DSC 40A

Theoretical Foundations of Data Science I

Foundations of Probability

Announcements

- Homework 5 will be released on Friday.
- No in-person discussion section on 11/11 - the groupwork will instead be asynchronous and due on Wed. 11/13 at 11:59PM.

* Midterm grades/solution released tomorrow

Agenda

- Overview: Probability and statistics.
- Complement, addition, and multiplication rules.

Note: There are no more DSC 40A-specific readings, but we've posted many probability resources on the resources tab of the course website. These will come in handy! Specific resources you should look at:

- The DSC 40A probability roadmap, written by Janine Tiefenbruck.
- The textbook Theory Meets Data, which explains many of the same ideas and contains more practice problems.

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.

Course Overview

Part 1: Learning from Data (Weeks 1 through 5)

- Summary statistics and loss functions; empirical risk minimization.
- Linear regression (including multiple variables); linear algebra.

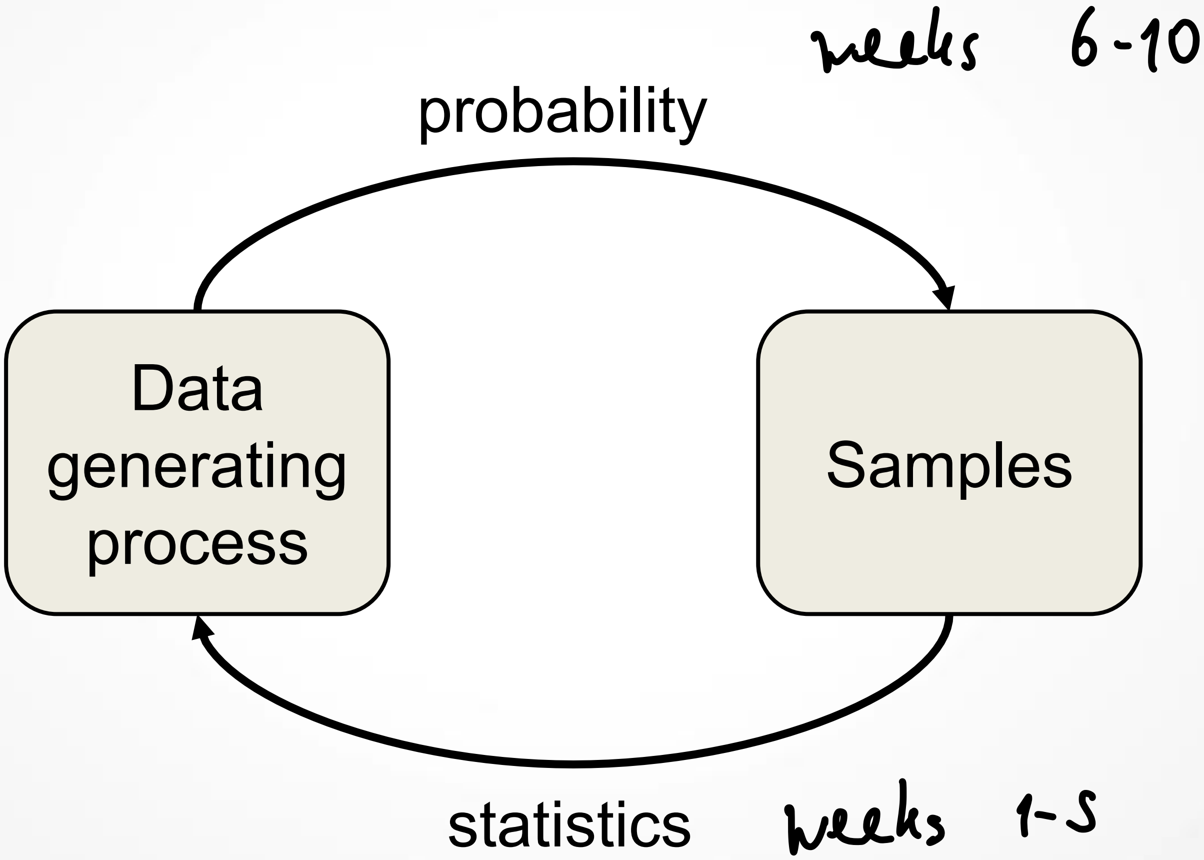
Part 2: Probability (Weeks 6 through 10)

- Set theory and combinatorics; probability fundamentals.
- Conditional probability and independence.
- The Naïve Bayes classifier.

Predicting from Samples

- So far in this class, we have made predictions based on a data set, or sample.
- This dataset can be thought of as a sample of some population.
- For a hypothesis function to be useful in the future, the sample that was used to create the hypothesis function needs to look similar to samples that we'll see in the future.

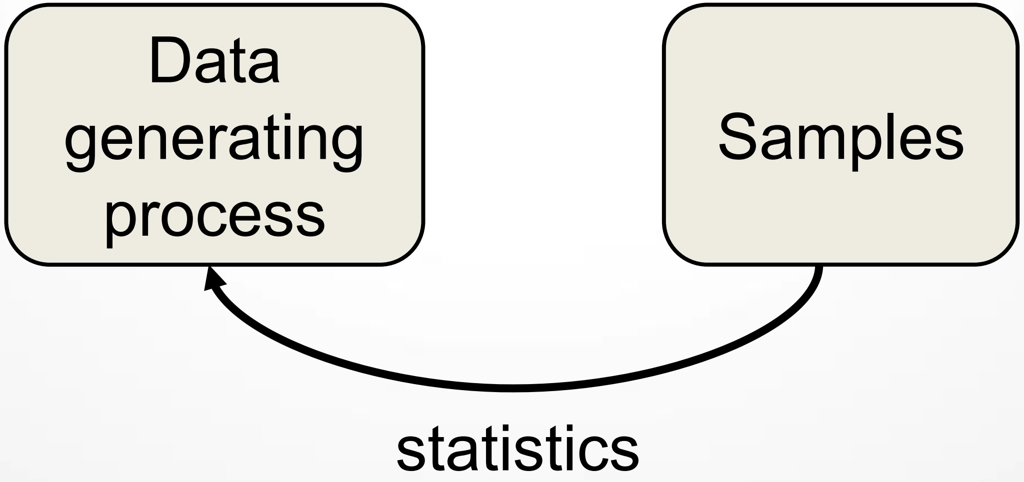
Probability and Statistics



Statistical Inference

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

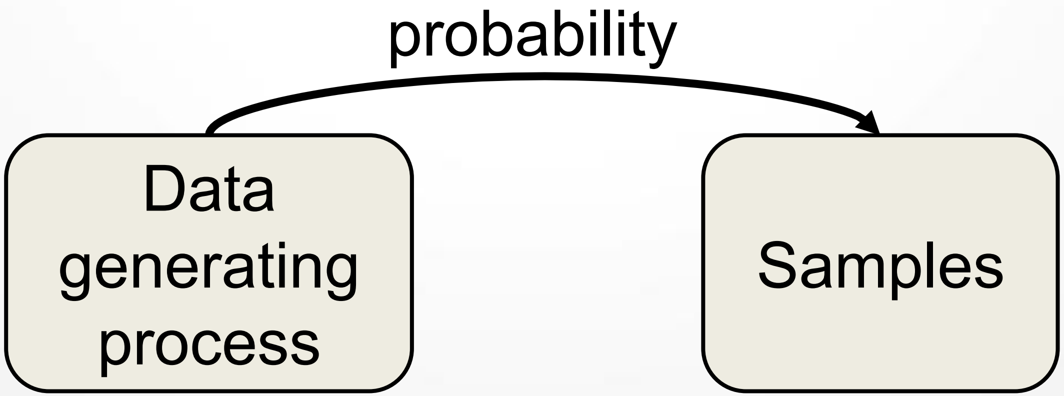
- predict other data generated from the same source
- know how different our sample could have been
- draw conclusions about whole population and not just observed sample - generalize



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

An **experiment** is some process whose outcome is random (e.g. flipping a coin, rolling a die).

A **set** is an unordered collection of items.

- Sets are usually denoted with { curly brackets }.
- $|A|$ denotes the number of elements in set A

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

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

Probability

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

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

- $0 <= p(s) <= 1$ for each s in S

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

Flip a coin

Sample Space

H T

1/2 1/2

Die roll

1 2 3

1/6 1/6 1/6

4 5 6

1/6 1/6 1/6

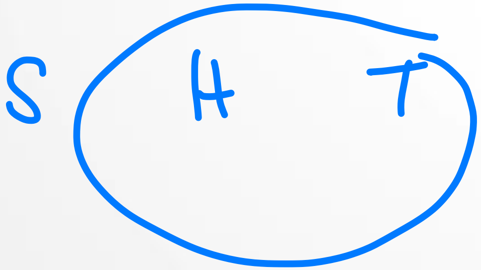
Probability

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

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

- $0 <= p(s) <= 1$ 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

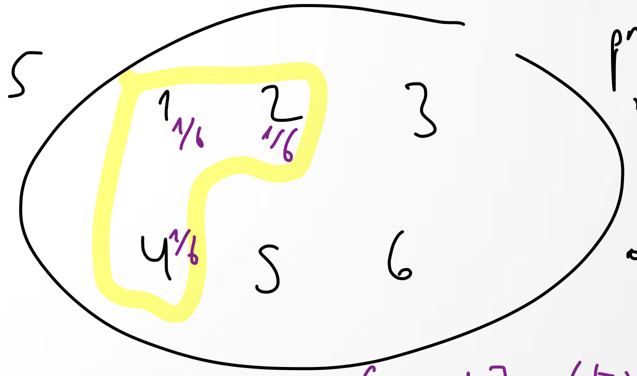
Sample space, *S*: (finite or countable) set of possible outcomes.

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

Dice roll

- $0 <= p(s) <= 1$ for each s in S

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



prob of rolling 1, 2, or 4?

Event, *E*: is a subset of the sample space

$p(E) = \sum_{s \in E} p(s) \Rightarrow 0 \leq p(E) \leq 1$

$E = \{1, 2, 4\}$ $p(E) = \frac{1}{2} = \frac{3}{6}$

$E = \phi \Rightarrow p(E) = 0$

$E = S \Rightarrow p(E) = 1$

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

$\sum_{s \in S} p(s) = 1$ $p(s) = p$ for all s in S

$= \sum_{s \in S} p = np = 1 \Rightarrow p = p(s \in S) = \frac{1}{n}$

When flipping a fair coin successively three times:

A. The sample space is {H, T}
B. The event $\underbrace{\text{\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}}}_{S=}$ has probability less than 1. $\cancel{p(E=S)} < 1$
C. The uniform distribution assigns probability 1/8 to each outcome. $p(s \in S) = \frac{1}{8}$
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:

$$p(E) = \sum_{s \in E} p(s) = \sum_{s \in E} \frac{1}{n} = \underbrace{\frac{1}{n} + \frac{1}{n} + \ldots + \frac{1}{n}}_{\text{size of } E} = \frac{\#\text{outcomes in } E}{\#\text{outcomes in } S}$$

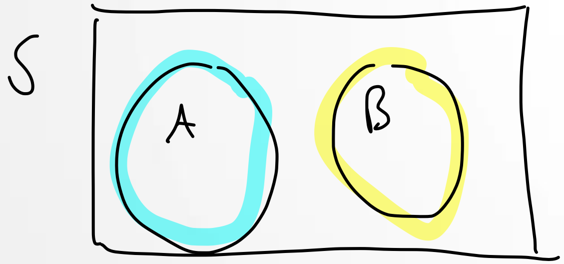
$$= \frac{|E|}{|S|}$$

$p(s) = \frac{1}{n}$

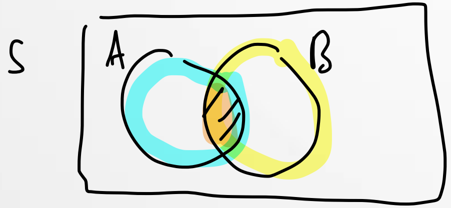
Addition Rule

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

$$p(A \text{ or } B) = \sum_{\substack{s\in A \\ \text{or} \\ s\in B}} p(s) = \sum_{s\in A} p(s) + \sum_{s\in B} p(s) = p(A) + p(B)$$



More generally (not mutually exclusive)



$$\underset{p(A\cup B)}{p(A \text{ or } B)} = \sum_{\substack{s\in A \\ \text{or } s\in B}} p(s) = p(A) + p(B) - \underset{p(A\cap B)}{p(A \text{ and } B)}$$

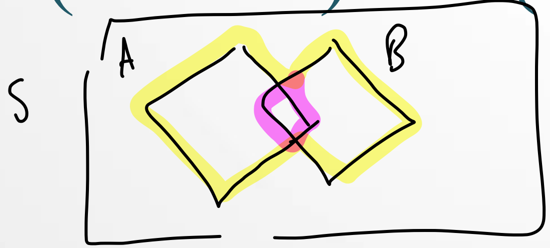
Addition Rule

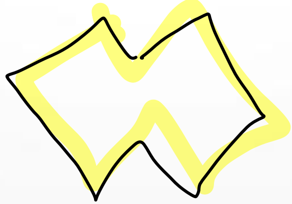
If A and B are mutually exclusive events (cannot happen simultaneously), then

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

In general

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





$A \cup B$

$A \cap B$

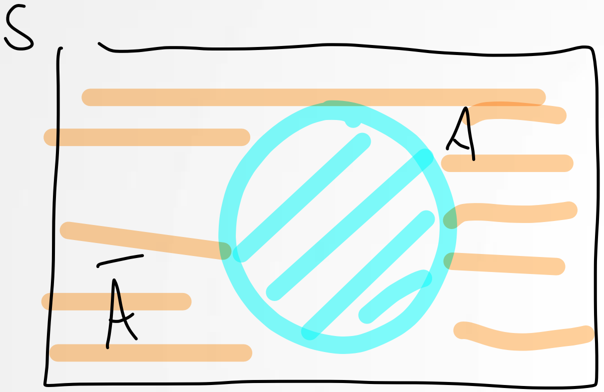
Multiplication Rule

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

conditional probability

$$= P(B) \cdot P(A|B) = P(B \cap A)$$

Complement Rule



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

not A

complement of A

$$p(\bar{A}) = \sum_{s \notin A} p(s) = \sum_{s \in S} p(s) - \sum_{s \in A} p(s) = 1 - p(A)$$

$$P(A) + P(\bar{A}) = P(S) = 1$$

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?

(i)

S: 1 2 3 4 5 6

$p(A \text{ and } B) =$

$p(A) \cdot p(B|A) = \frac{1}{6} \cdot 0 = 0$

(ii)

$S = \{1, 2, 3, 4, 5, 6\}$

$A = \{1\}$ $B = \{2\}$

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

$= \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$

(iii)

$S =$ all pairs of rolls

$A = \{1 \text{ for 1st roll}\}$

$B = \{2 \text{ for 2nd roll}\}$

$p(A \text{ and } B) = p(A)\,p(B|A) =$

$p(A)\,p(B) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$

from Theory Meets Data by Ani Adhikari, Chapter 4

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?

from Theory Meets Data by Ani Adhikari, Chapter 4

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.