Probability

Today's Topics

Last time:

Combinations
Permutations
Bucketing

TODAY: Probability!

Sample Spaces and Event Spaces
Axioms of Probability
Lots of Examples

Next week:

Conditional Probability

Counting Rules

Counting operations on *n* objects

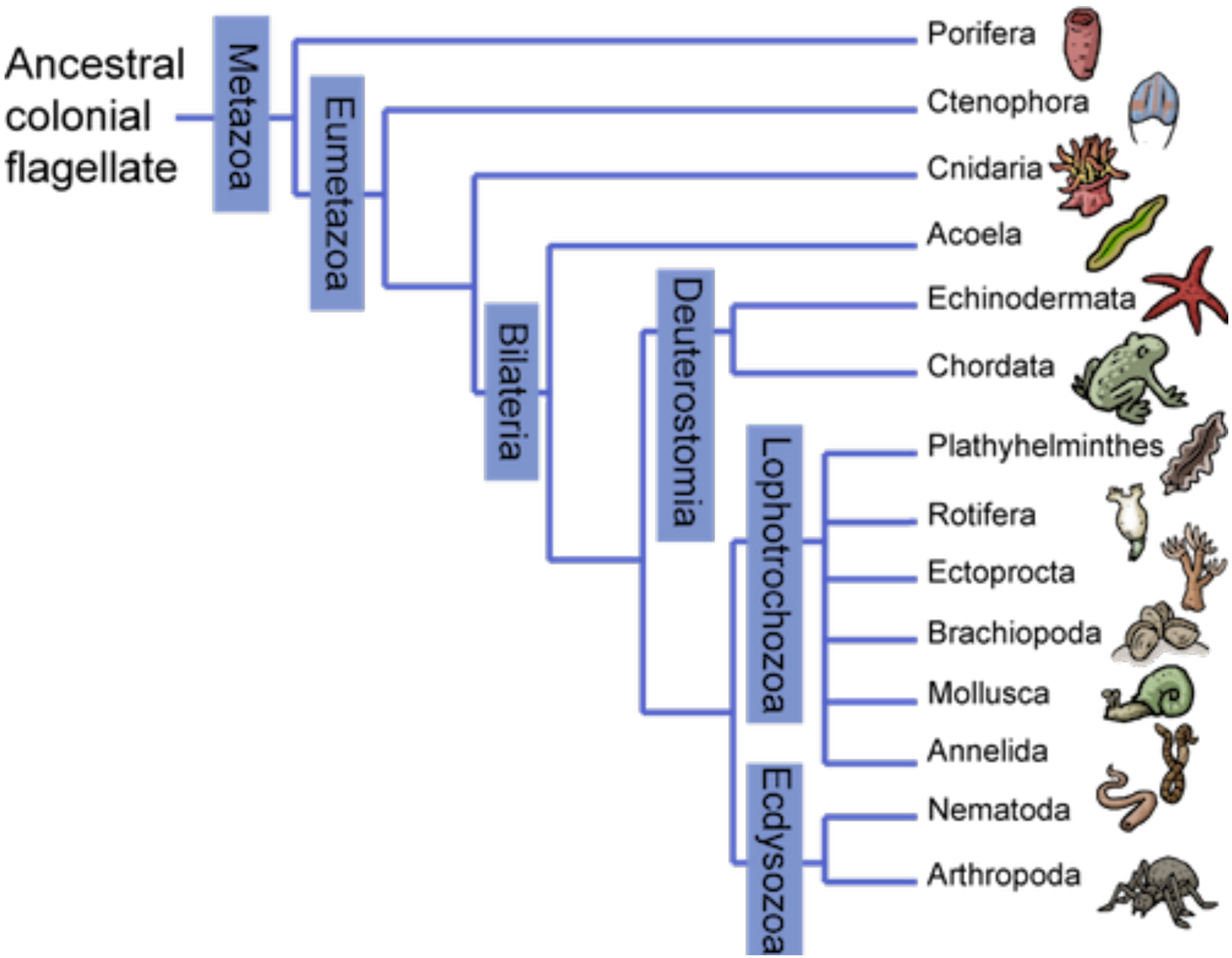
- Sort (order matters) (perms)
  - Distinct: $n!$
  - Some Distinct: $\dfrac{n!}{n_1!n_2!\ldots}$
- Choose *k* (combinations)
  - Distinct: $\dbinom{n}{k} = \dfrac{n!}{k!(n-k)!}$
- Put in *r* buckets
  - Distinct: r^n
  - None Distinct: $\dfrac{(n+r-1)!}{n!(r-1)!}$

Tool Review

Problem Archetype	Use this tool
4-smudge iPhone Codes	Permutations
Choose 2 Hunger Games tributes	Combinations (aka Binomial)
Choosing stats books	Combinations with cases
Mississippi	Permutations with indistinguishable elements (Multinomial)
Non-distinct strings into hashmaps	Permutations with divider-method

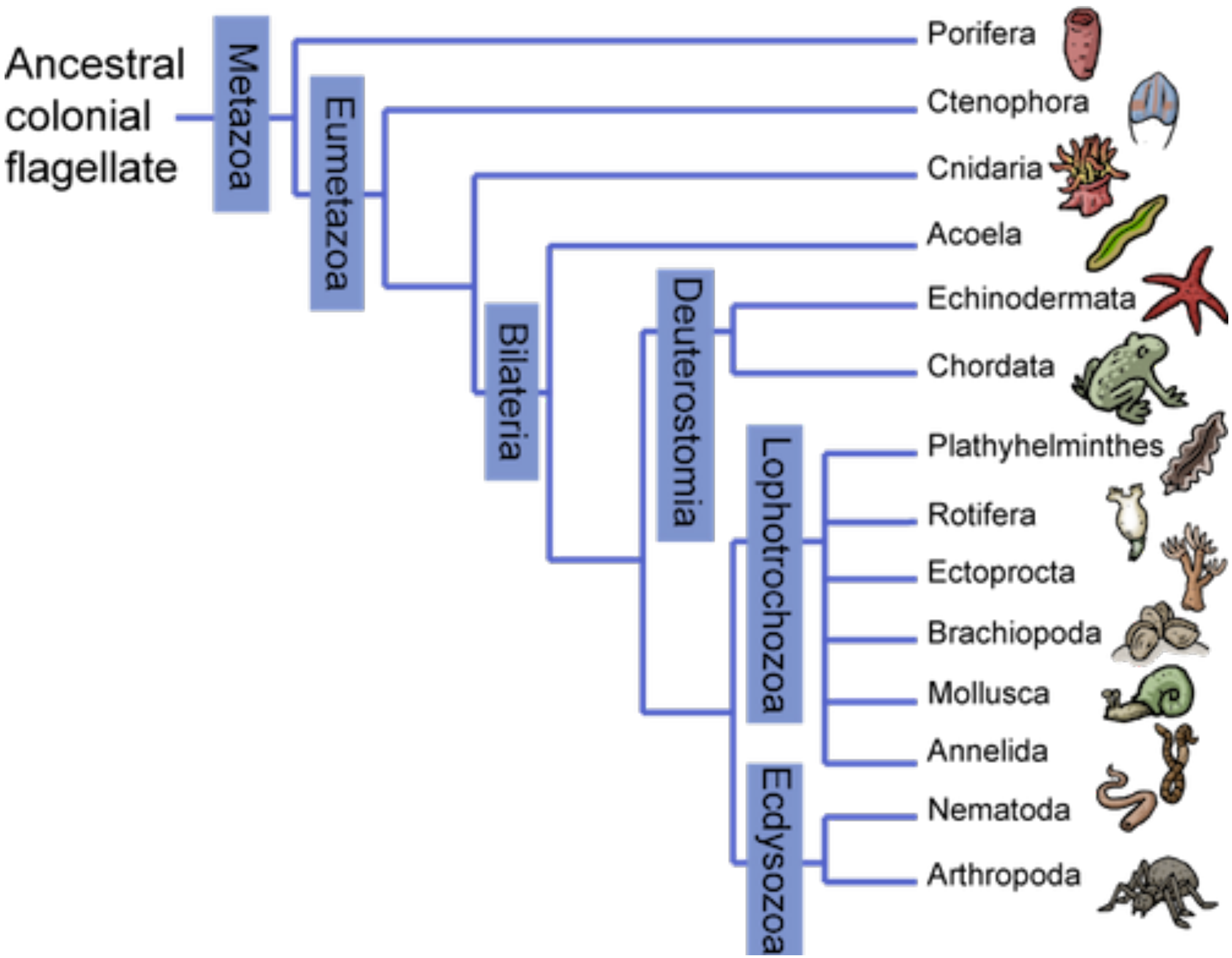
Counting Review

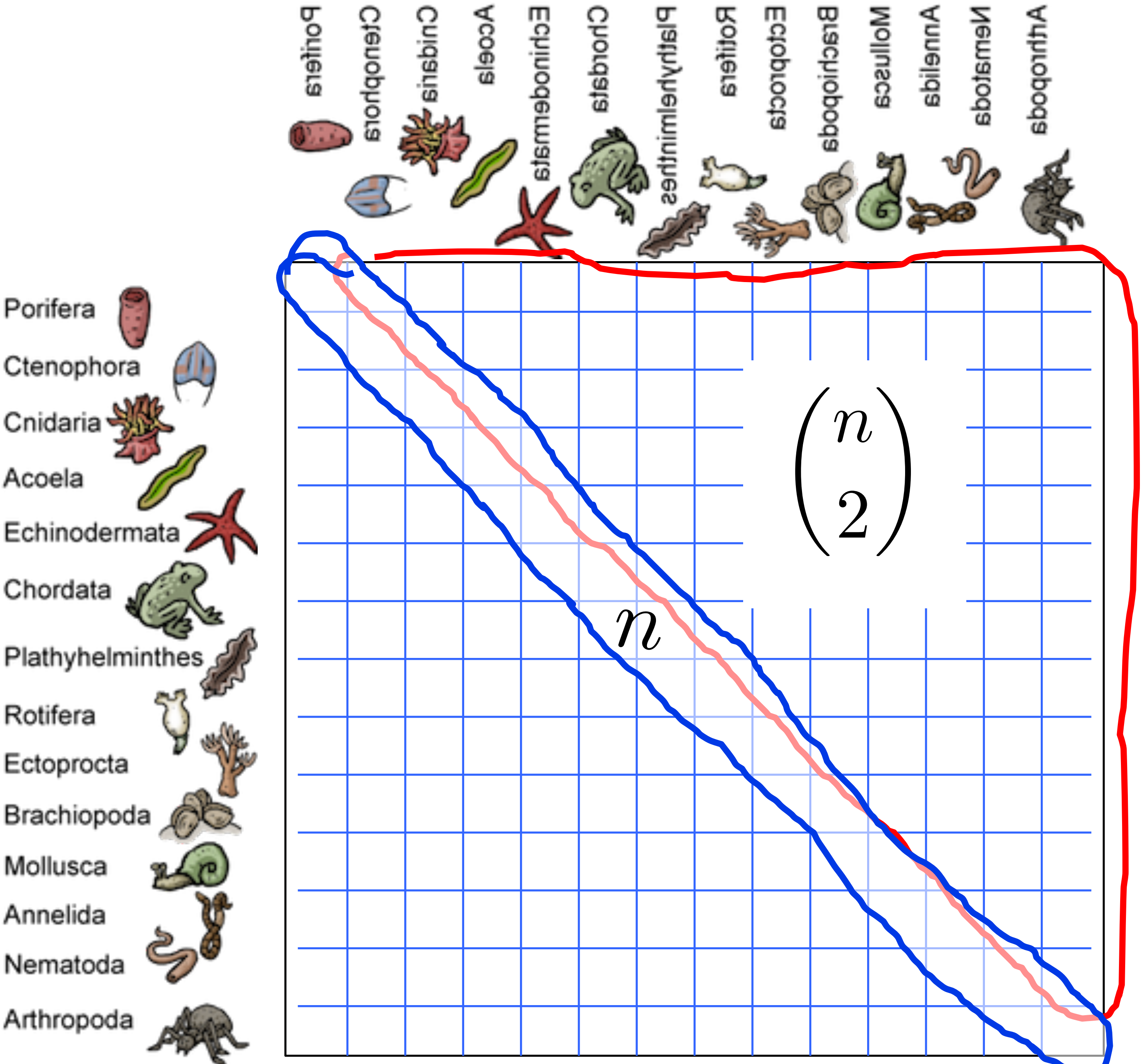
For a DNA tree we need to calculate the DNA distance between each pair of animals. How many calculations are needed?



Counting Review

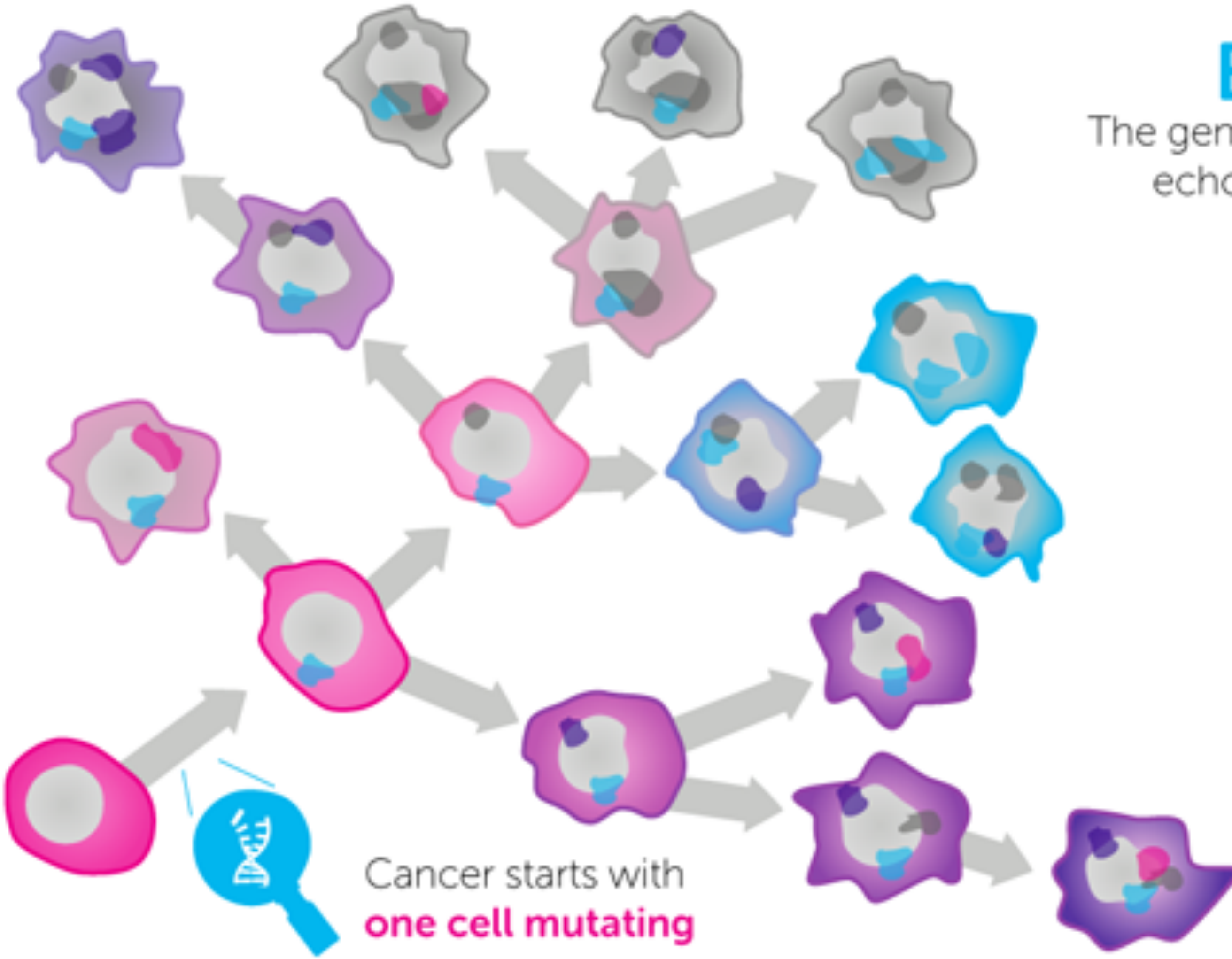
Q: There are n animals.
How many distinct pairs of animals are there?




Porifera
Ctenophora
Cnidaria
Acoela
Echinodermata
Chordata
Plathyhelminthes
Rotifera
Ectoprocta
Brachiopoda
Mollusca
Annelida
Nematoda
Arthropoda
n
$\binom{n}{2}$

BRANCHED EVOLUTION

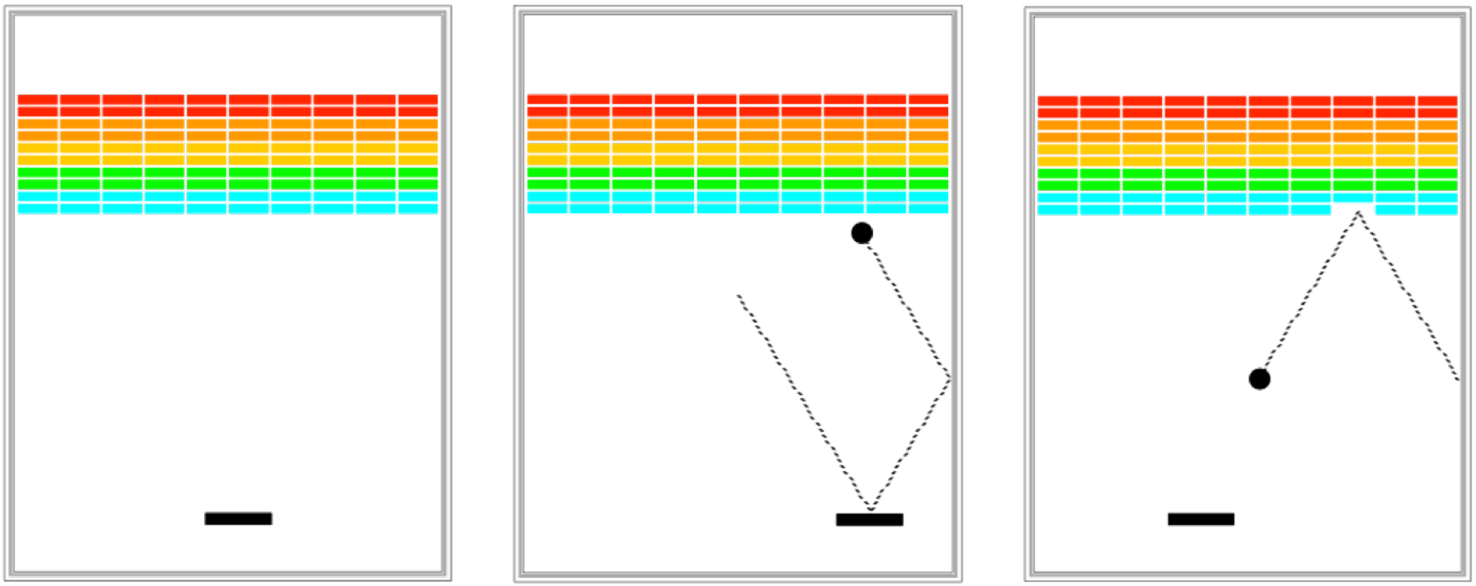
The genetic diversity in a tumour echoes Darwin's **Tree of Life**.



CANCER RESEARCH UK

Counting Review

Q: How many binary decisions make 10 million breakouts?



$$2^d = n$$

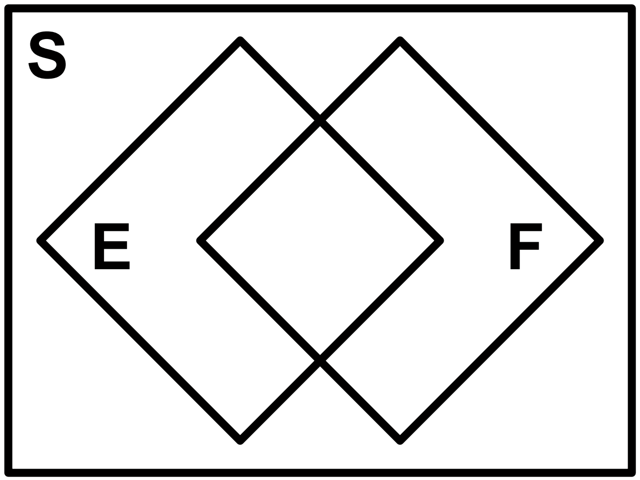
$$d = \log_2 n$$

$$\log_2 10,000,000 < 24$$

Sets Review

Set Operations Review

- Say E and F are subsets of S

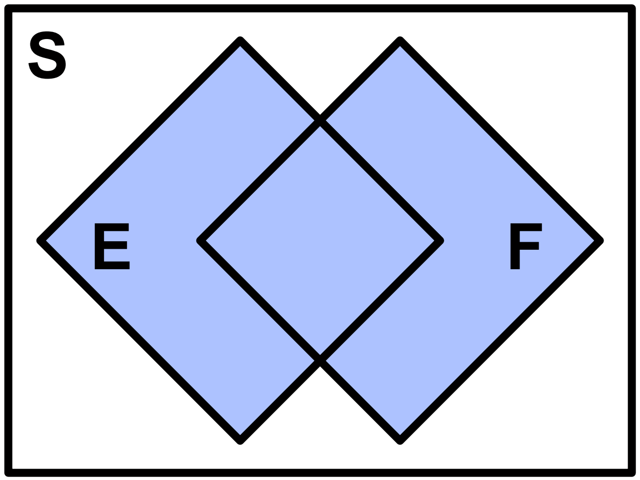


Set Operations Review

- Say E and F are events in S

Event that is in E <u>or</u> F

E ∪ F



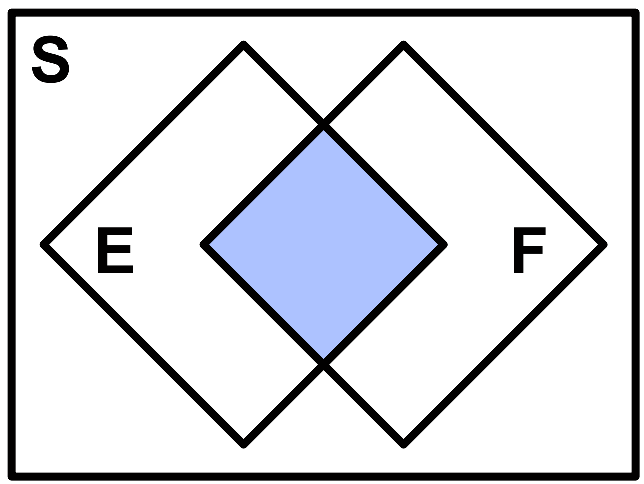
- S = {1, 2, 3, 4, 5, 6} die roll outcome
- E = {1, 2} F = {2, 3} E ∪ F = {1, 2, 3}

Set Operations Review

- Say E and F are events in S

Event that is in E <u>and</u> F

E ∩ F *or* **EF**



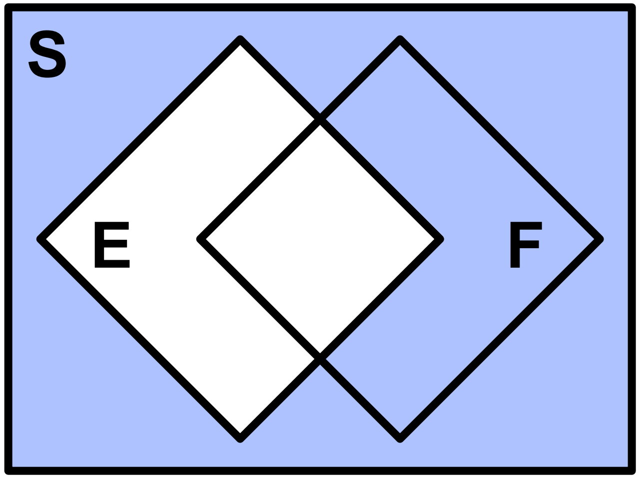
- S = {1, 2, 3, 4, 5, 6} die roll outcome
- E = {1, 2} F = {2, 3} E F = {2}
- **<u>Note:</u>** <u>*mutually exclusive*</u> events means E F = ∅

Set Operations Review

- Say E and F are events in S

  Event that is not in E (called complement of E)

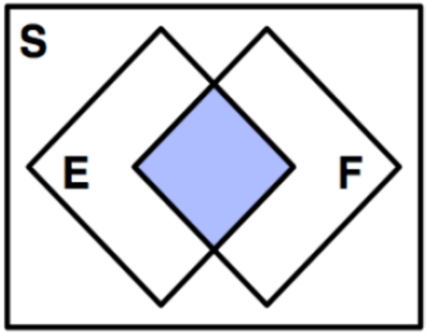
  E^c *or* **~E**



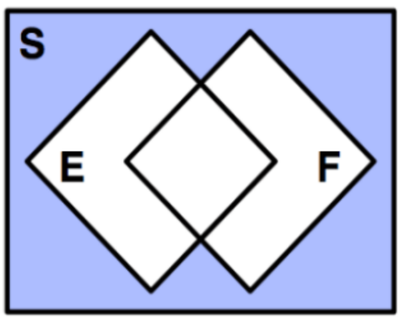
  - S = {1, 2, 3, 4, 5, 6}   die roll outcome
  - E = {1, 2} E^c = {3, 4, 5, 6}

Which is the correct picture for $\mathbf{E^c} \cap \mathbf{F^c}$

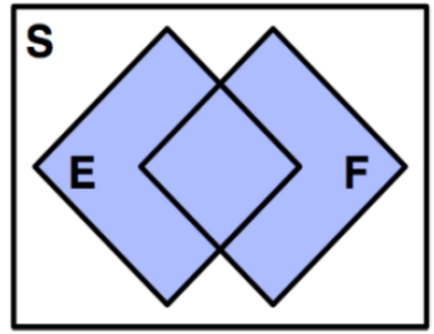
A



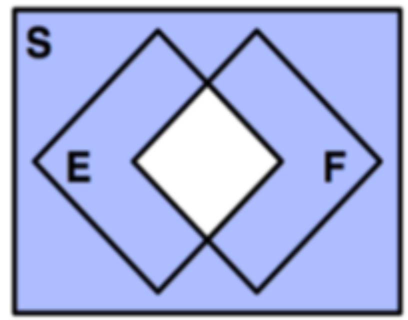
B



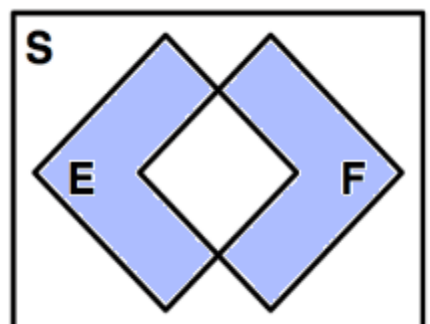
C



D



E

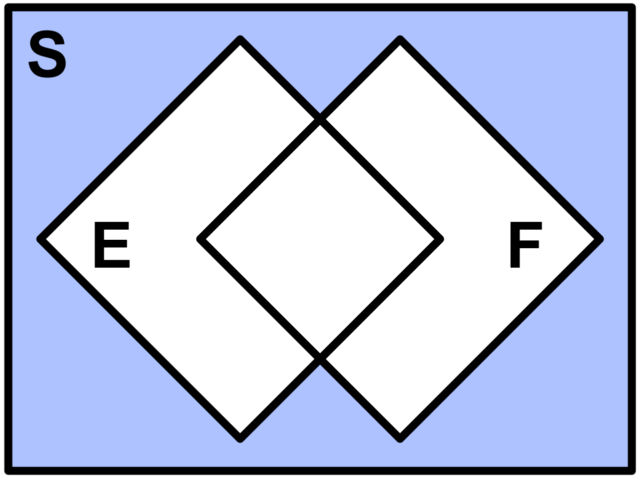


Set Operations Review

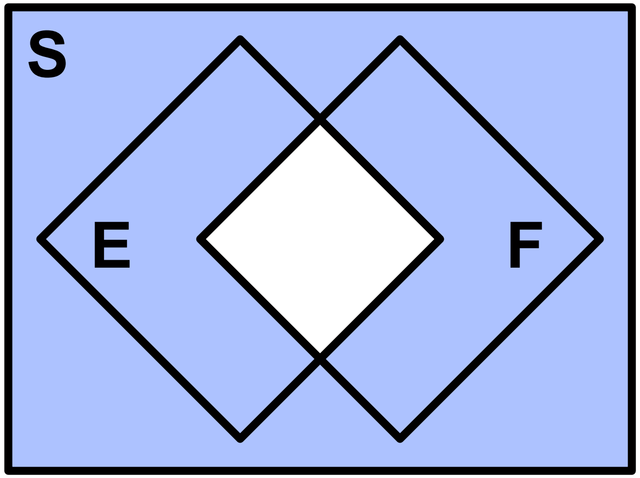
- Say E and F are events in S

DeMorgan's Laws

$(E \cup F)^c = E^c \cap F^c$



$(E \cap F)^c = E^c \cup F^c$



Sample Space

- **Sample space**, S, is set of all possible outcomes of an experiment
  - Coin flip: S = {Head, Tails}
  - Flipping two coins: S = {(H, H), (H, T), (T, H), (T, T)}
  - Roll of 6-sided die: S = {1, 2, 3, 4, 5, 6}
  - # emails in a day: $S = \{x \mid x \in \mathbf{Z}, x \geq 0\}$ (non-neg. ints)
  - YouTube hrs. in day: $S = \{x \mid x \in \mathbf{R}, 0 \leq x \leq 24\}$

Events

- **<u>Event</u>**, E, is some subset of S   ($E \subseteq S$)
  - Coin flip is heads: $E = \{\text{Head}\}$
  - $\geq$ 1 head on 2 coin flips: $E = \{(H, H), (H, T), (T, H)\}$
  - Roll of die is 3 or less: $E = \{1, 2, 3\}$
  - # emails in a day $\leq$ 20: $E = \{x \mid x \in \mathbf{Z}, 0 \leq x \leq 20\}$
  - Wasted day ($\geq$ 5 YT hrs.): $E = \{x \mid x \in \mathbf{R}, 5 \leq x \leq 24\}$

Note: When Ross uses: $\subset$, he really means: $\subseteq$

End Review

What is a probability?

Number between 0 and 1

To which we ascribe meaning…

What is a Probability

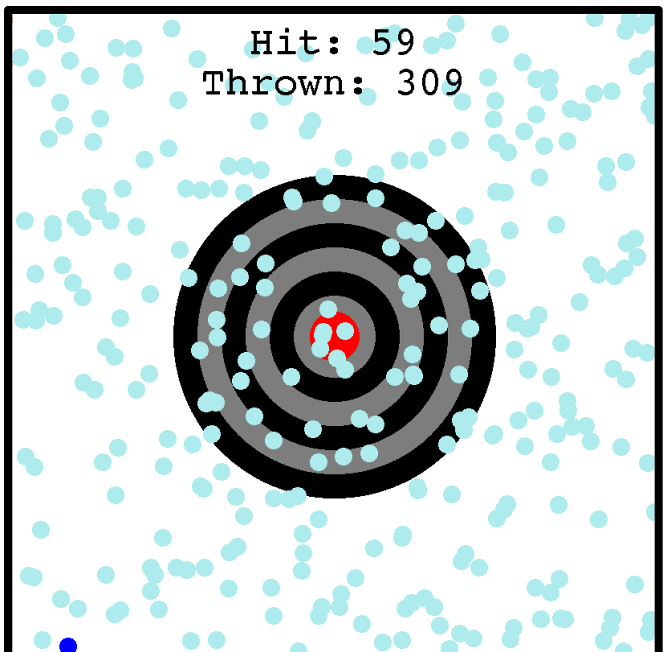
$$P(E)$$

* Our belief that an event E occurs

What is a Probability

What is a probability?

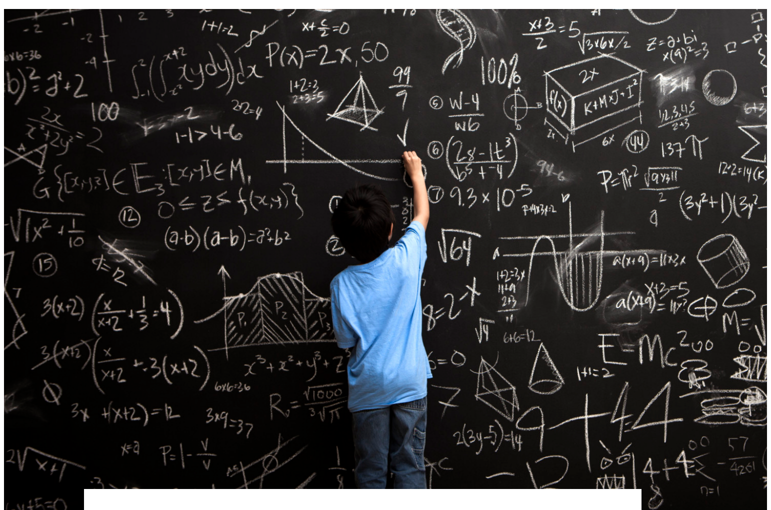
$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n}$$



Sources of Probability

1. Experimentation

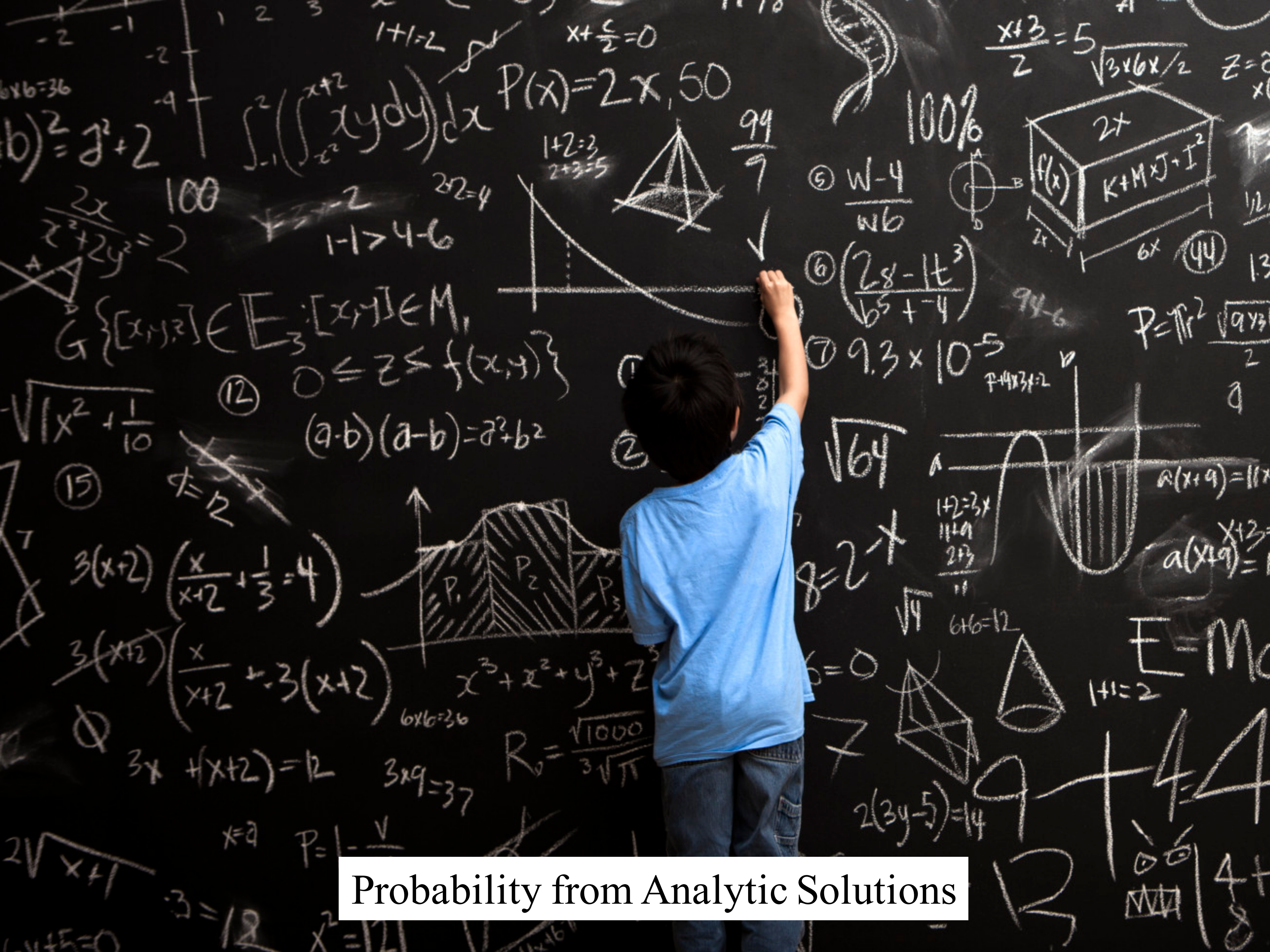
2. Dataset



3. Analytic Solution

4. Expert Opinion


90
MLB 040
VOLVO



Probability from Analytic Solutions

Axioms of Probability

Recall: S = all possible outcomes. E = the event.

- Axiom 1: $0 \leq P(E) \leq 1$

- Axiom 2: $P(S) = 1$

- Axiom 3: If E and F mutually exclusive ($E \cap F = \varnothing$), then $P(E) + P(F) = P(E \cup F)$

Implications of Axioms

- $P(E^c) = 1 - P(E)$ $\quad\quad$ $(= P(S) - P(E)\ )$
- If $E \subseteq F$, then $P(E) \leq P(F)$
- $P(E \cup F) = P(E) + P(F) - P(EF)$
  - This is just Inclusion-Exclusion Principle for Probability

General form of Inclusion-Exclusion Identity:

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{r=1}^{n} (-1)^{(r+1)} \sum_{i_1 < \ldots < i_r} P(E_{i_1} E_{i_2} \ldots E_{i_r})$$

Equally Likely Outcomes

- Some sample spaces have equally likely outcomes
  - Coin flip: S = {Head, Tails}
  - Flipping two coins: S = {(H, H), (H, T), (T, H), (T, T)}
  - Roll of 6-sided die: S = {1, 2, 3, 4, 5, 6}
- P(Each outcome) = $\frac{1}{|S|}$
- In that case, P(E) = $\frac{\text{number of outcomes in E}}{\text{number of outcomes in S}} = \frac{|E|}{|S|}$

Rolling Two Dice

- Roll two 6-sided dice.
  - What is P(sum = 7)?

- S = {(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)}
- E = {(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)}

- P(sum = 7) = |E|/|S| = 6/36 = 1/6

Mandarins and Grapefruit

- 4 Mandarins and 3 Grapefruit in a Bag. 3 drawn.
  - What is P(1 Mandarin and 2 Grapefruits drawn)?
- Ordered:
  - Pick 3 ordered items: |S| = 7 * 6 * 5 = 210
  - Pick Mandarin as either 1st, 2nd, or 3rd item:
    |E| = (4 * 3 * 2) + (3 * 4 * 2) + (3 * 2 * 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

Often make indistinct items distinct to get equally likely sample space outcomes

*You will need to use this "trick" with high probability

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(defective chip is in *k* selected chips)

$$= \frac{\binom{1}{1}\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{\frac{(n-1)!}{(k-1)!(n-k)!}}{\frac{n!}{k!(n-k)!}} = \frac{k}{n}$$

Any "Straight" Poker Hand

- Consider 5 card poker hands.
  - "straight" is 5 consecutive rank cards of any suit
  - What is P(straight)?
  - Note: this is a little different than the textbook
- |S| = $\binom{52}{5}$
- |E| = $10\binom{4}{1}^5$
- P(straight) = $\dfrac{10\binom{4}{1}^5}{\binom{52}{5}} \approx 0.00394$

Official “Straight” Poker Hand

- 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(straight, but not straight flush)?
- $|S| = \binom{52}{5}$
- $|E| = 10\binom{4}{1}^5 - 10\binom{4}{1}$
- $P(\text{straight}) = \dfrac{10\binom{4}{1}^5 - 10\binom{4}{1}}{\binom{52}{5}} \approx 0.00392$

Card Flipping

- 52 card deck. Cards flipped one at a time.
  - After first ace (of any suit) appears, consider next card
  - Is P(next card = Ace Spades) < P(next card = 2 Clubs)?
  - Initially, might think so, but consider the two cases:
  - First note: |S| = 52! (all cards shuffled)
- Case 1: Take Ace Spades out of deck
  - Shuffle left over 51 cards, add Ace Spades after first ace
  - |E| = 51! * 1 (only 1 place Ace Spades can be added)
- Case 2: Do same as case 1, but…
  - Replace "Ace Spades" with "2 Clubs" in description
  - |E| and |S| are the same as case 1
  - So P(next card = Ace Spade) **=** P(next card = 2 Clubs)

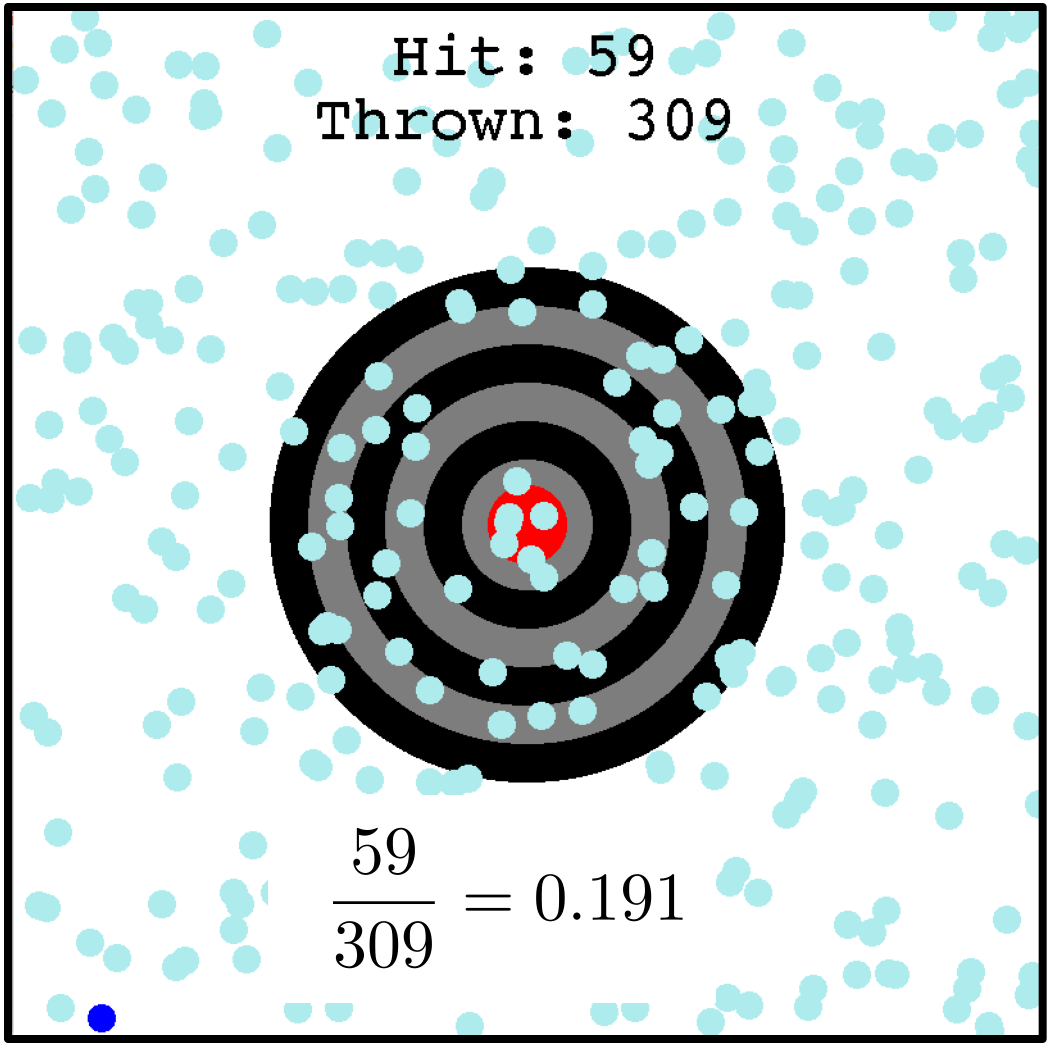
Selecting Programmers

- Say 28% of all students program in Java
  - 7% program in C++
  - 5% program in Java and C++
- What percentage of students do not program in Java or C++
  - Let A = event that a random student programs in Java
  - Let B = event that a random student programs in C++
  - $1 - P(A \cup B) = 1 - [P(A) + P(B) - P(AB)]$
    $= 1 - (0.28 + 0.07 - 0.05) = 0.7 \rightarrow 70\%$
- What percentage programs in C++, but not Java?
  - $P(A^c B) = P(B) - P(AB) = 0.07 - 0.05 = 0.02 \rightarrow 2\%$

When approaching a problem, start by defining events.

Then write the question in math notation.

Target Revisited



Screen size = 800x800
Radius of target = 200

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

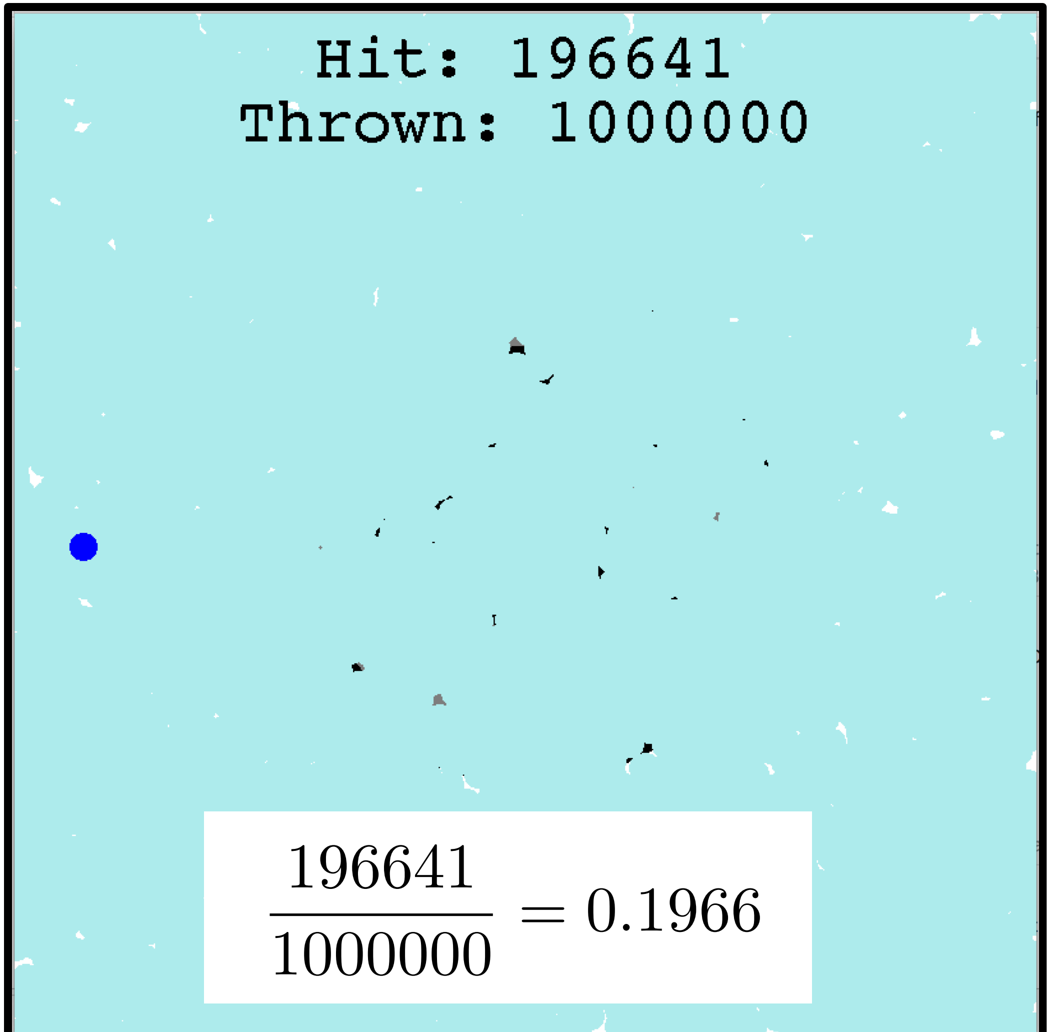
What is the probability of hitting the target?

$$|S| = 800^2$$

$$|E| = \pi 200^2$$

$$p(E) = \frac{\pi \cdot 200^2}{800^2} \approx 0.1963$$

Target Revisited



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?

$$|S| = 800^2$$

$$|E| = \pi 200^2$$

$$p(E) = \frac{\pi \cdot 200^2}{800^2} \approx 0.1963$$

SERENDIPITY

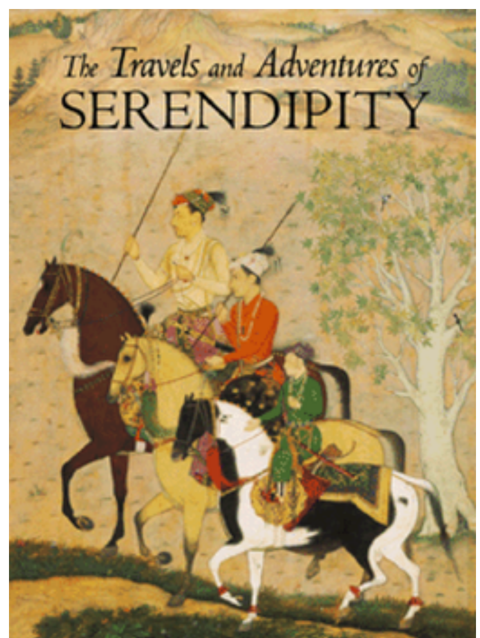
the effect by which one accidentally stumbles upon something truely wonderful, especially while looking for something entirely unrelated.

WHEN YOU MEET YOUR BEST FRIEND

Somewhere you didn't expect to.

Serendipity

- Say the population of Stanford is 21,000 people
  - You are friends with ?
  - Walk into a room, see 240 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 $P(E^C)$.

Birthdays

- What is the probability that of n people, none share the same birthday (regardless of year)?
  - $|S| = (365)^n$
  - $|E| = (365)(364)\ldots(365 - n + 1)$
  - P(no matching birthdays)

    $= (365)(364)\ldots(365 - n + 1)/(365)^n$
- Interesting values of n
  - $n = 23$: P(no matching birthdays) < ½ (least such n)
  - $n = 75$: P(no matching birthdays) < 1/3,000
  - $n = 100$: P(no matching birthdays) < 1/3,000,000
  - $n = 150$:

    P(no matching birthdays) < 1/3,000,000,000,000,000

Birthdays

- What is the probability that of *n* other people, none of them share the same birthday as <u>**you**</u>?
  - $|S| = (365)^n$
  - $|E| = (364)^n$
  - P(no birthdays matching yours) = $(364)^n/(365)^n$
- Interesting values of *n*
  - $n = 23$: P(no matching birthdays) $\approx$ 0.9388
  - $n = 150$: P(no matching birthdays) $\approx$ 0.6626
  - $n = 253$: P(no matching birthdays) $\approx$ 0.4995
    - Least such *n* for which P(no matching birthdays) < ½
  - Why are these probabilities much higher than before?
    - Anyone born on April 25th?
    - Is today anyone's birthday?

Crazy Version


Trailing the dovetail shuffle to it’s lair – Persi Diaconosis

Making History

- 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(no deck matching yours) = $(52!-1)^n/(52!)^n$

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

* Assumes 7 billion people have been shuffling cards once a second since cards were invented