Chapter 2. Probability

Biostatistics for Engineers

ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

Hansol Choi

August 26, 2023

Chaper 2. Probabiility

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.

- Each outcome in a sample space is called an **element** or a **member** of the sample space, or simply a **sample point**.

Example 2.1

Consider the experiment of tossing a die. If we are interested in the number that shows on the top face, the sample space is

$$S_1 = \{1, 2, 3, 4, 5, 6\}.$$

If we are interested only in whether the number is even or odd, the sample space is simply

$$S_2 = \{even,\ odd\}$$

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 $A \cap B = \emptyset$, 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.

Example 2.1-?

Consider the experiment of tossing a die. If we are interested in the number that shows on the top face, the sample space is

$$S = \{1, 2, 3, 4, 5, 6\}.$$

(1) How many events exist?
(2) Let E be an event composed of all even numbers. How many events exist which are disjoint with E?

The relationship between events and the corresponding sample space can be illustrated graphically by means of **Venn diagrams**.

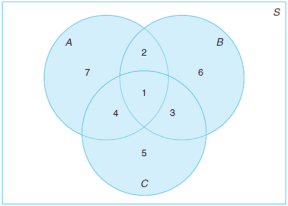

Figure 2.3: Events represented by various regions.

2.3 Counting Sample Points

In many cases, we shall be able to solve a probability problem by counting the number of points in the sample space without actually listing each element. The fundamental principle of counting is the multiplication rule.

Rule 2.1: Multiplication rule

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 2.13

How many sample points are there in the sample space when a pair of dice is thrown once?

Example 2.15

If a 22-member club needs to elect a chair and a treasurer, how many different ways can these two to be elected?

The generalized multiplication rule covering k operations is stated in the following.

Rule 2.2 : Generalized multiplication rule

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.

Example 2.16

Sam is going to assemble a computer by himself. He has the choice of chips from two brands, a hard drive from four, memory from three, and an accessory bundle from five local stores. How many different ways can Sam order the parts?

Example 2.17

How many even four-digit numbers can be formed from the digits 0, 1, 2, 5, 6, and 9 if each digit can be used only once?

Frequently, we are interested in a sample space that contains as elements all possible orders or arrangements of a group of objects. The different arrangements are called **permutations**.

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 $0! = 1$.

Definition 2.8

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

In general, n distinct objects taken r at a time can be arranged in $n(n-1)(n-2)\cdots(n-r+1)$ ways. We represent this product by,

Theorem 2.2

The number of permutations of n distinct objects taken r at a time is ${}_nP_r = \frac{n!}{(n-r)!}$

Theorem 2.1

The number of permutation of n objects is $n!$

Example 2.18

In one year, three awards (research, teaching, and service) will be given to a class of 25 graduate students in a statistics department. If each student can receive at most one award, how many possible selections are there?

Example 2.19

A president and a treasurer are to be chosen from a student club consisting of 50 people. How many different choices of officers are possible if
(1) there are no restrictions;
(2) A will serve only if he is president;
(3) B and C will serve together or not at all;
(4) D and E will not serve together?

Permutations that occur by arranging objects in a circle are called **circular permutations**.

Theorem 2.3

The number of permutations of n objects arranged in a circle is $(n-1)!$.

So far we have considered permutations of distinct objects. That is, all the objects were completely different or distinguishable. What if that's not the case?

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, $\cdots$, n_k of a kth kind is

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

Example 2.20

In a college football training session, the defensive coordinator needs to have 10 players standing in a row. Among these 10 players, there are 1 freshman, 2 sophomores, 4 juniors, and 3 seniors. How many different ways can they be arranged in a row if only their class level will be distinguished?

Often we are concerned with the number of ways of partitioning a set of n objects into r subsets called **cells**. A **partition** has been achieved if the intersection of every possible pair of the r subsets is the empty set $\emptyset$ and if the union of all subsets gives the original set. The order of the elements within a cell is of no importance.

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

$$\left(\begin{array}{c} n \\ n_1, n_2, \cdots, n_r \end{array} \right) = \frac{n!}{n_1! n_2! \cdots n_r!},$$

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

Example 2.21

In how many ways can 7 graduate students be assigned to 1 triple and 2 double hotel rooms during a conference?

In many problems, we are interested in the number of ways of selecting r objects from n without regard to order. These selections are called **combinations.**

Theorem 2.6

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

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

Example 2.22

A young boy asks his mother to get 5 Game-Boy^{TM} *cartridges from his collection of 10 arcade and 5 sports games. How many ways are there that his mother can get 3 arcade and 2 sports games?*

Example 2.23

How many different letter arrangements can be made from the letters in the word STATISTICS ?

2.4 Probability of an Event

Definition 2.9

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

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

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

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

Example 2.24

A coin is tossed twice. What is the probability that at least 1 head occurs?

Example 2.25

A die is loaded in such a way that an even number is twice as likely to occur as an odd number. If E is the event that a number less than 4 occurs on a single toss of the die, find $P(E)$.

Example 2.26

In Example 2.25, let A be the event that an even number turns up and let B be the event that a number divisible by 3 occurs. Find $P(A \cup B)$ and $P(A \cap B)$.

If the sample space for an experiment contains N elements, all of which are equally likely to occur, we assign a probability equal to $1/N$ to each of the N points. The probability of any event A containing n of these N sample points is then the ratio of the number of elements in A to the number of elements in S.

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

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

Equally likely to occur

- How can you know different events ”equally likely to occur?”
- Let’s assume you have tossed a coin for 1,000,000 times
- You’ve got Head 500,000 Tail 500,000
- Are these two events ’equally likely to occur’?
- What is the exact meaning of ’equally likely to occur’ in the real world?
- We are now dealing with a mathematical model. We assume that the events are equally likely to occur in our ’mathematical model’.

Example 2.27

A statistics class for engineers consists of 25 industrial, 10 mechanical, 10 electrical, and 8 civil engineering students. If a person is randomly selected by the instructor to answer a question, find the probability that the student chosen is (a) an industrial engineering major and (b) a civil engineering or an electrical engineering major.

Example 2.28

In a poker hand consisting of 5 cards, find the probability of holding 2 aces and 3 jacks.

2.5 Additive Rules

Theorem 2.7

If A and B are two events, then

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

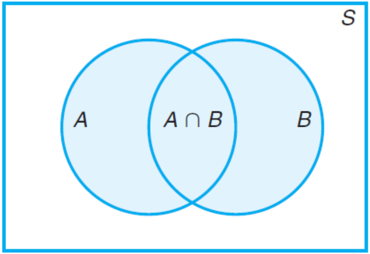

Figure 2.7: Additive rule of probability.

Corollary 2.1

If A and B are mutually exclusive, then

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

Corollary 2.2

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

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n) = \sum_{k=1}^{n} A_k.$$

Corollary 2.3

If $A_1, A_2, \cdots, A_n$ 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,

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ & P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned}$$

Theorem 2.9

If A and A' are complementary events, then

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

Example 2.32

If the probabilities that an automobile mechanic will service 3, 4, 5, 6, 7, or 8 or more cars on any given workday are, respectively, 0.12, 0.19, 0.28, 0.24, 0.10, and 0.07, what is the probability that he will service at least 5 cars on his next day at work?

Example 2.33

Suppose the manufacturer's specifications for the length of a certain type of computer cable are 2000 ± 10 millimeters. In this industry, it is known that small cable is just as likely to be defective (not meeting specifications) as large cable. That is, the probability of randomly producing a cable with length exceeding 2010 millimeters is equal to the probability of producing a cable with length smaller than 1990 millimeters. The probability that the production procedure meets specifications is known to be 0.99.

(a) What is the probability that a cable selected randomly is too large?

(b) What is the probability that a randomly selected cable is larger than 1990 millimeters?

2.6 Conditional Probability

The probability of an event B occurring when it is known that some event A has occurred is called a conditional probability and is denoted by $P(B|A)$. The symbol $P(B|A)$ is usually read "the probability that B occurs given that A occurs" or simply "the probability of B, given A."

Definition 2.10

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

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

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

Example 2.34

The probability that a regularly scheduled flight departs on time is $P(D) = 0.83$; the probability that it arrives on time is $P(A) = 0.82$; and the probability that it departs and arrives on time is $P(D \cap A) = 0.78$. Find the probability that a plane

(a) arrives on time, given that it departed on time, and

(b) departed on time, given that it has arrived on time.

Example 2.35

The concept of conditional probability has countless uses in both industrial and biomedical applications. Consider an industrial process in the textile industry in which strips of a particular type of cloth are being produced. These strips can be defective in two ways, length and nature of texture. For the case of the latter, the process of identification is very complicated. It is known from historical information on the process that 10% of strips fail the length test, 5% fail the texture test, and only 0.8% fail both tests. If a strip is selected randomly from the process and a quick measurement identifies it as failing the length test, what is the probability that it is texture defective?

Definition 2.11

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

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

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 2.36

Suppose that we have a fuse box containing 20 fuses, of which 5 are defective. If 2 fuses are selected at random and removed from the box in succession without replacing the first, what is the probability that both fuses are defective?

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 bag and placed unseen in the second bag. What is the probability that a ball now drawn from the second bag is black?

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.

Example 2.38

A small town has one fire engine and one ambulance available for emergencies. The probability that the fire engine is available when needed is 0.98, and the probability that the ambulance is available when called is 0.92. In the event of an injury resulting from a burning building, find the probability that both the ambulance and the fire engine will be available, assuming they operate independently.

Theorem 2.12

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

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

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

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

Example 2.40

Three cards are drawn in succession, without replacement, from an ordinary deck of playing cards. Find the probability that the event $A_1 \cap A_2 \cap A_3$ *occurs, where* A_1 *is the event that the first card is a red ace,* A_2 *is the event that the second card is a 10 or a jack, and* A_3 *is the event that the third card is greater than 3 but less than 7.*

Theorem 2.12

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

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

2.7 Bayes' Rule

Theorem 2.13 (theorem of total probability)

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

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

Example 2.41

In a certain assembly plant, three machines, B1 , B2 , and B3 , make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. Now, suppose that a finished product is randomly selected. What is the probability that it is defective?

Theorem 2.14 (Bayes' Rule)

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

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

for $r = 1, 2, \cdots, k$.

Example 2.42

With reference to Example 2.41, if a product was chosen randomly and found to be defective, what is the probability that it was made by machine B_3*?*

Example 2.43

A manufacturing firm employs three analytical plans for the design and development of a particular product. For cost reasons, all three are used at varying times. In fact, plans 1, 2, and 3 are used for 30%, 20%, and 50% of the products, respectively. The defect rate is different for the three procedures as follows: $P(D|P_1) = 0.01, P(D|P_2) = 0.03, P(D|P_3) = 0.02$, *where* $P(D|P_j)$ *is the probability of a defective product, given plan* j*. If a random product was observed and found to be defective, which plan was most likely used and thus responsible?*

TA hour