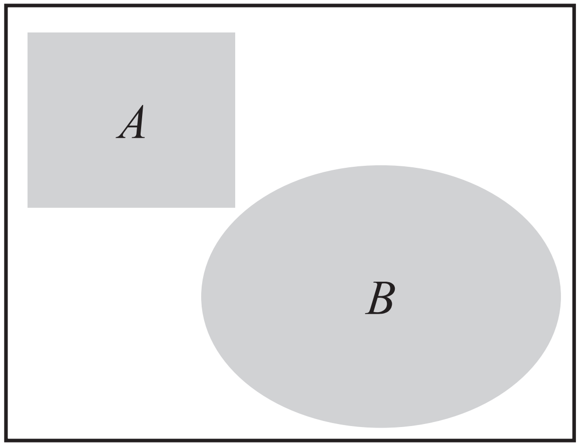
Section 1.1

Applying Set Theory to Probability

1.1 Comment: Mutually Exclusive Sets

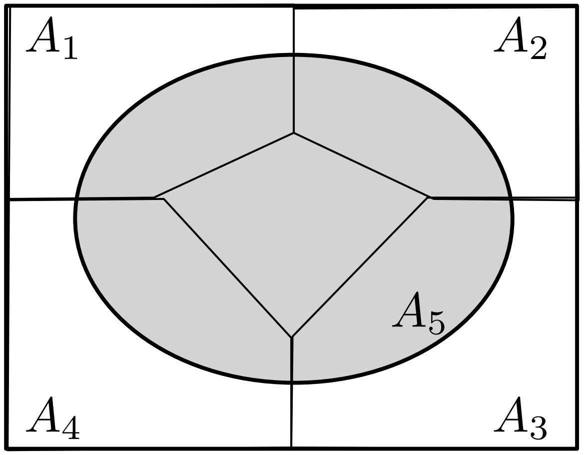


A collection of sets $A_1, \ldots, A_n$ is *mutually exclusive* if and only if

$$A_i \cap A_j = \varnothing, \qquad i \neq j. \tag{1.1}$$

The word *disjoint* is sometimes used as a synonym for mutually exclusive.

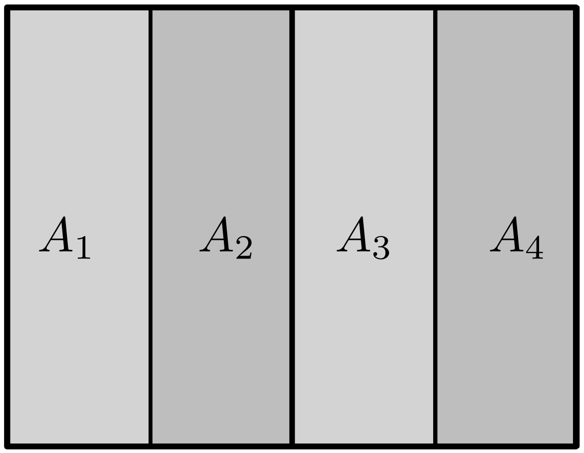
1.1 Comment: Collectively Exhaustive Sets



A collection of sets $A_1, \ldots, A_n$ issets,collectively exhaustive *collectively exhaustive* if and only if

$$A_1 \cup A_2 \cup \cdots \cup A_n = S. \tag{1.2}$$

1.1 Comment: Partitions



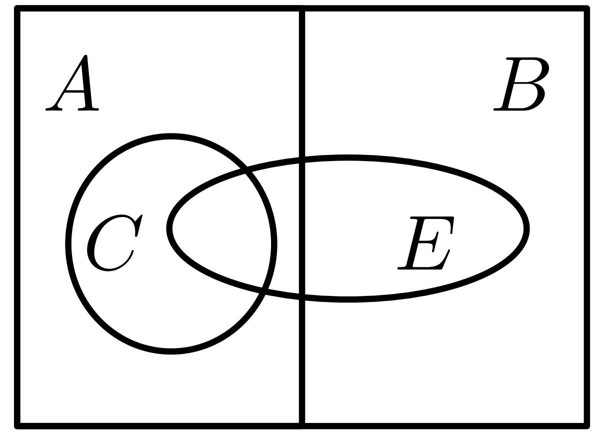
A collection of sets $A_1, \ldots, A_n$ is a *partition* if it is both mutually exclusive and collectively exhaustive.

Example 1.1 Problem

Phonesmart offers customers two kinds of smart phones, Apricot (A) and Banana (B). It is possible to buy a Banana phone with an optional external battery E. Apricot customers can buy a phone with an external battery (E) or an extra memory card (C) or both. Draw a Venn diagram that shows the relationship among the items A,B,C and E available to Phonesmart customers.

Example 1.1 Solution

Since each phone is either Apricot or Banana, A and B form a partition. Since the external battery E is available for both kinds of phones, E intersects both A and B. However, since the memory card C is available only to Apricot customers, $C \subset A$. A Venn diagram representing these facts is shown on the right.



1.1 Comment: Experiments

An experiment consists of a *procedure* and *observations*. There is uncertainty in what will be observed; otherwise, performing the experiment would be unnecessary. Some examples of experiments include

1. Flip a coin. Did it land with heads or tails facing up?

2. Walk to a bus stop. How long do you wait for the arrival of a bus?

3. Give a lecture. How many students are seated in the fourth row?

4. Transmit one of a collection of waveforms over a channel. What waveform arrives at the receiver?

5. Transmit one of a collection of waveforms over a channel. Which waveform does the receiver identify as the transmitted waveform?

Example 1.2

An experiment consists of the following procedure, observation, and model:

- Procedure: Monitor activity at a Phonesmart store.
- Observation: Observe which type of phone (Apricot or Banana) the next customer purchases.
- Model: Apricots and Bananas are equally likely. The result of each purchase is unrelated to the results of previous purchases.

Example 1.3

Monitor the Phonesmart store until three customers purchase phones. Observe the sequence of Apricots and Bananas.

Example 1.4

Monitor the Phonesmart store until three customers purchase phones. Observe the number of Apricots.

Definition 1.1 Outcome

An outcome *of an experiment is any possible observation of that experiment.*

Definition 1.2 Sample Space

The sample space *of an experiment is the finest-grain, mutually exclusive, collectively exhaustive set of all possible outcomes.*

Example 1.5

- The sample space in Example 1.2 is $S = \{a, b\}$ where a is the outcome "Apricot sold," and b is the outcome "Banana sold."

- The sample space in Example 1.3 is

$$S = \{aaa, aab, aba, abb, baa, bab, bba, bbb\} \tag{1.5}$$

- The sample space in Example 1.4 is $S = \{0, 1, 2, 3\}$.

Definition 1.3 Event

An event *is a set of outcomes of an experiment.*

Example 1.6

Observe the number of minutes a customer spends in the Phonesmart store. An outcome T is a nonnegative real number. The sample space is $S = \{T|T \geq 0\}$. The event "the customer stays longer than five minutes" is $\{T|T > 5\}$.

Example 1.7

Monitor three customers in the Phonesmart store. Classify the behavior as buying (b) if a customer purchases a smartphone. Otherwise the behavior is no purchase (n). An outcome of the experiment is a sequence of three customer decisions. We can denote each outcome by a three-letter word such as bnb indicating that the first and third customers buy a phone and the second customer does not. We denote the event that customer i buys a phone by B_i and the event customer i does not buy a phone by N_i. The event $B_2 = \{nbn, nbb, bbn, bbb\}$. We can also express an outcome as an intersection of events B_i and N_j. For example the outcome $bnb = B_1 N_2 B_3$.

Quiz 1.1

Monitor three consecutive packets going through a Internet router. Based on the packet header, each packet can be classified as either video (v) if it was sent from a Youtube server or as ordinary data (d). Your observation is a sequence of three letters (each letter is either v or d). For example, two video packets followed by one data packet corresponds to vvd. Write the elements of the following sets:

$$A_1 = \{\text{second packet is video}\}, \quad B_1 = \{\text{second packet is data}\},$$
$$A_2 = \{\text{all packets are the same}\}, \quad B_2 = \{\text{video and data alternate}\},$$
$$A_3 = \{\text{one or more video packets}\}, \quad B_3 = \{\text{two or more data packets}\}.$$

For each pair of events A_1 and B_1, A_2 and B_2, and so on, identify whether the pair of events is either mutually exclusive or collectively exhaustive or both.

Quiz 1.1 Solution

$$
\begin{aligned}
A_1 &= \{vvv, vvd, dvv, dvd\} \\
B_1 &= \{vdv, vdd, ddv, ddd\} \\
A_2 &= \{vvv, ddd\} \\
B_2 &= \{vdv, dvd\} \\
A_3 &= \{vvv, vvd, vdv, dvv, vdd, dvd, ddv\} \\
B_3 &= \{ddd, ddv, dvd, vdd\}
\end{aligned}
$$

Recall that A_i and B_i are collectively exhaustive if $A_i \cup B_i = S$. Also, A_i and B_i are mutually exclusive if $A_i \cap B_i = \phi$. Since we have written down each pair A_i and B_i above, we can simply check for these properties.

The pair A_1 and B_1 are mutually exclusive and collectively exhaustive. The pair A_2 and B_2 are mutually exclusive but *not* collectively exhaustive. The pair A_3 and B_3 are not mutually exclusive since dvd belongs to A_3 and B_3. However, A_3 and B_3 are collectively exhaustive.

Section 1.2

Probability Axioms

Definition 1.4 Axioms of Probability

A probability measure $\mathsf{P}[\cdot]$ *is a function that maps events in the sample space to real numbers such that*

Axiom 1 *For any event* A, $\mathsf{P}[A] \geq 0$.

Axiom 2 $\mathsf{P}[S] = 1$.

Axiom 3 *For any countable collection* $A_1, A_2, \ldots$ *of mutually exclusive events*

$$\mathsf{P}[A_1 \cup A_2 \cup \cdots] = \mathsf{P}[A_1] + \mathsf{P}[A_2] + \cdots .$$

Theorem 1.1

For mutually exclusive events A_1 and A_2,

$$P[A_1 \cup A_2] = P[A_1] + P[A_2].$$

Theorem 1.2

If $A = A_1 \cup A_2 \cup \cdots \cup A_m$ and $A_i \cap A_j = \varnothing$ for $i \neq j$, then

$$\mathsf{P}[A] = \sum_{i=1}^{m} \mathsf{P}[A_i].$$

Theorem 1.3

The probability measure $\mathsf{P}[\cdot]$ satisfies

(a) $\mathsf{P}[\varnothing] = 0$.

(b) $\mathsf{P}[A^c] = 1 - \mathsf{P}[A]$.

(c) For any A and B (not necessarily mutually exclusive),

$$\mathsf{P}[A \cup B] = \mathsf{P}[A] + \mathsf{P}[B] - \mathsf{P}[A \cap B].$$

(d) If $A \subset B$, then $\mathsf{P}[A] \leq \mathsf{P}[B]$.

Theorem 1.4

The probability of an event $B = \{s_1, s_2, \ldots, s_m\}$ is the sum of the probabilities of the outcomes contained in the event:

$$\mathsf{P}[B] = \sum_{i=1}^{m} \mathsf{P}[\{s_i\}].$$

Proof: Theorem 1.4

Each outcome s_i is an event (a set) with the single element s_i. Since outcomes by definition are mutually exclusive, B can be expressed as the union of m mutually exclusive sets:

$$B = \{s_1\} \cup \{s_2\} \cup \cdots \cup \{s_m\} \tag{1.6}$$

with $\{s_i\} \cap \{s_j\} = \varnothing$ for $i \neq j$. Applying Theorem 1.2 with $B_i = \{s_i\}$ yields

$$\mathrm{P}[B] = \sum_{i=1}^{m} \mathrm{P}[\{s_i\}]. \tag{1.7}$$

Theorem 1.5

For an experiment with sample space $S = \{s_1, \ldots, s_n\}$ in which each outcome s_i is equally likely,

$$\mathrm{P}[s_i] = 1/n \qquad 1 \leq i \leq n.$$

Proof: Theorem 1.5

Since all outcomes have equal probability, there exists p such that $P[s_i] = p$ for $i = 1, \ldots, n$. Theorem 1.4 implies

$$P[S] = P[s_1] + \cdots + P[s_n] = np. \tag{1.8}$$

Since Axiom 2 says $P[S] = 1$, $p = 1/n$.

Example 1.8 Problem

Roll a six-sided die in which all faces are equally likely. What is the probability of each outcome? Find the probabilities of the events: "Roll 4 or higher," "Roll an even number," and "Roll the square of an integer."

Example 1.8 Solution

The probability of each outcome is $\text{P}[i] = 1/6$ for $i = 1, 2, \ldots, 6$. The probabilities of the three events are

- $\text{P}[\text{Roll 4 or higher}] = \text{P}[4] + \text{P}[5] + \text{P}[6] = 1/2$.
- $\text{P}[\text{Roll an even number}] = \text{P}[2] + \text{P}[4] + \text{P}[6] = 1/2$.
- $\text{P}[\text{Roll the square of an integer}] = \text{P}[1] + \text{P}[4] = 1/3$.

Quiz 1.2

A student's test score T is an integer between 0 and 100 corresponding to the experimental outcomes $s_0, \ldots, s_{100}$. A score of 90 to 100 is an A, 80 to 89 is a B, 70 to 79 is a C, 60 to 69 is a D, and below 60 is a failing grade of F. If all scores between 51 and 100 are equally likely and a score of 50 or less never occurs, find the following probabilities:

(a) $P[\{s_{100}\}]$

(b) $P[A]$

(c) $P[F]$

(d) $P[T < 90]$

(e) P[a C grade or better]

(f) P[student passes]

Quiz 1.2 Solution

There are exactly 50 equally likely outcomes: s_{51} through s_{100}. Each of these outcomes has probability 1/50. It follows that

(a) $\text{P}[\{s_{100}\}] = 1/50 = 0.02$.

(b) $\text{P}[A] = \text{P}[\{s_{90}, s_{91}, \ldots, s_{100}\}] = 11/50 = 0.22$.

(c) $\text{P}[F] = \text{P}[\{s_{51}, \ldots, s_{59}\}] = 9/50 = 0.18$.

(d) $\text{P}[T < 90] = \text{P}[\{s_{51}, \ldots, s_{89}\}] = 39/50. = 0.78$.

(e) $\text{P}[C \text{ or better}] = \text{P}[\{s_{70}, \ldots, s_{100}\}] = 31 \times 0.02 = 0.62$.

(f) $\text{P}[\text{student passes}] = \text{P}[\{s_{60}, \ldots, s_{100}\}] = 41 \times 0.02 = 0.82$.

Section 1.3

Conditional Probability

Example 1.9

Consider an experiment that consists of testing two integrated circuits (IC chips) that come from the same silicon wafer and observing in each case whether a chip is accepted (a) or rejected (r). The sample space of the experiment is $S = \{rr, ra, ar, aa\}$. Let B denote the event that the first chip tested is rejected. Mathematically, $B = \{rr, ra\}$. Similarly, let $A = \{rr, ar\}$ denote the event that the second chip is a failure.

The chips come from a high-quality production line. Therefore the prior probability P$[A]$ is very low. In advance, we are pretty certain that the second circuit will be accepted. However, some wafers become contaminated by dust, and these wafers have a high proportion of defective chips. When the first chip is a reject, the outcome of the experiment is in event B and P$[A|B]$, the probability that the second chip will also be rejected, is higher than the *a priori* probability P$[A]$ because of the likelihood that dust contaminated the entire wafer.

Definition 1.5 Conditional Probability

The conditional probability of the event A given the occurrence of the event B is

$$\mathsf{P}[A|B] = \frac{\mathsf{P}[AB]}{\mathsf{P}[B]}.$$

Theorem 1.6

A conditional probability measure $P[A|B]$ has the following properties that correspond to the axioms of probability.

Axiom 1: $P[A|B] \geq 0$.

Axiom 2: $P[B|B] = 1$.

Axiom 3: If $A = A_1 \cup A_2 \cup \cdots$ with $A_i \cap A_j = \emptyset$ for $i \neq j$, then

$$P[A|B] = P[A_1|B] + P[A_2|B] + \cdots$$

Example 1.10 Problem

With respect to Example 1.9, consider the *a priori* probability model

$$\mathrm{P}[rr] = 0.01, \quad \mathrm{P}[ra] = 0.01, \quad \mathrm{P}[ar] = 0.01, \quad \mathrm{P}[aa] = 0.97. \qquad (1.9)$$

Find the probability of $A =$ “second chip rejected” and $B =$ “first chip reject
Also find the conditional probability that the second chip is a reject given that the first chip is a reject.

Example 1.10 Solution

We saw in Example 1.9 that A is the union of two mutually exclusive events (outcomes) rr and ar. Therefore, the a priori probability that the second chip is rejected is

$$\mathrm{P}[A] = \mathrm{P}[rr] + \mathrm{P}[ar] = 0.02 \tag{1.10}$$

This is also the a priori probability that the first chip is rejected:

$$\mathrm{P}[B] = \mathrm{P}[rr] + \mathrm{P}[ra] = 0.02. \tag{1.11}$$

The conditional probability of the second chip being rejected given that the first chip is rejected is, by definition, the ratio of $\mathrm{P}[AB]$ to $\mathrm{P}[B]$, where, in this example,

$$\mathrm{P}[AB] = \mathrm{P}[\text{both rejected}] = \mathrm{P}[rr] = 0.01 \tag{1.12}$$

Thus

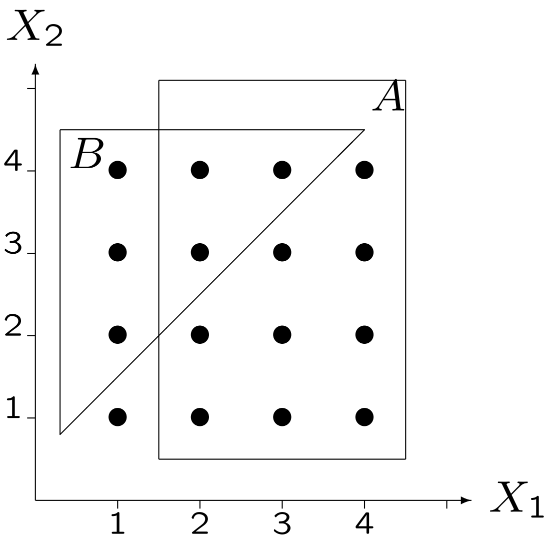
$$\mathrm{P}[A|B] = \frac{\mathrm{P}[AB]}{\mathrm{P}[B]} = 0.01/0.02 = 0.5. \tag{1.13}$$

The information that the first chip is a reject drastically changes our state of knowledge about the second chip. We started with near certainty, $\mathrm{P}[A] = 0.02$, that the second chip would not fail and ended with complete uncertainty about the quality of the second chip, $\mathrm{P}[A|B] = 0.5$.

Example 1.11 Problem

Roll two fair four-sided dice. Let X_1 and X_2 denote the number of dots that appear on die 1 and die 2, respectively. Let A be the event $X_1 \geq 2$. What is P$[A]$? Let B denote the event $X_2 > X_1$. What is P$[B]$? What is P$[A|B]$?

Example 1.11 Solution



We begin by observing that the sample space has 16 elements corresponding to the four possible values of X_1 and the same four values of X_2. Since the dice are fair, the outcomes are equally likely, each with probability 1/16. We draw the sample space as a set of black circles in a two-dimensional diagram, in which the axes represent the events X_1 and X_2. Each outcome is a pair of values (X_1, X_2). The rectangle represents A. It contains 12 outcomes, each with probability 1/16.

To find $\text{P}[A]$, we add up the probabilities of outcomes in A, so $\text{P}[A] = 12/16 = 3/4$. The triangle represents B. It contains six outcomes. Therefore $\text{P}[B] = 6/16 = 3/8$. The event AB has three outcomes, $(2, 3), (2, 4), (3, 4)$, so $\text{P}[AB] = 3/16$. From the definition of conditional probability, we write

$$\text{P}[A|B] = \frac{\text{P}[AB]}{\text{P}[B]} = \frac{1}{2}. \tag{1.14}$$

We can also derive this fact from the diagram by restricting our attention to the six outcomes in B (the conditioning event) and noting that three of the six outcomes in B (one-half of the total) are also in A.

Quiz 1.3

Monitor three consecutive packets going through an Internet router. Classify each one as either video (v) or data (d). Your observation is a sequence of three letters (each one is either v or d). For example, three video packets corresponds to vvv. The outcomes vvv and ddd each have probability 0.2 whereas each of the other outcomes vvd, vdv, vdd, dvv, dvd, and ddv has probability 0.1. Count the number of video packets N_V in the three packets you have observed. Describe in words and also calculate the following probabilities:

(a) $P[N_V = 2]$

(b) $P[N_V \geq 1]$

(c) $P[\{vvd\}|N_V = 2]$

(d) $P[\{ddv\}|N_V = 2]$

(e) $P[N_V = 2|N_V \geq 1]$

(f) $P[N_V \geq 1|N_V = 2]$

Quiz 1.3 Solution

(a) The probability of exactly two voice packets is

$$\mathsf{P}[N_V = 2] = \mathsf{P}[\{vvd, vdv, dvv\}] = 0.3. \tag{1}$$

(b) The probability of at least one voice packet is

$$\begin{aligned}\mathsf{P}[N_V \geq 1] &= 1 - \mathsf{P}[N_V = 0] \\ &= 1 - \mathsf{P}[ddd] = 0.8. \end{aligned} \tag{2}$$

(c) The conditional probability of two voice packets followed by a data packet given that there were two voice packets is

$$\begin{aligned}\mathsf{P}[\{vvd\} | N_V = 2] &= \frac{\mathsf{P}[\{vvd\}, N_V = 2]}{\mathsf{P}[N_V = 2]} \\ &= \frac{\mathsf{P}[\{vvd\}]}{\mathsf{P}[N_V = 2]} = \frac{0.1}{0.3} = \frac{1}{3}. \end{aligned} \tag{3}$$

[Continued]

Quiz 1.3 Solution (Continued 2)

(d) The conditional probability of two data packets followed by a voice packet given there were two voice packets is

$$P[\{ddv\} | N_V = 2] = \frac{P[\{ddv\}, N_V = 2]}{P[N_V = 2]} = 0.$$

The joint event of the outcome ddv and exactly two voice packets has probability zero since there is only one voice packet in the outcome ddv.

(e) The conditional probability of exactly two voice packets given at least one voice packet is

$$\begin{aligned} P[N_V = 2 | N_v \geq 1] &= \frac{P[N_V = 2, N_V \geq 1]}{P[N_V \geq 1]} \\ &= \frac{P[N_V = 2]}{P[N_V \geq 1]} = \frac{0.3}{0.8} = \frac{3}{8}. \end{aligned} \quad (4)$$

(f) The conditional probability of at least one voice packet given there were exactly two voice packets is

$$P[N_V \geq 1 | N_V = 2] = \frac{P[N_V \geq 1, N_V = 2]}{P[N_V = 2]} = \frac{P[N_V = 2]}{P[N_V = 2]} = 1. \quad (5)$$

Given two voice packets, there must have been at least one voice packet.

Partitions and the Law of Total Probability

Example 1.12 Problem

Flip four coins, a penny, a nickel, a dime, and a quarter. Examine the coins in order (penny, then nickel, then dime, then quarter) and observe whether each coin shows a head (h) or a tail (t). What is the sample space? How many elements are in the sample space?

Example 1.12 Solution

The sample space consists of 16 four-letter words, with each letter either h or t. For example, the outcome $tthh$ refers to the penny and the nickel showing tails and the dime and quarter showing heads. There are 16 members of the sample space.

Example 1.13

Continuing Example 1.12, let $B_i = \{\text{outcomes with } i \text{ heads}\}$. Each B_i is an event containing one or more outcomes. For example,

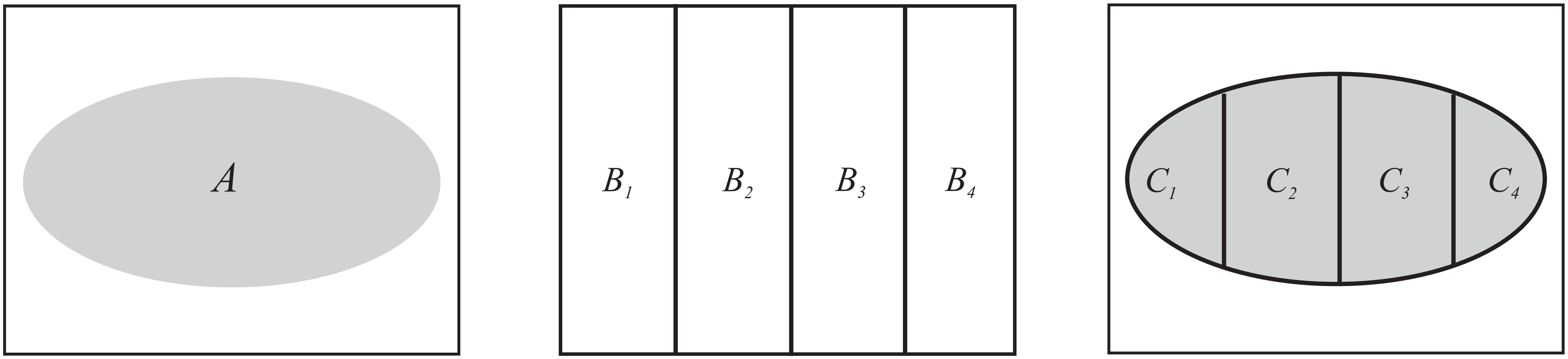
$$B_1 = \{ttth, ttht, thtt, httt\}$$

contains four outcomes. The set

$$B = \{B_0, B_1, B_2, B_3, B_4\}$$

is a partition. Its members are mutually exclusive and collectively exhaustive. It is not a sample space because it lacks the finest-grain property. Learning that an experiment produces an event B_1 tells you that one coin came up heads, but it doesn't tell you which coin it was.

Figure 1.1



In this example of Theorem 1.7, the partition is $B = \{B_1, B_2, B_3, B_4\}$ and $C_i = A \cap B_i$ for $i = 1, \ldots, 4$. It should be apparent that $A = C_1 \cup C_2 \cup C_3 \cup C_4$.

Theorem 1.7

For a partition $B = \{B_1, B_2, \ldots\}$ and any event A in the sample space, let $C_i = A \cap B_i$. For $i \neq j$, the events C_i and C_j are mutually exclusive and

$$A = C_1 \cup C_2 \cup \cdots .$$

Example 1.14

In the coin-tossing experiment of Example 1.12, let A equal the set of outcomes with less than three heads:

$$A = \{tttt, httt, thtt, ttht, ttth, hhtt, htht, htth, tthh, thth, thht\}. \qquad (1.15)$$

From Example 1.13, let $B_i = \{\text{outcomes with } i \text{ heads}\}$. Since $\{B_0, \ldots, B_4\}$ is a partition, Theorem 1.7 states that

$$A = (A \cap B_0) \cup (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3) \cup (A \cap B_4) \qquad (1.16)$$

In this example, $B_i \subset A$, for $i = 0, 1, 2$. Therefore $A \cap B_i = B_i$ for $i = 0, 1, 2$. Also, for $i = 3$ and $i = 4$, $A \cap B_i = \varnothing$ so that $A = B_0 \cup B_1 \cup B_2$, a union of mutually exclusive sets. In words, this example states that the event "less than three heads" is the union of events "zero heads," "one head," and "two heads."

Theorem 1.8

For any event A, and partition $\{B_1, B_2, \ldots, B_m\}$,

$$\mathsf{P}[A] = \sum_{i=1}^{m} \mathsf{P}[A \cap B_i].$$

Proof: Theorem 1.8

The proof follows directly from Theorem 1.7 and Theorem 1.2. In this case, the mutually exclusive sets are $C_i = \{A \cap B_i\}$.

Example 1.15

A company has a model of email use. It classifies all emails as either long (l), if they are over 10 MB in size, or brief (b). It also observes whether the email is just text (t), has attached images (i), or has an attached video (v). This model implies an experiment in which the procedure is to monitor an email and the observation consists of the type of email, t, i, or v, and the length, l or b. The sample space has six outcomes: $S = \{lt, bt, li, bi, lv, bv\}$. In this problem, each email is classifed in two ways: by length and by type. Using L for the event that an email is long and B for the event that a email is brief, $\{L, B\}$ is a partition. Similarly, the text (T), image (I), and video (V) classification is a partition $\{T, I, V\}$. The sample space can be represented by a table in which the rows and columns are labeled by events and the intersection of each row and column event contains a single outcome. The corresponding table entry is the probability of that outcome. In this case, the table is

	T	I	V
L	0.3	0.12	0.15
B	0.2	0.08	0.15

(1.17)

For example, from the table we can read that the probability of a brief image email is $\mathrm{P}[bi] = \mathrm{P}[BI] = 0.08$. Note that $\{T, I, V\}$ is a partition corresponding to $\{B_1, B_2, B_3\}$ in Theorem 1.8. Thus we can apply Theorem 1.8 to find the probability of a long email:

$$\mathrm{P}[L] = \mathrm{P}[LT] + \mathrm{P}[LI] + \mathrm{P}[LV] = 0.57. \tag{1.18}$$

Theorem 1.9 Law of Total Probability

For a partition $\{B_1, B_2, \ldots, B_m\}$ with $\mathrm{P}[B_i] > 0$ for all i,

$$\mathrm{P}[A] = \sum_{i=1}^{m} \mathrm{P}[A|B_i]\,\mathrm{P}[B_i]\,.$$

Proof: Theorem 1.9

This follows from Theorem 1.8 and the identity $\mathsf{P}[AB_i] = \mathsf{P}[A|B_i]\,\mathsf{P}[B_i]$, which is a direct consequence of the definition of conditional probability.

Example 1.16 Problem

A company has three machines B_1, B_2, and B_3 making 1 kΩ resistors. Resistors within 50 Ω of the nominal value are considered acceptable. It has been observed that 80% of the resistors produced by B_1 and 90% of the resistors produced by B_2 are acceptable. The percentage for machine B_3 is 60%. Each hour, machine B_1 produces 3000 resistors, B_2 produces 4000 resistors, and B_3 produces 3000 resistors. All of the resistors are mixed together at random in one bin and packed for shipment. What is the probability that the company ships an acceptable resistor?

Example 1.16 Solution

Let $A = \{\text{resistor is acceptable}\}$. Using the resistor accuracy information to formulate a probability model, we write

$$P[A|B_1] = 0.8, \quad P[A|B_2] = 0.9, \quad P[A|B_3] = 0.6. \tag{1.19}$$

The production figures state that $3000 + 4000 + 3000 = 10{,}000$ resistors per hour are produced. The fraction from machine B_1 is $P[B_1] = 3000/10{,}000 = 0.3$. Similarly, $P[B_2] = 0.4$ and $P[B_3] = 0.3$. Now it is a simple matter to apply the law of total probability to find the acceptable probability for all resistors shipped by the company:

$$\begin{aligned} P[A] &= P[A|B_1]P[B_1] + P[A|B_2]P[B_2] + P[A|B_3]P[B_3] \\ &= (0.8)(0.3) + (0.9)(0.4) + (0.6)(0.3) = 0.78. \end{aligned} \tag{1.20}$$

For the whole factory, 78% of resistors are within 50 Ω of the nominal value.

Theorem 1.10 Bayes' theorem

$$\mathsf{P}\left[B|A\right] = \frac{\mathsf{P}\left[A|B\right]\mathsf{P}\left[B\right]}{\mathsf{P}\left[A\right]}.$$

Example 1.17 Problem

In Example 1.16 about a shipment of resistors from the factory, we learned that:

- The probability that a resistor is from machine B_3 is $\text{P}[B_3] = 0.3$.
- The probability that a resistor is *acceptable* — i.e., within 50 Ω of the nominal value — is $\text{P}[A] = 0.78$.
- Given that a resistor is from machine B_3, the conditional probability that it is acceptable is $\text{P}[A|B_3] = 0.6$.

What is the probability that an acceptable resistor comes from machine B_3?

Example 1.17 Solution

Now we are given the event A that a resistor is within 50 Ω of the nominal value, and we need to find $\mathrm{P}[B_3|A]$. Using Bayes' theorem, we have

$$\mathrm{P}\left[B_3|A\right] = \frac{\mathrm{P}\left[A|B_3\right]\mathrm{P}\left[B_3\right]}{\mathrm{P}\left[A\right]}. \tag{1.23}$$

Since all of the quantities we need are given in the problem description, our answer is

$$\mathrm{P}\left[B_3|A\right] = (0.6)(0.3)/(0.78) = 0.23. \tag{1.24}$$

Similarly we obtain $\mathrm{P}[B_1|A] = 0.31$ and $\mathrm{P}[B_2|A] = 0.46$. Of all resistors within 50 Ω of the nominal value, only 23% come from machine B_3 (even though this machine produces 30% of all resistors). Machine B_1 produces 31% of the resistors that meet the 50 Ω criterion and machine B_2 produces 46% of them.

Quiz 1.4

Monitor customer behavior in the Phonesmart store. Classify the behavior as buying (B) if a customer purchases a smartphone. Otherwise the behavior is no purchase (N). Classify the time a customer is in the store as long (L) if the customer stays more than three minutes; otherwise classify the amount of time as rapid (R). Based on experience with many customers, we use the probability model $\text{P}[N] = 0.7$, $\text{P}[L] = 0.6$, $\text{P}[NL] = 0.35$. Find the following probabilities:

(a) $\text{P}[B \cup L]$

(b) $\text{P}[N \cup L]$

(c) $\text{P}[N \cup B]$

(d) $\text{P}[LR]$

Quiz 1.4 Solution

We can describe this experiment by the event space consisting of the four possible events NL, NR, BL, and BR. We represent these events in the table:

	N	B
L	0.35	?
R	?	?

Once we fill in the table, finding the various probabilities will be simple.

In a roundabout way, the problem statement tells us how to fill in the table. In particular,

$$\mathrm{P}[N] = 0.7 = \mathrm{P}[NL] + \mathrm{P}[NR],$$
$$\mathrm{P}[L] = 0.6 = \mathrm{P}[NL] + \mathrm{P}[BL].$$

Since $\mathrm{P}[NL] = 0.35$, we can conclude that $\mathrm{P}[NR] = 0.7 - 0.35 = 0.35$ and that $\mathrm{P}[BL] = 0.6 - 0.35 = 0.25$. This allows us to fill in two more table entries:

	N	B
L	0.35	0.25
R	0.35	?

The remaining table entry is filled in by observing that the probabilities must sum to 1.
[Continued]

Quiz 1.4 Solution (Continued 2)

This implies $\mathsf{P}[BR] = 0.05$ and the complete table is

	N	B
L	0.35	0.25
R	0.35	0.05

The various probabilities are now simple:

(a) $$\begin{aligned}\mathsf{P}[B \cup L] &= \mathsf{P}[NL] + \mathsf{P}[BL] + \mathsf{P}[BR] \\ &= 0.35 + 0.25 + 0.05 = 0.65.\end{aligned}$$

(b) $$\begin{aligned}\mathsf{P}[N \cup L] &= \mathsf{P}[N] + \mathsf{P}[L] - \mathsf{P}[NL] \\ &= 0.7 + 0.6 - 0.35 = 0.95.\end{aligned}$$

(c) $\mathsf{P}[N \cup B] = \mathsf{P}[S] = 1.$

(d) $\mathsf{P}[LR] = \mathsf{P}[LL^c] = 0.$

Section 1.5

Independence

Definition 1.6 Two Independent Events

Events A *and* B *are* independent *if and only if*

$$\mathsf{P}[AB] = \mathsf{P}[A]\,\mathsf{P}[B]\,.$$

1.5 Comment: Independent vs. Mutually Exclusive

Keep in mind that **independent and mutually exclusive are not synonyms**.

In some contexts these words can have similar meanings, but this is not the case in probability. Mutually exclusive events A and B have no outcomes in common and therefore $\text{P}[AB] = 0$. In most situations independent events are not mutually exclusive! Exceptions occur only when $\text{P}[A] = 0$ or $\text{P}[B] = 0$. When we have to calculate probabilities, knowledge that events A and B are *mutually exclusive* is very helpful. Axiom 3 enables us to *add* their probabilities to obtain the probability of the *union*. Knowledge that events C and D are *independent* is also very useful. Definition 1.6 enables us to *multiply* their probabilities to obtain the probability of the *intersection*.

Example 1.18 Problem

Suppose that for the experiment monitoring three purchasing decisions in Example 1.7, each outcome (a sequence of three decisions, each either buy or not buy) is equally likely. Are the events B_2 that the second customer purchases a phone and N_2 that the second customer does not purchase a phone independent? Are the events B_1 and B_2 independent?

Example 1.18 Solution

Each element of the sample space $S = \{bbb, bbn, bnb, bnn, nbb, nbn, nnb, nnn\}$ has probability 1/8. Each of the events

$$B_2 = \{bbb, bbn, nbb, nbn\} \quad \text{and} \quad N_2 = \{bnb, bnn, nnb, nnn\} \tag{1.26}$$

contains four outcomes, so $\mathrm{P}[B_2] = \mathrm{P}[N_2] = 4/8$. However, $B_2 \cap N_2 = \varnothing$ and $\mathrm{P}[B_2 N_2] = 0$. That is, B_2 and N_2 are mutually exclusive because the second customer cannot both purchase a phone and not purchase a phone. Since $\mathrm{P}[B_2 N_2] \neq \mathrm{P}[B_2]\,\mathrm{P}[N_2]$, B_2 and N_2 are not independent. Learning whether or not the event B_2 (second customer buys a phone) occurs drastically affects our knowledge of whether or not the event N_2 (second customer does not buy a phone) occurs. Each of the events $B_1 = \{bnn, bnb, bbn, bbb\}$ and $B_2 = \{bbn, bbb, nbn, nbb\}$ has four outcomes, so $\mathrm{P}[B_1] = \mathrm{P}[B_2] = 4/8 = 1/2$. In this case, the intersection $B_1 \cap B_2 = \{bbn, bbb\}$ has probability $\mathrm{P}[B_1 B_2] = 2/8 = 1/4$. Since $\mathrm{P}[B_1 B_2] = \mathrm{P}[B_1]\,\mathrm{P}[B_2]$, events B_1 and B_2 are independent. Learning whether or not the event B_2 (second customer buys a phone) occurs does not affect our knowledge of whether or not the event B_1 (first customer buys a phone) occurs.

Example 1.19 Problem

Integrated circuits undergo two tests. A mechanical test determines whether pins have the correct spacing, and an electrical test checks the relationship of outputs to inputs. We *assume* that electrical failures and mechanical failures occur independently. Our information about circuit production tells us that mechanical failures occur with probability 0.05 and electrical failures occur with probability 0.2. What is the probability model of an experiment that consists of testing an integrated circuit and observing the results of the mechanical and electrical tests?

Example 1.19 Solution

To build the probability model, we note that the sample space contains four outcomes:

$$S = \{(ma, ea), (ma, er), (mr, ea), (mr, er)\} \tag{1.27}$$

where m denotes mechanical, e denotes electrical, a denotes accept, and r denotes reject. Let M and E denote the events that the mechanical and electrical tests are acceptable. Our prior information tells us that $\mathsf{P}[M^c] = 0.05$, and $\mathsf{P}[E^c] = 0.2$. This implies $\mathsf{P}[M] = 0.95$ and $\mathsf{P}[E] = 0.8$. Using the independence assumption and Definition 1.6, we obtain the probabilities of the four outcomes:

$$\mathsf{P}[(ma, ea)] = \mathsf{P}[ME] = \mathsf{P}[M]\,\mathsf{P}[E] = 0.95 \times 0.8 = 0.76, \tag{1.28}$$

$$\mathsf{P}[(ma, er)] = \mathsf{P}[ME^c] = \mathsf{P}[M]\,\mathsf{P}[E^c] = 0.95 \times 0.2 = 0.19, \tag{1.29}$$

$$\mathsf{P}[(mr, ea)] = \mathsf{P}[M^cE] = \mathsf{P}[M^c]\,\mathsf{P}[E] = 0.05 \times 0.8 = 0.04, \tag{1.30}$$

$$\mathsf{P}[(mr, er)] = \mathsf{P}[M^cE^c] = \mathsf{P}[M^c]\,\mathsf{P}[E^c] = 0.05 \times 0.2 = 0.01. \tag{1.31}$$

Definition 1.7 Three Independent Events

A_1, A_2, *and* A_3 *are* mutually independent *if and only if*

(a) A_1 *and* A_2 *are independent,*

(b) A_2 *and* A_3 *are independent,*

(c) A_1 *and* A_3 *are independent,*

(d) $\mathsf{P}[A_1 \cap A_2 \cap A_3] = \mathsf{P}[A_1]\,\mathsf{P}[A_2]\,\mathsf{P}[A_3]$.

Example 1.20 Problem

In an experiment with equiprobable outcomes, the partition is $S = \{1, 2, 3, 4\}$. $P[s] = 1/4$ for all $s \in S$. Are the events $A_1 = \{1, 3, 4\}$, $A_2 = \{2, 3, 4\}$, and $A_3 = \varnothing$ mutually independent?

Example 1.20 Solution

These three sets satisfy the final condition of Definition 1.7 because $A_1 \cap A_2 \cap A_3 = \varnothing$, and

$$\mathrm{P}[A_1 \cap A_2 \cap A_3] = \mathrm{P}[A_1]\,\mathrm{P}[A_2]\,\mathrm{P}[A_3] = 0. \tag{1.32}$$

However, A_1 and A_2 are not independent because, with all outcomes equiprobable,

$$\mathrm{P}[A_1 \cap A_2] = \mathrm{P}[\{3,4\}] = 1/2 \neq \mathrm{P}[A_1]\,\mathrm{P}[A_2] = 3/4 \times 3/4. \tag{1.33}$$

Hence the three events are not mutually independent.

Definition 1.8 Events

If $n \geq 3$*, the events* $A_1, A_2, \ldots, A_n$ *are mutually independent if and only if*

(a) *all collections of* $n-1$ *events chosen from* $A_1, A_2, \ldots A_n$ *are mutually independent,*

(b) $\mathrm{P}[A_1 \cap A_2 \cap \cdots \cap A_n] = \mathrm{P}[A_1]\,\mathrm{P}[A_2] \cdots \mathrm{P}[A_n]$.

Quiz 1.5

Monitor two consecutive packets going through a router. Classify each one as video (v) if it was sent from a Youtube server or as ordinary data (d) otherwise. Your observation is a sequence of two letters (either v or d). For example, two video packets corresponds to vv. The two packets are independent and the probability that any one of them is a video packet is 0.8. Denote the identity of packet i by C_i. If packet i is a video packet, then $C_i = v$; otherwise, $C_i = d$. Count the number N_V of video packets in the two packets you have observed. Determine whether the following pairs of events are independent:

(a) $\{N_V = 2\}$ and $\{N_V \geq 1\}$

(b) $\{N_V \geq 1\}$ and $\{C_1 = v\}$

(c) $\{C_2 = v\}$ and $\{C_1 = d\}$

(d) $\{C_2 = v\}$ and $\{N_V \text{ is even}\}$

Quiz 1.5 Solution

In this experiment, there are four outcomes with probabilities

$$\mathrm{P}[\{vv\}] = (0.8)^2 = 0.64, \qquad \mathrm{P}[\{vd\}] = (0.8)(0.2) = 0.16,$$
$$\mathrm{P}[\{dv\}] = (0.2)(0.8) = 0.16, \qquad \mathrm{P}[\{dd\}] = (0.2)^2 = 0.04.$$

When checking the independence of any two events A and B, it's wise to avoid intuition and simply check whether $\mathrm{P}[AB] = \mathrm{P}[A]\,\mathrm{P}[B]$. Using the probabilities of the outcomes, we now can test for the independence of events.

(a) First, we calculate the probability of the joint event:

$$\mathrm{P}\left[N_V = 2, N_V \geq 1\right] = \mathrm{P}\left[N_V = 2\right] = \mathrm{P}\left[\{vv\}\right] = 0.64. \tag{1}$$

Next, we observe that $\mathrm{P}[N_V \geq 1] = \mathrm{P}[\{vd, dv, vv\}] = 0.96.$. Finally, we make the comparison

$$\mathrm{P}\left[N_V = 2\right] \mathrm{P}\left[N_V \geq 1\right] = (0.64)(0.96) \neq \mathrm{P}\left[N_V = 2, N_V \geq 1\right], \tag{2}$$

which shows the two events are dependent. [Continued]

Quiz 1.5 Solution (Continued 2)

(b) The probability of the joint event is

$$\mathrm{P}\left[N_V \geq 1, C_1 = v\right] = \mathrm{P}\left[\{vd, vv\}\right] = 0.80. \tag{3}$$

From part (a), $\mathrm{P}[N_V \geq 1] = 0.96$. Further, $\mathrm{P}[C_1 = v] = 0.8$ so that

$$\mathrm{P}\left[N_V \geq 1\right]\mathrm{P}\left[C_1 = v\right] = (0.96)(0.8) = 0.768 \neq \mathrm{P}\left[N_V \geq 1, C_1 = v\right]. \tag{4}$$

Hence, the events are dependent.

(c) The problem statement that the packets were independent implies that the events $\{C_2 = v\}$ and $\{C_1 = d\}$ are independent events. Just to be sure, we can do the calculations to check:

$$\mathrm{P}\left[C_1 = d, C_2 = v\right] = \mathrm{P}\left[\{dv\}\right] = 0.16. \tag{5}$$

Since $\mathrm{P}[C_1 = d]\,\mathrm{P}[C_2 = v] = (0.2)(0.8) = 0.16$, we confirm that the events are independent. Note that this shouldn't be surprising since we used the information that the packets were independent in the problem statement to determine the probabilities of the outcomes. [Continued]

Quiz 1.5 Solution (Continued 3)

(d) The probability of the joint event is

$$\mathsf{P}[C_2 = v, N_V \text{ is even}] = \mathsf{P}[\{vv\}] = 0.64. \tag{6}$$

Also, each event has probability

$$\mathsf{P}[C_2 = v] = \mathsf{P}[\{dv, vv\}] = 0.8, \tag{7}$$

$$\mathsf{P}[N_V \text{ is even}] = \mathsf{P}[\{dd, vv\}] = 0.68. \tag{8}$$

Thus,

$$\begin{aligned} \mathsf{P}[C_2 = v]\,\mathsf{P}[N_V \text{ is even}] &= (0.8)(0.68) \\ &= 0.544 \neq \mathsf{P}[C_2 = v, N_V \text{ is even}]. \end{aligned} \tag{9}$$

Thus the events are dependent.

Section 1.6

Matlab

Example 1.21

```
>> X=rand(1,4)
X =
  0.0879  0.9626  0.6627  0.2023
>> X<0.5
ans =
     1     0     0     1
```

Since `rand(1,4) < 0.5` compares four random numbers against 0.5, the result is a random sequence of zeros and ones that simulates a sequence of four flips of a fair coin. We associate the outcome 1 with {head} and 0 with {tail}.

Example 1.22 Problem

Use Matlab to generate 12 random student test scores T as described in Quiz 1.2.

Example 1.22 Solution

Since `randi(50,1,12)` generates 12 test scores from the set $\{1, \ldots, 50\}$, we need only to add 50 to each score to obtain test scores in the range $\{51, \ldots, 100\}$.

```
>> 50+randi(50,1,12)
ans =
    69    78    60    68    93    99    77    95    88    57    51    90
```

Example 1.23

```
>> s=rng;
>> 50+randi(50,1,12)
ans =
   89   76   80   80   72   92   58   56   77   78   59   58
>> rng(s);
>> 50+randi(50,1,12)
ans =
   89   76   80   80   72   92   58   56   77   78   59   58
```

Quiz 1.6

The number of characters in a tweet is equally likely to be any integer between 1 and 140. Simulate an experiment that generates 1000 tweets and counts the number of "long" tweets that have over 120 characters. Repeat this experiment 5 times.

Quiz 1.6 Solution

These two matlab instructions

```
>> T=randi(140,1000,5);
>> sum(T>120)
ans =
   126   147   134   133   163
```

simulate 5 runs of an experiment each with 1000 tweets. In particular, we note that `T=randi(140,1000,5)` generates a 1000×5 array `T` of pseudorandom integers between 1 and 140. Each column of `T` has 1000 entries representing an experimental run corresponding to the lengths of 1000 tweets. The comparison `T>120` produces a 5×1000 binary matrix in which each 1 marks a long tweet with length over 120 characters. Summing this binary array along the columns with the command `sum(T>120)` counts the number of long tweets in each experimental run.

The experiment in which we examine the length of one tweet has sample space $S = \{s_1, s_2, \ldots, s_{140}\}$ with s_i denoting the outcome that a tweet has length i. Note that $\mathrm{P}[s_i] = 1/140$ and thus

$$\mathrm{P}\left[\text{tweet length} > 120\right] = \mathrm{P}\left[\{s_{121}, s_{122}, \ldots, s_{140}\}\right] = \frac{20}{140} = \frac{1}{7}. \tag{1}$$

Thus in each run of 1000 tweets, we would expect to see about 1/7 of the tweets, or about 143 tweets, to be be long tweets with length of over 120 characters. However, because the lengths are random, we see that we observe in the neighborhood of 143 long tweets in each run.