Lecture 5: Conditional Probability

Ye Tian

Department of Statistics, Columbia University
Calculus-based Introduction to Statistics (S1201)

July 12, 2022
Recap: product rule and permutations

**Product rule:** If there are $k$ steps to finish a task, and:

- There are $n_1$ ways to do the first step
- Given any choice in the first step, there are $n_2$ ways to do the second step
- ...
- Given any choices in the first $(k-1)$ steps, there are $n_k$ ways to do the $k$-th step

Then finally there are $n_1 n_2 \cdots n_k$ ways to finish this task.

**Permutations:** Any ordered sequence of $k$ objects taken from a set of $n$ distinct objects is called a *permutation*. The total number of permutations equals

$$P_{k,n} = n(n - 1)(n - 2) \cdots (n - k + 1) = \frac{n!}{(n - k)!}.$$
Recap: combinations

Any of \( k \) objects (unordered) taken from a set of \( n \) distinct objects is called a combination. The number of combinations obtained when selecting \( k \) objects from \( n \) distinct objects to form a group is given by

\[
\binom{n}{k} = \frac{P_{k,n}}{k!} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\ldots(n-k+1)}{k(k-1)\ldots1}.
\]
Recap: permutation problem or combination problem?

(1) Choose 5 people from 100 to do a job
(2) Choose the winner of the first prize and second prize from 20 people
(3) Choose two winners of the first prize from 20 people
(4) Draw 4 cards from a 52-card deck (the order doesn't matter)
(5) Draw 4 cards from a 52-card deck (the order matters)
(6) Figure out a way to put 5 people in a line to get the vaccine
Recap: probability calculation

**Theorem:** In an experiment consisting of $N$ outcomes with equal probability, denote the number of outcomes that event $A$ contains as $N(A)$. Then for any event $A$,

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

Counting techniques help us to calculate $N(A)$ and $N$. 
Today's goal

- Understand conditional probabilities
- Know the multiplication rule and rule of total probability
- Understand Bayes Theorem and its statistical meanings
- Use all techniques we learn to calculate the probability
Conditional Probability
Conditional probability

◦ We have learned that given the experiment and the sample space $\Omega$, for an event $A$, $P(A)$ represents the probability that $A$ occurs, where $P(\cdot)$ is a function mapping any event $A \in \Omega$ to a number between 0 and 1.

◦ Here we want to discuss how the information "an event $B$ has occurred" affects the probability assigned to $A$

◦ Many events are actually related and the event that has happened can provide information about the event that hasn't occurred yet.

**Examples:**

◦ Two Columbia students, Claire and Jack, are sampled to do a Covid test.
  ▶ Claire feels good while Jack had a fever last night.
  ▶ Who is more likely to test positive?

◦ I randomly sample a person on the street in NYC and ask whether he/she knows me or not.
  ▶ Will the probability that this guy knows me be very high?
  ▶ What if I know this guy is a Columbia statistics Ph.D student?
### Conditional probability

**Definition:** Conditional probability of A given B is defined as

\[
P(A|B) = \frac{P(A \cap B)}{P(B)}.\]

**Remark:**

- It represents the probability that A occurs given that B has occurred.
- We require \(P(B) > 0\) to make it useful.
- **Fixing an event** \(B\), we can see \(P(·|B)\) as a probability function, which satisfies all properties of probability we described before. For example:
  - \(0 \leq P(A|B) \leq 1\) for any event \(A\), \(P(\emptyset|B) = 0\), \(P(\Omega|B) = 1\)
  - \(P(A \cup C|B) = P(A|B) + P(C|B) - P(A \cap C|B)\) for any \(A\) and \(C\)
  - \(P(\bigcup_{i=1}^{n} A_i|B) = \sum_{i=1}^{n} P(A_i|B)\) if \(A_1, \ldots, A_n\) are mutually exclusive
  - \(P(·|B)\) is a function mapping any event \(A \in \Omega\) to a number between 0 and 1.
- \(\Omega\) is the sample space when we study \(P(A)\), and \(\Omega \cap B = B\) can be seen as the new sample space when we study \(P(A|B)\)!
Example

**Conditional probability** of A given B is defined as

\[ P(A|B) = \frac{P(A \cap B)}{P(B)}. \]

**Example:** Suppose that of all individuals buying a certain digital camera:

- 60% include an optional memory card in their purchase
- 40% include an extra battery
- 30% include both a card and battery.

Consider randomly selecting a buyer, what's the probability that the buyer purchased the memory card given that he/she purchased the battery?

**Solution:** Let \( A = \{\text{memory card purchased}\} \) and \( B = \{\text{battery purchased}\} \). We want to know \( P(A|B) \).

Therefore, \( P(A) = 0.6 \), \( P(B) = 0.4 \) and \( P(A \cap B) = 0.3 \). Hence,

\[ P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.3}{0.4} = 0.75. \]
More examples

A news magazine publishes three columns entitled "Art" (A), "Books" (B), and "Cinema" (C). Reading habits of a randomly selected reader with respect to these columns are

<table>
<thead>
<tr>
<th>Read regularly</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>A ∩ B</th>
<th>A ∩ C</th>
<th>B ∩ C</th>
<th>A ∩ B ∩ C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>.14</td>
<td>.23</td>
<td>.37</td>
<td>.08</td>
<td>.09</td>
<td>.13</td>
<td>.05</td>
</tr>
</tbody>
</table>

What's the probability that the selected individual regularly reads the Art column given that he or she regularly reads at least one of the other two columns?

Solution:

\[
P(A|B \cup C) = \frac{P(A \cap (B \cup C))}{P(B \cup C)} = \frac{0.04 + 0.03 + 0.05}{0.2 + 0.04 + 0.08 + 0.05 + 0.03 + 0.07} = 0.255.
\]
The multiplication rule

**Multiplication rule:**

- \( P(A \cap B) = P(A)P(B|A) \)
- \( P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B) \)
- \( P(\bigcap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots A_{n-1}) \).

**Remark:**

- This is very useful when calculating the probability of intersection of events
- Intuition: process information "step by step"
Example

A box contains 10 tickets. Three of them are winning tickets. Pick three at random without replacement.

- What is the chance that the first three randomly picked tickets are the winning ones?
- What is the chance that the second ticket is a winning one?

**Solution 1:** Let \( A \) and \( B \) be the event of interest in two subproblems.

- View it as an **unordered** problem: Sample space
  \[ \Omega = \{\text{all ways to draw 3 from 10 tickets}\} \]
  We draw 3 tickets at the same time and the order doesn't matter.
  \[ N = N(\Omega) = \binom{10}{3} = 120 \]
  \[ N(A) = \binom{3}{3} = 1. \] Hence, \( P(A) = \frac{N(A)}{N} = 1/120. \)

- View it as an **ordered** problem: Sample space
  \[ \Omega = \{\text{all ways to draw 3 from 10 tickets in 3 steps}\} \]
  In each of three steps, we draw one ticket.
  By product rule: \( N = N(\Omega) = 10 \times 9 \times 8 = 720 \)
  \[ N(A) = 3 \times 2 \times 1 = 6. \] Hence, \( P(A) = \frac{N(A)}{N} = 1/120. \)
Example

A box contains 10 tickets. Three of them are winning tickets. Pick three at random **without replacement**.

- What is the chance that the first three randomly picked tickets are the winning ones?
- What is the chance that the second ticket is a winning one?

---

**Solution 2:** Let $W_i$ be the event that the $i$-th draw is a winning ticket.

\[
\mathbb{P}(W_1 \cap W_2 \cap W_3) = \mathbb{P}(W_1)\mathbb{P}(W_2|W_1)\mathbb{P}(W_3|W_1 \cap W_2)
\]

\[
= \frac{3}{10} \times \frac{2}{9} \times \frac{1}{8}
\]

\[
= \frac{1}{120}
\]

Think about how we obtain the values of $\mathbb{P}(W_1)$, $\mathbb{P}(W_2|W_1)$, $\mathbb{P}(W_3|W_1 \cap W_2)$. 
Example

A box contains 10 tickets. Three of them are winning tickets. Pick three at random without replacement.

- What is the chance that the first three randomly picked tickets are the winning ones?
- What is the chance that the second ticket is a winning one?

Solution 1: We have to view it as an ordered problem. Denote the event as $B$. Sample space $\Omega = \{\text{all possible ways to select 3 from 10 tickets in 3 steps}\}$. In each of three steps, we draw one ticket.

By product rule: $N = N(\Omega) = 10 \times 9 \times 8 = 720$

Event $B$ contains all outcomes in the form "$XWXX$", where $X$ can be any tickets while $W$ is the winning one.

$N(B) = P_{1,3} \times P_{2,10-1} = 3 \times 9 \times 8$. Hence,

$P(B) = \frac{N(B)}{N} = 3/10.$
Example

A box contains 10 tickets. Three of them are winning tickets. Pick three at random **without replacement**.

- What is the chance that the first three randomly picked tickets are the winning ones?
- What is the chance that the second ticket is a winning one?

---

**Solution 2:** Let $W_i$ be the event that the $i$-th draw is a winning ticket.

\[
\begin{align*}
\mathbb{P}(W_2) &= \mathbb{P}(W_1 \cap W_2) + \mathbb{P}(W_1^c \cap W_2) \\
&= \mathbb{P}(W_1)\mathbb{P}(W_2|W_1) + \mathbb{P}(W_1^c)\mathbb{P}(W_2|W_1^c) \\
&= \frac{3}{10} \times \frac{2}{9} + \frac{7}{10} \times \frac{3}{9} \\
&= \frac{3}{10}
\end{align*}
\]

Think about how we obtain the values of $\mathbb{P}(W_1)$, $\mathbb{P}(W_2|W_1)$, $\mathbb{P}(W_2|W_1^c)$. 

\[P(W_2) = P(W_1 \cap W_2) + P(W_1^c \cap W_2) = P(W_1)P(W_2|W_1) + P(W_1^c)P(W_2|W_1^c) = \frac{3}{10} \times \frac{2}{9} + \frac{7}{10} \times \frac{3}{9} = \frac{3}{10} \]
Rule of total probability

**Rule of total probability:** If $\Omega = A_1 \cup A_2 \cup \cdots \cup A_k$ and $A_1, \ldots, A_k$ are mutually exclusive (such $A_1, \ldots, A_k$ are called a *partition* of the sample space), then for any event $B$,

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

**Proof:** $B = \bigcup_{i=1}^{k} (A_k \cap B)$ where these $A_k \cap B$ are mutually exclusive (why?). Then $P(B) = \sum_{i=1}^{k} P(A_k \cap B) = \sum_{i=1}^{k} P(A_k)P(B|A_k)$.

**Remark:** Sometimes the conditional probability $P(B|A_k)$ is much straightforward to calculate than $P(B)$. Then the rule of total probability can be useful.
Example

An individual has 3 different email accounts.
- 70% of her messages come into account #1
- 20% come into account #2
- the remaining 10% into account #3

Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively. What is the probability that a randomly selected message is spam?

Solution: First we translate the problem by math language. Denote $A_i = \{\text{message is from account } \#i\}$, $i = 1, 2, 3$, $B = \{\text{message is spam}\}$. The problem implies that

$$
P(A_1) = 0.7, \quad P(A_2) = 0.2, \quad P(A_3) = 0.1$$

$$
P(B|A_1) = 0.01, \quad P(B|A_2) = 0.02, \quad P(B|A_3) = 0.05.$$ 

Finally by rule of total probability,

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) = 0.016.$$
Bayes' Theorem

Bayes' Theorem: If $A_1, \ldots, A_k$ are a partition of sample space $\Omega$, then for any event $B$,

$$
\mathbb{P}(A_j|B) = \frac{\mathbb{P}(A_j \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_j)\mathbb{P}(A_j)}{\sum_{i=1}^{k} \mathbb{P}(B|A_i)\mathbb{P}(A_i)}, \quad j = 1, \ldots, k.
$$

Remark:

- Proof: by rule of total probability

- In many cases, $B$ is the "consequence" while $A_1, \ldots, A_k$ are different "conditions" or "causes" of $B$. We may know each $\mathbb{P}(B|A_i)$ (given cause $A_i$, the probability that consequence $B$ happens), and would like to infer "the probability that the consequence $B$ is caused by $A_j"$, i.e. the probability that $A_j$ occurs given that $B$ has occurred (which is $\mathbb{P}(A_j|B)$).
Example (revisited)

An individual has 3 different email accounts.
- 70% of her messages come into account #1
- 20% come into account #2
- the remaining 10% into account #3

Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively. We randomly select a spam message. What's the probability that it is from account #2?

Solution: Denote $A_i = \{\text{message is from account } #i\}$, $i = 1, 2, 3$, $B = \{\text{message is spam}\}$. The problem implies that

\[
\begin{align*}
\mathbb{P}(A_1) &= 0.7, \quad \mathbb{P}(A_2) = 0.2, \quad \mathbb{P}(A_3) = 0.1 \\
\mathbb{P}(B|A_1) &= 0.01, \quad \mathbb{P}(B|A_2) = 0.02, \quad \mathbb{P}(B|A_3) = 0.05.
\end{align*}
\]

Then

\[
\mathbb{P}(A_2|B) = \frac{\mathbb{P}(B|A_2)\mathbb{P}(A_2)}{\mathbb{P}(B)} = \frac{0.02 \times 0.2}{0.016} = 0.25.
\]
**Example (revisited)**

Denote $A_i = \{\text{message is from account } \#i\}$, $i = 1, 2, 3$, $B = \{\text{message is spam}\}$. The problem implies that

\[
P(A_1) = 0.7, \quad P(A_2) = 0.2, \quad P(A_3) = 0.1
\]

\[
P(B|A_1) = 0.01, \quad P(B|A_2) = 0.02, \quad P(B|A_3) = 0.05.
\]

Then $P(A_2|B) = \frac{P(B|A_2)P(A_2)}{P(B)} = \frac{0.02 \times 0.2}{0.016} = 0.25$. 

Tree diagram may help!
More examples

Lab tests produce positive and negative results. Assume that for a lab test, 95% of patients with disease obtain positive results and only 2% of patients without disease obtain positive results. Assume that 1% of the population has the disease.

- Pick one person at random. What is the chance that lab test will be positive?
- Pick one person at random. The lab test shows positive result. What is the chance that the person really has the disease?
- What is the probability that, given that the lab test shows a negative result, the person does not have the disease?

Solution: First we translate the problem by math language. Denote \( A = \{ \text{test positive} \}, \ D = \{ \text{have the disease} \}. \) Then

\[
\mathbb{P}(A|D) = 0.95, \quad \mathbb{P}(A|D^c) = 0.02, \quad \mathbb{P}(D) = 0.01.
\]

- \( \mathbb{P}(A) = \mathbb{P}(D)\mathbb{P}(A|D) + \mathbb{P}(D^c)\mathbb{P}(A|D^c) = 0.01 \times 0.95 + 0.99 \times 0.02 = 2.93\% \)
More examples

Lab tests produce positive and negative results. Assume that for a lab test, 95% of patients with disease obtain positive results and only 2% of patients without disease obtain positive results. Assume that 1% of the population has the disease.

- Pick one person at random. What is the chance that lab test will be positive?
- Pick one person at random. The lab test shows positive result. What is the chance that the person really has the disease?
- What is the probability that, given that the lab test shows a negative result, the person does not have the disease?

**Solution:** First we translate the problem by math language. Denote $A = \{\text{test positive}\}$, $D = \{\text{have the disease}\}$. Then

$$\Pr(A|D) = 0.95, \quad \Pr(A|D^c) = 0.02, \quad \Pr(D) = 0.01.$$

- $\Pr(D|A) = \frac{\Pr(D \cap A)}{\Pr(A)} = \frac{\Pr(A|D)\Pr(D)}{\Pr(A)} = \frac{0.95 \times 0.01}{2.93\%} = 32.4\%$
More examples

Lab tests produce positive and negative results. Assume that for a lab test, 95% of patients with disease obtain positive results and only 2% of patients without disease obtain positive results. Assume that 1% of the population has the disease.

- Pick one person at random. What is the chance that lab test will be positive?
- Pick one person at random. The lab test shows positive result. What is the chance that the person really has the disease?
- What is the probability that, given that the lab test shows a negative result, the person does not have the disease?

**Solution:** First we translate the problem by math language. Denote $A = \{\text{test positive}\}$, $D = \{\text{have the disease}\}$. Then

\[ P(A|D) = 0.95, \quad P(A|D^c) = 0.02, \quad P(D) = 0.01. \]

\[ P(D^c|A^c) = \frac{P(D^c \cap A^c)}{P(A^c)} = \frac{P(A^c|D^c)P(D^c)}{P(A^c)} = \frac{[1-P(A|D^c)]P(D^c)}{1-P(A)} = 99.95\% \]
More examples: coffee purchase

The accompanying table gives information on the type of coffee selected by someone purchasing a single cup at a particular airport kiosk.

<table>
<thead>
<tr>
<th></th>
<th>Small</th>
<th>Medium</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular</td>
<td>14%</td>
<td>20%</td>
<td>26%</td>
</tr>
<tr>
<td>Decaf</td>
<td>20%</td>
<td>10%</td>
<td>10%</td>
</tr>
</tbody>
</table>

Consider randomly selecting such a coffee purchaser.

○ What is the probability that the individual purchased a small cup? A cup of decaf coffee?

We can see that \( S = (S \cap R) \cup (S \cap D) \) which are mutually exclusive, thus \( P(S) = P(S \cap R) + P(S \cap D) = 14\% + 20\% = 34\% \).

Similarly, \( D = (D \cap S) \cup (D \cap M) \cup (D \cap L) \), thus \( P(D) = P(D \cap S) + P(D \cap M) + P(D \cap L) = 20\% + 10\% + 10\% = 40\% \).
More examples: coffee purchase

The accompanying table gives information on the type of coffee selected by someone purchasing a single cup at a particular airport kiosk.

<table>
<thead>
<tr>
<th></th>
<th>Small</th>
<th>Medium</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular</td>
<td>14%</td>
<td>20%</td>
<td>26%</td>
</tr>
<tr>
<td>Decaf</td>
<td>20%</td>
<td>10%</td>
<td>10%</td>
</tr>
</tbody>
</table>

Consider randomly selecting such a coffee purchaser.

- If we learn that the selected individual purchased a small cup, what now is the probability that he/she chose decaf coffee?

\[
P(D|S) = \frac{P(D \cap S)}{P(S)} = \frac{20\%}{34\%} = 58.8\%\]
More examples: coffee purchase

The accompanying table gives information on the type of coffee selected by someone purchasing a single cup at a particular airport kiosk.

<table>
<thead>
<tr>
<th></th>
<th>Small</th>
<th>Medium</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular</td>
<td>14%</td>
<td>20%</td>
<td>26%</td>
</tr>
<tr>
<td>Decaf</td>
<td>20%</td>
<td>10%</td>
<td>10%</td>
</tr>
</tbody>
</table>

Consider randomly selecting such a coffee purchaser.

- If we learn that the selected individual purchased decaf, what now is the probability that a small size was selected?

\[
P(S \mid D) = \frac{P(S \cap D)}{P(D)} = \frac{20\%}{40\%} = 50\%\]
More examples: accident rate and seatbelt wearing

A recent Maryland highway safety study found that in 77% of all accidents the driver was wearing a seatbelt. Accident reports indicated that 92% of those drivers escaped serious injury (hospitalization or death), but only 63% of the nonbelted drivers were so fortunate. What's the probability that a driver who was seriously injured wasn't wearing a seatbelt?

Denote $B = \{\text{the driver was wearing a seatbelt}\}$ and $I = \{\text{serious injury or death}\}$. Then

$$\mathbb{P}(B) = 0.77, \quad \mathbb{P}(I|B) = 1 - 0.92 = 0.08, \quad \mathbb{P}(I|B^c) = 1 - 0.63 = 0.37.$$ 

Thus

$$\mathbb{P}(B^c|I) = \frac{\mathbb{P}(I \cap B^c)}{\mathbb{P}(I)} = \frac{\mathbb{P}(I|B^c)\mathbb{P}(B^c)}{\mathbb{P}(I|B^c)\mathbb{P}(B^c) + \mathbb{P}(I|B)\mathbb{P}(B)} = \frac{0.37 \times 0.23}{0.37 \times 0.23 + 0.08 \times 0.77} = 0.580$$
The last example: Monty Hall problem

You're given the choice of three doors: Behind one door is an iPhone 13; behind the others, goats. You pick a door, say No.1, and the host, who knows what's behind the doors, opens another door, say No.3, which has a goat. He then says to you, "Do you want to pick door No. 2?"

**Question:** Would you switch your choice?

![Diagram of three doors with one open showing a goat]

We have to make an assumption about the host: If the iPhone is behind the door we pick, then the host randomly picks one door from the remaining two to open.
Reading list (optional)

- "Probability and Statistics for Engineering and the Sciences" (9th edition):
  - Chapter 2.4
- "OpenIntro statistics" (4th edition, free online, download [here]):
  - Chapter 3.2 (skip Chapter 3.2.6 for now)
Many thanks to
- Chengliang Tang
- Joyce Robbins
- Yang Feng
- Owen Ward
- Wenda Zhou
- And all my teachers in the past 25 years