions of classical probability and how the Bayesian viewpoint is appropriate for certain types of problems. Both these viewpoints are still associated with the concept of probability as the relative frequency of an occurrence. At a broader context, one should distinguish between three kinds of probabilities:

- (i) *Objective or absolute probability* which is the classical one where it is interpreted as the “long run frequency”. This is the same for everyone (provided the calculation is done correctly!). It is an informed guess of an event which in its simplest form is a constant; for example, historical records yield the probability of flood occurring this year or of the infant mortality rate in the U.S.

Table 2.9 assembles probability estimates for the occurrence of natural disasters with 10 and 1000 fatalities per event (indicative of the severity level) during different time spans (1, 10 and 20 years). Note that floods and tornados have relatively small return times for small

Table 2.9 Estimates of absolute probabilities for different natural disasters in the United States. (Adapted from Barton and Nishenko 2008)

Exposure Times	10 fatalities per event				1000 fatalities per event			
	1 year	10 years	20 years	Return time (yrs)	1 year	10 years	20 years	Return time (yrs)
Earthquakes	0.11	0.67	0.89	9	0.01	0.14	0.26	67
Hurricanes	0.39	0.99	>0.99	2	0.06	0.46	0.71	16
Floods	0.86	>0.99	>0.99	0.5	0.004	0.04	0.08	250
Tornadoes	0.96	>0.99	>0.99	0.3	0.006	0.06	0.11	167

Table 2.10 Leading causes of death in the United States, 1992. (Adapted from Kolluru et al. 1996)

Cause	Annual deaths (× 1000)	Percent %
Cardiovascular or heart disease	720	33
Cancer (malignant neoplasms)	521	24
Cerebrovascular diseases (strokes)	144	7
Pulmonary disease (bronchitis, asthma..)	91	4
Pneumonia and influenza	76	3
Diabetes mellitus	50	2
Nonmotor vehicle accidents	48	2
Motor vehicle accidents	42	2
HIV/AIDS	34	1.6
Suicides	30	1.4
Homicides	27	1.2
All other causes	394	18
<i>Total annual deaths (rounded)</i>	<i>21,77</i>	<i>100</i>

events while earthquakes and hurricanes have relatively short times for large events. Such probability considerations can be determined at a finer geographical scale, and these play a key role in the development of codes and standards for designing large infrastructures (such as dams) as well as small systems (such as residential buildings).

- (ii) *Relative probability* where the chance of occurrence of one event is stated in terms of another. This is a way of comparing the effect of different types of adverse events happening on a system or on a population when the absolute probabilities are difficult to quantify. For example, the relative risk for lung cancer is (approximately) 10 if a person has smoked before, compared to a nonsmoker. This means that he is 10 times more likely to get lung cancer than a nonsmoker. Table 2.10 shows leading causes of death in the United States in the year 1992. Here the observed values of the individual number of deaths due to various causes are used to determine a relative risk expressed as % in the last column. Thus, heart disease which accounts for 33% of the total deaths is more than 16 times more likely than motor vehicle deaths. However, as a note of caution, these are values aggregated across the whole population, and need to be

interpreted accordingly. State and government analysts separate such relative risks by age groups, gender and race for public policy-making purposes.

- (iii) *Subjective probability* which differs from one person to another is an informed or best guess about an event which can change as our knowledge of the event increases. Subjective probabilities are those where the objective view of probability has been modified to treat two types of events: (i) when the occurrence is unique and is unlikely to repeat itself, or (ii) when an event has occurred but one is unsure of the final outcome. In such cases, one still has to assign some measure of likelihood of the event occurring, and use this in our analysis. Thus, a subjective interpretation is adopted with the probability representing a degree of belief of the outcome selected as having actually occurred. There are no “correct answers”, simply a measure reflective of our subjective judgment. A good example of such subjective probability is one involving forecasting the probability of whether the impacts on gross world product of a 3°C global climate change by 2090 would be large or not. A survey was conducted involving twenty leading researchers working on global warming issues but with different technical backgrounds, such as scientists, engineers, economists, ecologists, and politicians who were asked to assign a probability estimate (along with 10% and 90% confidence intervals). Though this was not a scientific study as such since the whole area of expert opinion elicitation is still not fully mature, there was nevertheless a protocol in how the questioning was performed, which led to the results shown in Fig. 2.31. The median, 10% and 90% confidence intervals predicted by different respondents show great scatter, with the ecologists estimating impacts to be 20–30 times higher (the two right most bars in the figure), while the economists on average predicted the chance of large consequences to have only a 0.4% loss in gross world product. An engineer or a scientist may be uncomfortable with such subjective probabilities, but there are certain types of problems where this is the best one can do with current knowledge. Thus, formal analysis methods have to accommodate such information, and it is here that Bayesian techniques can play a key role.

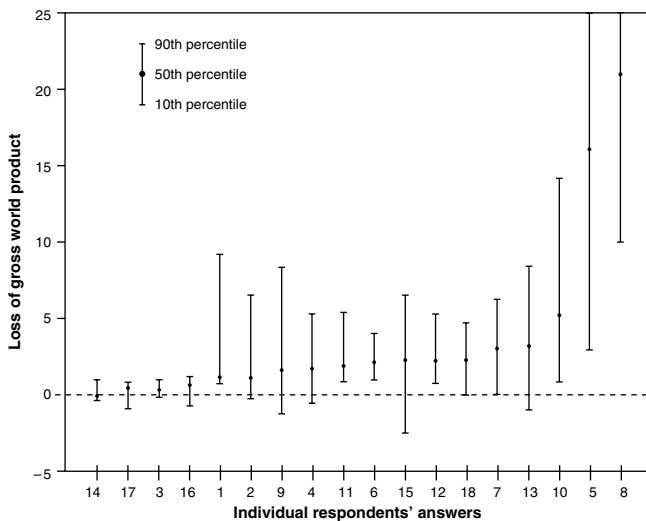


Fig. 2.31 Example illustrating large differences in subjective probability. A group of prominent economists, ecologists and natural scientists were polled so as to get their estimates of the loss of gross world product due to doubling of atmospheric carbon dioxide (which is likely to occur by the end of the twenty-first century when mean global temperatures increase by 3°C). The two ecologists predicted the highest adverse impact while the lowest four individuals were economists. (From Nordhaus 1994)

Problems

Pr. 2.1 An experiment consists of tossing two dice.

- List all events in the sample space
- What is the probability that both outcomes will have the same number showing up both times?
- What is the probability that the sum of both numbers equals 10?

Pr. 2.2 Expand Eq. 2.9 valid for two outcomes to three outcomes: $p(A \cup B \cup C) = \dots$

Pr. 2.3 A solar company has an inspection system for batches of photovoltaic (PV) modules purchased from different vendors. A batch typically contains 20 modules, while the inspection system involves taking a random sample of 5 modules and testing all of them. Suppose there are 2 faulty modules in the batch of 20.

- What is the probability that for a given sample, there will be one faulty module?
- What is the probability that both faulty modules will be discovered by inspection?

Pr. 2.4 A county office determined that of the 1000 homes in their area, 600 were older than 20 years (event A), that 500 were constructed of wood (event B), and that 400 had central air conditioning (AC) (event C). Further, it is found that events A and B occur in 300 homes, that all three events occur in 150 homes, and that no event occurs in 225 homes.

If a single house is picked, determine the following probabilities:

- that it is older than 20 years and has central AC?
- that it is older than 20 years and does not have central AC?
- that it is older than 20 years and is not made of wood?
- that it has central air and is made of wood?

Pr. 2.5 A university researcher has submitted three research proposals to three different agencies. Let E_1 , E_2 and E_3 be the events that the first, second and third bids are successful with probabilities: $p(E_1)=0.15$, $p(E_2)=0.20$, $p(E_3)=0.10$. Assuming independence, find the following probabilities:

- that all three bids are successful
- that at least two bids are successful
- that at least one bid is successful

Pr. 2.6 Consider two electronic components A and B with probability rates of failure of $p(A)=0.1$ and $p(B)=0.25$. What is the failure probability of a system which involves connecting the two components in (a) series and (b) parallel.

Pr. 2.7⁸ A particular automatic sprinkler system for a high-rise apartment has two different types of activation devices for each sprinkler head. Reliability of such devices is a measure of the probability of success, i.e., that the device will activate when called upon to do so. Type A and Type B devices have reliability values of 0.90 and 0.85 respectively. In case, a fire does start, calculate:

- the probability that the sprinkler head will be activated (i.e., at least one of the devices works),
- the probability that the sprinkler will not be activated at all, and
- the probability that both activation devices will work properly.

Pr. 2.8 Consider the two system schematics shown in Fig. 2.32. At least one pump must operate when one chiller is operational, and both pumps must operate when both chillers are on. Assume that both chillers have identical reliabilities of 0.90 and that both pumps have identical reliabilities of 0.95.

- Without any computation, make an educated guess as to which system would be more reliable overall when (i) one chiller operates, and (ii) when both chillers operate.
- Compute the overall system reliability for each of the configurations separately under cases (i) and (ii) defined above.

⁸ From McClave and Benson (1988) by permission of Pearson Education.

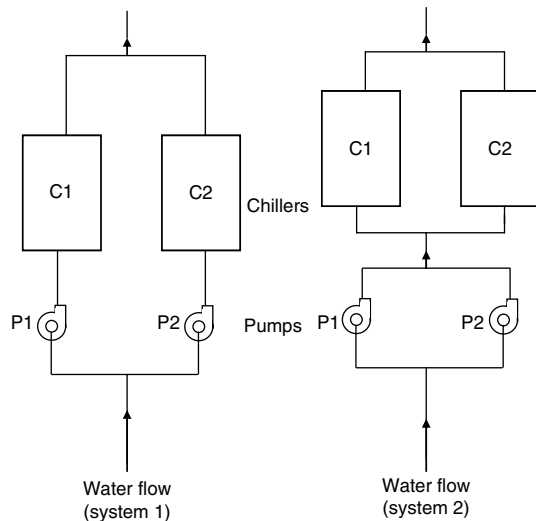


Fig. 2.32 Two possible system configurations

Pr. 2.9 Consider the following CDF:

$$F(x) = \begin{cases} 1 - \exp(-2x) & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

- Construct and plot the cumulative distribution function
- What is the probability of $x < 2$
- What is the probability of $3 < x < 5$

Pr. 2.10 The joint density for the random variables (X,Y) is given by:

$$f(x,y) = \begin{cases} 10xy^2 & 0 < x < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

- Verify that Eq. 2.19 applies
- Find the marginal distributions of X and Y
- Compute the probability of $0 < x < 1/2$, $1/4 < y < 1/2$

Pr. 2.11⁹ Let X be the number of times a certain numerical control machine will malfunction on any given day. Let Y be the number of times a technician is called on an emergency call. Their joint probability distribution is:

f(x,y)	X	1	2	3
Y	1	0.05	0.05	0.1
	2	0.05	0.1	0.35
	3	0	0.2	0.1

- Determine the marginal distributions of X and of Y
- Determine the probability $p(x < 2, y > 2)$

⁹ From Walpole et al. (2007) by permission of Pearson Education.

Pr. 2.12 Consider the data given in Example 2.2.6 for the case of a residential air conditioner. You will use the same data to calculate the flip problem using Bayes' law.

- During a certain day, it was found that the air-conditioner was operating satisfactorily. Calculate the probabilities that this was a "NH= not hot" day.
- Draw the reverse tree diagram for this case.

Pr. 2.13 Consider a medical test for a disease. The test has a probability of 0.95 of correctly or positively detecting an infected person (this is the *sensitivity*), while it has a probability of 0.90 of correctly identifying a healthy person (this is called the *specificity*). In the population, only 3% of the people have the disease.

- What is the probability that a person testing positive is actually infected?
- What is the probability that a person testing negative is actually infected?

Pr. 2.14 A large industrial firm purchases several new computers at the end of each year, the number depending on the frequency of repairs in the previous year. Suppose that the number of computers X purchased each year has the following PDF:

X	0	1	2	3
f(x)	0.2	0.3	0.2	0.1

If the cost of the desired model will remain fixed at \$ 1500 throughout this year and a discount of \$ $50x^2$ is credited towards any purchase, how much can this firm expect to spend on new computers at the end of this year?

Pr. 2.15 Suppose that the probabilities of the number of power failure in a certain locality are given as:

X	0	1	2	3
f(x)	0.4	0.3	0.2	0.1

Find the mean and variance of the random variable X.

Pr. 2.16 An electric firm manufacturers a 100 W light bulb, which is supposed to have a mean life of 900 and a standard deviation of 50 h. Assume that the distribution is symmetric about the mean. Determine what percentage of the bulbs fails to last even 700 h if the distribution is found to follow: (i) a normal distribution, (ii) a lognormal, (iii) a Poisson, and (iii) a uniform distribution.

Pr. 2.17 Sulfur dioxide concentrations in air samples taken in a certain region have been found to be well represented by a lognormal distribution with parameters $\mu = 2.1$ and $\sigma = 0.8$.

- What proportion of air samples have concentrations between 3 and 6?
- What proportion do not exceed 10?
- What interval (a,b) is such that 95% of all air samples have concentration values in this interval, with 2.5% have values less than a, and 2.5% have values exceeding b?

Pr. 2.18 The average rate of water usage (in thousands of gallons per hour) by a certain community can be modeled by a lognormal distribution with parameters $\mu=4$ and $\sigma=1.5$. What is the probability that the demand will:

- be 40,000 gallons of water per hour
- exceed 60,000 gallons of water per hour

Pr. 2.19 Suppose the number of drivers who travel between two locations during a designated time period is a Poisson distribution with parameter $\lambda=30$. In the long run, in what proportion of time periods will the number of drivers:

- Be at most 20?
- Exceed 25?
- Be between 10 and 20.

Pr. 2.20 Suppose the time, in hours, taken to repair a home furnace can be modeled as a gamma distribution with parameters $a=2$ and $\lambda=1/2$. What is the probability that the next service call will require:

- at most 1 h to repair the furnace?
- at least 2 h to repair the furnace?

Pr. 2.21¹⁰ In a certain city, the daily consumption of electric power, in millions of kilowatts-hours (kWh), is a random variable X having a gamma distribution with mean=6 and variance=12.

- Find the values of the parameters α and λ
- Find the probability that on any given day the daily power consumption will exceed 12 million kWh.

Pr. 2.22 The life in years of a certain type of electrical switches has an exponential distribution with an average life in years of $\lambda=5$. If 100 of these switches are installed in different systems,

- what is the probability that at most 20 will fail during the first year?
- How many are likely to have failed at the end of 3 years?

Pr. 2.23 *Probability models for global horizontal solar radiation.*

Probability models for predicting solar radiation at the surface of the earth was the subject of several studies in the last several decades. Consider the daily values of global (beam plus diffuse) radiation on a horizontal surface at a specified location. Because of the variability of the atmospheric conditions at any given location, this quantity can be viewed as

a random variable. Further, global radiation has an underlying annual pattern due to the orbital rotation of the earth around the sun. A widely adopted technique to filter out this deterministic trend is:

- to select the random variable not as the daily radiation itself but as the daily clearness index K defined as the ratio of the daily global radiation on the earth's surface for the location in question to that outside the atmosphere for the same latitude and day of the year, and
- to truncate the year into 12 monthly time scales since the random variable K for a location changes appreciably on a seasonal basis.

Gordon and Reddy (1988) proposed an expression for the PDF of the random variable $X = (K / \bar{K})$ where \bar{K} is the monthly mean value of the daily values of K during a month. The following empirical model was shown to be of general validity in that it applied to several locations (temperate and tropical) and months of the year with the same variance in K :

$$p(X; A, n) = AX^n[1 - (X/X_{\max})] \quad (2.55)$$

where A , n and X_{\max} are model parameters which have been determined from the normalization of $p(X)$, knowledge of \bar{K} (i.e., $\bar{X} = 1$) and knowledge of the variance of X or $\sigma^2(X)$. Derive the following expressions for the three model parameters:

$$\begin{aligned} n &= -2.5 + 0.5[9 + (8/\sigma^2(X))]^{1/2} \\ X_{\max} &= (n + 3)/(n + 1) \\ A &= (n + 1)(n + 2)/X_{\max}^{n+1} \end{aligned}$$

Note that because of the manner of normalization, the random variable selected can assume values greater than unity. Figure 2.33 shows the proposed distribution for a number of different variance values.

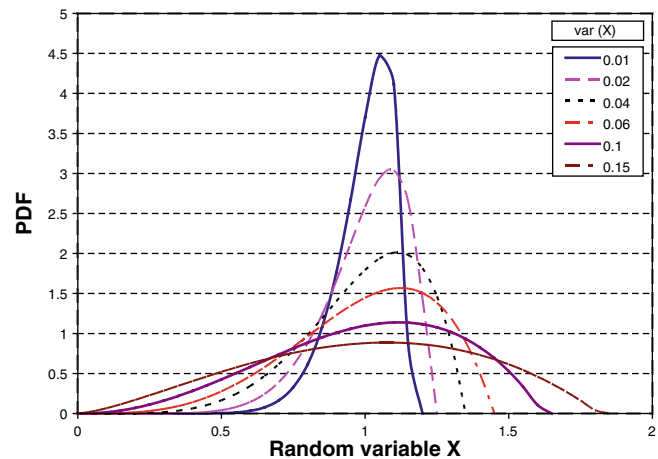


Fig. 2.33 Probability distributions of solar radiation given by Eq. 2.55 as proposed by Gordon and Reddy (1988)

¹⁰From Walpole et al. (2007) by permission of Pearson Education.

Pr. 2.24 Cumulative distribution and utilizability functions for horizontal solar radiation

Use the equations described below to derive the CDF and the utilizability functions for the Gordon-Reddy probability distribution function described in Pr. 2.23.

Probability distribution functions for solar radiation (as in Pr. 2.23 above) and also for ambient air temperatures are useful to respectively predict the long-term behavior of solar collectors and the monthly/annual heating energy use of small buildings. For example, the annual/monthly space heating load Q_{Load} (in MJ) is given by:

$$Q_{Load} = (UA)_{Bldg} \cdot DD \cdot (86,400 \frac{s}{day}) \cdot (10^{-6} \frac{MJ}{J})$$

where $(UA)_{Bldg}$ is the building overall energy loss/gain per unit temperature difference in $W/^\circ C$; and DD is the degree-day given by:

$$DD = \sum_{d=1}^N (18.3 - \bar{T}_d)^+ \quad \text{in } ^\circ C - \text{day}$$

where \bar{T}_d is the daily average ambient air temperature and $N=365$ if annual time scales are considered. The “+” sign indicates that only positive values within the brackets should contribute to the sum, while negative values should be set to zero. Physically this implies that only when the ambient air is lower than the design indoor temperature, which is historically taken as $18.3^\circ C$, would there be a need for heating the building. It is clear that the DD is the sum of the differences between each day’s mean temperature and the design temperature of the conditioned space. It can be derived from knowledge of the PDF of the daily ambient temperature values at the location in question.

A similar approach has also been developed for predicting the long-term energy collected by a solar collector either at the monthly or annual time scale involving the concept of *solar utilizability* (Reddy 1987). Consider Fig. 2.34a which shows the PDF function $P(X)$ for the normalized variable X described in Pr. 2.23, and bounded by X_{min} and X_{max} . The area of the shaded portion corresponds to a specific value X' of the CDF (see Fig. 2.34b):

$$F(X') = \text{probability}(X \leq X') = \int_{X_{min}}^{X'} P(X) dX \quad (2.56a)$$

The long-term energy delivered by a solar thermal collector is proportional to the amount of solar energy above a certain critical threshold X_c , and this is depicted as a shaded area in Fig. 2.34b). This area is called the solar utilizability, and is functionally described by:

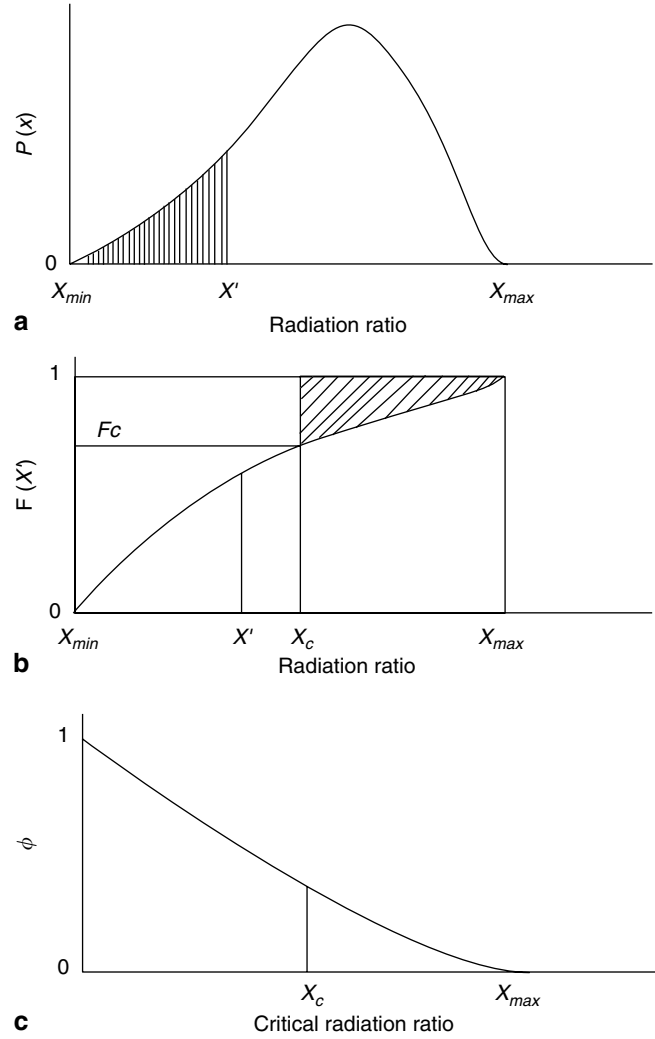


Fig. 2.34 Relation between different distributions. **a** Probability density curve (shaded area represents the cumulative distribution value $F(X')$). **b** Cumulative distribution function (shaded area represents utilizability fraction at X_c). **c** Utilizability curve. (From Reddy 1987)

$$\phi(X_c) = \int_{F_c}^1 (X' - X_c) dF = \int_{X_c}^{X_{max}} [1 - F(X')] dX' \quad (2.56b)$$

The value of the utilizability function for such a critical radiation level X_c is shown in Fig. 2.34c).

Pr. 2.25 Generating cumulative distribution curves and utilizability curves from measured data.

The previous two problems involved probability distributions of solar radiation and ambient temperature, and how these could be used to derive functions for quantities of interest such as the solar utilizability or the Degree-Days. If monitored data is available, there is no need to delve into such considerations of probability distributions, and one can calculate these functions numerically.

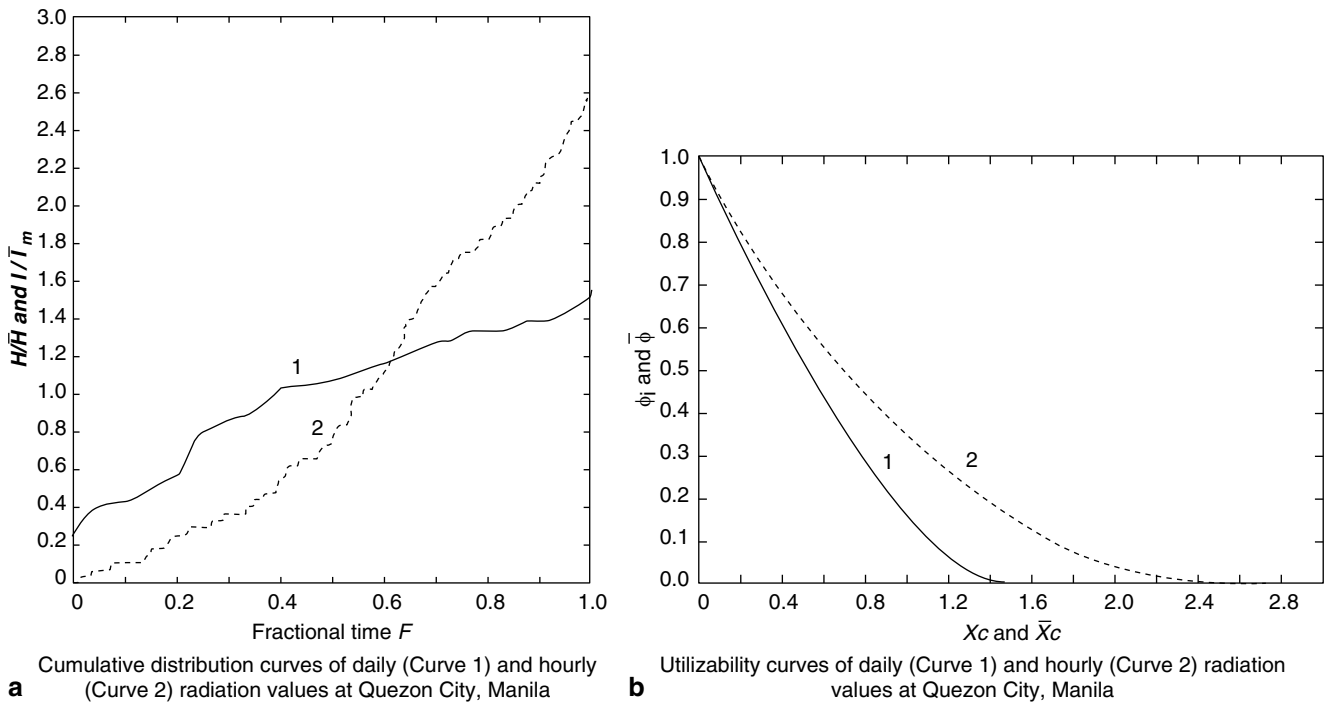


Fig. 2.35 Distribution for Quezon City, Manila during October 1980. (From Reddy 1987)

Consider Table P2.25 (in Appendix B) which assembles the global solar radiation on a horizontal surface at Quezon City, Manila during October 1980 (taken from Reddy 1987). You are asked to numerically generate the CDF and the utilizability functions (Eq. 2.56a, b) and compare your results to Fig. 2.35. The symbols I and H denote hourly and daily radiation values respectively. Note that instead of normalizing the radiation values by the extra-terrestrial solar radiation (as done in Pr. 2.23), here the corresponding average values \bar{I} (for individual hours of the day) and \bar{H} have been used.

References

- Ang, A.H.S. and W.H. Tang, 2007. *Probability Concepts in Engineering*, 2nd Ed., John Wiley and Sons, USA
- Barton, C. and S. Nishenko, 2008. *Natural Disasters: Forecasting Economic and Life Losses*, U.S. Geological Survey, Marine and Coastal Geology Program.
- Bolstad, W.M., 2004. *Introduction to Bayesian Statistics*, Wiley and Sons, Hoboken, NJ.
- Gordon, J. M., and T.A. Reddy, 1988. Time series analysis of daily horizontal solar radiation, *Solar Energy*, 41(3), pp.215-226
- Kolluru, R.V., S.M. Bartell, R.M. Pitblado, and R.S. Stricoff, R.S., 1996. *Risk Assessment and Management Handbook*, McGraw-Hill, New York.
- McClave, J.T. and P.G. Benson, 1988. *Statistics for Business and Economics*, 4th Ed., Dellen and Macmillan, London.
- Nordhaus, W.D., 1994. Expert opinion on climate change, *American Scientist*, 82: pp.45-51.
- Phillips, L.D., 1973. *Bayesian Statistics for Social Scientists*, Thomas Nelson and Sons, London, UK
- Reddy, T.A., 1987. *The Design and Sizing of Active Solar Thermal Systems*, Oxford University Press, Oxford, U.K.
- Walpole, R.E., R.H. Myers, S.L. Myers, and K. Ye, 2007, *Probability and Statistics for Engineers and Scientists*, 8th Ed., Prentice Hall, Upper Saddle River, NJ.



<http://www.springer.com/978-1-4419-9612-1>

Applied Data Analysis and Modeling for Energy
Engineers and Scientists

Reddy, T.A.

2011, XXI, 430 p., Hardcover

ISBN: 978-1-4419-9612-1