

DSC 40A

Theoretical Foundations of Data Science I

Random Sampling

Announcements

- Groupwork due tonight
- Homework 5 due Friday
- Upcoming homework schedule:
Homework 6 released Monday 11/18 and due 11/25

Agenda

- Conditional probability continued
- Sampling with and without replacement

Question

Answer at q.dsc40a.com

Remember, you can always ask questions at
q.dsc40a.com!

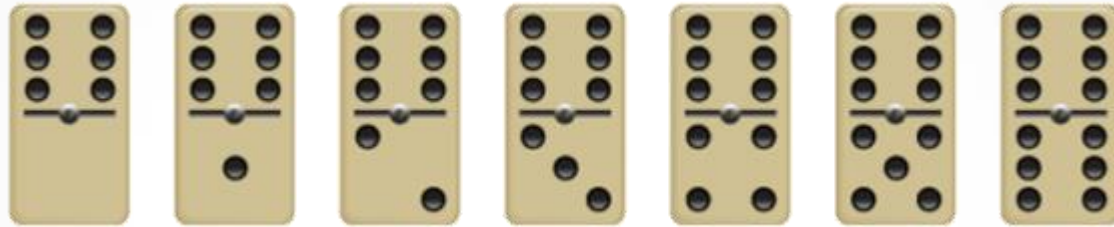
If the direct link doesn't work, click the "Lecture Questions" link in the top right corner of dsc40a.com.

Conditional probability continued



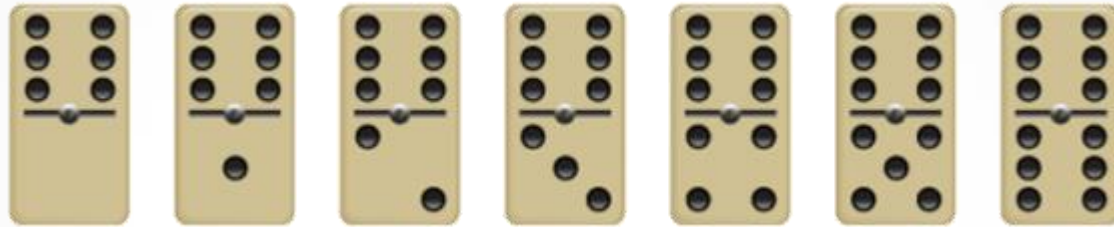
Dominoes

Question 3: Now you pick a random tile from the set and uncover only one side, revealing that it has 6 dots. What is the probability that this tile is a double, with 6 on both sides?



Dominoes

Question 3: Now you pick a random tile from the set and uncover only one side, revealing that it has 6 dots. What is the probability that this tile is a double, with 6 on both sides?



Try it out in [code!](#)

Conditional probabilities: Simpson's Paradox

	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%)

Which treatment is better?

- A. Treatment A for all cases.
- B. Treatment B for all cases.

- C. A for small and B for large.
- D. A for large and B for small.

Conditional probabilities: Simpson's Paradox

	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

"When the less effective treatment is applied more frequently to easier cases, it can appear to be a more effective treatment."

Random Sampling

The background features a series of overlapping, semi-transparent green triangles and polygons of various shades, ranging from light lime green to dark forest green. These shapes are primarily located on the right side of the slide, with some extending towards the center. The overall effect is a modern, abstract geometric design.

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:

What does *uniformly at random* mean?

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

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?

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?

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?

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?

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?

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?

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.

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?

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?

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?

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?

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