Lecture 1
The Concept of Probability

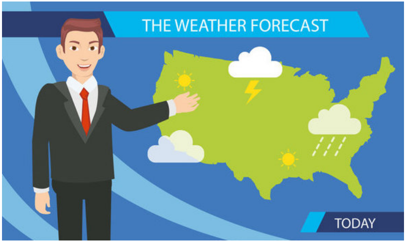
Chao Song

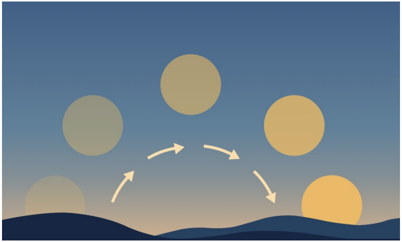
College of Ecology
Lanzhou University

September 1, 2025

What is probability?

Probability is the branch of mathematics concerning numerical descriptions of how likely an event is to occur; Probability focuses on events that are uncertain. Events that occurs with certainty is often not within the scope of probability.



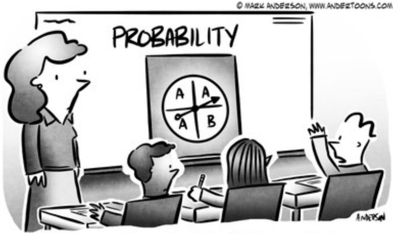


Events with uncertain outcomes, such as the weather tomorrow, is within the scope of probability while events with certain outcomes, such as the sunrise time tomorrow, is not.

What is probability?

We may interpret probability in various intuitive ways.

- **Frequentist view**: Probability measures how frequent an outcome occur if we perform the experiment repeatedly under the same condition;
- **Bayesian view**: Probability measures our belief on how likely a particular outcome is going to occur.



"I know mathematically that A is more likely, but I gotta say, I feel like B wants it more."

"I wish we hadn't learned probability 'cause I don't think our odds are good."

Essential terminologies to define probability

Random experiment: experiment for which the outcome cannot be predicted with certainty;

Sample space (*S*): the collection of all possible outcomes;

Event (*A*): a collection of outcomes that is part of the sample space, i.e., $A \subset S$.

When the random experiment is performed and the outcome of the experiment is in A, we say that **event A has occurred**.

Essential terminologies to define probability

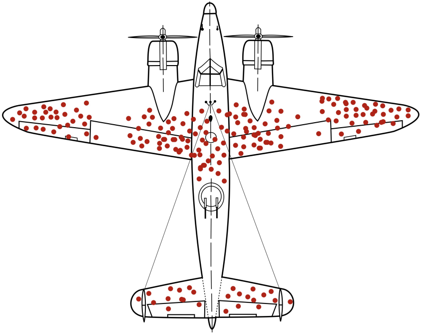
Example: Flip a coin 2 times and record the side facing up each time.

- Sample space $S = \{HH, HT, TH, TT\}$
- Event A_1: {get all heads}$= \{HH\}$
- Event A_2: {get at least one head}$= \{HH, HT, TH\}$
- Event A_3: {get both head and tail}$= \{HT, TH\}$

A note on sample space

Sample space is a seemingly intuitive concept. In practice, however, one needs to be careful in understanding what the sample space is.

Survival bias: the logical error of concentrating on entities that passed a selection process while overlooking those that did not.



Using patterns of damage on surviving planes to identify locations of reinforcement is an example of survival bias

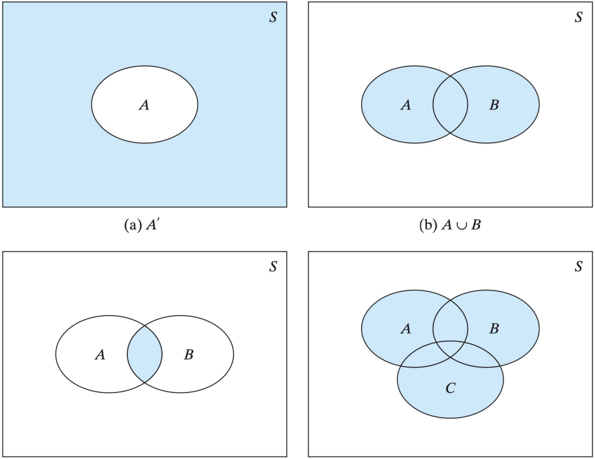
Probability and algebra of sets

In studying probability, the words **event** and **set** are often interchangable. Algebra of sets can be useful when calculating probability.

- **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 Morgan's laws**:
  $(A \cup B)' = A' \cap B'$
  $(A \cap B)' = A' \cup B'$

Probability and algebra of sets

Venn diagram is a useful tool to visualize relationship among events and help calculate probability.



(a) A' (b) $A \cup B$ (c) $A \cap B$ (d) $A \cup B \cup C$

Calculate probability

Example: Flip a fair coin 2 times and record the side facing up each time.

- Sample space $S = \{HH, HT, TH, TT\}$
- Event A_1: {get all heads}$= \{HH\}$
- Event A_2: {get at least one head}$= \{HH, HT, TH\}$
- Event A_3: {get both head and tail}$= \{HT, TH\}$

Each sampling point in the sample space occurs equally likely. We thus can calculate probability by counting the number of elements in the event and the sample space.

- $P(A_1) = \frac{1}{4}$;
- $P(A_2) = \frac{3}{4}$;
- $P(A_3) = \frac{2}{4}$

Classical model of probability

Let the sample space of a random experiment S be a finite set and $^{\#}$ denotes the number of sampling points in the set. If the chance of occurrence for every sampling point is equal, the probability of event A is

$$P(A) = \frac{^{\#}A}{^{\#}S}.$$

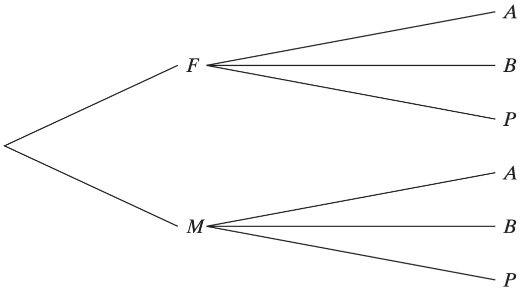
This model of probability is referred to as the **classical model of probability**. Its development is attributed to Jacob Bernoulli and Pierre-Simon Laplace.

Jacob Bernoulli (1655–1705) and Pierre-Simon Laplace (1749–1827)

Method of enumeration

Based on the definition of the classical model of probability, probability can be calculated by **method of enumeration**, i.e. counting number of sampling points in event and sample space.

Multiplication principle: Suppose that an experiment E_1 has n_1 outcomes, and for each of these possible outcomes, an experiment E_2 has n_2 possible outcomes. Then the composite experiment $E_1 E_2$ has $n_1 n_2$ outcomes.



Method of enumeration

Sample with replacement: the number of possible ordered samples of size r taken from a set of n objects is n^r when sampling with replacement.

Sample without replacement: the number of possible ordered samples of size r taken from a set of n objects is $n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}$ when sampling without replacement

Method of enumeration

Arrangement: number of ways of arranging or ordering r different objects chosen from n different objects.

$$\begin{aligned} A_n^r &= n(n-1)\cdots(n-r+1) \\ &= \frac{n!}{(n-r)!} \end{aligned}$$

Combination: number of ways of choosing r different objects from n different objects when the order of the r objects is disregarded.

$$\begin{aligned} C_n^r &= \frac{A_n^r}{A_r^r} \\ &= \frac{n!}{r!(n-r)!} \end{aligned}$$

Method of enumeration

Example: Three students (S) and six faculty members (F) attend a meeting. In how many ways can the nine participants be lined up at the table?

Answer: This is a simple arrangement problem. Number of ways to arrange 9 people is $A_9^9 = 9! = 362880$.

Example: Three students (S) and six faculty members (F) attend a meeting. How many line ups are possible, considering only the label S and F?

Answer: Once you determine which 3 seats are for students, you have your line up determined. Thus, the number of lines ups when only considering S and F is $C_9^3 = 84$.

Method of enumeration

Example: There are 6 female and 16 male athletes in the team. Randomly select 11 athletes to participate in the Olympics Game. What is the probability that exactly 3 females athletes are selected?

Let A denote the event that 3 female athletes are selected.

$$
\begin{aligned}
{}^{\#}S &= \mathbf{C}_{22}^{11} = \frac{22!}{11!(22-11)!} = 705432 \\
{}^{\#}A &= \mathbf{C}_{6}^{3}\mathbf{C}_{16}^{8} = 257400 \\
P(A) &= \frac{{}^{\#}A}{{}^{\#}S} = 0.365
\end{aligned}
$$

Method of enumeration

Example: A disk 20 centimeters in diameter is thrown at random on a tiled floor, where each tile is a square with sides 40 centimeters in length. Let C be the event that the disk lands entirely on one tile, what is the probability of C?

If the disk land entirely in one tile, the center of the disk must lie within a square having a side length of 20 centimeters and with its center coincident with the tile. Since the area of this square is 400 cm^2 and the area of the tile is 1600 cm^2, $P(C) = 400/1600 = 0.25$

Limitations of the classical model of probability

The classical model of probability only applies to scenarios when sampling points are equally likely. When sampling points do not occur with equal probability, the classical model of probability and the method of enumeration does not work any more.

Frequentist interpretation of probability: Let A be an event and f_N be the frequency of event A occurring in N repeated experiments. Probability of event A can be defined as $P(A) = \lim_{N \to \infty} f_N$.

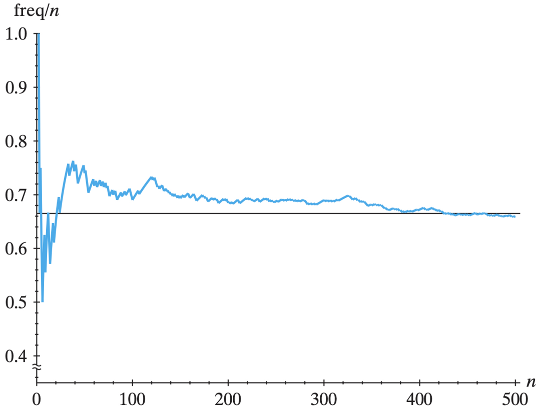
Frequentist interpretation of probability

The frequentists definition of probability has been extensively tested by statisticians using the coin flipping experiment. These experiments show that as N increases, frequency of getting heads approaches 0.5.

Experimenter	N	N_{head}	N_{heads}/N
De Morgan	2048	1061	0.5181
Buffon	4040	2048	0.5069
Kerrich	7000	3516	0.5023
Kerrich	8000	4034	0.5043
Kerrich	9000	4538	0.5042
Kerrich	10000	5067	0.5067
Feller	10000	4979	0.4979
Pearson	12000	6019	0.5016
Pearson	24000	12012	0.5005
Romanovsky	80640	40173	0.4982

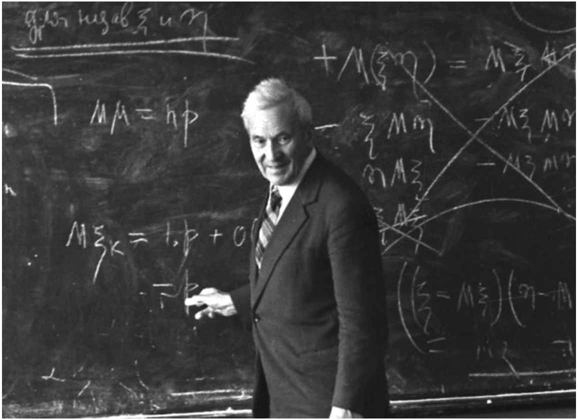
Frequentist interpretation of probability

A fair die is rolled six times. If the face numbered k is the outcome on roll k for $k = 1, 2, ..., 6$, we say that a match has occurred. The experiment is called a success if at least one match occurs. What is the probability of success?



Axiomatic definition of probability

The axiomatic definition of probability was introduced by **Andrey Kolmogorov** in his 1933 book Foundations of the Theory of Probability.



Andrew N. Kolmogorov (1903–1987)

Axiomatic definition of probability

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

- $P(A) \geqslant 0$;
- $P(S) = 1$;
- if $A_1, A_2, A_3 \dots$ are events and $A_i \cap A_j = \varnothing$, $i \neq j$, then

  $$P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_k) = P(A_1) + P(A_2) + P(A_3) + \ldots P(A_k)$$

  for each positive integer k and

  $$P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots$$

  for an infinite, but countable, number of events.

Properties of probability

Theorem: For each event A, $P(A) = 1 - P(A')$.
Proof: Because $A \cap A' = \varnothing$, we have $P(A \cup A') = P(A) + P(A')$
Also because $A \cup A' = S$, $P(A \cup A') = P(S) = 1$
Therefore, $P(A) = 1 - P(A')$.

Theorem: $P(\varnothing) = 0$.
Proof: Using the theorem above, take $A = \varnothing$ so that $A' = S$
$P(\varnothing) = 1 - P(S) = 0$

Properties of probability

Theorem: If events A and B are such that $A \subset B$, $P(A) \leqslant P(B)$;
Proof: We have $B = A \cup (B \cap A')$ and $A \cap (B \cap A') = \varnothing$.
Therefore, $P(B) = P(A) + P(B \cap A') \geqslant P(A)$,
because $P(B \cap A') \geqslant 0$.

Theorem: For each event A, $P(A) \leqslant 1$.
Proof: Because $A \subset S$, from the theorem above,
we have $P(A) \leqslant P(S) \leqslant 1$

Properties of probability

Theorem: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
Proof: The event $A \cup B$ can be represented as a union of disjoint events:

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

By the definition of probability,

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

Event B can be represented as the union of disjoint events:

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

By the definition of probability

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

Taken together, we have

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

Properties of probability

Based on the axiomatic definition of probability, we have:

- For each event A, $P(A) = 1 - P(A')$;
- $P(\varnothing) = 0$;
- If events A and B are such that $A \subset B$, $P(A) \leqslant P(B)$;
- For each event A, $P(A) \leqslant 1$;
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
- If events A and B are disjoint, $P(A \cup B) = P(A) + P(B)$.

Properties of probability

Example: A fair coin is slipped successively until the same face is observed on successive flips. Let A be the event that it will take 3 or more flips to observe the same face on two successive flips. What is $P(A)$?

Let $A = \{x : x = 3, 4, 5, \ldots\}$. Then $A' = \{x : x = 2\}$. In two flips of a coin, the possible outcomes are $\{HH, TT, HT, TH\}$. Each of the four outcomes has the same chance to be observed. Thus,

$$P(A') = P(HH, TT) = \frac{2}{4}$$

Thus, using the properties of probability,

$$P(A) = 1 - P(A') = 1 - \frac{2}{4} = \frac{1}{2}.$$

Properties of probability

Example: If A, B, and C are any three events, what is $P(A \cup B \cup C)$?

We first note $A \cup B \cup C = A \cup (B \cup C)$. Then

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

Using distributive laws of set operation, we have

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

We therefore have

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

Conditional probability: a motivating example

Example: A deck of cards (52 cards without jokers) is well shuffled and one card is drawn randomly. What is the probability that the card is a K?

$$P(K) = \frac{4}{52} = \frac{1}{13}$$

What is the probability of drawing a K if the card is known to be a face card (J, Q. K)?

$$P(K|J, Q, K) = \frac{4}{12} = \frac{1}{3}$$

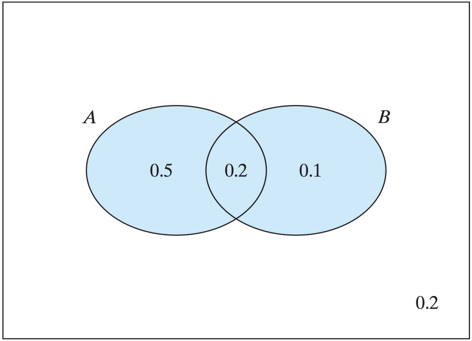
Why do the two probabilities differ ?

Current knowledge (face card) has changed or restricted the sample space.

Conditional probability

Given two events A and B. We denote the probability of event A happens **given** that event B is known to happen as $P(A|B)$.

We can think of “given B” as specifying the new sample space for which, to determine $P(A|B)$, we now want to calculate the probability of that part of A that is contained in B.



Conditional probability

The conditional probability of event A given event B can be calculated as

$$P(A|B) = \frac{P(AB)}{P(B)}$$

provided that $P(B) > 0$

Conditional probability satisfies the axioms for a probability function:

- $P(A|B) \geqslant 0$
- $P(B|B) = 1$
- if $A_1, A_2, A_3 \dots$ are mutually exclusive events, then

  $$P(A_1 \cup A_2 \cup \dots \cup A_k|B) = P(A_1|B) + P(A_2|B) + \dots + P(A_k|B)$$

  for each positive integer k, and

  $$P(A_1 \cup A_2 \cup \dots |B) = P(A_1|B) + P(A_2|B) + \dots$$

  for an infinite but countable number of events.

Conditional probability

Example: A common test for AIDS is called ELISA. Among 1 million people who are given the test, we obtain results in the following table

	B_1: AIDS	B_2: No AIDS	Totals
A_1: Positive	4885	73630	78515
A_2: Negative	115	921370	921485
Totals	5000	995000	1000000

Find the following probabilities: $P(B_1)$, $P(A_1)$, $P(A_1|B_2)$, $P(B_1|A_1)$.

General multiplication rule

Intuitively, we can view events A and B both occur as a two step process: event B occurs and then event A occur given that event B has already occurred.

The formula for conditional probability can be used the other way around. Multiplying both side by $P(B)$, we get the **general multiplication rule**:

$$P(AB) = P(B)P(A|B)$$

General multiplication rule for several events

The general multiplication rule can be extended to more than two events:

$P(ABC) = P(A) \times P(B|A) \times P(C|A, B)$;

$P(ABCD) = P(A) \times P(B|A) \times P(C|A, B) \times P(D|A, B, C)$;

$P(ABCDE) = P(A) \times P(B|A) \times P(C|A, B) \times P(D|A, B, C) \times P(E \mid A, B, C, D)$.

General multiplication rule

Example: A box contains 7 blue balls and 3 red balls. We randomly draw two balls successively without replacement. We want to compute the probability that the first draw is a red ball (A) and the second draw is a blue ball (B).

Using the definition of conditional probability, we have

$$P(AB) = P(A)P(B|A) = \frac{3}{10} \cdot \frac{7}{9} = \frac{7}{30}$$

Using the method of enumeration, we have

$$P(AB) = \frac{\mathbf{C}_3^1 \cdot \mathbf{C}_7^1}{\mathbf{A}_{10}^2} = \frac{7}{30}$$

Comment: We can compute probability by two seemingly different methods provided that our reasoning is consistent with the underlying assumptions.

Independent events

Two events are **independent** if knowing the outcome of one provides no useful information about the outcome of the other.

- **Independent**: Knowing that the coin landed on a head on the first toss does not provide any information for determining what the coin will land in the second toss.
- **Dependent**: Knowing that the first card drawn from a deck is an ace provide information on determining the probability of drawing an ace in the second draw.

Independent events

Definition: Events A and B are independent if and only if $P(AB) = P(A) \times P(B)$. In other words, $P(AB) = P(A) \times P(B)$ is a necessary and sufficient condition for A and B being independent.

- This comes from the general multiplication rule. $P(AB) = P(A) \times P(B|A)$ in which $P(B|A)$ reduces to $P(B)$ when events A and B are independent.
- More generally, events $A_1, A_2, \ldots A_k$ are **mutually independent** if and only if for any subset of the j events,

$$P(A_{i_1} A_{i_2} \ldots A_{i_j}) = P(A_{i_1}) \times P(A_{i_2}) \times \ldots P(A_{i_j})$$

Independent events

Question: If events A and B are independent, events B and C are independent, are events A and C necessarily independent?

Not necessary. For example, if A and C are the same events, and A and B are independent, we know that B and C are independent. But A and C are certainly not independent.

Question: If two events A and B are disjoint with $P(A) \neq 0$ and $P(B) \neq 0$, are they independent?

Disjoint events that each does not have a zero probability are never independent because knowing one event happening tells you for certain that the other will not happen. Formally, if A and B are disjoint, by definition of disjoint events, $P(AB) = 0$. Now $P(A) \neq 0$ and $P(B) \neq 0$. Thus $P(AB) \neq P(A)P(B)$.

Independent events

Example: A box contains four balls numbered 1, 2, 3, and 4. One ball is drawn at random from the box. Let events A, B, and C be defined by $A = \{1, 2\}$, $B = \{1, 3\}$, and $C = \{1, 4\}$. Then $P(A) = P(B) = P(C) = 1/2$. Furthermore,

$$P(AB) = \frac{1}{4} = P(A)P(B)$$
$$P(AC) = \frac{1}{4} = P(A)P(C)$$
$$P(BC) = \frac{1}{4} = P(B)P(C)$$

However, since $A \cap B \cap C = 1$, we have

$$P(ABC) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C)$$

Implications: Pairwise independence is not sufficient for complete independence of multiple events.

Independent events

As estimated in 2012, of the US population,

- 13.4% were 65 or older, and
- 52% of the population were male.

True or False: $0.134 \times 0.52 \approx 7\%$ of the US population were males aged 65 or older.

The answer is **false**

- Age and gender are not independent. On average women live longer than men. There are more old women than old men;
- According to survey data, among those 65 or older in the US, 44% are male, not 52%. Thus, $0.134 \times 0.44 \approx 5.9\%$ were males aged 65 or older in the US in 2012.

Law of total probability

Law of total probability: If events A_1, A_2, ..., A_n are pairwise disjoint events, $B \subset \bigcup_{k=1}^{n} A_k$, then

$$P(B) = \sum_{k=1}^{n} P(A_k)P(B|A_k)$$

Corollary 1: If events A_1, A_2, ..., A_n is a partition of the sample space, i.e., a set of pairwise disjoint events whose union is the entire sample space, then

$$P(B) = \sum_{k=1}^{n} P(A_k)P(B|A_k)$$

Corollary 2: For any events A and B, the probability of B can be calculated as $P(B) = P(A)P(B|A) + P(A')P(B|A')$.

Law of total probability

Seroprevalence adjustment: China Center for Disease Control and Prevention conducted a serological survey in April, 2020 in Wuhan and found 4.43% positive test. The test has a sensitivity of 90% and specificity of 98%. Does that mean that 4.43% of the Wuhan population were once infected?

Solution: Let A denote a positive test and B denote a true infection. Using the law of total probability, we have

$$P(A) = P(B)P(A|B) + P(B')P(A|B').$$

- $P(A) = 4.43\%$;
- Sensitivity $P(A|B) = 0.9$;
- Specificity $P(A'|B') = 0.98$;

Plug these values to the equation above, we get

$$0.0443 = P(B) \times 0.9 + (1 - P(B)) \times (1 - 0.98)$$

Solving the equation above for P(B), we arrive at $P(B) = 2.76\%$.

Bayes' theorem

Bayes' theorem, named after Thomas Bayes, describes the probability of an event, based on prior knowledge of conditions that might be related to the event:

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

where A and B are events and $P(B) \neq 0$.

Thomas Bayes (1702–1761)

Bayes' theorem

Disease survey problem: Wu et al. (2019) estimated that people living with human immunodeficiency virus (HIV) has risen to more than 1.25 million in China, roughly 0.09% of the total population. A blood test for HIV typically has 95% accuracy, i.e., the test correctly detects positive cases or negative case 95% times. If a person is tested positive, what is the probability that this person is infected with HIV?

Solution: Let A denote that a person has HIV and B denote a positive test. We know $P(A) = 0.0009$, $P(B|A) = 0.95$, and $P(B'|A') = 0.95$. We want to know $P(A|B)$.

Using Bayes' theorem, we have

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{0.0009 \times 0.95}{P(B)},$$

Bayes' theorem

Using law of total probability, we have

$$\begin{aligned} P(B) &= P(A)P(B|A) + P(A')P(B|A') \\ &= 0.0009 \times 0.95 + (1 - 0.0009) \times (1 - 0.95) \\ &= 0.05081 \end{aligned}$$

Finally, we have $P(A|B) = 1.68\%$

What does this tell us?

For a very rare disease, if you get a positive test, the chance that you indeed have that disease is very low.

Bayes' theorem

Disease survey problem: Will a more accurate test help? If we improve the test sensitivity and specificity to 99.9%. What is the probability of having HIV if a person tested positive?

Solution: Using Bayes' theorem and the law of total probability:

$$
\begin{aligned}
P(A|B) &= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A')P(B|A')} \\
&= \frac{0.0009 \times 0.999}{0.0009 \times 0.999 + (1 - 0.0009) \times (1 - 0.999)} \\
&= 47.36\%.
\end{aligned}
$$

Implications: We should be cautious when confirming a rare disease, even if the diagnostic test is highly accurate.