Probability
Review
Counting Rules

Counting operations on $n$ objects

- **Sort, order matters (perms)**
  - Distinct: $n!$
  - Some Distinct: $\frac{n!}{n_1!n_2! \ldots}$

- **Choose $k$ (combinations)**
  - Distinct: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

- **Put in $r$ buckets**
  - Distinct: $r^n$
  - None Distinct: $\frac{(n + r - 1)!}{n!(r - 1)!}$
Sample and Event Spaces

- **Sample Space**, $S$, is set of all possible outcomes of an experiment.
- **Event Space**, $E$, is some subset of $S$ ($E \subseteq S$).
What is a probability?

\[ P(E) = \lim_{n \to \infty} \frac{n(E)}{n} \]

The “event” \( E \) is that you hit the target.

\[ n \] is the number of trials.

\[ P(E) \approx 0.46 \]

Hit: 11
Thrown: 24
Recall: $S =$ all possible outcomes. $E =$ the event.

- Axiom 1: $0 \leq P(E) \leq 1$
- Axiom 2: $P(S) = 1$
- Axiom 3: $P(E^c) = 1 - P(E)$

Aside: axiom 3 is often stated as the probability of mutually exclusive events. We’ll come back to that later in the lecture…
Equally Likely Outcomes

- $P$(Each outcome) = \( \frac{1}{|S|} \)
- In that case, $P(E) = \frac{\text{number of outcomes in } E}{\text{number of outcomes in } S} = \frac{|E|}{|S|}$
End Review
Mandarins and Bananas

- 4 Mandarins and 3 Bananas in a Bag. 3 drawn.
  - What is P(1 Mandarin and 2 Bananas drawn)?

Equally likely sample space? Thought experiment
• 4 Mandarins and 3 Bananas in a Bag. 3 drawn.
  ▪ What is P(1 Mandarin and 2 Bananas drawn)?

• Ordered:
  ▪ Pick 3 ordered items: $|S| = 7 \times 6 \times 5 = 210$
  ▪ Pick Mandarin as either 1st, 2nd, or 3rd item:
    $|E| = (4 \times 3 \times 2) + (3 \times 4 \times 2) + (3 \times 2 \times 4) = 72$
  ▪ P(1 Mandarin, 2 Grapefruit) = $72/210 = 12/35$

• Unordered:
  ▪ $|S| = \binom{7}{3} = 35$
  ▪ $|E| = \binom{4}{1}\binom{3}{2} = 12$
  ▪ P(1 Mandarin, 2 Grapefruit) = $12/35$
Make indistinct items distinct to get equally likely sample space outcomes

*You will need to use this “trick” with high probability
Consider 5 card poker hands.

- “straight” is 5 consecutive rank cards of any suit
- What is $P(\text{straight})$?

$$|S| = \binom{52}{5}$$

$$|E| = 10 \cdot \binom{4}{1}^5$$

$$P(\text{straight}) = \frac{|E|}{|S|} = \frac{10 \cdot \binom{4}{1}^5}{\binom{52}{5}} \approx 0.00394$$

What is an example of one outcome?

Is each outcome equally likely?
Consider 5 card poker hands.

- “straight” is 5 consecutive rank cards of any suit
- “straight flush” is 5 consecutive rank cards of same suit
- What is \( P(\text{straight, but not straight flush}) \)?

\[
|S| = \binom{52}{5}
\]

\[
|E| = 10 \left( \binom{4}{1} \right)^5 - 10 \binom{4}{1}
\]

\[
P(\text{straight}) = \frac{|E|}{|S|} = \frac{10 \left( \binom{4}{1} \right)^5 - 10 \binom{4}{1}}{\binom{52}{5}} \approx 0.00392
\]
When approaching an “equally likely probability” problem, start by defining sample spaces and event spaces.
Chip Defect Detection

- $n$ chips manufactured, 1 of which is defective.
- $k$ chips randomly selected from $n$ for testing.
  - What is $P($defective chip is in $k$ selected chips$)$?

\[
|S| = \binom{n}{k}
\]

\[
|E| = \binom{1}{1}\binom{n-1}{k-1}
\]

\[
P(\text{defective chip is in } k \text{ selected chips}) = \frac{(n-1)!}{(k-1)!(n-k)!} = \frac{n!}{k!(n-k)!} = \frac{k}{n}
\]
Screen size = 800x800
Radius of target = 200

The dart is equally likely to land anywhere on the screen.

What is the probability of hitting the target?

\[
|S| = 800^2 \\
|E| = \pi 200^2
\]

\[
p(E) = \frac{\pi \cdot 200^2}{800^2} \approx 0.1963
\]
Screen size = 800x800
Radius of target = 200

The dart is equally likely to land anywhere on the screen.

What is the probability of hitting the target?

\[ |S| = 800^2 \]
\[ |E| = \pi 200^2 \]

\[ p(E) = \frac{\pi \cdot 200^2}{800^2} \approx 0.1963 \]
Let it find you.

SERENDIPITY

the effect by which one accidentally stumbles upon something truly wonderful, especially while looking for something entirely unrelated.
WHEN YOU MEET YOUR BEST FRIEND

Somewhere you didn't expect to.
Say the population of Stanford is 17,000 people

- You are friends with?
- Walk into a room, see 268 random people.
- What is the probability that you see someone you know?
- Assume you are equally likely to see each person at Stanford
Many times it is easier to calculate $P(E^C)$.
Trailing the dovetail shuffle to it’s lair – Persi Diaconosis
• What is the probability that in the \( n \) shuffles seen since the start of time, yours is unique?
  - \(|S| = (52!)^n\)
  - \(|E| = (52! - 1)^n\)
  - \(P(\text{no deck matching yours}) = (52!-1)^n/(52!)^n\)

  - For \( n = 10^{20} \),
    - \(P(\text{deck matching yours}) < 0.000000001\)

* Assume 7 billion people have been shuffling cards once a second since cards were invented
Back to Axiom 3
Recall: $S$ = all possible outcomes. $E$ = the event.

- Axiom 1: $0 \leq P(E) \leq 1$
- Axiom 2: $P(S) = 1$
- Axiom 3: $P(E^c) = 1 - P(E)$

Aside: axiom 3 is often stated as the probability of mutually exclusive events. We’ll come back to that later in the lecture…
Recall: \( S = \) all possible outcomes. \( E = \) the event.

- **Axiom 1:** \( 0 \leq P(E) \leq 1 \)
- **Axiom 2:** \( P(S) = 1 \)
- **Axiom 3:** If events \( E \) and \( F \) are mutually exclusive:
  \[
P(E \cup F) = P(E) + P(F)
  \]
Mutually Exclusive Events

$P(E \cup F) = P(E) + P(F)$

If events are mutually exclusive, probability of OR is simple:
If events are mutually exclusive, probability of OR is simple:

\[ P(E \cup F) = \frac{7}{50} + \frac{4}{5} = \frac{11}{50} \]
OR with Many Mutually Exclusive Events

\[
P(X_1 \cup X_2 \cup \cdots \cup X_n) = \sum_{i=1}^{n} P(X_i)
\]
If events are *mutually exclusive* probability of OR is easy!
\[ P(E^c) = 1 - P(E)? \]

\[ P(E \cup E^c) = P(E) + P(E^c) \]

Since \( E \) and \( E^c \) are mutually exclusive

\[ P(S) = P(E) + P(E^c) \]

Since everything must either be in \( E \) or \( E^c \)

\[ 1 = P(E) + P(E^c) \]

Axiom 2

\[ P(E^c) = 1 - P(E) \]

Rearrange
Stretch!
Conditional Probability
Why study probability?
• Roll two 6-sided dice, yielding values $D_1$ and $D_2$
• Let $E$ be event: $D_1 + D_2 = 4$
• What is $P(E)$?
  ▪ $|S| = 36$, $E = \{(1, 3), (2, 2), (3, 1)\}$
  ▪ $P(E) = 3/36 = 1/12$
• Let $F$ be event: $D_1 = 2$
• $P(E, \text{ given } F \text{ already observed})$?
  ▪ $S = \{(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6)\}$
  ▪ $E = \{(2, 2)\}$
  ▪ $P(E, \text{ given } F \text{ already observed}) = 1/6$
Dice – Our Misunderstood Friends

- Two people each roll a die, yielding $D_1$ and $D_2$. You win if $D_1 + D_2 = 4$

- Q: What do you think is the best outcome for $D_1$?
Conditional Probability

- **Conditional probability** is probability that E occurs *given* that F has already occurred “Conditioning on F”

- Written as $P(E|F)$
  - Means “P(E, given F already observed)”
  - Sample space, S, reduced to those elements consistent with F (i.e. $S \cap F$)
  - Event space, E, reduced to those elements consistent with F (i.e. $E \cap F$)
Conditional Probability

With equally likely outcomes:

\[ P(E \mid F) = \frac{\text{# of outcomes in } E \text{ consistent with } F}{\text{# of outcomes in } S \text{ consistent with } F} = \frac{\vert EF \vert}{\vert SF \vert} = \frac{\vert EF \vert}{\vert F \vert} \]

\[ P(E) = \frac{8}{50} \approx 0.16 \]

\[ P(E \mid F) = \frac{3}{14} \approx 0.21 \]
Conditional Probability

With equally likely outcomes:

\[ P(E \mid F) = \frac{\text{# of outcomes in } E \text{ consistent with } F}{\text{# of outcomes in } S \text{ consistent with } F} \]

\[ = \frac{|EF|}{|SF|} = \frac{|EF|}{|F|} \]

Shorthand notation for set intersection (aka set “and”)

\[ P(E) = \frac{8}{50} \approx 0.16 \]

\[ P(E \mid F) = \frac{3}{14} \approx 0.21 \]
Conditional Probability

• General definition:

\[ P(E \mid F) = \frac{P(EF)}{P(F)} \]

• Holds even when outcomes are not equally likely
• Implies: \( P(EF) = P(E \mid F) \cdot P(F) \) (chain rule)

• What if \( P(F) = 0 \)?
  ▪ \( P(E \mid F) \) undefined
  ▪ \textit{Congratulations! You observed the impossible!}
Generalized Chain Rule

• General definition of Chain Rule:

\[
P(E_1E_2E_3...E_n) = P(E_1)P(E_2 \mid E_1)P(E_3 \mid E_1E_2)...P(E_n \mid E_1E_2...E_{n-1})
\]
## Conditional Paradigm

<table>
<thead>
<tr>
<th>Name of Rule</th>
<th>Original Rule</th>
<th>Conditional Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>First axiom of probability</td>
<td>$0 \leq P(E) \leq 1$</td>
<td>$0 \leq P(E \mid G) \leq 1$</td>
</tr>
<tr>
<td>Complement Rule</td>
<td>$P(E) = 1 - P(E^C)$</td>
<td>$P(E \mid G) = 1 - P(E^C \mid G)$</td>
</tr>
<tr>
<td>Chain Rule</td>
<td>$P(EF) = P(E \mid F)P(F)$</td>
<td>$P(EF \mid G) = P(E \mid FG)P(F \mid G)$</td>
</tr>
</tbody>
</table>
What is the probability that a user will watch Life is Beautiful?

\[ P(E) \]

\[ S = \{ \text{Watch, Not Watch} \} \]

\[ E = \{ \text{Watch} \} \]

\[ P(E) = \frac{1}{2} ? \]
What is the probability that a user will watch Life is Beautiful?

\[ P(E) \]
What is the probability that a user will watch Life is Beautiful?

\[ P(E) \]

\[ P(E) = \lim_{n \to \infty} \frac{n(E)}{n} \approx \frac{\# \text{people who watched movie}}{\# \text{people on Netflix}} \]

\[ P(E) = \frac{10,234,231}{50,923,123} = 0.20 \]
Let $E$ be the event that a user watched the given movie:

- $P(E) = 0.19$
- $P(E) = 0.32$
- $P(E) = 0.20$
- $P(E) = 0.09$
- $P(E) = 0.23$

* These are the actual estimates
What is the probability that a user will watch Life is Beautiful, given they watched Amelie?

\[ P(E|F) = \frac{P(EF)}{P(F)} \]
What is the probability that a user will watch Life is Beautiful, given they watched Amelie?

\[ P(E|F) = \frac{P(EF)}{P(F)} = \frac{\text{#people who watched both}}{\text{#people on Netflix}} \times \frac{\text{#people who watched } F}{\text{#people on Netflix}} \]
What is the probability that a user will watch Life is Beautiful, given they watched Amelie?

\[ P(E|F) = \frac{P(EF)}{P(F)} = \frac{\text{#people who watched both}}{\text{#people who watched } F} \]

\[ P(E|F) = 0.42 \]
Netflix and Learn

Let $E$ be the event that a user watched the given movie,
Let $F$ be the event that the same user watched Amelie:

\[
P(E|F) = 0.14 \quad P(E|F) = 0.35 \quad P(E|F) = 0.20 \quad P(E|F) = 0.72 \quad P(E|F) = 0.49
\]
Machine Learning is:
Probability + Data + Computers