Lecture 2
Conditional Probability and Bayes' Theorem

Chao Song

College of Ecology

Lanzhou University

September 4, 2024

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)?

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

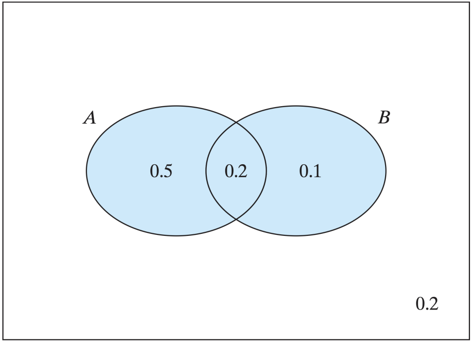
Why do the two probabilities differ ?

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 ... are mutually exclusive events, then

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

  for each positive integer k, and

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

  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, if events A_1, A_2, ... A_k are independent,

$$P(A_1 A_2 \ldots A_k) = P(A_1) \times P(A_2) \times \ldots P(A_k)$$

Independent events

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

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

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.