DSC 40A

Theoretical Foundations of Data Science I

In This Video

- What is the probability of a random sample having a certain property?

Sampling

Sampling with replacement:

1. Draw one element *uniformly at random* from list.
2. Return the element to the list.
3. Repeat

Sampling without replacement:

same, except skip step 2

What does *uniformly at random* mean? each element equally likely

Sampling

Sampling with or without replacement:

- All samples are equally likely.
- Uniform distribution!

$$P(\text{sample having a certain property}) =$$

Sampling

Sampling with or without replacement:

- All samples are equally likely.
- Uniform distribution!

$$P(\text{sample having a certain property}) = \frac{\#\text{ samples having property}}{\#\text{ possible samples}}$$

Practice Problems

1, 2, ..., 20

Example 5. There are 20 students in a class. A computer program selects a random sample of students by drawing 5 students at random **with replacement**. What is the chance that a particular student (17) is among the 5 selected students?

S = sequences of length 5 with entries in $\{1, 2, \ldots, 20\}$

ex.) 3, 12, 4, 15, 20

3, 3, 3, 8, 4

$$\frac{\#\text{ seq in } S \text{ that include } 17}{\#\text{ seq in } S}$$

Practice Problems

Part 1. Denominator. If you draw a sample of size 5 at random with replacement from a population of size 20, how many different sequences of individuals could you draw?

seq in S

ex.) 5, 10, 5, 3, 2 $\Rightarrow 20^5$

any # in $\{1, 2, \ldots, 20\}$ 20 options

$\Rightarrow$ 20 options

from Theory Meets Data by Ani Adhikari, Chapter 4

Practice Problems

Part 2. Numerator. If you draw a sample of size 5 at random with replacement from a population of size 20, how many different sequences of individuals include a particular person?

seq in S that include 17

ex.) 6, 4, 17, 3, 17

↑

20

from Theory Meets Data by Ani Adhikari, Chapter 4

Practice Problems

Using the complement. If you draw a sample of size 5 at random with replacement from a population of size 20, how many different sequences of individuals **do not** include a particular person?

seq in 5 that don't include 17

ex.) 16, 12, 14, 16, 11

19 options ⟹ 19^5

{1, ..., 20} but not 17

⇒ 19 options

from Theory Meets Data by Ani Adhikari, Chapter 4

Practice Problems

Example 5. There are 20 students in a class. A computer program selects a random sample of students by drawing 5 students at random **with replacement**. What is the chance that a particular student is among the 5 selected students?

$$\text{prob}(\text{include } 17) = \frac{\#\text{ seq in } S \text{ with } 17}{\#\text{ seq in } S} = \frac{\#\text{seq in } S - \#\text{seq in } S \text{ without } 17}{\#\text{ seq in } S}$$

$$= \frac{20^5 - 19^5}{20^5} = 1 - \frac{19^5}{20^5} = 1 - \left(\frac{19}{20}\right)^5$$

$$\text{prob}(\text{include } 17) = 1 - \text{prob}(\text{not include } 17)$$

$$= 1 - \left(\frac{19}{20}\right)^5 \approx 0.226$$

from Theory Meets Data by Ani Adhikari, Chapter 4

Practice Problems

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?

from Theory Meets Data by Ani Adhikari, Chapter 4

Practice Problems

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?

ex.) 5, 3, 2, 12, 14 must be different

Which probability will be higher?

A. Probability of including a particular student when sampling with replacement.

B. Probability of including a particular student when sampling without replacement.

C. Both probabilities are the same.

S = sequences of length 5 with entries $\{1, \cdots, 20\}$ with no repeats

from Theory Meets Data by Ani Adhikari, Chapter 4

Practice Problems

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

seq in S

ex.) 6, 7, 4, 12, 3

20 options, 19 options, 18, 17, 16

$\Rightarrow 20 \cdot 19 \cdot 18 \cdot 17 \cdot 16$

$= \frac{20!}{15!}$

from Theory Meets Data by Ani Adhikari, Chapter 4

Practice Problems

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

17

ex) 17, 6, 5, 14, 20

3, 12, 17, 2, 9

17 _ _ _ _ → 19·18·17·16

_ 17 _ _ _ → 19·18·17·16

_ _ 17 _ _

_ _ _ 17 _

_ _ _ _ 17

5·19·18·17·16

Practice Problems

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

ex.) 8, 12, 14, 16, 19

19 options, 18 options, 17, 16, 15

$\Rightarrow 19 \cdot 18 \cdot 17 \cdot 16 \cdot 15$

$= \frac{19!}{14!}$

from Theory Meets Data by Ani Adhikari, Chapter 4

Practice Problems

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?

$$\text{prob}(\text{include } 17) = \frac{\#\text{ include } 17}{\text{total } \#} = \frac{\text{total } \# - \#\text{ don't include } 17}{\text{total } \#}$$

$$= \frac{5 \cdot 19 \cdot 18 \cdot 17 \cdot 16}{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16} = \frac{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16 - 19 \cdot 18 \cdot 17 \cdot 16 \cdot 15}{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16}$$

$$= \boxed{\frac{1}{4}} = \frac{19 \cdot 18 \cdot 17 \cdot 16 (20 - 15)}{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16}$$

from Theory Meets Data by Ani Adhikari, Chapter 4

Summary

- When we sample uniformly, whether with or without replacement, each possible sample is equally likely.
- Probability questions become counting questions:

$$P(\text{sample having a certain property}) = \frac{\#\text{ samples having property}}{\#\text{ possible samples}}$$

- **Next time:** combinatorics, or counting principles