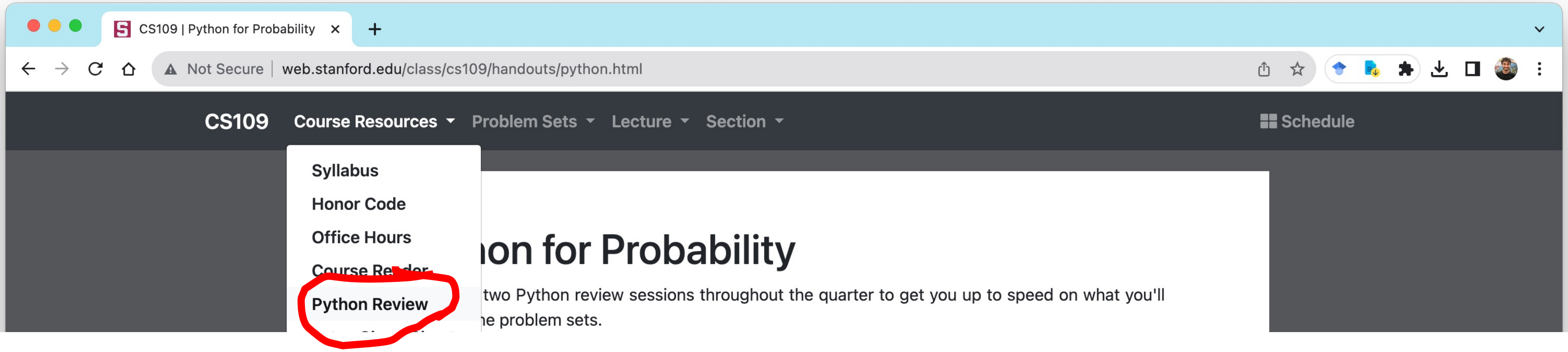
Probability

CS109

Python Review Session

Today at 5 @ Zoom with Joel!



Section Signups Are Due, In The Past?

We have virtual sections for SCPD students -- everyone gets to have a section!

Sign up before EOD tomorrow


We are going to make history today

Review

Counting & Combinatorics

Counting operations on *n* objects

- Sort, order matters {permutations}
  - 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: $\frac{(n+r-1)!}{n!(r-1)!}$

Counting & Combinatorics

Counting operations on *n* objects

- **Sort, order matters {permutations}**
  - Distinct: $n!$
  - Some Distinct: $\frac{n!}{n_1! n_2! \dots}$
- **Choose *k* {combinations}**
  - Distinct: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$
- **Put in *r* buckets**
  - Distinct: r^n
  - None Distinct: $\frac{(n+r-1)!}{n!(r-1)!}$

(5 options)	(4 options)	(3 options)	(2 options)	(1 option)
9	C	0	S	1

Counting & Combinatorics

Counting operations on *n* objects

Sort, order matters {permutations}

- 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: $\frac{(n+r-1)!}{n!(r-1)!}$

MISSISSIPPI

Counting & Combinatorics

Counting operations on *n* objects

Sort, order matters {permutations}

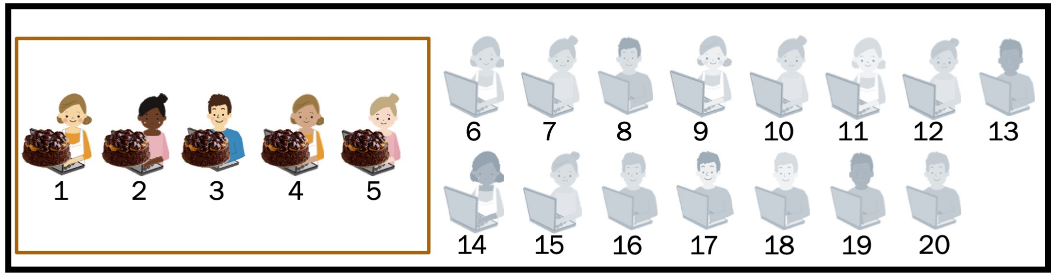
- Distinct: $n!$
- Some Distinct: $\frac{n!}{n_1!n_2!\ldots}$

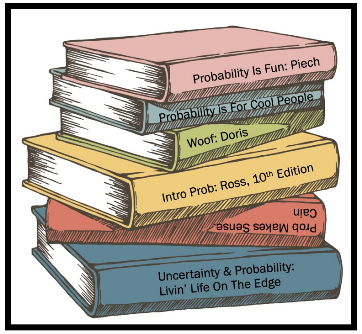
Choose *k* {combinations}

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

Put in *r* buckets

- Distinct: r^n
- None Distinct: $\frac{(n+r-1)!}{n!(r-1)!}$





?

Counting & Combinatorics

Counting operations on *n* objects

Sort, order matters {permutations}

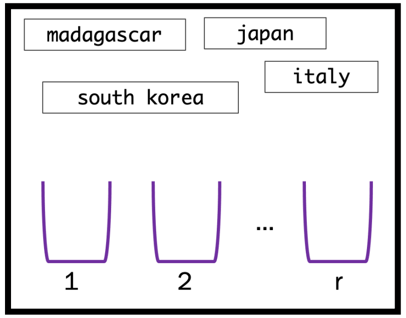
- Distinct: $n!$
- Some Distinct: $\frac{n!}{n_1!n_2!\dots}$

Choose *k* {combinations}

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

Put in *r* buckets

- Distinct: r^n
- None Distinct: $\frac{(n+r-1)!}{n!(r-1)!}$



Counting & Combinatorics

Counting operations on *n* objects

Sort, order matters {permutations}

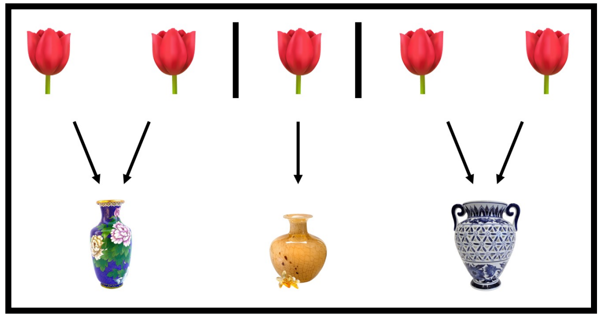
- Distinct: $n!$
- Some Distinct: $\frac{n!}{n_1!n_2!\dots}$

Choose *k* {combinations}

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

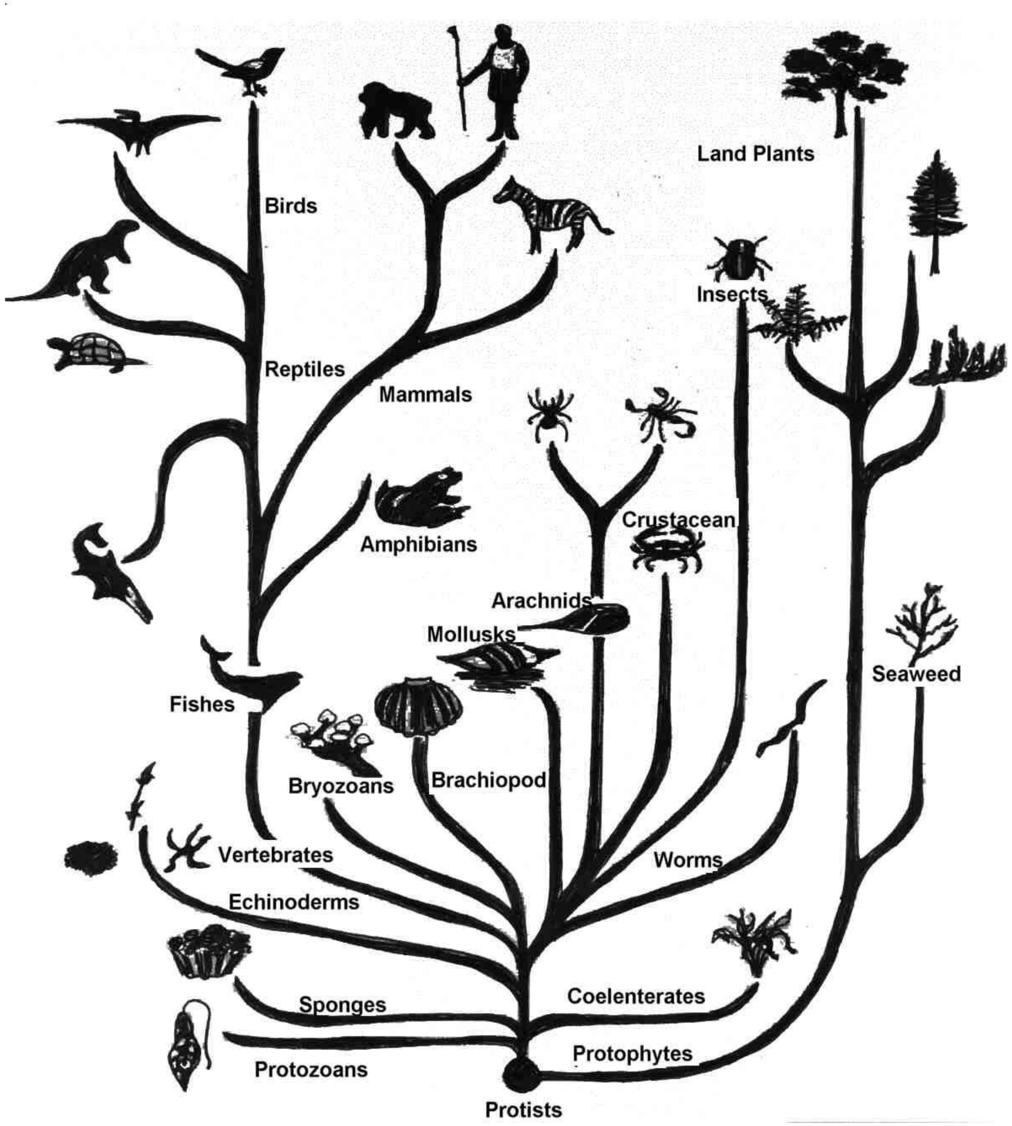
Put in *r* buckets

- Distinct: r^n
- None Distinct: $\frac{(n+r-1)!}{n!(r-1)!}$



$$x_1 + x_2 + \cdots + x_r = n,$$

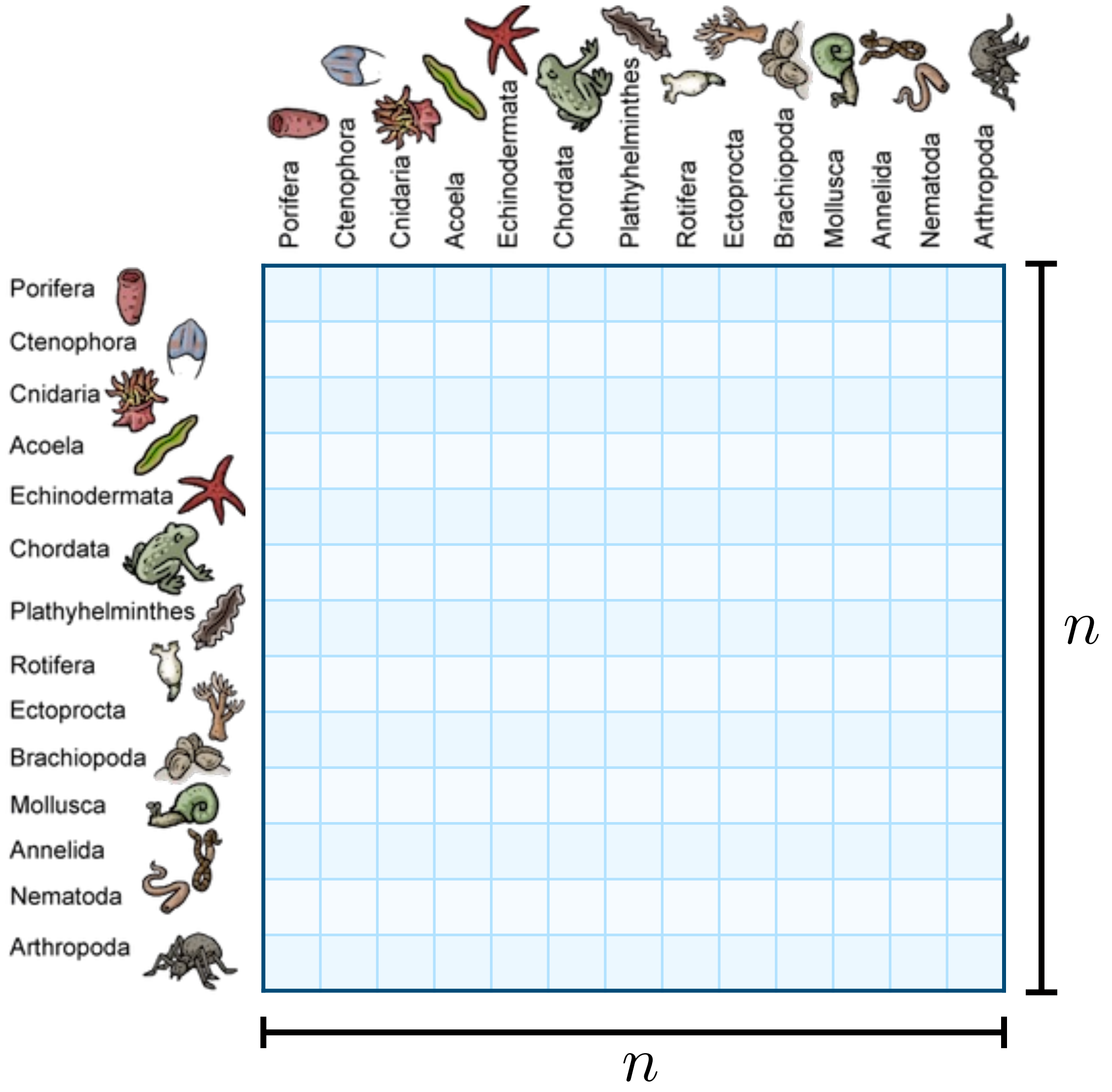
Counting Practice: Evolutionary Trees



Counting Practice: Evolutionary Trees

To construct an evolutionary tree between species, we have to compare data between all possible pairs of species.

If we have *n* species, how many unique pairs of species are there?

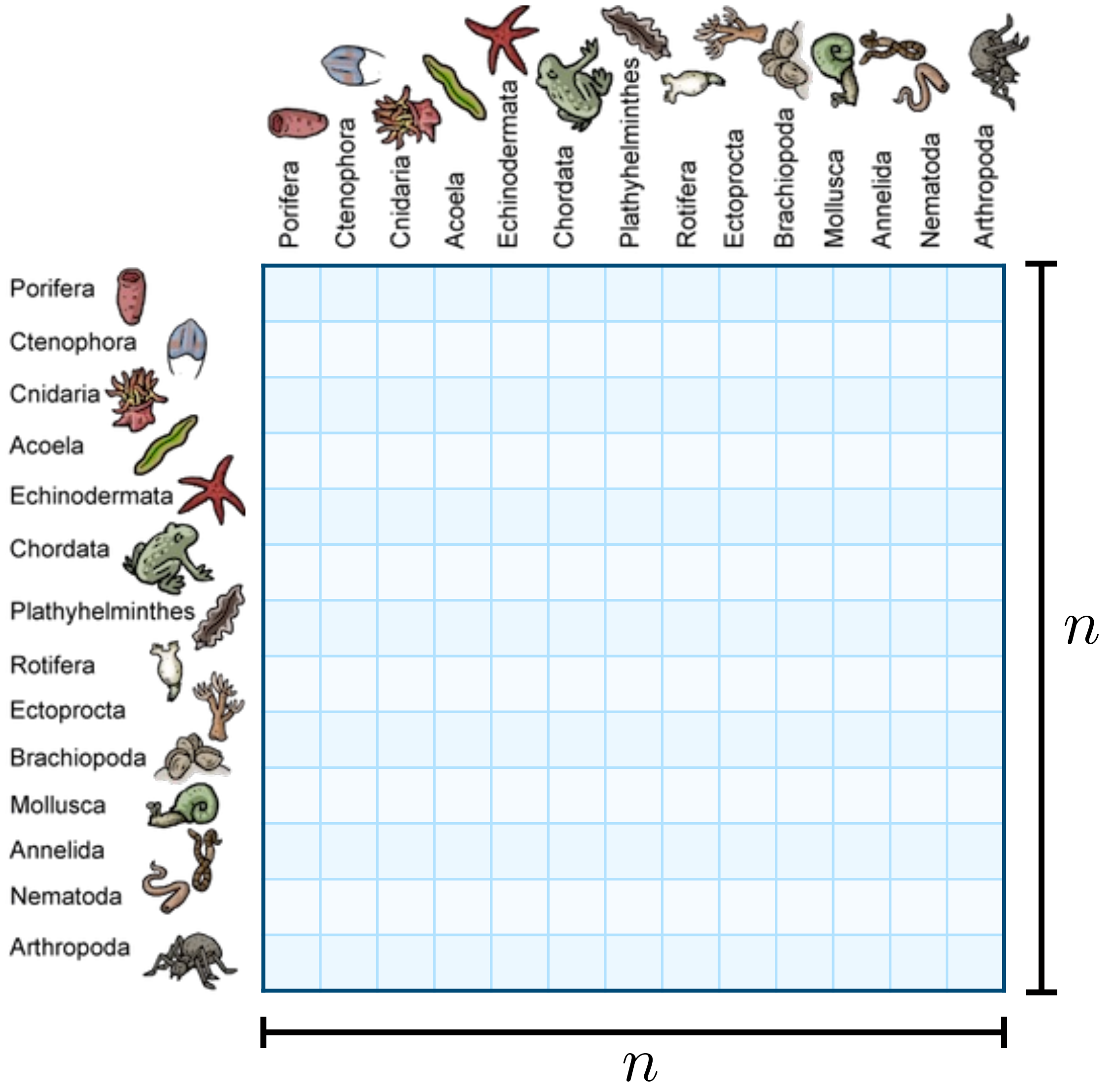


Counting Practice: Evolutionary Trees

To construct an evolutionary tree between species, we have to compare data between all possible pairs of species.

If we have *n* species, how many unique pairs of species are there?

Answer: $\binom{n}{2}$

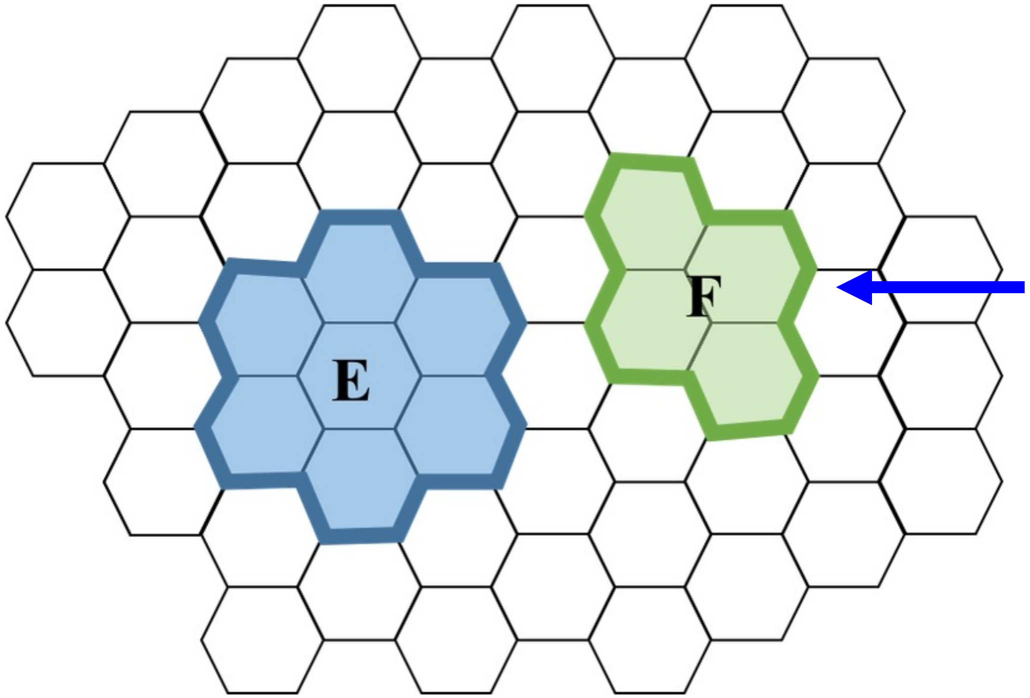


Events: Interesting Subsets of Outcomes

Experiments have *sets* of outcomes, containing groups of things that could possibly happen.

- **Event:** some subset of all possible outcomes that we care about

This is the entire **sample space**: all possible outcomes



Here is one event

Sample Space (S) vs. Event Space (E)

Experiment	Sample Space	Event	Event Space
Flipping a coin	{Heads, Tails}	Getting heads	{Heads}

$E \subseteq S$: Event spaces are always subsets of the sample space.

Sample Space (S) vs. Event Space (E)

Experiment	Sample Space	Event	Event Space
Flipping a coin	{Heads, Tails}	Getting heads	{Heads}
Rolling a dice	{1, 2, 3, 4, 5, 6}	At least 3	{3, 4, 5, 6}

$E \subseteq S$: Event spaces are always subsets of the sample space.

Sample Space (S) vs. Event Space (E)

Experiment	Sample Space	Event	Event Space
Flipping a coin	{Heads, Tails}	Getting heads	{Heads}
Rolling a dice	{1, 2, 3, 4, 5, 6}	At least 3	{3, 4, 5, 6}
Flipping two coins	{{H,H}, {H,T}, {T,H}, {T,T}}	One head	{{H,T}, {T,H}}

$E \subseteq S$: Event spaces are always subsets of the sample space.

Sample Space (S) vs. Event Space (E)

Experiment	Sample Space	Event	Event Space
Flipping a coin	{Heads, Tails}	Getting heads	{Heads}
Rolling a dice	{1, 2, 3, 4, 5, 6}	At least 3	{3, 4, 5, 6}
Flipping two coins	{{H,H}, {H,T}, {T,H}, {T,T}}	One head	{{H,T}, {T,H}}
# inches of rain	$\{x \mid x \in \mathbf{Z}, x \geq 0\}$	Drought	$\{x \mid x \in \mathbf{Z}, 0 \leq x \leq 2\}$

$E \subseteq S$: Event spaces are always subsets of the sample space.

Sample Space (S) vs. Event Space (E)

Experiment	Sample Space	Event	Event Space
Flipping a coin	{Heads, Tails}	Getting heads	{Heads}
Rolling a dice	{1, 2, 3, 4, 5, 6}	At least 3	{3, 4, 5, 6}
Flipping two coins	{{H,H}, {H,T}, {T,H}, {T,T}}	One head	{{H,T}, {T,H}}
# inches of rain	$\{x \mid x \in \mathbf{Z}, x \geq 0\}$	Drought	$\{x \mid x \in \mathbf{Z}, 0 \leq x \leq 2\}$
# hours slept	$\{x \mid x \in \mathbf{R}, 0 \leq x \leq 24\}$	Good sleep	$\{x \mid x \in \mathbf{R}, 7 \leq x \leq 12\}$

$E \subseteq S$: Event spaces are always subsets of the sample space.

What is a probability?

[suspense]

A number between 0 and 1

But it's a number we ascribe meaning to!

$$P(E)$$

…represents our belief that event E occurs.

Why Are Probabilities Beliefs?

The Formal, Technical Definition of Probability

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

$n(E)$ is the number of trials where event E happens

n is the number of trials

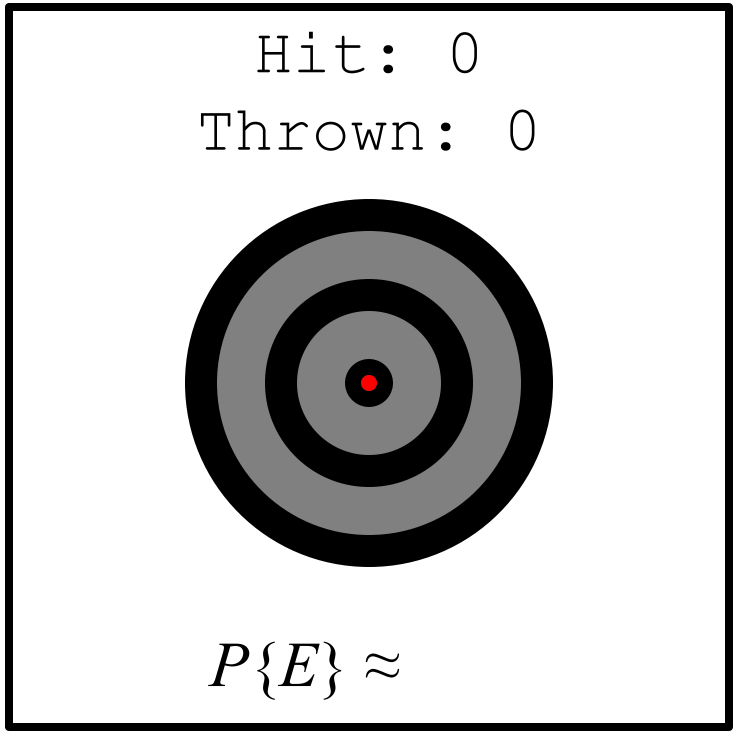
"If you repeated an experiment infinite times,

what fraction of the times does E happen?"

The Formal, Technical Definition of Probability

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

n is the number of darts thrown

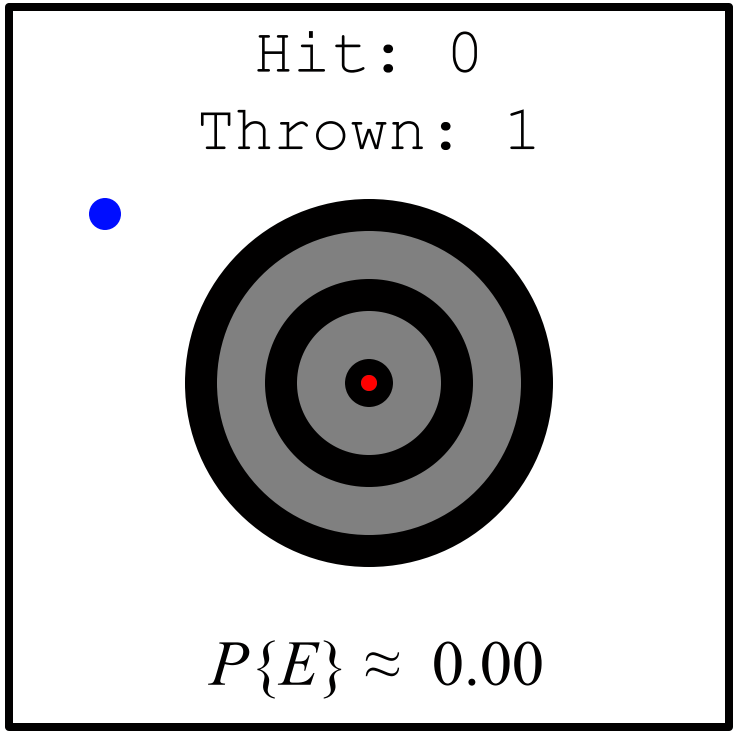


The target represents event E

The Formal, Technical Definition of Probability

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

n is the number of darts thrown

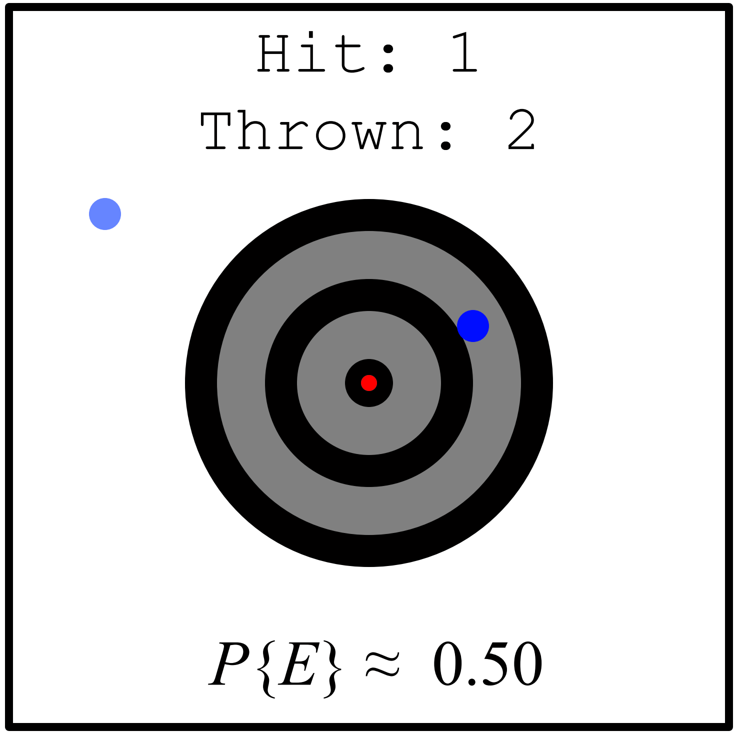


The target represents event E

The Formal, Technical Definition of Probability

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

n is the number of darts thrown

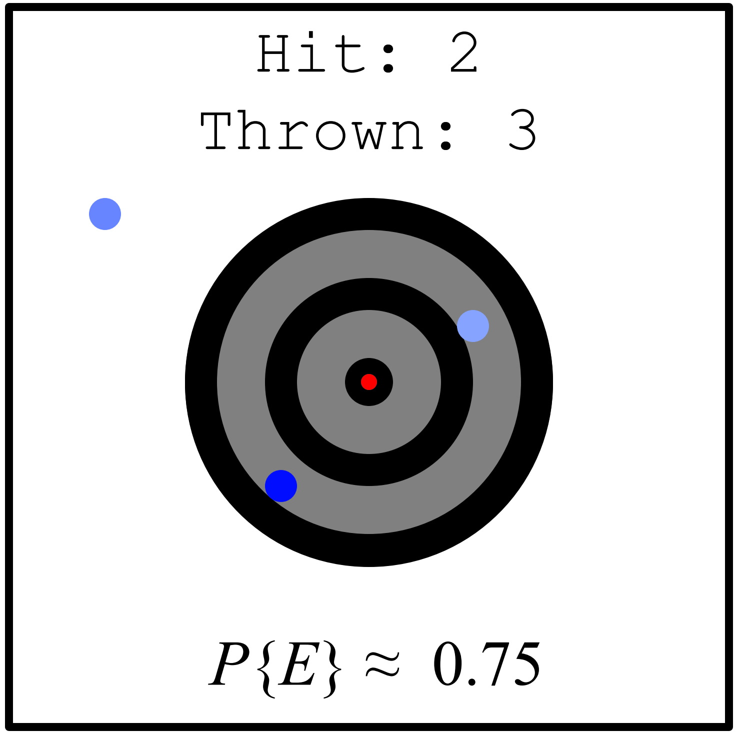


The target represents event E

The Formal, Technical Definition of Probability

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

n is the number of darts thrown

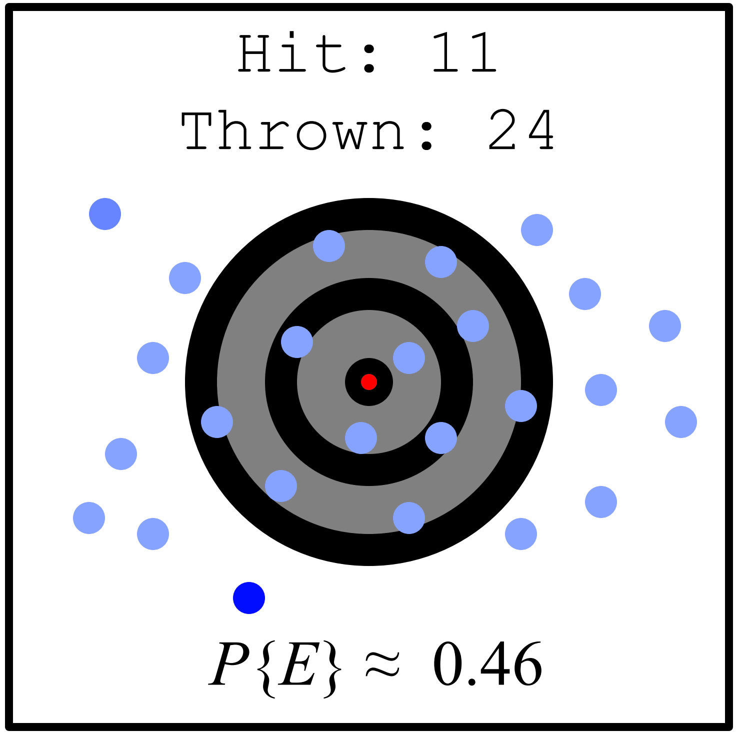


The target represents event E

The Formal, Technical Definition of Probability

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

n is the number of darts thrown



The target represents event E

Let's Simulate How This Formula Works: Coin Flips

```python
import random

# tip: don't go past 10000000

N_TRIALS = 10000000

def main():
    print("N_TRIALS: ", N_TRIALS)

    heads_count = 0
    for trial in range(N_TRIALS):
        result = flip_coin()
        if result == "heads":
            heads_count = heads_count + 1

    print("Estimated P(heads): ", heads_count / N_TRIALS)


def flip_coin():
    return random.choice(["heads", "tails"])
```

Calculating Probabilities From A Dataset

You're given a dataset of historical weather observations.

Day	Outcome
1	Rainy
2	Sunny
3	Rainy
4	Cloudy
5	Rainy
6	Sunny
7	Sunny
8	Sunny
	…
10000	Cloudy

Let E be the event that it is **Sunny**. What is P(E)?

Calculating Probabilities From A Dataset

You're given a dataset of historical weather observations.

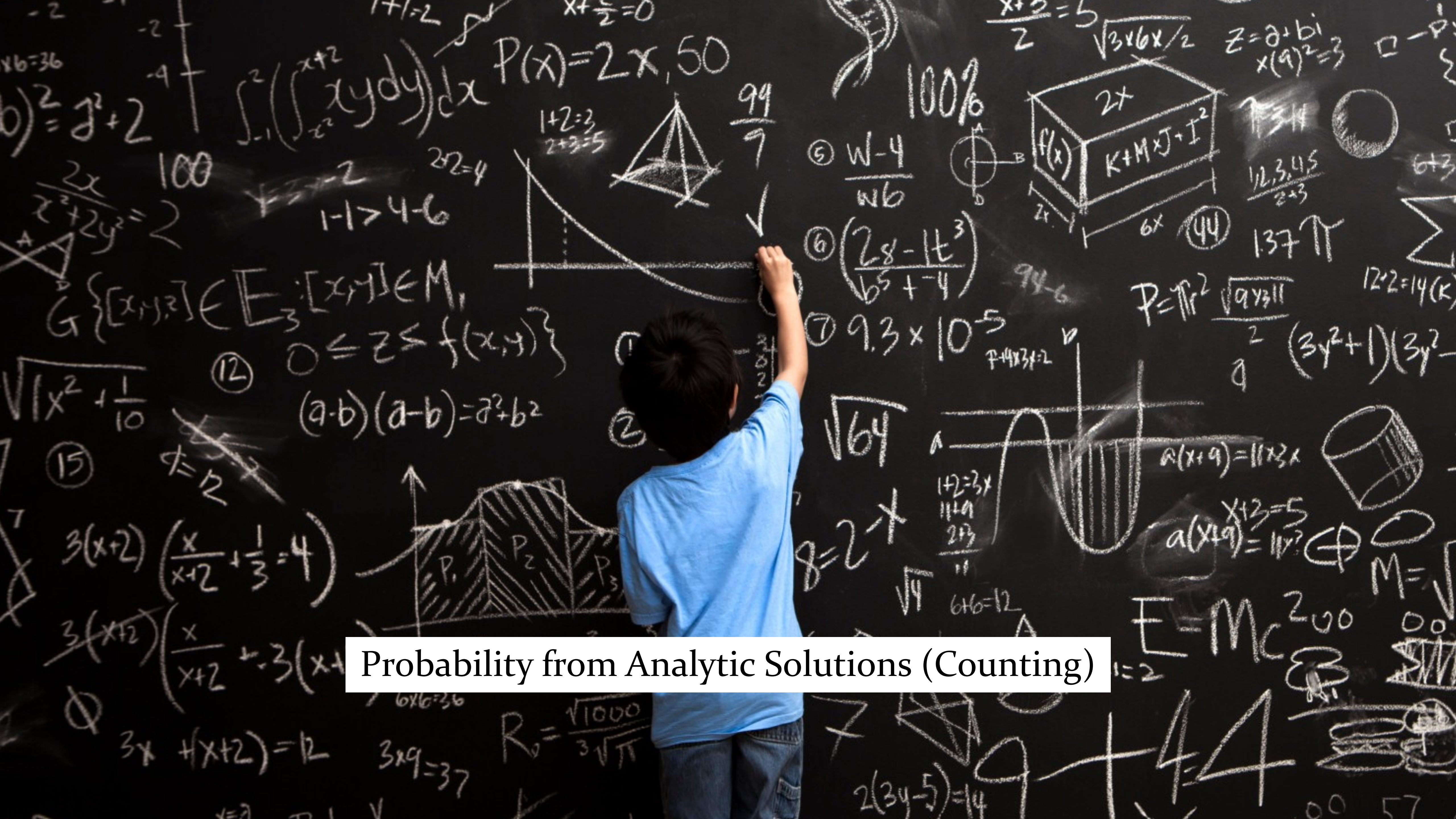
Day	Outcome
1	Rainy
2	Sunny
3	Rainy
4	Cloudy
5	Rainy
6	Sunny
7	Sunny
8	Sunny
	…
10000	Cloudy

Let E be the event that it is **Sunny**. What is P(E)?

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

$$\approx \frac{\text{Count}(E)}{10000}$$

$$\approx \frac{3332}{10000} \approx 0.3332$$

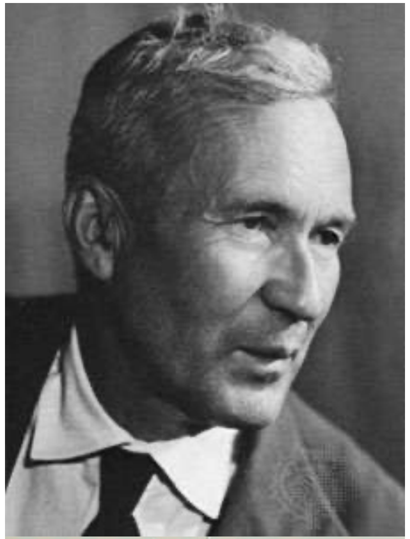


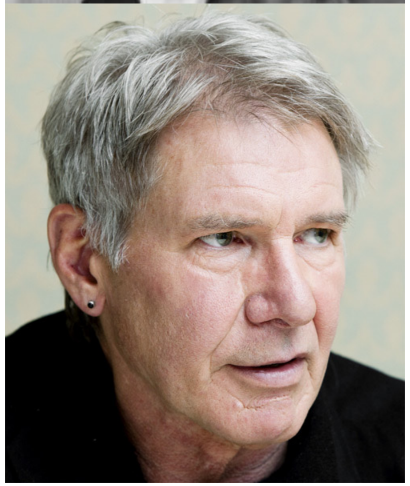
Probability from Analytic Solutions (Counting)

The Axioms of Probability

- Axiom 1: $0 \leq P(E) \leq 1$ All probabilities are between 0 and 1

Kolmogorov

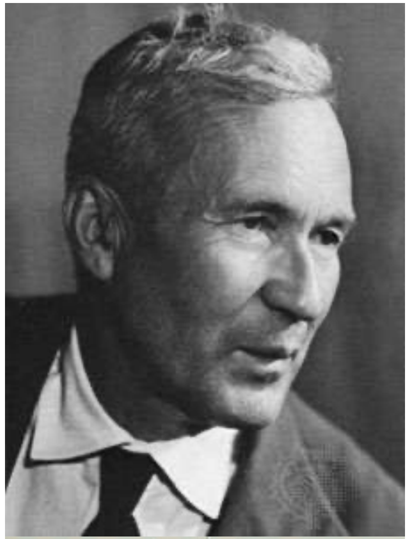


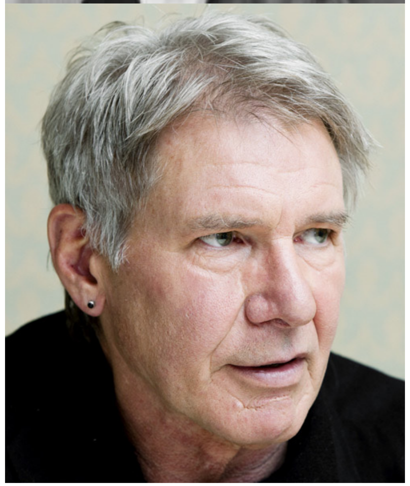


(looks like Harrison Ford?)

The Axioms of Probability

Kolmogorov



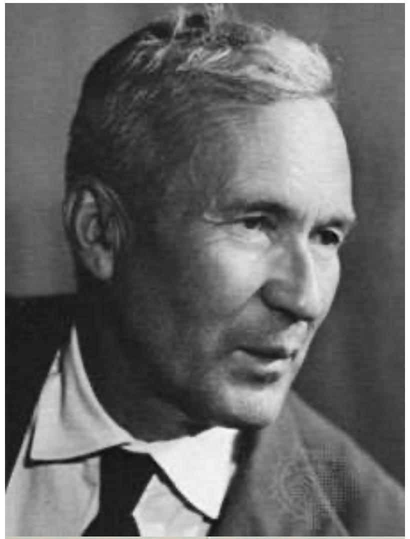


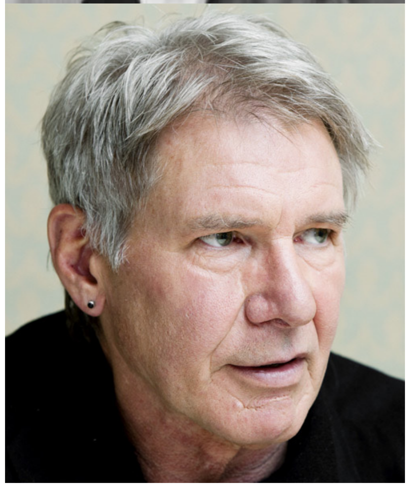
(looks like Harrison Ford?)

- Axiom 1: $0 \leq P(E) \leq 1$ — All probabilities are between 0 and 1
- Axiom 2: $P(S) = 1$ — The probability of the sample space is 1

The Axioms of Probability

Kolmogorov





(looks like Harrison Ford?)

- Axiom 1: $0 \leq P(E) \leq 1$ — All probabilities are between 0 and 1

- Axiom 2: $P(S) = 1$ — The probability of the sample space is 1

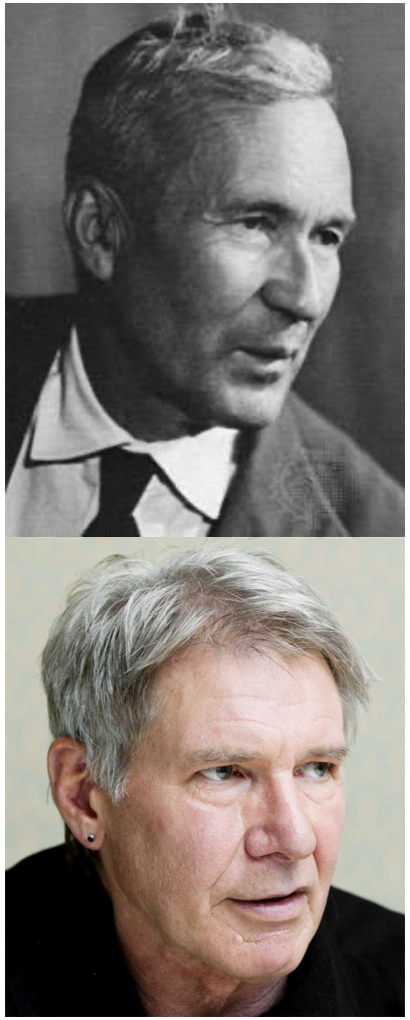
- Axiom 3: If events E and F are mutually exclusive,

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

Probability of event E *or* event F

The Axioms of Probability

Kolmogorov



(looks like Harrison Ford?)

- Axiom 1: $0 \leq P(E) \leq 1$ All probabilities are between 0 and 1

- Axiom 2: $P(S) = 1$ The probability of the sample space is 1

- Axiom 3: If events E and F are mutually exclusive,

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

Probability of event E *or* event F

- Identity 3*: $P(E^c) = 1 - P(E)$ Events either happen…or don't

"not E"

Equally Likely Outcomes

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

- Coin flip: S = {Head, Tails}
- Roll of 6-sided die: S = {1, 2, 3, 4, 5, 6}

If we have equally likely outcomes, then P(each outcome) $= \frac{1}{|S|}$

Equally Likely Outcomes

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

- Coin flip: S = {Head, Tails}
- Roll of 6-sided die: S = {1, 2, 3, 4, 5, 6}

*****If***** we have equally likely outcomes, then P(each outcome) $= \frac{1}{|S|}$

Therefore,

$$P(E) = \frac{\text{outcomes in } E}{\text{all outcomes}} = \frac{|E|}{|S|}$$

What Happens If Outcomes Aren't Equally Likely?

You've bought a lottery ticket. What is the probability that you win?

What Happens If Outcomes Aren't Equally Likely?

You've bought a lottery ticket. What is the probability that you win?

$$S = \{\text{Win, Lose}\}$$

$$E = \{\text{Win}\}$$

$$P(\text{Win}) = \frac{|E|}{|S|} = \frac{1}{2} = 0.5 \quad ?$$

Sometimes, Unequal Outcomes Are Less Obvious

Experiment	Sample Space	Event	Event Space
Flipping a coin	{Heads, Tails}	Getting heads	{Heads}
Rolling a dice	{1, 2, 3, 4, 5, 6}	At least 3	{1, 2, 3}
Flipping two coins	{{H,H}, {H,T}, {T,H}, {T,T}}	One head	{{H,T}, {T,H}}
# rainy days / year	$\{x \mid x \in \mathbf{Z}, x \geq 0\}$	Drought	$\{x \mid x \in \mathbf{Z}, 0 \leq x \leq 10\}$
# hours slept	$\{x \mid x \in \mathbf{R}, 0 \leq x \leq 24\}$	Good sleep	$\{x \mid x \in \mathbf{R}, 7 \leq x \leq 12\}$

$E \subseteq S$: Event spaces are always subsets of the sample space.

Sometimes, Unequal Outcomes Are Less Obvious

Option 1

Sample Space	Event Space
{{H,H}, {H,T}, {T,H}, {T,T}}	{{H,T}, {T,H}}

Option 2

Sample Space	Event Space
{{H,H}, {T,H}, {T,T}}	{{T,H}}

Sometimes, Unequal Outcomes Are Less Obvious

Option 1

Sample Space

{{H,H}, {H,T}, {T,H}, {T,T}}

Event Space

{{H,T}, {T,H}}

$$P(1\ \text{head}) = \frac{|E|}{|S|} = \frac{2}{4} = 0.5$$

Option 2

Sample Space

{{H,H}, {T,H}, {T,T}}

Event Space

{{T,H}}

$$P(1\ \text{head}) = \frac{|E|}{|S|} = \frac{1}{3} = 0.33$$

Which one is right?

To The Code: Simulating Two Coin Flips

```python
import random

# tip: don't go past 10000000

N_TRIALS = 10000000

def main():
    print("N_TRIALS: ", N_TRIALS)

    one_head_count = 0
    for trial in range(N_TRIALS):
        first_result = flip_coin()
        second_result = flip_coin()
        num_heads = first_result + second_result
        if num_heads == 1:
            one_head_count = one_head_count + 1

    print("Estimated P(one head): ", one_head_count / N_TRIALS)


def flip_coin():
    # 1 means heads, 0 means tails
    return random.choice([1, 0])
```

Sometimes, Unequal Outcomes Are Less Obvious

Option 1

Sample Space

{{H,H}, {H,T}, {T,H}, {T,T}}

Event Space

{{H,T}, {T,H}}

$$P(1 \text{ head}) = \frac{|E|}{|S|} = \frac{2}{4} = 0.5$$

Only this way has equally likely outcomes!

Option 2

Sample Space

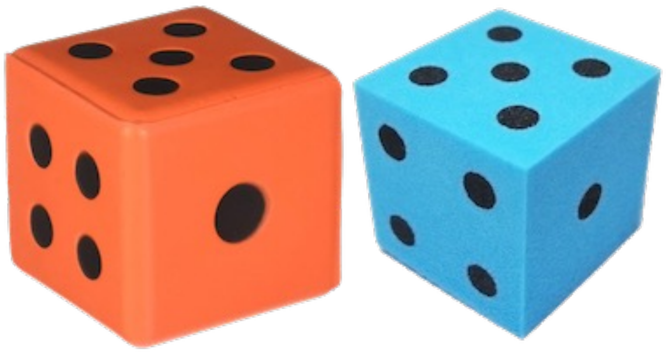
{{H,H}, {T,H}, {T,T}}

Event Space

{{T,H}}

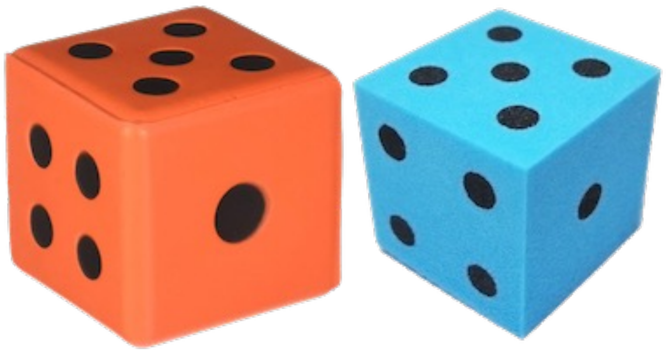
$$P(1 \text{ head}) = \frac{|E|}{|S|} = \frac{1}{3} = 0.33$$

Sum of Two Dice



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

Sum of Two Dice

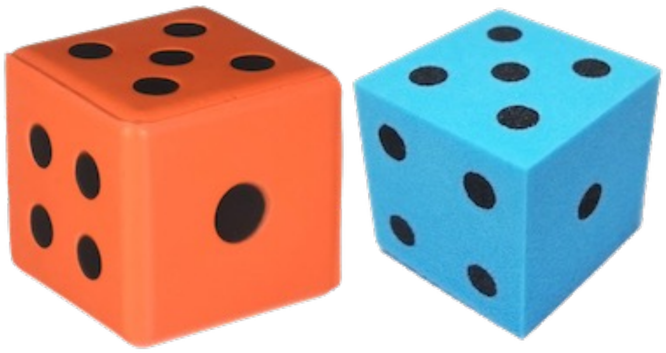


You roll two 6-sided dice. What is P(sum = 7)?

Imagine a few different ways of writing out outcomes.

What ways produce equally likely outcomes?

Sum of Two Dice

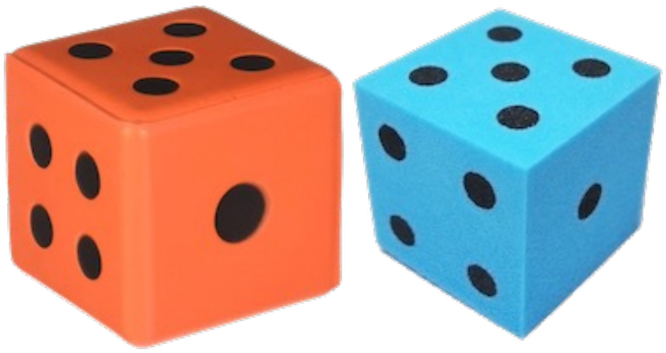


You roll two 6-sided dice. What is P(sum = 7)?

Outcomes are just possible sums

{2, 3, 4, 5,
6, **7**, 8, 9,
10, 11, 12}

Sum of Two Dice



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

Outcomes are just possible sums

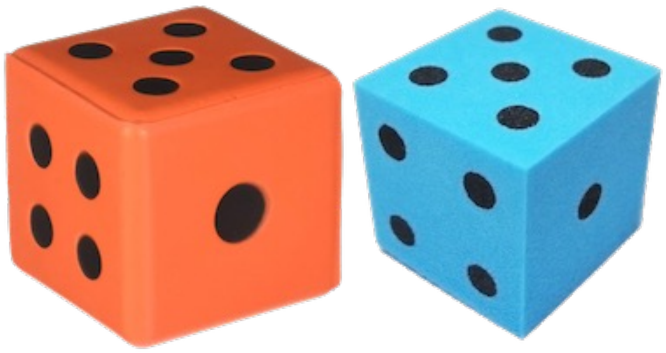
{2, 3, 4, 5,
6, **7**, 8, 9,
10, 11, 12}

Think of the dice as **distinct**

[5 , 5]

Value dice 1   Value dice 2

Sum of Two Dice



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

Outcomes are just possible sums

{2, 3, 4, 5, 6, **7**, 8, 9, 10, 11, 12}

Think of the dice as **distinct**

[5 , 5]

Value dice 1 — Value dice 2

Think of the dice as **indistinct**

{ 5 , 5 }

Value of a dice — Value of a dice

Sum of Two Dice: Distinct

You 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] }

Think of the dice as **distinct**

[5 , 5]

Value dice 1 | Value dice 2

Sum of Two Dice: Distinct

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

Think of the dice as **distinct**

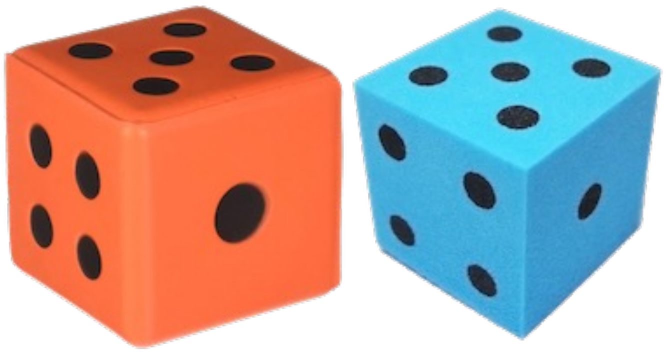
[5 , 5]

Value dice 1 — Value dice 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] }

$$P(E) = \frac{|E|}{|S|} = \frac{6}{36} = 0.16\bar{6}$$

Sum of Two Dice



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

Outcomes are just possible sums

{2, 3, 4, 5, 6, **7**, 8, 9, 10, 11, 12}

$$P(E) = \frac{1}{11} = 0.0\bar{9}$$

Think of the dice as **distinct**

[5 , 5]

Value dice 1 — Value dice 2

$$P(E) = \frac{6}{36} = 0.16\bar{6}$$

Think of the dice as **indistinct**

{ 5 , 5 }

Value of a dice — Value of a dice

Sum of Two Dice: Indistinct

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

Think of the dice as **indistinct**

$$\{\ 5\ ,\ 5\ \}$$

Value of a dice — Value of a dice

$$
\begin{aligned}
S = \{ & \{1,1\} \quad \{1,2\} \quad \{1,3\} \quad \{1,4\} \quad \{1,5\} \quad \{1,6\} \\
& \{2,2\} \quad \{2,3\} \quad \{2,4\} \quad \{2,5\} \quad \{2,6\} \\
& \{3,3\} \quad \{3,4\} \quad \{3,5\} \quad \{3,6\} \\
& \{4,4\} \quad \{4,5\} \quad \{4,6\} \\
& \{5,5\} \quad \{5,6\} \\
& \{6,6\} \}
\end{aligned}
$$

Sum of Two Dice: Indistinct

You roll two 6-sided dice. What is P(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} }

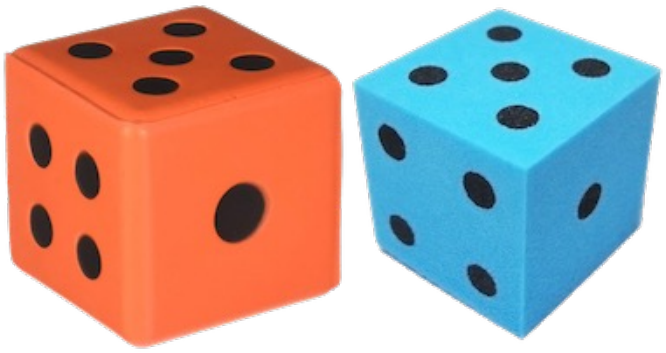
Think of the dice as **indistinct**

{ 5 , 5 }

Value of a dice — Value of a dice

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

Sum of Two Dice



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

Outcomes are just possible sums

{2, 3, 4, 5, 6, **7**, 8, 9, 10, 11, 12}

$$P(E) = \frac{1}{11} = 0.0\bar{9}$$

Think of the dice as **distinct**

[5 , 5]

Value dice 1 — Value dice 2

$$P(E) = \frac{6}{36} = 0.16\bar{6}$$

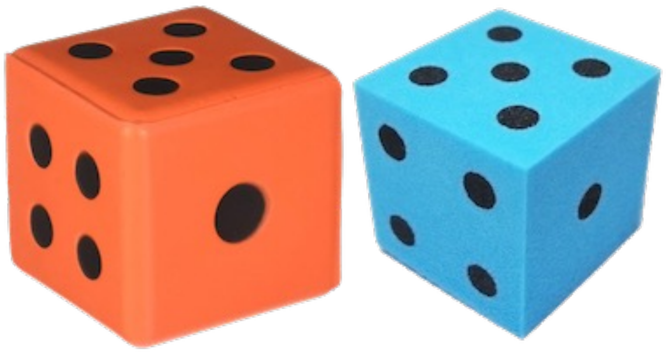
Think of the dice as **indistinct**

{ 5 , 5 }

Value of a dice — Value of a dice

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

Sum of Two Dice



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

Outcomes are just possible sums

3, 4

10 … 12}

$$ \ldots) = \frac{1}{11} = 0 \ldots $$

Think of the dice as **distinct**

$$ \left[\; 5 \;,\; 5 \;\right] $$

…lue … 1 … ue … dice 2

$$ P(E) = \frac{6}{36} = 0.16\bar{6} $$

Think of the dice as **indistinct**

Value o… …ue of

a di… …ice

$$ \ldots) = \frac{3}{20} = 0 \ldots $$

A Literal Toy Problem: Pigs and Cows

Concept Check

There are 4 cows and 3 pigs. You choose 3 at random.
What is P(1 pig and 2 cows)?

What is an equally likely sample space here?

The Choice of Sample Space is Yours!

Which choice will lead to equally likely outcomes?

	Distinct	Indistinct
Unordered		
Ordered		

The Choice of Sample Space is Yours!

Which choice will lead to equally likely outcomes?

	Distinct	Indistinct
Unordered	$\{cow_1, pig_2, pig_3\}$ $\{cow_1, cow_2, cow_3\}$	{2 cows, 1 pig} {3 cows}
Ordered	$[cow_1, pig_2, pig_3]$ $[cow_1, cow_2, cow_3]$	[cow, pig, cow] [cow, cow, cow]

The Choice of Sample Space is Yours!

Which choice will lead to equally likely outcomes?

	Distinct	Indistinct
Unordered	$\{cow_1, pig_2, pig_3\}$ $\{cow_1, cow_2, cow_3\}$	{2 cows, 1 pig} {3 cows}
Ordered	$[cow_1, pig_2, pig_3]$ $[cow_1, cow_2, cow_3]$	[cow, pig, cow] [cow, cow, cow]

A Literal Toy Problem: Pigs and Cows

Concept Check

There are 4 cows and 3 pigs. You choose 3 at random.
What is P(1 pig and 2 cows)?

(earlier version of this problem had a typo! The answer checker is looking for the typo answer, which is 12/35)

Ordered and Distinct

- Pick 3 ordered items: |S| = 7 * 6 * 5 = 210
- Pick pig as either 1st, 2nd, or 3rd item:
  |E| = (4 * 3 * 3) + (3 * 4 * 3) + (3 * 3 * 4) = 108

P(1 pig, 2 cows) = 108/210 = 18/35

Unordered and Distinct

- $|S| = \binom{7}{3} = 35$
- $|E| = \binom{4}{2}\binom{3}{1} = 18$

P(1 pig, 2 cows) = 18/35

Tips For Ensuring Equally Likely Outcomes

Start by imagining individual outcomes.

- Ask yourself: Should objects be distinct or indistinct? Ordered or unordered?
  - Distinct *usually* is correct – you can imagine indistinct objects as distinct
- Can you find two outcomes that are *not* equally likely? If so, try a different approach

Tips For Ensuring Equally Likely Outcomes

Start by imagining individual outcomes.

- Ask yourself: Should objects be distinct or indistinct? Ordered or unordered?
  - Distinct *usually* is correct – you can imagine indistinct objects as distinct
- Can you find two outcomes that are *not* equally likely? If so, try a different approach

Start with the sample space first, then the event space second.

- Can you imagine a “generative story” for building outcomes? Can you count them?
- Does this generative story produce ALL outcomes you want?

Tips For Ensuring Equally Likely Outcomes

Start by imagining individual outcomes.

- Ask yourself: Should objects be distinct or indistinct? Ordered or unordered?
  - Distinct *usually* is correct – you can imagine indistinct objects as distinct
- Can you find two outcomes that are *not* equally likely? If so, try a different approach

Start with the sample space first, then the event space second.

- Can you imagine a "generative story" for building outcomes? Can you count them?
- Does this generative story produce ALL outcomes you want?

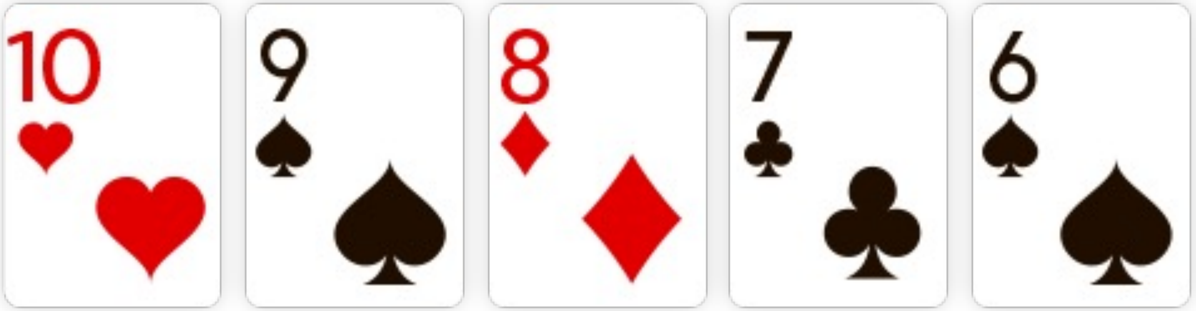
At the end, double-check:

- Are your sample space and event space counted the same way? (Is $E \subseteq S$?)

Probability of a Straight Poker Hand

A “straight” hand in poker is any 5 consecutive rank cards of any suit.

What is P(straight)?



What is a generative story for building straights?

Are outcomes equally likely?

Probability of a Straight Poker Hand

A “straight” hand in poker is any 5 consecutive rank cards of any suit.

What is P(straight)?

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

All possible ways to choose 5 cards from 52

Probability of a Straight Poker Hand

A "straight" hand in poker is any 5 consecutive rank cards of any suit.

What is P(straight)?

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

All possible ways to choose 5 cards from 52

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

10 choices for the start value, then rest are fixed

4 choices for the suit, need to choose for each card

Probability of a Straight Poker Hand

A "straight" hand in poker is any 5 consecutive rank cards of any suit.

What is P(straight)?

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

All possible ways to choose 5 cards from 52

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

10 choices for the start value, then rest are fixed

4 choices for the suit, need to choose for each card

$$P(\text{straight}) = \frac{|E|}{|S|} = \frac{10 \cdot \binom{4}{1}^5}{\binom{52}{5}} \approx 0.0039$$

Chip Defect Detection

Your company has manufactured *n* chips, 1 of which is defective.

k chips are randomly selected from *n* for testing.

What is the probability that the defective chip is in the *k* selected chips?

$$|S| =$$

$$|E| =$$

Chip Defect Detection

Your company has manufactured *n* chips, 1 of which is defective.

k chips are randomly selected from *n* for testing.

What is the probability that the defective chip is in the *k* selected chips?

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

$$|E| =$$

Chip Defect Detection

Your company has manufactured *n* chips, 1 of which is defective.

k chips are randomly selected from *n* for testing.

What is the probability that the defective chip is in the *k* selected chips?

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

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

Choose the defective chip, then choose k – 1 other chips

Chip Defect Detection

Your company has manufactured *n* chips, 1 of which is defective.

k chips are randomly selected from *n* for testing.

What is the probability that the defective chip is in the *k* selected chips?

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

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

Choose the defective chip, then choose k – 1 other chips

$$\text{P(defective chip is in } k \text{ selected chips)} = \frac{\binom{1}{1}\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{\frac{(n-1)!}{(k-1)!(n-k)!}}{\frac{n!}{k!(n-k)!}} = \frac{k}{n}$$


Let’s make history

Trailing the dovetail shuffle to its lair – Persi Diaconis

The Annals of Applied Probability
1992, Vol. 2, No. 2, 294–313

TRAILING THE DOVETAIL SHUFFLE TO ITS LAIR

By Dave Bayer[1] and Persi Diaconis[2]

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 $0 \le k \le n$. 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

Received January 1990; revised May 1991.

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

[2]Partially 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.

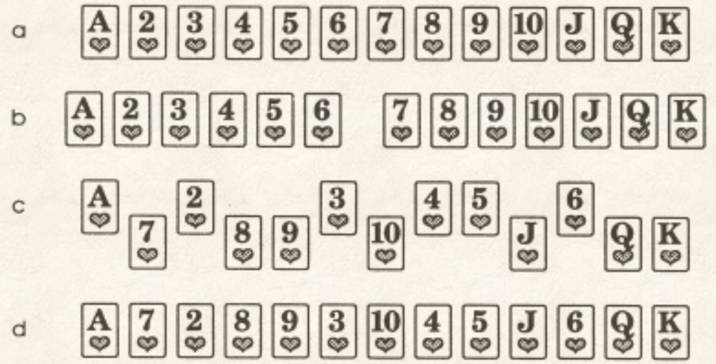


Fig. 1. *A riffle shuffle. (a) We begin with an ordered deck. (b) The deck is divided into two packets of similar size. (c) The two packets are riffled together. (d) The two packets can still be identified in the shuffled deck as two distinct "rising sequences" of face values.*

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 $x \to 2x \pmod 1$. 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

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

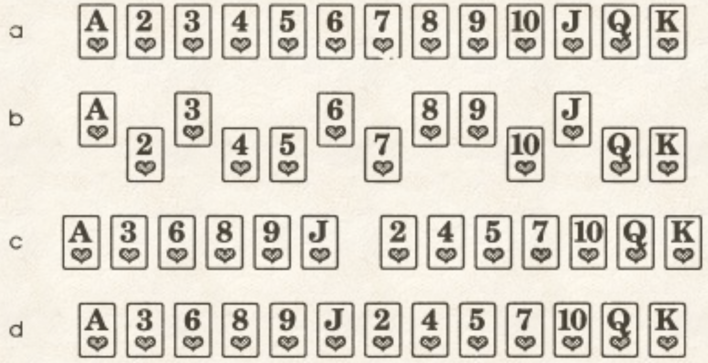


Fig. 2. *An inverse riffle shuffle. (a) We begin with a sorted deck. (b) Each card is moved one way or the other uniformly at random, to "pull apart" a riffle shuffle and retrieve two packets. (c) The two packets are placed in sequence. (d) The two packets can still be identified in the shuffled deck;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

Trailing the Dovetail Shuffle to Its Lair

You and one friend each shuffle your own decks of 52 cards.

What is the probability that the two decks are in different orders?

Trailing the Dovetail Shuffle to Its Lair

You and one friend each shuffle your own decks of 52 cards.

What is the probability that the two decks are in different orders?

$$|S| = 52!$$
$$|E| = 52! - 1$$

$$P(\text{different}) = \frac{|E|}{|S|} = \frac{52! - 1}{52!} > 0.999999999...$$

...about 67 “9”s

Trailing the Dovetail Shuffle to Its Lair

You and **two friends** each shuffle your own decks of 52 cards.

What is the probability that your friends' decks are in a different order from yours?

Trailing the Dovetail Shuffle to Its Lair

You and **two friends** each shuffle your own decks of 52 cards.

What is the probability that your friends' decks are in a different order from yours?

$$|S| = 52!^2$$

$$|E| = (52!–1)^2$$

$$P(\text{different}) = \frac{|E|}{|S|} = \frac{(52! - 1)^2}{52!^2}$$

Trailing the Dovetail Shuffle to Its Lair

You shuffle a deck of 52 cards.

What is the probability that the order of your deck has **never been seen before?**

Trailing the Dovetail Shuffle to Its Lair

You shuffle a deck of 52 cards.

What is the probability that the order of your deck has **never been seen before?**

$$|S| = 52!^n$$

$$|E| = (52!–1)^n$$

$$P(\text{different}) = \frac{|E|}{|S|} = \frac{(52! - 1)^n}{52!^n}$$

Trailing the Dovetail Shuffle to Its Lair

You shuffle a deck of 52 cards.

What is the probability that the order of your deck has **never been seen before?**

$$|S| = 52!^n$$

$$|E| = (52! - 1)^n$$

$$P(\text{different}) = \frac{|E|}{|S|} = \frac{(52! - 1)^n}{52!^n}$$

Assume 7 billion people have been shuffling cards once a second since 52-card decks were invented.

For $n = 10^{20}$,
P(any deck ever matching yours) $< 10^{-47}$

Next time: The Core Probability Toolkit™

Have a great weekend!