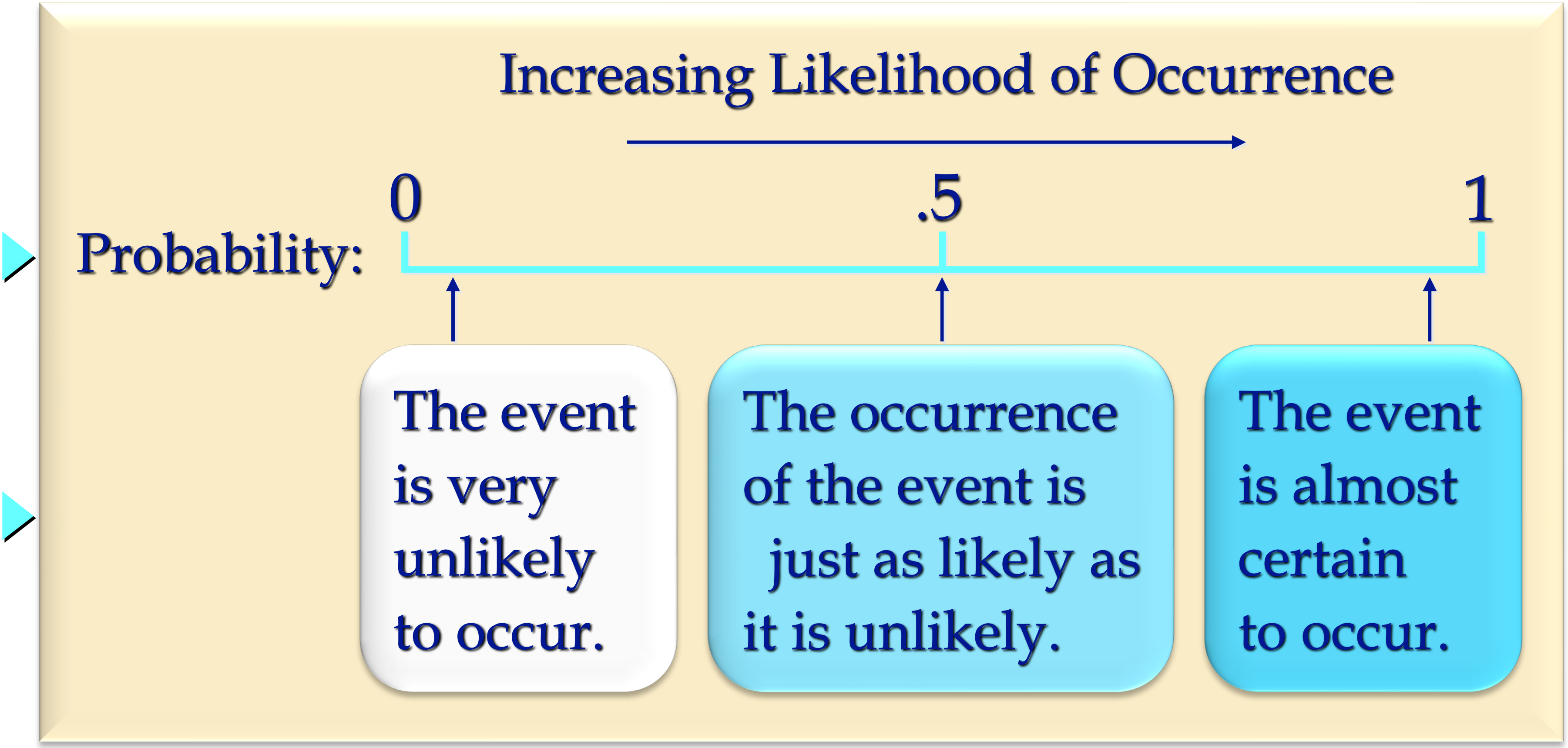
統計學

Fall 2025

政治大學統計系余清祥
2025年9月23日
第四章：機率論
http://csyue.nccu.edu.tw

Chapter Contents

Probability as a Numerical Measure of the Likelihood of Occurrence



4.1 Random Experiment and Sample Space

A **Random experiment** is a process that generates *experimental outcomes*, and possesses the following properties (*see notes):

1. The experimental outcomes are well-defined and may also be listed prior to conducting the experiment.
2. On any single repetition (*trial*) of the experiment, one and only one of the possible experimental outcomes will occur.
3. The experimental outcome that occurs on any trial is determined solely by chance.

An experimental outcome is also called a **sample point**. The **sample space**, *S*, for an experiment is the set of all experimental outcomes.

Example: consider the random experiment of tossing a coin.

- The experimental outcomes (sample points) are Head and Tail.
- The sample space is: $S = \{\text{Head}, \text{Tail}\}$

An Experiment and Its Sample Space

Experiment	Experiment Outcomes
Toss a coin	Head, tail
Inspection a part	Defective, non-defective
Conduct a sales call	Purchase, no purchase
Roll a die	1, 2, 3, 4, 5, 6
Play a football game	Win, lose, tie

4.1 Counting Rule for Multiple-Step Experiments

A **multiple-step experiment** can be described as a sequence of *k* steps with n_1 possible outcomes on the first step, n_2 possible outcomes on the second step, and so on.

The total number of experimental outcomes is given by $(\boldsymbol{n_1})(\boldsymbol{n_2})\ldots(\boldsymbol{n_k})$.

As an example, consider the random experiment of tossing two coins, one at a time.

The first tossed coin has $n_1 = 2$ outcomes, and the second tossed coin has also $n_2 = 2$ outcomes.

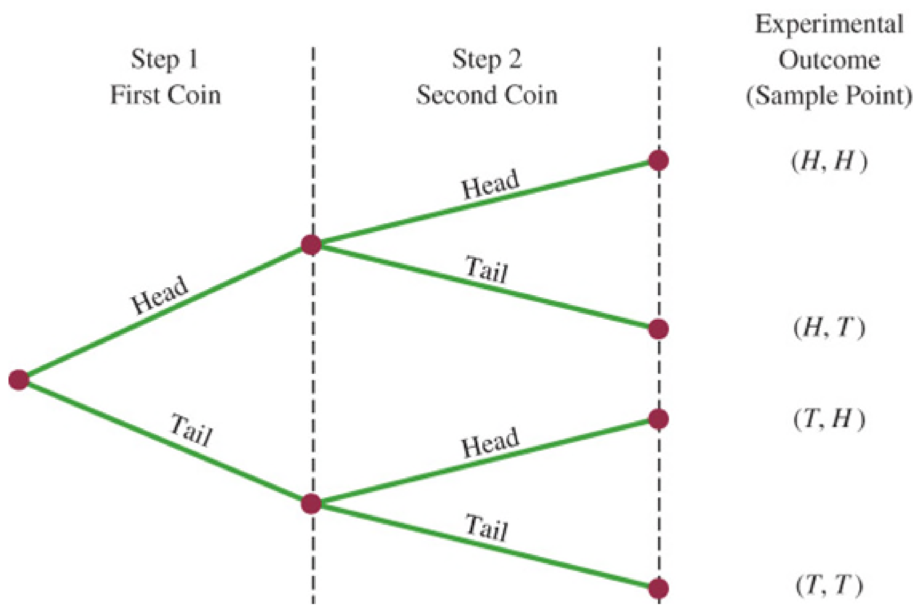
From the counting rule, we have:

$(n_1)(n_2) = (2)(2) = 4$ experimental outcomes

The sample space is :

$S = \{(\mathrm{H,H}),(\mathrm{H,T}),(\mathrm{T,H}),(\mathrm{T,T})\}$

The **tree diagram** for the two-coin toss shown to the right is a graphical representation that helps visualizing a multiple-step experiment.



CENGAGE

4.1 The KP&L Project: a Multiple-Step Experiment

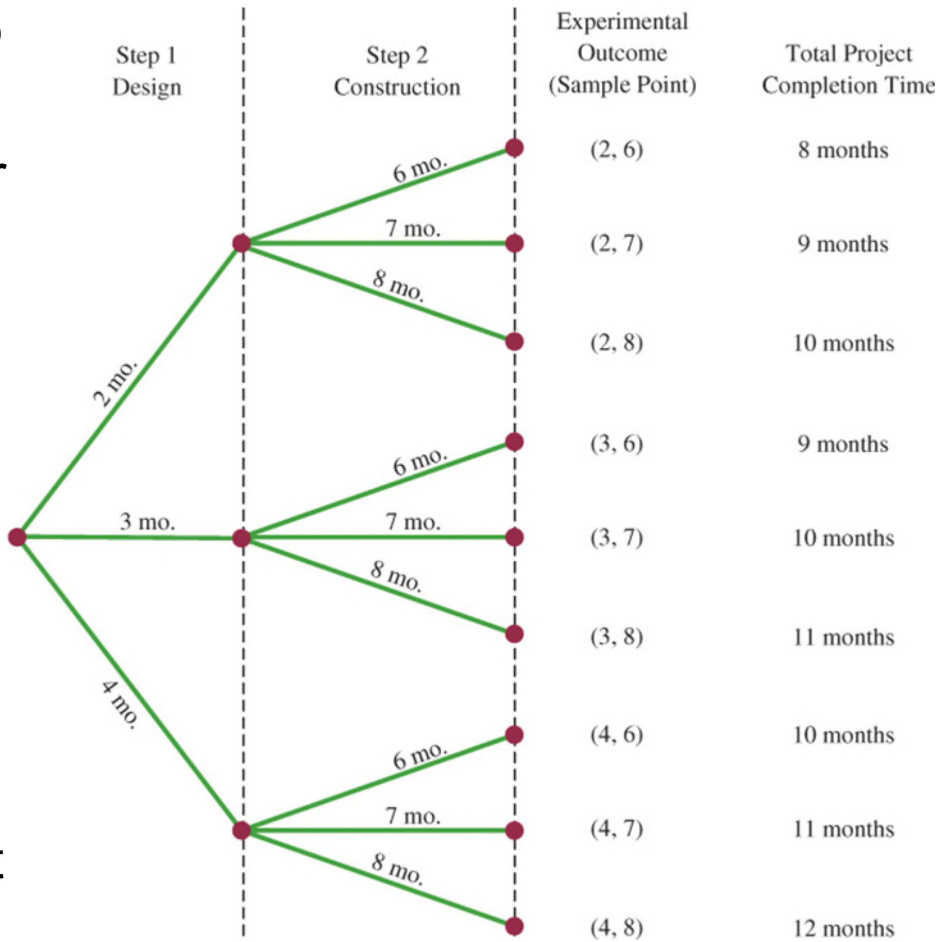
The KP&L capacity expansion project consists of two sequential stages: design and construction.

Management expects the design stage to last 2, 3, or 4 months, the construction stage 6, 7, or 8 months, with an overall completion goal of 10 months.

Application of the counting rule for multiple-step experiments reveals a total of (3)(3) = 9 experimental outcomes.

The tree diagram to the right shows how the 9 outcomes (sample points) occur and the project completion time for each outcome.

We need to assign probabilities to each of the 9 outcomes before we can compute the probability that the project is completed on time.



An Experiment and Its Sample Space

Example: Bradley Investments

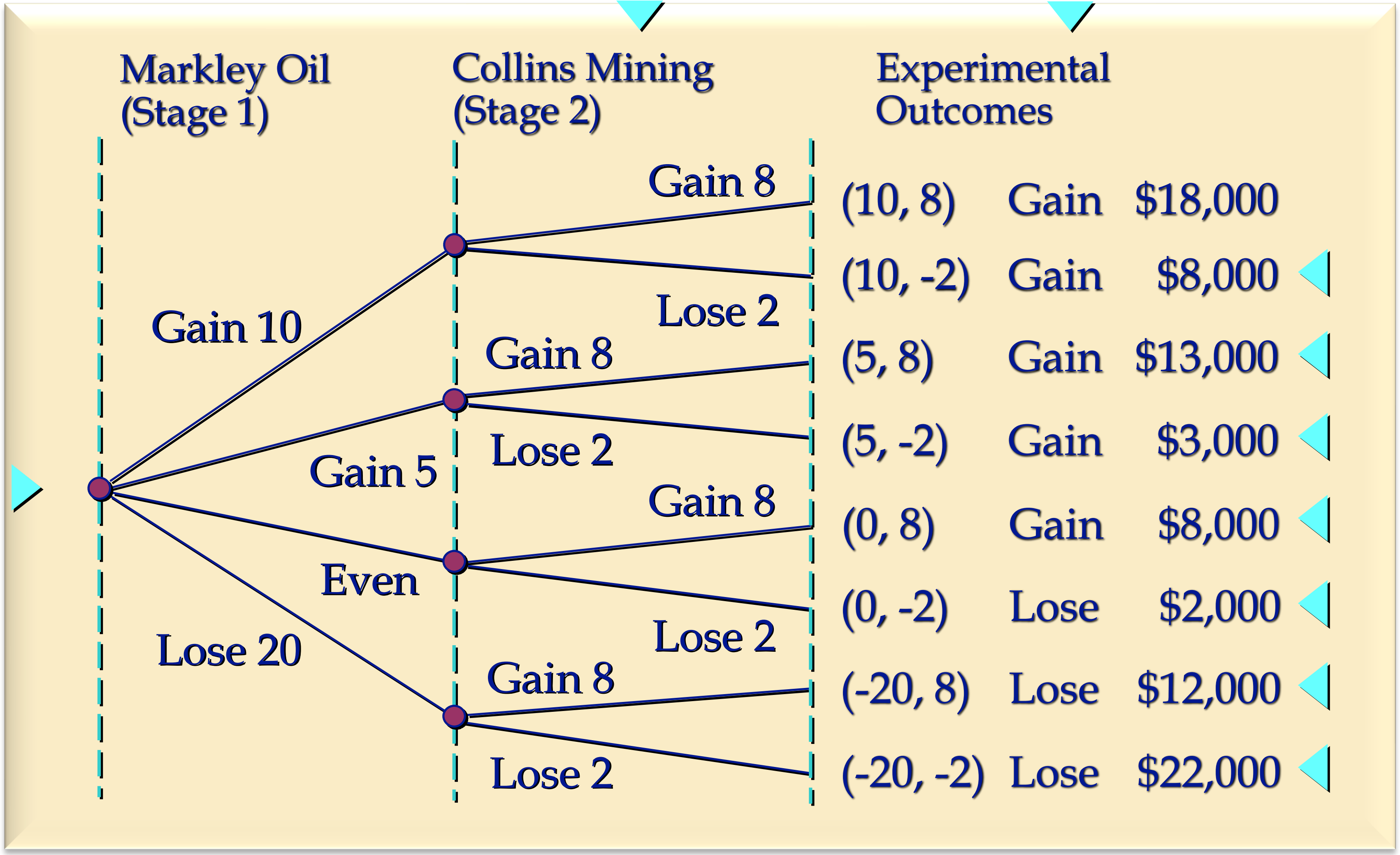
Bradley has invested in two stocks, Markley Oil and Collins Mining. Bradley has determined that the possible outcomes of these investments three months from now are as follows.

Investment Gain or Loss in 3 Months (in $000)

Markley Oil	Collins Mining
10	8
5	–2
0	
–20	

Tree Diagram

- Example: Bradley Investments



CENGAGE

4.1 Counting Rule for Combinations

The **counting rule for combinations** dictates that the number of combinations of *n* objects taken *x* at a time is

$$\boldsymbol{C_x^n = \binom{n}{x} = \frac{n!}{x!\,(n - x!)}}$$

Example: consider a quality control in which an inspector randomly selects two of five parts to test for defects. In a group of five parts, how many combinations of two parts may be selected?

The counting rule for combinations shows

$$C_2^5 = \binom{5}{2} = \frac{5!}{2!\,(5 - 2!)} = \frac{(5)(4)(3)(2)(1)}{(2)(1)(3)(2)(1)} = \frac{120}{12} = 10 \text{ outcomes}$$

If we label the five parts A, B, C, D, and E, the sample space of the combinations can be represented as

$$S = \{(\mathrm{A, B}),(\mathrm{A, C}), (\mathrm{A, D}), (\mathrm{A, E}),(\mathrm{B, C}), (\mathrm{B, D}), (\mathrm{B, E}), (\mathrm{C, D}), (\mathrm{C, E}), (\mathrm{D, E})\}$$

CENGAGE

4.1 Counting Rule for Permutations

When the order of selection matters, we can use the **counting rule for permutations**.

The number of permutations of *n* objects taken *x* at a time is

$$\boldsymbol{P_x^n = \frac{n!}{(n - x!)}}$$

Example: consider again the inspector's quality control process. In a group of five parts, how many permutations of two parts may be selected?

The counting rule for permutations shows

$$P_2^5 = \frac{5!}{(5 - 2!)} = \frac{5!}{3!} = \frac{(5)(4)(3)(2)(1)}{(3)(2)(1)} = \frac{120}{6} = 20 \text{ outcomes}$$

The counting rule for permutations is closely related to the one for combinations because every selection of x objects can be ordered in $x!$ different ways.

CENGAGE

4.1 Assigning Probabilities

The basic requirements for assigning probabilities are

1. The probability assigned to each experimental outcome must be between 0 and 1, inclusively.

   $$\mathbf{0 \leq P(E_i) \leq 1}$$

   where E_i is the ith experimental outcome and $P(E_i)$ its probability.

2. The sum of the probabilities for all experimental outcomes must equal 1.

   $$\boldsymbol{P(E_1) + P(E_2) + \cdots P(E_n) = 1}$$

   where n is the number of experimental outcomes.

There are three methods of assigning probabilities:

- The **classical method**, is based on the assumption of equally likely outcomes.
- The **empirical method**, based on relative frequencies from experimental or historical data.
- The **subjective method**, based on experience or intuition when no data are available.

4.1 Assigning Probabilities for the KP&L Project

Using experience and judgment, KP&L management concluded that not all the outcomes were equally likely. Thus, the classical method could not be used.

A study on the completion times of 40 similar projects undertaken over the past three years revealed the frequency distribution shown.

Using the empirical method, we can assign the probabilities to each of the nine outcomes.

For example, sample point (2, 6) on the top row, with an expected completion time of 8 months, has probability:

$$P(2,6) = 6/40 = 0.15$$

Completion Time (months)					
Stage 1 Design	Stage 2 Design	Sample Point	Number of Past Projects	Project Completion Time (months)	Probability of Sample Point
2	6	(2, 6)	6	8	$P(2, 6) = 6/40 = 0.15$
2	7	(2, 7)	6	9	$P(2, 7) = 6/40 = 0.15$
2	8	(2, 8)	2	10	$P(2, 8) = 2/40 = 0.05$
3	6	(3, 6)	4	9	$P(3, 6) = 4/40 = 0.10$
3	7	(3, 7)	8	10	$P(3, 7) = 8/40 = 0.20$
3	8	(3, 8)	2	11	$P(3, 8) = 2/40 = 0.05$
4	6	(4, 6)	2	10	$P(4, 6) = 2/40 = 0.05$
4	7	(4, 7)	4	11	$P(4, 7) = 4/40 = 0.10$
4	8	(4, 8)	6	12	$P(4, 8) = 6/40 = 0.15$
			40		

4.2 Events

An **event** is a collection of sample points.

If we can identify all the sample points of an experiment and assign a probability to each, we can compute the probability of an event.

In the KP&L example, management was interested in the probability that the project is completed in 10 months or less.

From the table in the previous slide, we see that there are six sample points that provide a project completion time of 10 months or less.

Let *C* denote the event that the project is completed in 10 months or less. Thus, we have:

$$C = \{(2, 6),(2, 7), (2, 8), (3, 6),(3, 7), (4, 6)\}$$

Other events of interest may be:

L, the event that the project will complete in less than 10 months: $L = \{(2, 6),(2, 7), (3, 6)\}$

M, the event that the project will complete in more than 10 months: $M = \{(3, 8),(4, 7), (4, 8)\}$

4.2 Probability of an Event

The **probability of an event** is equal to the sum of the probabilities of the sample points in the event.

We can calculate the probability that the KP&L project is completed in 10 months or less by adding the probabilities of the six sample points belonging to event *C*.

$$P(C) = P(2,6) + P(2,7) + P(2,8) + P(3,6) + P(3,7) + P(4,6)$$

If we refer to the sample point probabilities listed in the previous table, we have

$$P(C) = 0.15 + 0.15 + 0.05 + 0.10 + 0.20 + 0.05 = 0.7$$

Similarly, the probability that the KP&L project is completed in less than 10 months is

$$P(L) = P(2,6) + P(2,7) + P(3,6) = 0.15 + 0.15 + 0.10 = 0.4$$

And the probability that the KP&L project is completed in more than 10 months is

$$P(M) = P(3,8) + P(4,7) + P(4,8) = 0.05 + 0.10 + 0.15 = 0.3$$

4.3 Some Basic Relationships of Probability

In an experiment with a large number of sample points, the identification of all the sample points, as well as the determination of their associated probabilities, can become an extremely cumbersome task.

In this section, we present some basic probability relationships that can be used to compute the probability of an event without knowledge of all the sample point probabilities.

These basic relationships of probability are:

- The complement of an event
- The union of two events
- The intersection of two events
- The addition law
- The multiplication law

4.3 Complement of an Event

The **complement of event $\boldsymbol{A}$** is defined as the event consisting of all sample points that are *not* in A.

The complement of A is denoted by A^C.

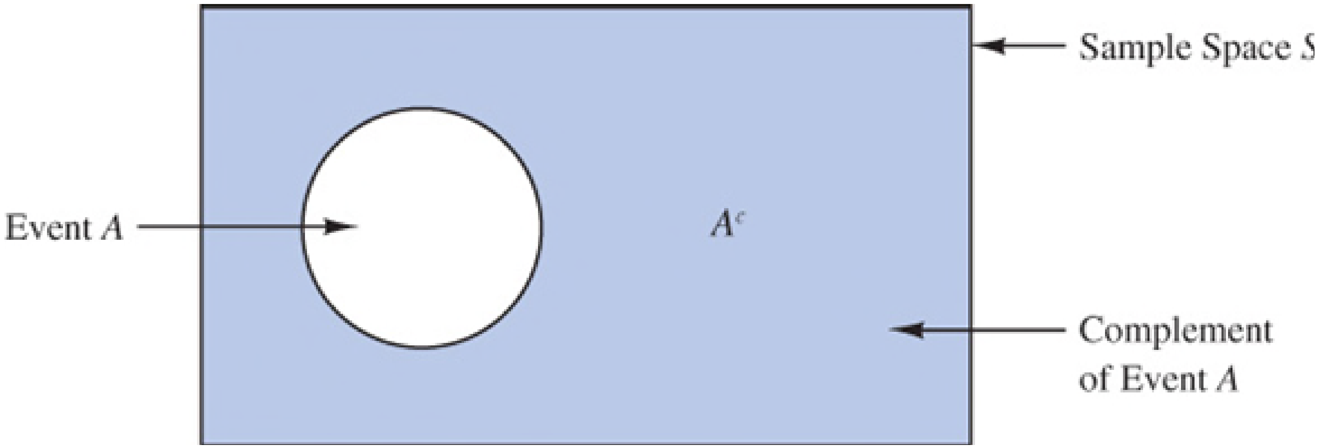
In any probability application, either event A or its complement A^C must occur, as shown in the Venn diagram to the right.

Therefore, we have

$$P(A) + P(A^C) = 1$$

Solving for $P(A)$, we obtain the formula for the probability of the complement of an event

$$\boldsymbol{P(A) = 1 - P\left(A^C\right)}$$

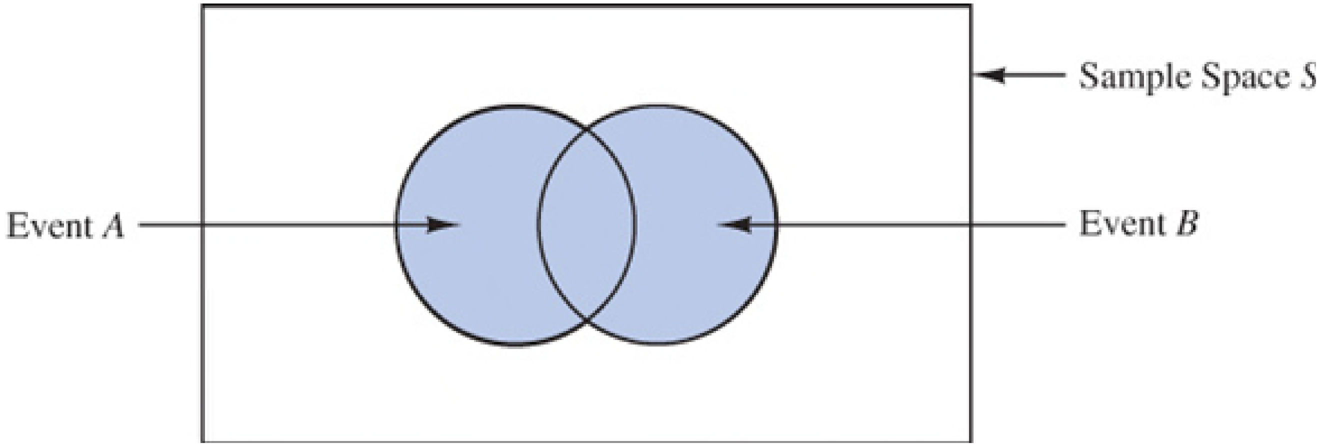


4.3 Union of Two Events

The **union of events A and B** is the event containing all sample points that are in A and B or both.

The union of events A and B is denoted by $A \cup B$.

Note that the two circles in the Venn diagram overlap, indicating that some sample points are contained in both A and B.



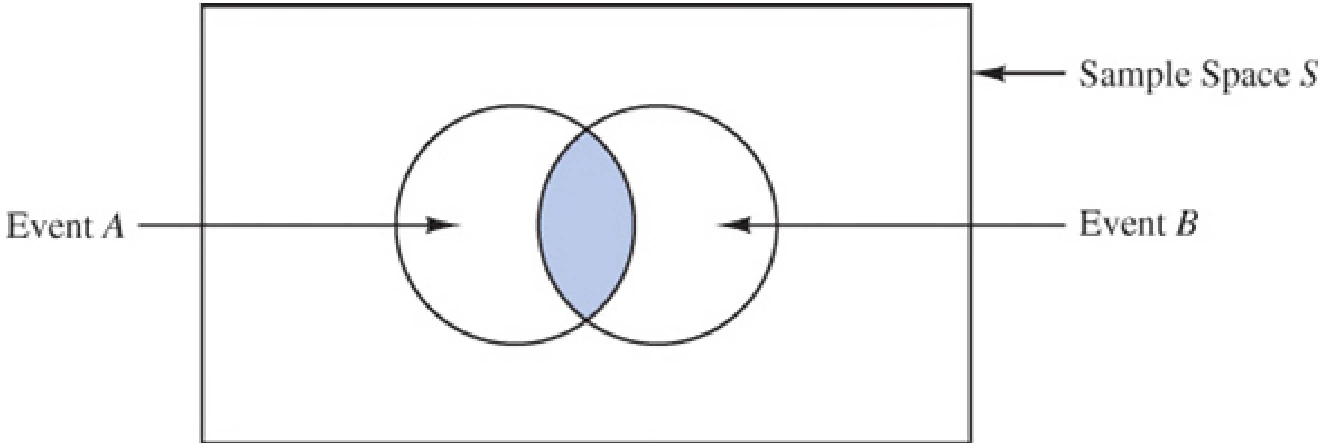
4.3 Intersection of Two Events

The **intersection of events $\boldsymbol{A}$ and $\boldsymbol{B}$** is the event containing all sample points belonging to both A and B.

The intersection of events A and B is denoted by $A \cap B$.

Note that the area in the Venn diagram where the two circles overlap is the intersection.

The intersection contains all the sample points that are in both A and B.



4.3 Addition Law

The **addition law** provides a way to compute the probability of the union of two events:

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

To understand the addition law intuitively, consider the following:

- The sample points in the intersection $A \cap B$ are in both A and B.
- When we compute $P(A) + P(B)$, we are in effect counting each of the sample points in $A \cap B$ twice.
- We correct for this overcounting by subtracting $P(A \cap B)$.

In the next slide, we consider an example as an application of the addition law.

4.3 An Application of the Addition Law

Consider a group of 50 software engineers who work in online banking.

At the end of an evaluation period, a bank manager found that

- 5 engineers completed work late (event *L*)
- 6 engineers produced code that contains errors (event *E*)
- 2 engineers completed work late and produced code that contains errors (event $L \cap E$)

Thus, we have:

$$P(L) = 5/50 = 0.10$$

$$P(E) = 6/50 = 0.12$$

$$P(L \cap E) = 2/50 = 0.04$$

The probability that an engineer completed work late or produced code that contains errors (event $L \cup E$) can be calculated using the addition law as

$$\mathrm{P}(L \cup E) = P(L) + P(E) - P(L \cap E) = 0.10 + 0.12 - 0.04 = 0.18$$

CENGAGE

4.3 Mutually Exclusive Events

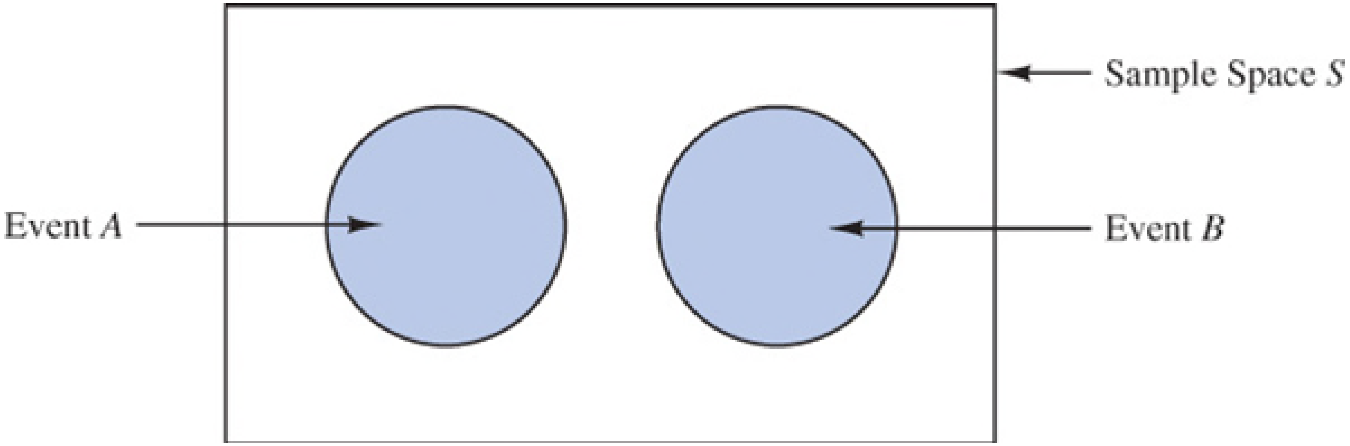
Two events are said to be **mutually exclusive** when they have no sample points in common.

In other words, events A and B are mutually exclusive if, when one event occurs, the other cannot occur.

Thus, the intersection of A and B does not contain any points, and $P(A \cap B) = 0$.

When two events are mutually exclusive, the addition law simplifies to

$$\boldsymbol{P(A \cup B) = P(A) + P(B)}$$

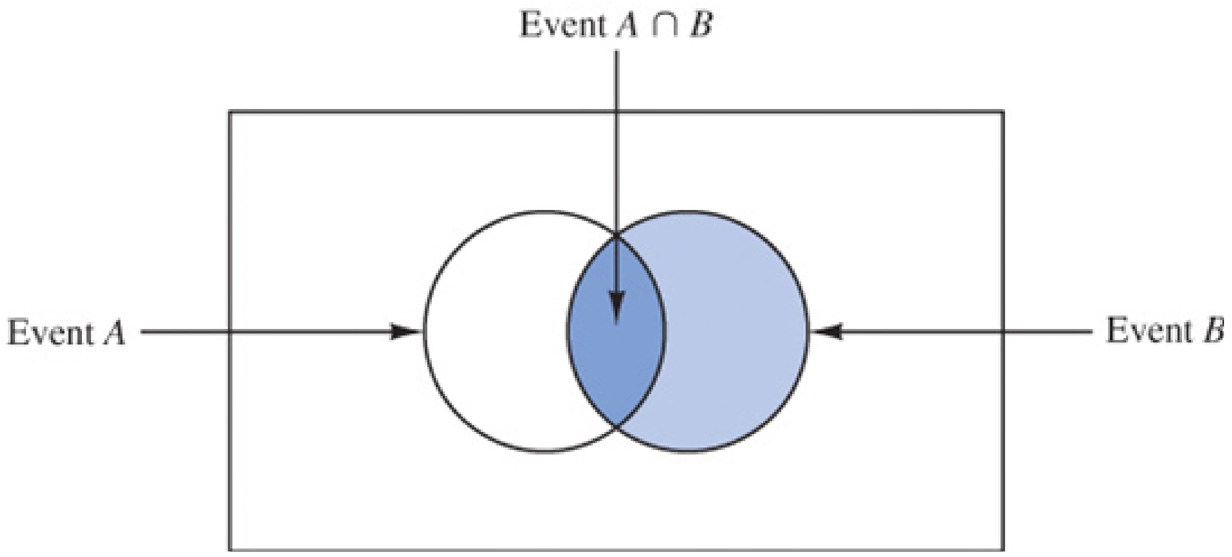


4.4 Conditional Probability

We use the notation $P(A \mid B)$ to denote the **conditional probability** of event A, given that event B has occurred.

We calculate the probability of event A given B as

$$\boldsymbol{P(A \mid B) = \frac{P(A \cap B)}{P(B)}}$$



We calculate the probability of event *B* given *A* as

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

Note that

$$P(B \mid A) \neq P(A \mid B)$$

4.4 Application: A Discrimination Case on the Promotion of Police Officers

The police force of a major metropolitan area in the eastern U.S. consists of 1200 officers, 960 self-identified males, and 240 self-identified females.

The specific breakdown of the promotions of 288 male and 36 female police officers over the past two years is summarized in the crosstabulation shown below.

After reviewing the promotion record, a committee of female officers raised a discrimination case on the basis of the low promotion of female officers when compared to male officers.

The police administration argued that the lower number of promotions for female officers was due to the relatively lower number of female officers in the police force.

	Male	Female	Total
Promoted	288	36	324
Not Promoted	672	204	876
Total	960	240	1,200

CENGAGE

4.4 Joint Probability Table

Dividing the data values in the crosstabulation by the total of 1,200 officers enables us to summarize the available information on promotions with a **joint probability table**.

In the body of the table, we have the **joint probabilities**, which are the intersection probabilities of the events of whether a police officer was promoted (A) or not (A^c) with whether the police officer is male (M) or female (F).

The joint probability that an officer was promoted (A) and is a male (M)

$$P(A \cap M) = 288/1200 = 0.24$$

On the margins of the table, we have the **marginal probabilities**, providing the probabilities of each separate event.

The probability that an officer is male (M)

$$P(M) = 960/1200 = 0.80$$

	Male (*M*)	**Female (*F*)**	**Total**
Promoted (*A*)	0.24	0.03	0.27
Not Promoted (A^c)	0.56	0.17	0.73
Total	0.80	0.20	1.00

CENGAGE

4.4 Conditional Probability Analysis

We note that the marginal probabilities are found by summing the joint probabilities in the corresponding row or column of the joint probability table. For example

$$P(A) = P(A \cap M) + P(A \cap F) = 0.24 + 0.03 = 0.27$$

The conditional probability that an officer was promoted (*A*) given that the officer is male (*M*)

$$P(A \mid M) = \frac{P(A \cap M)}{P(M)} = \frac{288/1200}{960/1200} = \frac{288}{960} = 0.30$$

Note that this could have also been calculated directly from the corresponding values in the crosstabulation using the conditional distribution (column) for *M* as $P(A \mid M) = 288/960 = 0.30$.

The conditional probability that an officer was promoted (*A*) given that the officer is female (*F*)

$$P(A \mid F) = \frac{P(A \cap F)}{P(F)} = \frac{36/1200}{240/1200} = \frac{36}{240} = 0.15$$

Thus, because $P(A \mid M) > P(A \mid F)$, the analysis supports the female officers' argument.

Simpson's Paradox in University admission

UC Berkeley admitted 44% of males and 35% of females who applied in 1973. Data from the six largest departments.

Department	Male acceptance rate	Female acceptance rate
A	62%	82%
B	63%	68%
C	37%	34%
D	33%	35%
E	28%	24%
F	6%	7%

	Male		Female	
	Applicants	%	Applicants	%
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	373	6%	341	7%

CE

4.4 Independent Events

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

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

Otherwise, the events are *dependent* if

$$P(A \mid B) \neq P(A) \quad \text{or} \quad P(B \mid A) \neq P(B)$$

In the analysis of the promotion of police officers, we found

$$P(A) = 0.27,\ P(A \mid M) = 0.30, \text{ and } P(A \mid F) = 0.15$$

Because $P(A \mid M) \neq P(A)$, that is, the probability of whether an officer is promoted (A) is altered or affected by knowing whether the officer is male (M), we conclude that the events A and M are dependent.

Similarly, because $P(A \mid F) \neq P(A)$, we would say that events A and F are also dependent (*see notes.)

4.4 Multiplication Law

The **multiplication law** is based on the definition of conditional probability, and it provides a way to compute the probability of the intersection of two events

$$\boldsymbol{P(A \cap B) = P(A \mid B)P(B)}$$

or

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

In the case of independent events, $P(A \mid B) = P(A)$ or $P(B \mid A) = P(B)$, and the **multiplication law for independent events** becomes

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

or

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

Therefore, two events *A* and *B* are dependent when $P(A \cap B) \neq P(A)P(B)$.

4.5 Bayes' Theorem

The **Bayes' theorem (two-event case)** states that

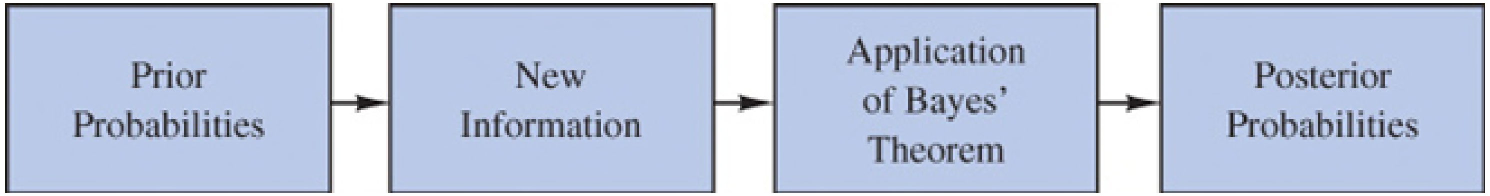
$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2)} \quad \text{with } i = 1,2$$

Where:

$P(A_1)$ and $P(A_2)$ are the initial or **prior probabilities**, available at the beginning.

$P(B \mid A_1)$ and $P(B \mid A_2)$ are conditional probabilities calculated from a sample, special report, or product test.

Bayes' theorem provides a mean to calculate the revised or **posterior probabilities** $P(A_1 \mid B)$ and $P(A_2 \mid B)$.



4.5 An Application of Bayes' Theorem

Consider a manufacturing firm that purchases 65% of its parts from supplier 1 and 35% from supplier 2.

If we let A_1 denote that event that a part is from supplier 1, and A_2 the event that the part is from supplier 2, we can write the *prior probabilities* as

$$P(A_1) = 0.65 \qquad P(A_2) = 0.35$$

Historical data for the quality ratings of the two suppliers are shown in the table below.

Thus, we can write the *conditional probabilities* of receiving a good (G) or bad (B) part from either supplier as:

$$P(G \mid A_1) = 0.98 \qquad P(B \mid A_1) = 0.02$$

$$P(G \mid A_2) = 0.95 \qquad P(B \mid A_2) = 0.05$$

	Percentage Good Parts	Percentage Bad Parts
Supplier 1	98	2
Supplier 2	95	5

4.5 Posterior Probabilities

Using Bayes’ theorem and the prior and conditional probabilities, we can compute the *posterior probabilities* that either supplier ships a bad part

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2)} = \frac{0.65(0.02)}{0.65(0.02) + 0.35(0.05)} = \frac{0.130}{0.130 + 0.175} = 0.426$$

$$P(A_2 \mid B) = \frac{P(A_2)P(B \mid A_2)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2)} = \frac{0.35(0.05)}{0.65(0.02) + 0.35(0.05)} = \frac{0.175}{0.130 + 0.175} = 0.574$$

Note that, for supplier 1, we began with an initial probability of $P(A_1) = 0.65$. However, given the information that the part was bad, the revised probability drops to $P(A_1 \mid B) = 0.426$.

Conversely, for supplier 2, we began with an initial probability of $P(A_2) = 0.35$ and when we include the information that the part was bad, the revised probability grows to $P(A_2 \mid B) = 0.574$.

In conclusion, despite the fact more parts come from supplier 1 than supplier 2, there is more than a 50-50 chance that a bad part comes from supplier 2 (*see notes.)

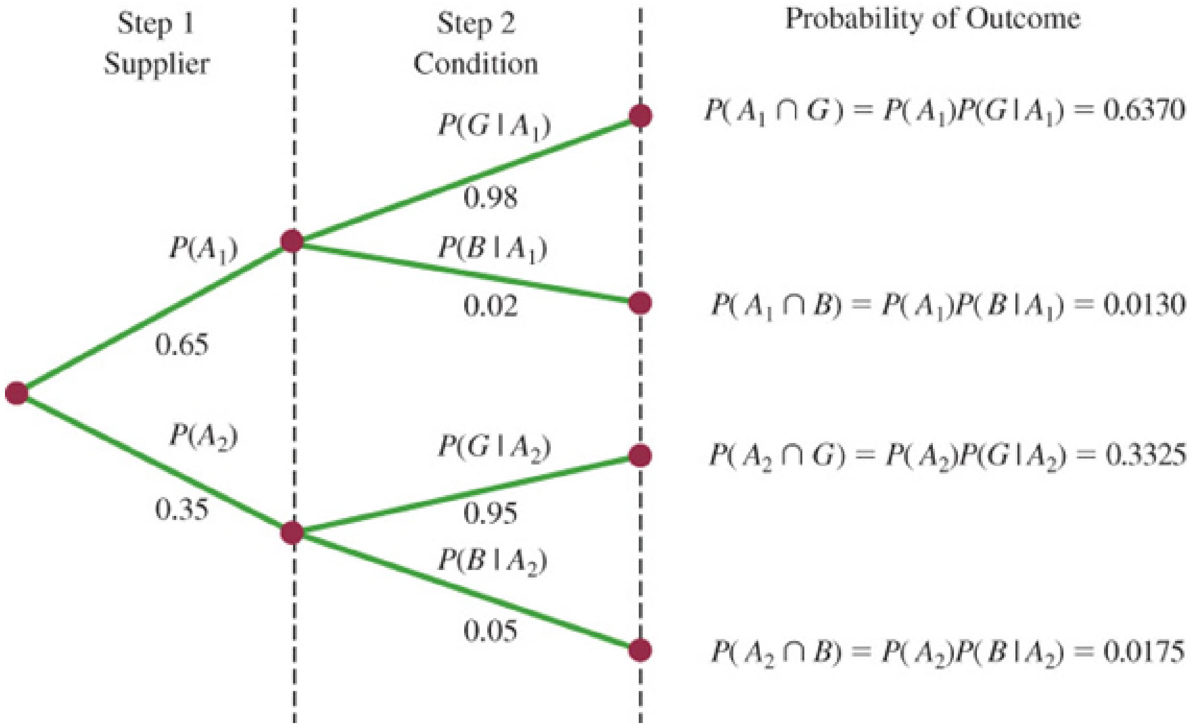
4.5 Bayes' Theorem with a Probability Tree

Bayes' theorem can be visualized with a **probability tree**. From left to right:

Step 1 includes the two branches for the prior probabilities of the two suppliers.

Step 2 branches out the four conditional probabilities of receiving a good (G) or bad (B) part from either supplier 1 (A_1) or 2 (A_2).

The resulting four outcomes represent the possible intersections of supplier (A_1) or (A_2) with either a good (G) or bad (B) part.



Because $P(B) = P(A_1 \cap B) + P(A_2 \cap B)$, the posterior probabilities in Bayes' theorem can be calculated from the joint probabilities as

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i \cap B)}{P(A_1 \cap B) + P(A_2 \cap B)} = \frac{P(A_i)P(B \mid A_i)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2)} \quad \text{with } i = 1,2$$

4.5 Tabular Approach to Bayes' Theorem

A tabular approach is helpful in conducting Bayes' theorem calculations (see table below.)

Step 1: Create three columns: (1) with the two events A_1 and A_2 for which posterior probabilities are desired, (2) with the prior probabilities $P(A_1)$ and $P(A_2)$, and (3) with the conditional probabilities from the available data for event B.

Step 2: In column (4), for each of the two rows compute the joint probabilities $P(A_i \cap B)$ by multiplying the values in columns (1) and (2).

Step 3: Sum the joint probabilities in column (4) to obtain $P(B)$.

Step 4: In column (5), compute the posterior probabilities using the values from column (4):

$$P(A_1 \mid B) = \frac{P(A_1 \cap B)}{P(B)}$$

$$P(A_2 \mid B) = \frac{P(A_2 \cap B)}{P(B)}$$

(1) Events A_i	(2) Prior Probabilities $P(A_i)$	(3) Conditional Probabilities $P(B\|A_i)$	(4) Joint Probabilities $P(A_i \cap B)$	(5) Posterior Probabilities $P(A_i\|B)$
A_1	0.65	0.02	0.0130	0.0130 / 0.0305 = 0.4262
A_2	0.35	0.05	0.0175	0.0175 / 0.0305 = 0.5738
	1.00		0.0305	1.0000

CENGAGE

Summary

- In this chapter, we introduced basic probability concepts and illustrated how probability analysis can be used to provide helpful information for decision-making.
- We described how probability can be interpreted as a numerical measure of the likelihood that an event will occur.
- In addition, we saw that the probability of an event can be computed by either:
  - Summing the probabilities of the experimental outcomes (sample points) that comprise the event.
  - Using the relationships established by different laws of probability, such as:
    - addition law
    - conditional probability
    - multiplication law
- For cases in which additional information is available, we showed how Bayes' theorem can be used to obtain revised or posterior probabilities.

直覺與理性判斷：醫學檢驗

→A patient takes a lab test and the result comes back positive. The test returns a correct positive result in 99% of the cases in which the disease is actually present, and a correct negative result in 98% of the cases in which the disease is not present. Furthermore, .001 of all people have this cancer.

P(cancer) =	.001	P(~ cancer) =	.999
P(+ | cancer) =	.99	P(– | cancer) =	.01
P(+ | ~cancer) =	.02	P(– | ~cancer) =	.98

$$P(\text{cancer} \mid +) = \frac{P(+ \mid cancer)\,P(cancer)}{P(+)} = \mathbf{.047}$$

CENGAGE

計算細節：

- 假設某地區有一百萬人：

  → 999,000人健康，1,000人罹患癌症

  → 檢查出陽性反應者：

  (1) 健康者中有 $999{,}000 \times 2\% = 19{,}980$ 偽陽性

  (2) 癌症患者中有 $1{,}000 \times 99\% = 990$

因此，陽性反應者中罹患癌症的比例：

$$P(cancer \mid +) = \frac{990}{19{,}980 + 990} = \frac{990}{20{,}970} \cong 4.72\%$$

Suppose a second test for the same patient returns a positive result as well. What are the posterior probabilities for cancer?

$P(\text{cancer}) =$	$.001$	$P(\sim\text{cancer}) =$	$.999$
$P(+ \mid \text{cancer}) =$	$.99$	$P(- \mid \text{cancer}) =$	$.01$
$P(+ \mid \sim\text{cancer}) =$	$.02$	$P(- \mid \sim\text{cancer}) =$	$.98$

$$P(\text{cancer} \mid +_1 +_2) = \frac{P(+_1 +_2 \mid cancer)\,P(cancer)}{P(+_1 +_2)} = \mathbf{.710}$$