CS70 @ UC Berkeley, Spring 2026

Lecture 16
Introduction to Discrete Probability

March 17, 2026

Introduction

Ubiquity of Probabilistic Modeling

- Probability theory has its origins in gambling – analyzing card games, dice, roulette wheels, and so on. Today it is an essential tool in diverse areas of science.

- **Computer Science**
  - Machine learning / AI — probabilistic models, Bayesian inference (CS188, CS189)
  - Randomized algorithms — hashing, sampling, approximation algorithms (CS174)
  - Cryptography — probabilistic security guarantees
  - Distributed systems — probabilistic consensus, fault tolerance
  - Information theory — entropy and channel capacity

- **Other Engineering Systems**
  - Signal processing — noise models, filtering
  - Communications — error-correcting codes, channel noise
  - Control theory — stochastic control and Kalman filtering
  - Wireless networks — fading channels and interference models
  - Queueing theory — internet traffic, call centers
  - Operations research — stochastic optimization
  - Robotics — probabilistic localization

Ubiquity of Probabilistic Modeling

- **Everyday Technology**
  - Search engines — ranking algorithms
  - Recommendation systems — Netflix, Amazon
  - Navigation — probabilistic GPS filtering
  - Spam detection — Bayesian classification
- **Economics & Finance**
  - Financial markets — models of market forecast
  - Risk management — probabilistic modeling of uncertainty
  - Game theory — mixed strategies
- **Biology & Medicine**
  - Evolutionary dynamics — fixation probabilities
  - Neuroscience — stochastic neural firing
  - Epidemiology — stochastic disease spread models
  - Single-cell biology — gene expression noise
  - Clinical trials — statistical inference and uncertainty

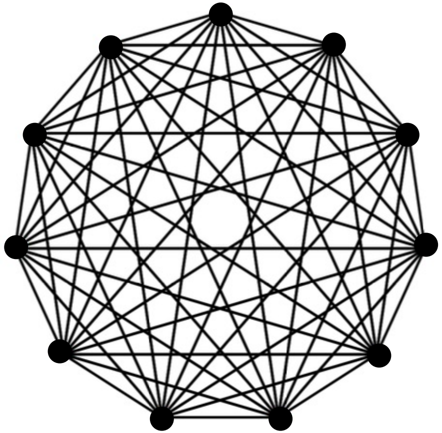
Ubiquity of Probabilistic Modeling

- **Physics**
  - Quantum mechanics — wave functions give probability amplitudes
  - Statistical mechanics — macroscopic laws from random particle motion
  - Thermodynamics — entropy and microstate probabilities
- **Earth & Environmental Sciences**
  - Climate modeling — stochastic dynamics
  - Seismology — earthquake probabilities
- **Mathematics**
  - Probabilistic method — existence proof
  - Random graphs — network structure and phase transitions
  - Stochastic processes — Brownian motion, Markov chains
  - Random matrices — physics, statistics, number theory

The Probabilistic Method

- **Key Idea:** Prove existence by showing that there is positive probability that a randomly chosen object has the desired property.
- This technique, pioneered by **Paul Erdős**, is covered in depth in CS174. In this course, we will get a glimpse of this powerful method.

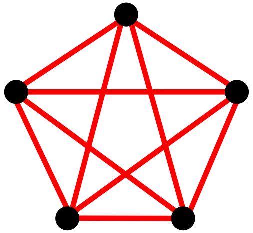
K_{11}



Is it possible to color the edges of K_{11} using two colors such that there exists **no** monochromatic K_5?

Not allowed

K_5



We will later prove the following result using the probabilistic method:

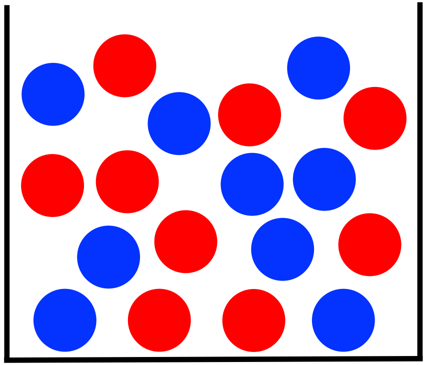
Theorem: For $m < n$, if $\binom{n}{m} < 2^{\binom{n}{m}-1}$, then there exists a 2-coloring of the edges of K_n such that it contains no monochromatic K_m subgraph.

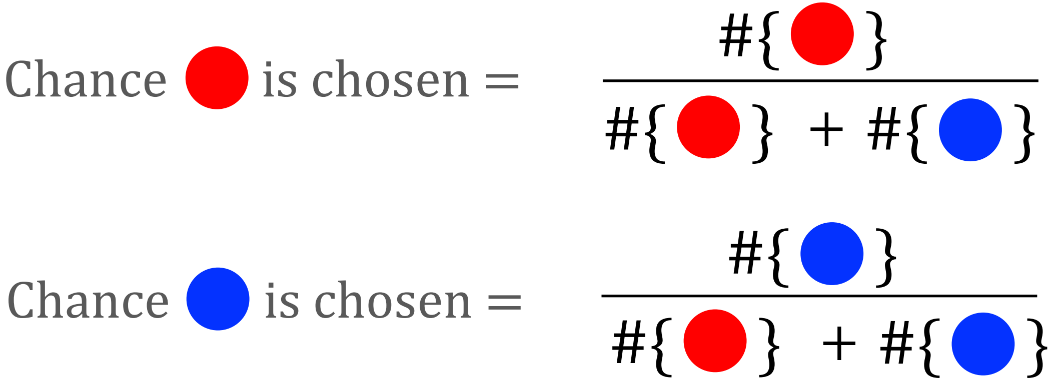
What is Probability?

- **Intuitive answer**: a measure of likelihood or chance.

0 Impossible — 1 Certain

- **Probability as proportion:**
  - Mix well and pick a ball without looking
  - Assume that each ball is equally likely to be chosen. Then,





How to estimate the number of fish in a lake

How to estimate the number of fish in a lake

Capture-Mark-Recapture Method

- Catch r fish, mark them red, and release them.
- Some time later, catch n fish. Let R_n = number of fish with red marks.
- What can be said about the total (unknown) number N of fish in the lake?

- Heuristics: proportion of red fish in sample ≈ proportion in lake:

$$\frac{R_n}{n} \approx \frac{r}{N}$$

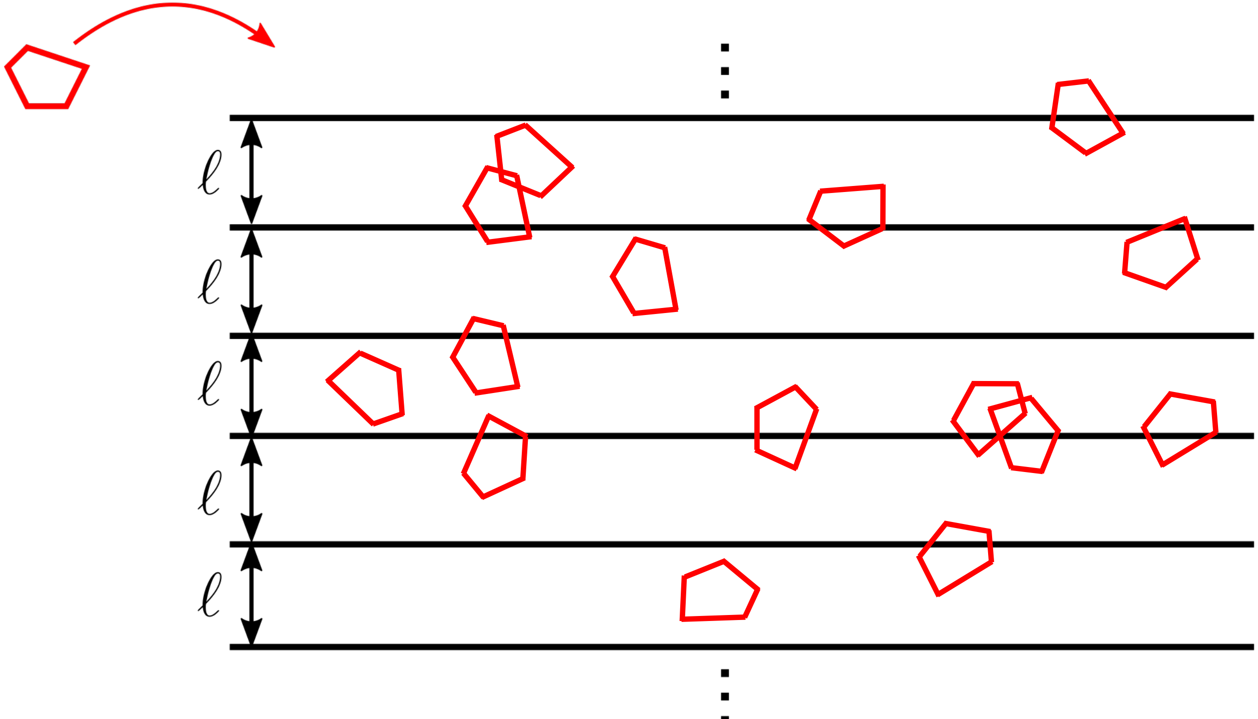
- So, we expect $\frac{nr}{R_n}$ to be a good estimator of N.
- In Note 16, we will derive a formula for the probability that $R_n = k$, where $k \in \{0, \dots, r\}$, in terms of r, N, k, n, and show that $N = \frac{nr}{k}$ maximizes that probability.

Probability as Frequency

- Suppose you have a biased coin that shows up **Heads** with probability p, where $0 < p < 1$, when you toss it randomly.
- What does this mean?

Convex Polygon Throws

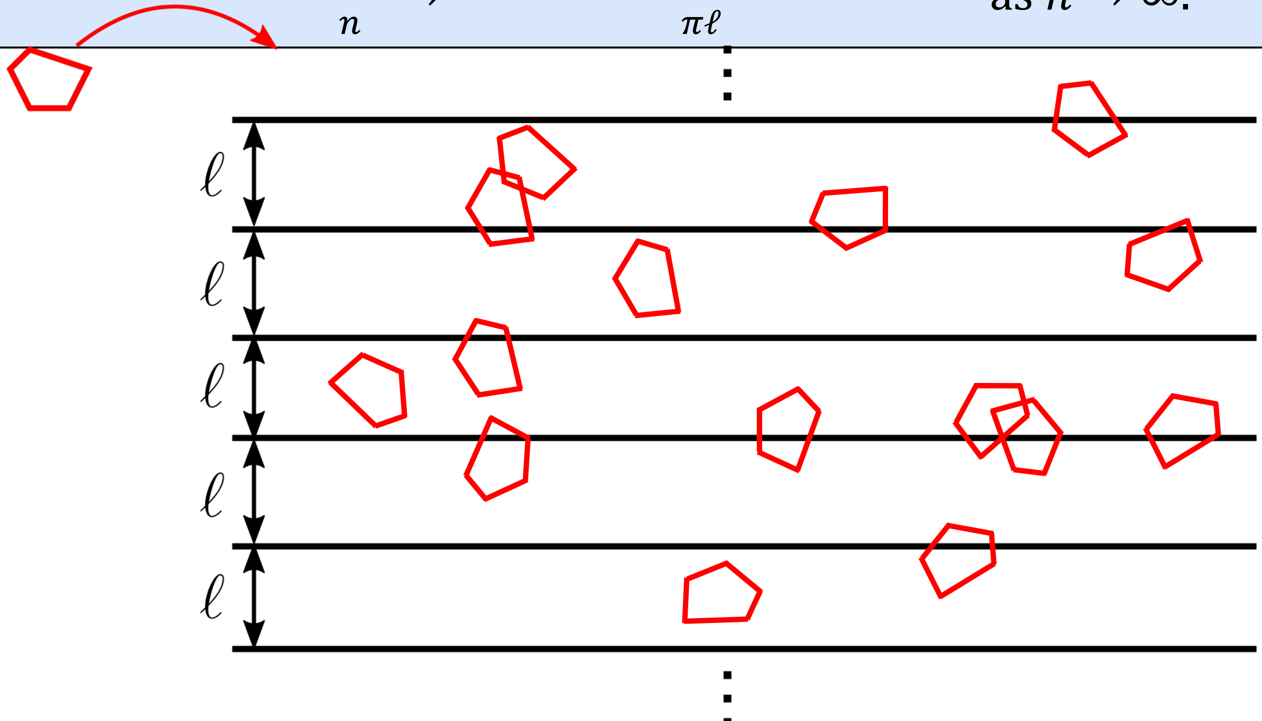
Consider an infinite array of parallel lines with gap size ℓ and a **sufficiently small** (so it cannot cross more than one line at once) convex polygon.
If the polygon is thrown at random, what is the probability that it touches the grid?



Convex Polygon Throws

Theorem. Let X_n denote the number of times that the convex polygon (sufficiently small so it cannot cross more than one line at once) intersects the grid in n random throws. Then,

$$\frac{X_n}{n} \longrightarrow \frac{\text{circumference of the polygon}}{\pi\ell} \quad \text{as } n \to \infty.$$

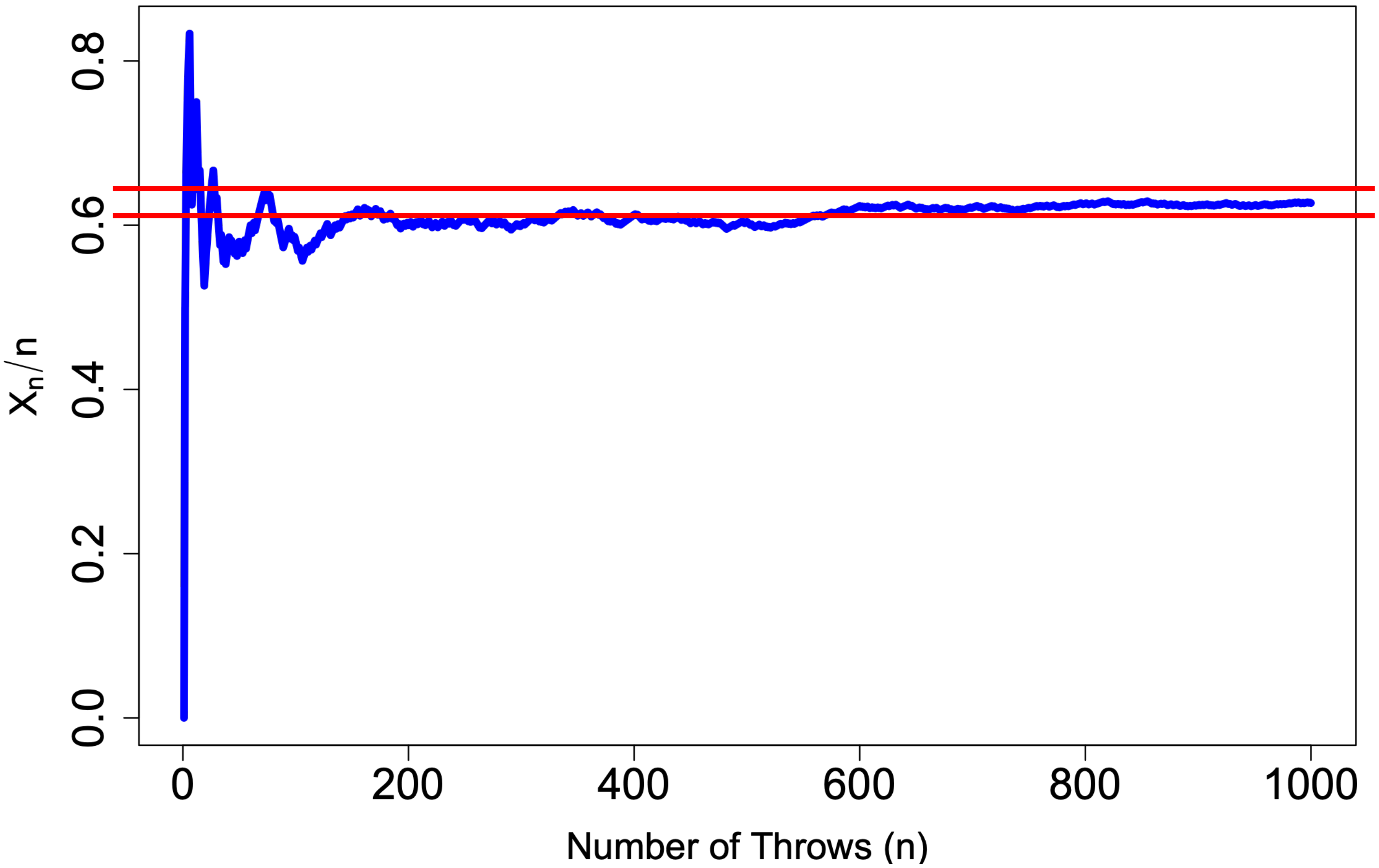


We will prove this later in the course.

For a line segment of length ℓ, circumference $= 2\ell$, so

$$\frac{X_n}{n} \longrightarrow \frac{2}{\pi} \text{ as } n \to \infty.$$

Needle Throws



$$\frac{2}{\pi} = 0.63662$$

Later we will learn about the rate of convergence and the typical size of fluctuation about the limiting value.

Fundamentals

A Random Experiment

Sample k elements from some set S. (Recall counting)

Definition: The set of all possible outcomes is called the sample space and is denoted by Ω.

Example: Experiment = Toss a coin twice.
Sample twice from $S = \{H, T\}$ with replacement.

$$\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$$

Definition: An event in an experiment is a subset of Ω.

Example: Experiment = Toss a coin twice.

The event E_1 = "H in the first toss" $= \{(H,H),(H,T)\} \subset \Omega$.

The event E_2 = "observe different sides" $= \{(H,T),(T,H)\} \subset \Omega$.

Discrete Probability Space

- **Key question:** What is the probability of an event?
- Mathematically, given an event $E \subseteq \Omega$, assign a number $\mathbb{P}(E) \in [0,1]$.
- $\mathbb{P}$ is called a probability measure.
- **Technical details** (not required for CS70):
  - $\mathbb{P}: \mathcal{F} \to [0,1]$ is a function on a certain set $\mathcal{F}$ (called σ-field) of subsets of Ω.
  - We first need to define $(\Omega, \mathcal{F})$, which imposes constraints on possible probability measures that can be defined on $\mathcal{F}$ in a mathematically consistent manner.

Definition (CS70 version): Discrete probability space $(\Omega, \mathbb{P})$.

1. Sample space Ω.
2. A probability $\mathbb{P}(\{\omega\})$ for each $\omega \in \Omega$ such that
   i. $\mathbb{P}(\{\omega\}) \in [0,1], \forall \omega \in \Omega$.
   ii. $\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = 1$.

Given an event $E \subseteq \Omega$, we have $\mathbb{P}(E) = \sum_{\omega \in E} \mathbb{P}(\{\omega\})$.

Rolling a Fair Die

Example: Experiment = Roll a fair die twice.

- Sample twice from $S = \{1, 2, 3, 4, 5, 6\}$ with replacement.
- The sample splace is $\Omega = \{(a, b): a, b \in S\}$.
- Event E = "sum $\equiv 0 \pmod 4$". What is $\mathbb{P}(E)$?

$$E = \{(1,3), (3,1), (2,2), (2,6), (6,2), (3,5), (5,3), (4,4), (6,6)\}$$

- $|\Omega| = 36.$
- $\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|} = \frac{1}{36}, \forall \omega \in \Omega$
- $\mathbb{P}(E) = \sum_{\omega \in E} \mathbb{P}(\{\omega\}) = \frac{9}{36} = \frac{1}{4}$

In general, for finite Ω, if all $\omega \in \Omega$ are equally likely, then $\mathbb{P}(E) = \frac{|E|}{|\Omega|}$.

Biased Coin Tosses

Example: A biased coin with probability of Heads $= p \in [0,1]$ (probability of Tails $= 1 - p$) is tossed n times.

- $S = \{H, T\}$.
- $\Omega = \{$All length-n strings over $S\}$.
- $|\Omega| = 2^n$.
- Event $E_k =$ "get k Heads". What is $\mathbb{P}(E_k)$?

$$E_k = \{\text{All length-}n\text{ strings over } S \text{ with } k \;\; H\text{s and } n-k \;\; T\text{s}\}.$$

- Recall counting Case 3: Sample k positions from $\{1, \dots, n\}$ without replacement and ignore order. Put Hs at those positions and Ts at the remaining positions.
- $|E_k| = \binom{n}{k}$.
- For all $\omega \in E_k$, $\mathbb{P}(\{\omega\}) = p^k (1-p)^{n-k}$.
- $\mathbb{P}(E_k) = \sum_{\omega \in E_k} \mathbb{P}(\{\omega\}) = \binom{n}{k} p^k (1-p)^{n-k}$.

We will explain why this is true when we discuss independence later.

This is called the Binomial distribution.

Biased Coin Tosses

- Event E_k = "get k Heads"
- $\mathbb{P}(E_k) = \sum_{\omega \in E_k} \mathbb{P}(\{\omega\}) = \binom{n}{k} p^k (1-p)^{n-k}$

$$\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = \sum_{k=0}^{n} \mathbb{P}(E_k)$$

$$= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1$$

Why does this equality hold?

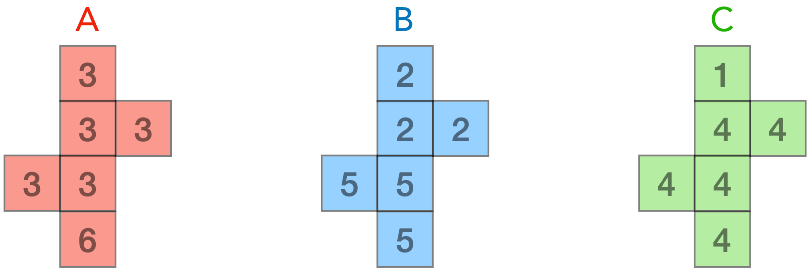
Every $\omega \in \Omega$ belongs to exactly one E_k and $E_0 \cup E_1 \cup \cdots \cup E_n = \Omega$.

By the Binomial Theorem from Lecture 15:

Theorem 1 (Binomial Theorem). *For all $n \in \mathbb{N}$,*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Probabilistic Rock-Paper-Scissors



- **Bob** chooses a die first.
- **Alice** then chooses a die from the remaining two dice.
- Each person rolls their die and the person with a higher number wins the round.
- 11 rounds will be played with the same chosen dice.
- Who has the advantage?

Event "A beats B" $= \{(6,5),(6,2),(3,2)\}$

$$\mathbb{P}[\{(6,5)\}] = \frac{1}{6}\times\frac{3}{6} = \frac{1}{12} = \mathbb{P}[\{(6,2)\}]$$

$$\mathbb{P}[\{(3,2)\}] = \frac{5}{6}\times\frac{3}{6} = \frac{5}{12}$$

$$\mathbb{P}[\text{"A beats B"}] = \frac{7}{12} \approx 0.58$$

Event "B beats C" $= \{(5,4),(5,1),(2,1)\}$

$$\mathbb{P}[\text{"B beats C"}] = \frac{1}{2}+\frac{1}{2}\times\frac{1}{6} = \frac{7}{12} \approx 0.58$$

Event "C beats A" $= \{(4,3)\}$

$$\mathbb{P}[\text{"C beats A"}] = \frac{5}{6}\times\frac{5}{6} = \frac{25}{36} \approx 0.69$$

Non-transitive Dice

Die A

$\swarrow \qquad \nwarrow$

Die B > Die C

Probabilistic Rock-Paper-Scissors

$$\mathbb{P}[\text{Alice wins the game}] = \sum_{k=6}^{11} \binom{11}{k} p^k (1-p)^{11-k}$$

where $p =$ probability of winning a single round.

- If Bob chooses die A, Alice can choose die C
  $$p = \frac{25}{36} \implies \mathbb{P}[\text{Alice wins the game}] \approx 0.915$$
- If Bob chooses die B, Alice can choose die A
  $$p = \frac{7}{12} \implies \mathbb{P}[\text{Alice wins the game}] \approx 0.715$$
- If Bob chooses die C, Alice can choose die B
  $$p = \frac{7}{12} \implies \mathbb{P}[\text{Alice wins the game}] \approx 0.715$$

Event "A beats B" $= \{(6,5),(6,2),(3,2)\}$

$$\mathbb{P}[\{(6,5)\}] = \frac{1}{6} \times \frac{3}{6} = \frac{1}{12} = \mathbb{P}[\{(6,2)\}]$$

$$\mathbb{P}[\{(3,2)\}] = \frac{5}{6} \times \frac{3}{6} = \frac{5}{12}$$

$$\mathbb{P}[\text{“A beats B”}] = \frac{7}{12} \approx 0.58$$

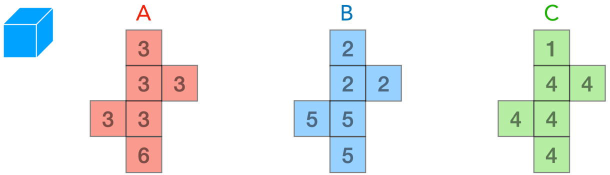
Event "B beats C" $= \{(5,4),(5,1),(2,1)\}$

$$\mathbb{P}[\text{“B beats C”}] = \frac{1}{2} + \frac{1}{2} \times \frac{1}{6} = \frac{7}{12} \approx 0.58$$

Event "C beats A" $= \{(4,3)\}$

$$\mathbb{P}[\text{“C beats A”}] = \frac{5}{6} \times \frac{5}{6} = \frac{25}{36} \approx 0.69$$

What would happen as the number of rounds per game increases?

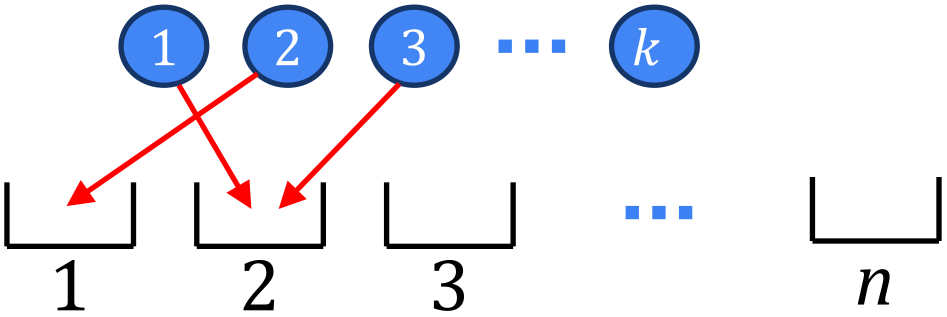


Die A

$<$ $>$

Die B $>$ Die C

Probabilistic Balls and Bins

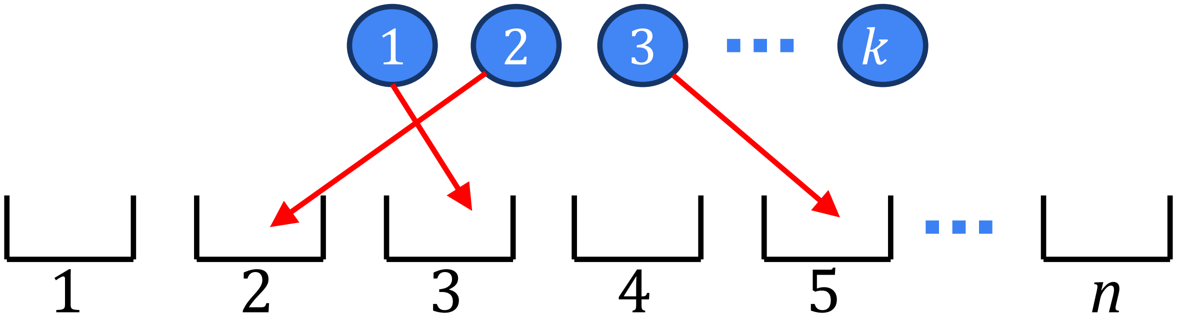


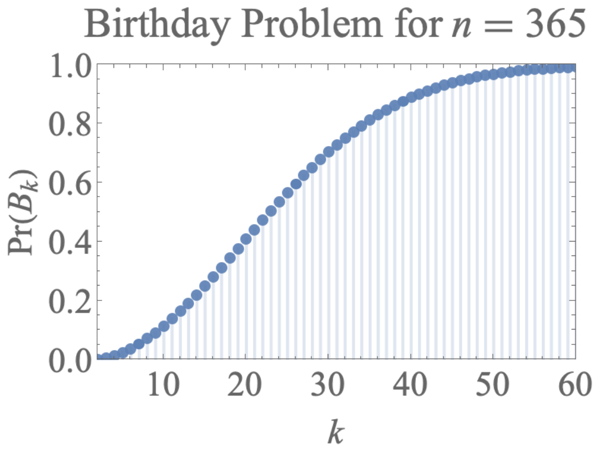
Throw k labeled balls into n labeled bins.

- **Assume:** Probability that ball i lands in bin j $= \frac{1}{n}$ for all $i \in \{1, \dots, k\}$ and for all $j \in \{1, \dots, n\}$.
- Sample k times **with replacement** from $S = \{1, \dots, n\}$.
- $\Omega = \{(a_1, \dots, a_k): a_i \in \{1, \dots, n\} \text{ for all } i\}$
- $|\Omega| = n^k$ and $\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|} = \frac{1}{n^k}$ for all $\omega \in \Omega$.
- Event E_j = "Bin j is empty" $= \{(a_1, \dots, a_k) \in \Omega: a_i \neq j \text{ for all } i = 1, \dots, k\}$
- $\mathbb{P}(E_j) = \frac{|E_j|}{|\Omega|} = \frac{(n-1)^k}{n^k} = \left(1 - \frac{1}{n}\right)^k$. Note that this does not depend on j.

Birthday Paradox

- B_k = event that at least two people have the same birthday in a group of k people.
- $B_k^c = \Omega \backslash B_k$ (no collision)





$$\mathbb{P}(B_k^c) = \frac{|B_k^c|}{|\Omega|} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} \approx e^{-\binom{k}{2}\frac{1}{n}} \text{ for large } n.$$

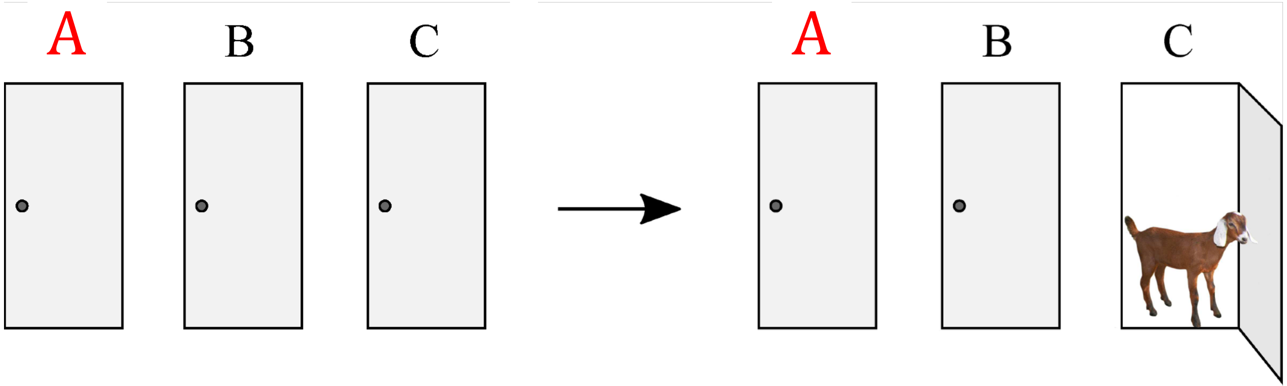
For $n = 365$, $\mathbb{P}(B_{23}) \approx 0.507$

$\mathbb{P}(\Omega) = \mathbb{P}(B_k) + \mathbb{P}(B_k^c) = 1$, so

$$\mathbb{P}(B_k) = 1 - \mathbb{P}(B_k^c) = 1 - \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} \approx 1 - e^{-\binom{k}{2}\frac{1}{n}} \text{ for large } n.$$

Monty Hall

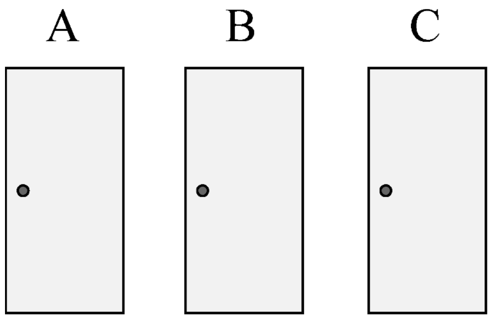
- A famous game show from 1970s, hosted by Monty Hall.
- One door has a car behind it. The other two doors have goats.



Question: Does the contestant have a better chance of winning if he/she switches doors?

- The contestant picks a door (A in this example), but doesn't open it.
- Monty Hall's assistant (Carol) opens one of the two remaining doors to reveal a goat (since Carol knows where the prize is, she can always do this).
- The contestant is then given the option of sticking with his/her current door, or switching to the other unopened one.

Monty Hall



Notation: (Prize door, Contestant's choice, Carol's choice)

$$\Omega = \{(A, A, B), (A, A, C), (B, B, A), (B, B, C), (C, C, A), (C, C, B),$$
$$(A, B, C), (A, C, B), (B, A, C), (B, C, B), (C, A, B), (C, B, A)\}$$

- For each $\omega \in \Omega$ in the first row, $\mathbb{P}(\{\omega\}) = \frac{1}{3} \times \frac{1}{3} \times \frac{1}{2} = \frac{1}{18}$
- For each $\omega \in \Omega$ in the second row, $\mathbb{P}(\{\omega\}) = \frac{1}{3} \times \frac{1}{3} \times 1 = \frac{1}{9}$
- So, the contestant would be better off switching, since $\frac{\mathbb{P}[\{(B,A,C)\}]}{\mathbb{P}[\{(A,A,C)\}]} = 2.$