CS109: Probability for Computer Scientists

Lisa Yan
April 6, 2020
<table>
<thead>
<tr>
<th>3</th>
<th>Introduction + Intro to counting</th>
<th>LIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>Counting II</td>
<td>01b_counting_ii</td>
</tr>
<tr>
<td>73</td>
<td>Pigeonhole Principle</td>
<td>01c_pigeonhole</td>
</tr>
<tr>
<td>79</td>
<td>Permutations I</td>
<td>01d_permutations</td>
</tr>
</tbody>
</table>

Today’s discussion thread: [https://us.edstem.org/courses/109/discussion/24490](https://us.edstem.org/courses/109/discussion/24490)
Welcome to CS109!
Lecture with zoom

- Turn on your camera if you are able, mute your mic in the big room
- Virtual backgrounds are encouraged (classroom-appropriate)
Yes, my undergrad was here...

...But now I’m here!!!

Received PhD 2019
Now:
Stanford’s newest CS lecturer

My interests over time

Create technology

Teaching

Help people

Networks, Data Science

Education Tools

Create technology to help people
Why I like probability

• I like data

• I want to help people

• Probability helps me help people with data

• Also Pokemon

\[ a = \frac{(3 \times HP_{\text{max}} - 2 \times HP_{\text{current}}) \times \text{rate} \times \text{bonus}_{\text{ball}}}{3 \times HP_{\text{max}}} \times \text{bonus}_{\text{status}} \]
What makes this quarter important

We are seeing a huge surge in **statistics, predictions, and probabilistic models** shared through global news, governing bodies, and social media.

Global cases of COVID-19 as of April 1st (JHU)
https://coronavirus.jhu.edu/map.html

Predicted Hospital Resource Use in United States (IHME)
https://covid19.healthdata.org/projections

Cases per 100K in NY, NJ, and CA counties (my dad)
https://app.flourish.studio/login
What makes this quarter important

We are seeing a huge surge in statistics, predictions, and probabilistic models shared through global news, governing bodies, and social media.

The challenge of delivering Stanford-class education online reflects our university’s commitment to fostering a diverse body of students.

126 survey responses
What makes this quarter important

We are seeing a huge surge in statistics, predictions, and probabilistic models shared through global news, governing bodies, and social media.

The challenge of delivering Stanford-class education online reflects our university’s commitment to fostering a diverse body of students.

The technological and social innovation we develop during this time will strongly impact how we approach truly world-class education.
What makes this quarter important

We are seeing a huge surge in statistics, predictions, and probabilistic models shared through global news, governing bodies, and social media. The challenge of delivering Stanford-class education online reflects our university’s commitment to fostering a diverse body of students. The technological and social innovation we develop during this time will strongly impact how we approach truly world-class education.

The S/NC grading guidelines means that you have the freedom to set your own learning goals and learn for the sake of learning.
What makes this quarter important

We are seeing a huge surge in statistics, predictions, and probabilistic models shared through global news, governing bodies, and social media.

The challenge of delivering Stanford-class education online reflects our university’s commitment to fostering a diverse body of students.

The technological and social innovation we develop during this time will strongly impact how we approach truly world-class education.

The S/NC grading guidelines means that you have the freedom to set your own learning goals and learn for the sake of learning.

My goals this quarter (at minimum)

To teach you how probability applies to real life
To help you foster and maintain human connections throughout this course
that being said...
What makes this quarter important

These are extraordinary circumstances.

The teaching staff and I realize that this quarter cannot replace an in-person, on-campus experience. Your diverse backgrounds amplify this difference.

All our situations may change.

We are committed to working through this version of this course together and adapting as a class and as a community. We welcome your thoughts.

Thank you in advance for being patient with necessary changes to make this educational experience fulfilling, meaningful, and equitable.
The CS109 teaching team
What about you?

...first, some Breakout Room guidelines...
Lecture with Zoom

- Turn on your camera if you are able, mute your mic in the big room
- Virtual backgrounds are encouraged (classroom-appropriate)

Breakout Rooms for meeting your classmates
- Just like sitting next to someone new

We will use Ed instead of Zoom chat
- Like raising your hand in the classroom, except with a lower barrier to entry
- You can upvote your classmates’ posts
- Persistent copy: Teaching staff and I can answer questions during and after lecture
- Better threading/reply support, copy/paste, LaTeX math mode, emojis

Join discussion forum here: [https://us.edstem.org/join/BmUE24](https://us.edstem.org/join/BmUE24)
Today’s discussion thread: [https://us.edstem.org/courses/109/discussion/24490](https://us.edstem.org/courses/109/discussion/24490)
Post or upvote some thoughts on Ed:
• What is something you hope to get out of this quarter?
• What are you worried about this quarter?
• What are your hopes for CS109, given that it is online and S/NC?

Join discussion forum here: https://us.edstem.org/join/BmUE24

Today's discussion thread: https://us.edstem.org/courses/109/discussion/24490
Introduce yourself! (name, major, year)

Then check out the responses your classmates wrote, and comment/discuss!
- What is something you hope to get out of this quarter?
- What are you worried about this quarter?
- What are your hopes for CS109, given that it is online and S/NC?

Join discussion forum here: https://us.edstem.org/join/BmUE24

Today’s discussion thread: https://us.edstem.org/courses/109/discussion/24490
Course mechanics
Course mechanics (light version)

- For more info, read the Administrivia handout and FAQ


- Canvas (only for posting videos/recordings)
Prerequisites

**CS106B/X**
- Programming
- Recursion
- Hash tables
- Binary trees

**MATH 51/CME 100**
- Multivariate differentiation
- Multivariate integration
- Basic facility with linear algebra (vectors)

**CS103**
(co-requisite OK)
- Proofs (induction)
- Set theory
- Math maturity

Important!
How many units should I take?

Hours per week = Units × 3
Average about 10 hours / week for assignments

Are you an undergrad?

No

Do you want to take CS109 for fewer units?

Yes

3 Units -or- 4 Units

No

Start Here

Yes

5 Units
Will this class count towards my CS degree?

Yes.

“For CS-MS, CS-BS, and CS-Minor students:

All classes taken Spring quarter will satisfy requirements as if taken for a letter grade. This applies to CS-MS requirements, CS-BS requirements, CS-Minor requirements, and the SoE requirements for the CS major.”

If you are an undergraduate, you still must take this course for 5 units.
Staff contact

- Discussion forum: [https://us.edstem.org/courses/109/discussion/](https://us.edstem.org/courses/109/discussion/)
- Staff email cs109@cs.stanford.edu
- **Working** office hours: For all timezones (starting later this week)
- Contact mailing list for course level issues, extensions, etc.
Lecture format

”Probability is a number between 0 and 1”

Short pre-recorded lecture (several 5-10 min videos)

”What is the definition of probability? (select one)”

Concept check quiz on Gradescope (part of grade, submit infinitely many times)

“What is the probability that you get exactly 3 heads in 5 coin flips?”

50-min in-person, discussion-oriented lecture MWF 10:30am-11:20am PT (note 50 min, not 80 min)
Where you learn

Pre-recorded lectures

Live lectures recordings posted to Canvas

Discussion Section starting Week 2

Lecture notes on website

Textbook readings optional

Problem Sets 6

Quizzes 2

Optional, open-ended contest 1
S/NC Class breakdown

60%  6 Problem Sets

25%  Quizzes
     • Thursday, April 30
     • Thursday, May 20

15%  Participation
     • Concept checks on pre-recorded material
     • Section participation (alternatives provided)
60% Problem Sets

“Passing work” 60% on each problem set

Late Policy
+5% grade for on-time submission
+0% bonus for 1 class day late
cap 80% for 2 class days late
cap 60% for 3 class days (1 week) late

Review session #1 this Friday 4/10 (time TBA)

Optional but encouraged, tutorial online TBD
Quizzes, Participation

Quizzes

- Ideally, 1-2 hours of individual work
- 24-hour take-home window

Participation (full policy on website)

1. (10%) Concept checks: Submit for pre-lecture recording, unlimited submissions/autograder before each lecture
2. (5%) Section participation
CS109 Contest

• Announced mid-quarter
• A meaningful submission will replace your section grade, stronger submissions replace problem sets for passing work

Your baseline is CS109, and the sky is the limit.

Previous winning submissions:
• Recidivism Risk: Algorithmic Prediction and Racial Bias
• A Better Way to Reform the Electoral College
• Monte Carlo Tree Search for Tic Tac Toe
Stanford Honor Code

Permitted
• Talk to the course staff ✓
• Talk with classmates (cite collaboration) ✓
• Look up general material online ✓

NOT permitted:
• Copy answers:
  from classmates
  from former students
  from previous quarters

• Copy answers from the internet
  Besides, these are usually incorrect
Why you should take CS109
Traditional View of Probability
CS view of probability

http://www.site.com

But wait...
There's MORE!!
Machine Learning
= Machine (compute power)
+ Probability
+ Data
Machine Learning Algorithm

Data $\rightarrow$ Build a *probabilistic model* $\rightarrow$ Do one thing
Classification
Where is this useful?

A machine learning algorithm performs better than the best dermatologists.

Developed in 2017 at Stanford.

Image tagging
Decision-making: The last remaining board game
Augmented Reality Machine Translation

Automatic machine translation on Google Translate
Voice assistants

What can I help you with?

Alexa    Siri    Google Now    Cortana
Probability is *more* than just machine learning.
Probability and medicine

Predicted Hospital Resource Use in United States (IHME)
https://covid19.healthdata.org/projections

How do COVID-19 testing rates in a region correlate with the actual spread of the disease?
Probability and art
Probability and climate
Probabilistic analysis of algorithms
Probability in practice

Frequently bought together

Total price: $117.03

Lisa Yan, CS109, 2020
Probability at your fingertips
Probability and philosophy
Probability for good

How do we identify systemic biases in our data and incorporate human judgment into our probabilistic models?

Algorithms of Oppression, Safiya Umoja Noble. 2018
We’ll get there!
Probability is not always intuitive.
A patient takes a virus test that returns positive. What is the probability that they have the virus?

- 0.03% of people have the virus
- Test has 99% positive rate for people with the virus
- Test has 7% positive rate for people without the virus

Correct answer: 0.42%
Probability = Important + Needs Studying
Counting I
What is Counting?

An experiment in probability:

Counting: How many possible outcomes can occur from performing this experiment?
What is Counting?

Roll

6

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

Roll even only

3

$\{2, 4, 6\}$

Roll

36

$\{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}$
Sum Rule of Counting

If the outcome of an experiment can be either from Set $A$, where $|A| = m$, or Set $B$, where $|B| = n$, where $A \cap B = \emptyset$, Then the number of outcomes of the experiment is $|A| + |B| = m + n$. 

One experiment

\[ A = \{1, 3, 5, 8\} \]

\[ B = \{2, 4, 6\} \]

\[ 3 + 3 = 6 \]
Product Rule of Counting

If an experiment has two parts, where

The first part’s outcomes are from Set $A$, where $|A| = m$,

and the second part’s outcomes are from Set $B$, where $|B| = n$,

Then the number of outcomes of the experiment is

$|A||B| = mn$. 

Two-step experiment
Let’s try it out

Sum Rule, Product Rule, or something else? How many outcomes?

1. Video streaming application
   • Your application has distributed servers in 2 locations (SJ: 100, Boston: 50).
   • If a web request is routed to a server, how large is the set of servers it can get routed to?

2. Dice
   • How many possible outcomes are there from rolling two 6-sided dice?

3. Strings
   • How many different orderings of letters are possible for the string BOBA?

Think, pair, and we’ll come back as a group.
Post any questions here: https://us.edstem.org/courses/109/discussion/24490
Let’s try it out

Sum Rule, Product Rule, or something else? How many outcomes?

1. Video streaming application
   • Your application has distributed servers in 2 locations (SJ: 100, Boston: 50).
   • If a web request is routed to a server, how large is the set of servers it can get routed to?

2. Dice
   • How many possible outcomes are there from rolling two 6-sided dice?

3. Strings
   • How many different orderings of letters are possible for the string BOBA?
For next time

- Watch pre-recorded lectures for today (Monday 4/6) and Wednesday 4/8 to be posted this afternoon PT
- Complete one concept check that covers both lectures to be posted this afternoon PT

http://cs109.stanford.edu/
Questions?
Counting II

Gradescope quiz, blank slide deck, etc.
(Available Monday 4/6 evening PT)
http://web.stanford.edu/class/cs109/
**TOP DEFINITION**

**kick it up a notch**

To make things more intense, exciting, or interesting.

(introduced by **chef Emeril Lagasse** in reference to **spicing** up his recipes )
Inclusion-Exclusion Principle

If the outcome of an experiment can be either from Set $A$ or set $B$, where $A$ and $B$ may overlap, then the total number of outcomes of the experiment is

$$|A \cup B| = |A| + |B| - |A \cap B|.$$
Transmitting bytes over a network

An 8-bit string is sent over a network.
• The receiver only accepts strings that either start with 01 or end with 10.

How many 8-bit strings will the receiver accept?

Define

$A$: 8-bit strings starting with 01

$B$: 8-bit strings ending with 10
Transmitting bytes over a network

An 8-bit string is sent over a network.

- The receiver only accepts strings that either start with 01 or end with 10.

How many 8-bit strings will the receiver accept?

Define

\[ A : \text{8-bit strings starting with } 01 \]

\[ B : \text{8-bit strings ending with } 10 \]

\[ A \cap B : \text{8-bit strings starting with } 01 \text{ and ending with } 10 \]

\[ |A| = 2^6 \]

\[ |B| = 2^6 \]

\[ |A \cup B| = |A| + |B| - |A \cap B| \]

\[ = 2^6 + 2^6 - 2^4 \]

\[ = 112 \]
General Principle of Counting

If an experiment has \( r \) steps, such that

Step \( i \) has \( n_i \) outcomes for all \( i = 1, \ldots, r \),

Then the number of outcomes of the experiment is

\[
\prod_{i=1}^{r} n_i.
\]

Multi-step experiment

Product Rule of Counting: A special case
License plates

How many CA license plates are possible if...

(pre-1982)

(present day)
License plates

How many CA license plates are possible if...

(pre-1982)

\[
\frac{26 \cdot 26 \cdot 10 \cdot 10 \cdot 10}{A-Z \quad A-Z \quad A-Z \quad 0-9 \quad 0-9 \quad 0-9} = 17,576,000
\]

(present day)

\[
\begin{align*}
\text{Soln 1} : & \quad 10 \cdot 26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 = 17,576,000 \\
\text{Soln 2} : & \quad 10 \cdot \frac{17,576,000}{0-9} = 1,760,000
\end{align*}
\]
Pigeonhole Principle

Gradescope quiz, blank slide deck, etc.
http://cs109.stanford.edu/
Floors and ceilings

Floor function

\[ [x] \]

The largest integer \( \leq x \)

Ceiling function

\[ [x] \]

The smallest integer \( \geq x \)

Check it out:

\[
\begin{align*}
\end{align*}
\]
Pigeonhole Principle

For positive integers $m$ and $n$,
if $m$ objects are placed in $n$ buckets,
then at least one bucket must contain
at least $\lfloor m/n \rfloor$ objects.

Example:

$m$ objects = 10 pigeons
$n$ buckets = 9 pigeonholes

At least one pigeonhole must contain $\lfloor m/n \rfloor = 2$ pigeons.

Bounds: an important part of CS109
Balls and urns

$n$ balls

$r$ urns (buckets)
Balls and urns  Hash Tables and strings

Consider a hash table with 100 buckets.
950 strings are hashed and added to the table.

1. Is it guaranteed that at least one bucket contains \textit{at least} 10 entries?
2. Is it guaranteed that at least one bucket contains \textit{at least} 11 entries?
3. Is it possible to have a bucket with \textit{no entries}?
Balls and urns  Hash Tables and strings

Consider a hash table with 100 buckets. 950 strings are hashed and added to the table.

1. Is it guaranteed that at least one bucket contains \textit{at least} 10 entries?

2. Is it guaranteed that at least one bucket contains \textit{at least} 11 entries?

3. Is it possible to have a bucket with \textit{no entries}?
Permutations I

Gradescope quiz, blank slide deck, etc.
http://cs109.stanford.edu/
Unique 6-digit passcodes with **six** smudges

How many unique 6-digit passcodes are possible if a phone password uses each of **six** distinct numbers?
Sort $n$ indistinct objects
Sort $n$ distinct objects

Ayesha  Tim  Irina  Joey  Waddie
Sort \( n \) distinct objects

Steps:

1. Choose 1\(^{\text{st}}\) can \(5\) options
2. Choose 2\(^{\text{nd}}\) can \(4\) options
   ...
5. Choose 5\(^{\text{th}}\) can \(1\) option

Total \( = 5 \times 4 \times 3 \times 2 \times 1 \)
   \( = 120 \)
Permutations

A permutation is an ordered arrangement of objects.

The number of unique orderings (permutations) of $n$ distinct objects is

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1.$$
Unique 6-digit passcodes with six smudges

How many unique 6-digit passcodes are possible if a phone password uses each of six distinct numbers?

Total = 6!

= 720 passcodes
Unique 6-digit passcodes with five smudges

How many unique 6-digit passcodes are possible if a phone password uses each of five distinct numbers?