Lecture 19 – Combinatorics

DSC 40A, Fall 2022 @ UC San Diego

Dr. Truong Son Hy, with help from **many others**

Agenda

- Conditional probability (continued).
- Sequences, permutations, and combinations.
- Practice problems.

Example: rolling a die

- Suppose we roll the dice n times. What is the probability that only the faces 2, 4, and 5 appear?

$$\left(\frac{1}{2}\right)^n$$

- Suppose we roll the dice twice. What is the probability that the two rolls have different faces?

$$\frac{5}{6}$$

Conditional probability

- The probability of an event may **change** if we have additional information about outcomes.

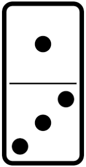
- Starting with the multiplication rule, $P(A \cap B) = P(A)P(B|A)$, we have that

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

assuming that $P(A) > 0$.

Example: dominoes (source: 538)

In a set of dominoes, each tile has two sides with a number of dots on each side: zero, one, two, three, four, five, or six. There are 28 total tiles, with each number of dots appearing alongside each other number (including itself) on a single tile.



Example: dominoes

Question 1: What is the probability of drawing a “double” from a set of dominoes — that is, a tile with the same number on both sides?

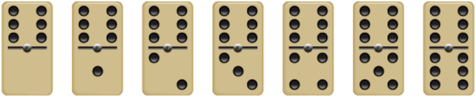
Example: dominoes

Question 1: What is the probability of drawing a “double” from a set of dominoes — that is, a tile with the same number on both sides?

$$\frac{7}{28} = \frac{1}{4}$$

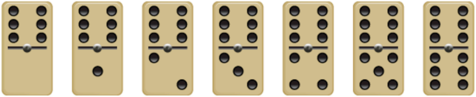
Example: dominoes

Question 2: Now your friend picks a random tile from the set and tells you that at least one of the sides is a 6. What is the probability that your friend's tile is a double, with 6 on both sides?



Example: dominoes

Question 2: Now your friend picks a random tile from the set and tells you that at least one of the sides is a 6. What is the probability that your friend's tile is a double, with 6 on both sides?



$$\frac{1}{7}$$

Simpson's Paradox (source: nih.gov)

	Treatment A	Treatment B
Small kidney stones	81 successes / 87 (93%)	234 successes / 270 (87%)
Large kidney stones	192 successes / 263 (73%)	55 successes / 80 (69%)
Combined	273 successes / 350 (78%)	289 successes / 350 (83%)

Discussion Question

Which treatment is better?

A) Treatment A for all cases.
B) Treatment B for all cases.
C) Treatment A for small stones and B for large stones.
D) Treatment A for large stones and B for small stones.

Simpson's Paradox (source: nih.gov)

Let A be a random variable taking value True if treatment A is effective, or False otherwise. Let X be a random variable taking values, small or large, denoting the size of the kidney stone.

By the **Law of Total Probability**, We have:

$$P(A = \text{True}) = P(A = \text{True}|X = \text{small}) \cdot P(X = \text{small}) +$$

$$P(A = \text{True}|X = \text{large}) \cdot P(X = \text{large})$$

Theat is equal to:

$$P(A = \text{True}) = \frac{81}{87} \cdot \frac{87}{350} + \frac{192}{263} \cdot \frac{263}{350} = \frac{273}{350} = 78\%$$

Simpson's Paradox (source: nih.gov)

Let B be a random variable taking value True if treatment B is effective, or False otherwise. Let Y be a random variable taking values, small or large, denoting the size of the kidney stone. We use Y not X because for each experiment for each treatment, 350 different people.

By the **Law of Total Probability**, We have:

$$P(B = \text{True}) = P(B = \text{True}|Y = \text{small}) \cdot P(Y = \text{small}) +$$

$$P(B = \text{True}|Y = \text{large}) \cdot P(Y = \text{large})$$

Theat is equal to:

$$P(B = \text{True}) = \frac{234}{270} \cdot \frac{270}{350} + \frac{55}{80} \cdot \frac{80}{350} = \frac{289}{350} = 83\%$$

Simpson's Paradox (source: nih.gov)

It is called a **paradox** because:

$$P(B = \text{True}|Y = \text{small}) < P(A = \text{True}|X = \text{small})$$

$$P(B = \text{True}|Y = \text{large}) < P(A = \text{True}|X = \text{large})$$

But

$$P(B = \text{True}) > P(A = \text{True}).$$

The problem lies in the fact that distributions of X and Y are approximations (based on sampling) of the actual distribution of patients with small or large kidney stones.

How can we fix this?

We need to make a better approximation of the distribution of patients with small or large stones.

Simpson's Paradox (source: nih.gov)

There are totally 700 = 350 + 350 patients in which:

- 87 + 270 = 357 have small stones: 357/700 = 51%, denoted by $P(\text{small})$

- 263 + 80 = 343 have large stones: 343/700 = 49%, denoted by $P(\text{large})$

Simpson's Paradox (source: nih.gov)

There are totally 700 = 350 + 350 patients in which:

- 87 + 270 = 357 have small stones: 357/700 = 51%, denoted by $P(\text{small})$

- 263 + 80 = 343 have large stones: 343/700 = 49%, denoted by $P(\text{large})$

By the **Law of Total Probability**, we have the actual effectiveness of A is:

$$P(A = \text{True}) \approx P(A = \text{True}|\text{small}) \cdot P(\text{small}) + P(A = \text{True}|\text{large}) \cdot P(\text{large})$$

That equals to:

$$P(A = \text{True}) \approx 93\% \cdot 51\% + 73\% \cdot 49\% = 83.2\%$$

Simpson's Paradox (source: nih.gov)

By the **Law of Total Probability**, we have the actual effectiveness of B is:

$$P(B = \text{True}) \approx P(B = \text{True}|\text{small}) \cdot P(\text{small}) + P(B = \text{True}|\text{large}) \cdot P(\text{large})$$

That equals to:

$$P(B = \text{True}) \approx 87\% \cdot 51\% + 69\% \cdot 49\% = 81.24\%$$

Now, we can conclude that treat A is better in general.

Simpson's Paradox (source: nih.gov)

	Treatment A	Treatment B
Small kidney stones	81 successes / 87 (93%)	234 successes / 270 (87%)
Large kidney stones	192 successes / 263 (73%)	55 successes / 80 (69%)
Combined	273 successes / 350 (78%)	289 successes / 350 (83%)

Simpson's Paradox occurs when an association between two variables exists when the data is divided into subgroups, but reverses or disappears when the groups are combined.

- See more in DSC 80.

Sequences, permutations, and combinations

Motivation

- Many problems in probability involve counting.
  - Suppose I flip a fair coin 100 times. What's the probability I see 34 heads?
  - Suppose I draw 3 cards from a 52 card deck. What's the probability they all are all from the same suit?
- In order to solve such problems, we first need to learn how to count.
- The area of math that deals with counting is called **combinatorics**.

Selecting elements (i.e. sampling)

- Many experiments involve choosing k elements randomly from a group of n possible elements. This group is called a **population**.
  - If drawing cards from a deck, the population is the deck of all cards.
  - If selecting people from DSC 40A, the population is everyone in DSC 40A.
- Two decisions:
  - Do we select elements with or without **replacement**?
  - Does the **order** in which things are selected matter?

Sequences

- A **sequence** of length k is obtained by selecting k elements from a group of n possible elements **with replacement**, such that **order matters**.

- **Example:** Draw a card (from a standard 52-card deck), put it back in the deck, and repeat 4 times.

Sequences

- A **sequence** of length k is obtained by selecting k elements from a group of n possible elements **with replacement**, such that **order matters**.

- **Example:** Draw a card (from a standard 52-card deck), put it back in the deck, and repeat 4 times.

$$52^4$$

- **Example:** A UCSD PID starts with "A" then has 8 digits. How many UCSD PIDs are possible?

Sequences

- A **sequence** of length k is obtained by selecting k elements from a group of n possible elements **with replacement**, such that **order matters**.

- **Example:** Draw a card (from a standard 52-card deck), put it back in the deck, and repeat 4 times.

$$52^4$$

- **Example:** A UCSD PID starts with “A” then has 8 digits. How many UCSD PIDs are possible?

$$10^8 = 100,000,000$$

Sequences

In general, the number of ways to select k elements from a group of n possible elements such that **repetition is allowed** and **order matters** is n^k.

(Note: We mentioned this fact in the first lecture on clustering!)

Permutations

- A **permutation** is obtained by selecting k elements from a group of n possible elements **without replacement**, such that **order matters**.
- **Example:** How many ways are there to select a president, vice president, and secretary from a group of 8 people?

Permutations

- A **permutation** is obtained by selecting k elements from a group of n possible elements **without replacement**, such that **order matters**.
- **Example:** How many ways are there to select a president, vice president, and secretary from a group of 8 people?

Permutations

- A **permutation** is obtained by selecting k elements from a group of n possible elements **without replacement**, such that **order matters**.
- **Example:** How many ways are there to select a president, vice president, and secretary from a group of 8 people?
  President: 8 choices
  Vice: 7 choices
  Secretary: 6 choices
  Total: $8 \cdot 7 \cdot 6 = 336$

Permutations

- In general, the number of ways to select k elements from a group of n possible elements such that **repetition is not allowed** and **order matters** is

$$P(n,k) = (n)(n-1)\ldots(n-k+1)$$

- To simplify: recall that the definition of $n!$ is

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

- Given this, we can write

$$P(n,k) = \frac{n!}{(n-k)!}$$

Discussion Question

UCSD has 7 colleges. How many ways can I rank my top 3 choices?

A) 21
B) 210
C) 343
D) 2187
E) None of the above

Discussion Question

UCSD has 7 colleges. How many ways can I rank my top 3 choices?

A) 21
B) 210
C) 343
D) 2187
E) None of the above

Answer: B) $7 \cdot 6 \cdot 5 = 210$

Special case of permutations

- Suppose we have n people. The total number of ways I can rearrange these n people in a line is

$$n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1 = n!$$

- This is consistent with the formula

$$P(n,n) = \frac{n!}{(n-n)!} = \frac{n!}{0!} = \frac{n!}{1} = n!$$

Combinations

- A **combination** is a set of k items selected from a group of n possible elements **without replacement**, such that **order does not matter**.

- **Example:** There are 24 ice cream flavors. How many ways can you pick two flavors?

Combinations

- A **combination** is a set of k items selected from a group of n possible elements **without replacement**, such that **order does not matter**.

- **Example:** There are 24 ice cream flavors. How many ways can you pick two flavors?

$$\frac{24 \cdot 23}{2} = 276$$

From permutations to combinations

- There is a close connection between:
  - the number of permutations of k elements selected from a group of n, and
  - the number of combinations of k elements selected from a group of n

$$\text{\# combinations} = \frac{\text{\# permutations}}{\text{\# orderings of } k \text{ items}}$$

- Since # permutations = $\frac{n!}{(n-k)!}$ and # orderings of k items = $k!$, we have

$$C(n,k) = \binom{n}{k} = \frac{n!}{(n-k)!k!}$$

Combinations

In general, the number of ways to select k elements from a group of n elements such that **repetition is not allowed** and **order does not matter** is

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

The symbol $\binom{n}{k}$ is pronounced "n choose k", and is also known as the **binomial coefficient**.

Example: committees

- How many ways are there to select a president, vice president, and secretary from a group of 8 people?

Example: committees

- How many ways are there to select a president, vice president, and secretary from a group of 8 people?

$$8 \cdot 7 \cdot 6 = 336$$

- How many ways are there to select a committee of 3 people from a group of 8 people?

Example: committees

- How many ways are there to select a president, vice president, and secretary from a group of 8 people?

$$8 \cdot 7 \cdot 6 = 336$$

- How many ways are there to select a committee of 3 people from a group of 8 people?

$$\binom{8}{3} = \frac{8 \cdot 7 \cdot 6}{6} = 56$$

- If you're ever confused about the difference between permutations and combinations, **come back to this example**.

Discussion Question

A domino consists of two faces, each with anywhere between 0 and 6 dots. A set of dominoes consists of every possible combination of dots on each face.

How many dominoes are in the set of dominoes?

A) $\binom{7}{2}$

B) $\binom{7}{1} + \binom{7}{2}$

C) $P(7, 2)$

D) $\frac{P(7,2)}{P(7,1)} 7!$

Discussion Question

A domino consists of two faces, each with anywhere between 0 and 6 dots. A set of dominoes consists of every possible combination of dots on each face.

How many dominoes are in the set of dominoes?

A) $\binom{7}{2}$

B) $\binom{7}{1} + \binom{7}{2}$

C) $P(7, 2)$

D) $\frac{P(7,2)}{P(7,1)} 7!$

Answer: A

Summary

Summary

- A **sequence** is obtained by selecting k elements from a group of n possible elements with replacement, such that order matters.
  - Number of sequences: n^k.

- A **permutation** is obtained by selecting k elements from a group of n possible elements without replacement, such that order matters.
  - Number of permutations: $P(n, k) = \frac{n!}{(n-k)!}$.

- A **combination** is obtained by selecting k elements from a group of n possible elements without replacement, such that order does not matter.
  - Number of combinations: $\binom{n}{k} = \frac{n!}{(n-k)!k!}$.