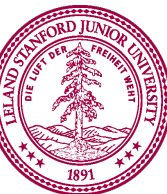


Probability

Announcements

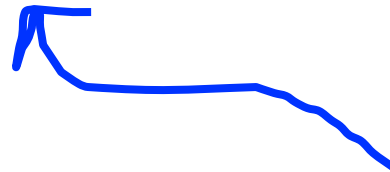
- It is Friday. Story time!
- One week extension to CS109 online.
- Sign up for sections by Sunday.
- PSet FAQ.
- Python Review Session today.
- We are going to make history today.



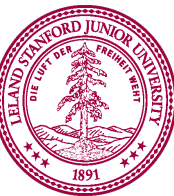
Stanford

JANUARY 7, 2022

- All other undergraduate courses will resume in-person on **Monday, Jan. 24**, at the beginning of Week 4.



That is us.
One week extension.



Sign up for Sections by Sunday noon

The screenshot shows the CS109 course website. The main navigation bar includes 'Course', 'Problem Sets', 'Lecture', 'Section', and 'Resources'. The 'Section' dropdown menu is open, highlighting 'Section Sign Up'. The page title is 'CS109: Probability for Computer Scientists' for Winter 2022, with a section meeting on Monday, Wednesday, Friday 1:30pm - 2:50pm. A 'Section Sign Up' button is visible. The page is divided into several sections: 'YOUR WEEK 1 TODO' with tasks like 'Come to Lecture!', 'Join Ed', 'Sign up for Section', and 'Start PSet 1'; 'TEACHING TEAM' featuring Professor Chris Piech; 'Schedule Coming Up' with a table of dates, lectures, and outside class activities; 'Sign up for section' with a link to the sign-up form; 'Python Tutorial this Friday'; 'RESOURCES' including Discussion, Course Reader, Syllabus, Schedule, and Office Hours; and 'KEY DATES' for Midterm, Final, and Last lecture. The URL for the sign-up form is <https://forms.gle/wZDokMYg5PTzDbHc6>.

Date	Lecture	Outside Class
Mon, Jan 3	Welcome to CS109	
Wed, Jan 5	Combinatorics	PSet 1 out
Fri, Jan 7	Probability	Section sign up, Python review

The screenshot shows the Google Form titled 'CS109 section sign-up form (Win 2022)'. The form asks for personal information to help with course pacing. The fields include: 'Email *' (with a note that it's required and a link to switch accounts), 'My preferred name is... *' (for full name), 'My pronouns are...' (with a note that it's optional), and 'My class year is... *'. The form is hosted on docs.google.com/forms.



Problem Set #1 is out

new app

The screenshot shows a web browser window with the URL `cs109psets.netlify.app/win22/pset1/randompasswords`. The page title is "PS1 5. Random Choice". The question asks: "What is the probability that both users will get the same randomly generated password? Provide an answer to three decimal places!". Below the question is a code editor with the following Python code:

```
import random

def main():
    user_1_password = generate_password()
    user_2_password = generate_password()

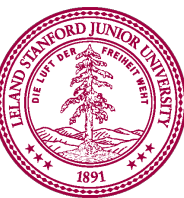
def generate_password():
    part_1 = random.choice([
        'red',
        'funky',
        'smelly'
    ])
    part_2 = random.choice([
        'apple',
        'pear',
        'pineapple'
    ])
    return part_1 + '-' + part_2
```

On the right side of the browser, there is an "Answer Editor" section. It contains a "Numeric Answer:" field with the value "97" and a "Check Answer" button. Below this is an "Explanation:" section with a rich text editor. The text in the editor is "I am so excited! What a good time!". The rich text editor has a toolbar with options for "Block LaTeX", "Image", "B", "I", and "U".

Check your answer

Insert LaTeX

Auto Submission



PSet FAQ

Q: Can I check my answers (many) times?

A: Yes. No penalty. We will know if you try all solutions.

Q: Will you just grade us on if the answers are correct?

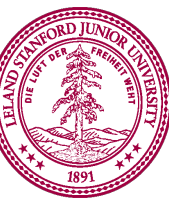
A: No! You must explain your answer. Graded on style and correctness.

Q: How much explanation should we provide?

A: As detailed as the course reader [worked examples](#)

Q: Must I do all my work on the app?

A: No



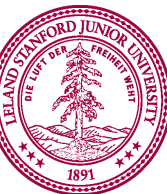
PSet Example Answer

Enigma Machine

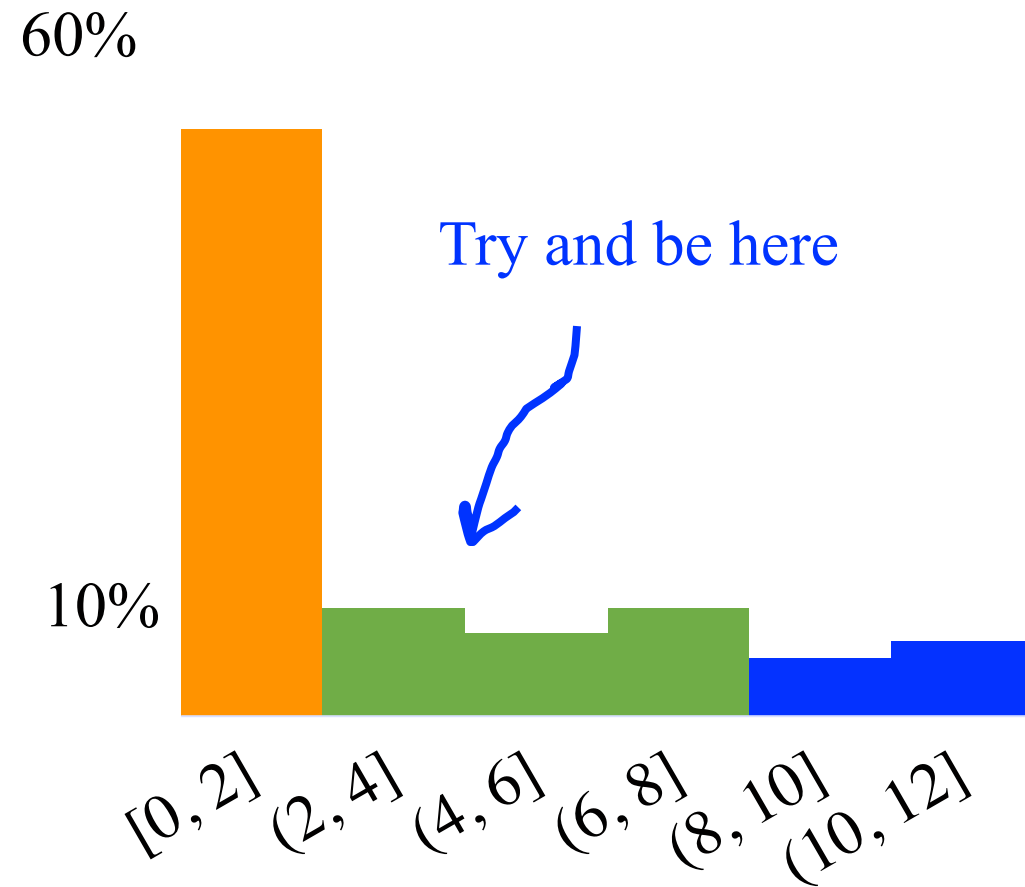
Two wires: How many ways are there to place exactly two wires? Recall that wires are not considered distinct. Each letter can have at most one wire connected to it, thus you couldn't have a wire connect 'K' to 'L' and another one connect 'L' to 'X'

There are $\binom{26}{2}$ ways to place the first wire and $\binom{24}{2}$ ways to place the second wire. However, since the wires are indistinct, we have double counted every possibility. Because every possibility is counted twice we should divide by 2:

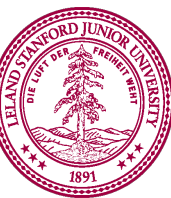
$$\text{Total} = \frac{\binom{26}{2} \cdot \binom{24}{2}}{2} = 44,850$$



PSet Status



About half the class has started



Python Review Session



Friday (today) at 6pm PT
(online)
With Tim Gianitsos

Find links, recordings, and setup here

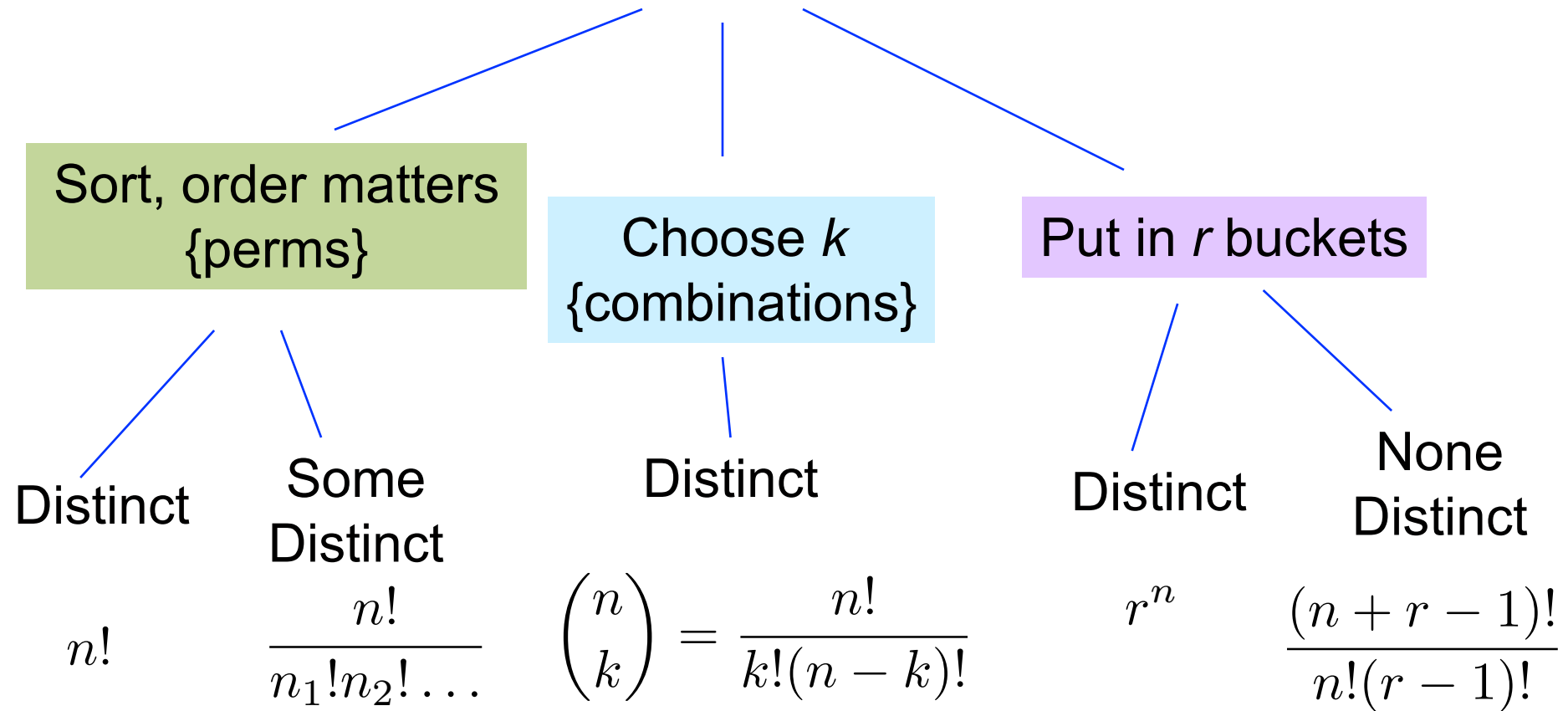
A screenshot of a web browser displaying the CS109 course website. The browser's address bar shows 'web.stanford.edu/class/cs109/'. The website header includes 'CS109' and navigation links for 'Course', 'Problem Sets', 'Lecture', 'Section', and 'Resources'. A dropdown menu is open under 'Resources', listing 'Course Reader', 'Python Review', and 'Latex Cheat Sheet'. The main content area features the course title 'CS109: Probability for Computer Scientists' and the semester 'Winter 2022'. The Stanford University logo is visible in the bottom right corner of the page.



Review

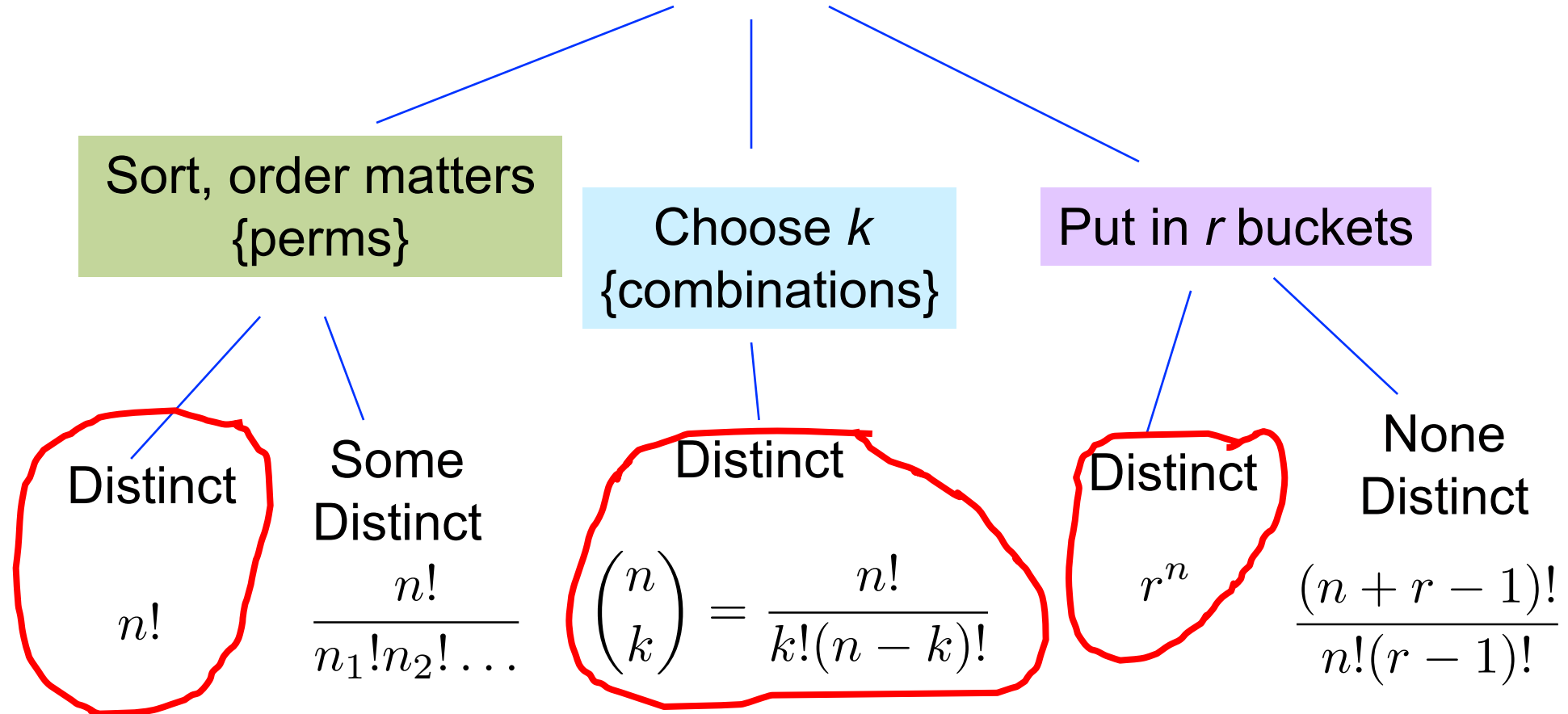
Counting Rules

Counting operations on n objects



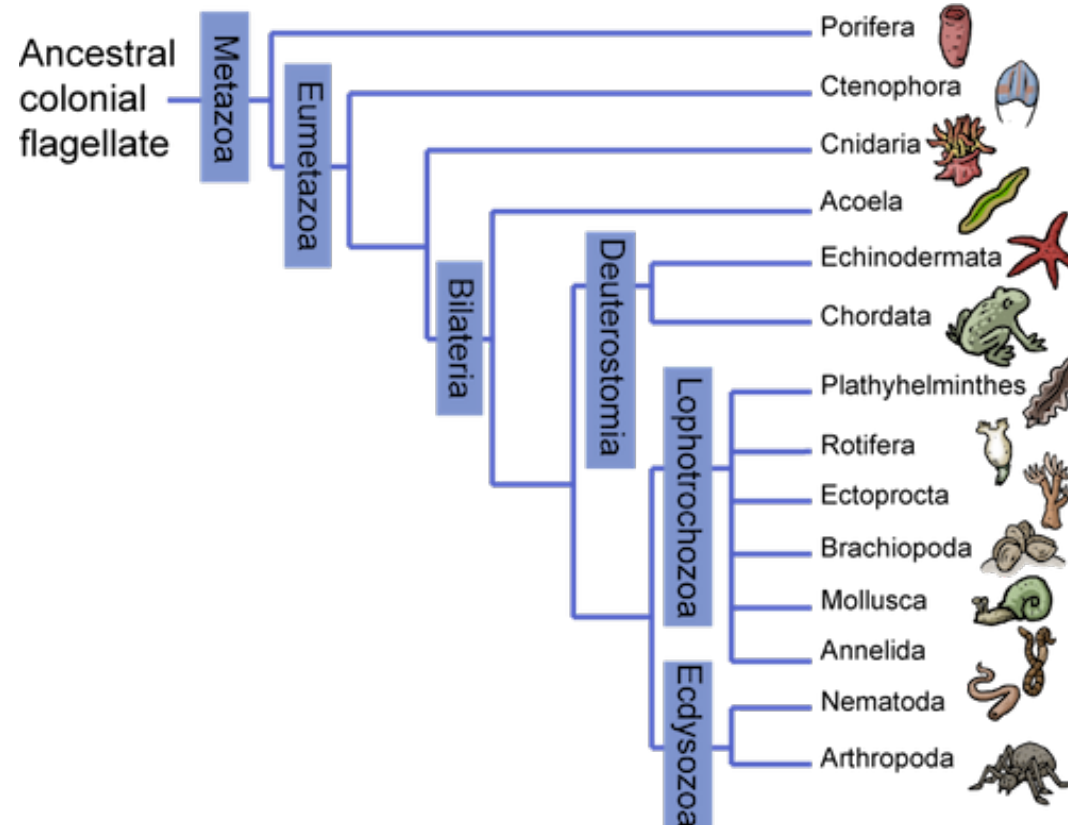
Counting Rules

Counting operations on n objects



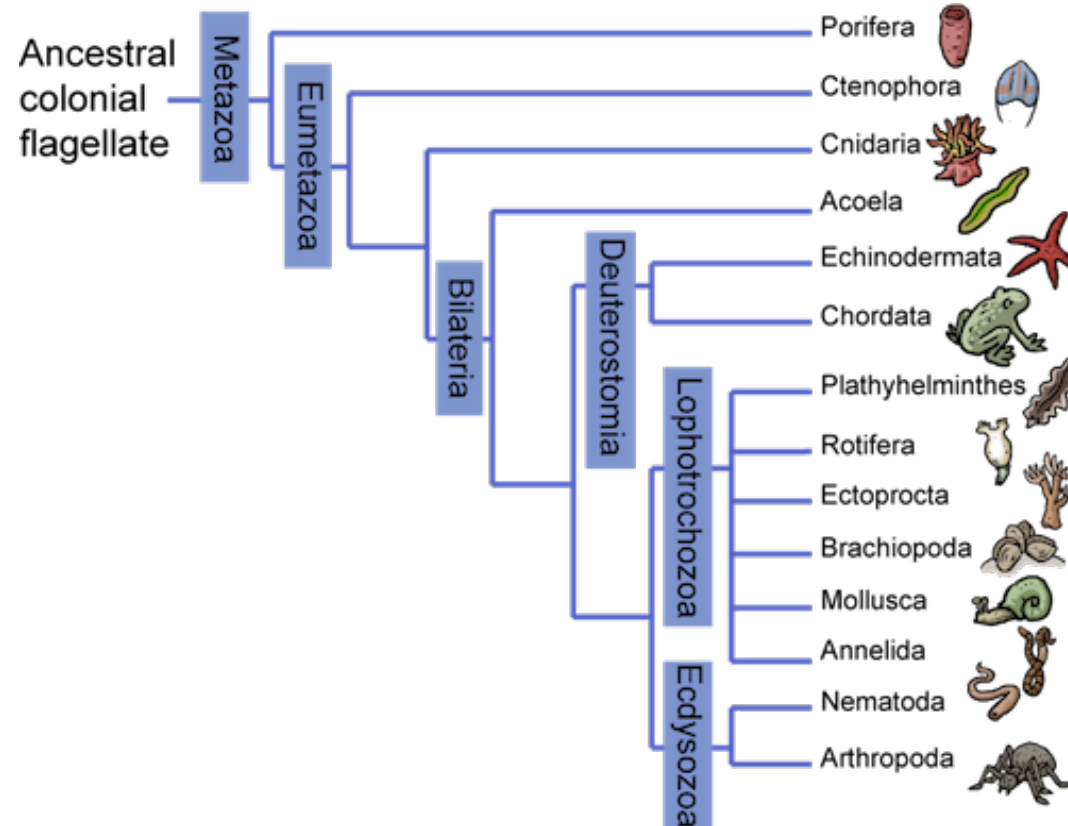
Counting Review

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



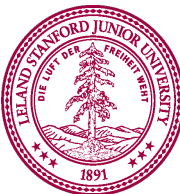
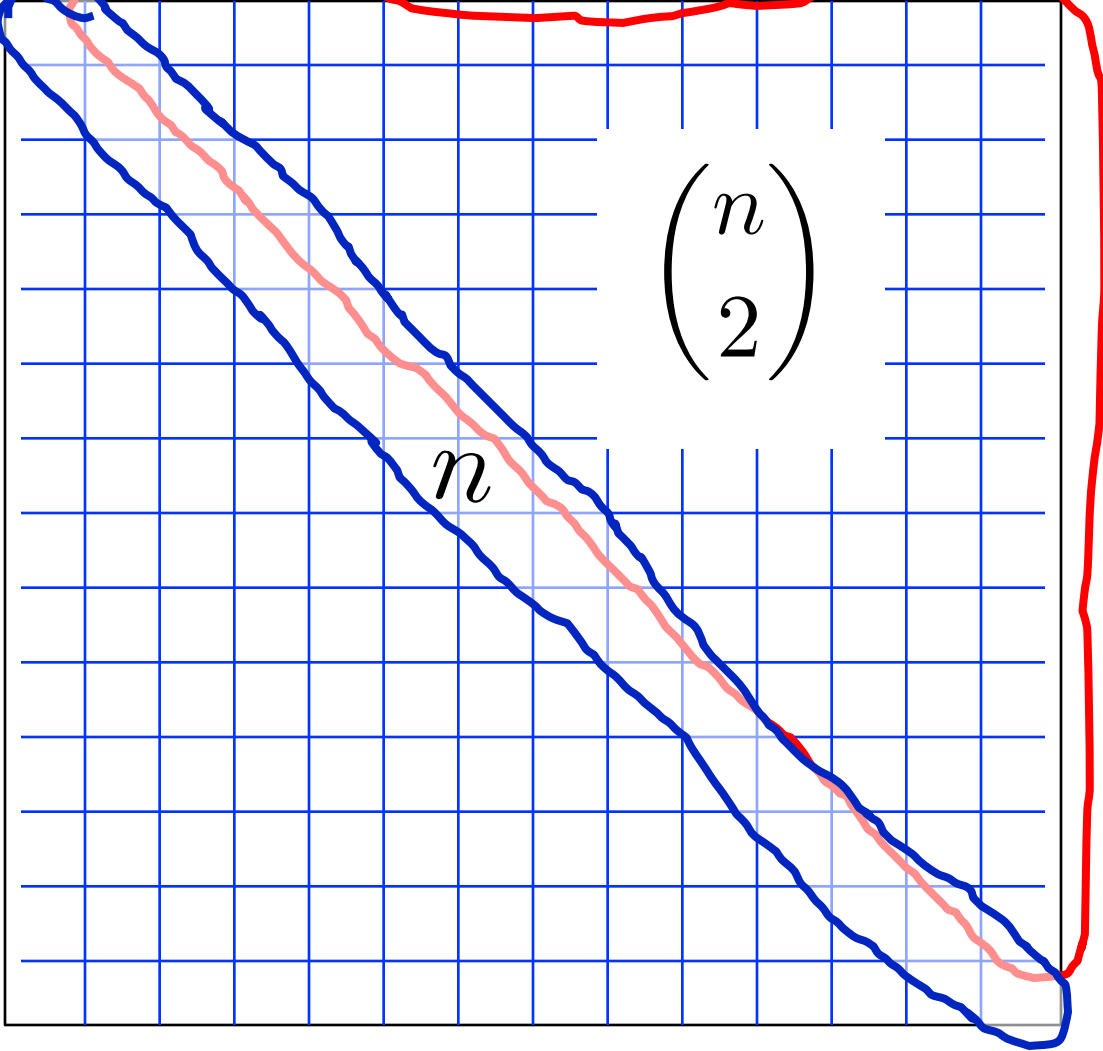
Counting Review

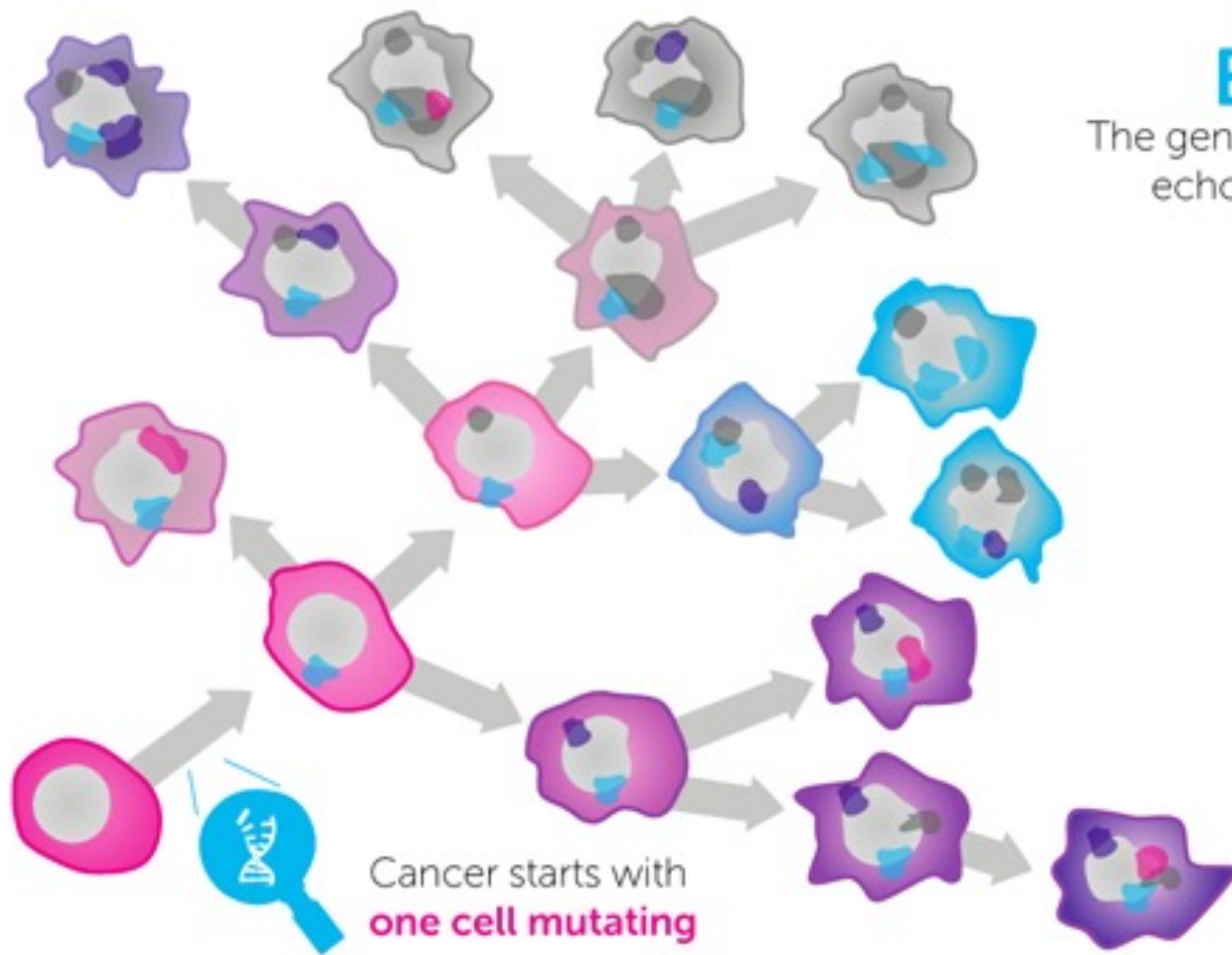
Q: There are n animals.
How many distinct pairs of animals are there?



- Porifera 
- Ctenophora 
- Cnidaria 
- Acoela 
- Echinodermata 
- Chordata 
- Platyhelminthes 
- Rotifera 
- Ectoprocta 
- Brachiopoda 
- Mollusca 
- Annelida 
- Nematoda 
- Arthropoda 

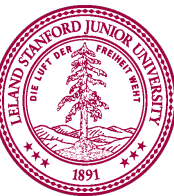
- Arthropoda 
- Nematoda 
- Annelida 
- Mollusca 
- Brachiopoda 
- Ectoprocta 
- Rotifera 
- Platyhelminthes 
- Chordata 
- Echinodermata 
- Acoela 
- Cnidaria 
- Ctenophora 
- Porifera 





BRANCHED EVOLUTION

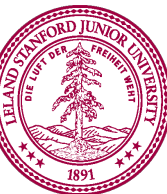
The genetic diversity in a tumour echoes Darwin's **Tree of Life**.



End Review

Sample Space

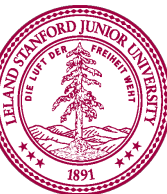
- **Sample space**, S , is set of all possible outcomes of an experiment
 - Coin flip: $S = \{\text{Head, Tails}\}$
 - Flipping two coins: $S = \{\{H, H\}, \{H, T\}, \{T, H\}, \{T, T\}\}$
 - Roll of 6-sided die: $S = \{1, 2, 3, 4, 5, 6\}$
 - # emails in a day: $S = \{x \mid x \in \mathbf{Z}, x \geq 0\}$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



Event Space

- **Event**, E , is some subset of S $\{E \subseteq S\}$
 - Coin flip is heads: $E = \{\text{Head}\}$
 - ≥ 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 ≤ 20 : $E = \{x \mid x \in \mathbf{Z}, 0 \leq x \leq 20\}$
 - Wasted day $\{\geq 5 \text{ YT hrs.}\}$: $E = \{x \mid x \in \mathbf{R}, 5 \leq x \leq 24\}$

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



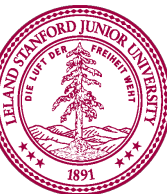
Event Space

Sample Space, S

- Coin flip
 $S = \{\text{Heads, Tails}\}$
- 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
 $S = \{x \mid x \in \mathbb{Z}, x \geq 0\}$
- TikTok hours in a day
 $S = \{x \mid x \in \mathbb{R}, 0 \leq x \leq 24\}$

Event, E

- Flip lands heads
 $E = \{\text{Heads}\}$
- ≥ 1 head on 2 coin flips
 $E = \{(H,H), (H,T), (T,H)\}$
- Roll is 3 or less:
 $E = \{1, 2, 3\}$
- Low email day (≤ 20 emails)
 $E = \{x \mid x \in \mathbb{Z}, 0 \leq x \leq 20\}$
- Wasted day (≥ 5 TT hours):
 $E = \{x \mid x \in \mathbb{R}, 5 \leq x \leq 24\}$



What is a probability?

Number between 0 and 1

A number to which we ascribe meaning

$$P(E)$$

* Our belief that an event E occurs

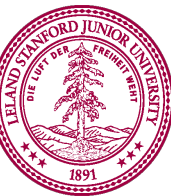


A number to which we ascribe meaning



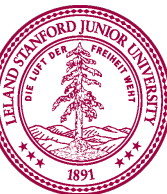
$\text{Pr}(E)$

* Our belief that an event E occurs



What is a Probability?

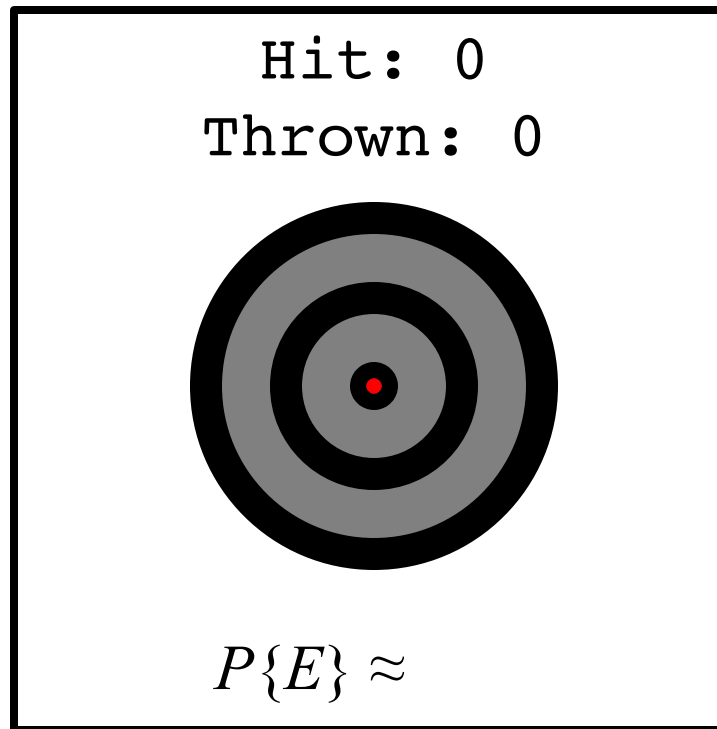
$$P(E) = \lim_{n \rightarrow \infty} \frac{n(E)}{n}$$



What is a Probability?

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

n is the number
of trials

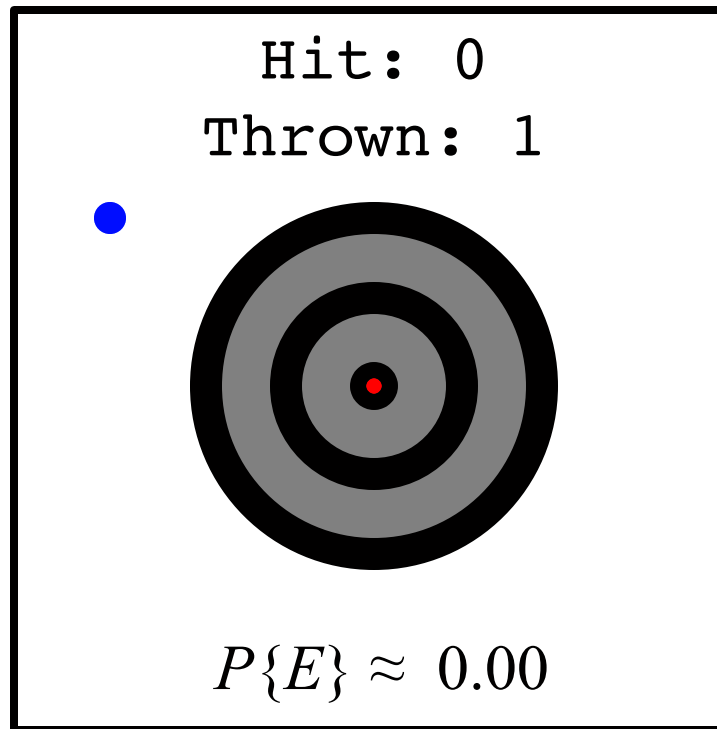


The “event” E
is that you hit
the target

What is a Probability?

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

n is the number
of trials

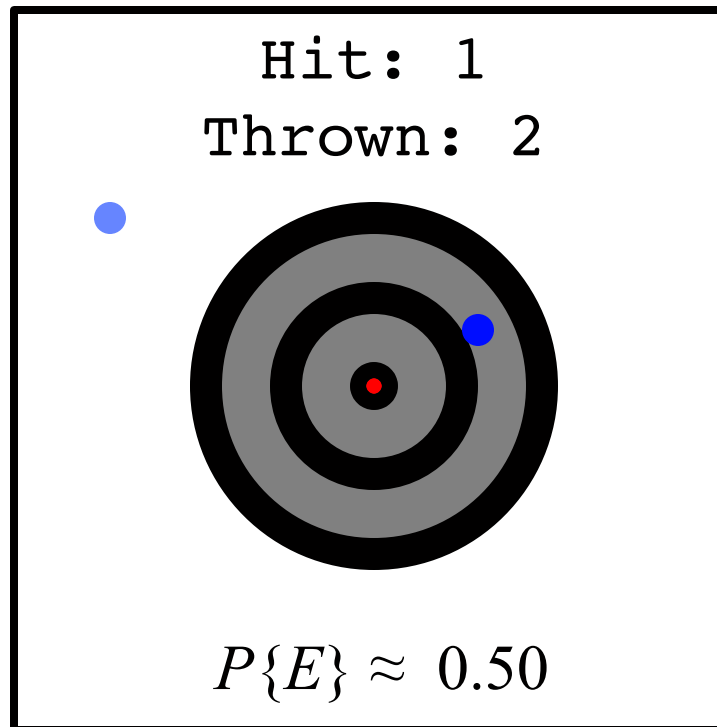


The “event” E
is that you hit
the target

What is a Probability?

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

n is the number
of trials

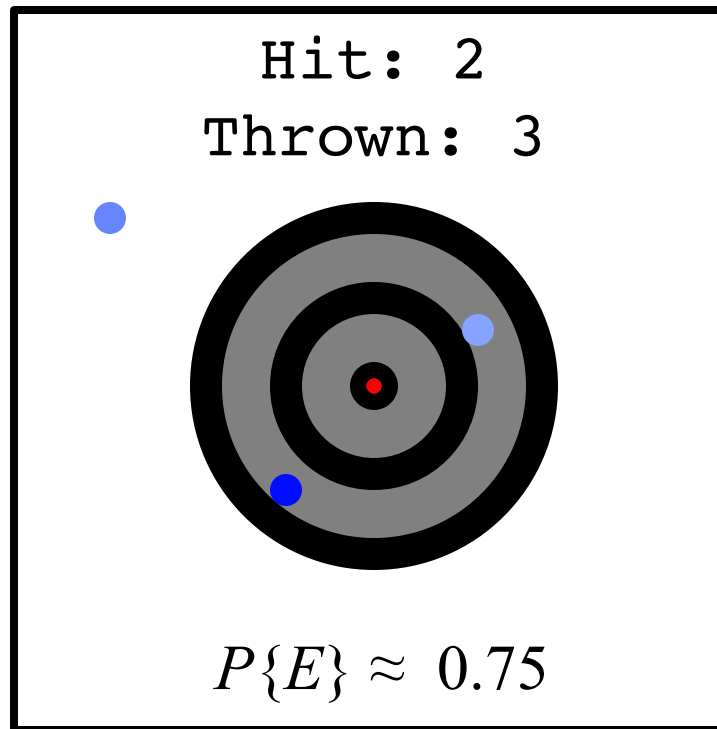


The “event” E
is that you hit
the target

What is a Probability?

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

n is the number
of trials

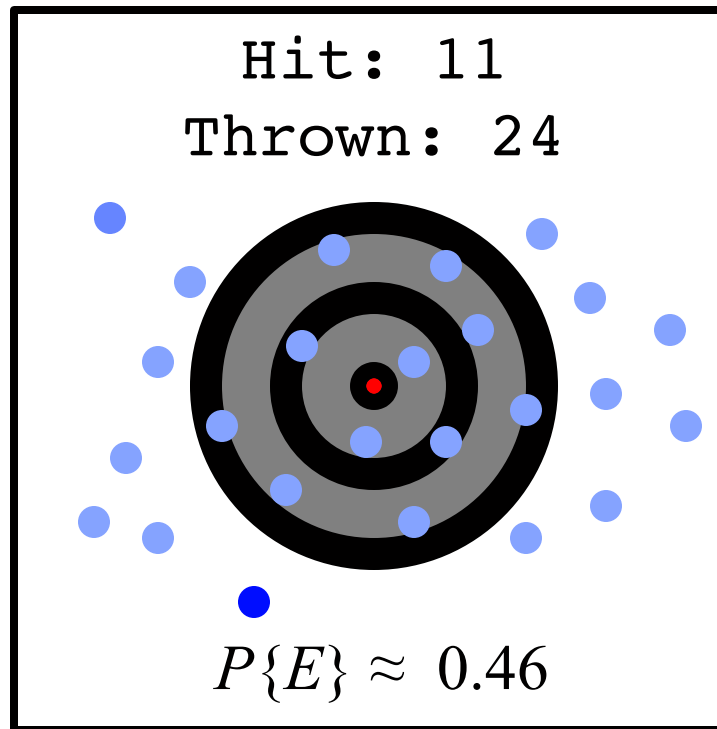


The “event” E
is that you hit
the target

What is a Probability?

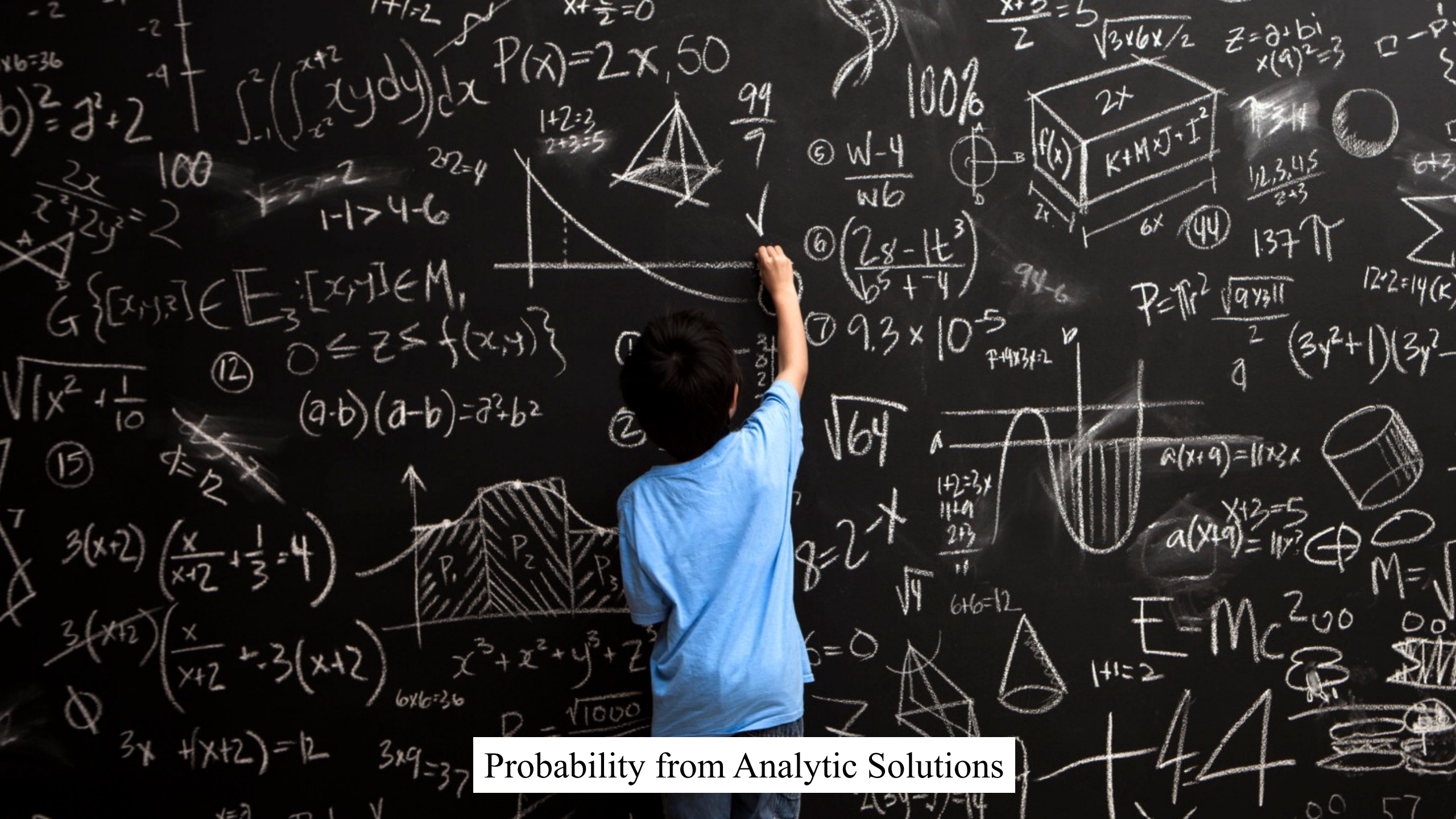
$$P(E) = \lim_{n \rightarrow \infty} \frac{n(E)}{n}$$

n is the number
of trials



The “event” E
is that you hit
the target





Probability from Analytic Solutions

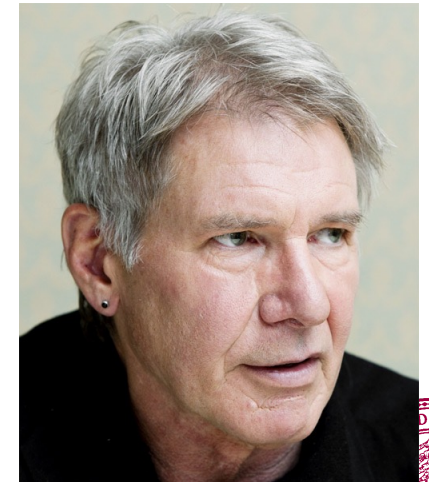
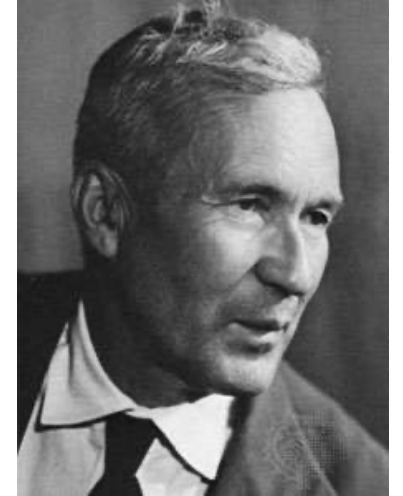
Axioms of Probability

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

- Axiom 1: $0 \leq P(E) \leq 1$
- Axiom 2: $P(S) = 1$
- Axiom 3: If events E and F are mutually exclusive:

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

Kolmogorov



Harrison Ford



Core Rules of Probability

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

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

Technically Identity 3 can be proved from the 3 axioms.



Special Case of
Analytic Probability

Equally Likely Outcomes

Equally Likely Outcomes

Some sample spaces have **equally likely outcomes**.

- Coin flip: $S = \{\text{Head}, \text{Tails}\}$
- Flipping two coins: $S = \{\{H, H\}, \{H, T\}, \{T, H\}, \{T, T\}\}$
- Roll of 6-sided die: $S = \{1, 2, 3, 4, 5, 6\}$

If we have equally likely outcomes, then $P\{\text{Each outcome}\} = \frac{1}{|S|}$

Therefore
$$P(E) = \frac{\# \text{ outcomes in } E}{\# \text{ outcomes in } S} = \frac{|E|}{|S|} \quad \{\text{by Axiom 3}\}$$

Not Everything is Equally Likely

- Play lottery.
 - What is $P\{\text{Win}\}$?
-

- $S = \{\text{Lose}, \text{Win}\}$
- $E = \{\text{Win}\}$
- $P\{\text{Win}\} = |E|/|S| = 1/2 = 50\%$



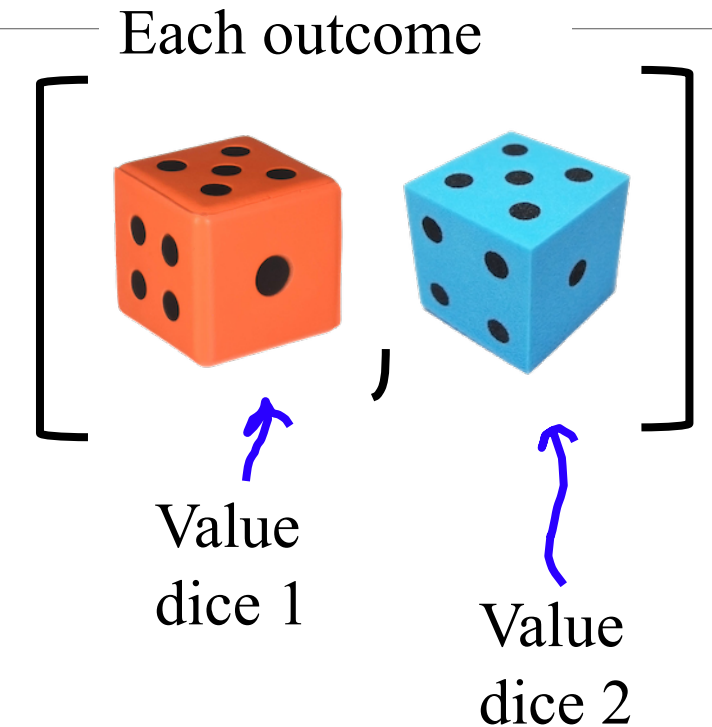
Sum of Two Die = 7?

Roll two 6-sided dice. What is $P[\text{sum} = 7]$?

$S = \{$

[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]

$\}$



Sum of Two Die = 7?

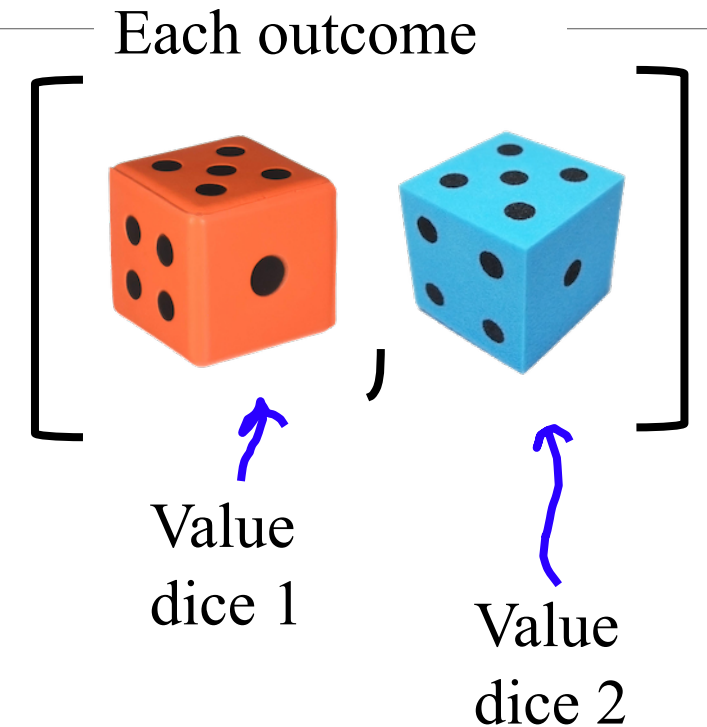
Roll two 6-sided dice. What is probability the sum = 7?

Let E be the event that the sum is 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 = *in blue*

$$P(E) = \frac{|E|}{|S|} = \frac{6}{36} = 0.1\overline{6}$$



Is it correct?

$$P(E) = \frac{|E|}{|S|} = \frac{6}{36} = 0.1\overline{6}$$

Sum of Two Die = 7?

```
1 import random
2
3 # getting close to infinity
4 n_trials = 10000000
5
6 def main():
7     n_events = 0
8     for i in range(n_trials):
9         outcome = roll_dice()
10        if is_event(outcome):
11            n_events += 1
12    pr_e = n_events / n_trials
13    print(f'after {n_trials} trials')
14    print('P(E) = ', pr_e)
15
16 def is_event(outcome):
17     # was it a 6?
18     return outcome == 6
19
20 def roll_dice():
21     # give me a random dice roll
22     # alternatively random.randint(1, 7)
23     return random.choice([1,2,3,4,5,6])
24
```

$$P(E) = \frac{|E|}{|S|} = \frac{6}{36} = 0.1\overline{6}$$

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

```
piech@Chriss-MBP-2 3 % python dice_soln.py
after 10000000 trials
P(E) = 0.1666913
```

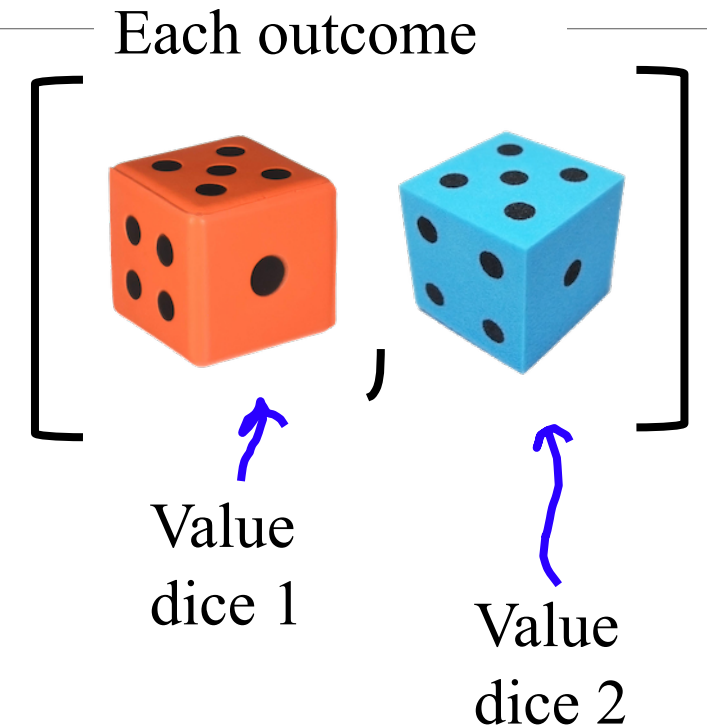
Sum of Two Die = 2?

Roll two 6-sided dice. What is probability the sum = 7?

Let E be the event that the sum is 7

$$S = \{ \begin{array}{cccccc} [1,1] & [1,2] & [1,3] & [1,4] & [1,5] & [1,6] \\ [2,1] & [2,2] & [2,3] & [2,4] & [2,5] & [2,6] \\ [3,1] & [3,2] & [3,3] & [3,4] & [3,5] & [3,6] \\ [4,1] & [4,2] & [4,3] & [4,4] & [4,5] & [4,6] \\ [5,1] & [5,2] & [5,3] & [5,4] & [5,5] & [5,6] \\ [6,1] & [6,2] & [6,3] & [6,4] & [6,5] & [6,6] \end{array} \}$$

E =



Sum of Two Die = 2?

Roll two 6-sided dice. What is probability the sum = 2?

Let E be the event that the sum is 2

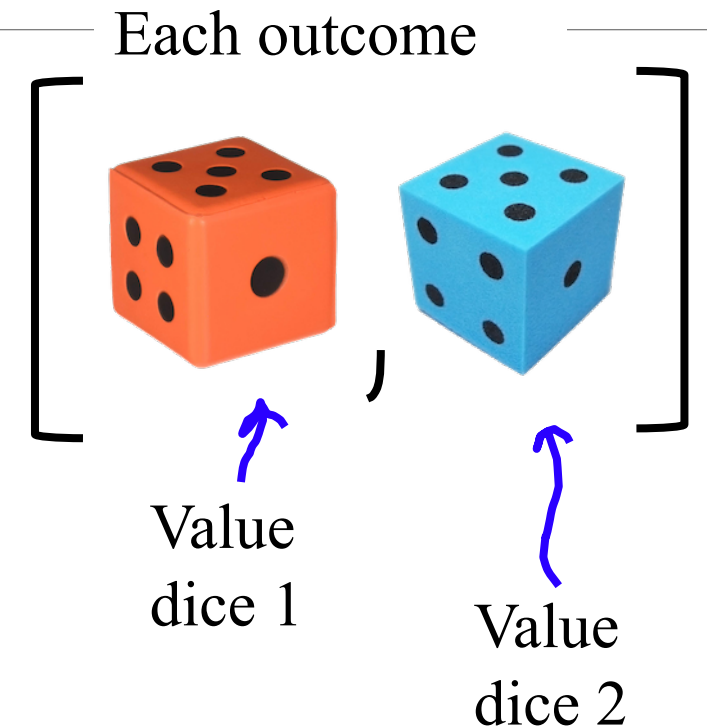
$S = \{$

[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]

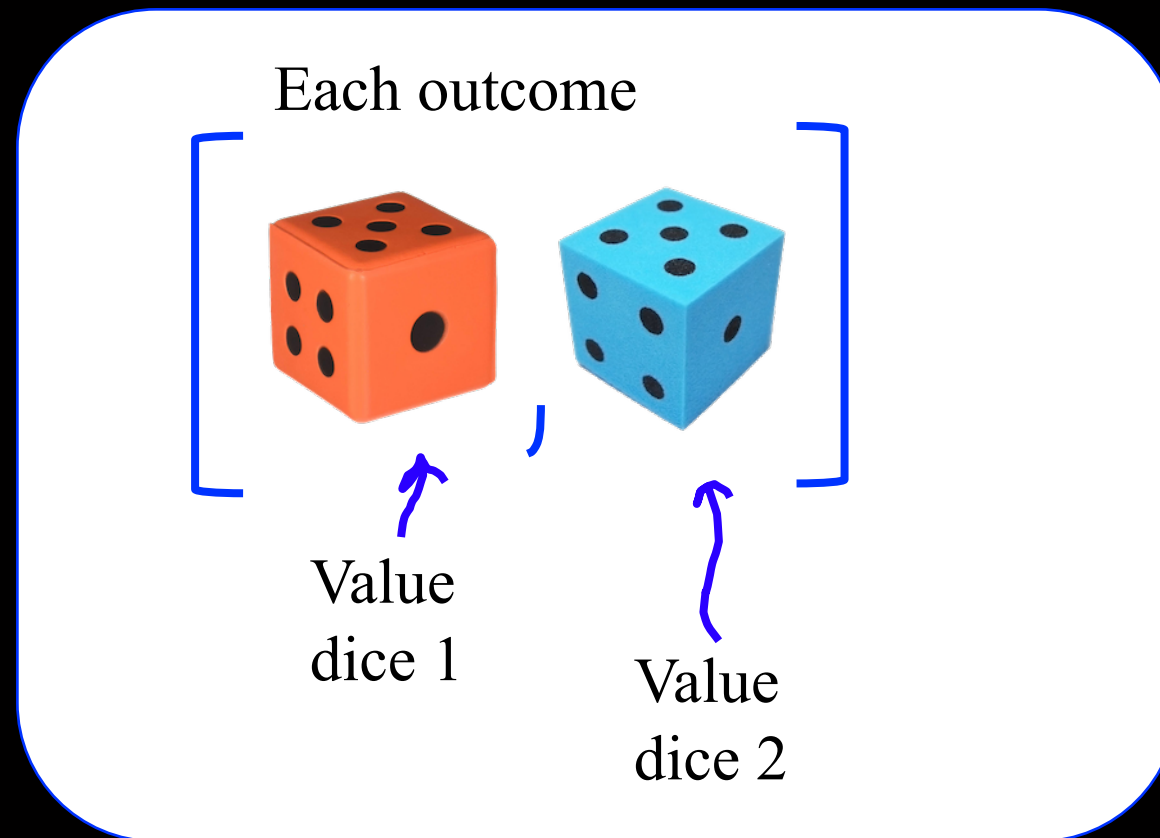
$\}$

E = *in red*

$$P(E) = \frac{|E|}{|S|} = \frac{1}{36} = 0.02\bar{7}$$



Other ways to make a Sample Space?



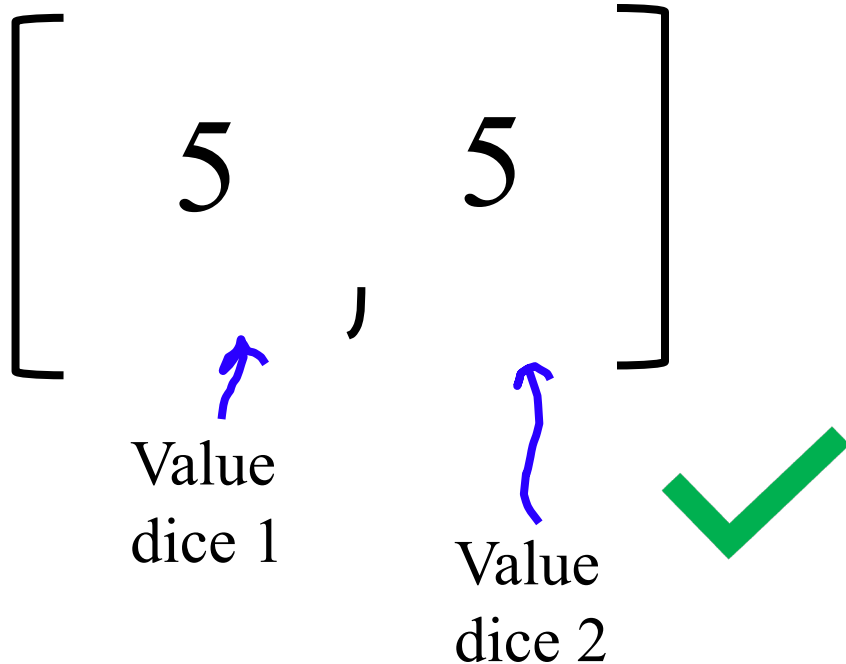
Sum of Two Die: Three options for the sample space

Value
dice 1

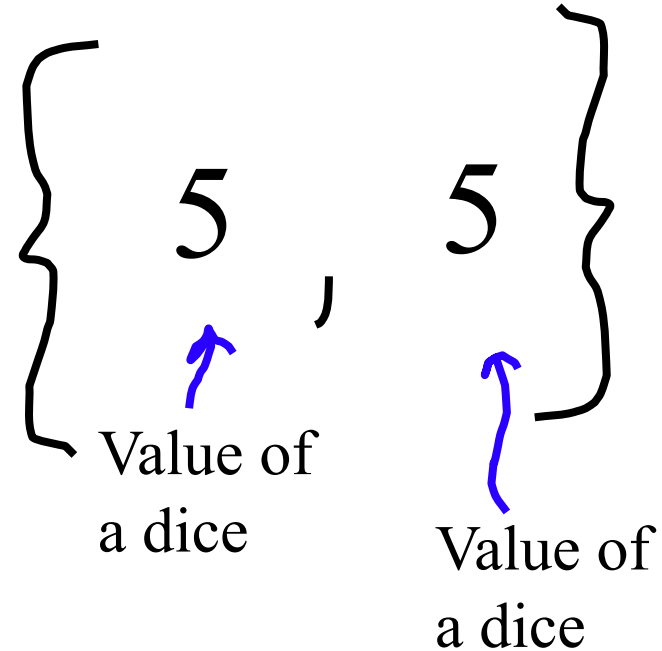


Value
dice 2

Think of the die as **distinct**



Think of the die as **indistinct**



Just look at the sum

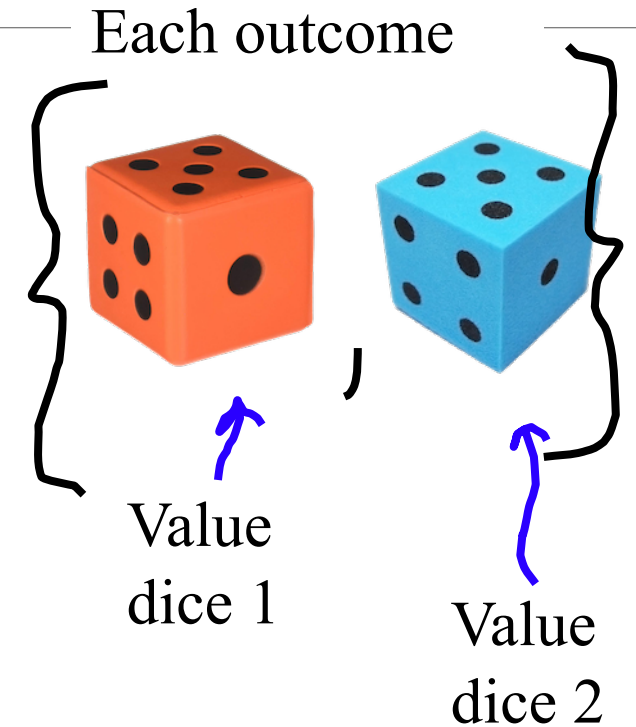
10

Sum of Two Die = 7? Bug: Die are Indistinct

Roll two 6-sided dice. What is probability the sum = 7?

Let E be the event that the sum is 7

S = { 1 2 3 4 5 6
 7 8 9 10 11 12 }



E = *in red*

$$P(E) = \frac{|E|}{|S|} = \frac{1}{12} = 0.08\bar{3}?$$

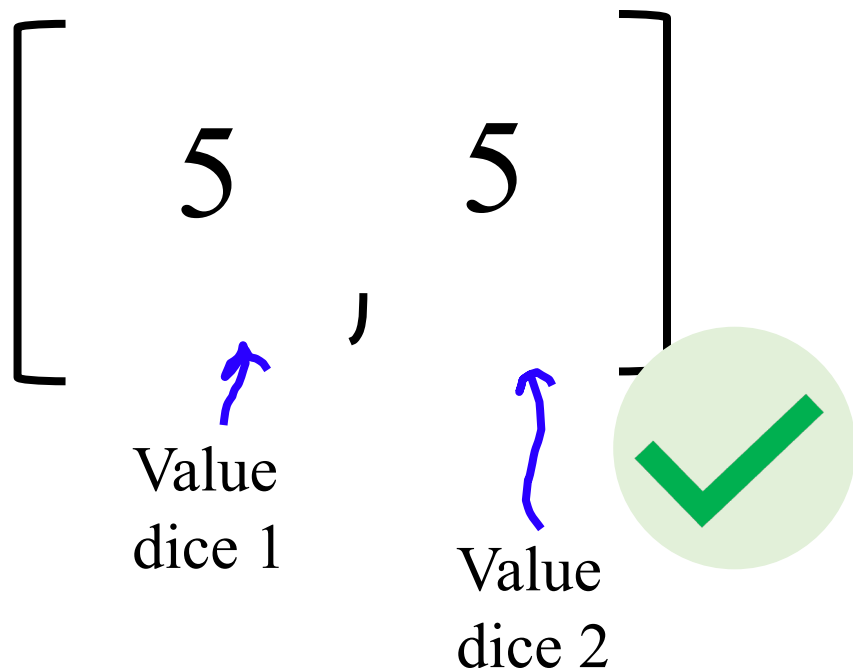
Sum of Two Die: Three options for the sample space

Value
dice 1

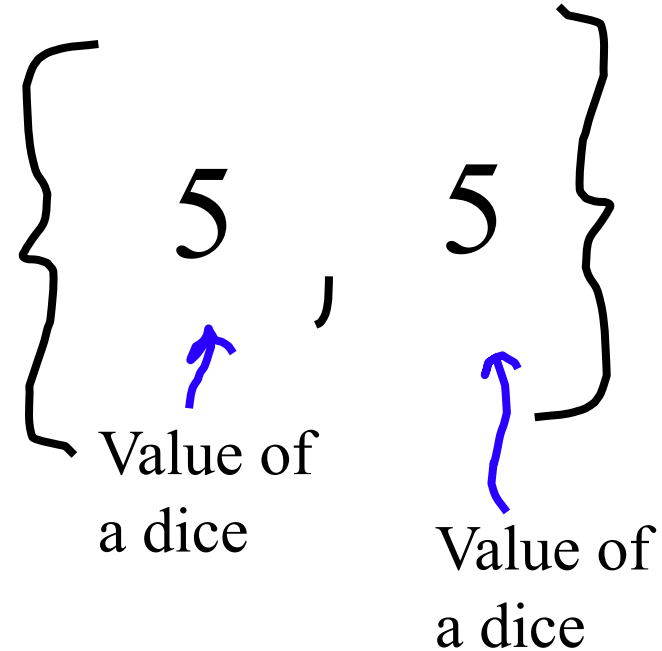


Value
dice 2

Think of the die as **distinct**



Think of the die as **indistinct**



Just look at the sum

10



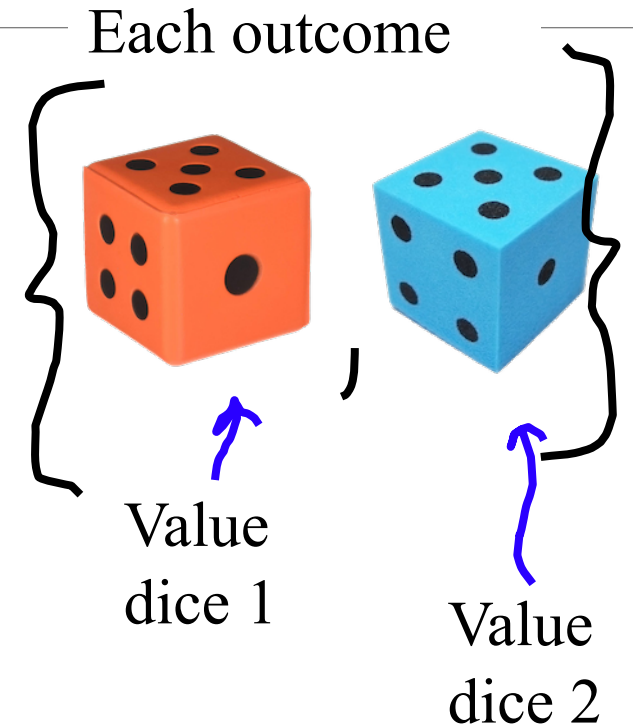
Sum of Two Die = 7? Bug: Die are Indistinct

Roll two 6-sided dice. What is $P(\text{sum} = 7)$?

$S = \{$

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

$\}$



$E =$ *in blue*

$$P(E) = \frac{|E|}{|S|} = \frac{3}{20} = 0.15?$$

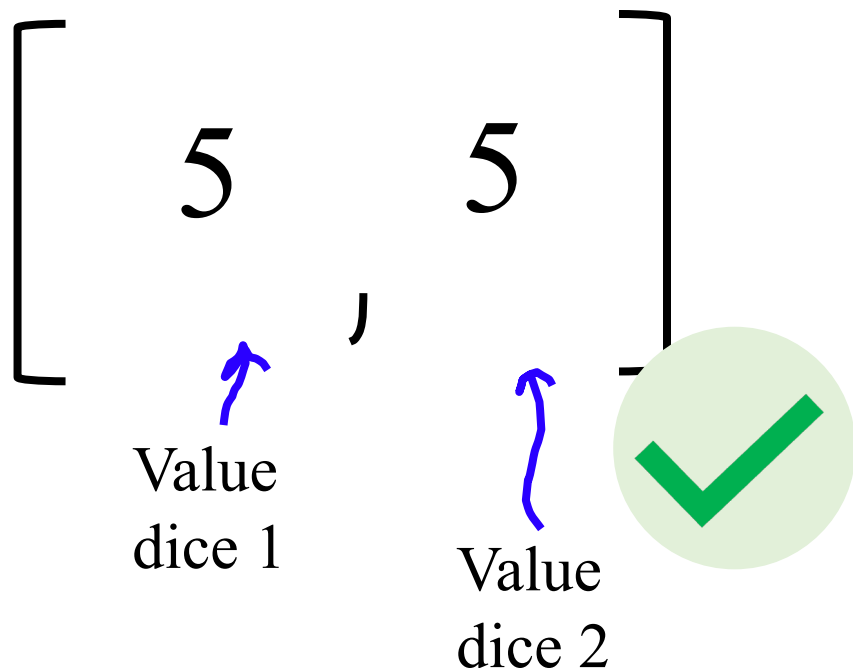
Sum of Two Die: Three options for the sample space

Value
dice 1

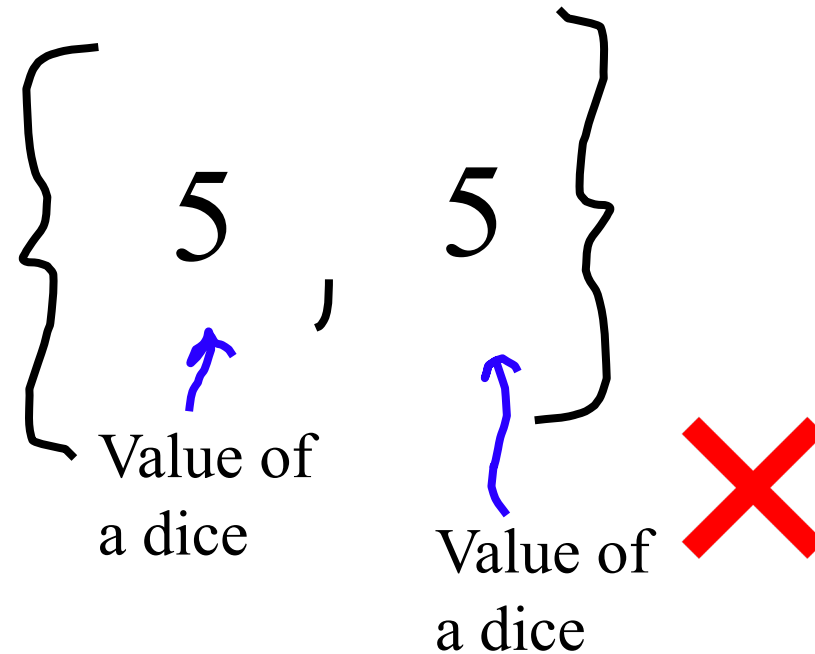


Value
dice 2

Think of the die as **distinct**



Think of the die as **indistinct**



Just look at the sum

10



pigs and cows

- 4 cows and 3 pigs in a toy box. 3 drawn.
 - What is $P(1 \text{ cow and } 2 \text{ pigs drawn})$?

Equally likely sample space? Thought experiment



4 cows

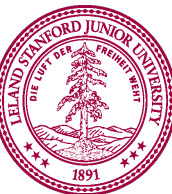


3 pig

The Choice of Sample Space is Yours!

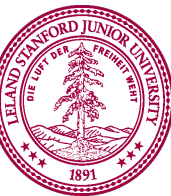
	Distinct	Indistinct
Unordered	$\{C_1, P_2, P_3\}$ $\{C_1, C_2, C_3\}$	$\{2 \text{ cows, } 1 \text{ pig}\}$ $\{3 \text{ cows}\}$
Ordered	$[C_1, P_2, P_3]$ $[C_1, C_2, C_3]$	$[\text{cow, pig, cow}]$ $[\text{cow, cow, cow}]$

Which choice will lead to equally likely outcomes?



pigs and cows

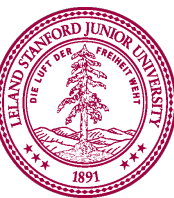
- 4 pigs and 3 cows in a Bag. 3 drawn.
 - What is $P(1 \text{ pig and } 2 \text{ cows drawn})$?
- Ordered and Distinct:
 - Pick 3 ordered items: $|S| = 7 * 6 * 5 = 210$
 - Pick pig as either 1st, 2nd, or 3rd item:
 $|E| = \{4 * 3 * 2\} + \{3 * 4 * 2\} + \{3 * 2 * 4\} = 72$
 - $P(1 \text{ pig, } 2 \text{ cows}) = 72/210 = 12/35$
- Unordered:
 - $|S| = \binom{7}{3} = 35$
 - $|E| = \binom{4}{1} \binom{3}{2} = 12$
 - $P(1 \text{ pig, } 2 \text{ cows}) = 12/35$





Make indistinct items
distinct to get equally
likely sample space
outcomes

*You will need to use this “trick” with high probability



Straight Poker Hand

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



Straight Poker Hand

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

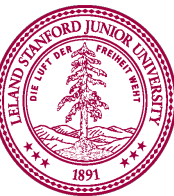
$$|S| = \binom{52}{5}$$

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

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

What is an example
of one outcome?

Is each outcome
equally likely?



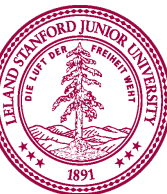
Straight Poker Hand

- 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 $P(\text{straight, but not straight flush})$?

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

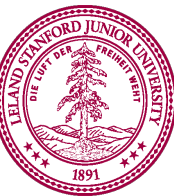
$$|E| = 10 \binom{4}{1}^5 - 10 \binom{4}{1}$$

$$P(\text{straight}) = \frac{|E|}{|S|} = \frac{10 \binom{4}{1}^5 - 10 \binom{4}{1}}{\binom{52}{5}} \approx 0.00392$$





When approaching an
“**equally likely probability**”
problem, start by defining
sample spaces and
event spaces.



Chip Defect Detection

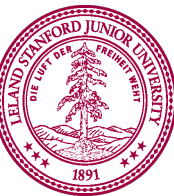
- n chips manufactured, 1 of which is defective.
- k chips randomly selected from n for testing.
 - What is $P\{\text{defective chip is in } k \text{ selected chips}\}$?

- $|S| = \binom{n}{k}$

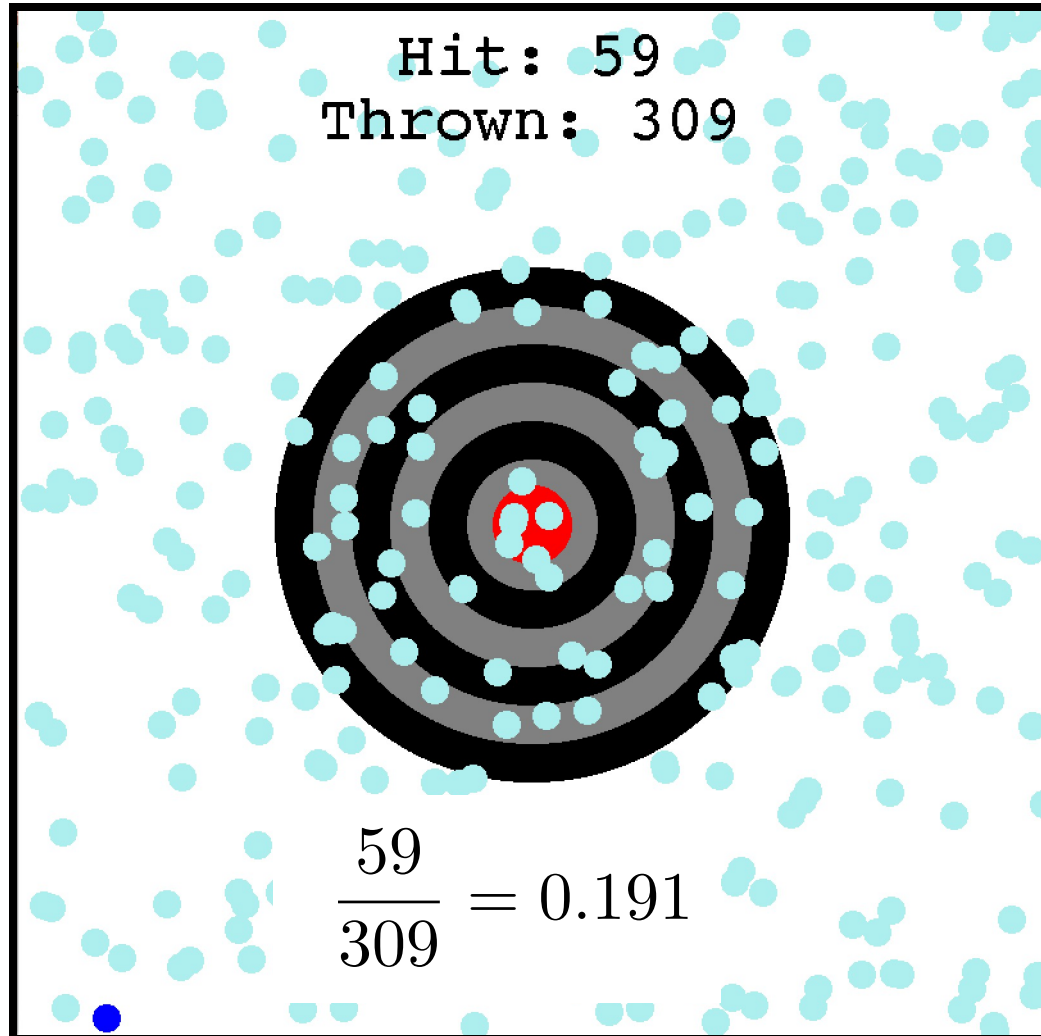
- $|E| = \binom{1}{1} \binom{n-1}{k-1}$

- $P(\text{defective chip is in } k \text{ selected chips})$

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



Target Revisited



Screen size = 800×800

Radius of target = 200

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

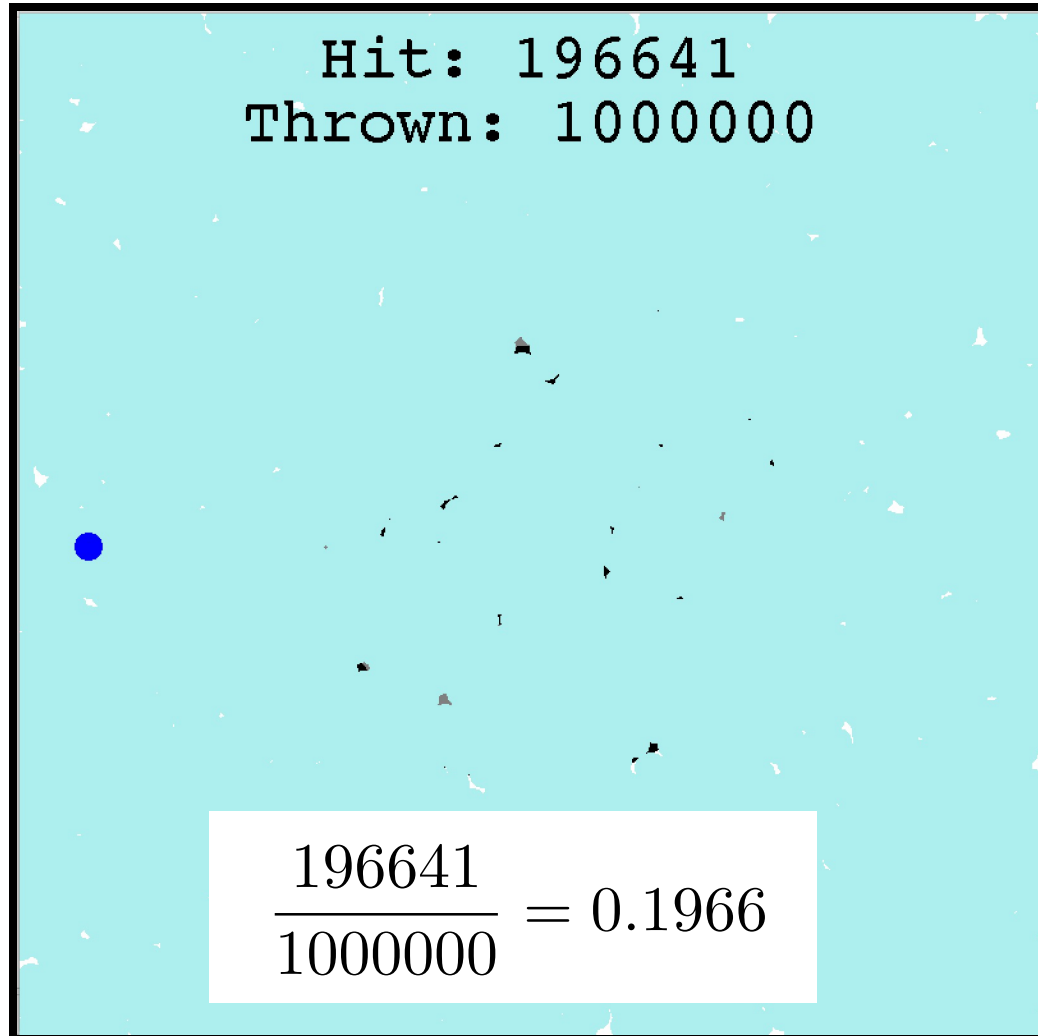
What is the probability of hitting the target?

$$|S| = 800^2$$

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

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

Target Revisited



Screen size = 800×800

Radius of target = 200

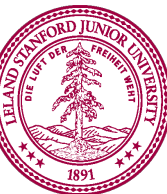
The dart is equally likely to land anywhere on the screen.

What is the probability of hitting the target?

$$|S| = 800^2$$

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

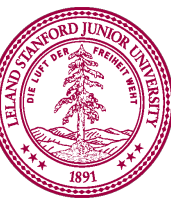
$$p(E) = \frac{\pi \cdot 200^2}{800^2} \approx 0.1963$$



Let it find you.

SERENDIPITY

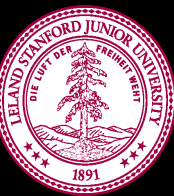
the effect by which one accidentally stumbles upon something truly wonderful, especially while looking for something entirely unrelated.





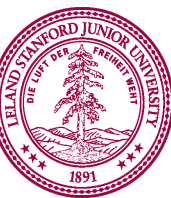
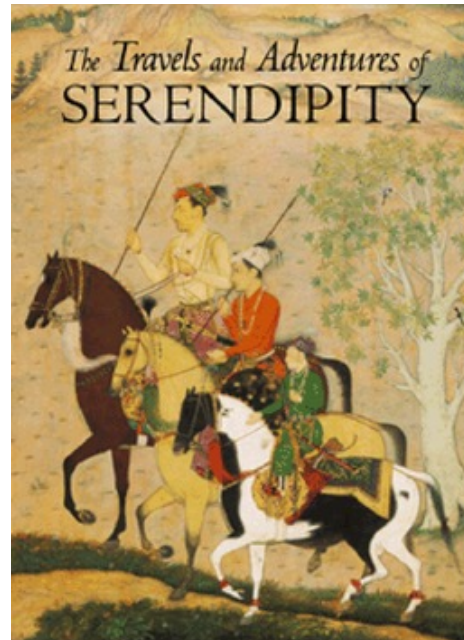
WHEN YOU MEET YOUR BEST FRIEND

Somewhere you didn't expect to.



Serendipity

- 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 $P(E^C)$.

Back to Axiom 3



Axioms of Probability

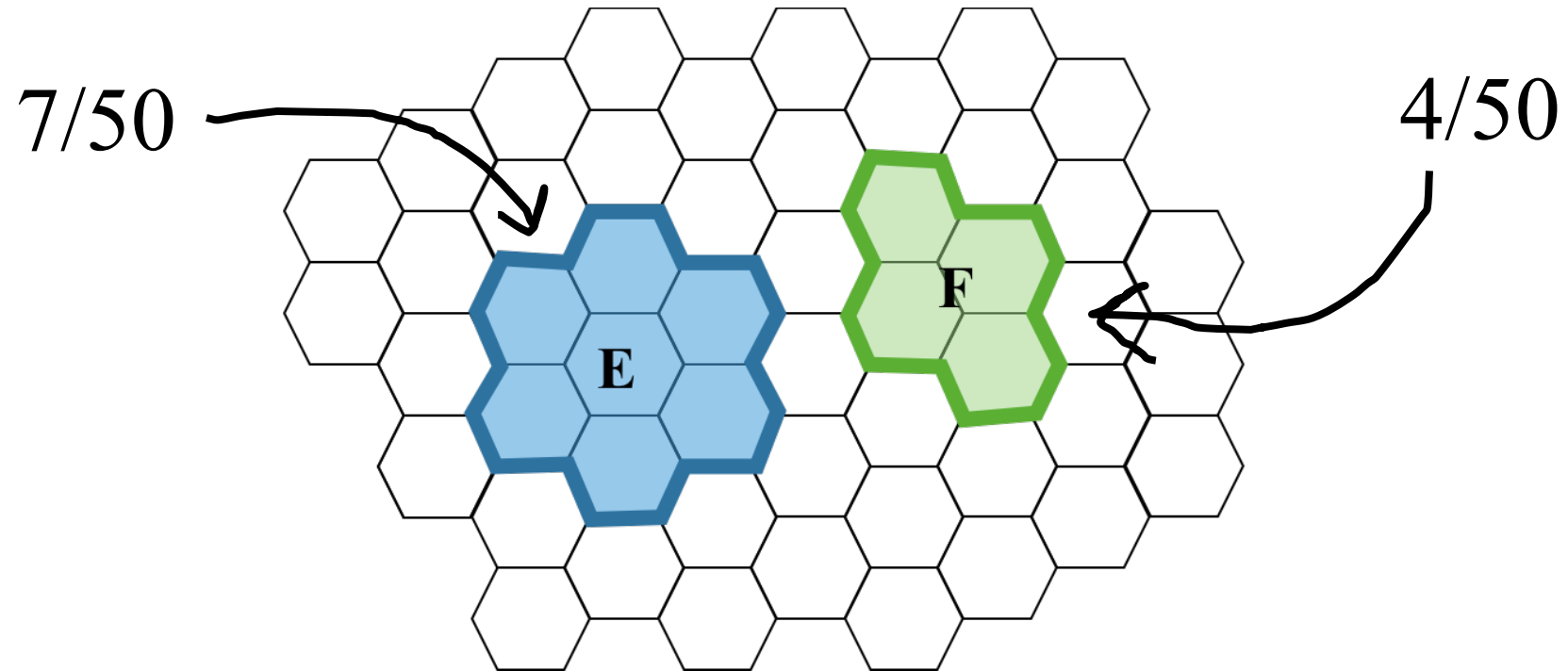
Recall: S = all possible outcomes. E = the event.

- Axiom 1: $0 \leq P(E) \leq 1$
- Axiom 2: $P(S) = 1$
- Axiom 3: If events E and F are mutually exclusive:

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

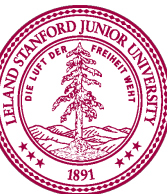


Mutually Exclusive Events

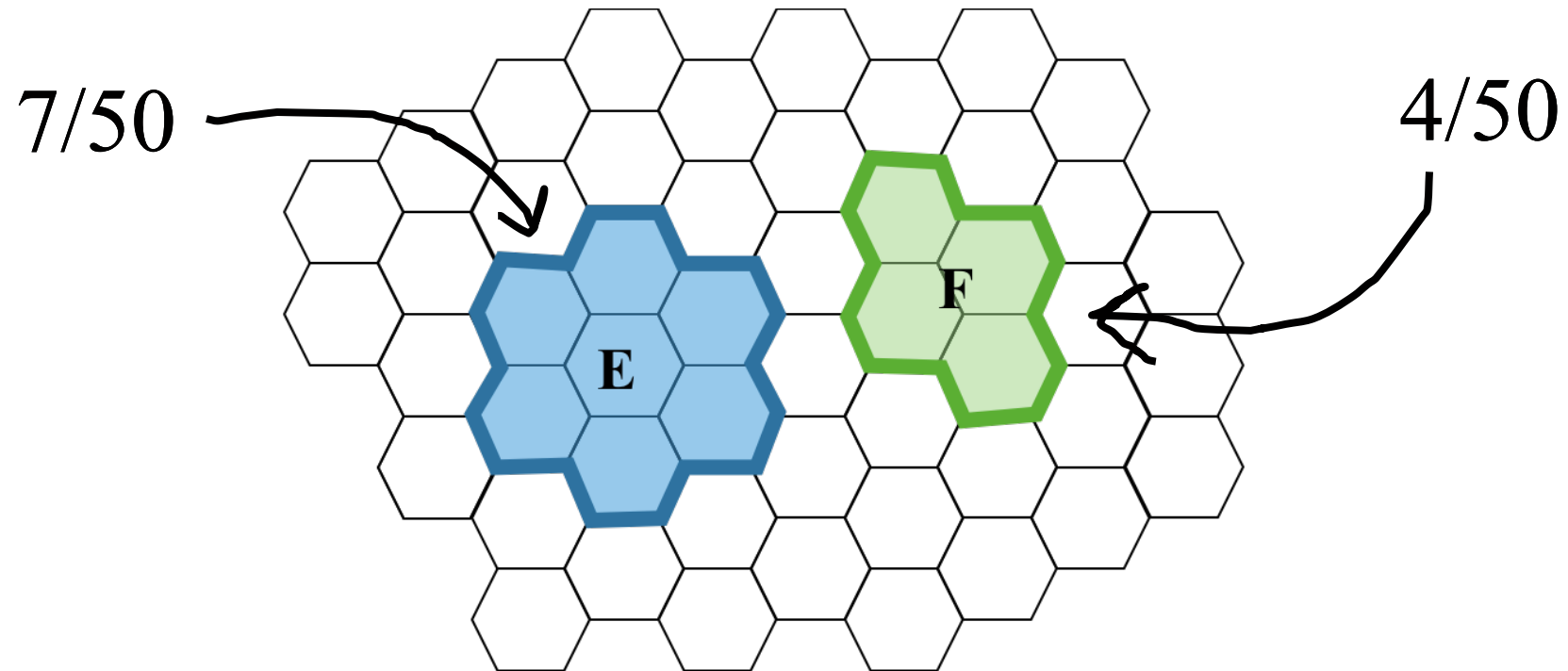


If events are mutually exclusive, probability of OR is simple:

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



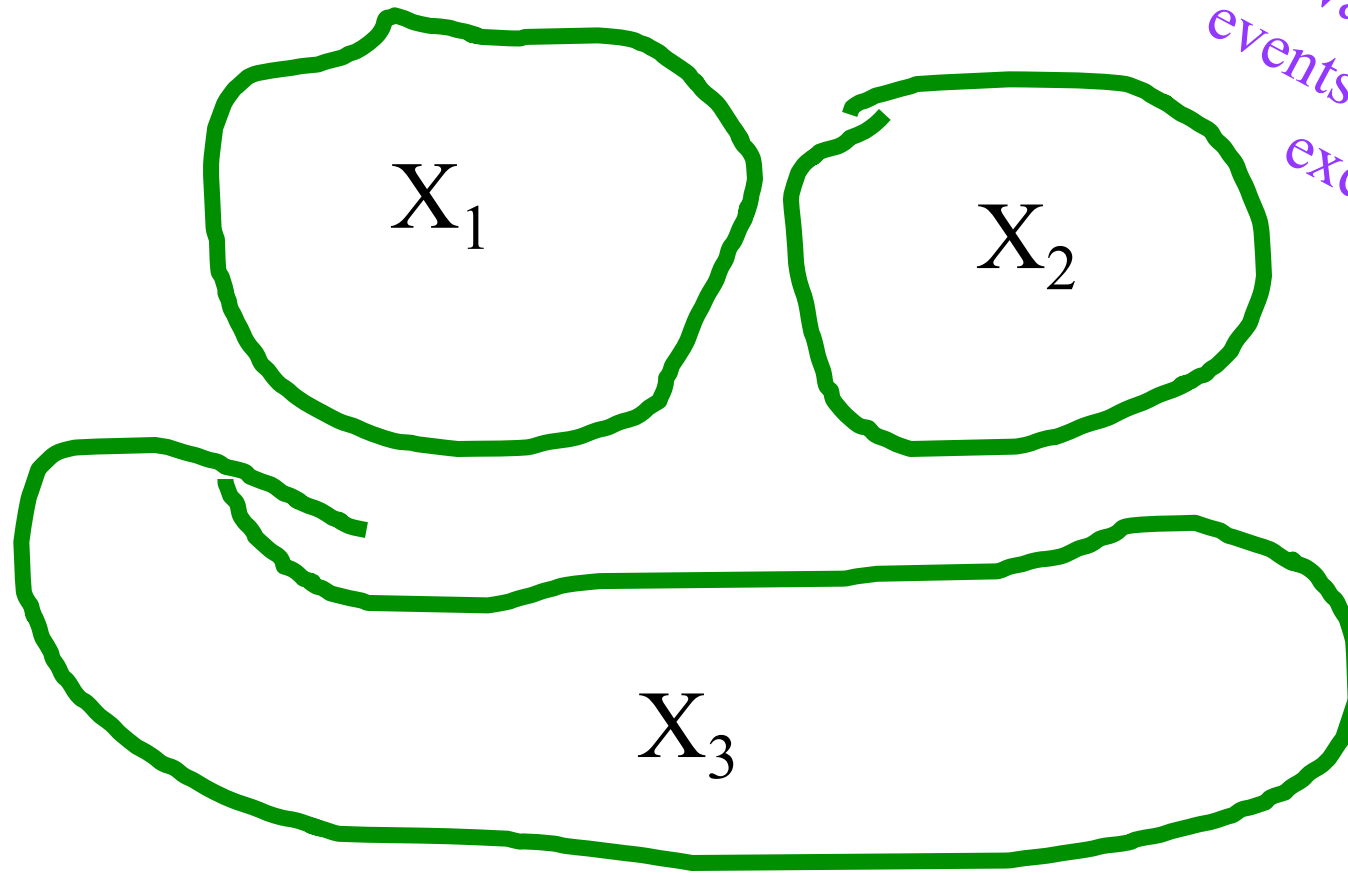
Mutually Exclusive Events



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

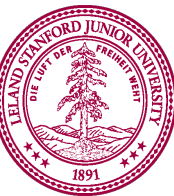
$$P(E \cup F) = \frac{7}{50} + \frac{4}{50} = \frac{11}{50}$$

Probability of "or"



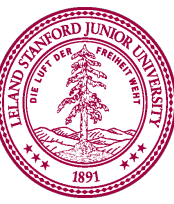
Wahoo! All my
events are mutually
exclusive

$$P(X_1 \cup X_2 \cup \cdots \cup X_n) = \sum_{i=1}^n P(X_i)$$





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

$$P(S) = P(E) + P(E^c)$$

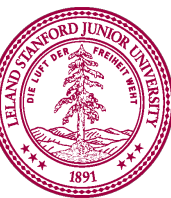
Since everything must either be in E or E^c

$$1 = P(E) + P(E^c)$$

Axiom 2

$$P(E^c) = 1 - P(E)$$

Rearrange





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

Trailing the Dovetail Shuffle to Its Lair

TRAILING THE DOVETAIL SHUFFLE TO ITS LAIR

BY DAVE BAYER¹ AND PERSI DIACONIS²

Columbia University and Harvard University

We analyze the most commonly used method for shuffling cards. The main result is a simple expression for the chance of any arrangement after any number of shuffles. This is used to give sharp bounds on the approach to randomness: $\frac{3}{2} \log_2 n + \theta$ shuffles are necessary and sufficient to mix up n cards.

Key ingredients are the analysis of a card trick and the determination of the idempotents of a natural commutative subalgebra in the symmetric group algebra.

1. Introduction. The dovetail, or riffle shuffle is the most commonly used method of shuffling cards. Roughly, a deck of cards is cut about in half and then the two halves are riffled together. Figure 1 gives an example of a riffle shuffle for a deck of 13 cards.

A mathematically precise model of shuffling was introduced by Gilbert and Shannon [see Gilbert (1955)] and independently by Reeds (1981). A deck of n cards is cut into two portions according to a binomial distribution; thus, the chance that k cards are cut off is $\binom{n}{k}/2^n$ for $0 \leq k \leq n$. The two packets are then riffled together in such a way that cards drop from the left or right heaps with probability proportional to the number of cards in each heap. Thus, if there are A and B cards remaining in the left and right heaps, then the chance that the next card will drop from the left heap is $A/(A+B)$. Such shuffles are easily described backwards: Each card has an equal and independent chance of being pulled back into the left or right heap. An inverse riffle shuffle is illustrated in Figure 2.

Experiments reported in Diaconis (1988) show that the Gilbert–Shannon–Reeds (GSR) model is a good description of the way real people shuffle real cards. It is natural to ask how many times a deck must be shuffled to mix it up. In Section 3 we prove:

THEOREM 1. *If n cards are shuffled m times, then the chance that the deck is in arrangement π is $\binom{2^m + n - r}{n} / 2^{mn}$, where r is the number of rising sequences in π .*

Rising sequences are defined and illustrated in Section 2 through the analysis of a card trick. Section 3 develops several equivalent interpretations of

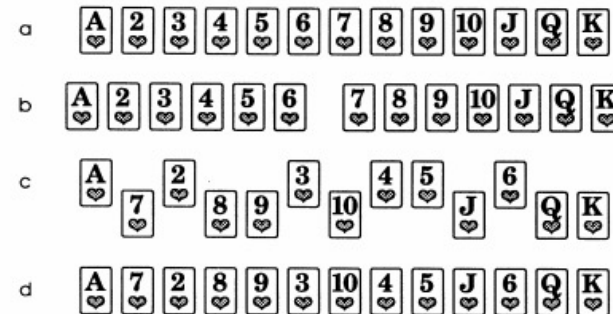


FIG. 1. A riffle shuffle. (a) We begin with an ordered deck. (b) The deck is divided into two packets of similar size. (c) The two packets are riffled together. (d) The two packets can still be identified in the shuffled deck as two distinct “rising sequences” of face values.

the GSR distribution for riffle shuffles, including a geometric description as the motion of n points dropped at random into the unit interval under the baker’s transformation $x \rightarrow 2x \pmod{1}$. This leads to a proof of Theorem 1.

Section 3 also relates shuffling to some developments in algebra. A permutation π has a descent at i if $\pi(i) > \pi(i+1)$. A permutation π has r rising sequences if and only if π^{-1} has $r-1$ descents. Let

$$A_k = \sum_{\pi \text{ has } k \text{ descents}} \pi$$

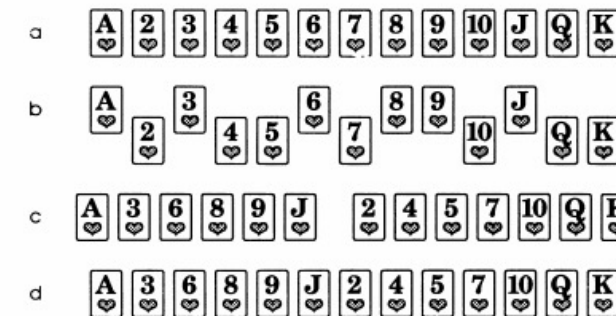


FIG. 2. An inverse riffle shuffle. (a) We begin with a sorted deck. (b) Each card is moved one way or the other uniformly at random, to “pull apart” a riffle shuffle and retrieve two packets. (c) The two packets are placed in sequence. (d) The two packets can still be identified in the shuffled deck; they are separated by a “descent” in the face values. This shuffle is inverse to the shuffle diagrammed in Figure 1.

Received January 1990; revised May 1991.

¹Partially supported by the Alfred P. Sloan Foundation, by ONR contract N00014-87-K0214 and by NSF Grant DMS-90-06116.

²Partially supported by NSF Grant DMS-89-05874.

AMS 1980 subject classifications. 20B30, 60B15, 60C05, 60F99.

Key words and phrases. Card shuffling, symmetric group algebra, total variation distance.



Probability of "or"

- 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 $n = 10^{20}$,
 - $P\{\text{deck matching yours}\} < 0.0000000001$

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

