

**DSC 40A**

*Theoretical Foundations of Data Science I*

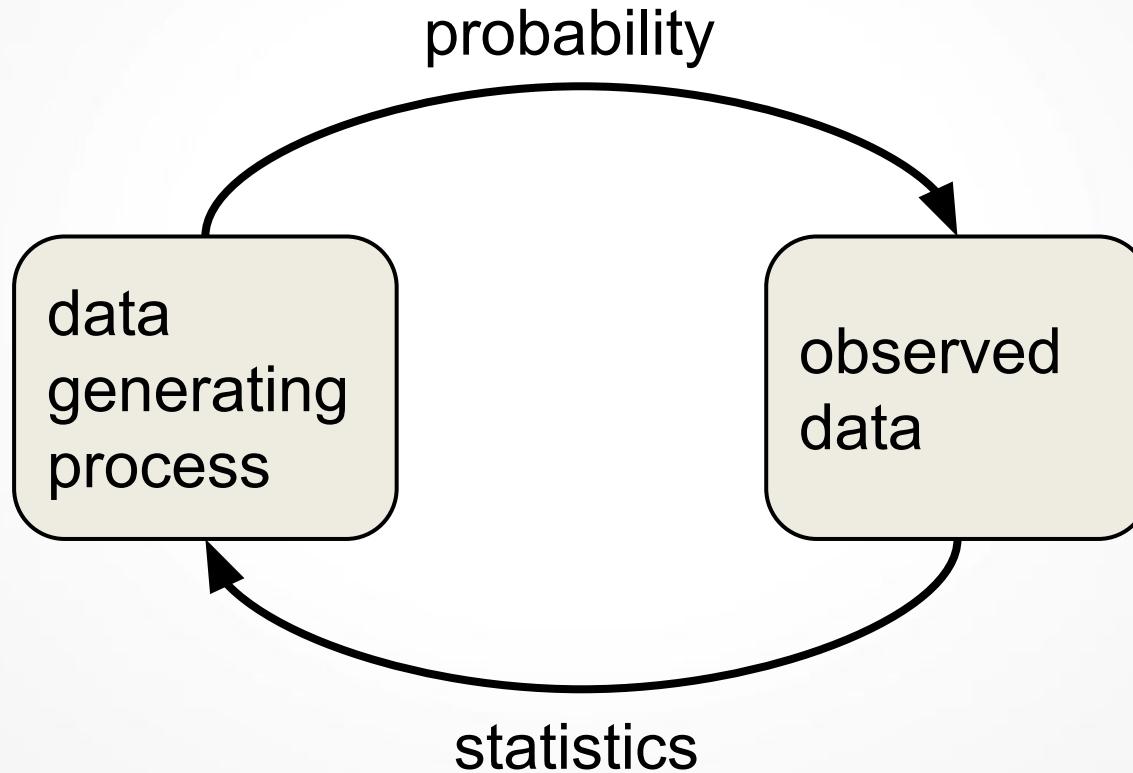
# Predicting from Samples

- So far in this class, we have made predictions based on a data set, or sample.
- For our predictions to be any good, we need to know that other samples generated in the same way will be similar.

## In This Video

- We'll study the basic definitions and rules of discrete probability.

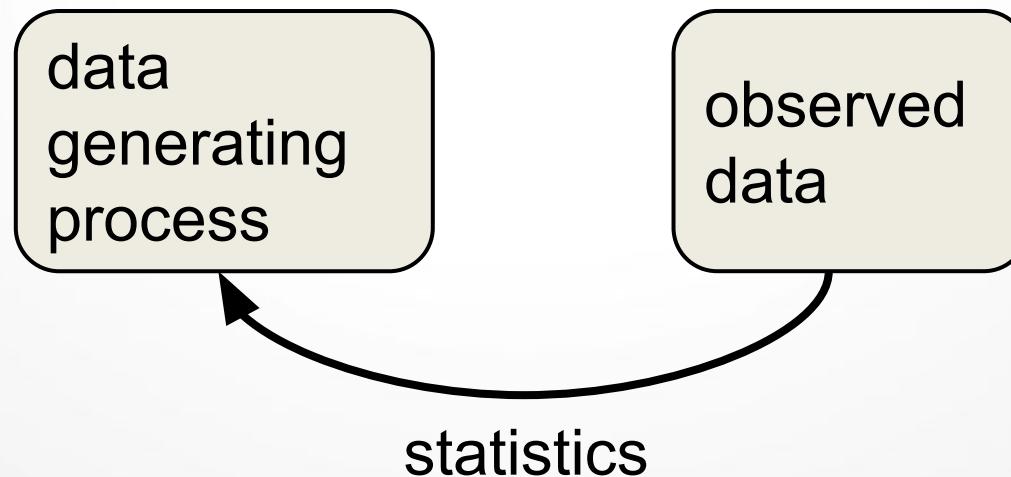
# Probability and Statistics



# Statistical Inference

Given observed data, we want to know how it was generated or where it came from. Maybe we want to

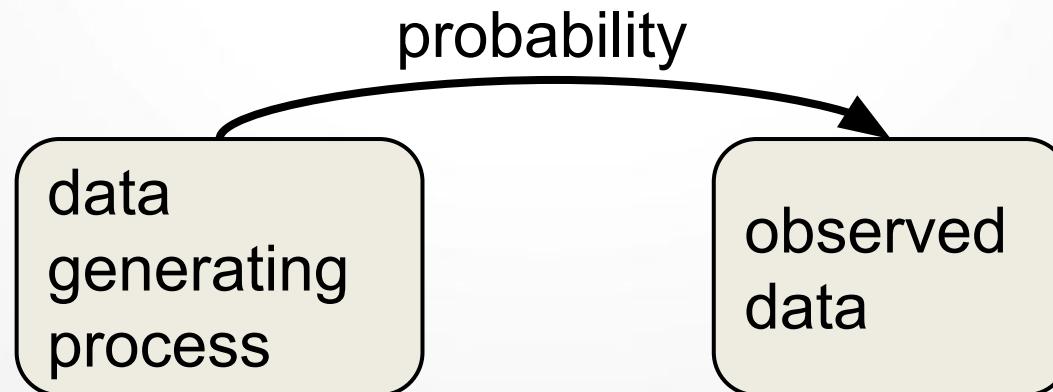
- predict other data generated from the same source
- know how different our sample could have been
- draw conclusions about whole population and not just observed sample - generalize



# Probability

Given a certain model for data generation, what kind of data do you expect the model to produce? How similar is it to the data you have? Probability is the tool to answer these questions.

- expected value versus sample mean
- variance versus sample variance
- likelihood of producing exact observed data



# Probability

**Sample space,  $S$ :** (finite or countable) set of possible outcomes.

**Probability distribution,  $p$ :** assignment of probabilities to outcomes in  $S$  so that

# Probability

**Sample space, S:** (finite or countable) set of possible outcomes.

**Probability distribution, p:** assignment of probabilities to outcomes in S so that

- $0 \leq p(s) \leq 1$  for each  $s$  in  $S$
- Sum of probabilities is 1,  $\sum_{s \in S} p(s) = 1$

Compare flipping a **fair coin** and **biased coin**:

- A. Different sample spaces, different probability distributions.
- B. Different sample spaces, same probability distributions.
- C. Same sample spaces, different probability distributions.
- D. Same sample spaces, same probability distributions.

# Probability

**Sample space,  $S$ :** (finite or countable) set of possible outcomes.

**Probability distribution,  $p$ :** assignment of probabilities to outcomes  $s$  in  $S$  so that

- $0 \leq p(s) \leq 1$  for each  $s$  in  $S$
- Sum of probabilities is 1,  $\sum_{s \in S} p(s) = 1$

**Event,  $E$ :**

# Uniform distribution

For sample space  $S$  with  $n$  elements, **uniform distribution** assigns the probability  $1/n$  to each element of  $S$ .

- flipping fair coin 3 times in a row
- rolling a die

When flipping a fair coin successively three times:

- A. The sample space is {H, T}
- B. The event {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT} has probability less than 1.
- C. The uniform distribution assigns probability 1/8 to each outcome.
- D. None of the above.

# Uniform distribution

For sample space  $S$  with  $n$  elements, **uniform distribution** assigns the probability  $1/n$  to each element of  $S$ .

- flipping fair coin 3 times in a row
- rolling a die

For uniform distribution, the probability of an event  $E$  is:

# Addition Rule

If A and B are mutually exclusive events (cannot happen simultaneously), then

## Addition Rule

If A and B are mutually exclusive events (cannot happen simultaneously), then

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$$

In general

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

# Multiplication Rule

# Multiplication Rule

$P(A \text{ and } B) = P(A \cap B)$   
=  $P(A)*P(B \text{ given that } A \text{ has happened})$   
=  $P(A)*P(B|A)$

# Complement Rule

# Complement Rule

$$P(\bar{A}) = 1 - P(A)$$

# Practice Problems

**Example 1. Rolling A Die.** A fair 6-sided die has numbers from 1 to 6. Each time it is rolled, the outcome will be a number from 1 to 6. The probability of getting any of the six numbers is the same, which is  $1/6$ . No roll affects the outcome of any other roll.

- (i) Suppose the die is rolled once. What is the probability of rolling a 1 and a 2?
- (ii) If the die is rolled once, what is the probability of rolling a 1 or a 2?
- (iii) If the die is rolled twice, what is the probability of rolling a 1 on the first roll and a 2 on the second roll?

# Practice Problems

**Example 2.** A die is rolled 3 times. What is the probability that the face 1 never appears in any of the rolls?

# Practice Problems

**Example 3.** A die is rolled  $n$  times. What is the chance that only faces 2, 4 or 6 appear?

# Practice Problems

**Example 4.** A die is rolled two times. What is the probability that the two rolls had different faces?

# Summary

- We saw the basic definitions and rules in probability:
  - addition rule
  - multiplication rule
  - complement rule
- **Next time:** We'll learn about conditional probability, the probability of one event given that another has occurred.