1.0 WHAT YOU NEED TO KNOW ABOUT PROBABILITY

Statistical inferences will often be expressed as probabilities. Probability can be defined as our degree of belief that an event will occur (e.g., I think I have an even chance of finishing this book on schedule). In statistical analysis, however, probability is usually defined as the expected frequency that an outcome will occur (e.g., nine times out of ten the actual value of a product's failure rate will be within a range calculated by a particular method). This frequency may be stated as a percentile (e.g., 90% of the time the failure rate will actually be within the calculated range) or its decimal equivalent (e.g., there is a probability of 0.90 that the failure rate is really in the calculated range). Probability may also be referred to as "confidence" (e.g., there is a 90% confidence that the failure rate is within the calculated range).

In dealing with probabilities, some useful relationships can be applied, depending on certain assumptions.

1 When Events are Independent

If we know the probabilities of two events happening, and can assume that the events are independent (i.e., the occurrence of one does not increase or decrease the probability that the other will occur), then the probability of both events happening is:

$$P(\text{a and b}) = P(\text{a})P(\text{b}) \tag{1-1}$$

where:

P(a and b)	=	probability of both event "a" and event "b" happening
P(a)	=	probability of event "a" happening
P(b)	=	probability of event "b" happening

For example, suppose an airplane uses a satellite receiver to track its position from the Global Positioning System (GPS) and can also track its position from a radio direction finder (RDF) receiver. If the GPS receiver fails once in 100 flights, the probability of losing the GPS tracking capability is 0.01 per flight. If the RDF receiver fails once every fifty flights, the probability of losing the RDF tracking capability is 0.02 per flight. If we can assume that failure of one does not affect the other (this is not a trivial assumption: both could fail simultaneously from some common cause, for example, lightning hitting the aircraft) then the probability of losing both position tracking systems on the same flight is:

$$P = (0.01)(0.02) = 0.0002 \tag{1-2}$$

Or, two times in 10,000 flights both position tracking systems will be out of service.

1.2 When Events are Mutually Exclusive

When probabilities are mutually exclusive (i.e., the occurrence of one event precludes the other), the probability of either of two events happening is:

$$P(a \text{ or } b) = P(a) + P(b) \quad (1\text{-}3)$$

A useful fact is that the sum of the probabilities of all possible outcomes of an event must equal unity. Further, the probability that an event will occur (P) plus the probability that it will not occur (Q) must also equal one, since there are no other possibilities. Thus:

$$P + Q = 1 \text{ or } P = (1 - Q) \text{ or } Q = (1 - P)$$

If the probability of a GPS receiver failure is 0.01, then the probability of no failure is (1 - 0.01) or 0.99.

Often, it is much easier to calculate one of the parameters (P or Q) than the other. Since P + Q = 1, one parameter can always be found from the other.

Continuing the position tracking example, and the assumption of independence, the probability of a flight without total loss of position tracking capability would be:

$$P(s) = P(g)P(r) + P(g)Q(r) + Q(g)P(r)$$

where:

P(s) = probability of success (no total loss of all position tracking systems)
P(g) = probability of no failure in GPS receiver = (1 - 0.01) = 0.99
Q(g) = probability of failure in GPS receiver = 0.01
P(r) = probability of no failure in RDF receiver = (1 - 0.02) = 0.98
Q(r) = probability of failure in RDF receiver = 0.02

Hence:

$$P(s) = (0.99)(0.98) + (0.99)(0.02) + (0.01)(0.98) = 0.9702 + 0.0198 + 0.0098 = 0.9998 \quad (1\text{-}6)$$

Each of these events is mutually exclusive and they constitute all the "successful" situations. The other possibility is Q(g)Q(r), the probability that both the GPS and RDF receivers will fail, which was computed by Equation 1-2 as equal to 0.0002. From Equation 1-4:

$$P(s) = 1.0 - Q(g)Q(r) = 1.0 - 0.0002 = 0.9998$$

Note: in this example, an event can be defined either as the occurrence of a failure or as the lack of a failure. Hence P(i) can be the probability of no failure or the probability of failure in the component identified as (i). By convention, P(i) is usually the probability of success (no failure), and Q(i) the probability of failure, when both notations are used in one formula. P(i) is generally used when only one notation is needed, whether it refers to a failure event or a non-failure event.

We are following this convention, even though this reverses the meaning of P(i) from the previous example. Q(g)Q(r) in Equation 1-7 is identical in meaning to P(a)P(b) in Equation 1-1.

Finally, consideration of mutually exclusive events leads to another solution for the case of independent events, shown in Equation 1-8.

$$P(s) = P(g) + P(r) - P(g)P(r) \tag{1-8}$$

where all terms are as defined for Equation 1-5.

The rationale for this is that P(g) includes all of the cases in which the GPS receiver is operating, including both the times that the RDF receiver is operating and the times that it has failed. Similarly, P(r) includes all of the cases in which the RDF receiver is operating, including both the times the GPS receiver is operating and the times it fails. Thus, P(g) + P(r) twice counts the times that both the GPS and RDF receivers are operating, and so these times must be subtracted to yield P(s). This is easily proven by decomposing P(g) and P(r) into mutually exclusive events:

$$P(g) = P(g)P(r) + P(g)Q(r) \tag{1-9}$$

$$P(r) = P(r)P(g) + P(r)Q(g) = P(g)P(r) + Q(g)P(r) \tag{1-10}$$

Substituting Equations 1-9 and 1-10 into Equation 1-8 we get:

$$\begin{aligned} P(s) &= P(g)P(r) + P(g)Q(r) + P(g)P(r) + Q(g)P(r) - P(g)P(r) \\ &= P(g)P(r) + P(g)Q(r) + Q(g)P(r) \end{aligned} \tag{1-11}$$

which is the same result as Equation 1-5

1.3 When Events are Not Independent

In the examples given in Sections 1.1 and 1.2, a failure of the GPS receiver does not change the probability that the RDF receiver will also fail. This is not always true. Suppose one-tenth of all failures in the GPS are due to external events, like lightning strikes, which also take out the RDF. Then, our calculation of the probability of both receivers failing becomes more complicated. First, we need a new term: P(b|a), defined as the conditional probability that event "b" will occur, given that event "a" has occurred. Then:

$$P(a \text{ and } b) = P(a)P(b|a) \tag{1-12}$$

where:

P(a and b)	=	the probability that both events "a" and "b" will occur
P(a)	=	the probability that event "a" will occur
P(b|a)	=	the probability that event "b" will occur, given that event "a" occurs

Since "a" and "b" are arbitrary labels, Equation 1-12 can also be written:

$$P(a \text{ and } b) = P(b)P(a|b) \tag{1-13}$$

If the events were independent, $P(b|a) = P(b)$ and $P(a|b) = P(a)$, and Equations 1-12 and 1-13 would be identical in form to Equation 1-1. Since the events are not independent, we must do a little more work.

If P(a and b) is the probability of both the GPS and RDF receivers failing, and P(a) is the probability of the GPS failing, then P(b|a) is the probability of the RDF failing on a flight when the GPS failed. We know that one-tenth (10% or 0.10) of all GPS failures are caused by factors that also kill the RDF (probability of RDF failure = 1.0). This means that for 90% of the GPS failures, any RDF failures must be from other causes, which have some probability of occurrence whether or not the GPS has failed. To determine this probability, we could search our records using just those flights when there was no failure of the GPS, thus eliminating any effects of GPS failure. Suppose we found that the failure rate for the RDF, using the censored data, was 19 failures in 1,000 flights = 0.019. Hence:

$$P(b|a) = 0.10(1) + 0.90(0.019) \tag{1-14}$$

Using Equations 1-12 and 1-14, the probability of both receivers failing is:

$$P(a \text{ and } b) = P(a)\,P(b|a) = P(a)[0.10\,(1) + 0.90(0.019)] = 0.01\,[0.10 + 0.90(0.019)]$$

$$= 0.001\ [0.\ + 0.017] = 0.0001 + 0.00017 = 0.00027 \tag{1-15}$$

This result is higher than the 0.0002 found by Equation 1-2, where we assumed independence, even though the RDF failure rate, exclusive of simultaneous failures, is higher in Equation 1-2 than it is in Equation 1-15. When some failure mechanisms take out both units simultaneously, the overall probability of both units failing must go up.

1.3.1 Bayes' Theorem

A noted derivation from conditional probabilities is Bayes' Theorem. For our discussion, let us consider a radar installed in an aircraft. Assume we have gathered some statistics as shown in Table 1-1.

Table 1-1: Known Data

Mission Profile	**Percent of Sorties Using Mission Profile**	**Probability of Radar Failure During Mission**
Combat	0.20	0.20
	0.20	0.10
	0.60	0.05

Letting event a_i represent the probability of a sortie using a specific mission profile, and event b represent a radar failure, we can convert the data in Table 1-1 to terms of probabilities and conditional probabilities. Table 1-2 shows the converted data.

Table 1-2: Converted Data

i	Mission Profile	$P(a_i)$	$P(b\|a_i)$
1	Combat	0.20	0.20
2	Training	0.20	0.10
3	Transport	0.60	0.05

Suppose we are given the information that an aircraft came back from a sortie with a failed radar. We can use the information to calculate the probability that the sortie was a combat mission. To do so, we must derive Bayes' Theorem.

Since $P(a \text{ and } b) = P(a|b)\,P(b) = P(b|a)\,P(a)$, it follows that:

$$P(a_i)\,P(b) = P(a_i|b)\,P(b) = P(b|a_i)\,P(a_i) \tag{1-16}$$

Therefore:

$$P(a_i|b) = P(b|a_i)\,P(a_i)/P(b) \tag{1-17}$$

Since $P(a_1)$ is the probability of a combat mission, the solution we seek is:

$$P(a_1|b) = P(b|a_1)\,P(a_1)/P(b) \tag{1-18}$$

Since event "a" is a set of mutually exclusive events, and there is a different conditional probability of event "b" happening for each event in set "a", the total probability of event "b" happening is:

$$P(b) = \Sigma\, P(b|a_i)\,P(a_i) \tag{1-19}$$

Substituting Equation 1-19 into Equation 1-18:

$$P(a_1 \mid b) = \frac{P(b \mid a_1)\,P(a_1)}{\sum P(b \mid a_i)\,P(a_i)} \tag{1-20}$$

In the equation, $P(a_1|b)$ is the probability of the returned airplane having flown a combat profile, given that it came back with a failed radar. The other terms are quantified in Table 1-2. Equation 1-20 is Bayes' Theorem.

Substituting the data from Table 1-2 into Equation 1-20:

$$P(a_1 \mid b) = \frac{(0.20)(0.20)}{(0.20)(0.20)+(0.20)(0.10)+(0.60)(0.05)} = \frac{0.04}{0.04+0.02+0.03}$$

$$\frac{0.04}{0.09} = 0.44 \tag{1-21}$$

Without the knowledge of the radar failure, we would have estimated the probability of the aircraft having just flown a combat profile at 0.20, based on the data in the second column of Table 1-1. The information that the radar failed raises our estimate to 0.44. The 0.20 figure is called the "prior" estimate, because it comes before the gathering of additional data (i.e., that the radar failed on the mission), and the 0.44 figure is called the "posterior" estimate, because it comes after the new data is considered.

Bayes' Theorem is the foundation of both useful and dubious analyses combining "prior" data with information from statistical sampling to produce a "posterior" estimate. The theorem is quite correct. Some applications that depend on "subjective priors" (i.e., the known information is an opinion or assumption rather than a conclusion from a set of statistics) can be questionable.

1.4 In Summary

Table 1-3 summarizes the material presented in this section.

Table 1-3: Summary of Section

When events are:	And the following apply:	You can use:
Independent	The occurrence of one event has no effect on the occurrence of the other	P(a and b) = P(a) P(b) P(a or b) = P(a) + P(b) - P(a and b)
Mutually exclusive	The occurrence of one event precludes the other	P(a or b) = P(a) + P(b)
	P(a) = probability event "a" occurs Q(a) = probability event "a" does not occur	P(a) + Q(a) = 1 P(a) = 1 - Q(a) Q(a) = 1 - P(a)
Not independent	The occurrence of one event may affect the other	P(a and b) = P(a) P(b|a) = P(b) P(a|b)
	One event may have several different outcomes, each affecting the other event differently	$P(a_1 \mid b) = \frac{P(b \mid a_1) P(a_1)}{\sum P(b \mid a_i) P(a_i)}$