Probability

1 Introduction and background

Probability theory provides a mathematical framework to describe model, analyze, and solve problems involving random phenomena. Gambling served as a motivation for early mathematicians to develop probability theory, and the connection between gambling and probability can be traced back to the 17th century. It is now used in many applications in finance, engineering, biology, chemistry, mathematics, statistics, and so on.

2 Random experiment

Random experiment is an experiment that even if it is carried out under identical conditions could have a different outcome. Consider an experiment whose outcome is not predictable with certainty. However, although the outcome of the experiment will not be known in advance, let us suppose that the set of all possible outcomes is known. **This set of all possible outcomes of an experiment is known as the sample space of the experiment and is denoted by S.**

Example: Number of car accident on a single day at MSU: $S = \{0, 1, 2, 3, \ldots\}$

Value of a single dice roll: $S = \{1, 2, 3, 4, 5, 6\}$

If the experiment consists of flipping two coins: $S = \{(H, H), (H, T), (T, T), (T, T)\}$

3 Random event

A possible outcome of a random experiment is called an random event.

Example: The sample of the random experiment "throw a die" has 6 random events.

The random experiment of flipping two coins has 4 random events.

If the outcome of an experiment consists in the determination of the sex of a newborn child, then $S = \{girl, boy\}$. Then then one event is that the child is a girl and another event is that the child is a boy.

4 What is probability?

Consider an experiment whose sample space is S. For each event E of the sample space S, we assume that a number P(E) (probability of E) is defined and satisfies the following three properties:

- With probability 1, the outcome will be a point in the sample space S.
- For any point outside of the sample space, the probability of that event is 0.
- $P(E)$ is between 0 and 1 i.e. the probability that the outcome of the experiment is an outcome in E is some number between 0 and 1.

Example: If you toss a coin once, tossing the coin is a random experiment. The sample space is $\{H, T\}$. Your random events are "heads" or "tails". Then,

- with probability 1, the outcome of tossing a coin will be either "heads" or "tails".
- for any event other than "heads" or "tails", the probability is 0.
- the probability that the outcome of the experiment is either "heads" or "tails" is some number between 0 and 1.

5 How do I calculate probability?

Approach one:

One can simply calculate probability based on the ratio of the number of times an event occurs to the number of attempts made.

Refer back to our previous example of tossing a coin: If you toss a coin 10 times, and you get heads 6 times, the probability of having a heads is $\frac{6}{10} = 0.6$

If you roll a die 10 times, and you get a six 2 times, the probability of rolling a six on a die is $\frac{2}{10} = 0.2$

This type of probability is called **empirical probability or experimental probability**.

Approach two:
Consider an experiment whose sample space S is $S = \{1, 2, ..., N\}$. In many experiments, it is natural to assume that all outcomes in the sample space are equally likely to occur. For example, when you flip a coin, heads and tails are equally likely. When you roll a die, 1, 2, 3, 4, 5, and 6 are all equally likely, assuming, of course, that the die is balanced correctly. Then, for any event E in a sample space we can list the probability of any event E equals:

P(E)=$\frac{\text{Number of outcomes in E}}{\text{Number of all possible outcomes in S}}$

This type of probability is called **theoritical probability**.

Refer back to our previous example of tossing a coin: If you toss a coin, the sample space $S = \{H, T\}$. Then, let an event E be getting a tail on the toss. Thus,

P(E)=$\frac{\text{Number of outcomes in E}}{\text{Number of all possible outcomes in S}} = \frac{1}{2}$.

What is the probability of rolling a six on a die?
The sample space $S = \{1, 2, 3, 4, 5, 6\}$. Then, let an event E be getting a six. Thus,

P(E)=$\frac{\text{Number of outcomes in E}}{\text{Number of all possible outcomes in S}} = \frac{1}{6}$.

What is the probability of rolling either one or six on a die?
The sample space $S = \{1, 2, 3, 4, 5, 6\}$. Then, event E is getting an one or a six. Thus,

P(E)=$\frac{\text{Number of outcomes in E}}{\text{Number of all possible outcomes in S}} = \frac{2}{6}$.

I have a deck of well shuffled cards and one card is drawn at a random? What is the probability that the card drawn is a red card?

The sample space $S = 52$ cards. Then, event E is getting a red card. There are 26 possible red cards in a deck of well shuffled cards. Thus,

P(E)=$\frac{\text{Number of outcomes in E}}{\text{Number of all possible outcomes in S}} = \frac{26}{52} = \frac{1}{2}$.

6 But why did empirical probability and theoretical probability differ?

Class Activity: I asked everyone in class to toss a coin 10 times and report the number of times coin landed "heads". The result was a lots of 6s and 4s, few 3s, 5s, and 7s. Thus, on repeating the experiment 10 times, the probability of a coin landing "heads" had a lot of variability.

Then, I asked you to add your group sample and the result was 18 heads in 40 tosses, 33 heads in 60 tosses, and 28 heads in 60 tosses. Thus, as we repeated the experiment few more times, the probability of a coin landing "heads" looked like 0.45, 0.55, and 0.47. An observation that as the number of times an experiment is increased, experimental probability has less variability.

If I add the class sample up, then the result is 79 heads in 160 coin tosses which gives us an experimental probability of 0.49375. The experimental probability is converging to theoretical probability when the number of experiments increases.

Conclusion - If the random experiment is repeated infinitely many times, empirical probability = theoretical probability.

7 Two simple propositions

Mutually exclusive events: By definition, two events E and F are mutually exclusive whenever $E \cap F = 0$. In simpler words, two events are mutually exclusive event when they cannot occur at the same time. For example, if we flip a coin it can only show a head OR a tail, not both.

(i) The probability that an event does not occur is 1 minus the probability that it does occur. For any event E,
P(E does *not* occur) = 1-P(E does occur).
E^c is the complement of E. P(E^c)=1-P(E)

(ii) Let E and F be two random events. The probability that event E happens OR event F happens is (the probability that E happens) plus (the probability that F happens) minus (the probability that both E AND F happen)

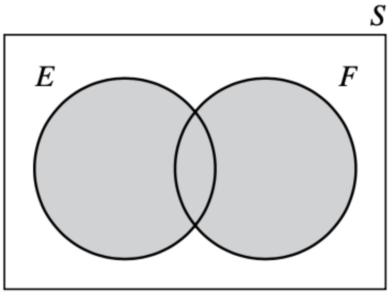
P(E OR F) = P(E) + P(F) - P(E AND F).

The symbol for "or" is $\cup$ and "and" is $\cap$. Then, it can be written as

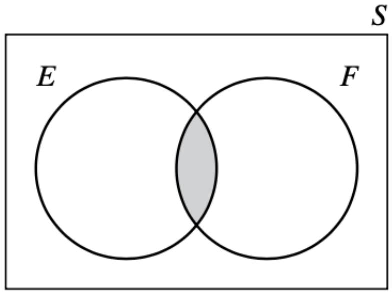
$$P(E \cup F) = P(E) + P(F) - P(E \cap F).$$

If the two events are mutually exclusive,

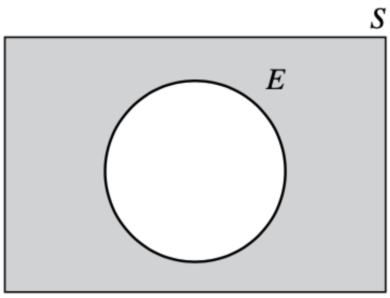
$$P(E \cup F) = P(E) + P(F).$$



(a) Shaded region: $E \cup F$.



(b) Shaded region: EF.



(c) Shaded region: E^c.

Example: Rolling a six-sided die. Roll a fair, six-sided die:
a. Find the probability that the die shows an even number OR a number greater than 4 on top.
Let E be an event that the die shows an even number i.e. $E = \{2, 4, 6\}$. Let F be an event that the die shows a number greater than 4 on top i.e. $F = \{5, 6\}$. Then, $E \cap F = \{6\}$. Thus,

$$P(E \cup F) = P(E) + P(F) - P(E \cap F) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{4}{6}.$$

b. Find the probability that the die shows an even number OR the number 5 on top.

Let E be an event that the die shows an even number i.e. $E = \{2, 4, 6\}$. Let F be an event that the die shows a number 5 on top i.e. $F = \{5\}$. Then, $E \cap F = \{\}$. Thus,

$$P(E \cup F) = P(E) + P(F) - P(E \cap F) = \frac{3}{6} + \frac{1}{6} - \frac{0}{6} = \frac{4}{6}.$$

Example: A total of 28 percent of American males smoke cigarettes, 7 percent smoke cigars, and 5 percent smoke both cigars and cigarettes.

- What percentage of males smoke neither cigars nor cigarettes?
  Let E be an event that the American males smoke cigarettes, F be an event that the the American males smoke cigars. We have, $P(E) = 0.28$, $P(F) = 0.07$, $P(E \cap F) = 0.05$.

  $$P(E \cup F) = P(E) + P(F) - P(E \cap F) = 0.28 + 0.07 - 0.05 = 0.3$$

  Thus, 30% of American males smoke either cigarettes or cigars. So the percentage of males that smoke neither cigars nor cigarettes is $1 - 0.3 = 0.7$ or 70%

- What percentage smoke cigars but not cigarettes?
  Let E be an event that the American males smoke cigars only. Then, $P(E) = 0.07 - 0.05 = 0.02$
  Thus, 2% males smoke cigars but not cigarettes.

Using a venn diagram is always helpful to answer such questions. Similarly, once you get enough practice and feel confident, you do not need to be super formal in your solutions like my notes.

Example: In January 2024, a survey was conducted among 800 New York City residents and asked a number of subway-related questions. One of the questions asked was: "Based on your experience, how would you rate the cleanliness of the New York City subway system? Very clean? Somewhat clean? Somewhat dirty? Or very dirty?" The results are reported in this table:

Survey on Cleanliness of New York City Subway System (January 2024)

How would you rate the cleanliness of the New York City subway system?	**Manhattan**	**Brooklyn**	**Queens**	**Bronx**	**Staten Island**	**Total**
Very Clean	45	35	25	15	10	130
Somewhat Clean	80	90	70	50	20	310
Somewhat Dirty	60	70	60	40	10	240
Very Dirty	40	45	35	20	5	145
Not Sure	10	5	10	5	5	35
Total	**235**	**245**	**200**	**130**	**50**	**800**

- What is the probability that a randomly selected New York resident believes that the subway system is very dirty?

$$P(\text{Very Dirty}) = 145/800$$

- What is the probability that a randomly selected New York City resident lives in Queens?

$$P(\text{Queens}) = 200/800$$

- What is the probability that a randomly selected New York City resident does not believe the subway system is very clean?

$$P(\text{not "Very Clean"}) = 1 - (130/800) = 670/800$$

- What is the probability that a randomly selected New York City resident believes the subway system is very clean AND lives in Manhattan?

$$P(\text{Very Clean AND Manhattan}) = 45/800$$

- What is the probability that a randomly selected New York City resident believes the subway system is very clean **OR** somewhat clean?

$$P(\text{Very Clean OR Somewhat Clean}) = (130/800) + (310/800) = 440/800$$

- Are the two events in the previous question mutually exclusive?

  Yes, because a person cannot believe the subway system is very clean and somewhat clean at the same time.

More example: Suppose we will select a person at random from the collection of 665 people categorized in table below.

Education Level	Single	Married	Divorced	Widow/Widower	Total
Less HS	17	70	10	28	125
High school	68	240	59	30	397
College or higher	27	98	15	3	143
Total	112	408	84	61	665

a. What is the probability that the person is married?

$$P(\text{married}) = 408/655$$

b. What is the probability that the person has a college education or higher?

$$P(\text{college education or higher}) = 143/655$$

c. What is the probability that the person is married OR has a college education or higher?

P(married OR college education or higher) = P(married)+ P(college education or higher) - P(married AND college education or higher)

$$= 408/655 + 143/655 - 98/655 = 453/655$$

8 Conditional probability

Sometimes random events depend on each other since the occurrence of one changes the occurrence of the other.

Let's consider some examples to understand this.

- Imagine you want to know the chances of it raining in the next hour. You could just guess based on general patterns of rain around the world over many years. But a better way to predict the weather would be to look outside right now. If the sky is clear and the sun is shining, you'd probably say there's a low chance of rain soon. This is because the current weather conditions give you important information to make a better guess.
- Say you want to search for valuable shipwrecks in the ocean. You could assume an a priori probability of it being located anywhere in the ocean with uniform likelihood, or you could use historical facts about its disappearance to narrow the shipwreck down to a nontrivial distribution over say a dozen square kilometers.
- If you want to count cards in a card game to gain an advantage, the whole point is to keep track of the current state of the deck (what cards it has and doesn't have) in order to have more accurate predictions about what could be in your competitors' hands.

These are all examples of real world conditional probability.

For example: After taking one card out from the deck there are less cards available, so the probabilities change. Probability of an event A on condition that event B occurs is written as $P(A \mid B)$ and read as "Probability of A given B."

Example: Let's roll a pair of die. Then the following is the list of all 36 possible outcomes.

	1	**2**	**3**	**4**	**5**	**6**
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Now let's calculate the probability of getting at least one '6' while rolling a pair of die.

Now we have the following 11 possible scenarios:

	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Hence the probability of getting at least one '6' while rolling a pair of die is $\frac{11}{36}$. Now imagine a scenario where I told you that you already have a '6' on one of the dies. What is the probability that you get two '6's given that you already have a '6'?

Now your sample space reduces to only these 11 possibilities.

	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

And you only have one outcome where both dies have a '6.' Hence the probability is $\frac{1}{11}$.

This type of probability is called conditional probability and it is calculated using

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

The same example I illustrated above can be calculated using this formula.

What is the probability that you get two sixes when you roll a pair of die given that you already have a six?

Let A be an event I get two sixes while rolling a pair of die and B be an event that I have a six on a roll of die. Now,

$P(A) = \frac{1}{36}$ and $P(B) = \frac{11}{36}$.

$P(A \cap B) = \frac{1}{36}$

$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/36}{11/36} = \frac{1}{11}$.

Example: A coin is flipped twice. Assuming that all four points in the sample space S = $\{(h, h), (h, t), (t, h), (t, t)\}$ are equally likely, what is the probability that both flips land on heads, given that the first flip lands on heads?

Let A = $\{(h, h)\}$ be the event that both flips land on heads; let B = $\{(h, h), (h, t)\}$ be the event that the first flip lands on heads. Now we can calculate the probability $P(A \mid B)$ as

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

Example: A scientist decides to study pregnancy probabilities from the locally available pregnancy test kit at CVS. She observed the test results 1000 participants and attached below are the results.

	Positive Test	Negative Test	Total
Pregnant	270	30	300
Not Pregnant	7	693	700
Total	277	723	1,000

- If a randomly selected study participant receives a positive test, what is the probability they are truly pregnant?

  P(pregnant GIVEN positive test) $= \frac{P(\text{pregnant and positive test})}{P(\text{positive test})} = \frac{270/1000}{277/1000} = 270/277 = 0.9747$

- If a randomly selected study participant receives a negative test, what is the probability they are truly not pregnant?

  P(not pregnant GIVEN negative test) $= \frac{P(\text{not pregnant and negative test})}{P(\text{negative test})} = \frac{693/1000}{723/1000} = 693/723$ $= 0.9585$

- If a randomly selected study participant is truly pregnant, what is the probability that the test will show the participant is pregnant?

  P(+ve test GIVEN pregnant) $= \frac{P(\text{+ve test and pregnant})}{P(\text{pregnant})} = \frac{270/1000}{300/1000} = 270/300 = 0.9$

9 Fun conditional probability problems

Eaxmple 1: Suppose the test for HIV is 99% accurate in both directions and 0.3% of the population is HIV positive. If someone tests positive, what is the probability they actually are HIV positive?

Let D be the event that a person is HIV positive, and T be the event that the person tests positive. So, we want to calculate the probability that the person is HIV positive given that they tested positive ($P(D \mid T)$ which can be written as:

$$P(D \mid T) = \frac{P(D \cap T)}{P(T)}$$

We can easily calculate the probability that a person is HIV positive and tested positive $P(D \cap T)$ as $0.99 \times 0.003 = 0.00297$. Now the little tricky part is how to calculate probability that a person has tested positive.

Let's recall that if a person is tested HIV positive, they might actually be HIV positive or it was just a false positive. So, $P(T)$ is calculated as,

$$\begin{aligned} P(T) &= P(\text{Person has tested HIV positive \& is actually HIV positive}) \\ &\quad + P(\text{Person has tested HIV positive but is not actually HIV positive}) \\ &= (0.99 \times 0.003) + (0.01 \times 0.997) = 0.01294 \end{aligned}$$

Now we substitute the values in our formula as

$$P(D \mid T) = \frac{P(D \cap T)}{P(T)} = \frac{0.00297}{0.01294} = 0.23$$

It's just 23%! A short reason why this surprising result holds is that the error in the test is much greater than the percentage of people with HIV. A detailed reason involves Bayes' theorem which is beyond the scope of this course.

Eaxmple 2: The king comes from a family of 2 children. What is the probability that the other child is his sister?

Let K be the event that we have a King, and S be the event that the king has a sister. So, we want to calculate the probability as ($P(S \mid K)$ which can be written as:

$$P(S \mid K) = \frac{P(S \cap K)}{P(K)}$$

So we need to calculate the probability $P(S \cap K)$ that we have a sister and a king. Recall that we already have a king. Now the only possibility to have a King and a sister is either (King, younger sister), or (Older Sister, King). Similarly, to calculate the probability of a king $P(K)$, we have just three choices i.e. (King, younger sister), (Older Sister, King), or a (King, younger brother).

Hence the probability is $\frac{2}{3}$.

10 Independent events

Let's recall our definition of conditional probability.

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

Now we can write $P(A \cap B)$ as $P(A \cap B) = P(A \mid B) \times P(B)$

If there is a scenario such that random event B does not affect the outcome of A, then we get $P(A \mid B) = P(A)$. We call those type of events independent events, and the formula simplifies to

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

Example: You flip a fair coin twice, what is the probability that you get a "heads" on the second toss given that it landed on "heads" for the first time?

The random event of a coin landing on "heads" for the first time has no effect whatsoever when you toss the coin for the 2nd time. So the probability is still $\frac{1}{2}$.

Example: You flip a **fair coin** hundred time, what is the probability that you get a "heads" on the hundredth toss given that it landed on "tails" for the first 99 times?

The random event of a coin landing on "tails" for the first 99 time has no effect whatsoever when you toss the coin for the hundredth time if it is a fair coin. So the probability is still $\frac{1}{2}$. Sounds absurd, but the theory of independent event still holds.

Example: You enter into a casino and see a game of roulette where there are numbers from 1 to 38. Numbers from 1 to 19 are marked in red whereas numbers from 20 to 38 are marked in black. A person is gambling and they notice that it has been landing only on black numbers for the last 8 times. They decide that for the next spin it is guaranteed to land on a red number as it can't keep landing on black all the time. Is the person right on thinking that assuming that roulette is a fair one?

No. The random event of a roulette landing on black numbers for the last 8 times has no effect whatsoever on the 9th spin. The probability is still $\frac{1}{2}$. The theory of independent event still holds!

Example: The probability if having a non-communicable disease is 0.1. What is the probability that a student in the class will have the disease provided that 5 of their classmates have that disease?

The probability is still 0.1. The theory of probability of independent events like in previous examples hold true.

Example: You are at a grocery store, and you pick a carton of eggs. The probability of finding a cracked egg in the carton is 5%. What is the probability that the next carton you pick will also have a cracked egg if the first carton you picked had one?

The random event of finding a cracked egg in the first carton has no effect on the probability of finding a cracked egg in the next carton. So, the probability is still 5%. The theory of independent events holds true here as well.

Example: You flip a fair coin eight times. Which one of these two events is more likely?

Event A: The probability of getting all heads i.e. $\{H, H, H, H, H, H, H, H\}$

Event B: Heads and tails in the order $\{H, T, H, T, T, H, H, T\}$

Both of them have an equal probability of occurrence. Every flip of a fair coin eight times is an independent event and don't rely on the prior flip. To answer the question posed above, both the events have an equal probability of 1/256. In general, all the 256 different possible outcomes have an equal probability of 1/256. It is no different to getting all heads in a row, or all tails in a row, or alternating heads and tails.