CS145: Probability & Computing

Lecture 1: Sets & Events, Counting & Combinatorics

Instructors: Alessio Mazzetto and Eli Upfal

Figure credits:
*Bertsekas & Tsitsiklis, **Introduction to Probability**, 2008*
*Pitman, **Probability**, 1999*

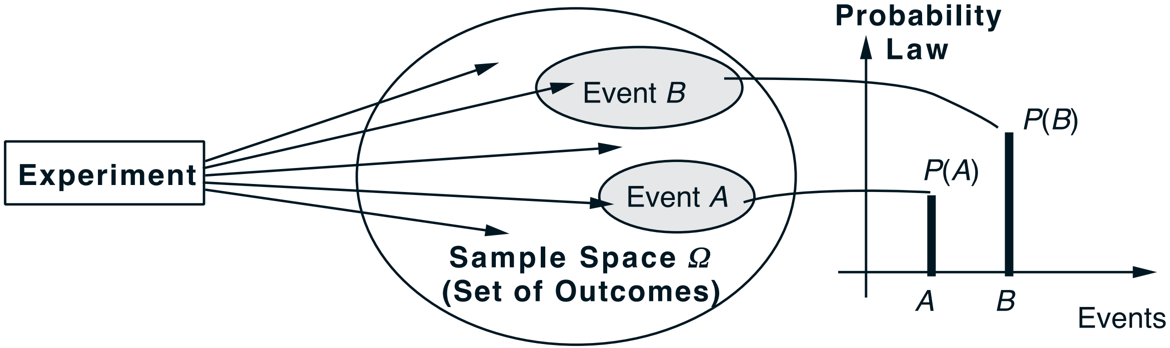
The Birthday "Paradox"

Suppose there are *m* students in a class.
What is the probability that at least two students in the class have the same birthday?

In a class of 70 students, this probability is about 99.87%

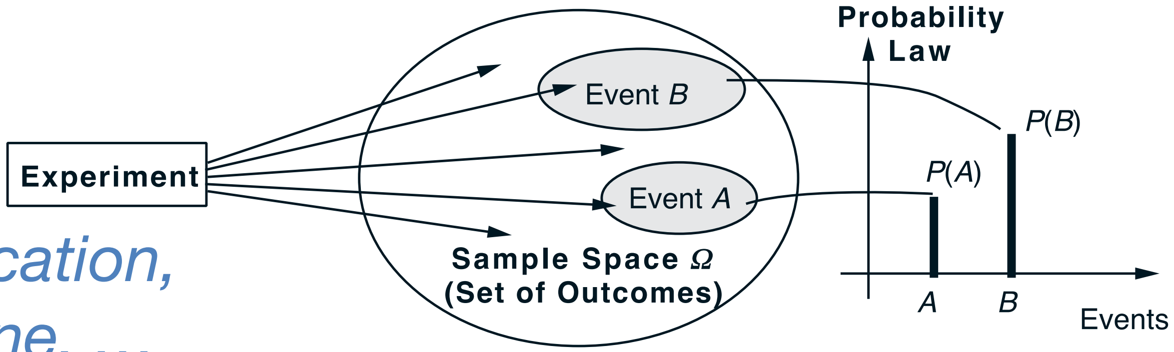
CS145: Lecture 1 Outline

- ➢ Sample spaces: Sets of possible outcomes
- ➢ Probability: Counting and the Discrete Uniform Law
- ➢ Example: The birthday paradox



Defining a Probabilistic Space/Model

flip a coin,
roll a die,
take a medication,
test an engine, …



Elements of a Probabilistic Model

- The **sample space** Ω, which is the set of all possible outcomes of an experiment.
- The **probability law**, which assigns to a set A of possible outcomes (also called an **event**) a nonnegative number $\mathbf{P}(A)$ (called the **probability** of A) that encodes our knowledge or belief about the collective "likelihood" of the elements of A. The probability law must satisfy certain properties to be introduced shortly.

Background: Sets

- A set is a collection of objects, which are elements of the set.
- A set can be finite, $S = \{1, 2, \ldots, n\}$. Cardinality (size): $|S| = n$
- A set can be countably infinite:

$$\begin{aligned} S &= \{x \mid x = 2k+1 \text{ or } x = -2k+1, k \text{ integer}\} \\ &= \{1, -1, 3, -3, 5, -5, \ldots\}. \end{aligned}$$

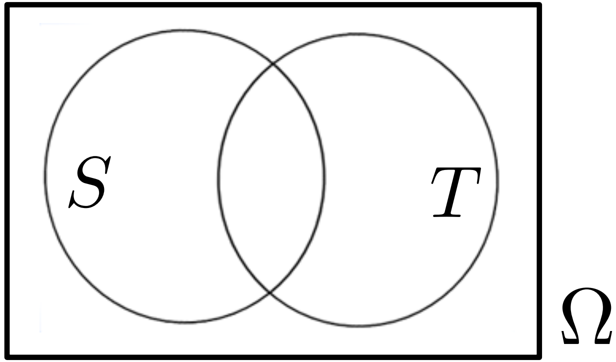
- A set can be uncountable, $S = \{x \mid x \in [0, 1]\}$.
- A set can be empty $S = \emptyset$.

Sets: Elements & Relationships

- $x \in S$ - the element x is a member of the set S
- $x \notin S$ - the element x is not a member of the set S
- $\exists x$ - there exists x...
- $\forall$ - for all elements x ...
- $T \subseteq S$ - $\forall x \in T, x \in S$
- $T \subset S$ - $\forall x \in T, x \in S$ AND $\exists x \in S$ such that $x \notin T$.

Sets: Combination & Manipulation

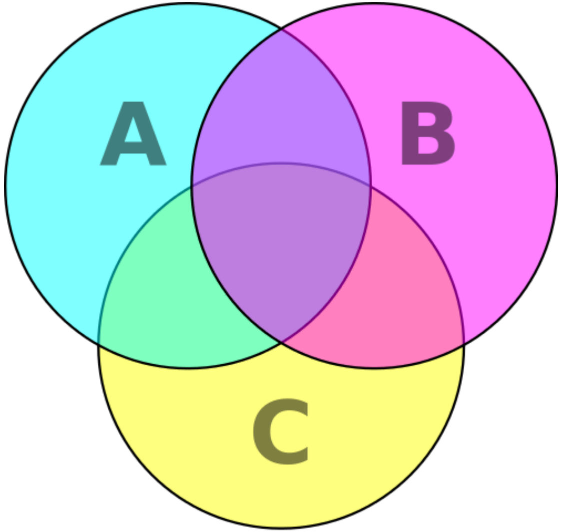
- A base set Ω, all sets are subsets of Ω
- Basic operations: for $S, T \subseteq \Omega$,
  - $S \cup T = \{x \mid x \in S \text{ or } x \in T\}$ ➢ *union*
  - $S \cap T = \{x \mid x \in S \text{ and } x \in T\}$ ➢ *intersection*
  - $\bar{S} = S^c = \{x \mid x \notin S\}$ ➢ *complement*
- De Morgan's laws:
  - $(S \cup T)^c = \bar{S} \cap \bar{T}$
  - $(S \cap T)^c = \bar{S} \cup \bar{T}$

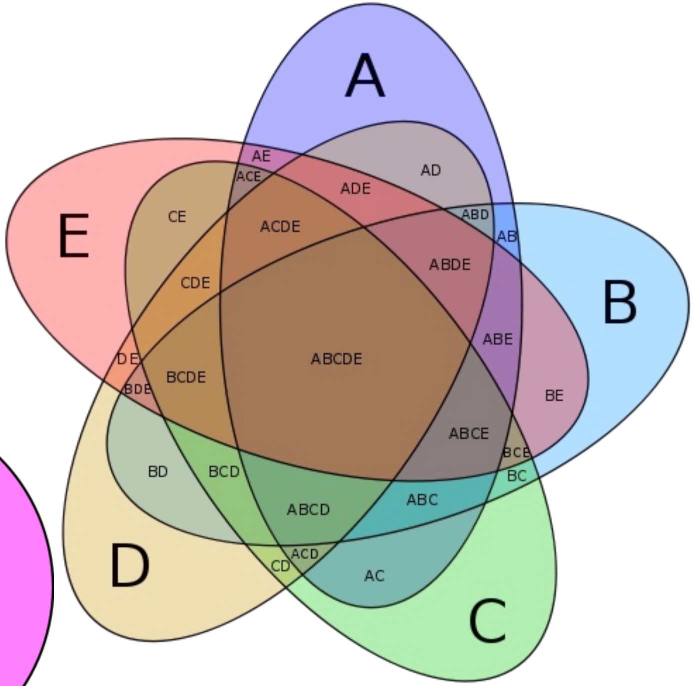


Venn Diagram

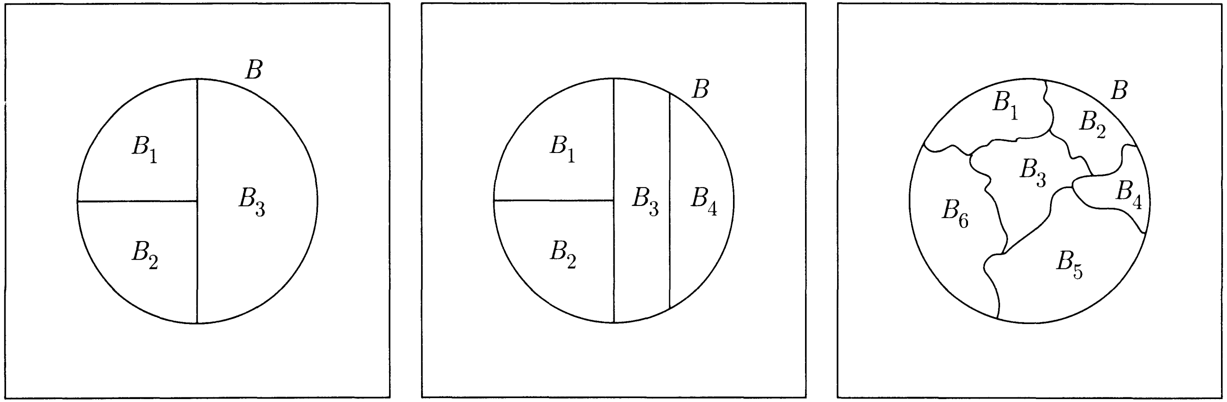
Visualizing Sets: Venn Diagrams

- Basic operations: for $S, T \subseteq \Omega$,
  - $S \cup T = \{x \mid x \in S \text{ or } x \in T\}$
  - $S \cap T = \{x \mid x \in S \text{ and } x \in T\}$
  - $\bar{S} = S^c = \{x \mid x \notin S\}$
- De Morgan's laws:
  - $(S \cup T)^c = \bar{S} \cap \bar{T}$
  - $(S \cap T)^c = \bar{S} \cup \bar{T}$
  - $\left(\bigcup_{i \in I} S_i\right)^c = \bigcap_{i \in I} \bar{S}_i$
  - $\left(\bigcap_{i \in I} S_i\right)^c = \bigcup_{i \in I} \bar{S}_i$





Partitions of a Set



A set B is *partitioned* into *n* subsets if:

$$B_1 \cup B_2 \cup \cdots \cup B_n = B$$

$$B_i \cap B_j = \emptyset \text{ for any } i \neq j$$

mutually disjoint

The Sample Space

$\Omega =$ a set of possible outcomes of some random (not deterministic) experiment

"Omega"

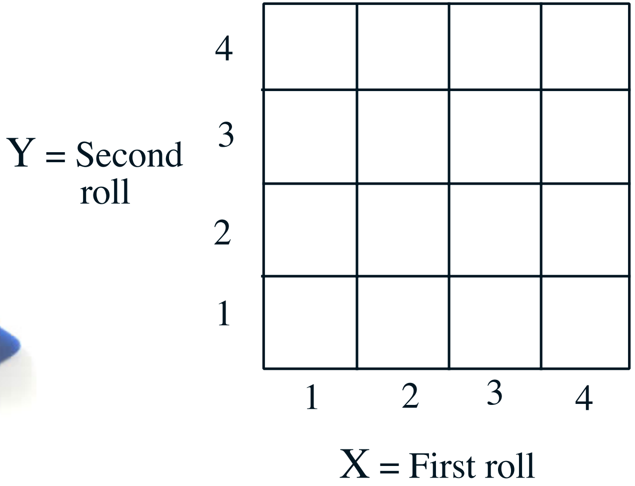
The list defining the sample space must be:

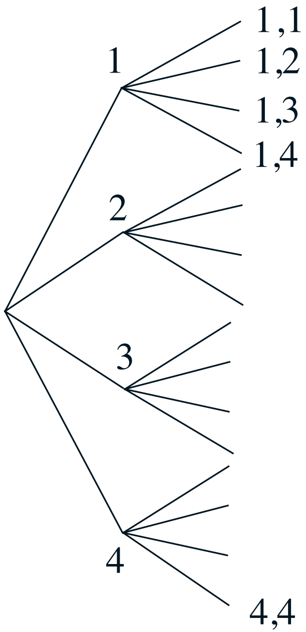
- *Mutually exclusive:* Each outcome of the experiment is represented once in the sample space.
- *Collectively exhaustive:* All outcomes are elements of the sample space.
- *An art:* Choosing the "right" granularity, to capture the phenomenon of interest as simply as possible.
  Modeling in science and engineering involves tradeoffs between accuracy, simplicity, & tractability.

A Finite Sample Space

You roll a tetrahedral (4-sided) die 2 times.

$$\Omega = \{(x, y) \mid x \in \{1, 2, 3, 4\}, y \in \{1, 2, 3, 4\}\}$$





- Formally, sample space is a set of $4^2=16$ discrete outcomes
- Can also model outcome via *tree-based sequential description*

Alternative Samples Spaces

You roll a tetrahedral (4-sided) die 2 times.

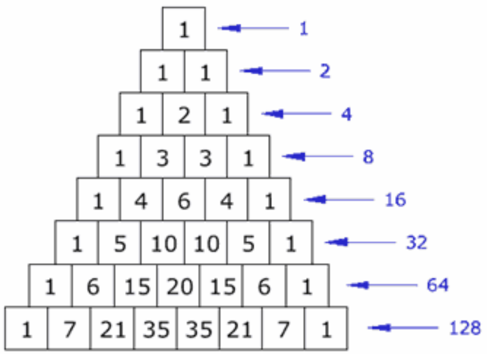
$$\Omega = \{(x, y) \mid x \in \{1, 2, 3, 4\}, y \in \{1, 2, 3, 4\}\}$$

$$\Omega_{sum} = \{x + y \mid x \in \{1, 2, 3, 4\}, y \in \{1, 2, 3, 4\}$$

- Sample space Ω is a set of *4^2=16* discrete outcomes
- Sample space Ω_{sum} = {2,3,4,5,6,7,8} is a set of 7 discrete outcomes
- The two sample spaces represent the same experiment
- Elements in Ω_{sum} don't have the same "probability"

CS145: Lecture 1 Outline

- ➢ Sample spaces: Sets of possible outcomes
- ➢ Probability: Counting and the Discrete Uniform Law
- ➢ Example: The birthday paradox



The Discrete Uniform Law (Combinatorics)

Formalizes the idea of "uniformly random" sampling.

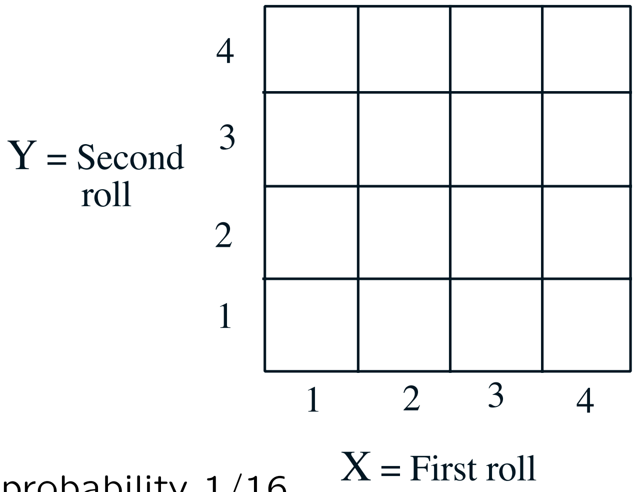
- Let all outcomes be equally likely
- Then,

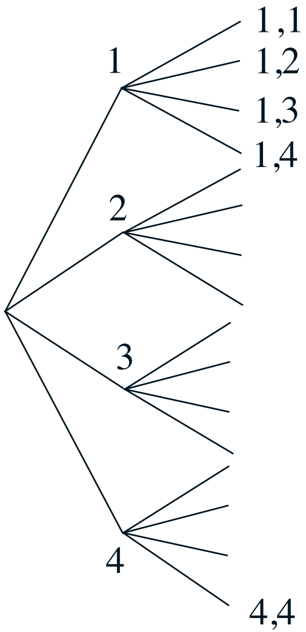
$$\mathbf{P}(A) = \frac{\text{number of elements of } A}{\text{total number of sample points}}$$

- Computing probabilities $\equiv$ counting
- Defines fair coins, fair dice, well-shuffled decks

Uniform Law for a Finite Sample Space

You roll a tetrahedral (4-sided) die 2 times.





- Let every possible outcome have probability 1/16
  - $\mathbf{P}((X,Y)$ is (1,1) or (1,2)$) =$
  - $\mathbf{P}(\{X = 1\}) =$
  - $\mathbf{P}(X + Y$ is odd$) =$
  - $\mathbf{P}(\min(X,Y) = 2) =$

The Basic Counting Principle

Consider a process that consists of r stages. Suppose that:

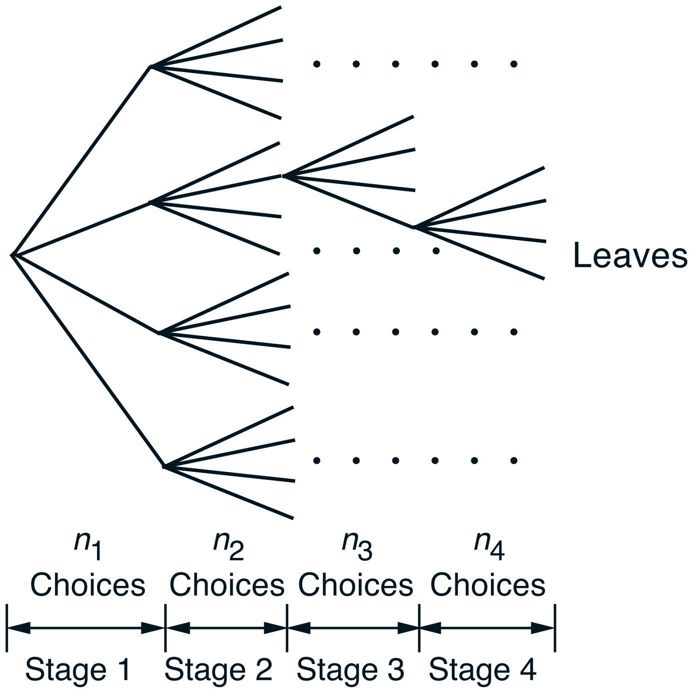
(a) There are n_1 possible results for the first stage.

(b) For every possible result of the first stage, there are n_2 possible results at the second stage.

(c) More generally, for all possible results of the first $i-1$ stages, there are n_i possible results at the ith stage.

Then, the total number of possible results of the r-stage process is

$$\boxed{n_1 \cdot n_2 \cdots n_r.}$$



Simple examples:

- Number of license plates with 3 letters and 4 digits $=$

- . . . if repetition is prohibited $=$

The set of choices at each stage can depend on previous choices, as long as the number of choices at each stage is constant.

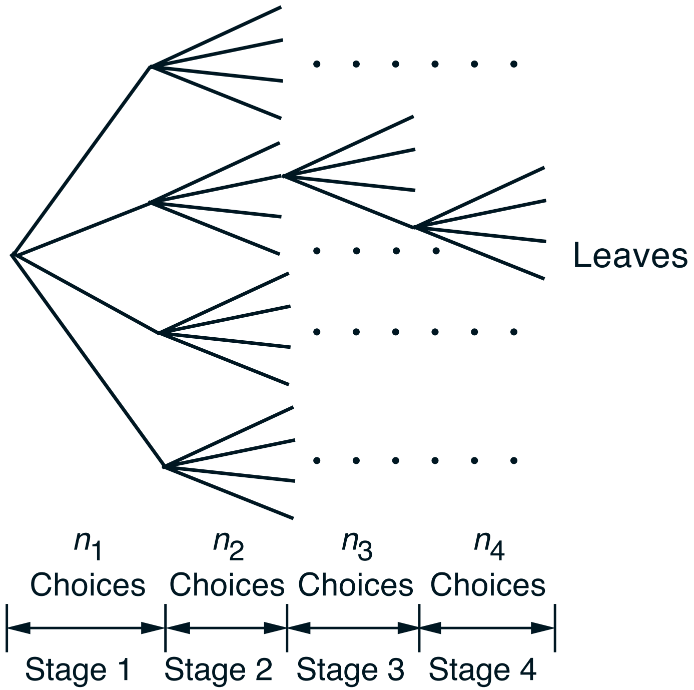
Permutations and Subsets

- **Permutations:** Number of ways of ordering n elements is:

$$n(n-1)(n-2)\cdots 1 = \prod_{i=1}^{n} i = n!$$

- Number of subsets of $\{1, \ldots, n\}$ =

$$\underbrace{2 \cdot 2 \cdots 2}_{n \text{ times}} = 2^n$$



Combinations

- $\binom{n}{k}$: number of k-element subsets of a given n-element set

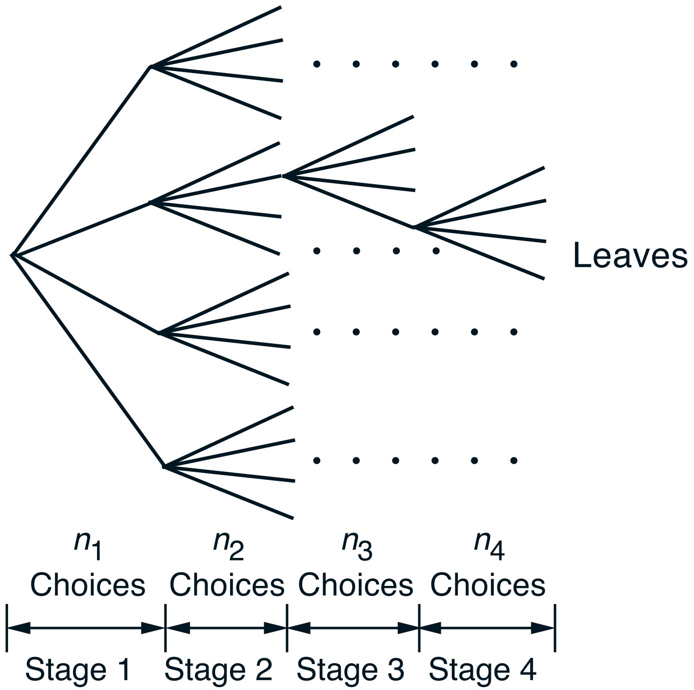
"n choose k"

- Two ways of constructing an ordered sequence of k **distinct** items:
  - Choose the k items one at a time:
    $n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$ choices
  - Choose k items, then order them ($k!$ possible orders)

- Hence:

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

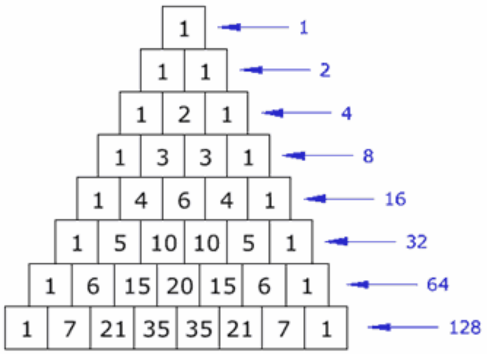
$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$



- The total number of subsets:

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n$$

Pascal's Triangle



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

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n$$

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

Pascal’s Triangle

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1
1 7 21 35 35 21 7 1
1 8 28 56 70 56 28 8 1
1 9 36 84 126 126 84 36 9 1
1 10 45 120 210 252 210 120 45 10 1
1 11 55 165 330 462 462 330 165 55 11 1
1 12 66 220 495 792 924 792 495 220 66 12 1
1 13 78 286 715 1287 1716 1716 1287 715 286 78 13 1
1 14 91 364 1001 2002 3003 3432 3003 2002 1001 364 91 14 1
1 15 105 455 1365 3003 5005 6435 6435 5005 3003 1365 455 105 15 1
1 16 120 560 1820 4368 8008 11440 12870 11440 8008 4368 1820 560 120 16 1

Painting of Pascal made by François II Quesnel in 1691

http://www.mathwarehouse.com/

Binomial Probabilities

If I toss a coin n times, what is the probability that I see k heads?

- n independent coin tosses
  - $\mathbf{P}(H) = p = \frac{1}{2}$

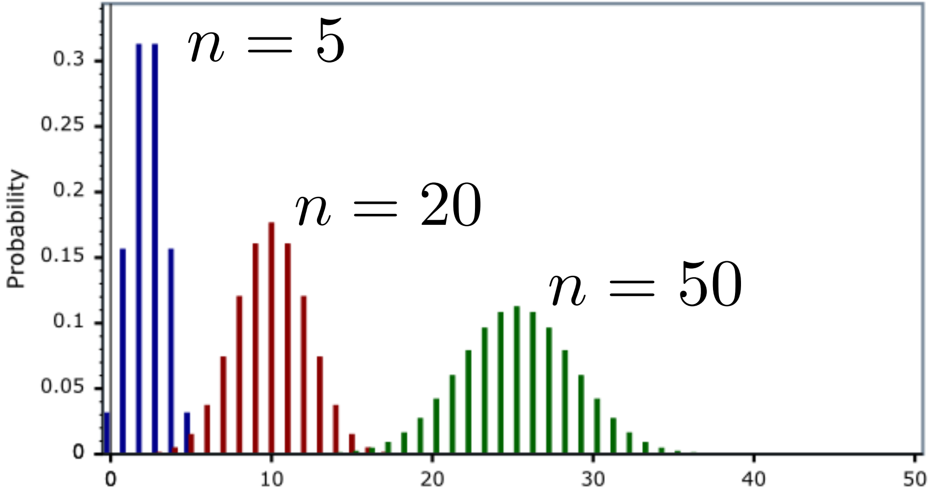
- $\mathbf{P}(HTTHHH) =$

- $\mathbf{P}(\text{sequence}) = p^{\#\text{ heads}}(1-p)^{\#\text{ tails}}$

$$\mathbf{P}(k\text{ heads}) = \sum_{k-\text{head seq.}} \mathbf{P}(\text{seq.})$$

$$= (\#\text{ of } k-\text{head seqs.}) \cdot p^k(1-p)^{n-k}$$

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



CS145: Lecture 1 Outline

- Sample spaces: Sets of possible outcomes
- Probability: Counting and the Discrete Uniform Law
- Example: The birthday paradox

The Birthday "Paradox"

Suppose there are *m* students in a class.
What is the probability that at least two students in the class have the same birthday?

In a class of 70 students, this probability is about 99.87%

Assumptions:

- Birthdays are equally likely to occur on any of *N*=*365* days
- No dependence between birthdays of different students

Not completely true, but fairly accurate approximations.

The Birthday "Paradox"

- N possible birthdays, m students in the class.
- We can compute this probability by counting elements of the *sample space of all N^m possible birthday patterns*:

$$\Omega = \{(b_1, \ldots, b_m) \mid b_i \in \{1, \ldots, N\}\}$$

- The number of *birthday patterns with all pairs distinct*:

$$D_m = \{(b_1, \ldots, b_m) \mid b_i \neq b_j \text{ for all } i \neq j\}$$

$$|D_m| = N(N-1)(N-2)\cdots(N-m+1) = \frac{N!}{(N-m)!}$$

The Birthday "Paradox"

Sample Space: $\Omega = \{(b_1, \ldots, b_m) \mid b_i \in \{1, \ldots, N\}\}$

Set of items in the sample space with all distinct birthdays:

$$D_m = \{(b_1, \ldots, b_m) \mid b_i \neq b_j \text{ for all } i \neq j\}$$

In the uniform probability model:

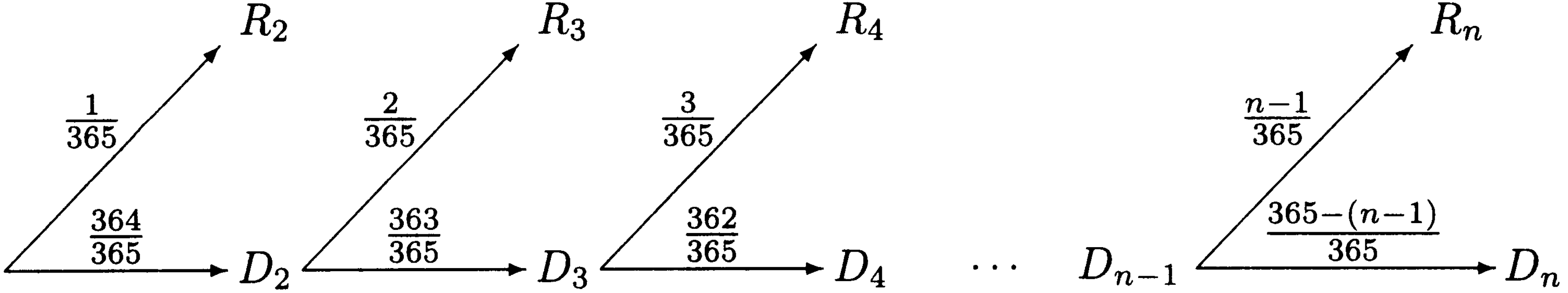
$$\frac{|Dm|}{|\Omega|} = P(D_m) = \frac{N!}{(N-m)!N^m} = \prod_{i=0}^{m-1} \left(\frac{N-i}{N}\right)$$

The Birthday "Paradox"

- Sort students in arbitrary order (say, alphabetical by name)
- Define two events for student *j* in the list of *m* students:

$$R_j \rightarrow \text{ birthday } j \text{ is a repeat of some previous student}$$
$$D_j \rightarrow \text{ all of the first } j \text{ birthdays are distinct}$$

- Tree diagram of event probabilities for *N=365* days:



The Birthday "Paradox"

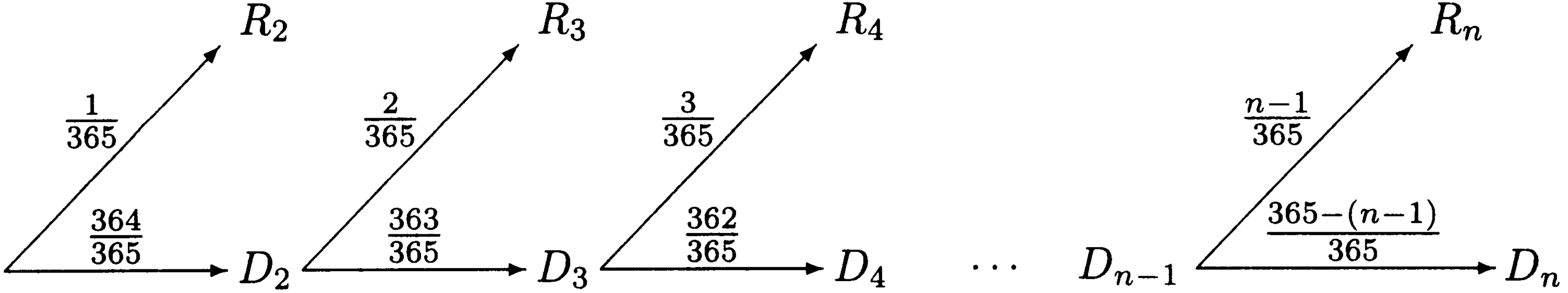
- The probability that the first *m* birthdays are distinct is then:

$$P(D_m) = \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right)\cdots\left(1 - \frac{m-1}{N}\right) = \prod_{i=0}^{m-1}\left(1 - \frac{i}{N}\right)$$

$R_j \rightarrow$ birthday j is a repeat of some previous student

$D_j \rightarrow$ all of the first j birthdays are distinct

- Tree diagram of event probabilities for *N=365* days:



The Birthday "Paradox"

- The probability that the *m* birthdays are distinct:

$$P(D_m) = \frac{N!}{(N-m)!N^m} = \prod_{i=0}^{m-1} \left(\frac{N-i}{N} \right)$$

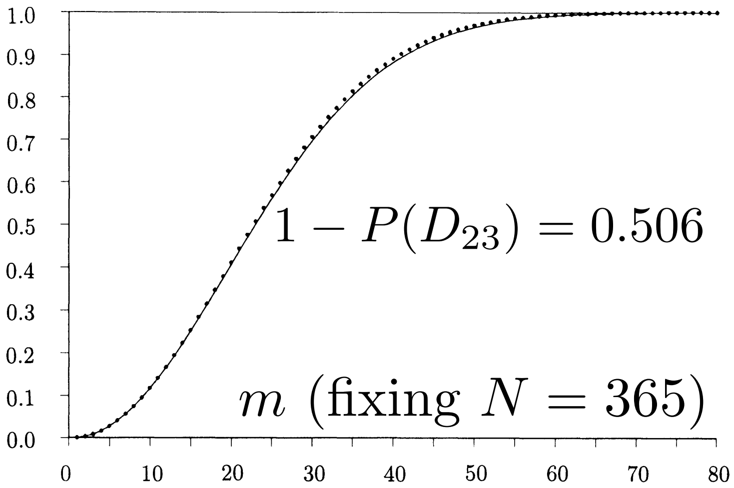
$$P(D_m) = \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right) \cdots \left(1 - \frac{m-1}{N}\right) = \prod_{i=0}^{m-1} \left(1 - \frac{i}{N}\right)$$

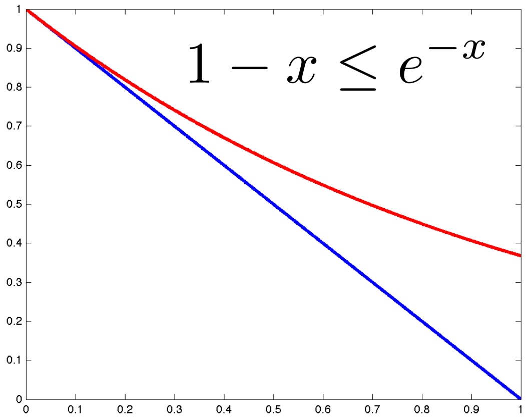
The Birthday "Paradox"

➢ The probability that *m* birthdays on *N* days are distinct:

$$P(D_m) = \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right)\cdots\left(1 - \frac{m-1}{N}\right) = \prod_{i=0}^{m-1}\left(1 - \frac{i}{N}\right)$$

$1 - P(D_m)$



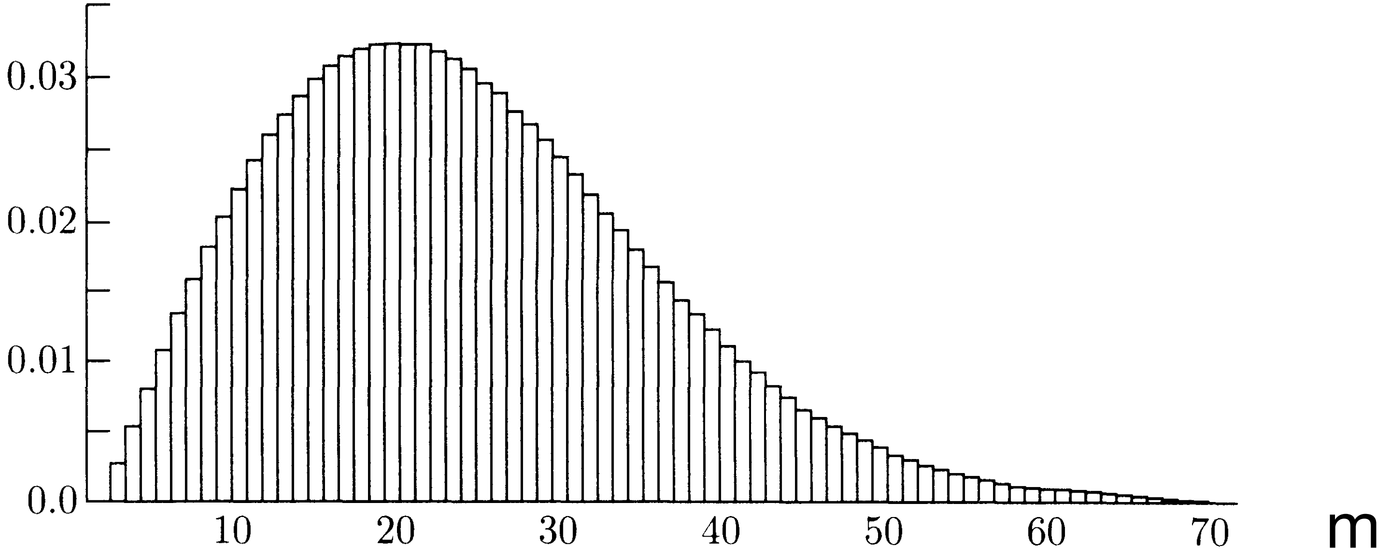


(curve) $$P(D_m) \leq \prod_{i=0}^{m-1} e^{-\frac{i}{N}} = e^{-\sum_{i=0}^{m-1}\frac{i}{N}} = e^{-\frac{m(m-1)}{2N}}$$

The Birthday "Paradox"

- The probability that *m* birthdays on *N* days are distinct:

$$P(D_m) = \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right)\cdots\left(1 - \frac{m-1}{N}\right) = \prod_{i=0}^{m-1}\left(1 - \frac{i}{N}\right)$$

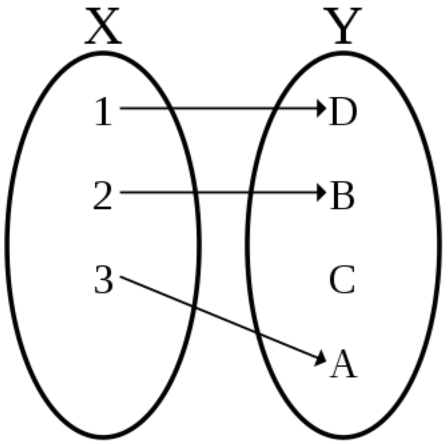


Probability that first repeated birthday is found with student *j*

Countable and Uncountable Infinite

Advanced topic not covered in homeworks or exams!

- A finite set is countable
- The set of natural numbers N = {1, 2, 3, ...} is countable
- A set S is countable if there is injective mapping from S to N
- The set of all integers is countable 1, -1, 2, -2, 3, -3,.....
- The set of all rationales is countable

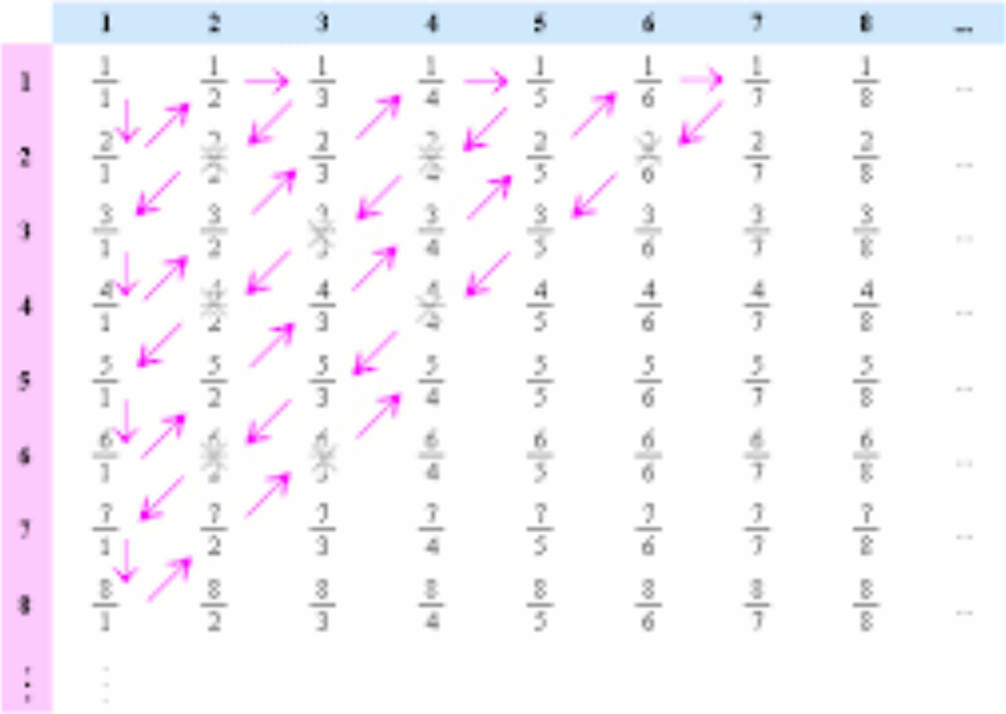


The Rational numbers in [0,1] is Countable

Advanced topic not covered in homeworks or exams!

- Each rational number $\frac{a}{b}$ in [0,1] has a place in the table

- There is an order in the table that associates each rational number with a natural number.



Source: www.homeschoolmath.net/teaching/rational-numbers-countable.php

The Real Numbers in [0,1] are Uncountable

Advanced topic not covered in homeworks or exams!

Cantor's diagonal argument

- Assume that the real numbers in [0,1] were countable.
- Write each real number in [0,1] in binary fraction
- If countable can be ordered in a list
- Construct a real number s such that its i-th bit is the complement the i-th bit of the i-th number in the list.
- The number s doesn't appear in the list. For i>0 the number s is different from the i-th number in its i-th bit.

$$
\begin{array}{ll}
s_1 & = 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\ldots \\
s_2 & = 1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\ldots \\
s_3 & = 0\,1\,0\,1\,0\,1\,0\,1\,0\,1\,0\ldots \\
s_4 & = 1\,0\,1\,0\,1\,0\,1\,0\,1\,0\,1\ldots \\
s_5 & = 1\,1\,0\,1\,0\,1\,1\,0\,1\,0\,1\ldots \\
s_6 & = 0\,0\,1\,1\,0\,1\,1\,0\,1\,1\,0\ldots \\
s_7 & = 1\,0\,0\,0\,1\,0\,0\,0\,1\,0\,0\ldots \\
s_8 & = 0\,0\,1\,1\,0\,0\,1\,1\,0\,0\,1\ldots \\
s_9 & = 1\,1\,0\,0\,1\,1\,0\,0\,1\,1\,0\ldots \\
s_{10} & = 1\,1\,0\,1\,1\,1\,0\,0\,1\,0\,1\ldots \\
s_{11} & = 1\,1\,0\,1\,0\,1\,0\,0\,1\,0\,0\ldots \\
\vdots & \quad \vdots\,\vdots\,\vdots\,\vdots\,\vdots\,\vdots\,\vdots\,\vdots\,\vdots\,\vdots\,\vdots\,\ddots
\end{array}
$$

$$
s = 1\,0\,1\,1\,1\,0\,1\,0\,0\,1\,1\ldots
$$

https://en.wikipedia.org/wiki/Cantor%27s_diagonal_argument

Discrete vs. Continuous Spaces

- Countable sample space $\longrightarrow$ discrete probability model
- Uncountable sample space $\longrightarrow$ continuous probability model