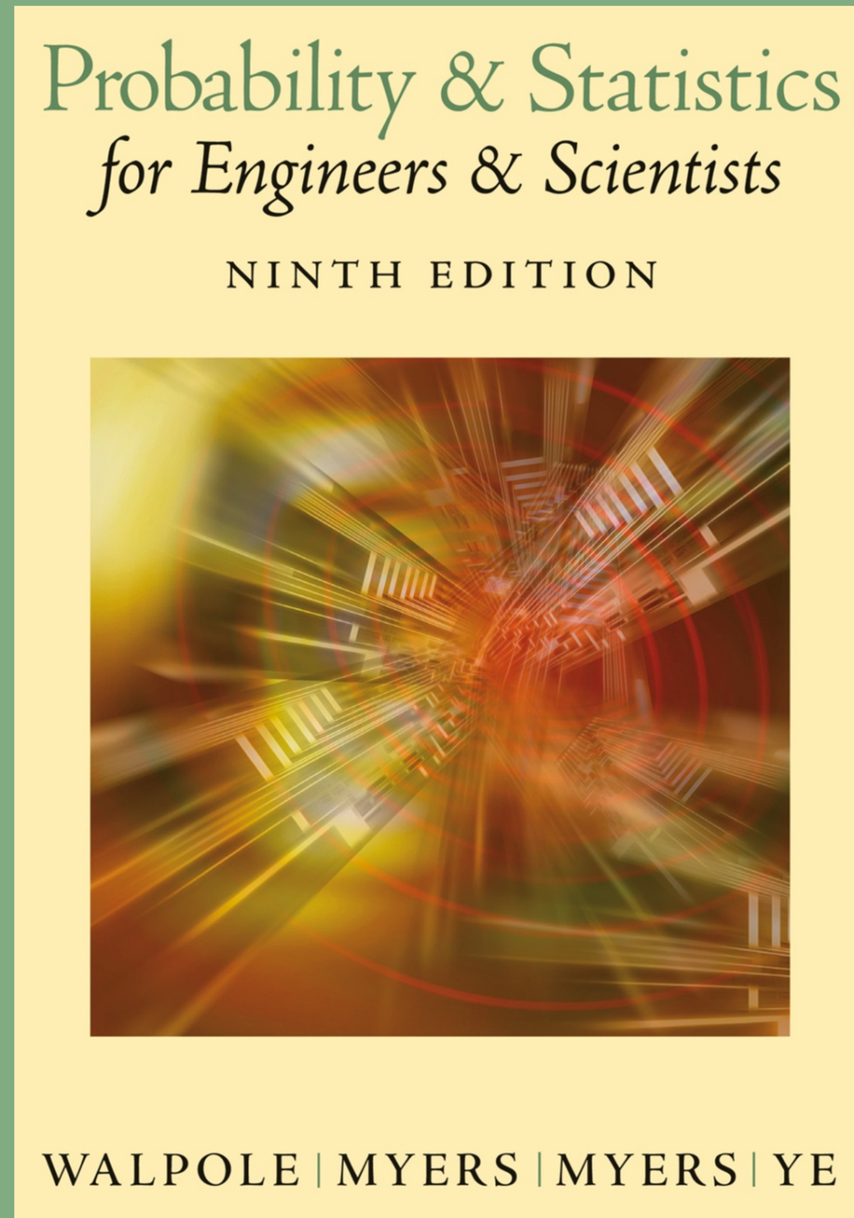
Chapter 2

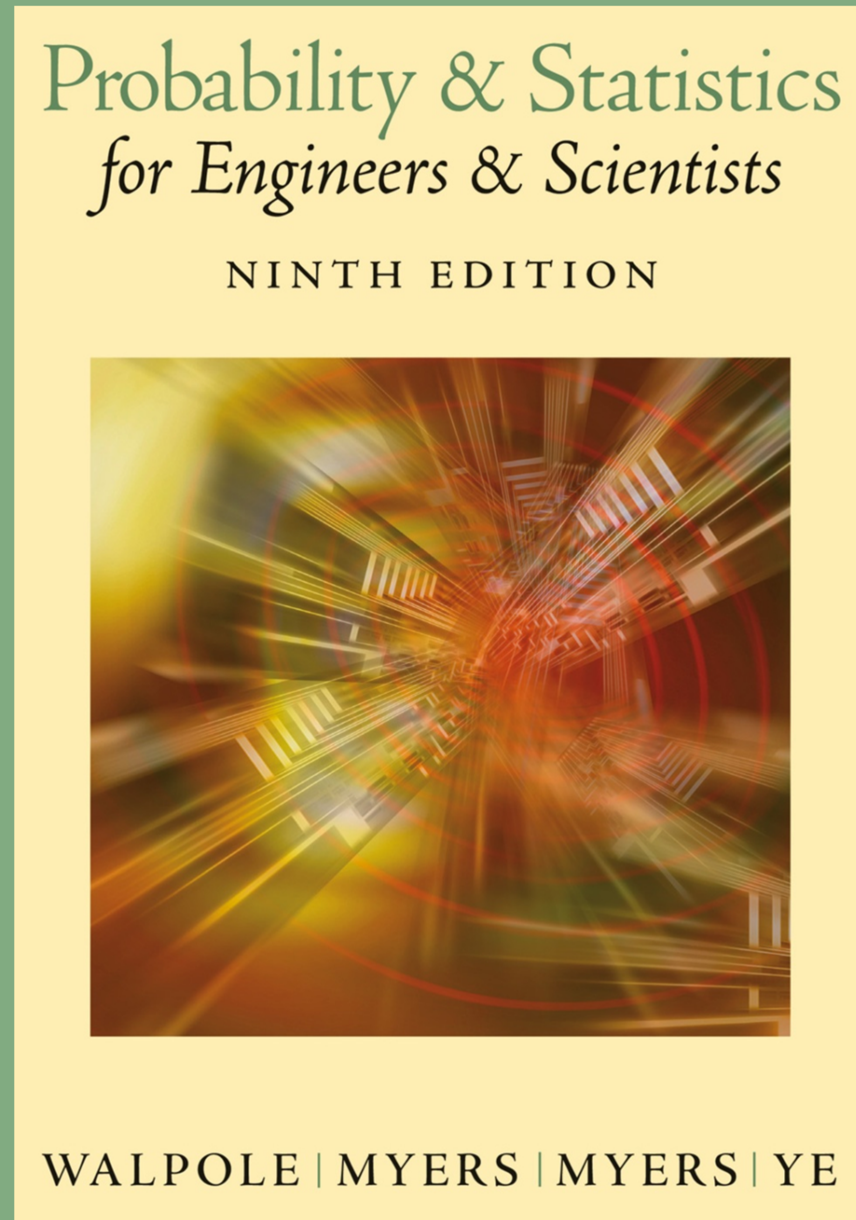
Probability



PEARSON
Addison Wesley

Section 2.1

Sample Space



Definition 2.1

The set of all possible outcomes of a statistical experiment is called the **sample space** and is represented by the symbol S.

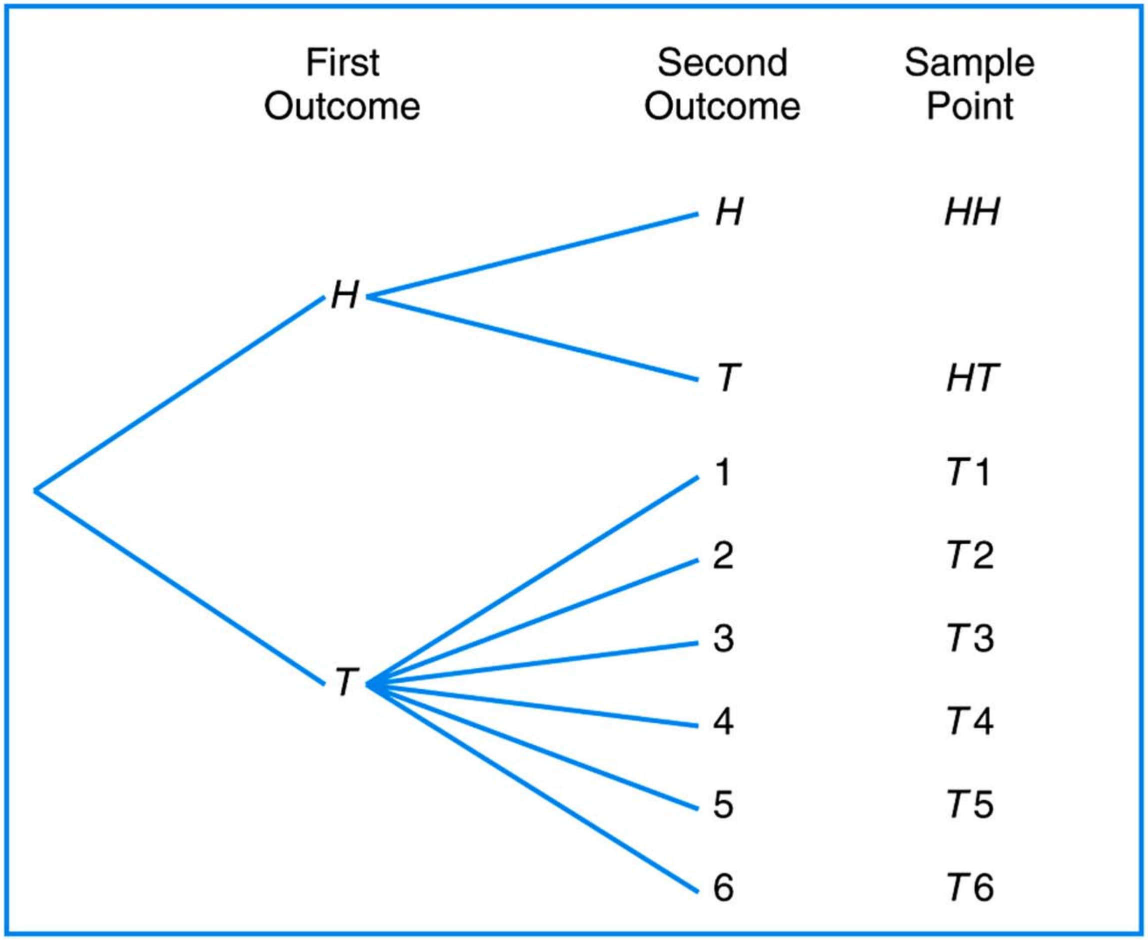
Example 2.2

An experiment consists of

- Flipping a coin, then
- Flip it again if the first one is heads, or
- Toss a die if the first toss is a tail

List the sample space

Figure 2.1 Tree diagram for Example 2.2



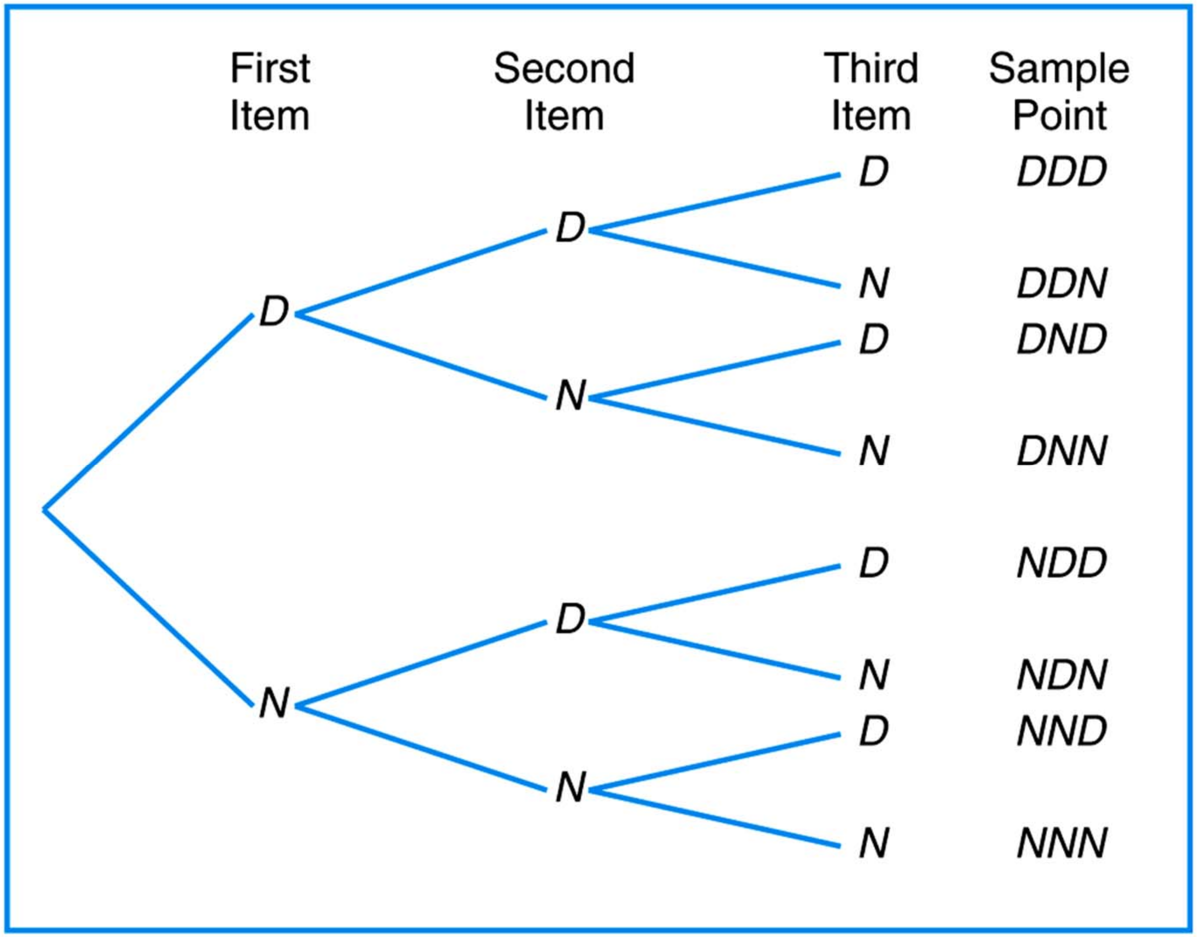
Example 2.3

Three items are selected from a manufacturing process

Each item is tested and then classified as D for defective or N for nondefective

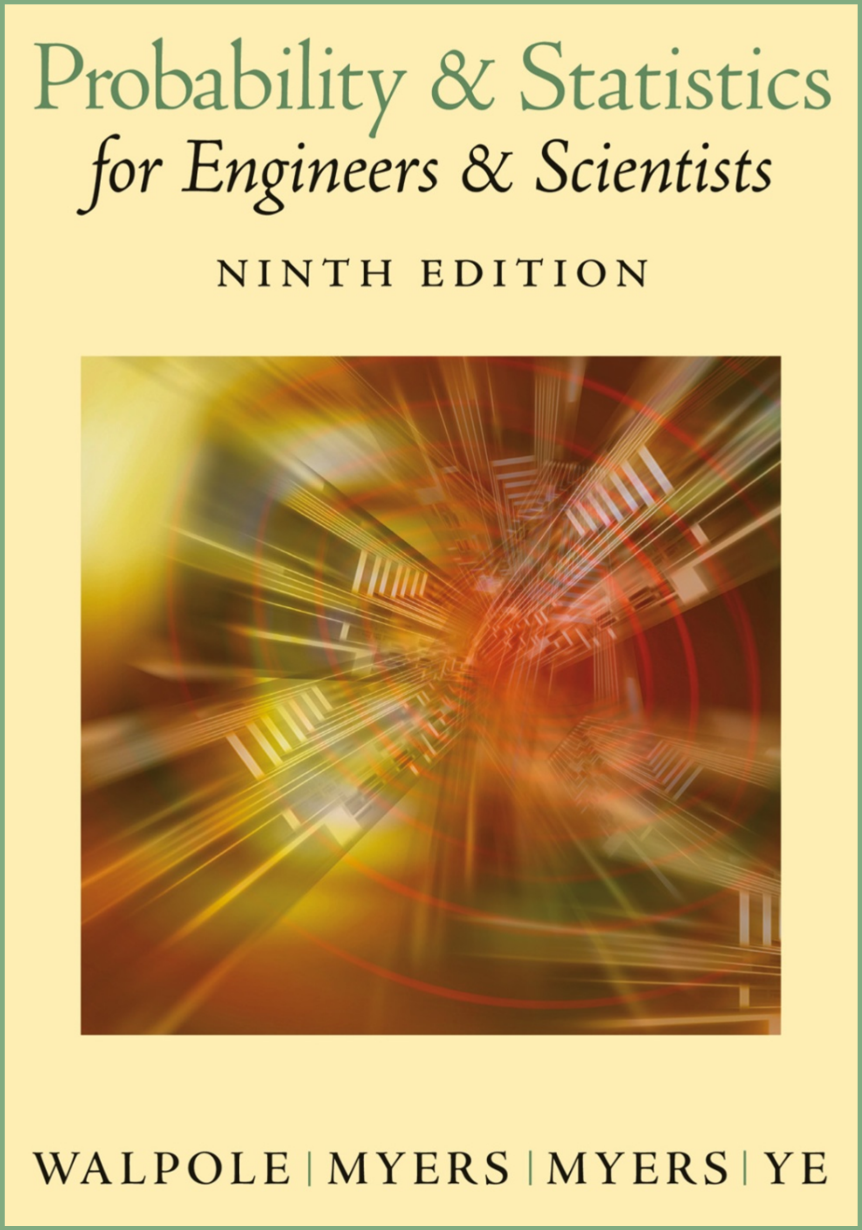
List the elements of the sample space

Figure 2.2 Tree Diagram for Example 2.3



Section 2.2

Events



PEARSON Addison Wesley

Definition 2.2

An **event** is a subset of a sample space.

Definition 2.3

The **complement** of an event A with respect to S is the subset of all elements of S that are not in A. We denote the complement of A by the symbol A'.

Definition 2.4

The **intersection** of two events A and B, denoted by the symbol $A \cap B$, is the event containing all elements that are common to A and B.

Definition 2.5

Two events A and B are **mutually exclusive**, or **disjoint**, if $A \cap B = \phi$, that is, if A and B have no elements in common.

Definition 2.6

The **union** of the two events A and B, denoted by the symbol $A \cup B$, is the event containing all the elements that belong to A or B or both.

Figure 2.3 Events represented by various regions

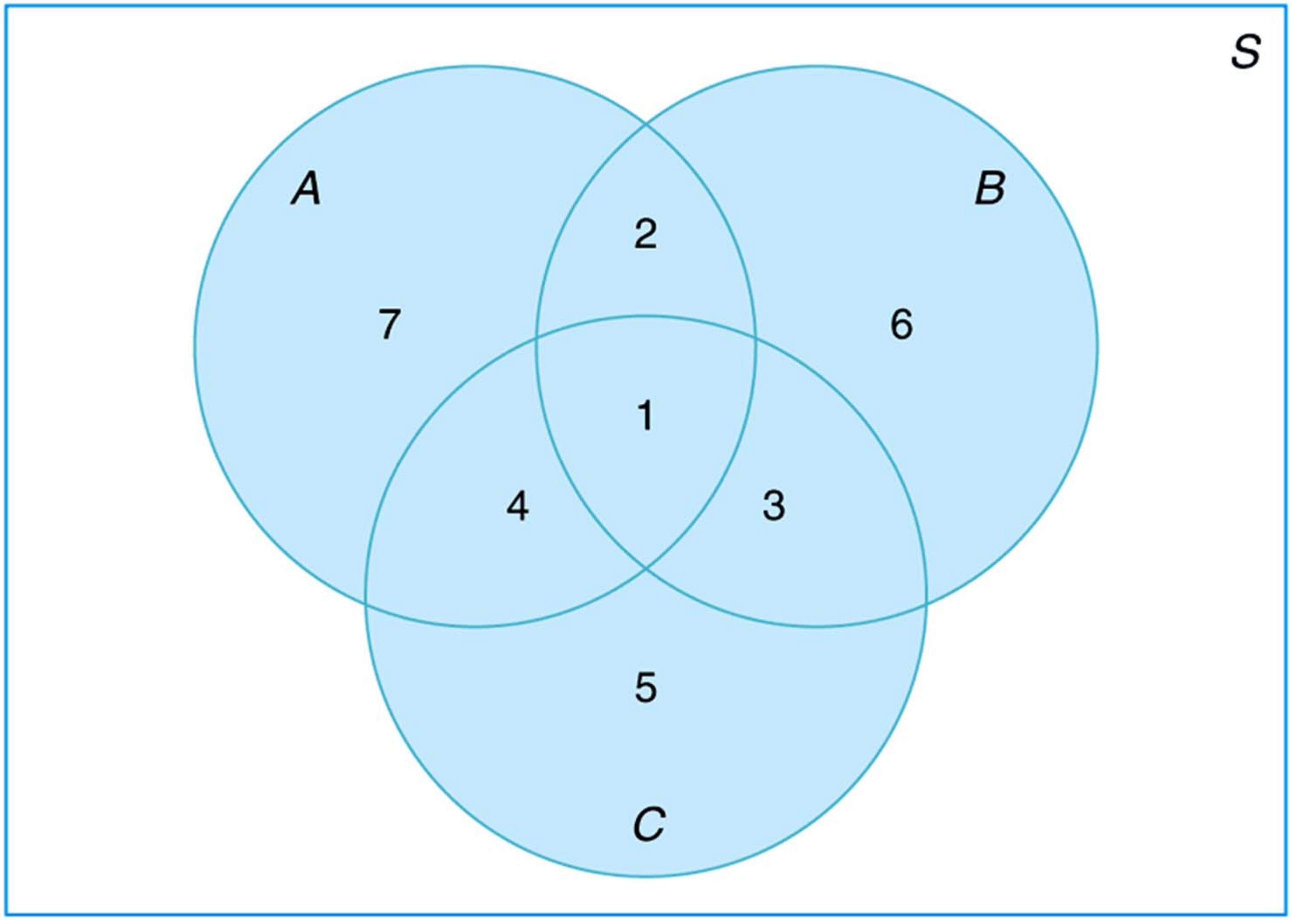


Figure 2.4 Events of the sample space *S*

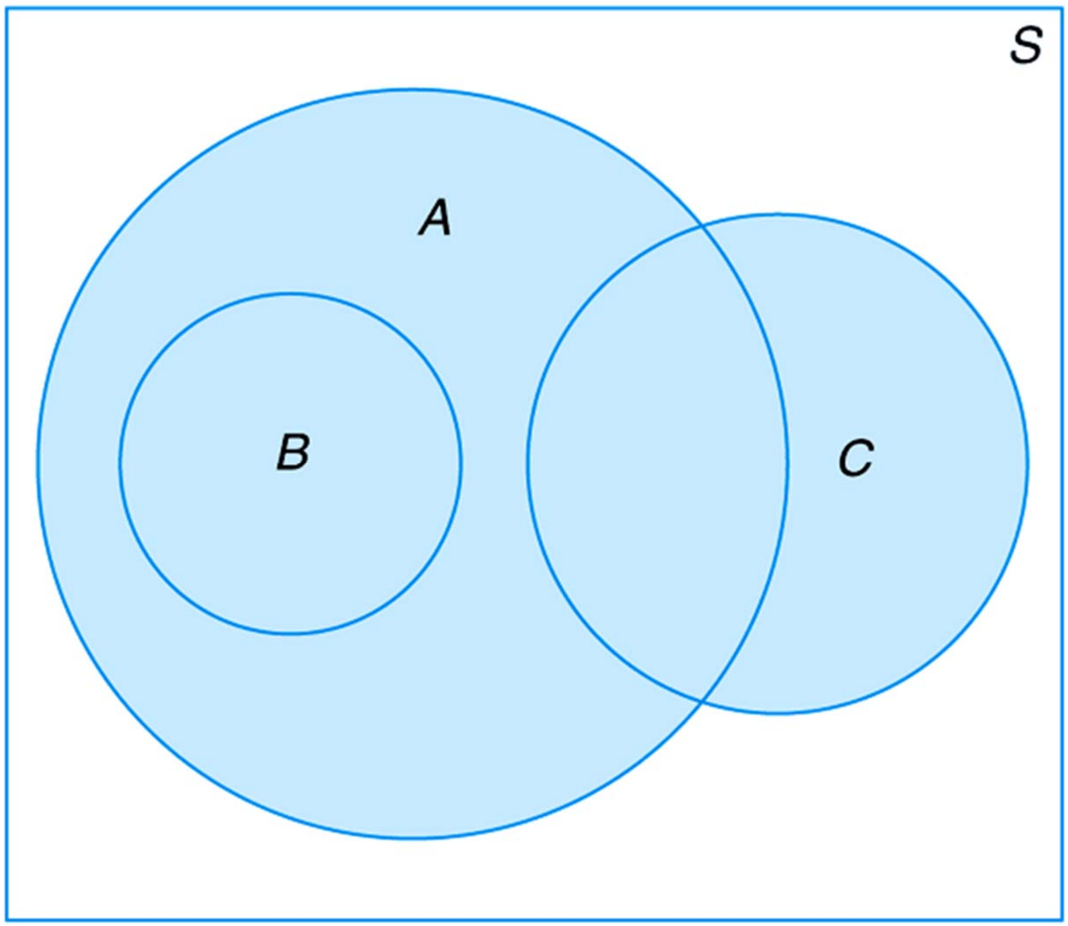
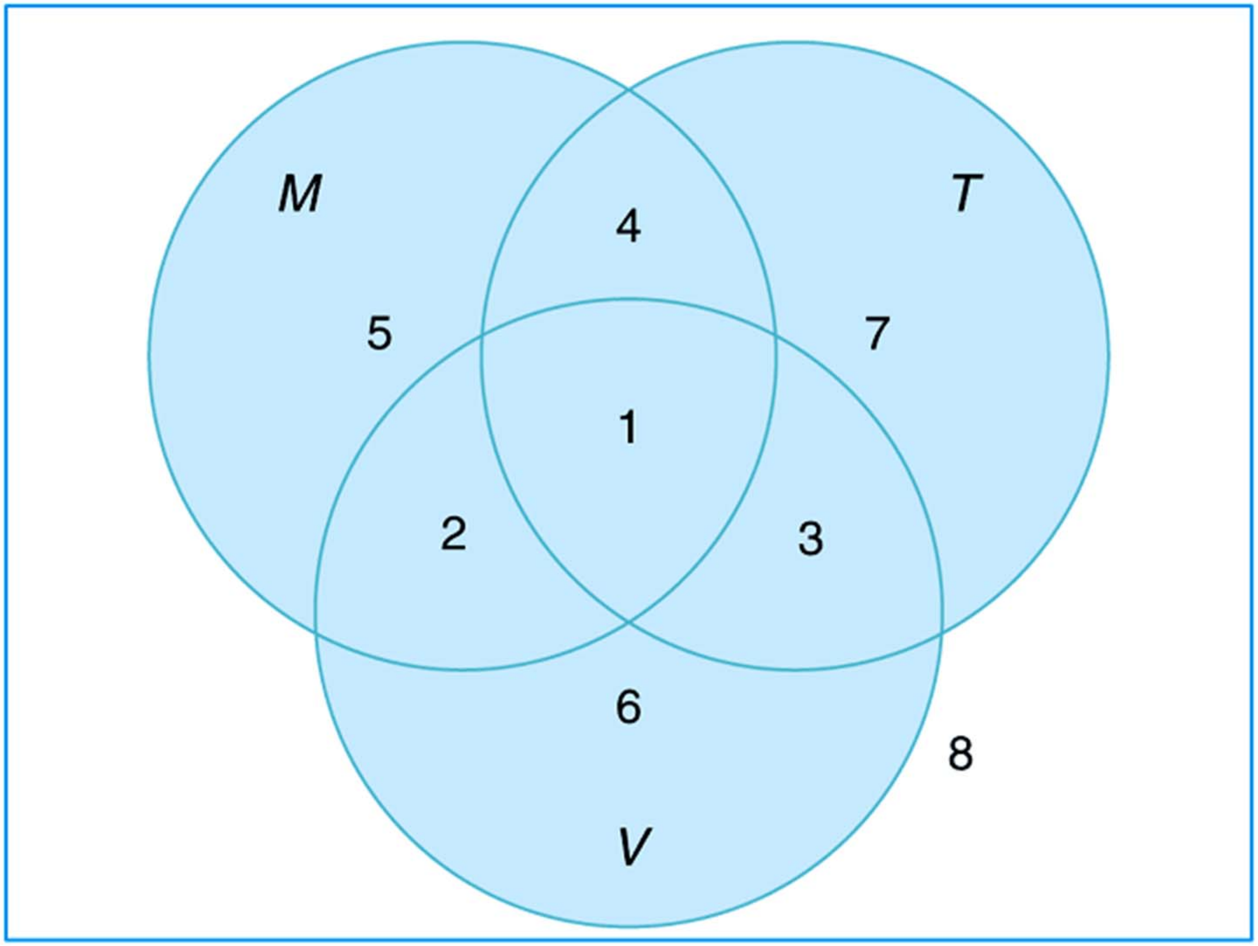
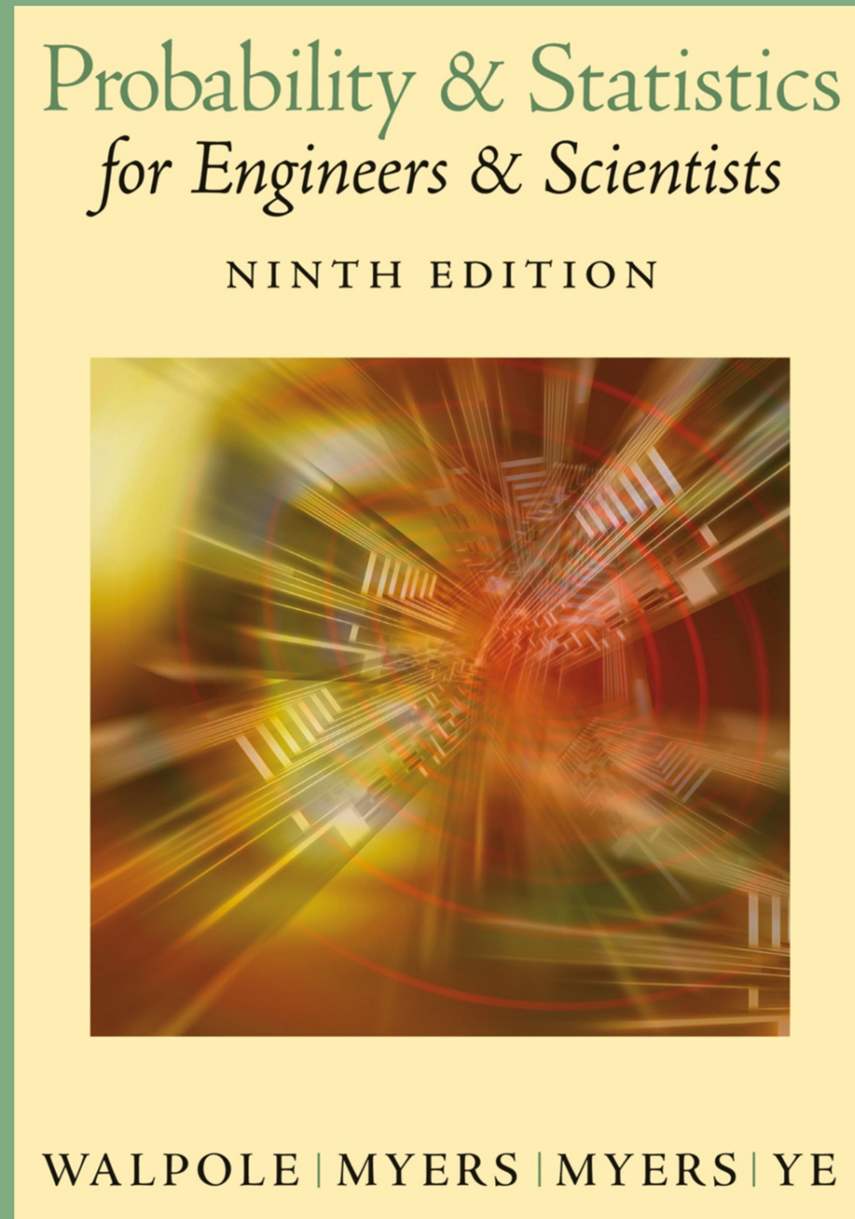


Figure 2.5 Venn diagram for Exercises 2.19 and 2.20



Section 2.3

Counting Sample Points



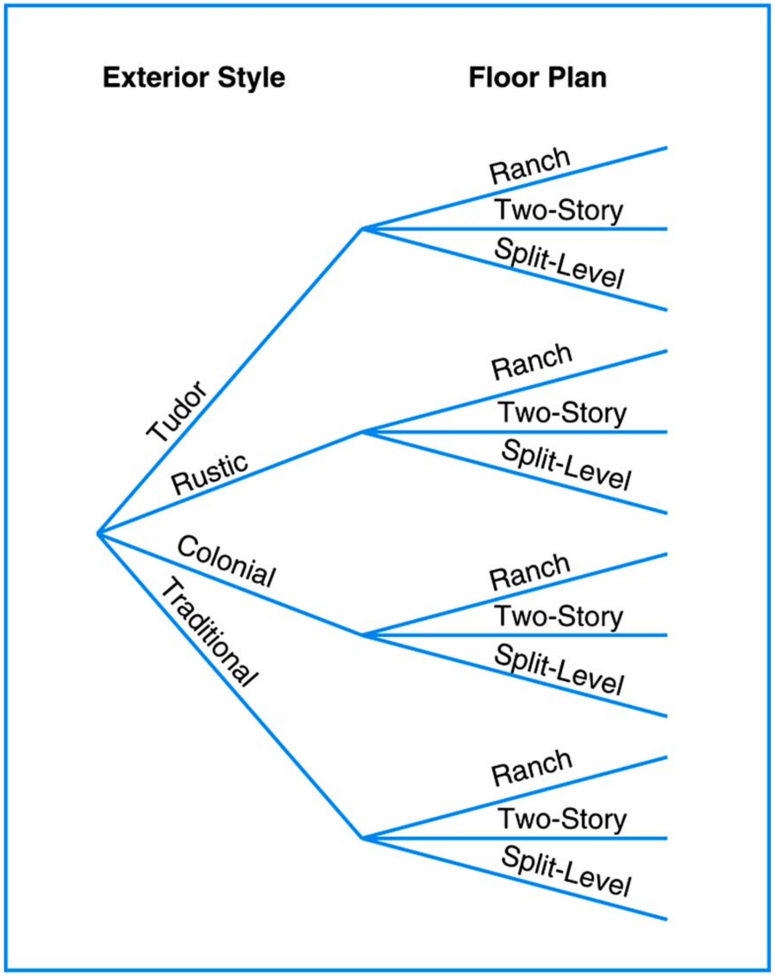
Rule 2.1

If an operation can be performed in n_1 ways, and if for each of these ways a second operation can be performed in n_2 ways, then the two operations can be performed together in $n_1 n_2$ ways.

Example 2.14

- A developer of a a new subdivision offers prospective home buyers a choice of Tudor, rustic, colonial, and traditional exterior styling in ranch, two-story, and split level floor plans.
- How many different choices does a buyer have?

Figure 2.6 Tree diagram for Example 2.14



Rule 2.2

If an operation can be performed in n_1 ways, and if for each of these a second operation can be performed in n_2 ways, and for each of the first two a third operation can be performed in n_3 ways, and so forth, then the sequence of k operations can be performed in $n_1 n_2 \cdots n_k$ ways.

Definition 2.7

A **permutation** is an arrangement of all or part of a set of objects.

Definition 2.8

For any non-negative integer n, $n!$, called "n factorial," is defined as

$$n! = n(n-1)\cdots(2)(1),$$

with special case $0! = 1$.

Theorem 2.1

The number of permutations of n objects is $n!$.

Theorem 2.2

The number of permutations of n distinct objects taken r at a time is

$$_nP_r = \frac{n!}{(n-r)!}.$$

Theorem 2.3

$\boxed{\text{The number of permutations of } n \text{ objects arranged in a circle is } (n-1)!.}$

Theorem 2.4

The number of distinct permutations of n things of which n_1 are of one kind, n_2 of a second kind, $\ldots$, n_k of a kth kind is

$$\frac{n!}{n_1!n_2!\cdots n_k!}.$$

Theorem 2.5

The number of ways of partitioning a set of n objects into r cells with n_1 elements in the first cell, n_2 elements in the second, and so forth, is

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

where $n_1 + n_2 + \cdots + n_r = n$.

Theorem 2.6

The number of combinations of n distinct objects taken r at a time is

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

Exercise 2.36

a) How many digit numbers can be formed from the digits 0, 1, 2, 3, 4, 5, and 6 if each digit can be used only once?

b) How many of these are odd numbers?

c) How many are greater than 330?

Exercise 2.43

- In how many ways can 5 different trees be planted in a circle?

Exercise 2.45

How many distinct permutations can be made from the letters of the word:

- OPEN
- BOOK?
- INFINITY?

Exercise 2.46

- In how many ways can 3 oaks, 4 pines, and 2 maples be arranged along a property line if one does not distinguish among trees of the same kind?

Exercise 2.47

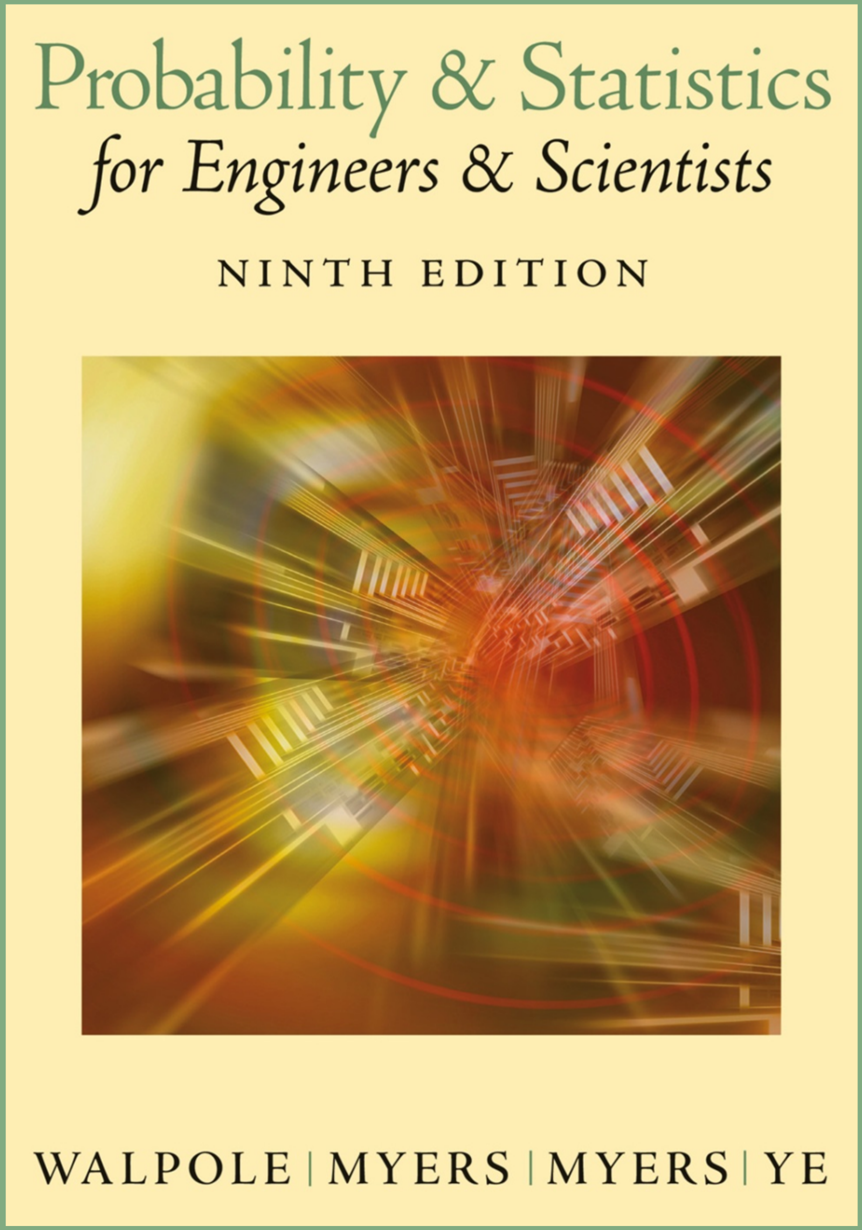
- How many ways are there to select 3 candidates from 8 equally qualified recent graduates for openings in an accounting firm?

Exercise 2.48

- How many ways are there hat no two students will have the same birth date in a class of size 60?

Section 2.4

Probability of an Event



Interpretations of Probability

Subjective probability: assigning a value based on one's belief of how likely a certain event (that can not be repeated)

- What is the chance that Joe will marry Zoe?
- What is the chance that the current government will stay until the end of its full term?

Interpretations of Probability

Frequentist's probability:

- Repeat an experiment a large number N of times
- Let n(A) be the number of times that event A was observed
- the chance of A is approximately n(A)/N

The law of large numbers says:

$$\lim n(A)/N = P[A]$$

Definition 2.9

The **probability** of an event A is the sum of the weights of all sample points in A. Therefore,

$$0 \leq P(A) \leq 1, \quad P(\phi) = 0, \quad \text{and} \quad P(S) = 1.$$

Furthermore, if $A_1, A_2, A_3, \ldots$ is a sequence of mutually exclusive events, then

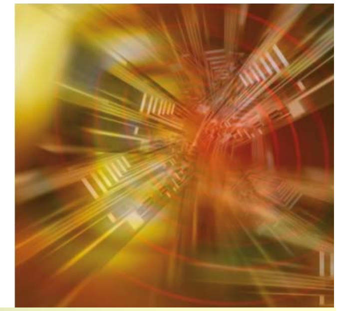
$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots .$$

Rule 2.3

If an experiment can result in any one of N different equally likely outcomes, and if exactly n of these outcomes correspond to event A, then the probability of event A is

$$P(A) = \frac{n}{N}.$$

Example 2.28



- In a poker hand (i.e. five cards taken out of a deck of 52 cards), find the probability of holding 2 aces and 3 jacks?

Example 2.28

- The number of ways we can choose 2 aces is ${}_4C_2$
- The number of ways we can choose 3 jacks is ${}_4C_3$
- The number of ways we can choose 3 jacks is ${}_{52}C_5$
- Hence the required answer is

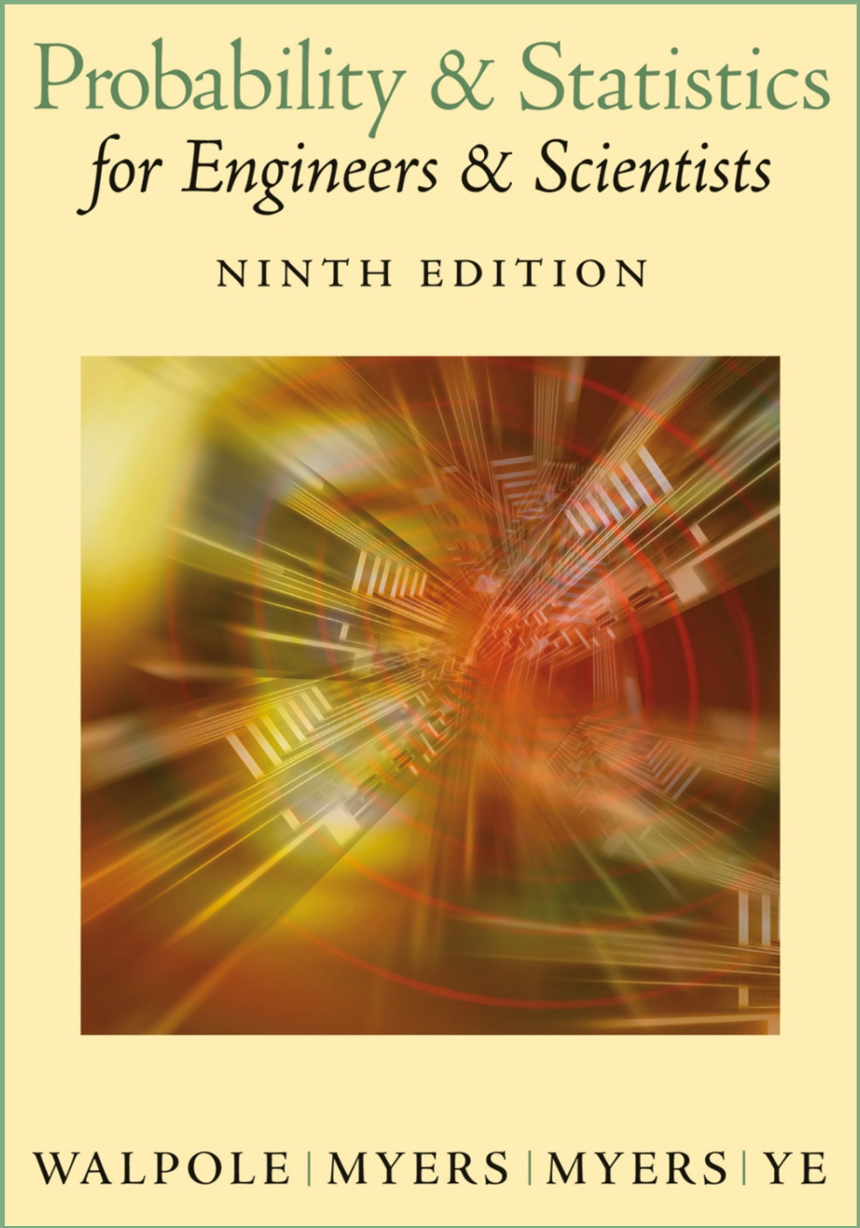
$$ {}_4C_2 * {}_4C_3 / {}_{52}C_5 $$

Exercise

- In a poker hand, find the probability of holding at least 2 aces?

Section 2.5

Additive Rules

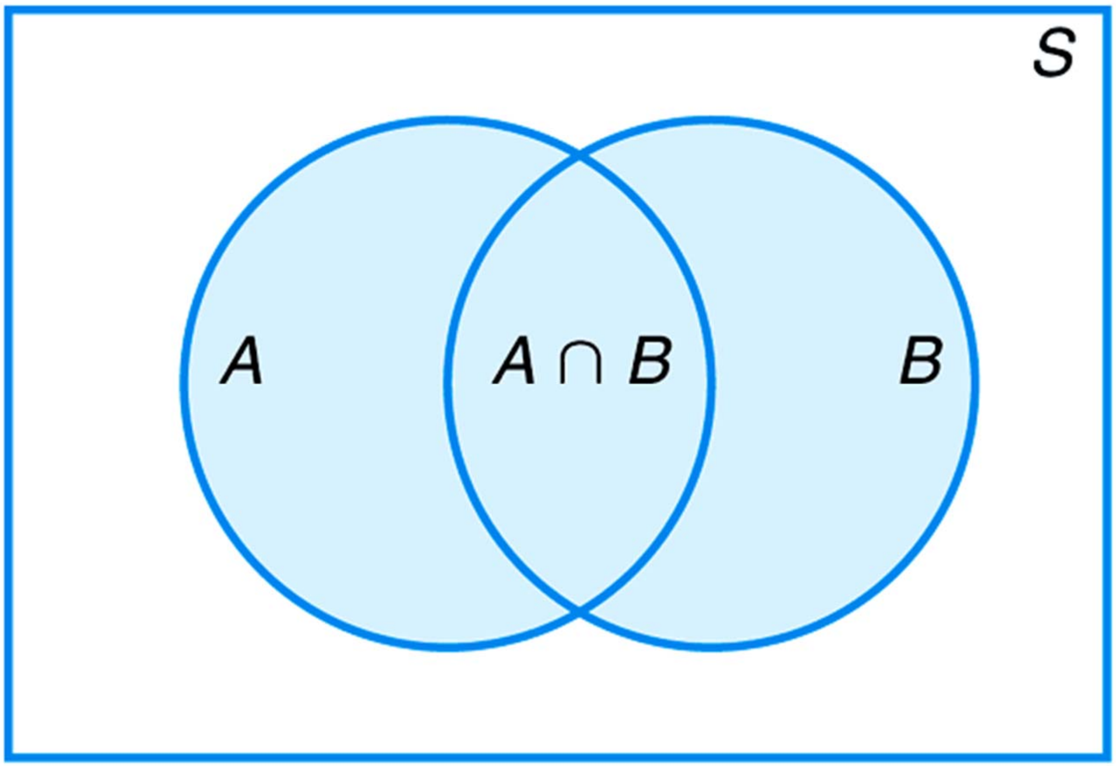


Theorem 2.7

If A and B are two events, then

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

Figure 2.7 Additive rule of probability



Corollary 2.1

If A and B are mutually exclusive, then

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

Corollary 2.2

If $A_1,\ A_2, \ldots, A_n$ are mutually exclusive, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n).$$

Corollary 2.3

If $A_1,\ A_2, \ldots, A_n$ is a partition of sample space S, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n) = P(S) = 1.$$

Theorem 2.8

For three events A, B, and C,

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

Theorem 2.9

If A and A' are complementary events, then

$$P(A) + P(A') = 1.$$

Exercise 2.52

- In a senior class of 500 students, 210 smoke, 258 drink alcoholic beverages, 216 eat between meals, 122 smoke and drink, 83 eat between meals and drink, 97 smoke and eat between meals. 52engage in all of these bad health practices.
  Find the probability that a randomly chosen senior student:

Exercise 2.52

- In a senior class of 500 students, 210 smoke, 258 drink alcoholic beverages, 216 eat between meals, 122 smoke and drink, 83 eat between meals and drink, 97 smoke and eat between meals. 52engage in all of these bad health practices.
  Find the probability that a randomly chosen senior student:

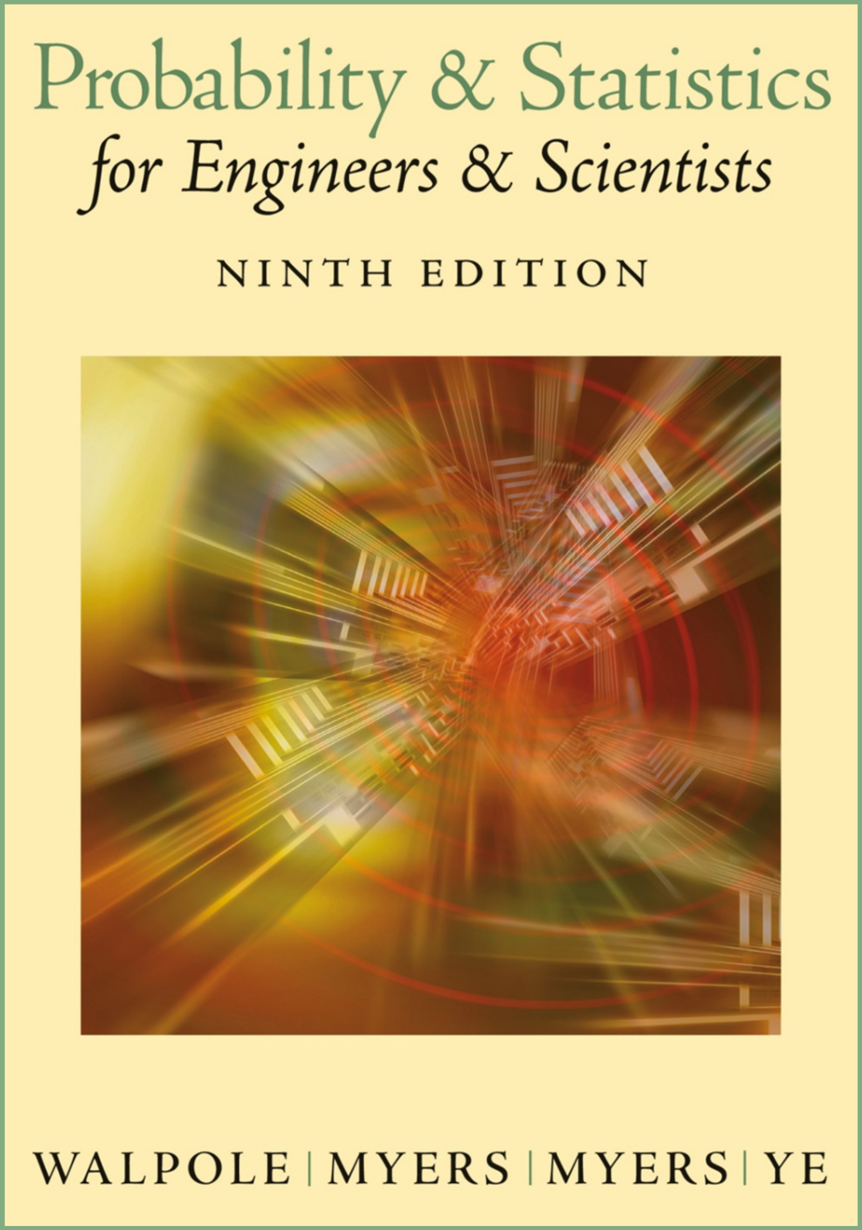
a) Smokes but does not drink

b) Eats between meals and drinks but does not smoke

c) Neither smokes nor eats between meals

Section 2.6

Conditional Probability, Independence, and the Product Rule



Definition 2.10

The conditional probability of B, given A, denoted by $P(B|A)$, is defined by

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

Explanation of Def. 2.10

- Knowing that A is given, we know that the outcome that was observed must be in A
- It follows that the new "reduced" sample space is A (any outcome outside A becomes impossible)
- Then for B to occur, the outcome must be in both A and B
- The division by P(A) must be done so that $P(A|A) = 1$

Table 2.1 Categorization of the Adults in a Small Town

	Employed	Unemployed	Total
Male	460	40	500
Female	140	260	400
Total	600	300	900

Definition 2.11

Two events A and B are **independent** if and only if

$$P(B|A) = P(B) \quad \text{or} \quad P(A|B) = P(A),$$

assuming the existences of the conditional probabilities. Otherwise, A and B are **dependent**.

Theorem 2.10

If in an experiment the events A and B can both occur, then

$$P(A \cap B) = P(A)P(B|A), \text{ provided } P(A) > 0.$$

Example 2.37

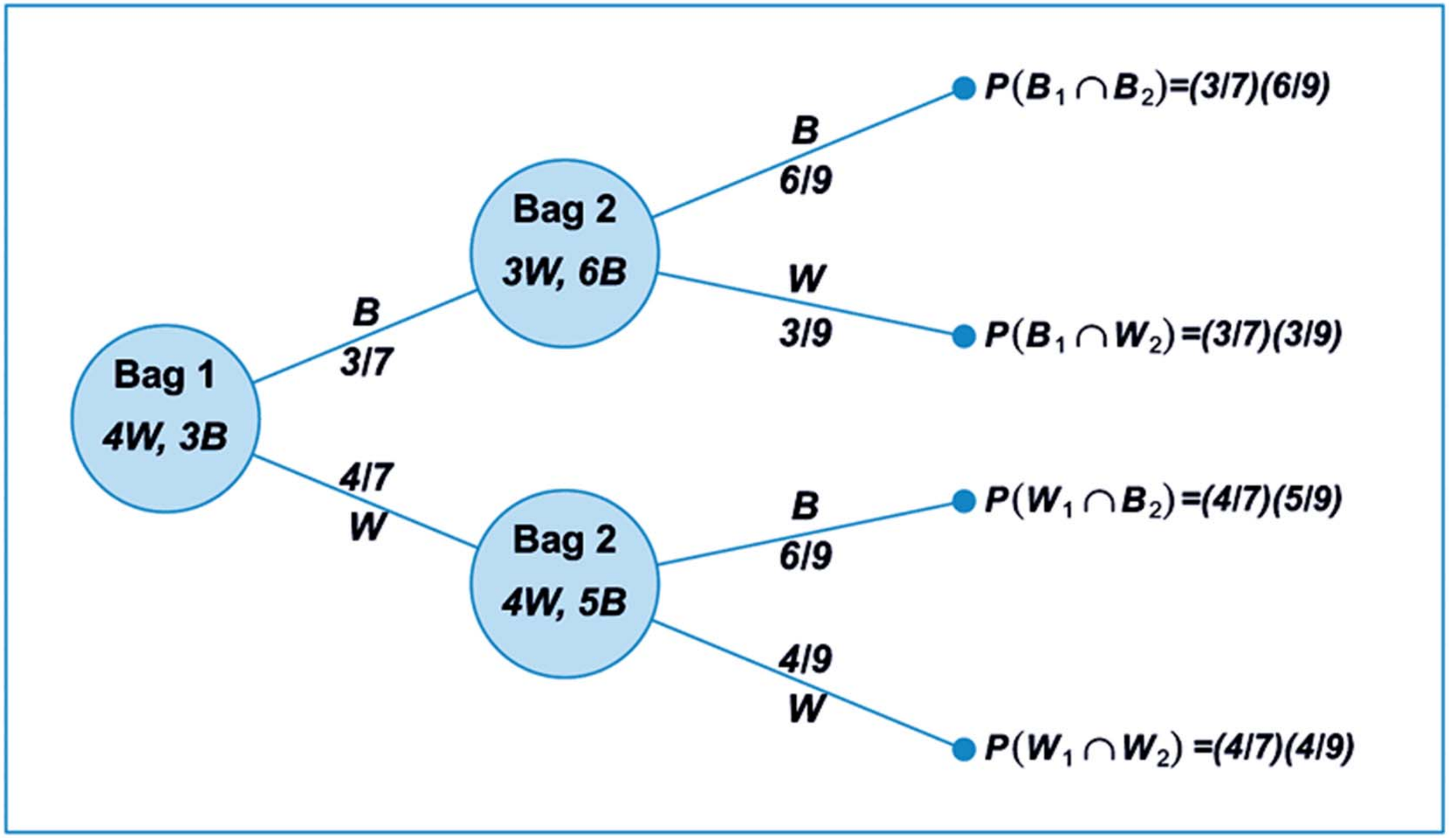
- One bag contains 4 white balls and 3 black balls, and a second bag contains 3 white balls and 5 black balls. One ball is drawn from the first ball and placed <u>unseen</u> in the second bag. What is the probability that a ball now drawn from the second bag is black?

Example 2.37

- One bag contains 4 white balls and 3 black balls, and a second bag contains 3 white balls and 5 black balls. One ball is drawn from the first ball and placed <u>unseen</u> in the second bag. What is the probability that a ball now drawn from the second bag is black?

- Define the events
  B1: a black ball is drawn from bag 1
  B2: a black ball is drawn from bag 2
  W1: a white ball is drawn from bag 1
- We want $P[B2]= P[B1 \cap B2]+P[W1 \cap B2]$

Figure 2.8 Tree diagram for Example 2.37



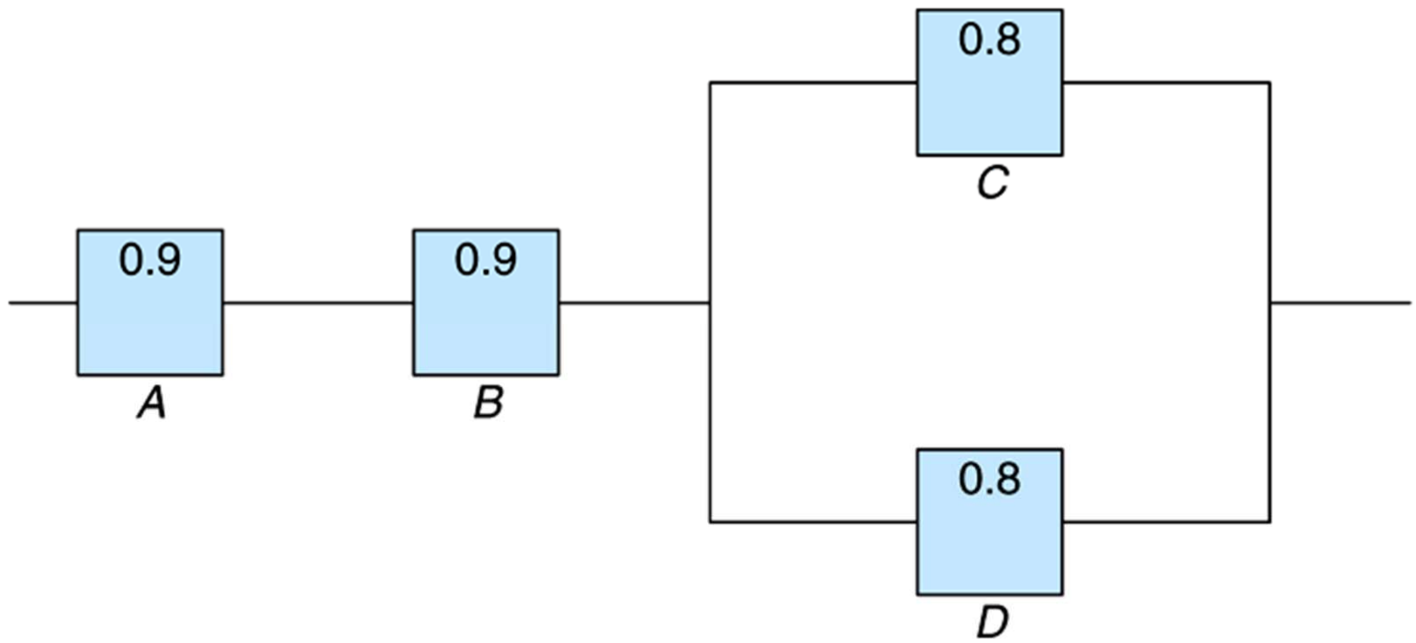
Theorem 2.11

Two events A and B are independent if and only if

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

Therefore, to obtain the probability that two independent events will both occur, we simply find the product of their individual probabilities.

Figure 2.9 An electrical system for Example 2.39



Theorem 2.12

If, in an experiment, the events $A_1, A_2, \ldots, A_k$ can occur, then

$$P(A_1 \cap A_2 \cap \cdots \cap A_k)$$
$$= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_k|A_1 \cap A_2 \cap \cdots \cap A_{k-1}).$$

If the events $A_1, A_2, \ldots, A_k$ are independent, then

$$P(A_1 \cap A_2 \cap \cdots \cap A_k) = P(A_1)P(A_2) \cdots P(A_k).$$

Definition 2.12

A collection of events $\mathcal{A} = \{A_1, \ldots, A_n\}$ are mutually independent if for any subset of $\mathcal{A}$, $A_{i_1}, \ldots, A_{i_k}$, for $k \leq n$, we have

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}).$$

A counting Problem

- How many equations are needed to hold so that n events are mutually independent?

A counting Problem

- How many equations are needed to hold so that n events are mutually independent?

Answer:

${}_nC_2$ equations to check that every two events are independent.

${}_nC_3$ equations to check that every three events are independent.

etc.

${}_nC_n = 1$ to check the equation involving all events.

We get ${}_nC_2 + {}_nC_3 + \ldots + {}_nC_n$

Exercise 2.86

- In 1970, 11% of Americans completed 4 years of college; 43% of them were women. In 1990, 22% of Americans completed 4 years of college; 53% of them were women.

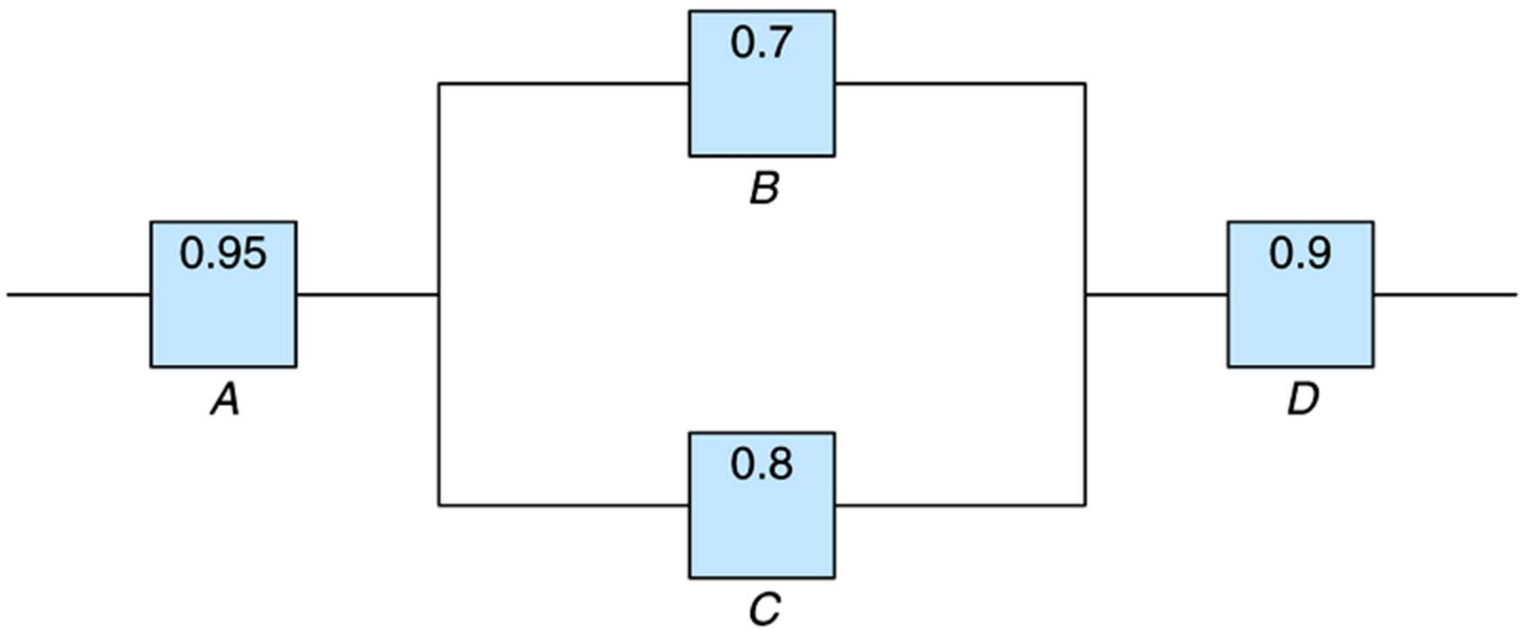
a) Given that a person completed 4 years of college, what is the probability that the person was a woman?

b) What is the probability that a woman finished 4 years of college in 1990?

Exercise 2.92

- An electrical system is given in the diagram below. Assuming that components function independently, find the probability that the system works.

Figure 2.10 Diagram for Exercise 2.92



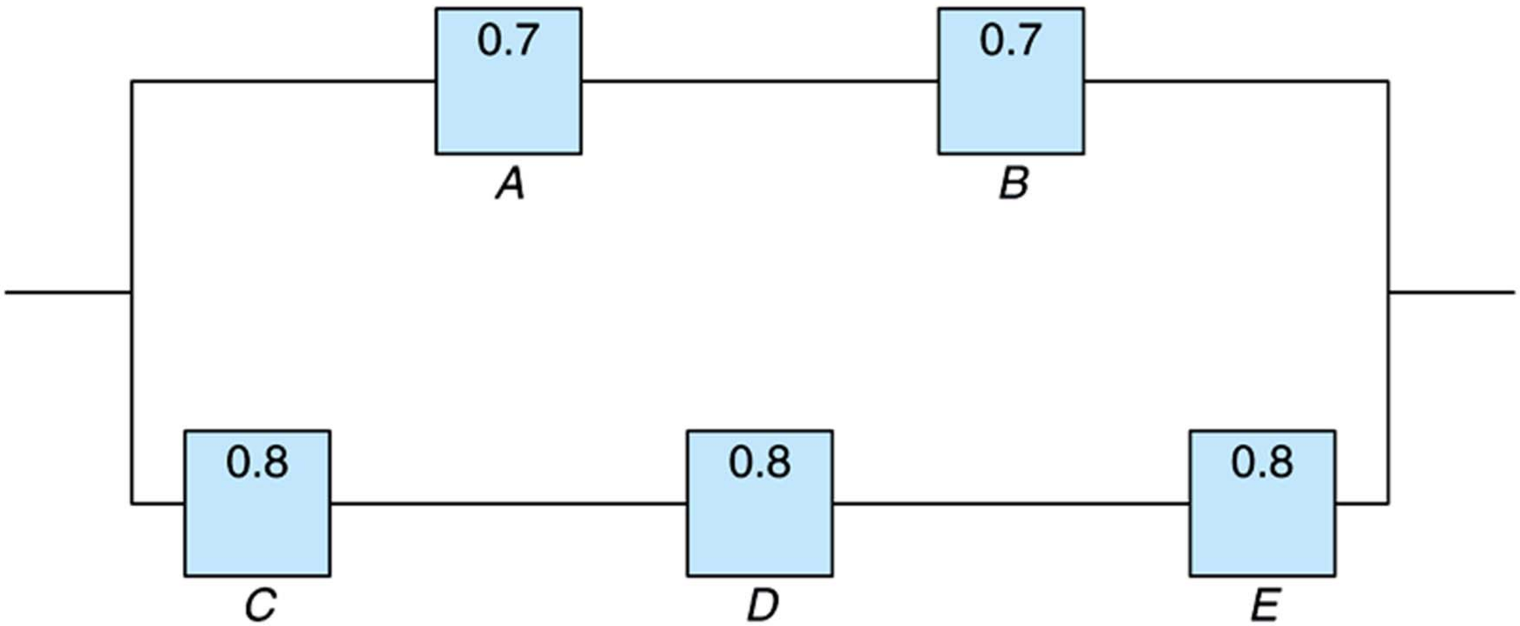
Exercise 2.93

- An electrical system is given in the diagram below. Assuming that components function independently.

a) Find the probability that the system works.

b) Given that the system works, find the probability that that component A is not working.

Figure 2.11 Diagram for Exercise 2.93



Section 2.7

Bayes' Rule

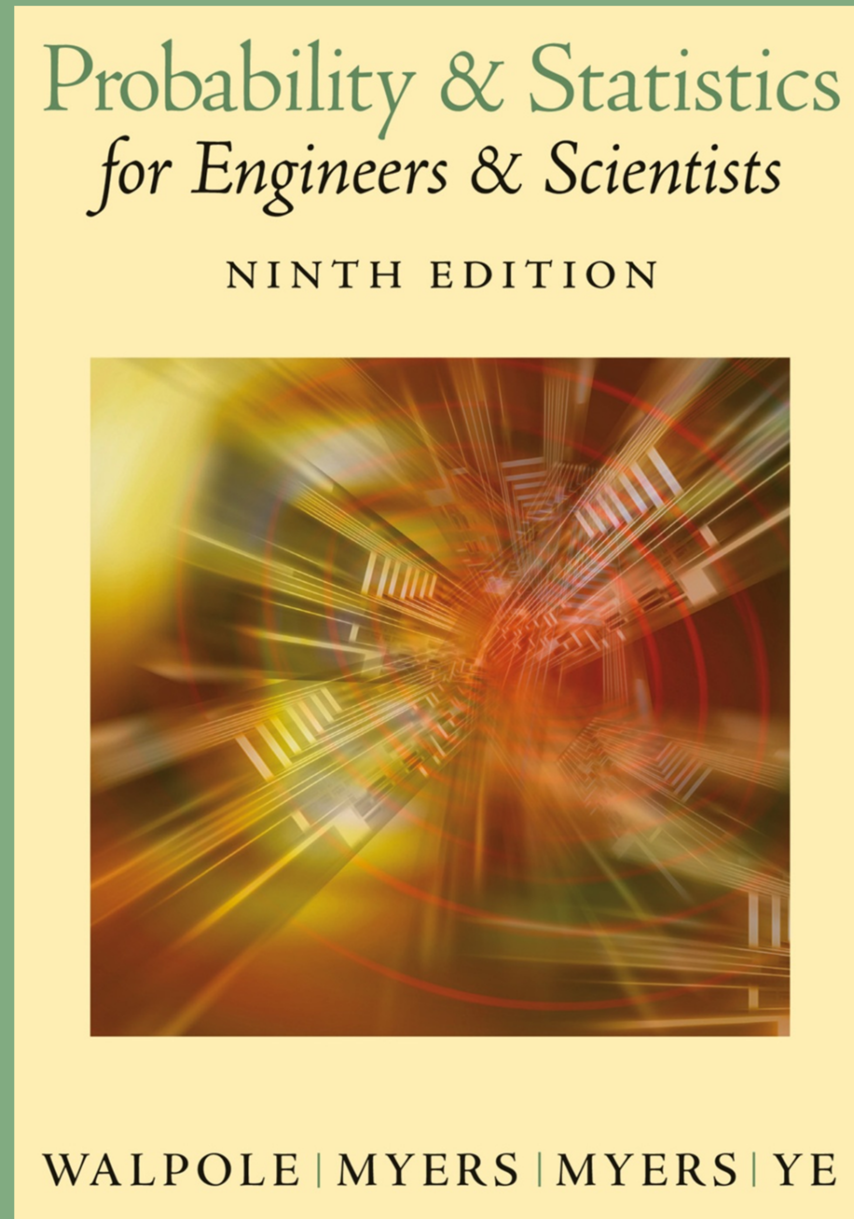


Figure 2.12 Venn diagram for the events *A*, *E* and *E*´

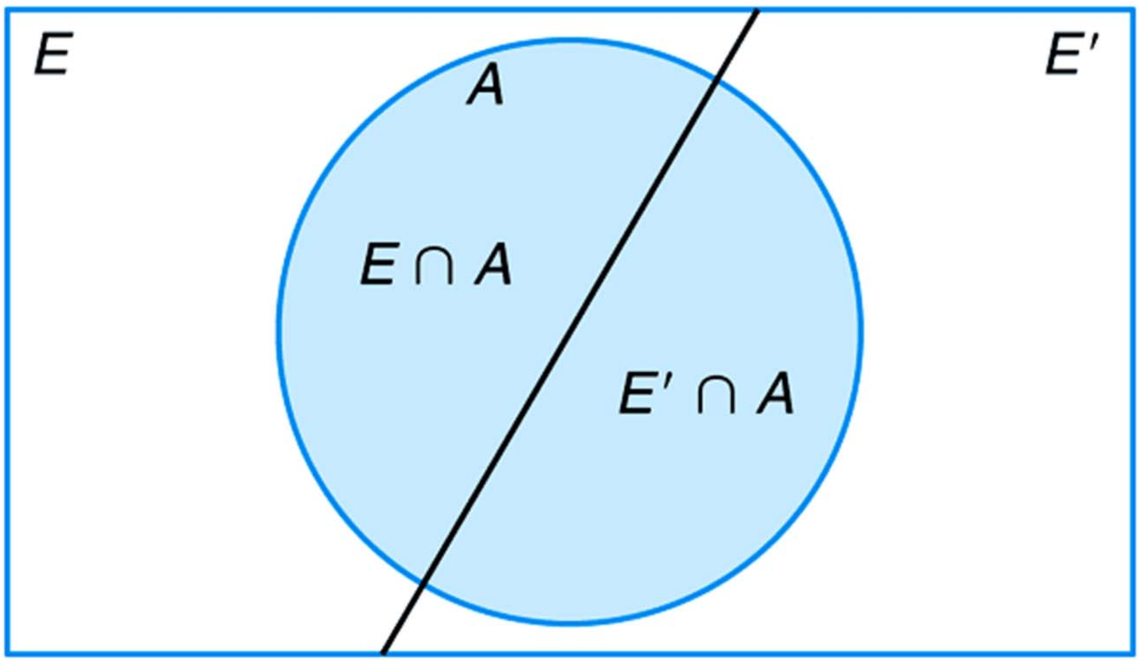
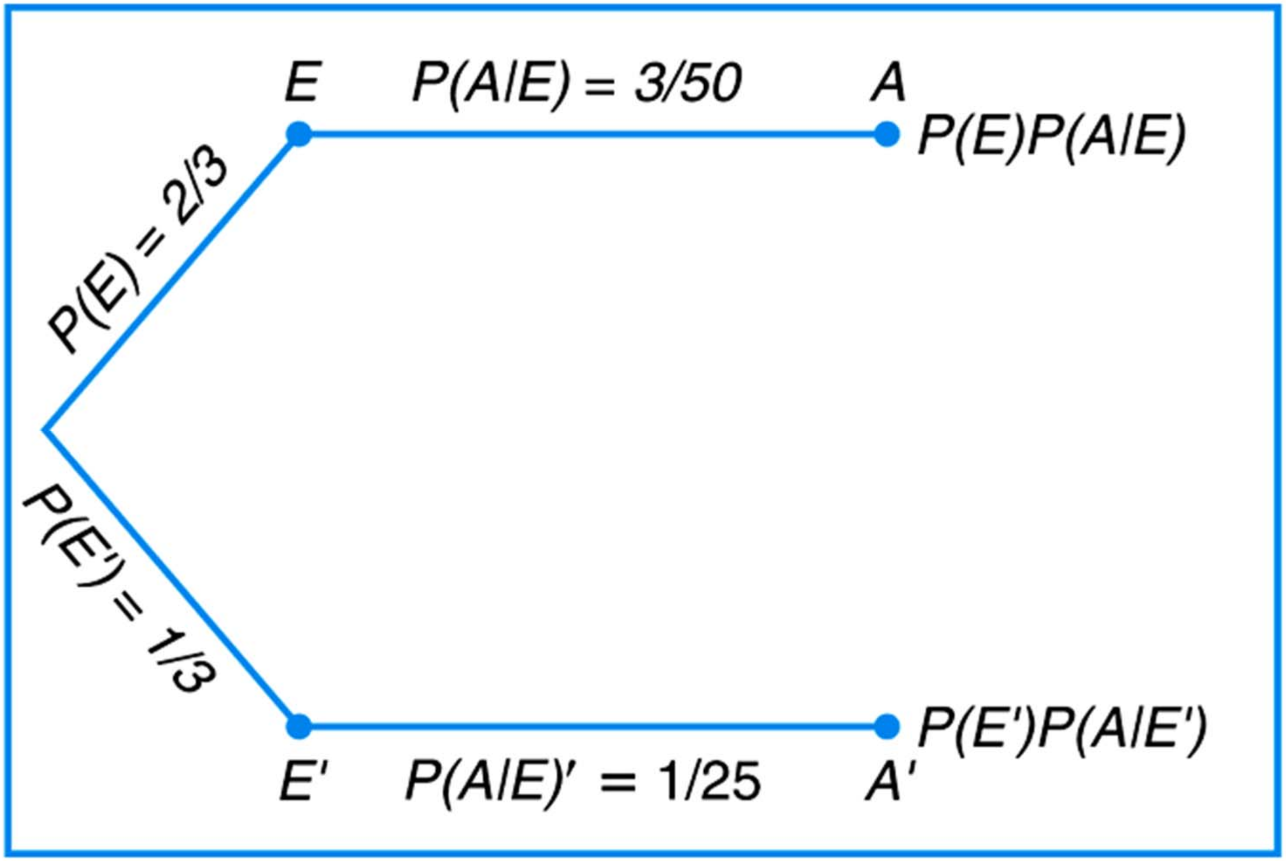


Figure 2.13 Tree diagram for the data on page 63, using additional information on page 72



Theorem 2.13

If the events $B_1, B_2, \ldots, B_k$ constitute a partition of the sample space S such that $P(B_i) \neq 0$ for $i = 1, 2, \ldots, k$, then for any event A of S,

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

Figure 2.14 Partitioning the sample space *S*

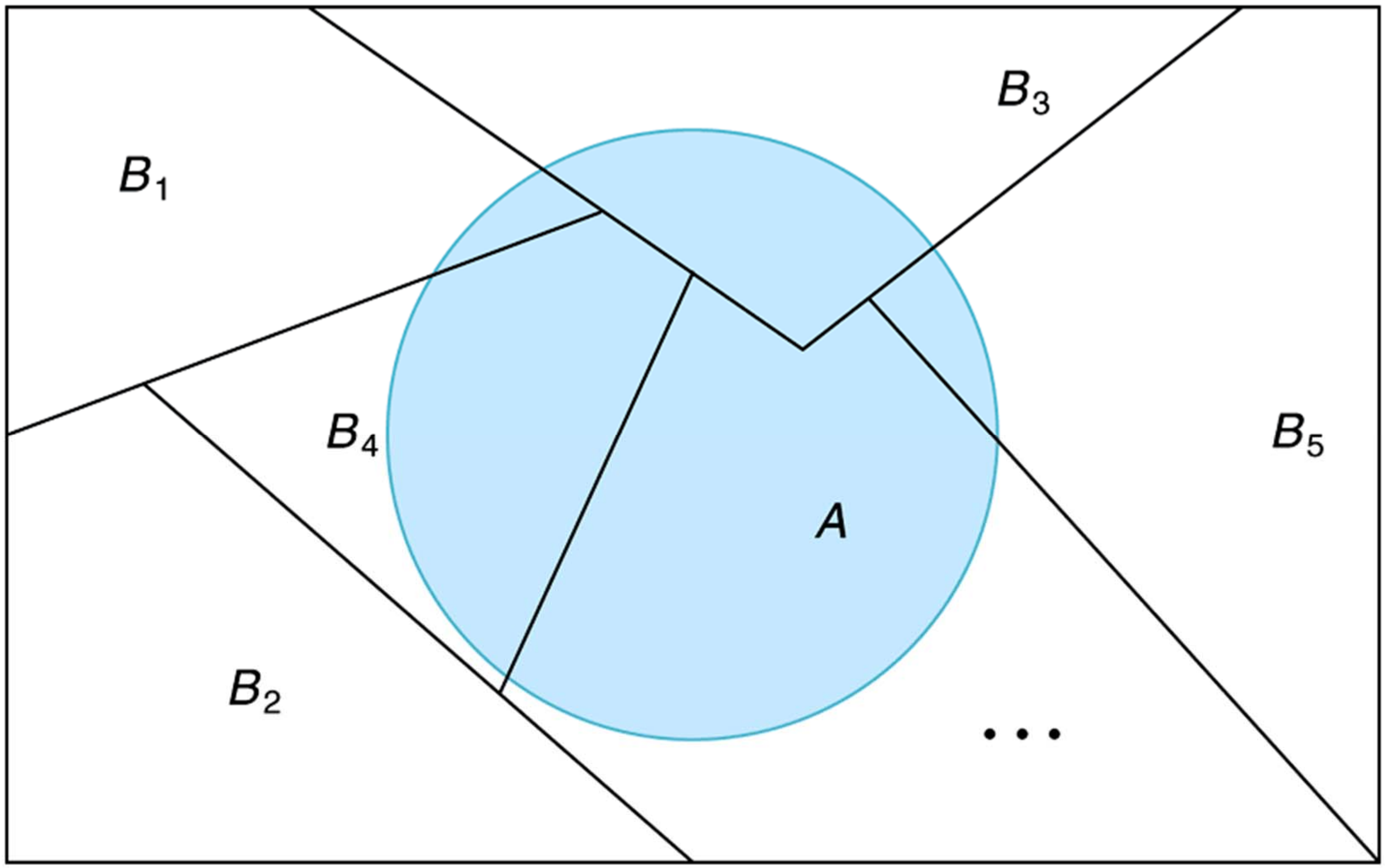
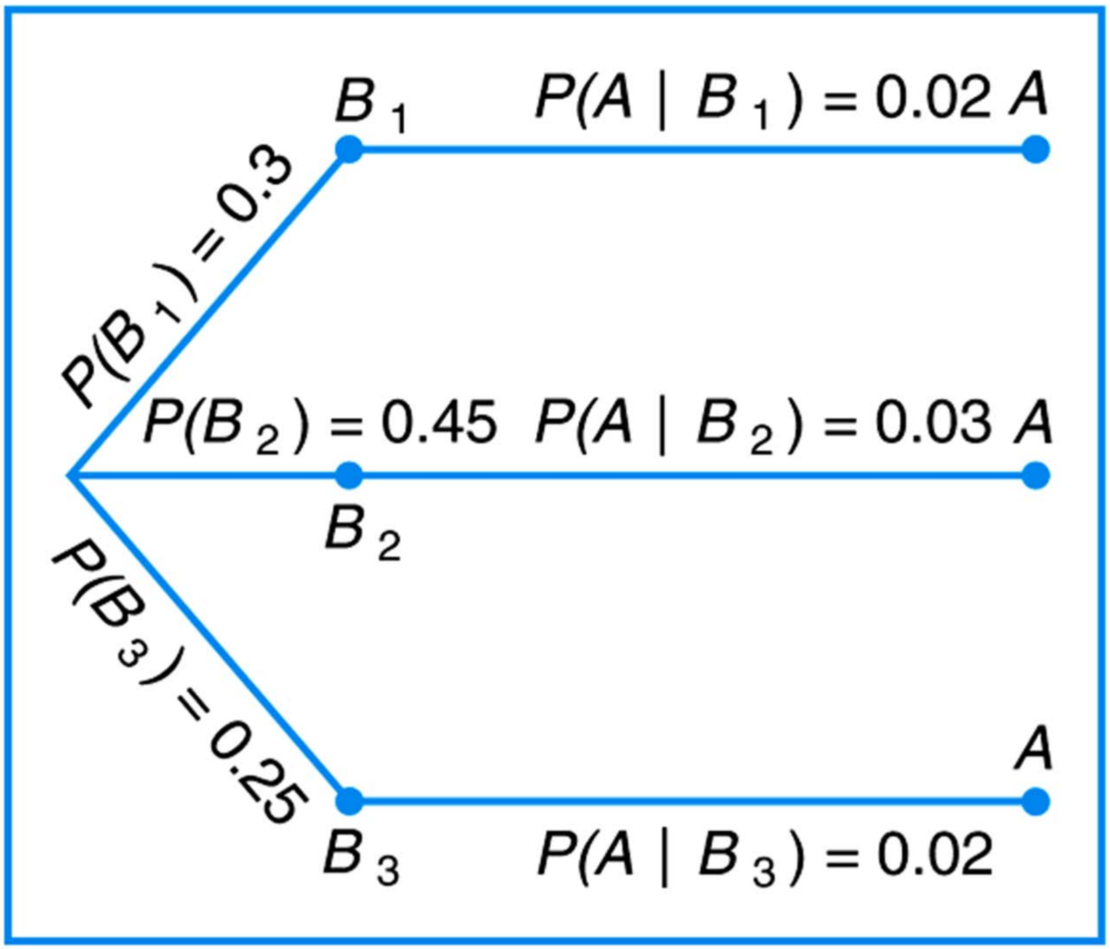


Figure 2.15 Tree diagram for Example 2.41



Theorem 2.14

(Bayes' Rule) If the events $B_1, B_2, \ldots, B_k$ constitute a partition of the sample space S such that $P(B_i) \neq 0$ for $i = 1, 2, \ldots, k$, then for any event A in S such that $P(A) \neq 0$,

$$P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^{k} P(B_i)P(A|B_i)} \quad \text{for } r = 1, 2, \ldots, k.$$

Example 1

Marbles

A box contains 10 marbles of which exactly 6 are red. Two marbles are drawn one by one *without replacement*.

What is the probability that the second marble is red?

Example 1 Solution

Marbles

What is the probability that the second marble is red?

Two cases for first marble:
case1: red or
case2: not red

P(case1) = 6/10 = 0.6

P(case2) = 4/10 = 0.4

P(second red | case1) = 5/9 = 0.5555

P(second red | case2) = 6/9 = 0.6666

Example 1 Solution

Marbles

What is the probability that the second marble is red?

So P(second is red) =

P(case1) P(second is red | case 1)
+ P(case2) P(second is red | case 2)

= (6/10)(5/9) + (4/10)(6/9) = (30+24)/90

= 54 /90

= 6/10

Example 2

More marbles

- Box 1 contains 4 green and 6 red marbles.
- Box 2 contains 7 green and 3 red marbles.
- Draw one marble randomly from box 1; and place it in box 2 without looking at it.
- Now draw a marble from box 2. What is the probability that this second marble is red?

Example 3

Example 1 Revisited

- In the first example, what is the probability that the first was red, if we know that the second marble was red?

Example 3

Example 1 Revisited

- In the first example, what is the probability that the first was red, if we know that the second marble was red?

Answer:

P(first is red | second is red)

P(first red AND second red) / P(second is red)

Example 3

Example 1 Revisited

- In the first example, what is the probability that the first was red, if we know that the second marble was red?

Answer:

P(first is red | second is red)

P(first red AND second red) / P(second is red)

Example 4

Genetic Disease

- A rare genetic disease is discovered. Although only one in a million people carry it, you consider getting screened. You are told that the genetic test is extremely good; it is 99% sensitive (i.e. It detects the disease correctly if you have the disease) and 95% specific (it gives a false positive result only 5% of the time).
- Suppose you take the test and it comes out positive, what is the chance that you actually have the disease?

Answer to Example 4

We want P[Disease | Test is positive]

We have: $P[D]=10^{-6}$; $P[D^c]= 1- 10^{-6}$

$P[\ T+ \mid D] = 0.99$; {correct decision}

$P[\ T+ \mid D^c] = 0.05$; {false positive}

So $P[T+] = P[D]P[\ T+ \mid D] + P[D^c]P[\ T+ \mid D^c]$

$= 0.05000099$

and $P[D \mid T+] = 0.00000099/0.05000094 =$ $19.7*10^{-6}$. This is about 20 times begger than the prior assessment

Answer to Example 4

So $P[T+] = P[D]P[\ T+ \mid D] + P[D^c]P[\ T+ \mid D^c]$

$= 0.05000099$

Therefore

$P[D \mid T+] = 0.00000099/0.05000094 = 19.7*10^{-6.}$

Eventhough this probability is still tiny, it is about 20 times bigger than the prior assessment of the chance of having the disease! Further action must be taken..