CS109: Probability for Computer Scientists

Lisa Yan
June 25, 2018
Lisa Yan

Yes, I graduated from here... (UC Berkeley B.S., EECS)

...But now I’m here!!! (5th year PhD student in EE)

PhD: CS education in undergrad classes
More about me

Childhood: New Jersey, the place of dreams!!
Adulthood: California, with 19 times the dreams!!
What about you?
Course administrivia

Course website:

cs109.stanford.edu

Awesome Teaching Assistants:

Elliott Chartock       Ya Le       Andrew Davis
## Prerequisites

<table>
<thead>
<tr>
<th>CS106B/X</th>
<th>CS103</th>
<th>MATH 51/CME 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programming</td>
<td>Proofs (induction)</td>
<td>Multivariate differentiation</td>
</tr>
<tr>
<td>Recursion</td>
<td>Set theory</td>
<td>Multivariate integration</td>
</tr>
<tr>
<td>Hash tables</td>
<td>Math maturity</td>
<td>Basic facility with linear algebra (vectors)</td>
</tr>
<tr>
<td>Binary trees</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Important!*
Coding in CS109

Review session TBA (end of this week or next)
Readings

Lecture Notes
(a summary)

Textbook
(optional)
Staff contact

Piazza

Working office hours

Email cs109@cs.stanford.edu

Contact Lisa for course level issues.
CS109 Units

Start Here

Are you an Undergrad?

Yes 5 Units
No

Do you want to take CS109 for fewer units?

No

Yes

3 Units -or- 4 Units

Hours per week = Units × 3

Average about 10 hours / week for assignments
## Grades

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Component</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>45%</td>
<td>6 Assignments</td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>Midterm</td>
<td>Tuesday July 24(^{th}), 7-9pm</td>
</tr>
<tr>
<td>30%</td>
<td>Final</td>
<td>Friday August 17(^{th}), 3:30-6:30pm</td>
</tr>
<tr>
<td>5%</td>
<td>(Optional) participation</td>
<td></td>
</tr>
</tbody>
</table>
Optional Participation (5%)

These lectures are recorded...

1. mvideox.stanford.edu
2. Canvas (live streaming)

Participation:
- Attend ≥80% of lectures in person (SCPD: watch within 24hr)
- Give lecture feedback on 2 lectures
Late Days

2

(class days)
**Honor Code**

**Allowed to:**
- Talk to the course staff
- Talk with classmates *(cite collaboration)*
- Look up general material online

**NOT allowed to:**
- Copy answers:
  - from classmates
  - from former students
  - from previous quarters
- Copy answers from the internet
  Besides, these are usually incorrect
Probability

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

Me, circa 2003

The likelihood of catching Pokemon in the game
Traditional view of probability
CS view of probability

http://www.site.com

But wait... There’s MORE!!!
Machine Learning
= Machine (compute power)
+ Probability
+ Data
Voice assistants
Self-driving cars
Social media
Image tagging
Machine Learning
= Machine
  + Probability
  + Data
Counting
Goals for today

Combinatorial Analysis
- The Sum and Product Rules of Counting
- Inclusion-Exclusion Principle and General Principle of Counting
- General Pigeonhole Principle
- Permutations
- BSTs
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

$|A| + |B| = m + n$. 
Video streaming application

Problem:
Distributed servers:

If a server request is sent to the application, how large is the set of servers it can get routed to?

Solution:
Set A: San Jose servers, $|A| = 100$, or
Set B: Boston servers, $|B| = 50$
where $A \cap B = \emptyset$

$|A| + |B| = 150$ servers
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

$|A| \cdot |B| = mn$. 
Dice

Problem:

Two 6-sided dice are rolled.

How many possible outcomes of the roll are there?

Solution:

Set A: first die’s outcomes, \(|A| = 6\), and
Set B: second die’s outcomes, \(|B| = 6\)

\(|A| \cdot |B| = 36\) outcomes
Inclusion-Exclusion Principle

If the outcome of an experiment can be either from

Set $A$, where $|A| = m$, or
Set $B$, where $|B| = n$,
where $A$ and $B$ may overlap,

Then the number of outcomes of the experiment is

$$|A \cup B| = |A| + |B| - |A \cap B|.$$
Bit strings

How many ways can 4 bits be ordered that start or end with 1?

Solution:

|A|: \[ 2^3 = 8 \text{ strings that start with 1} \]
|B|: \[ 2^3 = 8 \text{ strings that end with 1} \]
|A \cap B|: \[ 2^2 = 4 \text{ that start and end with 1} \]

\[ |A \cup B| = 8 + 8 - 4 = 12 \text{ ways} \]
Break!
Two **functions**

<table>
<thead>
<tr>
<th>Floor function</th>
<th>Ceiling function</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lfloor x \rfloor )</td>
<td>( \lceil x \rceil )</td>
</tr>
<tr>
<td>The largest integer ( \leq x )</td>
<td>The smallest integer ( \geq x )</td>
</tr>
</tbody>
</table>

Check it out:

\[
\begin{align*}
\lfloor 1/2 \rfloor &= 0 & \lfloor 2.9 \rfloor &= 2 & \lceil 8.0 \rceil &= 8 & \lceil -1/2 \rceil &= -1 \\
\lfloor 1/2 \rfloor &= 1 & \lfloor 2.9 \rfloor &= 3 & \lceil 8.0 \rceil &= 8 & \lceil -1/2 \rceil &= 0 \\
\end{align*}
\]
General 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 $\left\lfloor \frac{m}{n} \right\rfloor$ objects.

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

“At least one bucket must contain $\left\lfloor \frac{m}{n} \right\rfloor$ objects”
$= 1$ hole has 2 pigeons
Hash tables

Problem:
A hash table has 100 buckets. \( n = 100 \)
You hash 950 strings to the table. \( m = 950 \)

Questions:
1. Is there a bucket that has \textit{at least 10 entries}?
   Pigeonhole Principle: \( \left\lceil \frac{950}{100} \right\rceil = \left\lceil 9.5 \right\rceil = 10 \) entries
   Yes!

2. Is there a bucket that has \textit{at least 11 entries}?
   No!

3. Can there be a bucket \textit{with no entries}?
   Sure!
General Principle of Counting

If \( r \) experiments are performed such that:

- Part 1 has \( n_1 \) outcomes
- Part 2 has \( n_2 \) outcomes
- Part \( i \) has \( n_i \) outcomes,

Then the total number of outcomes is:

\[
\prod_{i=1}^{r} n_i = n_1 \times n_2 \times \ldots \times n_r.
\]
License plates

Problem:
How many license plates are there if...

Solution:
\[26 \times 26 \times 26 \times 10 \times 10 \times 10 = 17,576,000\]

\[10 \times 26 \times 26 \times 26 \times 10 \times 10 \times 10 = 175,760,000\]
Binary functions

Problem:
How many binary functions $f$ are defined on $n$ points?
Define $f(i) \rightarrow \{0, 1\}$ for $i = 1, 2, \ldots, n$

Solution:

$$2 \times 2 \times 2 \times 2 \times 2 \times \cdots \times 2 \times 2 = 2^n$$

$\Rightarrow$ Number of bit strings of length $n$, where bit $i$ is value of $f(i)$
Permutations

A permutation is an ordered arrangement of distinct objects. $n$ distinct objects can be permuted in:

$$n \times (n - 1) \times (n - 2) \times \ldots \times 2 \times 1 = n!$$

ways

Problem:
4 PCs, 3 Macs, 2 Linux computers are to be scheduled for maintenance. How many ways can they be scheduled if:

1. ...they can be arranged in any order? $9! = 362,880$

2. ...the PCs, Macs, Linuxes each need to be scheduled together?
Permutations, cont.

4 PCs, 3 Macs, 2 Linux computers scheduled for maintenance.
How many ways can they be scheduled if:

2. ...the PCs, Macs, Linuxes each need to be scheduled together?

All PCs together: \(4!\)
All Macs together: \(3!\)
All Linuxes together: \(2!\)
Ways to permute groups: \(3!\)

\((4!3!2!)3! = 1728\)
Phone passcodes

Problem:

iPhones have 4-digit passcodes. How many distinct passcodes are possible if:

...there are 4 smudges over 4 digits on the screen?

Solution:

4!
Phone passcodes

Problem:
iPhones have 4-digit passcodes.
How many distinct passcodes are possible if:

...there are 3 smudges over 3 digits on the screen?

(for next time)
A **binary search tree** (BST), is a binary tree where for *every* node $n$ in the tree:

- $n$'s value is **greater** than all the values in its left subtree.
- $n$'s value is **less** than all the values in its right subtree.
- both $n$'s left and right subtrees are binary search trees.
Binary Search Trees

Problem:
How many possible BSTs containing values 1, 2, and 3 have degenerate structure (i.e., each node in the BST has at most one child)?

Solution:
3! ways to order 1, 2, and 3 for insertion

There are 4 degenerate BSTs possible.
Summary

One experiment, multiple outcomes

Sum Rule of Counting
(general form) Inclusion-Exclusion Principle

Series of experiments

Product Rule of Counting
(general form) General Principle of Counting

General Pigeonhole Principle

Permutations

Σ (sum)

Π (product)