DSC 40A

Theoretical Foundations of Data Science I

In This Video

- Many probability questions can be solved by counting, or combinatorics.
- We’ll learn how to count sequences and sets.

Sequences vs. Sets

Sequences lists/tuples	Sets collection of elements
Order matters	Order does not matter
Repetitions allowed	No repetitions allowed
Elements listed in order	Elements listed in no particular order within curly braces
Ex: 2, 4, 5 ≠ 4, 2, 5	Ex: {2, 4, 5} = {4, 2, 5}
Ex: 2, 2, 2 ≠ 2, 2	Ex: $\{2, \cancel{2, 2}\} = \{2, \cancel{2}\} = \underline{\{2\}}$
Ex: 1, 3, 4 = 1, 3, 4	Ex: {1, 3, 4} = {1, 3, 4}

Sequences

Sequences
Order matters ←
Repetitions allowed
Elements listed in order
Ex: 2, 4, 5 ≠ 4, 2, 5
Ex: 2, 2, 2 ≠ 2, 2
Ex: 1, 3, 4 = 1, 3, 4

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

A. 8^{10}

[✓ B. 10^8]

C. 8!

D. 10*9*8*7*6*5*4*3

A _ _ _ _ _ _ _ _

↑ ↑ ↑ . . . . .

10 digits

Sequences

Sequences
Order matters
Repetitions allowed
Elements listed in order
Ex: 2, 4, 5 ≠ 4, 2, 5
Ex: 2, 2, 2 ≠ 2, 2
Ex: 1, 3, 4 = 1, 3, 4

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

1, . . . , 8

for president : 8 options

for VP : 7 options

for secretary : 6 options

$8 * 7 * 6$

$$\underset{p}{\underline{5}}\ \underset{VP}{\underline{3}}\ \underset{sec}{\underline{2}}$$

$$\neq \underset{p}{\underline{3}}\ \underset{VP}{\underline{2}}\ \underset{sec}{\underline{5}}$$

Sequences

Sequences
Order matters
Repetitions allowed ←
Elements listed in order
Ex: 2, 4, 5 ≠ 4, 2, 5
Ex: 2, 2, 2 ≠ 2, 2
Ex: 1, 3, 4 = 1, 3, 4

How many ways to select a president, vice president, and secretary from a group of 8 people? $8 * 7 * 6$

$n=8$ ($n = \#$ elements to choose from)

$k=3$ ($k = \#$ distinct elements to choose)

$$P(8,3) = 8*7*6$$

$$P(n,k) = \underbrace{(n)*(n-1)* \dots * \overset{n-k+1}{(n-(k-1))}}_{k \text{ terms}}$$

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

Sequences where repetitions are not allowed are permutations.

Sets

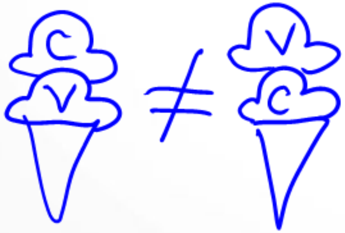
There are 24 ice cream flavors. How many ways can you pick 2 different flavors?

A. 24
C. 24*24
B. 24*23
[✓ D. 12*23]

Sets
Order does not matter
No repetitions allowed
Elements listed in no particular order within curly braces
Ex: {2, 4, 5} = {4, 2, 5}
Ex: {2, 2, 2} = {2, 2} = {2}
Ex: {1, 3, 4} = {1, 3, 4}

first: count sequences

counts



$$\frac{24*23}{2}$$

$$\#\text{sets} = \frac{\#\text{sequences}}{\#\text{orderings}} \Rightarrow \#\text{sequences} = \#\text{sets} * \#\text{orderings}$$

Sets

$k! = P(k,k)$

How many ways to select a committee of 3 from a group of 8?

$$\#\text{sets} = \frac{\#\text{sequences}}{\#\text{orderings}} = \frac{8*7*6}{3*2*1}$$

$$= 8*7$$

$$= 56$$

$n=8$ (#elements to choose from)

$k=3$ (# elements to select)

$$C(8,3) = \frac{8*7*6}{3*2*1} = \frac{8!/5!}{3!} = \frac{8!}{5!\,3!}$$

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

Sets
Order does not matter
No repetitions allowed
Elements listed in no particular order within curly braces
Ex: {2, 4, 5} = {4, 2, 5}
Ex: {2, 2, 2} = {2, 2} = {2}
Ex: {1, 3, 4} = {1, 3, 4}

Permutations vs. Combinations

Permutations	Combinations
Order matters	Order does not matter
No repetitions allowed	No repetitions allowed
Counts the number of **sequences of k distinct elements** chosen from n possible elements	Counts the number of **sets of size k** chosen from n possible elements
$\underline{P}(n,k) = (\underline{n})(\underline{n-1})\ldots(n-k+1) = \frac{n!}{(n-k)!}$	"n choose k" $\underline{C}(n,k) = \boxed{\binom{n}{k}} = \frac{n!}{k!(n-k)!}$
☆ How many ways to select a president, vice president, and secretary from a group of 8 people? $\boxed{P(8,3)}$	How many ways to select a committee of 3 from a group of 8? $\boxed{C(8,3)} = \binom{8}{3}$

Sampling Without Replacement

Example 6. There are 20 students in a class. A computer program selects a random sample of students by drawing 5 students at random **without replacement.** What is the chance that a 17 particular student is among the 5 selected students?

last time: sample space was sequences

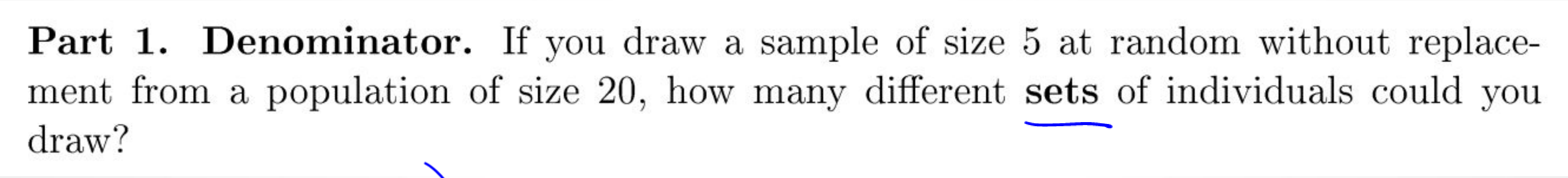
today: sample space will be sets

S = sets of 5 students, chosen from 20

$$\text{prob}\left(\begin{matrix}\text{student}\\ \text{17 in}\\ \text{sample}\end{matrix}\right) = \frac{\#\text{ sets include person 17}}{\#\text{ sets of 5, chosen from 20}}$$

from Theory Meets Data by Ani Adhikari, Chapter 4

Sampling Without Replacement



Part 1. Denominator. If you draw a sample of size 5 at random without replacement from a population of size 20, how many different **sets** of individuals could you draw?

$C(20,5)$

from [Theory Meets Data] by Ani Adhikari, Chapter 4

Sampling Without Replacement

Part 2. Numerator. If you draw a sample of size 5 at random without replacement from a population of size 20, how many different **sets** of individuals include a particular person?

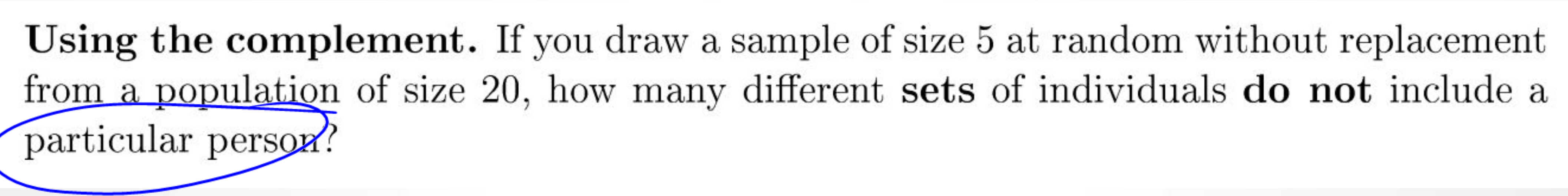
# of sets of 5, chosen from 20, including person 17

Key: this is the same as # of sets of 4, chosen from the other 19

$= C(19,4)$

from Theory Meets Data by Ani Adhikari, Chapter 4

Sampling Without Replacement



Using the complement. If you draw a sample of size 5 at random without replacement from a population of size 20, how many different **sets** of individuals **do not** include a particular person?

17

sets of size 5, chosen from 20, not including 17

$= C(19,5)$

from Theory Meets Data by Ani Adhikari, Chapter 4

Sampling Without Replacement

Example 6. There are 20 students in a class. A computer program selects a random sample of students by drawing 5 students at random **without replacement**. What is the chance that a particular student is among the 5 selected students?

$$\text{prob} = \frac{\#\text{ sets with } 17}{\text{total }\#\text{ sets}} = \frac{\text{total }\#\text{ sets} - \#\text{ sets without } 17}{\text{total }\#\text{ sets}}$$

$$= \frac{C(19,4)}{C(20,5)} \checkmark \qquad = \frac{C(20,5) - C(19,5)}{C(20,5)}$$

$$= \frac{\cancel{\dfrac{19!}{(4! * \cancel{15!})}}}{\cancel{\dfrac{20!}{(5! * \cancel{15!})}}} = \frac{19!}{4!} \times \frac{5!}{20!} = \frac{5}{20} = \boxed{1/4}$$

from Theory Meets Data by Ani Adhikari, Chapter 4

Summary

- Sequences vs. sets
- When elements are distinct: permutations vs. combinations

$$P(n,k) = (n)(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$$

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

- **Next time:** more examples