Probability and Statistics I

STAT 3600 – Fall 2021

Le Chen
lzc0090@auburn.edu

Last updated on

July 4, 2021

Auburn University
Auburn AL

Chapter 1. Probability

Sample space

Definition 1.1-1 Experiments for which the outcome cannot be predicted with certainty are called *random experiments*.

Definition 1.1-2 The collection of all possible outcomes is called the *sample space*, denoted as S.

Example 1.1-1 Find the sample space for the experiment of tossing a coin for (a) once (b) twice.

Example 1.1-2 Find the sample space for the experiment of tossing a coin repeatedly and of counting the number of tosses required until the first head appears.

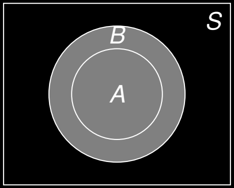
Example 1.1-3 Find the sample space for the experiment of measuring (in hours) the lifetime of a transistor.

Event

Definition 1.1-3 An *event*, denoted as A, is a collection of outcomes in S, that is, $A \subset S$. In particular, $S \subset S$.

Definition 1.1-4 When a random experiment is performed and an outcome of the experiment is in A, we say that *event A has occurred*.

$A \subset B$



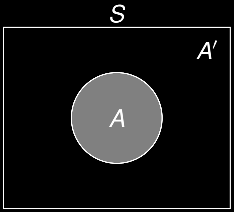
Example 1.1-4 Consider the experiments in Examples 1.1-1 and 1.1-2.
(a) Let A be the event that the number of tosses required until the first head appears is even.
(b) Let B be the event that the number of tosses required until the first head appears is odd.
(c) Let C be the event that the number of tosses required until the first head appears is less than 6.

Algebra of sets

Definition 1.1-5 (Equality) Two sets A and B are *equal*, denoted $A = B$, if and only if $A \subset B$ and $B \subset A$.

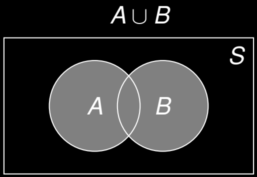
Definition 1.1-6 (Complementation) Suppose $A \subset S$. The *complement* of set A, denoted A' or $\overline{A}$, is the set containing all elements in S but not in A, namely,

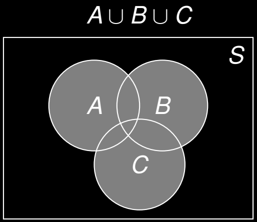
$$A' := \{x : x \in S \text{ and } x \notin A\} .$$



Definition 1.1-7 (Union) The *union* of sets A and B, denoted $A \cup B$, is the set containing all elements in either A or B or both, i.e.,

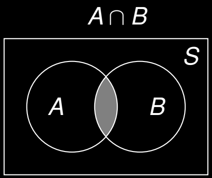
$$A \cup B := \{x : x \in A \text{ or } x \in B\}.$$

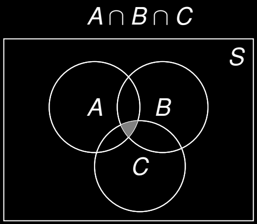




Definition 1.1-8 (Intersection) The *intersection* of sets A and B, denoted $A \cap B$, is the set containing all elements in both A and B.

$$A \cap B = \{x : x \in A \text{ and } B\}$$



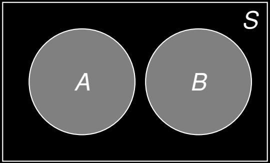


Definition 1.1-9 (Null set) The set containing no element is called the *null set*, denoted $\emptyset$. Note that

$$\emptyset = S'.$$

Definition 1.1-10 (Disjoit sets) Two sets A and B are called *disjoint* or *mutually exclusive* if they contain no common element, that is, if $A \cap B = \emptyset$.

A and B are disjoint



The definitions of the union and intersection of two sets can be extended to any finite number of sets as follows:

$$\begin{aligned}
\bigcup_{i=1}^{n} A_i &= A_1 \cup A_2 \cup \cdots \cup A_n \\
&= \{x : x \in A_1 \text{ or } A_2 \text{ or } \cdots \text{ or } x \in A_n\}, \\
\bigcap_{i=1}^{n} A_i &= A_1 \cap A_2 \cap \cdots \cap A_n \\
&= \{x : x \in A_1 \text{ and } A_2 \text{ and } \cdots \text{ and } x \in A_n\},
\end{aligned}$$

and even to countably infinite many sets:

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup A_3 \cup \cdots$$

$$\bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap A_3 \cap \cdots$$

Interpretation of sets in probability

1.

$$S = \text{the certain event}$$
$$\emptyset = \text{the impossible event}$$

2. If A and B are events in S, then

$$A' = \text{the event that } A \text{ did not occur}$$
$$A \cup B = \text{the event that either } A \text{ or } B \text{ or both occurred}$$
$$A \cap B = \text{the event that both } A \text{ and } B \textit{ occurred}$$

3. Similarly, if $A_1, A_2, ..., A_n$ are sequence of events in S, then

$$\bigcup_{i=1}^{n} A_i = \text{the event that at least one of the } A_i \text{ occurred}$$
$$\bigcap_{i=1}^{n} A_i = \text{the event that all of the } A_i \text{ occurred}$$

Definition 1.1-11 (Mutually exclussive and exhaustive events) Events $A_1, A_2, ..., A_k$ are *mutually exclusive and exhaustive* if

1. $A_i \cap A_j = \emptyset, i \neq j$, and (mutually exclusive)

2. $A_1 \cup A_2 \cup \cdots \cup A_k = S$. (exhaustive)

Identities

By the above set of definitions, we obtain the following identities:

$$S' = \emptyset$$

$$A'' = A$$

$$S \cup A = S$$

$$S \cap A = A$$

$$A \cup A' = S$$

$$A \cap A' = \emptyset$$

Commutative Laws:

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

Associative Laws:

$$A \cup (B \cup C) = (A \cup B) \cup C$$

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

Distributive Laws:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

De Morrgan's Laws:

$$(A \cup B)' = A' \cap B'$$

$$(A \cap B)' = A' \cup B'$$

The distributive laws can be extended as follows:

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

$$A \cup \left(\bigcap_{i=1}^{n} B_i \right) = \bigcap_{i=1}^{n} (A \cup B_i)$$

Similarly, De Morgan's laws also can be extended as follows:

$$\left(\bigcup_{i=1}^{n} A_i \right)' = \bigcap_{i=1}^{n} A_i'$$

$$\left(\bigcap_{i=1}^{n} A_i \right)' = \bigcup_{i=1}^{n} A_i'$$

Example 1.1-5 An experiment consists of tossing two dice.

(a) Find the sample space S.

(b) Find the event A that the sum of the dots on the dice equals 7.

(c) Find the event B that the sum of the dots on the dice is greater than 10.

(d) Find the event C that the sum of the dots on the dice is greater than 12.

Example 1.1-6 Consider the experiment of Example 1.1-2. We define the events

$$A = \{k : k \text{ is odd}\}$$
$$B = \{k : 4 \leq k \leq 7\}$$
$$C = \{k : 1 \leq k \leq 10\}$$

where k is the number of tosses required until the first H (head) appears. Determine the events A', B', C', $A \cup B$, $B \cup C$, $A \cap B$, $A \cap C$, $B \cap C$, and $A' \cap B$.

Properties of Probability

Definition 1.1-12 (Relative frequency definition) Suppose that the random experiment is repeated n times. If event A occurs $\mathcal{N}(A)$ times, then the probability of event A, denoted by $\mathbb{P}(A)$, is defined as

$$\mathbb{P}(A) = \lim_{n\to\infty} \frac{\mathcal{N}(A)}{n},$$

where $\mathcal{N}(A)/n$ is relative frequency of an event A.

Remark 1.1-1 Note that this limit may not exist.

It is clear that for any event A, the relative frequency of A has the following properties:

1. $0 \leq \mathcal{N}(A)/n \leq 1$, where $\mathcal{N}(A)/n = 0$ if A occurs in none of the repeated trials and $\mathcal{N}(A)/n = 1$ if A occurs in all of the n repeated trials.

2. If A and B are mutually exclusive events, then

$$\mathcal{N}(A \cup B) = \mathcal{N}(A) + \mathcal{N}(B)$$

and

$$\frac{\mathcal{N}(A \cup B)}{n} = \frac{\mathcal{N}(A)}{n} + \frac{\mathcal{N}(B)}{n}.$$

Example 1.1-7 A fair six-sided die is rolled six times. If the face numbered k is the outcome on roll k for $k = 1, 2, \cdots, 6$, we say that a match has occurred. The experiment is called a *success* if at least one match occurs during the six trials. Otherwise, the experiment is called a *failure*. The sample space is

$$S = \{\text{success}, \ \text{failure}\}.$$

Let $A = \{\text{success}\}$. We would like to assign a value to $\mathbb{P}(A)$. Accordingly, this experiment was simulated on a computer as follows:

No. of trials n	$\mathcal{N}(A)$	$\mathcal{N}(A)/n$
10	7	0.700
100	69	0.690
500	336	0.672
2000	1307	0.653
46656	31042	0.665

The probability of event A is not intuitively obvious, but it will be shown latter that

$$\mathbb{P}(A) = 1 - \left(1 - \frac{1}{6}\right)^6 = \frac{31031}{46656} \approx 0.665.$$

This assignment is certainly supported by the simulation (although not proved by it).

```python
#!/usr/bin/env python3

import random
random.seed(10)

def Experiment(trials):
    success = 0
    for k in range(0, trials):
        for i in range(1, 7):
            n = random.randint(1, 6)
            if n == i:
                success = success + 1
                break
    print("Number and ratio of success among {} tries are {} and {:0.3}, respectively".
         format(trials, success, success/trials))


def main():
    Experiment(10)
    Experiment(100)
    Experiment(500)
    Experiment(2000)
    Experiment(46656)


if __name__ == "__main__":
    main()
```

```
Number and ratio of success among 10 tries are 7 and 0.7, respectively
Number and ratio of success among 100 tries are 69 and 0.69, respectively
Number and ratio of success among 500 tries are 336 and 0.672, respectively
Number and ratio of success among 2000 tries are 1307 and 0.653, respectively
Number and ratio of success among 46656 tries are 31042 and 0.665, respectively
```

Axiomatic definition

Definition 1.1-13 (Probability) *Probability* is a real-valued set function $\mathbb{P}$ that assigns, to each event A in the sample space S, a number $\mathbb{P}(A)$, called the *probability* of the event A, such that the following properties are satisfied:

(a) $P(A) \geq 0$;

(b) $P(S) = 1$;

(c) If $A_1, A_2, A_3, \cdots$ are events and $A_i \cap A_j = \emptyset$, $i \neq j$, then

$$\mathbb{P}\left(A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_k\right) = P(A_1) + P(A_2) + P(A_3) + \cdots + P(A_k)$$

for each positive integer K, and

$$\mathbb{P}\left(A_1 \cup A_2 \cup A_3 \cup \cdots\right) = P(A_1) + P(A_2) + P(A_3) + \cdots$$

for an infinite, but countable, number of events.

Example 1.1-8 (Rolling a die once) Consider rolling a die once. For each subset A of $S = \{1, 2, 3, 4, 5, 6\}$, let $\mathbb{P}(A)$ be the number of elements of A divided by 6. It is trivial to see that this satisfies the first two axioms. There are only finitely many distinct collections of nonempty disjoint events. It is not difficult to see that Axiom (c) is also satisfied by this example.

Example 1.1-9 (Loaded die) If we believe that the die is loaded, we might believe that some sides have different probabilities of turning up. To be specific, suppose that we believe that 6 is three times as likely to come up as each of the other five sides. We could set $p_i = 1/8$ for $i = 1, 2, 3, 4, 5$ and $p_6 = 3/8$. Then, for each event A, define $\mathbb{P}(A)$ to be the sum of all p_i such that $i \in A$. For example, if $A = \{1, 4, 5\}$, then

$$\mathbb{P}(A) = p_1 + p_4 + p_5 = \frac{3}{8}.$$

It is not difficult to check that this also satisfies all three axioms.

Theorem 1.1-1 For each event A, we have $\mathbb{P}(A) = 1 - \mathbb{P}(A')$.

Theorem 1.1-2 $\mathbb{P}(\emptyset) = 0$.

Theorem 1.1-3 If events A and B are such that $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Theorem 1.1-4 For each event A, it holds that $0 \leq \mathbb{P}(A) \leq 1$.

Theorem 1.1-5 If A and B are any two events, then

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}\left(A \cap B\right).$$

Theorem 1.1-6 If A, B, and C are any three events, then

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

Definition 1.1-14 Consider a finite sample space S with m elements

$$S = \{e_1, e_2, ..., e_m\},$$

where e_i is a possible outcomes of the experiment. If each of these outcomes has the same probability of occurring, we say that the m outcomes are *equally likely*, that is,

$$\mathbb{P}(e_i) = \frac{1}{m}, i = 1, 2, ..., m,$$

and

$$\mathbb{P}(A) = \frac{\mathcal{N}(A)}{m},$$

where $\mathcal{N}(A)$ is the number of outcomes belonging to event A and $m = \mathcal{N}(S)$ is the number of sample points in S.

Definition 1.1-15 Let $\mathbb{P}(A) = 0.9, \mathbb{P}(B) = 0.8$. Show that $\mathbb{P}(A \cap B) \geq 0.7$.

Example 1.1-10 Given that $\mathbb{P}(A) = 0.9, \mathbb{P}(B) = 0.8$, and $\mathbb{P}(A \cap B) = 0.75$, find (a) $\mathbb{P}(A \cup B)$; (b) $\mathbb{P}(A \cap B')$; (c) $\mathbb{P}(A' \cap B')$.

Example 1.1-11 Let A, B, and C be three events in S. If

$$\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{4}, \quad \mathbb{P}(C) = \frac{1}{3}, \quad \mathbb{P}\left(A \cap B\right) = \frac{1}{8}, \quad \mathbb{P}\left(A \cap C\right) = \frac{1}{6}$$

and $\mathbb{P}\left(B \cap C\right) = 0$, find $\mathbb{P}\left(A \cup B \cup C\right)$.

Example 1.1-12 Prove that

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \mathbb{P}(A_i) \quad \text{and} \quad \mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) \ge 1 - \sum_{i=1}^{n} \mathbb{P}(A'),$$

where the second inequality above is known as the *Bonferroni inequality*.

Example 1.1-13 During a visit to a primary care physician's office, the probability of having neither lab work nor referral to a specialist is 0.21. Of those coming to that office, the probability of having lab work is 0.41 and the probability of having referral is 0.53. What is the probability of having both lab work and a referral?

Example 1.1-14 Draw one card at random from a standard deck of cards. The sample space S is the collection of the 52 cards. Assume that the probability set function assigns $1/52$ to each of the 52 outcomes. Let

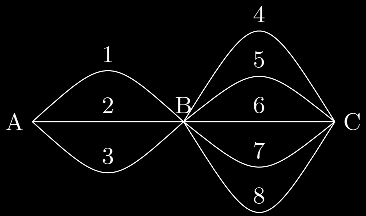
$$\begin{aligned}
A &= \{x : x \text{ is a jack, queen, or king}\}, \\
B &= \{x : x \text{ is a } 9, 10, \text{ or jack and } x \text{ is red}\}, \\
C &= \{x : x \text{ is a club}\}, \\
D &= \{x : x \text{ is a diamond, a heart, or a spade}\}.
\end{aligned}$$

Find (a) $\mathbb{P}(A)$; (b) $\mathbb{P}(A \cap B)$; (c) $\mathbb{P}(A \cup B)$; (d) $\mathbb{P}(C \cup D)$; (e) $\mathbb{P}(C \cap D)$.

Example 1.1-15 Consider two events A and B such that $\mathbb{P}(A) = 1/3$ and $\mathbb{P}(B) = 1/2$. Determine the value of $\mathbb{P}(B \cap A')$ for each of the following conditions: (a) A and B are disjoint; (b) $A \subset B$; (c) $\mathbb{P}(A \cap B) = 1/8$.

Exercises from textbook: Section 1.1: 1, 2, 4, 5, 6, 7, 8, 9, 12, 13.

Example 1.2-1 (Routes between Cities) Suppose that there are three different routes from city A to city B and five different routes from city B to city C. The cities and routes are depicted in the figure, with the routes numbered from 1 to 8. We wish to count the number of different routes from A to C that pass through B. For example, one such route is 1 followed by 4, which we can denote (1, 4). Similarly, there are the routes $(1, 5), (1, 6), ..., (3, 8)$. It is not difficult to see that the number of different routes $3 \times 5 = 15$.



Theorem 1.2-1 (Multiplication Principle) Suppose that an experiment (or procedure) E_1 has n_1 outcomes and, for each of these possible outcomes, an experiment (procedure) E_2 has n_2 possible outcomes. Then the has $n_1 n_2$ possible outcomes.

More generally, suppose that an experiment has k parts $(k \geq 2)$, that the i-th part of the experiment can have n_i possible outcomes $(i = 1, ..., k)$, and that all of the outcomes in each part can occur regardless of which specific outcomes have occurred in the other parts. Then the sample space S of the experiment will contain all vectors of the form $(u_1, \cdots, u_k)$, where u_i is one of the n_i possible outcomes of part $i (i = 1, ..., k)$. The total number of these vectors in S will be equal to the product $n_1 n_2 \cdots n_k$.

Example 1.2-2 A boy found a bicycle lock for which the combination was unknown. The correct combination is a four-digit number, $d_1 d_2 d_3 d_4$, where d_i, $i = 1, 2, 3, 4$, is selected from $1, 2, 3, 4, 5, 6, 7$, and 8. How many different lock combinations are possible with such a lock?

Suppose that n positions are to be filled with n different objects. There are n choices for filling the first position, $n-1$ for the second, ..., and 1 choice for the last position. So, by multiplication rule, there are

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

possible arrangements. The symbol $n!$ is read "n factorial."

Definition 1.2-1 (Permutation) Each of the $n!$ arrangements (in a raw) of n different objects is called a *permutation* of the n objects.

Example 1.2-3 Order 7 books on a shelf = 7! permutations.

If only r positions are to be filled with objects selected from n different objects, $r \leq n$, then the number of possible ordered arrangements is

$$_nP_r = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!}$$

Definition 1.2-2 Each of the $_nP_r$ arrangements is called a *permutation of n objects taken r at a time.*

Remark 1.2-1 Sampling without replacement, one at a time, order is important!

Example 1.2-4 (Choosing Officers) Suppose that a club consists of 25 members and that a president and a secretary are to be chosen from the membership. We shall determine the total possible number of ways in which these two positions can be filled.

Since the positions can be filled by first choosing one of the 25 members to be president and then choosing one of the remaining 24 members to be secretary, the possible number of choices is ${}_{25}P_2 = (25)(24) = 600$.

Sampling with Replacement

Consider a box that contains n balls numbered $1, ..., n$. First, one ball is selected at random from the box and its number is noted. This ball is then put back in the box and another ball is selected (it is possible that the same ball will be selected again). As many balls as desired can be selected in this way. This process is called sampling with replacement. It is assumed that each of the n balls is equally likely to be selected at each stage and that all selections are made independently of each other.

Suppose that a total of r selections are to be made, where r is a given positive integer. Then the sample space S of this experiment will contain all vectors of the form $(x_1, ..., x_k)$, where x_i is the outcome of the ith selection ($i = 1, ..., k$). Since there are n possible outcomes for each of the r selections, the total number of vectors in S is n^r. Furthermore, from our assumptions it follows that S is a equally likely sample space. Hence, the probability assigned to each vector in S is $1/n^r$.

Example 1.2-5 (Birthday problem) In a group of k people, what is the probability that at least 2 people will have the same birthday? Assume $n = 365$ and that birthdays are equally distributed throughout the year, no twins, etc.

Solution.

1. Since there are 365 possible birthdays for each of k people, the sample space S will contain 365^k outcomes, all of which will be equally probable.
2. If $k > 365$, there are not enough birthdays for every one to be different, and hence at least two people must have the same birthday. (*Pigeonhole principle*)
3. So, we assume below that $k \leq 365$.
4. Denote

$$A_k = \{\text{at least 2 people have the same birthday in a group of } k \text{ people}\}$$

and hence,

$$A'_k = \{\text{all } k \text{ people have distinct birthdays}\}.$$

Solution(continued).

5. Counting the number of outcomes in A_k is tedious. However, the number of outcomes in S for which all k birthdays will be different, namely, the number of outcomes in A'_k, is easy.
6. Indeed, $\mathcal{N}(A'_k) =_{365} P_k$, since the first person's birthday could be any one of the 365 days, the second person's birthday could then be any of the other 364 days, and so on.
7. Hence,

$$\mathbb{P}\left(A'_k\right) = \frac{{}_{365}P_k}{365^k},$$

and therefore,

$$\mathbb{P}(A_k) = 1 - \mathbb{P}\left(A'_k\right) = 1 - \frac{{}_{365}P_k}{365^k} = 1 - \frac{365!}{(365-k)!\,365^k}.$$

$\square$

k	2	3	10	20	60
$\mathbb{P}(A_k)$	$\dfrac{1}{365}$	$\dfrac{1093}{133225}$	$\dfrac{2689423743942044098153}{22996713557917153515625}$	–	–
Approx.	0.002704	0.008204	0.1169	0.4114	0.9941

Sampling without replacement

Sample a subset of size r from n different objects, if we aren't concerned with order, the number of subsets $= {}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$.

Definition 1.2-3 (Combinations) Each of the ${}_nC_r$ unordered subsets is called a *combination of n objects taken r at a time*, where

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

Example 1.2-6 A committee of 5 persons is to selected randomly from a group of 5 men and 10 women.

(a) Find the probability that the committee consists of 2 men and 3 women.

(b) Find the probability that the committee consists of all women.

Theorem 1.2-2 (Binomial theorem) For $n \geq 0$, it holds that

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

$n=0$	$\binom{0}{0}=1$	1
$n=1$	$\binom{1}{0}=1 \quad \binom{1}{1}=1$	$x+y$
$n=2$	$\binom{2}{0}=1 \quad \binom{2}{1}=2 \quad \binom{2}{2}=1$	$x^2+2xy+y^2$
$n=3$	$\binom{3}{0}=1 \quad \binom{3}{1}=3 \quad \binom{3}{2}=3 \quad \binom{3}{3}=1$	$x^3+3x^2y+3xy^2+y^3$
$n=4$	$\binom{4}{0}=1 \quad \binom{4}{1}=4 \quad \binom{4}{2}=6 \quad \binom{4}{3}=4 \quad \binom{4}{4}=1$	$x^4+4x^3y+6x^2y^2+4xy^3+y^4$

Example 1.2-7 Prove $\sum_{r=0}^{n}(-1)^r\binom{n}{r} = 0$ and $\sum_{r=0}^{n}\binom{n}{r} = 2^n$.

Split objects into m groups of various sizes:

Suppose that in a set of n objects, n_1 are similar, n_2 are similar, ..., n_m are similar, where $n_1 + n_2 + \cdots + n_m = n$. Then the number of distinguishable permutations of the n objects is

$$\binom{n}{n_1, n_2, ..., n_m} = \frac{n!}{n_1!n_2! \cdots n_m!}$$

These numbers are called multinomial coefficient because of the following multinomial expansion:

$$(x_1 + x_2 + \cdots + x_k)^n = \sum_{k_1+k_2+\cdots+k_m=n} \binom{n}{k_1, k_2, \cdots, k_m} x_1^{k_1} x_2^{k_2} \cdots x_m^{k_m}.$$

Example 1.2-8 20 members of a club need to be split into 3 committees (A, B, C) of 8, 8, and 4 people, respectively. How many ways are there to split the club into these committees?

Example 1.2-9 Suppose that three runners from team A and three runners from team B participate in a race. If all six runners have equal ability and there are no ties, what is the probability that the three runners from team A will finish first, second, and third, and the three runners from team B will finish fourth, fifth, and sixth?

Exercises from textbook: Section 1.2: 1, 3, 4, 5, 7, 8, 9, 11, 16, 17.

A major use of probability in statistical inference is the updating of probabilities when certain events are observed.

The updated probability of event A after we learn that event B has occurred is the conditional probability of A given B.

Example 1.3-1 (Lottery) Consider a state lottery game in which six numbers are drawn without replacement from a bin containing the numbers $1 - 30$.

Each player tries to match the set of six numbers that will be drawn without regard to the order in which the numbers are drawn.

Suppose that you hold a ticket in such a lottery with the numbers

$$1, 14, 15, 20, 23, \text{ and } 27.$$

You turn on your television to watch the drawing but all you see is one number, 15, being drawn when the power suddenly goes off in your house. You don't even know whether 15 was the first, last, or some in-between draw.

However, now that you know that 15 appears in the winning draw, the probability that your ticket is a winner must be higher than it was before you saw the draw. How do you calculate the revised probability?

Example 1.3-1 is typical of the following situation. An experiment is performed for which the sample space S is given (or can be constructed easily) and the probabilities are available for all of the events of interest.

We then learn that some event B has occurred, and we want to know how the probability of another event A changes after we learn that B has occurred.

In Example 1.3-1, the event that we have learned is

$$B = \{\text{one of the numbers drawn is 15}\}.$$

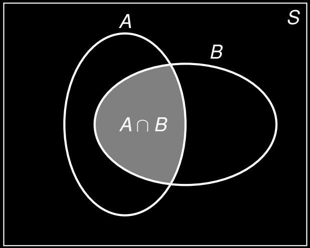
We are certainly interested in the probability of

$$A = \{\text{the numbers } 1, 14, 15, 20, 23, \text{ and 27 are drawn}\},$$

and possibly other events.

If we know that the event B has occurred, then we know that the outcome of the experiment is one of those included in B. Hence, to evaluate the probability that A will occur, we must consider the set of those outcomes in B that also result in the occurrence of A.

As sketched in the figure, this set is precisely the set $A \cap B$. It is therefore natural to calculate the revised probability of A according to the following definition.



Definition 1.3-1 (Conditional probability) *Conditional probability of event A given event B*, is defined by

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

Similarly,

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

Using the above two equations, we have the following multiplication rule

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)P(B) = \mathbb{P}(B|A)P(A).$$

This often quite useful in computing the joint probability of events A and B.

In general, suppose that $A_1, A_2, \cdots, A_n$ are events such that $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$. Then

$$\begin{gathered} P(A_1 \cap A_2 \cap \cdots \cap A_n) \\ \| \\ P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \end{gathered}$$

Solution (Example 1.3-1). In Example 1.3-1, you learned that the event $B = \{\text{one of the numbers drawn is } 15\}$ has occurred. You want to calculate the probability of the event A that your ticket is a winner. Both events A and B are expressible in the sample space that consists of the ${}_{30}C_6 = 30!/(6!24!)$ possible combinations of 30 items taken six at a time, namely, the unordered draws of six numbers from $1 - 30$. The event B consists of combinations that include 15. Since there are 29 remaining numbers from which to choose the other five in the winning draw, there are ${}_{29}C_5 = 29!/(5!24!)$ outcomes in B. It follows that

$$\mathbb{P}(B) = \frac{{}_{29}C_5}{{}_{30}C_6} = 0.2.$$

The event A that your ticket is a winner consists of a single outcome that is also in B, so $A \cap B = A$, and

$$\mathbb{P}(A \cap B) = P(A) = \frac{1}{{}_{30}C_6} = 1.68 \times 10^{-6}.$$

It follows that the conditional probability of A given B is

$$\mathbb{P}(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1.68 \times 10^{-6}}{0.2} = 8.4 \times 10^{-6}.$$

This is five times as large as $\mathbb{P}(A)$ before you learned that B had occurred. $\square$

Example 1.3-2 Consider an experiment consists of observing the sum of the dice when two fair dice are thrown; you are then informed that the sum is not greater than 3.

(a) Find the probability of the event that two faces are the same without the given information.

(b) Find the probability of the same event with the information given.

Example 1.3-3 A lot of 100 semiconductor chips contains 20 that are defective. Two chips are selected at random, without replacement, from the lot.

(a) What is the probability that the first one selected is defective?

(b) What is the probability that the second one selected is defective given that the first one was defective?

(c) What is the probability that both are defective?

Example 1.3-4 A number is selected at random from $\{1, 2, 3, ..., 100\}$. Given that the number selected is divisible by 2, find the probability that it is divisible by 3 or 5.

Example 1.3-5 An urn contains four colored balls; two orange and two blue. Two balls are selected at random without replacement, and you are told that at least one of them is orange. What is the probability that the other ball is also orange?

Example 1.3-6 Suppose that four balls are selected one at a time, without replacement, from a box containing r red balls and b blue balls ($r \geq 2, b \geq 2$). Determine the probability of obtaining the sequence of outcomes red, blue, red, blue.

Example 1.3-7 For any two events A and B with $\mathbb{P}(B) > 0$, prove that $\mathbb{P}(A'|B) = 1 - \mathbb{P}(A|B)$.

Example 1.3-8 For any three events A, B, and D, such that $\mathbb{P}(D) > 0$, prove that $\mathbb{P}(A \cup B|D) = \mathbb{P}(A|D) + \mathbb{P}(B|D) - \mathbb{P}(A \cap B|D)$.

Exercises from textbook. Section 1.3: 1, 3, 4, 5, 6, 8, 9, 11, 12a, 15, 16.

For certain pairs of events, the occurrence of one of them may or may not change the probability of the occurrence of the other. In the latter case, they are said to be independent events. However, before giving the formal definition of independence, let us consider an example.

Example 1.4-1 Flip a fair coin twice and observe the sequence of heads and tails. The sample space is then

$$S = \{HH, HT, TH, TT\}.$$

It is reasonable to assign a probability $1/4$ to each of these four outcomes (equally likely events). Let

$$\begin{aligned} A &= \{\text{heads on the first flip}\} = \{HH, HT\}, \\ B &= \{\text{tails on the second flip}\} = \{HT, TT\}, \\ C &= \{\text{tails on the both flips}\} = \{TT\}. \end{aligned}$$

Then $\mathbb{P}(B) = 2/4 = 1/2$.

Now, on the one hand, if we are given that C has occurred, then $\mathbb{P}(B|C) = 1$, because $C \subset B$. That is, the knowledge of the occurrence of C has changed the probability of B.

On the other hand, if we are given that A has occurred, then

$$\mathbb{P}(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/4}{2/4} = \frac{1}{2} = P(B).$$

So the occurrence of A has not changed the probability of B. Hence, the probability of B does not depend upon knowledge about event A, so we say that A and B are independent events. That is, events A and B are independent if the occurrence of one of them does not affect the probability of the occurrence of the other.

A more mathematical way of saying this is

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

provided $\mathbb{P}(A) > 0$ or, in the latte case, $\mathbb{P}(B) > 0$. With the first of these equalities and the multiplication rule, we have

$$\mathbb{P}(A \cap B) = P(A)P(B|A) = P(A)P(B).$$

The second of these equalities, namely, $\mathbb{P}(A|B) = P(A)$, gives us the same result

$$\mathbb{P}(A \cap B) = P(B)P(A|B) = P(B)P(A).$$

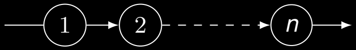
Definition 1.4-1 Events A and B are *independent* if and only if $\mathbb{P}(A \cap B) = P(A)P(B)$. Otherwise A and B are called *dependent* events.

Remark 1.4-1 It follows immediately that if A and B are independent, then

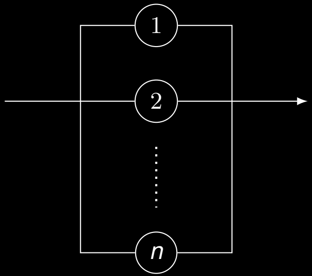
$$\mathbb{P}(A|B) = \mathbb{P}(A) \qquad \text{and} \qquad \mathbb{P}(B|A) = \mathbb{P}(B).$$

Theorem 1.4-1 If A and B are independent events, then the following pairs of events are also independent:
(a) A and B'. (b) A' and B. (c) A' and B'.

Example 1.4-2 A system consisting of n separate components is said to be a *series* system if it functions when all n components function. Assume that the components fail independently and that the probability of failure of component i is p_i, $i = 1, 2, \cdots, n$. Find the probability that the system functions.



Example 1.4-3 A system consisting of n separate components is said to be a *parallel* system if it functions when at least one of components functions. Assume that the components fail independently and that the probability of failure of component i is p_i, $i = 1, 2, \cdots, n$. Find the probability that the system functions.



Example 1.4-4 (Tossing a coin until a head appears) Suppose that a fair coin is tossed until a head appears for the first time, and assume that the outcomes of the tosses are independent. We shall determine the probability p_n that exactly n tosses will be required.

Solution. The desired probability is equal to the probability of obtaining $n - 1$ tails in succession and then obtaining a head on the next toss. Since the outcomes of the tosses are independent, the probability of this particular sequence of n outcomes is $p_n = (1/2)^n$. $\square$

Remark 1.4-2 The probability that a head will be obtained sooner or later (or, equivalently, that tails will not be obtained forever) is

$$\sum_{n=1}^{\infty} p_n = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 1.$$

Since the sum of the probabilities p_n is 1, it follows that the probability of obtaining an infinite sequence of tails without ever obtaining a head must be 0, which is equivalent to the limit:

$$\lim_{n\to\infty} \left(\frac{1}{2}\right)^n = 0.$$

Definition 1.4-2 (Mutually independence) Events A, B, and C are *mutually independent* if and only if the following two conditions hold:

(a) A, B, and C are pairwise independent, that is,

$$\mathbb{P}(A \cap B) = P(A)P(B)$$
$$P(A \cap C) = P(A)P(C)$$
$$\mathbb{P}(B \cap C) = P(B)P(C);$$

and (b)

$$\mathbb{P}(A \cap B \cap C) = P(A)P(B)P(C).$$

Example 1.4-5 In the experiment of throwing two fair dice, let A be the event that the first die is odd, B be the event that the second die is odd, and C be the event that the sum is odd. Show that events A, B, and C are pairwise independent, but A, B, and C are not independent.

Exercises from textbook: Section 1.4: 1, 2, 3, 5, 7, 8, 9, 11, 12, 13, 16.

Chapter 1. Probability

§ 1.1 Properties of Probability

§ 1.2 Methods of Enumerations

§ 1.3 Conditional Probability

§ 1.4 Independent Events

§ 1.5 Bayes Theorem

Suppose that we are interested in which of several disjoint events $B_1, \ldots, B_k$ will occur and that we will get to observe some other event A. If $\mathbb{P}(A|B_i)$ is available for each i, then Bayes' theorem is a useful formula for computing the conditional probabilities of the B_i events given A, namely, $\mathbb{P}(B_i|A)$.

Example 1.5-1 (Test for a disease) Suppose that you are walking down the street and notice that the Department of Public Health is giving a free medical test for a certain disease. The test is 90 percent reliable in the following sense: If a person has the disease, there is a probability of 0.9 that the test will give a positive response; whereas, if a person does not have the disease, there is a probability of only 0.1 that the test will give a positive response.

Data indicate that your chances of having the disease are only 1 in $10,000$. However, since the test costs you nothing, and is fast and harmless, you decide to stop and take the test. A few days later you learn that you had a positive response to the test. Now, what is the probability that you have the disease?

The last question in Example 1.5-1 is a prototype of the question for which Bayes' theorem was designed. We have at least two disjoint events ("you have the disease" and "you do not have the disease") about which we are uncertain, and we learn a piece of information (the result of the test) that tells us something about the uncertain events. Then we need to know how to revise the probabilities of the events in the light of the information we learned.

From multiplication rule $\mathbb{P}(B \cap A) = \mathbb{P}(A|B)P(B)$, we can obtain the following Bayes' rule:

Theorem 1.5-1 (Bayes' rule)

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)} \qquad \text{provided} \qquad \mathbb{P}(A) \neq 0.$$

Let the events $B_1, B_2, ..., B_m$ are mutually exclusive and exhaustive events, that is,

$$S = B_1 \cup B_2 \cup \cdots \cup B_m \qquad \text{and} \qquad B_i \cap B_j = \emptyset,\ i \neq j.$$

Furthermore, suppose the prior probability of the events B_i is positive, that is, $\mathbb{P}(B_i) > 0,\ i = 1, \cdots, m$.

If A is an event, then A is the union of m mutually exclusive events, namely,

$$A = (B_1 \cap A) \cup (B_2 \cap A) \cup \cdots \cup (B_m \cap A).$$

Thus,

$$\mathbb{P}(A) = \sum_{i=1}^{m} \mathbb{P}(B_i \cap A) = \sum_{i=1}^{m} \mathbb{P}(B_i)\mathbb{P}(A|B_i),$$

which is known as the total probability of event A.

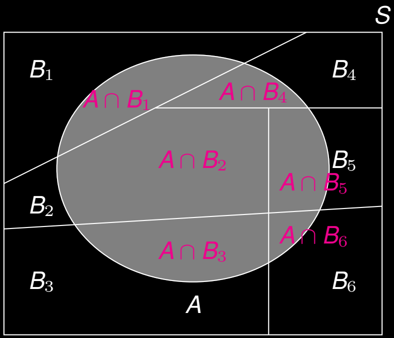
If $\mathbb{P}(A) > 0$, then

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(B_k \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B_k)\mathbb{P}(A|B_k)}{\mathbb{P}(A)} \qquad k = 1, 2, \cdots, m.$$

Using total probability equation for $\mathbb{P}(A)$, we obtain

Theorem 1.5-2 (Bayes' theorem) Let the events $B_1, B_2, ..., B_m$ are mutually exclusive and exhaustive events. If $\mathbb{P}(A) > 0$, then

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(B_k)\mathbb{P}(A|B_k)}{\sum_{i=1}^{m} \mathbb{P}(B_i)\mathbb{P}(A|B_i)}, \qquad k = 1, \cdots, m.$$


S
B_1
B_4
$A \cap B_1$
$A \cap B_4$
$A \cap B_2$
B_5
$A \cap B_5$
B_2
$A \cap B_6$
$A \cap B_3$
B_3
B_6
A

Solution(Example 1.5-1). Let us return to the example with which we began this section. We have just received word that we have tested positive for a disease. The test was 90 percent reliable in the sense that we described in Example 1.5-1. We want to know the probability that we have the disease after we learn that the result of the test is positive. Some of you may feel that this probability should be about 0.9. However, this feeling completely ignores the small probability of 0.0001 that you had the disease before taking the test. We shall let B_1 denote the event that you have the disease, and let B_2 denote the event that you do not have the disease. The events B_1 and B_2 form a partition. Also, let A denote the event that the response to the test is positive. The event A is information we will learn that tells us something about the partition elements. Then, by Bayes' theorem,

$$\begin{aligned} P(B_1|A) &= \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} \\ &= \frac{(0.9)(0.0001)}{(0.9)(0.0001) + (0.1)(0.9999)} = 0.0009. \end{aligned}$$

□

Remark 1.5-1 Thus, the conditional probability that you have the disease given the test result is approximately only 1 in 1000. Of course, this conditional probability is approxi- mately 9 times as great as the probability was before you were tested, but even the conditional probability is quite small.

Another way to explain this result is as follows: Only one person in every $10,000$ actually has the disease, but the test gives a positive response for approximately one person in every 10. Hence, the number of positive responses is approximately 1000 times the number of persons who actually have the disease. In other words, out of every 1000 persons for whom the test gives a positive response, only one person actually has the disease. This example illustrates not only the use of Bayes' theorem but also the importance of taking into account all of the information available in a problem.

Example 1.5-2 A company producing electric relays has three manufacturing plants producing $50, 30$, and 20 percent, respectively, of its product. Suppose that the probabilities that a relay manufactured by these plants is defective are 0.02, 0.05, and 0.01, respectively.

(a) If a relay is selected at random from the output of the company, what is the probability that it is defective?

(b) If a relay selected at random is found to be defective, what is the probability that it was manufactured by plant 2?

Example 1.5-3 (Test for a disease again) There is a new diagnostic test for a disease that occurs in about 0.05% of the population. The test is not perfect, but will detect a person with the disease 99% of the time. It will, however, say that a person without the disease has the disease about 3% of the time. A person is selected at random from the population, and the test indicates that this person has the disease. What are the conditional probabilities that
(a) the person has the disease?
(b) the person does not have the disease?

Example 1.5-4 In a certain city, 30 percent of the people are Conservatives, 50 percent are Liberals, and 20 percent are Independents. Records show that in a particular election, 65 percent of the Conservatives voted, 82 percent of the Liberals voted, and 50 percent of the Independents voted. If a person in the city is selected at random and it is learned that she did not vote in the last election, what is the probability that she is a Liberal?

Exercises from textbook: Section 1.5: 1, 3, 4, 5, 7, 12.