1 Introduction

What is statistics?

Statistics is a science that quantifies the uncertainty inherent in conclusions drawn from less than complete information.

The mathematical theory of probability is the main tool used in quantifying uncertainty.

Examples of statistical problems:

- A prescribed amount of a hormone is administered to a mouse. Does this affect the expression of a particular gene in the mouse's genome?
- If 450 people out of 1000 in a survey say they want more gun control, what can we say about the percentage of *all* people who want more gun control?

In the last problem, here's an example of a conclusion stated in statistical terms:

One may be 95% confident that the percentage of all U.S. adults who favor more gun control is between 42% and 48%.

Two main components of the statistical paradigm:

- Population
- Sample

The *population* is a collection of numbers about which one wants to draw a conclusion or make an *inference*.

The *sample* is a subset of the population.

The problem of interest:

Draw a conclusion about the population based on information in a sample.

Typically, the population is so large that it is too time-consuming and/or expensive to determine every number in the population. So, we look at just a subset of the population, and usually a relatively small subset.

For example, there are more than 100 million adults in the US, but a survey may only consider 1000 of them to estimate the proportion having a given opinion.

$$\frac{1000}{100,000,000} \times 100\% = 0.001\%$$

Somewhat surprisingly, if it is obtained in a prescribed way, a sample containing less than one thousandth of a percent of the population can actually provide very accurate results about the entire population.

Conclusions (about a population) based on information in a sample are marked by uncertainty, at least when the sample is a *proper* subset of the population. Such conclusions are called **inductive**.

- *Induction* – Reasoning from specific to general
- *Deduction* – Reasoning from general to specific

In our statistical paradigm:

$$\begin{array}{rcl} \text{sample} & \Longleftrightarrow & \text{specific} \\ \text{population} & \Longleftrightarrow & \text{general} \end{array}$$

Probability is the tool used in quantifying the uncertainty in inductive statistical conclusions.

2 Experiments and Events

So, we now begin our study of probability. One can make the study of probability completely abstract, just as with any other mathematical discipline. Instead, I will try to use examples that show how probability is used in statistics.

Some necessary definitions

1. *Experiment* – some process having a number of possible outcomes, all of which are known, but whose ultimate result is not known.
2. *Sample space* – the set of all possible outcomes of the experiment. The sample space is denoted $\mathcal{S}$ and a single element of $\mathcal{S}$ is s.
3. *Event* – a subset of $\mathcal{S}$.

Example 1 Roll two dice. We assume the dice are distinguishable; one is red and the other green. An outcome is denoted (i, j), where i is the red outcome and j the green.

$$\mathcal{S} = \{(i, j) : i = 1, \ldots, 6; j = 1, \ldots, 6\}$$

Some events:

"Red die is 1 and green is 5"$= \{(1, 5)\}$

"A 1 and a 5"$= \{(1, 5), (5, 1)\}$

"Total is less than 5"$= \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1)\}$

The notation $A \subset B$ means that the event A is a subset of B.

The *empty set* is the set containing no elements, and is denoted $\emptyset$.

Finite and *infinite* sets are ones with a finite and infinite number of elements, respectively.

There are two kinds of infinite sets: *countable* and *uncountable*.

A *countably infinite set* is one whose elements can be put into 1 to 1 correspondence with the integers $1, 2, \ldots$.

A set is *uncountable* if it is neither finite nor countable. An example of an uncountable set is all the real numbers in the interval $(0, 1)$.

Example 2 *Experiment with an infinite but countable sample space*

The experiment is to keep tossing a coin until it comes up heads. The outcome of the experiment is the number of coin tosses needed to get the first head. The sample space is $\mathcal{S} = \{1, 2, 3, \ldots\}$.

Example 3 *Experiment with an uncountable sample space*

The experiment is to turn on an electrical device and observe how long it operates. The sample space is

$$\mathcal{S} = \{s : s \geq 0\},$$

where an individual outcome s is the length of time (in minutes) until the device quits operating. The length of time could be any nonnegative number.

3 Set Operations

- *Union* – The *union* of events A and B, denoted $A \cup B$, is an event containing all the elements that are in A but not B, B but not A, or both A and B.

- *Intersection* – The *intersection* of A and B, denoted $A \cap B$, is an event containing all elements that are in both A and B.

- *Complement* – The *complement* of an event A, denoted A^c, is the set of all elements in the sample space that are *not* in A.

- *Mutually exclusive* – Events A and B are said to be *mutually exclusive (or disjoint)* if they have no elements in common. In other words, $A \cap B = \emptyset$. Several sets, say $A_1, \ldots, A_k$, are mutually exclusive if A_i and A_j are disjoint for every $i \neq j$.

- *DeMorgan's Laws*

$$\text{(i) } A_1^c \cap \cdots \cap A_k^c = (A_1 \cup \cdots \cup A_k)^c$$
$$\text{(ii) } A_1^c \cup \cdots \cup A_k^c = (A_1 \cap \cdots \cap A_k)^c$$

4 Definition of Probability

We will use the notation $P(A)$ to denote "probability of event A."

Intuitive notions of probability:

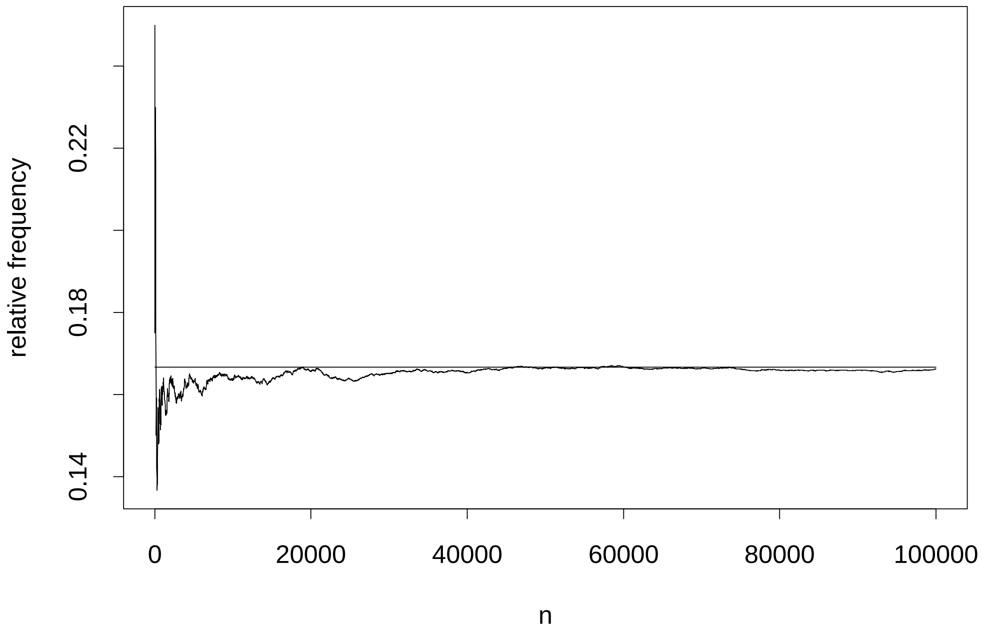
- $P(A)$ represents a degree of belief that event A will happen.
- $P(A)$ represents the proportion of the time A would occur if the experiment were repeated a large number of times.

Illustration of Long-Run Relative Frequency

Suppose a die is tossed repeatedly, and we count the number of times that the toss results in six spots. We then plot the proportion of times that the toss results in a six versus the number of tosses.

n	$n(A)$	$\frac{n(A)}{n}$
10	2	0.20000
100	23	0.23000
1000	160	0.16000
10000	1639	0.16390
100000	16618	0.16618

Relative Frequency of Tosses of Die Resulting in a Six



4.1 Axiomatic Definition of Probability

A probability measure on $\mathcal{S}$ is a function P from subsets of $\mathcal{S}$ to the real line satisfying the following:

Axiom 1: For every event A, $P(A) \geq 0$.

Axiom 2: $P(\mathcal{S}) = 1$.

Axiom 3: For every infinite sequence of disjoint events $A_1, A_2, \ldots$,

$$P(A_1 \cup A_2 \cup \cdots) = \sum_{i=1}^{\infty} P(A_i).$$

As usual with axioms, we take these to be true without proof. However, many interesting properties are logical consequences of just these three axioms. Probabilists are *still* discovering new consequences. We'll talk about a few of the simple ones.

4.2 Properties of Probability (provable using only the axioms)

1. $P(\emptyset) = 0$

2. For any *finite* sequence of disjoint events $A_1, \ldots, A_k$,

$$P(A_1 \cup \cdots \cup A_k) = \sum_{i=1}^{k} P(A_i).$$

3. For any event A, $P(A) = 1 - P(A^c)$.

Proof: For any event A, $A \cup A^c = \mathcal{S}$. Since A and A^c are disjoint, property 2 says that

$$P(A) + P(A^c) = P(\mathcal{S}).$$

Axiom 2 says that $P(\mathcal{S}) = 1$, and so

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

4. For any event A, $P(A) \leq 1$.

Proof: From Property 3,

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

Axiom 1 says that the probability of any event is nonnegative, implying that

$$P(A^c) \geq 0 \Longrightarrow 1 - P(A^c) \leq 1 \Longrightarrow P(A) \leq 1.$$

5. If $A \subset B$, then $P(A) \leq P(B)$.

6. For any two events A and B,

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

Proof: We may express the event $A \cup B$ as

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

implying that $P(A \cup B) = P(A) + P(B \cap A^c)$. (Why?) Also,

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

and so

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

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

The proof is done by substituting $P(B) - P(A \cap B)$ for $P(A^c \cap B)$ in the last expression for $P(A \cup B)$.

5 Finite Sample Spaces and Counting Rules

A finite sample space is one with a finite number of elements, i.e., outcomes. (By definition, these outcomes are mutually exclusive.) Let $s_1, \ldots, s_n$ denote the outcomes in a finite sample space $\mathcal{S}$. Then it must be true that

$$\sum_{i=1}^{n} P(\{s_i\}) = 1. \qquad \text{Why?}$$

Any nonempty event A may be expressed as

$$A = \{s_{i_1}, s_{i_2}, \ldots, s_{i_k}\},$$

implying that

$$P(A) = \sum_{j=1}^{k} P(\{s_{i_j}\}).$$

In words, *the probability of any event A is the sum of the probabilities of the outcomes it contains.*

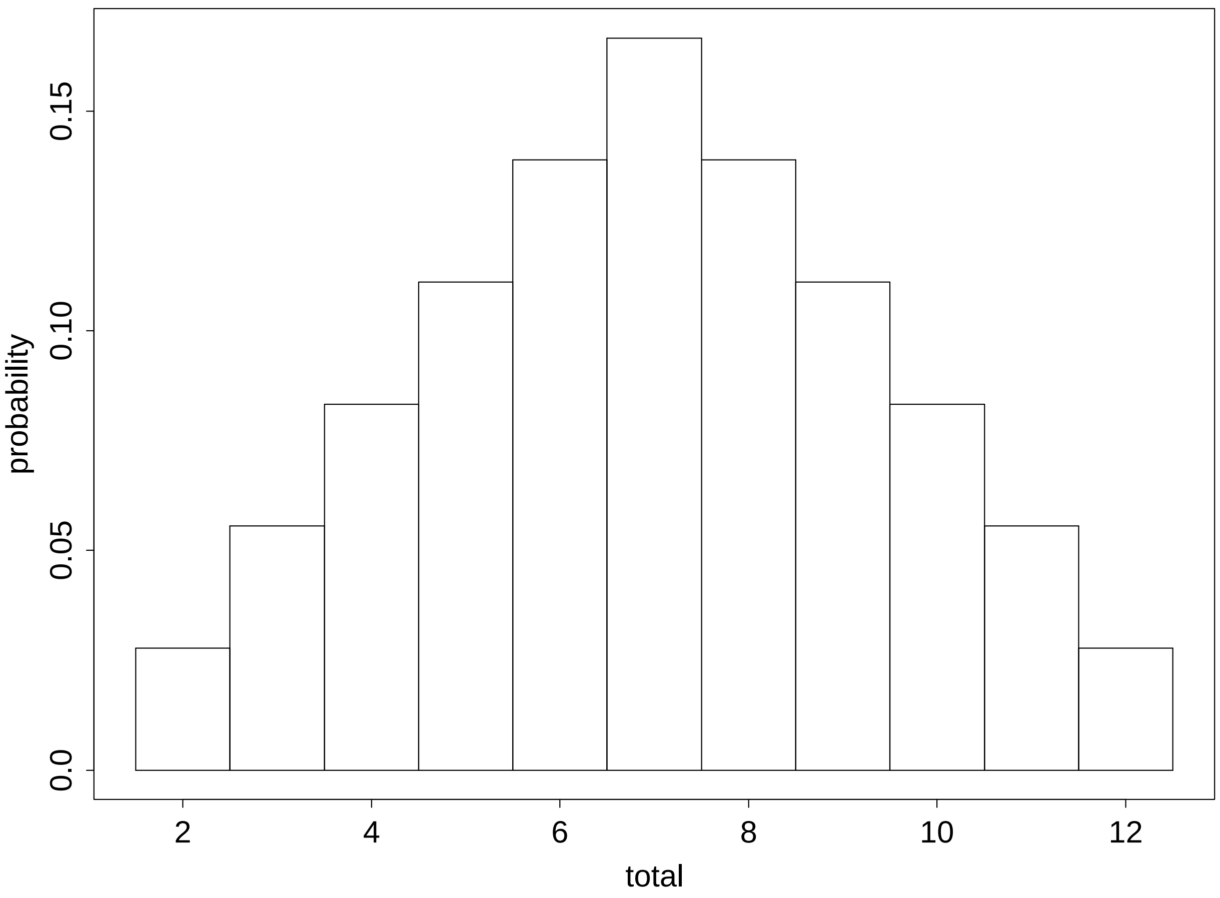
Example 4 Consider again our dice experiment. If the dice are balanced, then the probability of each of the 36 different outcomes is the same. In this case, for each (i, j)

$$P(\{(i, j)\}) = \frac{1}{36}. \qquad \text{Why?}$$

What's the probability of rolling a total of 7? To figure this out, we just need to count the number of outcomes that make up the event. The event of rolling 7 is

$$\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}.$$

Since this has six outcomes, the probability of 7 is $6/36 = 1/6$. We can likewise find the probability of any of the possible totals. A graph of the probabilities is given below.


0.0
0.05
0.10
0.15
probability
2
4
6
8
10
12
total

5.1 Counting Rules

For any finite sample space where all outcomes are equally likely, we find the probability of any event by counting the number of outcomes in that event, and then dividing this count by the total number of outcomes.

Sometimes the sample space has lots of outcomes (thousands or millions) and it becomes unwieldy (if not impossible) to list all the outcomes in an event. In such cases, *counting rules* come in handy.

5.2 Multiplication rule:

Suppose there are r different ways in which the first stage of an experiment can result. If, regardless of what happens at the first stage, there are s ways in which a second stage can result, then the total possible number of outcomes after stage 2 is $r \times s$.

Example 5 The experiment is to draw two cards from a standard deck of 52 cards. How many possible *ordered* hands of two cards are there? (By ordered we mean that $(2\diamondsuit, 5\spadesuit)$ is considered different from $(5\spadesuit, 2\diamondsuit)$.)

On the first draw, there are 52 possibilities. Regardless of which card is drawn, there are 51 possibilities for the second draw. So, the total number of hands is $52(51) = 2652$.

If we don't care about the order in which we get the cards, how many hands are there?

Answer: Just divide the first answer by 2, yielding $2652/2 = 1326$.

5.3 Permutations

Suppose we have n distinct objects and want to arrange them in a row. One possible rearrangement of the objects is known as a *permutation*. The number of possible permutations of n objects is $n! = n(n-1)(n-2)\cdots 2 \cdot 1$.

The previous result is proven using the multiplication rule.

How many ways can we permute the "objects" 1,2,3,4,5? The answer is $5! = 5 \cdot 4 \cdot 3 \cdot 2 = 120$. This illustrates the value of counting rules. It would be time consuming to start listing the possibilities:

$$
\begin{gathered}
1,\ 2,\ 3,\ 4,\ 5 \\
1,\ 2,\ 3,\ 5,\ 4 \\
1,\ 2,\ 5,\ 4,\ 3 \\
\vdots
\end{gathered}
$$

5.4 Sampling from a Finite Population

Consider a population consisting of N distinct elements. There are two ways to draw a sample from the population: with and without replacement.

Sampling with replacement. Suppose a sample of n elements is drawn sequentially in such a way that the element gotten on the $(k-1)$st draw is replaced before the kth element is drawn, $k = 2, \ldots, n$.

Number of possible samples: N^n Why?

Sampling without replacement. A sample of n elements ($n \leq N$) is drawn sequentially without replacing any elements along the way.

Number of possible samples:

$$P_{n,N} = N(N-1)(N-2)\cdots(N-n+1) = \frac{N!}{(N-n)!}$$

This is called the number of permutations of n objects selected out of N objects and is denoted $P_{n,N}$.

Example 6: What is the probability that at least two people in a group of size k have the same birthday?

Solution: Let us suppose that the 365 birthdays (excluding leap years) are equally likely. First we calculate the probability no two have the same birthday:

$$\frac{P_{k,365}}{365^k}$$

The probability at least 2 have the same birthday:

$$P(k) = 1 - \frac{P_{k,365}}{365^k}$$

k	$P(k)$	k	$P(k)$
10	.117	25	.569
15	.253	30	.706
20	.411	40	.891
21	.444	50	.970
22	.476	60	.994
23	.507		

5.5 Combinations–Unordered Samples

In many applications, the *order* in which population elements are drawn is irrelevant. The important thing is *which* elements are drawn. [Example: *card games.* What I care about is the hand of cards I get, and not the order in which they were dealt to me.]

Suppose we draw a sample of size n without replacement from a population of N distinct elements. How many *unordered* samples of size n are possible? This is equivalent to asking "*How many subsets of size n are there in a set of size N?*"

We know there are $N!/(N-n)!$ ordered samples. For any subset of n elements, there are $n!$ ways to permute the objects in that subset. So, the number of unordered samples must be

$$\frac{N!/(N-n)!}{n!} = \frac{N!}{n!(N-n)!} \stackrel{\text{def}}{=} \binom{N}{n}.$$

The number $\binom{N}{n}$ is called a *binomial coefficient*. For any real numbers x and y, the binomial theorem says that

$$(x+y)^N = \sum_{n=0}^{N} \binom{N}{n} x^n y^{N-n},$$

where $0!$ is defined to be 1.

Note that

$$\binom{N}{n} = \binom{N}{N-n}.$$

Why does this make sense?

We'll encounter binomial coefficients again when we discuss the *binomial distribution* and the *hypergeometric distribution*. These distributions occur when sampling from a dichotomous population.

Example 7: The Department of Statistics has 35 faculty members of whom 10 are assistant professors, 5 are associate professors, and 20 are full professors. A committee of 3 faculty members is chosen at random. Obtain the probability that all three members of the committee have the same rank and also the probability that all three members have different ranks.

We first obtain the number of three-person committees: $\binom{35}{3} = 6545$.

We next count the number of committees where all three individuals have the same rank: $\binom{10}{3} + \binom{5}{3} + \binom{20}{3} = 120 + 10 + 1140 = 1270$,

Thus, the probability that all three members have the same rank equals

$$\frac{\binom{10}{3} + \binom{5}{3} + \binom{20}{3}}{\binom{35}{3}} = \frac{1270}{6545} = 0.194.$$

We next count the number of committees where all three individuals have different ranks: $10 \times 5 \times 20 = 1000$. Thus, the probability that all three members have different ranks equals

$$\frac{10 \times 5 \times 20}{6545} = \frac{1000}{6545} = 0.153.$$

Extension to r Classes: The number of ways that n objects can be grouped into r classes with n_i in the i^{th} class is

$$\binom{n}{n_1 n_2 \cdots n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}$$

Example 7 continued:

Consider the 10 assistant professors. Suppose that you want to assign each of them to exactly one committee from committees A, B, and C. Suppose that committee A has 5 members, committee B has 3 members, and committee C has 2 members.

The number of ways that you can assign exactly 5 assistant professors to committee A, 3 to committee B, and 2 to committee C is

$$\binom{10}{5, 3, 2} = \frac{10!}{5!3!2!} = 2520.$$

6 Conditional Probability

START 9/6/21

Suppose our experiment has already been conducted. We don't know the specific outcome that occurred, but we do know that the outcome is in B.

How would we define the probability of an event A in this situation? We use the notation $P(A|B)$, which is read "probability of A given B."

Intuition based on relative frequency:

If an experiment is repeated many times, say n times, and an event A occurs $n(A)$ we approximate the probability of A by $n(A)/n$. $\approx P(A)$

How would we approximate $P(A|B)$? $= \dfrac{P(A \cap B)}{P(B)}$

$$P(A|B) \approx \frac{n(A \cap B)}{n(B)}$$

Example 8 In the 1950s a vaccine for polio was developed. Define the events:

$A =$ the event that a child gets polio, $B =$ the event that a child is vaccinated

The results of the trial were approximately as follows:

	Vaccinated		
	Yes	No	Total
Polio	33	110	143
No Polio	201,196	200,635	401,831
Total	201,229	200,745	401,974

If we select at child at random from the children involved in the trial, the probability that the child gets polio is

$$P(A) = \frac{143}{401974} = 0.00356.$$

We are more interested in comparing the probability that an unvaccinated child gets polio to the probability that a vaccinated child gets polio.

Suppose B occurs $n(B)$ times out of the n repetitions. Of the times B occurs, we could count the number of times A also occurs. Call this number $n(A \cap B)$. It would then make sense to approximate the conditional probability of A given B, $P(A|B)$, by

$$\frac{n(A \cap B)}{n(B)} = \frac{n(A \cap B)/n}{n(B)/n} \approx \underbrace{\frac{P(A \cap B)}{P(B)} = P(A|B)}$$

$$* \; P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

Example 8

The probability of polio for vaccinated children is

$$P(A|B) = \frac{n(A \cap B)}{n(B)} = \frac{33}{201229} = \frac{\frac{33}{401974}}{\frac{201229}{401974}} = 0.000164.$$

The probability of polio for unvaccinated children is

$$P(A|B^c) = \frac{n(A \cap B^c)}{n(B^c)} = \frac{110}{200745} = \frac{\frac{110}{401974}}{\frac{200745}{401974}} = 0.000548.$$

6.1 Definition of Conditional Probability

We will *define* the conditional probability of A given B to be as in the intuitive motivation.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{where we assume } P(B) > 0.$$

Remark: Suppose $P(B) > 0$. Then the definition of conditional probability satisfies the axioms of probability:

1. $P(A|B) \geq 0$ for any event A
2. $P(\mathcal{S}|B) = 1$ is this the LOTP?
3. For every infinite sequence of disjoint events $A_1, A_2, \ldots$,

$$P(A_1 \cup A_2 \cup \cdots |B) = \sum_{i=1}^{\infty} P(A_i|B).$$

Example 9 Suppose $P(A) = 0.5$, $P(B) = 0.6$ and $P(A \cap B) = 0.2$. Find $P(A|B)$ and $P(A|B^c)$.

note: $P(A|B) + P(A|B^c) \neq 1$

$P(A|B) + P(A^c|B) = 1$

Solution:

$$P(A|B) = \frac{0.2}{0.6} = \frac{1}{3}$$

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

$$\begin{aligned} P(A|B^c) &= \frac{P(A \cap B^c)}{P(B^c)} \\ &= \frac{P(A) - P(A \cap B)}{1 - P(B)} \\ &= \frac{0.5 - 0.2}{1 - 0.6} = 0.75 \end{aligned}$$

Multiplication rule

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

note: still valid if P(B) = 0.

This rule provides an alternative way of determining $P(A \cap B)$. It's especially useful in cases where an experiment proceeds in stages.

Suppose event B is determined by a first stage and A by a second stage. It often is easier to figure out $P(B)$ and $P(A|B)$ than figuring out $P(A \cap B)$ directly. The following problem provides such an example.

Example 10   Suppose we randomly select a sample of size n without replacement from a population of size N. What's the probability of selecting element # 4 on the first draw and element # 10 on the second?

Let B be the event of "4" on the first draw and A the event of "10" on the second draw. We could use counting techniques to find $P(A \cap B)$.

Obviously, though,

$$P(B) = \frac{1}{N} \quad \text{and} \quad P(A|B) = \frac{1}{N-1},$$

and hence

$$P(A \cap B) = \frac{1}{N} \cdot \frac{1}{N-1} = \frac{1}{N(N-1)}.$$

More general multiplication rule

As a shorthand we'll write

$$P(A_0 \cap A_1 \cap \cdots \cap A_n) = P(A_0 A_1 \cdots A_n).$$

Let $A_0, A_1, \ldots, A_n$ be $n+1$ events for which $P(A_0 A_1 \cdots A_n) > 0$. Then

$$\begin{aligned} P(A_0 A_1 \cdots A_n) &= P(A_0) P(A_1 | A_0) \\ &\quad \times P(A_2 | A_0 A_1) \cdots \\ &\quad \times P(A_n | A_0 A_1 \cdots A_{n-1}). \end{aligned}$$

A classic example of using multiplication rule to simplify probability calculations is the *birthday problem*.

Example 7 *Birthday problem* Given a group of n (randomly selected) persons, what is the probability that at least two people have the same birthday (i.e., same month and day)?

Let A be the event of interest. We'll find $P(A^c)$, where A^c is event that all n people have different birthdays. Line the people up in a row and go down the line determining birthdays.

A_i is the event that first i people have distinct birthdays, $i = 1, \ldots, n$.

$$A^c = A_n = A_1 \cap A_2 \cap \cdots \cap A_n$$

$$\begin{aligned} P(A^c) &= P(A_1 A_2 \cdots A_n) \\ &= P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots \\ &\quad \times P(A_n|A_1 \cdots A_{n-1}) \\ &= 1\left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\cdots\left(\frac{365-(n-1)}{365}\right) = \frac{P_{n,365}}{365^n} \end{aligned}$$

6.2 Bayes Theorem

Bayes theorem provides a method of reversing the order of conditioning in conditional probabilities.

Suppose we have events $B_1, B_2, \ldots$ that are mutually exclusive and exhaustive. By exhaustive, we mean that

$$\bigcup_{i=1}^{\infty} B_i = \mathcal{S}.$$

In other words, the union of the B_i's exhausts all the possible outcomes of the experiment.

We call $\{B_i\}$ a *partition* of the sample space.

Law of total probability: For any event A,

$$P(A) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i).$$

Proof: Obviously $A = A \cap \mathcal{S}$, implying that

$$A = A \cap \left(\bigcup_{i=1}^{\infty} B_i \right) = \bigcup_{i=1}^{\infty} (A \cap B_i).$$

Since the events $A \cap B_1, A \cap B_2, \ldots$ are mutually exclusive,

$$\begin{aligned} P(A) &= \sum_{i=1}^{\infty} P(A \cap B_i) \\ &= \sum_{i=1}^{\infty} P(A|B_i)P(B_i). \end{aligned}$$

Bayes Theorem

Suppose that A and B are events with $P(A) > 0$ and $P(B) > 0$. Then

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

Remark: We use Bayes rule to compute $P(A|B)$ when $P(A),\ P(B|A)$, and $P(B|A^c)$ are known. We can extend this to situation where we know the conditional probabilities of B given each member A_i of a partition of $\mathcal{S}$.

Suppose $A_1, A_2, \ldots$ are mutually exclusive and exhaustive (just like in the law of total probability), and that $P(A_i) > 0$ for each i. Then for any event B with $P(B) > 0$ and for each k,

$$P(A_k|B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{\infty} P(B|A_i)P(A_i)}.$$

Example 11 A company that manufactures digital camcorders produces a basic model and a deluxe model. Over the past year, 40% of the camcorders sold have been of the basic model. Of those buying the basic model, 30% purchase an extended warranty, whereas 50% of all deluxe purchasers do so. If you learn that a randomly selected purchaser has an extended warranty, how likely is it that he/she has a basic model?

$$B = \text{Basic} \qquad W = \text{Warranty}$$

$$P(W) = P(B \cap W) + P(B^c \cap W) = 0.12 + 0.30 = 0.42$$

$$P(B|W) = \frac{P(B \cap W)}{P(W)} = \frac{0.12}{0.42} = 0.2858$$

Example 12 *Paradox of the lie detector*

Suppose that 6% of a lawyer's clients lie when asked their names in a polygraph test. Among people who lie, the polygraph so indicates 95% of the time. Among people who don't lie, the polygraph says they are lying 5% of the time. If the polygraph indicates that a person is lying, what is the probability that the person really is lying?

First step: Define events!

$B_1 =$ event that a person is lying
$B_2 = B_1^c$
$A =$ event that polygraph indicates person is lying

Second step: Write down probabilities given in problem statement.

$$P(B_1) = 0.06 \qquad P(B_2) = 0.94$$

$$P(A|B_1) = 0.95 \qquad P(A|B_2) = 0.05$$

Third step: Apply Bayes rule to find the probability asked for.

$$P(B_1|A) = \frac{(0.95)(0.06)}{(0.95)(0.06) + (0.05)(0.94)} = 0.548$$

This answer is a little surprising. In spite of the apparent accuracy of the polygraph, in cases where the polygraph says someone is lying, there really isn't a strong likelihood that the person really *is* lying.

What about $P(B_2|A^c)$?

$$\begin{aligned} P(B_2|A^c) &= \frac{P(A^c|B_2)P(B_2)}{\sum_{i=1}^{2} P(A^c|B_i)P(B_i)} \\ &= \frac{(0.95)(0.94)}{(0.05)(0.06) + (0.95)(0.94)} = 0.997 \end{aligned}$$

So, the polygraph is quite reliable in cases where it says someone is telling the truth.

7 Independent Events

We say two events A and B are *independent* if

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

Suppose $P(B) > 0$. Then the condition for independence is equivalent to

$$P(A|B) = P(A).$$

Likewise, if $P(A) > 0$, the independence condition is equivalent to

$$P(B|A) = P(B).$$

So, whenever the conditional probabilities are defined, A and B are independent if and only if a conditional probability equals its corresponding unconditional probability.

In a certain sense, independence means that A doesn't affect B and vice versa.

Events $A_1, \ldots, A_k$ are said to be *mutually independent* (or simply independent) if for every subset $\{i_1, \ldots, i_m\}$ of $\{1, \ldots, k\}$,

$$P\left(\bigcap_{j=1}^{m} A_{i_j}\right) = \prod_{j=1}^{m} P(A_{i_j})$$

Consider three events A, B and C. These events are mutually independent iff

$$P(A \cap B) = P(A)P(B), \quad P(A \cap C) = P(A)P(C),$$
$$P(B \cap C) = P(B)P(C) \quad \text{and} \quad P(A \cap B \cap C) = P(A)P(B)P(C).$$

Events $A_1, \ldots, A_k$ are said to be *pairwise independent* if

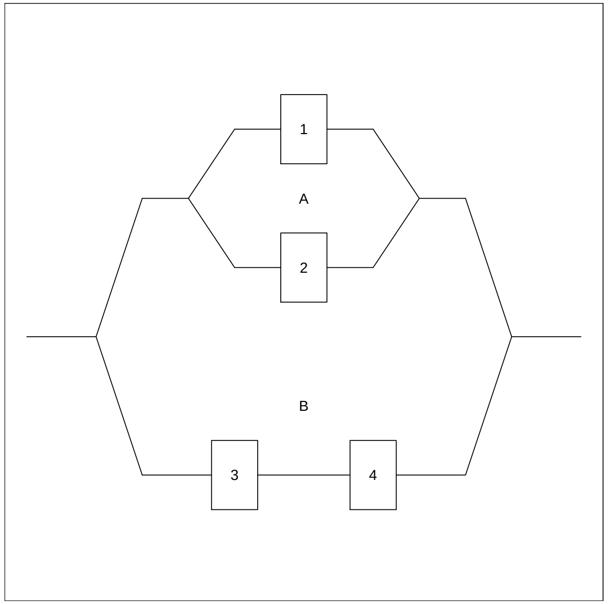
$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

for every pair of distinct events A_i and A_j.

Note: Independence $\Rightarrow$ pairwise independence, but events can be pairwise independent without being independent.

Application: Often engineering systems can be represented as comprised of independent components connected in parallel or in series.

Consider the system of components in the accompanying figure. Components 1 and 2 are connected in parallel (the subsystem works if either component works). Components 3 and 4 are connected in series (the subsystem works if both components work). The two subsystems are connected in parallel.



Let $C_i =$ the event that component i works. Then

$$\begin{aligned} P[\text{System A Works}] &= P[C_1 \cup C_2] = 1 - P[C_1^c \cap C_2^c] \\ &= 1 - P[C_1^c]P[C_2^c] = 1 - (1 - P(C_1))(1 - P(C_2)) \\ P[\text{System B Works}] &= P[C_3 \cap C_4] = P[C_3] \times P[C_4] \end{aligned}$$

