Chapter 2

Elements of Probability Theory

2.1 Introduction

Whether referring to a storm's intensity, an arrival time, or the success of a decision, the word "probable," or "likely," has long been part of our language. Most people have an appreciation for the impact of chance on the occurrence of an event. In the last 350 years, the theory of probability has evolved to explain the nature of chance and how it can be studied.

Probability theory is the formal study of events whose outcomes are uncertain. Its origins trace to 17th-century gambling problems. Games that involved playing cards, roulette wheels, and dice provided mathematicians with a host of interesting problems. The solutions to many of these problems yielded the first principles of modern probability theory. Today, probability theory is of fundamental importance in science, engineering, and business.

Engineering risk management aims to identify and manage events whose outcomes are uncertain. In particular, its focus is on events that, if they occur, have unwanted impacts or consequences to a project or program. The phrase "*if they occur*" means these events are probabilistic in nature. Thus, understanding them in the context of probability concepts is essential. This chapter presents an introduction to these concepts and illustrates how they apply to managing risks in engineering systems.

2.2 Interpretations and Axioms

We begin this discussion with the traditional look at dice. If a six-sided die is tossed, there clearly are six possible outcomes for the number that appears on the upturned face. These outcomes can be listed as elements in a set $\{1, 2, 3, 4, 5, 6\}$.

The set of all possible outcomes of an experiment, such as tossing a six-sided die, is called the *sample space*, which we will denote by Ω. The individual outcomes of Ω are called sample points, which we will denote by ω.

An *event* is any subset of the sample space. An event is *simple* if it consists of exactly one outcome. Simple events are also referred to as *elementary* events or elementary outcomes. An event is *compound* if it consists of more than one outcome. For instance, let A be the event an odd number appears and B be the event an even number appears in a single toss of a die. These are compound events that can be expressed by the sets $A = \{1, 3, 5\}$ and $B = \{2, 4, 6\}$. Event A occurs *if and only if* one of the outcomes in A occurs. The same is true for event B.

Seen in this discussion, events can be represented by sets. New events can be constructed from given events according to the rules of set theory. The following presents a brief review of set theory concepts.

Union. For any two events A and B of a sample space, the new event $A \cup B$ (which reads A union B) consists of all outcomes either in A or in B or in both A and B. The event $A \cup B$ occurs if either A or B occurs. To illustrate the union of two events, consider the following: if A is the event an odd number appears in the toss of a die and B is the event an even number appears, then the event $A \cup B$ is the set $\{1, 2, 3, 4, 5, 6\}$, which is the sample space for this experiment.

Intersection. For any two events A and B of a sample space Ω, the new event $A \cap B$ (which reads A *intersection* B) consists of all outcomes that are in *both* A and in B. The event $A \cap B$ occurs *only if both* A *and* B *occur*. To illustrate the intersection of two events, consider the following: if A is the event a 6 appears in the toss of a die, B is the event an odd number appears, and C is the event an even number appears, then the event $A \cap C$ is the simple event $\{6\}$; on the other hand, the event $A \cap B$ contains no outcomes. Such an event is called the *null event*. The null event is traditionally denoted by $\emptyset$. In general, if $A \cap B = \emptyset$, we say events A and B are *mutually exclusive* (*disjoint*). For notation convenience, the intersection of two events A and B is sometimes written as AB, instead of $A \cap B$.

Complement. The complement of event A, denoted by A^c, consists of all outcomes in the sample space Ω that are not in A. The event A^c occurs *if and only if* A does not occur. The following illustrates the complement of an event. If C is the event an even number appears in the toss of a die, then C^c is the event an odd number appears.

Subset. Event A is said to be a subset of event B if all the outcomes in A are also contained in B. This is written as $A \subset B$.

In the preceding discussion, the sample space for the toss of a die was given by $\Omega = \{1, 2, 3, 4, 5, 6\}$. If we *assume* the die is fair, then any outcome in the sample space is as likely to appear as any other. Given this, it is reasonable to conclude the proportion of time each outcome is expected to occur is 1/6. Thus, the probability of each simple event in the sample space is

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}$$

Similarly, suppose B is the event an odd number appears in a single toss of the die. This compound event is given by the set $B = \{1, 3, 5\}$. Since there are three ways event B can occur out of six possible, the probability of event B is

$$P(B) = \frac{3}{6} = \frac{1}{2}$$

The following presents a view of probability known as the equally likely interpretation.

Equally Likely Interpretation. In this view, if a sample space Ω consists of a finite number of outcomes n, which are all equally likely to occur, then the probability of each simple event is $1/n$. If an event A consists of m of these n outcomes, then the probability of event A is

$$P(A) = \frac{m}{n} \tag{2.1}$$

In the above, it is assumed the sample space consists of a *finite* number of outcomes and all outcomes are equally likely to occur. What if the sample space is finite but the outcomes are *not* equally likely? In these situations, probability might be measured in terms of how frequently a particular outcome occurs when the experiment is repeatedly performed under identical conditions. This leads to a view of probability known as the frequency interpretation.

Frequency Interpretation. In this view, the probability of an event is the limiting proportion of time the event occurs in a set of n repetitions of the experiment. In particular, we write this as

$$P(A) = \lim_{n \to \infty} \frac{n(A)}{n}$$

where $n(A)$ is the number of times in n repetitions of the experiment the event A occurs. In this sense $P(A)$ is the limiting frequency of event A. Probabilities measured by the frequency interpretation are referred to as *objective probabilities*. In many circumstances it is appropriate to work with objective probabilities. However, there are limitations with this interpretation of probability. It restricts events to those that can be subjected to repeated trials conducted under *identical conditions*. Furthermore, it is not clear how many trials of an experiment are needed to obtain an event's limiting frequency.

Axiomatic Definition. In 1933, the Russian mathematician A.N. Kolmogorov* presented a definition of probability in terms of three axioms [1]. These axioms define probability in a way that encompasses the *equally likely and frequency interpretations* of probability. It is known as the axiomatic definition of probability. It is the view of probability adopted in this book. Under this definition, it is assumed for each event A, in the sample space Ω, there is a real number $P(A)$ that denotes the probability of A. In accordance with Kolmogorov's axioms, a probability is simply a numerical measure that satisfies the following:

Axiom 1 $0 \leq P(A) \leq 1$ for any event A in Ω

Axiom 2 $P(\Omega) = 1$

Axiom 3 For any sequence of mutually exclusive events $A_1, A_2, \ldots$ defined on Ω it follows that $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$
For any *finite sequence* of mutually exclusive events $A_1, A_2, \ldots, A_n$ defined on Ω it follows that $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$

The first axiom states the probability of any event is a non-negative number in the interval zero to unity. In axiom 2, the sample space Ω is sometimes referred to as the *sure* or *certain event*; therefore, we have $P(\Omega)$ equal to one. Axiom 3 states for any sequence of mutually exclusive events, the probability of at least one of these events' occurring is the sum of the probabilities associated with each event A_i. In axiom 3, this sequence may also be finite. From these axioms come five basic theorems of probability.

*A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, *Ergeb. Mat. und ihrer Grenzg.*, vol. 2, no. 3, 1933. Translated into English by N. Morrison, *Foundations of the Theory of Probability*, New York (Chelsea), 1956 [1].

Theorem 2.1 *The probability event A occurs is one minus the probability it will not occur; that is,*

$$P(A) = 1 - P(A^c)$$

Theorem 2.2 *The probability associated with the null event ∅ is zero; that is,*

$$P(\emptyset) = 0$$

Theorem 2.3 *If events A_1 and A_2 are mutually exclusive, then*

$$P(A_1 \cap A_2) \equiv P(A_1 A_2) = 0$$

Theorem 2.4 *For any two events A_1 and A_2*

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

Theorem 2.5 *If event A_1 is a subset of event A_2 then*

$$P(A_1) \leq P(A_2)$$

Measure of Belief Interpretation. From the axiomatic view, probability need only be a numerical measure satisfying the three axioms stated by Kolmogorov. Given this, it is possible for probability to reflect a "measure of belief" in an event's occurrence. For instance, an engineer might assign a probability of 0.70 to the event "the radar software for the Advanced Air Traffic Control System (AATCS) will not exceed 100K delivered source instructions." We consider this event to be non-repeatable. It is not practical, or possible, to build the AATCS n-times (and under identical conditions) to determine whether this probability is indeed 0.70. When an event such as this arises, its probability may be assigned. Probabilities assigned on the basis of personal judgment, or measure of belief, are known as *subjective probabilities*.

Subjective probabilities are the most common in engineering system projects. Such probabilities are typically assigned by expert technical judgment. The engineer's probability assessment of 0.70 is a subjective probability. Ideally, subjective probabilities should be based on available evidence and previous experience with similar events. Subjective probabilities become suspect if they are premised on limited insights or no prior experience. Care is also needed in soliciting subjective probabilities. They must certainly be plausible *and* they must be *consistent* with Kolmogorov's axioms and the theorems of probability, which stem from these axioms. Consider the following:

The XYZ Corporation has offers on two contracts A and B. Suppose the proposal team made the following subjective probability assignments. The chance of winning contract A is 40%, the chance of winning contract B is 20%, the chance of winning *contract A or contract B* is 60%, and the chance of winning *both contract A and contract B* is 10%. It turns out this set of probability assignments is *not* consistent with the axioms and theorems of probability. Why is this?* If the chance of winning contract B was changed to 30%, then this *set of probability assignments* would be consistent.

Kolmogorov's axioms, and the resulting theorems of probability, do not suggest how to assign probabilities to events. Instead, they provide a way to verify that probability assignments are consistent, whether these probabilities are objective or subjective.

Risk versus Uncertainty. There is an important distinction between the terms *risk* and *uncertainty*. Risk is the chance of loss or injury. In a situation that includes favorable and unfavorable events, risk is the *probability an unfavorable event occurs*. Uncertainty is the *indefiniteness about the outcome of a situation*. We analyze uncertainty for the purpose of measuring risk. In systems engineering the analysis might focus on measuring the risk of: (1) failing to achieve performance objectives, (2) overrunning the budgeted cost, or (3) delivering the system too late to meet user needs. Conducting the analysis often involves degrees of subjectivity. This includes defining the events of concern and, when necessary, subjectively specifying their occurrence probabilities. Given this, it is fair to ask whether it is meaningful to apply rigorous mathematical procedures to such analyses. In a speech before the 1955 Operations Research Society of America meeting, Charles J. Hitch (RAND) addressed this question. He stated [2, 3]:

> Systems analyses provide a framework which permits the judgment of experts in many fields to be combined to yield results that transcend any individual judgment. The systems analyst may have to be content with better rather than optimal solutions; or with devising and costing sensible methods of hedging; or merely with discovering critical sensitivities. We tend to be worse, in an absolute sense, in applying analysis or scientific method to broad context problems; but unaided intuition in such problems is also much worse in the absolute sense. Let's not deprive ourselves of any useful tools, however short of perfection they may fail.

---

*The answer can be seen from theorem 2.4.

2.3 Conditional Probability and Bayes' Rule

In many circumstances, the probability of an event is conditioned on knowing another event has taken place. Such a probability is known as a *conditional probability*. *Conditional probabilities* incorporate information about the occurrence of another event. The conditional probability of event A given event B has occurred is denoted by $P(A\,|B)$. If a pair of dice is tossed, then the probability the sum of the toss is *even* is 1/2. This probability is known as a *marginal* or *unconditional probability*.

How would this unconditional probability change (i.e., be conditioned) if it was *known* the sum of the toss was a number less than 10? This is discussed in the following example.

Example 2.1
A pair of dice is tossed and the sum of the toss is a number less than 10. Given this, compute the probability this sum is an even number.

Solution
Suppose we define events A and B as follows:

A: The sum of the toss is even

B: The sum of the toss is a number less than 10

The sample space Ω contains 36 possible outcomes; however, in this case we want the subset of Ω containing *only* those outcomes whose toss yielded a sum less than 10. This subset is shown in Table 2.1. It contains 30 outcomes. Within Table 2.1, only 14 outcomes are associated with the event "the sum of the toss is even given the sum of the toss is a number less than 10."

$$\left\{\begin{matrix}(1,1),(1,3),(1,5),(2,2),(2,4),(2,6),(3,1),(3,3),(3,5)\\(4,2),(4,4),(5,1),(5,3),(6,2)\end{matrix}\right\}$$

Therefore, the probability of this event is $P(A\,|B) = 14/30$

In Example 2.1, observe that $P(A\,|B)$ was obtained directly from a subset of the sample space Ω and that $P(A\,|B) = 14/30 < P(A) = 1/2$.

If A and B are events in the same sample space Ω, then $P(A\,|B)$ is the probability of event A within the subset of the sample space defined by event B.

TABLE 2.1: Outcomes Associated with Event B

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	**(2,2)**	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	**(3,3)**	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	**(4,4)**	(4,5)	
(5,1)	(5,2)	(5,3)	(5,4)		
(6,1)	(6,2)	(6,3)			

A subset of Ω that contains only those outcomes whose toss yielded a sum less than 10

Formally, the *conditional probability of event A given event B has occurred*, where $P(B) > 0$, is defined as

$$P(A\,|B) = \frac{P(A \cap B)}{P(B)} \tag{2.2}$$

Likewise, the conditional probability of event B given event A has occurred, where $P(A) > 0$, is defined as

$$P(B\,|A) = \frac{P(B \cap A)}{P(A)} \tag{2.3}$$

Example 2.2

A proposal team from XYZ Corporation has offers on two contracts A and B. The team made subjective probability assignments on the chances of winning these contracts. They assessed a 40% chance on the event winning contract A, a 50% chance on the event's winning contract B, and a 30% chance on the event winning both contracts. Given this, what is the probability of:

(a) Winning at least one of these contracts?

(b) Winning contract A and not winning contract B?

(c) Winning contract A if the proposal team has won at least one contract?

Solution

(a) Winning at least one contract means winning either contract A *or* contract B or both contracts. This event is represented by the set $A \cup B$. From theorem 2.4

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

therefore

$$P(A \cup B) = 0.40 + 0.50 - 0.30 = 0.60$$

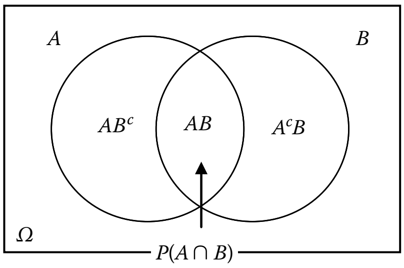


Figure 2.1: Venn diagram for $P(A) = P((A \cap B^c) \cup (A \cap B))$.

(b) The event winning contract A and not winning contract B is represented by the set $A \cap B^c$. From the Venn diagram in Figure 2.1, observe that

$$P(A) = P((A \cap B^c) \cup (A \cap B))$$

Since the events $A \cap B^c$ and $A \cap B$ are mutually exclusive (disjoint), from theorem 2.3 and theorem 2.4 we have

$$P(A) = P(A \cap B^c) + P(A \cap B)$$

This is equivalent to

$$P(A \cap B^c) = P(A) - P(A \cap B)$$

therefore,

$$P(A \cap B^c) = P(A) - P(A \cap B) = 0.40 - 0.30 = 0.10$$

(c) If the proposal team has won one of the contracts, the probability of winning contract A must be revised (or conditioned) on this information. This means we must compute $P(A \,|A \cup B\,)$. From Equation 2.2

$$P(A \,|A \cup B\,) = \frac{P(A \cap (A \cup B))}{P(A \cup B)}$$

Since

$$P(A) = P(A \cap (A \cup B))$$

we have

$$P(A \,|A \cup B\,) = \frac{P(A \cap (A \cup B))}{P(A \cup B)} = \frac{P(A)}{P(A \cup B)} = \frac{0.40}{0.60} = \frac{2}{3} \approx 0.67$$

A consequence of conditional probability is obtained if we multiply Equations 2.2 and 2.3 by $P(B)$ and $P(A)$, respectively. This multiplication yields

$$P(A \cap B) = P(B)P(A\,|B) = P(A)P(B\,|A) \tag{2.4}$$

Equation 2.4 is known as the *multiplication rule*. The multiplication rule provides a way to express the probability of the intersection of two events in terms of their conditional probabilities. An illustration of this rule is presented in example 2.3.

Example 2.3

A box contains memory chips of which 3 are defective and 97 are non-defective. Two chips are drawn at random, one after the other, without replacement. Determine the probability:

(a) Both chips drawn are defective.

(b) The first chip is defective and the second chip is non-defective.

Solution

(a) Let A and B denote the event the first and second chips drawn from the box are *defective*, respectively. From the multiplication rule, we have

$$\begin{aligned} P(A \cap B) &= P(A)P(B\,|A) \\ &= P(\text{1st chip defective})\,P(\text{2nd chip defective}|\text{1st chip defective}) \\ &= \frac{3}{100}\left(\frac{2}{99}\right) = \frac{6}{9900} \end{aligned}$$

(b) To determine the probability the first chip drawn is defective and the second chip is *non-defective*, let C denote the event the second chip drawn is non-defective. Thus,

$$\begin{aligned} P(A \cap C) &= P(AC) = P(A)P(C\,|A) \\ &= P(\text{1st chip defective})\,P(\text{2nd chip nondefective}|\text{1st chip defective}) \\ &= \frac{3}{100}\left(\frac{97}{99}\right) = \frac{291}{9900} \end{aligned}$$

In this example the sampling was performed *without replacement*. Suppose the chips sampled were *replaced*; that is, the first chip selected was replaced before the second chip was selected. In that case, the probability of a defective chip's being selected on the second drawing is independent of the outcome of the first chip drawn. Specifically,

$$P(\text{2nd chip defective}) = P(\text{1st chip defective}) = 3/100$$

so

$$P(A \cap B) = \frac{3}{100}\left(\frac{3}{100}\right) = \frac{9}{10000}$$

and

$$P(A \cap C) = \frac{3}{100}\left(\frac{97}{100}\right) = \frac{291}{10000}$$

Independent Events

Two events A and B are said to be *independent* if and only if

$$P(A \cap B) = P(A)P(B) \tag{2.5}$$

and *dependent* otherwise. Events $A_1, A_2, \ldots, A_n$ are (mutually) *independent* if and only if for every set of indices $i_1, i_2, \ldots, i_k$ between 1 and n, inclusive,

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \ldots P(A_{i_k}), \quad (k = 2, \ldots, n)$$

For instance, events A_1, A_2, and A_3, are independent (or mutually independent) if the following equations are satisfied

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) \tag{2.5a}$$

$$P(A_1 \cap A_2) = P(A_1)P(A_2) \tag{2.5b}$$

$$P(A_1 \cap A_3) = P(A_1)P(A_3) \tag{2.5c}$$

$$P(A_2 \cap A_3) = P(A_2)P(A_3) \tag{2.5d}$$

It is possible to have three events A_1, A_2, and A_3 for which Equations 2.5b through 2.5d hold but Equation 2.5a does not hold. Mutual independence implies pairwise independence, in the sense that Equations 2.5b through 2.5d hold, but the converse is not true.

There is a close relationship between independent events and conditional probability. To see this, suppose events A and B are independent. This implies

$$P(AB) = P(A)P(B)$$

From this, Equations 2.2 and 2.3 become, respectively, $P(A\,|B) = P(A)$ and $P(B\,|A) = P(B)$. Thus, when two events are independent the occurrence of one event has no impact on the probability the other event occurs.

To illustrate the concept of independence, suppose a fair die is tossed. Let A be the event an odd number appears. Let B be the event one of these numbers {2, 3, 5, 6} appears. From this,

$$P(A) = 1/2$$

and

$$P(B) = 2/3$$

Since $A \cap B$ is the event represented by the set $\{3, 5\}$, we can readily state $P(A \cap B) = 1/3$. Therefore, $P(A \cap B) = P(AB) = P(A)P(B)$ and we conclude events A and B are independent.

Dependence can be illustratedby tossing two fair dice. Suppose A is the event the sum of the toss is odd and B is the event the sum of the toss is even. Here, $P(A \cap B) = 0$ and $P(A)$ and $P(B)$ were each 1/2. Since $P(A \cap B) \neq P(A)P(B)$ we would conclude events A and B are dependent, in this case.

It is important not to confuse the meaning of independent events with mutually exclusive events as shown in Figure 2.2. If events A and B are mutually exclusive, the event A *and* B is empty; that is, $A \cap B = \emptyset$. This implies $P(A \cap B) = P(\emptyset) = 0$. If events A and B are *independent* with $P(A) \neq 0$ and $P(B) \neq 0$, then A and B cannot be mutually exclusive since $P(A \cap B) = P(A)P(B) \neq 0$.

Bayes' Rule

Suppose we have a collection of events A_i representing possible conjectures about a topic. Furthermore, suppose we have some initial probabilities associated with the "truth" of these conjectures. Bayes' rule* provides a way to update (or revise) initial probabilities when new information about these conjectures is evidenced.

*Named in honor of Thomas Bayes (1702–1761), an English minister and mathematician.

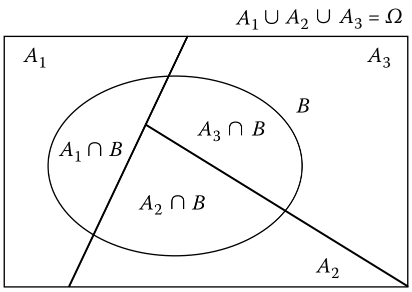


Figure 2.2: Partitioning Ω into three mutually exclusive sets.

Bayes' rule is a consequence of conditional probability. Suppose we partition a sample space Ω into a finite collection of three mutually exclusive events as shown in Figure 2.2. Define these events as A_1, A_2, and A_3 where $A_1 \cup A_2 \cup A_3 = \Omega$. Let B denote an arbitrary event contained in Ω. We can write the event B as

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B)$$

Since the events $(A_1 \cap B)$, $(A_2 \cap B)$, $(A_3 \cap B)$ are mutually exclusive, we can apply axiom 3 and write

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$$

From the multiplication rule given in Equation 2.4, $P(B)$ can be expressed in terms of conditional probability as

$$P(B) = P(A_1)P(B\,|A_1) + P(A_2)P(B\,|A_2) + P(A_3)P(B\,|A_3)$$

This equation is known as the *total probability law*. Its generalization is

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B\,|A_i)$$

where $\Omega = \bigcup_{i=1}^{n} A_i$ and $A_i \cap A_j = \emptyset$ and $i \neq j$.

The conditional probability for each event A_i given event B has occurred is

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

When the total probability law is applied to this Equation we have

$$P(A_i\,|B) = \frac{P(A_i)P(B\,|A_i)}{\sum_{i=1}^{n} P(A_i)P(B\,|A_i\,)} \tag{2.6}$$

Equation 2.6 is known as *Bayes' Rule*.

Example 2.4

The ChipyTech Corporation has three divisions, D_1, D_2, and D_3, that each manufacture a specific type of microprocessor chip. From the total annual output of chips produced by the corporation, D_1 manufactures 35%, D_2 manufactures 20%, and D_3 manufactures 45%. Data collected from the quality control group indicate 1% of the chips from D_1 are defective, 2% of the chips from D_2 are defective, and 3% of the chips from D_3 are defective. Suppose a chip was randomly selected from the total annual output produced and it was found to be defective. What is the probability it was manufactured by D_1? By D_2? By D_3?

Solution

Let A_i denote the event the selected chip was produced by division $D_i (i = 1, 2, 3)$. Let B denote the event the selected chip is defective. To determine the probability the defective chip was manufactured by D_i we must compute the conditional probability $P(A_i\,|B)$ for $i = 1, 2, 3$. From the information provided, we have

$$P(A_1) = 0.35, \quad P(A_2) = 0.20, \quad \text{and} \quad P(A_3) = 0.45$$
$$P(B|A_i) = 0.01, \quad P(B|A_2) = 0.02, \quad P(B|A_3) = 0.03$$

The total probability law and Bayes' rule will be used to determine $P(A_i\,|B)$ for each i = 1, 2, and 3. Recall from Equation 2.9 that $P(B)$ can be written as

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3)$$
$$P(B) = 0.35(0.01) + 0.20(0.02) + 0.45(0.03) = 0.021$$

and from Bayes' rule we can write

$$P(A_i\,|\,B) = \frac{P(A_i)P(B\,|A_i)}{\sum_{i=1}^{n} P(A_i)P(B\,|A_i\,)} = \frac{P(A_i)P(B\,|A_i)}{P(B)}$$

TABLE 2.2: Bayes' Probability: Example 2.4 Summary

i	$P(A_i)$	$P(A_i \mid B)$
1	0.35	0.167
2	0.20	0.190
3	0.45	0.643

from which

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{P(B)} = \frac{0.35(0.01)}{0.021} = 0.167$$

$$P(A_2 \mid B) = \frac{P(A_2)P(B \mid A_2)}{P(B)} = \frac{0.20(0.02)}{0.021} = 0.190$$

$$P(A_3 \mid B) = \frac{P(A_3)P(B \mid A_3)}{P(B)} = \frac{0.45(0.03)}{0.021} = 0.643$$

Table 2.2 provides a comparison of $P(A_i)$ with $P(A_i \mid B)$ for each $i = 1, 2, 3$.

The probabilities given by $P(A_i)$ are the probabilities the selected chip will have been produced by division D_i before it is randomly selected and before it is known whether the chip is defective. Therefore, $P(A_i)$ are the *prior*, or *a priori* (before-the-fact) probabilities. The probabilities given by $P(A_i \mid B)$ are the probabilities the selected chip was produced by division D_i after it is known the selected chip is defective. Therefore, $P(A_i \mid B)$ are the *posterior*, or *a posteriori* (after-the-fact) probabilities. Bayes' rule provides a means for the computation of posterior probabilities from the known prior probabilities $P(A_i)$ and the conditional probabilities $P(B \mid A_i)$ for a particular situation or experiment.

Bayes' rule established areas of study that became known as *Bayesian inference* and *Bayesian decision theory*. These areas play important roles in the application of probability theory to systems engineering problems. In the total probability law, we can think of A_i as representing possible states of nature to which an engineer assigns subjective probabilities. These subjective probabilities are the prior probabilities, which are often premised on personal judgments based on past experience. In general, Bayesian methods offer a powerful way to revise or update probability assessments as new information becomes available.

2.4 Applications to Engineering Risk Management

Chapter 2 concludes with an expanded discussion of Bayes' rule in terms of its application to the analysis of risks in the engineering of systems. In addition, a best-practice protocol for expressing risk in terms of its occurrence probability and consequences is introduced.

2.4.1 Probability Inference — An Application of Bayes' Rule

This discussion presents a technique known as *Bayesian inference*. Bayesian inference is a way to examine how an initial belief in the truth of a hypothesis H may change when evidence e relating to it is observed. This is done by an application of Bayes' rule, which we illustrate in the discussion below.

Suppose an engineering firm has been awarded a project to develop a software application. Suppose a number of challenges are associated with this, among them (1) staffing the project, (2) managing multiple development sites, and (3) functional requirements that continue to evolve.

Given these challenges, suppose the project's management team believes it has a 50% chance of completing the software development in accordance with the customer's planned schedule. From this, how might management use Bayes' rule to monitor whether the *chance* of completing the project on schedule is increasing or decreasing?

As mentioned above, *Bayesian inference* is a procedure that takes evidence, observations, or indicators as they emerge and applies Bayes' rule to infer the truthfulness or falsity of a hypothesis in terms of its probability. In this case, the hypothesis H is *Project XYZ will experience significant delays in completing its software development*.

Suppose at time t_1 the project's management comes to recognize that *project XYZ has been unable to fully staff to the number of software engineers needed for this effort*. In Bayesian inference, we treat this as an observation or evidence that has some bearing on the truthfulness of H. This is illustrated in Figure 2.3. Here, H is the hypothesis "node" and e_1 is the evidence node contributing to the truthfulness of H.

Given the evidence-to-hypothesis relationship in Figure 2.3 we can form the following equations from Bayes' rule.

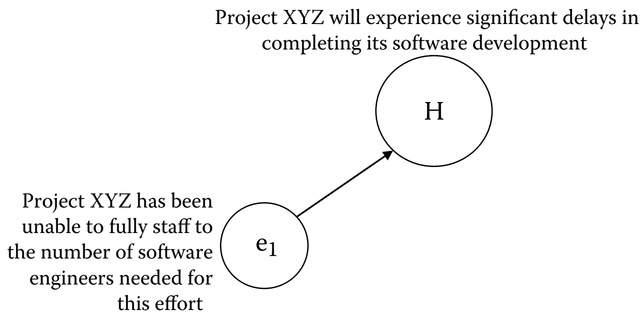


Figure 2.3: Evidence observed at time t_1.

$$P(H\,|e_1) = \frac{P(H)P(e_1\,|H)}{P(H)P(e_1\,|H) + P(H^c)P(e_1\,|H^c)} \tag{2.7}$$

$$P(H\,|e_1) = \frac{P(H)P(e_1\,|H)}{P(H)P(e_1\,|H) + (1 - P(H))P(e_1\,|H^c)} \tag{2.8}$$

Here, $P(H)$ is the team's initial or *prior* subjective (judgmental) probability that Project XYZ will be completed in accordance with the customer's planned schedule. Recall from the above discussion this was $P(H) = 0.50$. The other terms in Equation 2.7 (or Equation 2.8) are defined as follows: $P(H\,|e_1)$ is the probability H is true given evidence e_1, the term $P(e_1\,|H)$ is the probability evidence e_1 would be observed given H is *true*, and the term $P(e_1\,|H^c)$ is the probability evidence e_1 would be observed given H is *not true*.

Suppose this team's experience with e_1 is that staffing shortfalls is a factor that contributes to delays in completing software development projects. Given this, suppose they judge $P(e_1\,|H)$ and $P(e_1\,|H^c)$ to be 0.60 and 0.25, respectively.

From the evidence e_1 and the team's probability assessments related to e_1 we can compute a revised probability that Project XYZ will experience significant delays in completing its software development. This revised probability is given by Equation 2.9.

$$\begin{aligned} P(H\,|e_1) &= \frac{P(H)P(e_1\,|H)}{P(H)P(e_1\,|H) + (1 - P(H))P(e_1\,|H^c)} \\ &= \frac{(0.50)(0.60)}{(0.50)(0.60) + (1 - 0.50)(0.25)} = 0.70589 \end{aligned} \tag{2.9}$$

Notice the effect evidence e_1 has on increasing the probability that Project XYZ will experience a significant schedule delay. We've gone from the initial or *prior* probability of 50% to a *posterior* probability of just over 70%.

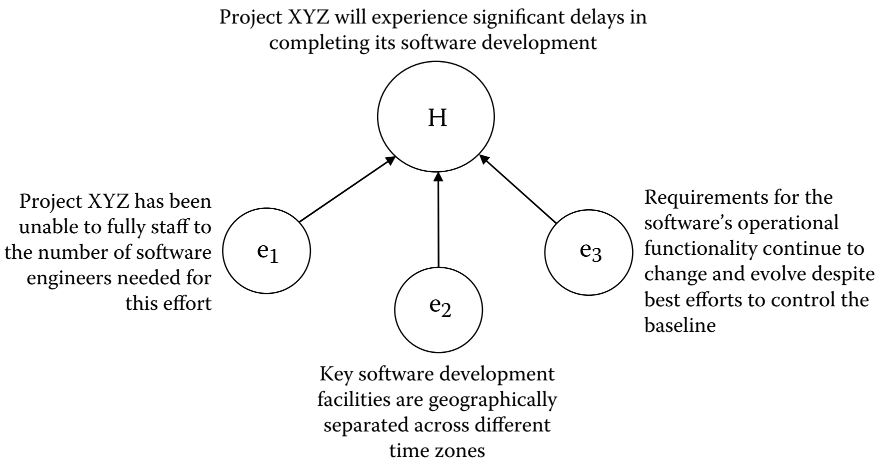


Figure 2.4: Evidence e_2 and e_3 observed at time t_2.

In the Bayesian inference community this is sometimes called *updating*; that is, updating the "belief" in the truthfulness of a hypothesis in light of observations or evidence that adds new information to the initial or prior assessments.

Next, suppose the management team observed two more evidence nodes at time t_2. Suppose these are in addition to the continued relevance of evidence node e_1. Suppose the nature of evidence nodes e_2 and e_3 are described in Figure 2.4. Now, what is the chance Project XYZ will experience a significant schedule delay given all the evidence collected in the set shown in Figure 2.4? Bayesian updating will again be used to answer this question.

Here, we will show how Bayesian updating is used to sequentially revise the *posterior* probability computed in Equation 2.9, to account for the observation of new evidence nodes e_2 and e_3. We begin by writing the following:

$$P(H\,|e_1 \cap e_2) \equiv P(H\,|e_1 e_2) \tag{2.10}$$

$$P(H\,|e_1 e_2) = \frac{P(H\,|e_1)\;P(e_2\,|H)}{P(H\,|e_1)\;P(e_2\,|H) + (1 - P(H\,|e_1))P(e_2\,|H^c)} \tag{2.11}$$

$$P(H\,|e_1 \cap e_2 \cap e_3) \equiv P(H\,|e_1 e_2 e_3) \tag{2.12}$$

$$P(H\,|e_1 e_2 e_3) = \frac{P(H\,|e_1 e_2)\;P(e_3\,|H)}{P(H\,|e_1 e_2)\;P(e_3\,|H) + (1 - P(H\,|e_1 e_2)\,|e_1))P(e_3\,|H^c)} \tag{2.13}$$

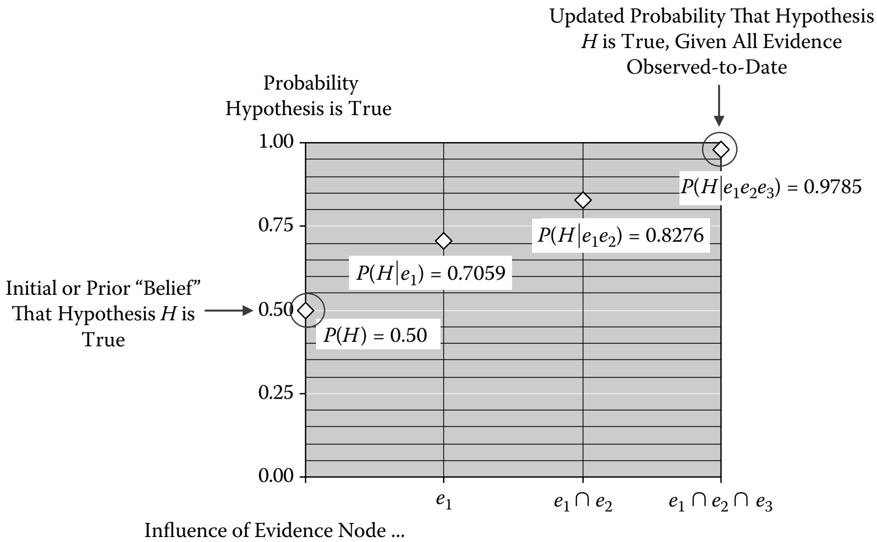


Figure 2.5: Bayesian updating: truthfulness of hypothesis H.

Suppose the management team made the following assessments.

$$P(e_2 \,|H) = 0.90, \quad P(e_2 \,|H^c) = 0.45$$
$$P(e_3 \,|H) = 0.95, \quad P(e_3 \,|H^c) = 0.10$$

Substituting them first into Equation 2.11 and then into 2.13 yields the following:

$$P(H|e_1e_2) = 0.8276 \quad \text{and} \quad P(H|e_1e_2e_3) = 0.9785$$

Thus, given the influence of *all* the evidence observed we can conclude hypothesis H is almost certain to occur. Figure 2.5 illustrates the findings from this analysis.

2.4.2 Writing a Risk Statement

Fundamentally, probability is a measure of the chance an *event* may or may not occur. Furthermore, all probabilities are conditional in the broadest sense that one can always write the following*:

$$Prob(A|\Omega) = Prob(A)$$

where A is an event (a subset) contained in the sample space Ω.

*This result derives from the fact that $Prob(\Omega|A) = 1$.

In a similar way, one can consider subjective or judgmental probabilities as conditional probabilities. The conditioning event (or events) may be experience with the occurrence of events known to have a bearing on the occurrence probability of the future event. Conditioning events can also manifest themselves as evidence, as discussed in the previous section on Bayesian inference.

Given these considerations, a "best practice" for expressing an identified risk is to write it in a form known as the *risk statement*. A risk statement aims to provide clarity and descriptive information about the identified risk so a reasoned and defensible assessment can be made on the risk's occurrence probability and its areas of impact (if the risk event occurs).

A protocol for writing a risk statement is the *Condition-If-Then* construct. This protocol applies in all risk management processes designed for any systems engineering environment. It is a recognition that a risk event is, by its nature, a probabilistic event and one that, if it occurs, has unwanted consequences.

What is the *Condition-If-Then* construct? The *Condition* reflects what is known today. It is the *root cause* of the identified risk event. Thus, the *Condition* is an event that has occurred, is presently occurring, or will occur with certainty. Risk events are future events that may occur *because* of the *Condition* present. Below is an illustration of this protocol.

Suppose we have the following two events. Define the *Condition* as event B and the *If* as event A (the risk event)

B = {Current test plans are focused on the components of the subsystem and not on the subsystem as a whole}

A = {Subsystem will not be fully tested when integrated into the system for full-up system-level testing}

The risk statement is the *Condition-If* part of the construct; specifically,

> **Risk Statement:** {*The subsystem will not be fully tested when integrated into the system for full-up system-level testing, because current test plans are focused on the components of the subsystem and not on the subsystem as a whole.*}

From the above, we see the *Condition-If* part of the construct is equivalent to a probability event; formally, we can write

$$0 < P(A \mid B) = \alpha < 1$$

where α is the probability risk event A occurs given the conditioning event B (the root cause event) has occurred. Why do you think $P(A \mid B)$ here is written as a strict inequality?

In the above, it was explained why a risk event is equivalent to a probability event; that is, the *Condition-If* part of the construct. The *Then* part of the construct contains additional information; that is, information on the risk's consequences. An example is shown on the right side of Figure 2.6.

In summary, a best-practice formalism for writing a risk is to follow the *Condition-If-Then* construct. Here, the *Condition* is the same as described above (i.e., it is the

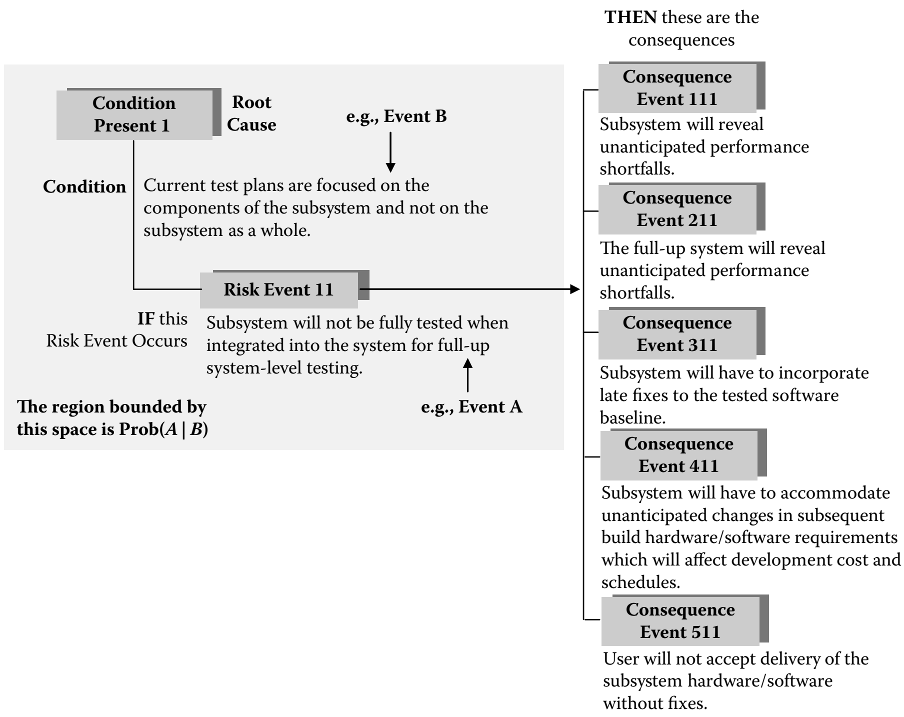


Figure 2.6: An Illustration of the *Condition-If-Then* construct [4].

root cause). The *If* is the associated risk event. The *Then* is the consequence, or set of consequences, that will impact the engineering system project if the risk event occurs. Figure 2.6 illustrates the *Condition-If-Then* construct for this example.

Questions and Exercises

1. State the interpretation of probability implied by the following:

   (a) The probability a tail appears on the toss of a fair coin is 1/2.
   (b) After recording the outcomes of 50 tosses of a fair coin, the probability a tail appears is 0.54.
   (c) It is with certainty the coin is fair.
   (d) The probability is 60% that the stock market will close 500 points above yesterday's closing count.
   (e) The design team believes there is less than a 5% chance the new microchip will require more than 12,000 gates.

2. A sack contains 20 marbles exactly alike in size but different in color. Suppose the sack contains 5 blue marbles, 3 green marbles, 7 red marbles, 2 yellow marbles, and 3 black marbles. Picking a single marble from the sack and then replacing it, what is the probability of choosing the following:

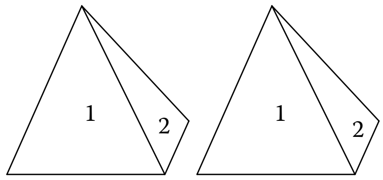
   (a) Blue marble? (b) Green marble? (c) Red marble?
   (d) Yellow marble? (e) Black marble? (f) Non-blue marble
   (g) Red or non-red marble?

3. If a fair coin is tossed, what is the probability of not obtaining a head? What is the probability of the event: (a head or not a head)?

4. Suppose A is an event (a subset) contained in the sample space Ω. Given this, are the following probability statements true or false, and why?

   (a) $P(A \cup A^c) = 1$ (b) $P(A\,|\Omega) = P(A)$

5. Suppose two tetrahedrons (4-sided polygons) are randomly tossed. Assuming the tetrahedrons are weighted fair, determine the set of all possible outcomes Ω. Assume each face is numbered 1, 2, 3, and 4.



Two tetrahedrons for Exercise 5.

Let the sets A, B, C, and D represent the following events

A: The sum of the toss is even

B: The sum of the toss is odd

C: The sum of the toss is a number less than 6

D: The toss yielded the same number on each upturned face

(a) Find $P(A)$, $P(B)$, $P(C)$, $P(A \cap B)$, $P(A \cup B)$, $P(B \cup C)$, and $P(B \cap C \cap D)$.

(b) Verify $P((A \cup B)^c) = P(A^c \cap B^c)$.

6. The XYZ Corporation has offers on two contracts A and B. Suppose the proposal team made the following subjective probability assessments: the chance of winning contract A is 40%, the chance of winning contract B is 20%, the chance of winning contract A or contract B is 60%, the chance of winning both contracts is 10%.

   (a) Explain why the above set of probability assignments is *inconsistent* with the axioms of probability.

   (b) What must $P(B)$ equal such that it and the set of other assigned probabilities specified above are consistent with the axioms of probability?

7. Suppose a coin is balanced such that tails appears three times more frequently than heads. Show the probability of obtaining a tail with such a coin is 3/4. What would you expect this probability to be if the coin was fair — that is, equally balanced?

8. Suppose the sample space of an experiment is given by $\Omega = A \cup B$. Compute $P(A \cap B)$ if $P(A) = 0.25$ and $P(B) = 0.80$.

9. If A and B are disjoint subsets of Ω show that

   (a) $P(A^c \cup B^c) = 1$

   (b) $P(A^c \cap B^c) = 1 - [P(A) + P(B)]$

10. Two missiles are launched. Suppose there is a 75% chance missile A hits the target and a 90% chance missile B hits the target. If the probability missile A hits the target is *independent* of the probability missile B hits the target, determine the probability missile A or missile B hits the target. Find the probability needed for missile A such that if the probability of missile B's hitting the target remains at 90%, the probability missile A or missile B hits the target is 0.99.

11. Suppose A and B are independent events. Show that
    (a) The events A^c and B^c are independent.
    (b) The events A and B^c are independent.
    (c) The events A^c and B are independent.

12. Suppose A and B are independent events with $P(A) = 0.25$ and $P(B) = 0.55$. Determine the probability
    (a) At least one event occurs.
    (b) Event B occurs but event A does not occur.

13. Suppose A and B are independent events with $P(A) = r$ and the probability that at least A or B occurs is s. Show the only value for $P(B)$ is the product $(s - r)(1 - r)^{-1}$.

14. At a local sweet shop, 10% of all customers buy ice cream, 2% buy fudge, and 1% buy both ice cream and fudge. If a customer selected at random bought fudge, what is the probability the customer bought an ice cream? If a customer selected at random bought ice cream, what is the probability the customer bought fudge?

15. For any two events A and B, show that $P(A \mid A \cap (A \cap B)) = 1$.

16. A production lot contains 1000 microchips of which 10% are defective. Two chips are successively drawn at random without replacement. Determine the probability
    (a) Both chips selected are non-defective.
    (b) Both chips are defective.
    (c) The first chip is defective and the second chip is non-defective.
    (d) The first chip is non-defective and the second chip is defective.

17. Suppose the sampling scheme in exercise 16 was with replacement, that is, the first chip is returned to the lot before the second chip is drawn. Show how the probabilities computed in exercise 16 change.

18. Spare power supply units for a communications terminal are provided to the government from three different suppliers A_1, A_2, and A_3. Suppose 30% come from A_1, 20% come from A_2, and 50% come from A_3. Suppose these units occasionally fail to perform according to their specifications and the following has been observed: 2% of those supplied by A_1 fail, 5% of those supplied by A_2 fail, and 3% of those supplied by A_3 fail. What is the probability any one of these units provided to the government will perform *without* failure?

19. In a single day, ChipyTech Corporation's manufacturing facility produces 10,000 microchips. Suppose machines A, B, and C individually produce 3000, 2500, and 4500 chips daily. The quality control group has determined the output from machine A has yielded 35 defective chips, the output from machine B has yielded 26 defective chips, and the output from machine C has yielded 47 defective chips.
    (a) If a chip was selected at random from the daily output, what is the probability it is defective?
    (b) What is the probability a randomly selected chip was produced by machine A? By machine B? By machine C?
    (c) Suppose a chip *was* randomly selected from the day's production of 10,000 microchips and it was found to be defective. What is the probability it was produced by machine A? By machine B? By machine C?

20. From section 2.4.1, show that Bayes' rule is the basis for the equations below.
    (a) $P(H\,|e_1) = \dfrac{P(H)P(e_1\,|H)}{P(H)P(e_1\,|H) + (1 - P(H))P(e_1\,|H^c)}$
    (b) $P(H\,|e_1e_2) = \dfrac{P(H\,|e_1)\;P(e_2\,|H)}{P(H\,|e_1)\;P(e_2\,|H) + (1 - P(H\,|e_1))P(e_2\,|H^c)}$
    (c) $P(H\,|e_1e_2e_3) = \dfrac{P(H\,|e_1e_2)\;P(e_3\,|H)}{P(H\,|e_1e_2)\;P(e_3\,|H) + (1 - P(H\,|e_1e_2)\,|e_1))P(e_3\,|H^c)}$