DSC 40A

Theoretical Foundations of Data Science I

Foundations of Probability

Announcements

- Homework 5 will be released on Friday.
- Midterm grades and solutions to be released by Friday.

Resources

- 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.

Agenda

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

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 6)

- Summary statistics and loss functions; empirical risk minimization.   Statistics
- 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.

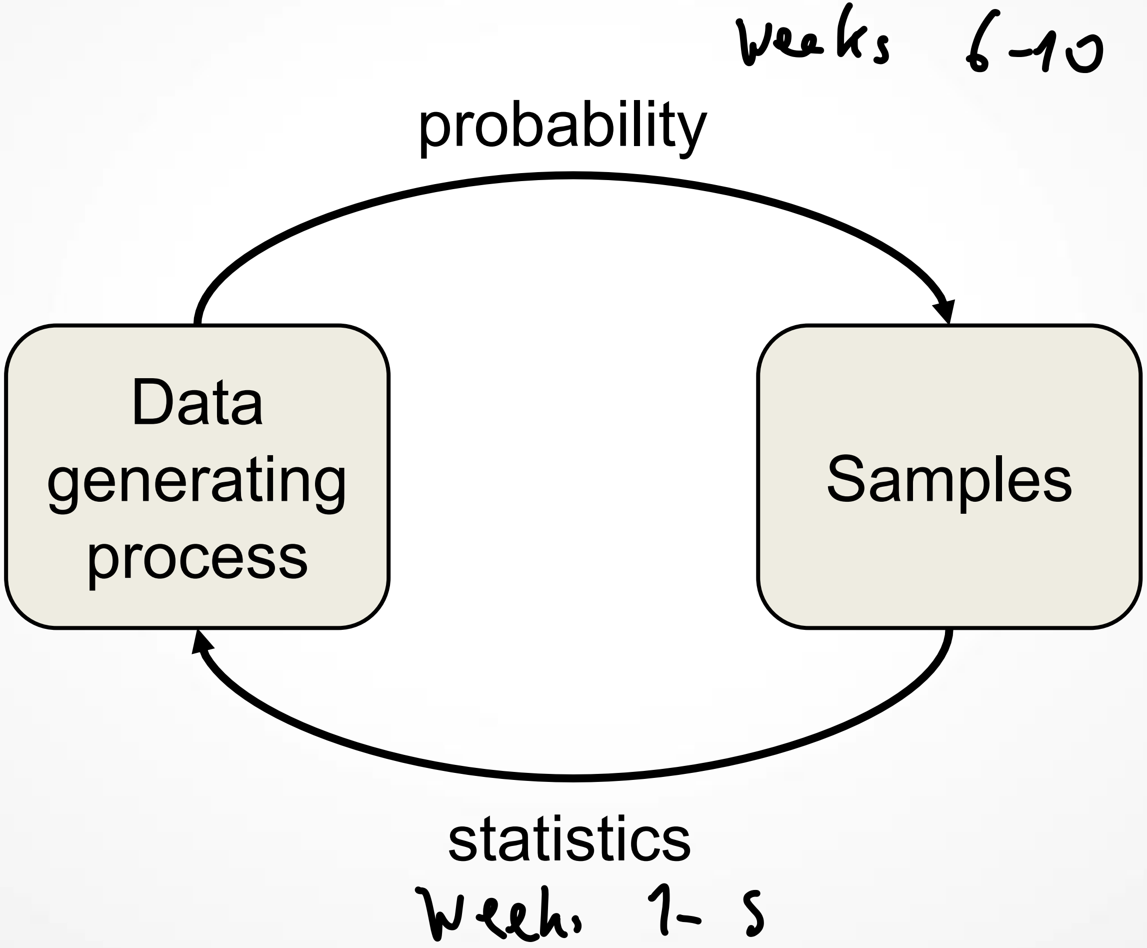
Predicting from Samples

- 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.

Today

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

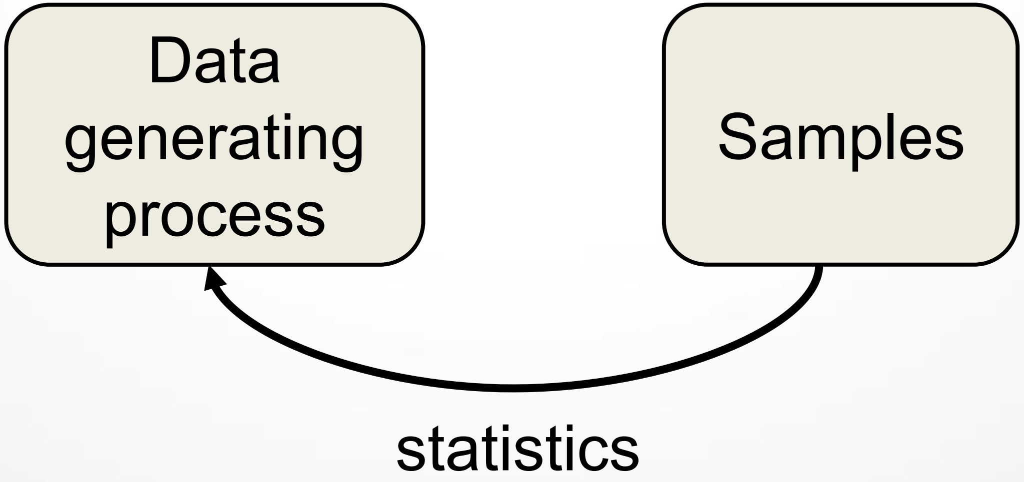
Probability and Statistics



Statistical Inference

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

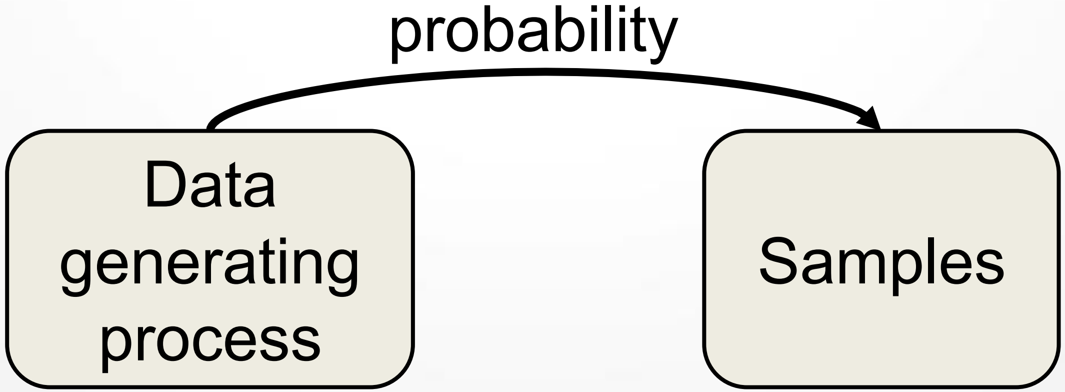
- predict other data generated from the same source
- know how different our sample could have been
- draw conclusions about whole population and not just observed sample - generalize



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   Bayes classifier



Probability

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

prob = 1/2 1/2
{ 'H', 'T' } coin

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

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

← cardinality

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

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

{1, 2, 3, 4, 5, 6} die

prob 1/6 for all outcomes

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

Probability

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

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

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

S	H	T
p(fair)	$\frac{1}{2}$	$\frac{1}{2}$
p(bias)	$\frac{1}{10}$	$\frac{9}{10}$

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.

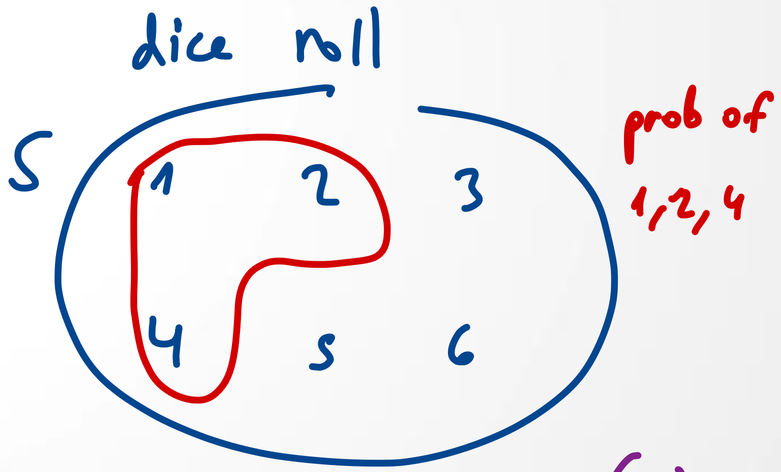
Probability

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

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

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

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



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

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

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

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

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 all"

$\sum_{s \in S} p(s) = 1 \qquad p(s) = p \quad \forall s \in S$

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

When flipping a fair coin successively three times:

A. The sample space is {H, T} → sample space for one coin toss
B. The event {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT} has probability less than 1. → $E = S \Rightarrow p(E) = 1$
C. The uniform distribution assigns probability 1/8 to each outcome.
D. None of the above.

$|S| = 8$

$p = \frac{1}{8}$

S = {HHH, TTT, HHT, HTH, THH, TTH, THT, HTT}

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:

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

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

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 prob

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.