Week 1

Andrew Yao (姚期智)

2024-2-27

Math for CS and AI, Spring 2024

Teachers:

Weeks 1-8, math for general CS, Andrew Yao (姚期智)

Weeks 9-16, math focused on AI, Jingzhou Zhang (张景昭)

TAs: Homework sessions/Grading

Format: Weekly homework sets – 50%

Midterm exam – 25%

Final exam – 25%

- All exams are in-class, and open-book

Contents for 1st Half (Weeks 1-8)

Tentative Plan:

- Probability Theory: 3 weeks
- Graphs/Combinatorics: 2 weeks
- Geometry/Advanced Topics: 2 weeks

  -- complexity, geometry, topology

- Midterm exam: April 15
- Reference book:

  Discrete Mathematics, by Lovasz, Pelikan and Vesztergombi, 2003;

  elementary, supplemental reading (not required)

- Other reading materials as needed

Probability Theory

In the face of uncertainties, we often need to estimate how likely something occurs.

Throw a pair of unbiased dice, let the result be (i, k)

- Question: *What is the probability that i+k = 8?*

  Intuitively, there are 36 possible values of (i,k), all equally likely to occur.

- There are 5 of these that satisfy i+k=8.

  Thus, the probability must be 5/36 $\approx$14%

- How about more complex questions? Say, if one performs the above experiment 100 times, what is the probability that the outcome i+k=8 occured 35 times?
- As the question gets more complicated, we need a precise mathematical definition of what probability means!

Definition: A *probability space* **P** = (U, *p*) consists of:

-- universe U : finite non-empty set

-- probability function *p* : U → [0,1] such that $\sum_{u \in U} p(u) = 1$

An event is T ⊆ U

The *probability of* T is defined to be Pr{T} = $\sum_{u \in T} p(u)$

* Intuition of event T: Pick a random point u in U according to p,

Pr{T} is the chance that u falls into subset T

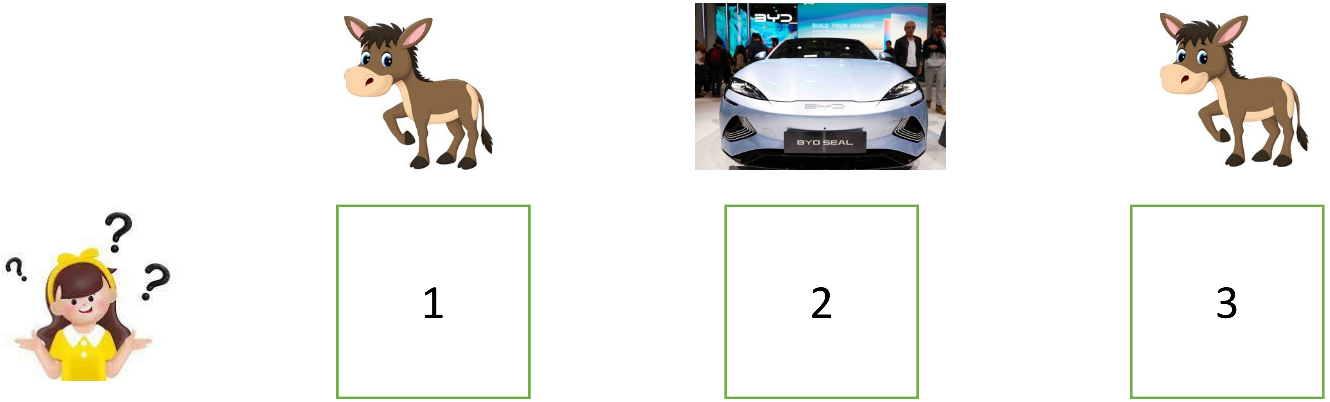
For throwing two unbiased dice, formalize it as:

- $U = \{ (i,k) \mid 1 \le i,k \le 6 \}$
- $T = \{ (i,k) \mid 1 \le i,k \le 6, \; i+k=8 \}$
- $p(u) = 1/|U|$ for all $u \in U$. $\Pr\{T\}=|T|/|U|=5/36$

Example 1: Monte Hall Problem

<u>Mr. Monte Hall</u> hosts a game show *"Let's make a deal"* on TV.

One such game involves 3 closed doors. Behind one (randomly chosen) door is a beautiful sports car, while the other 2 doors each has a donkey behind it



Game format: Guest is invited to try to win the car as follows:

- Guest: Picks a random door i.
- Monte: Opens a <u>different</u> door $j \neq i$ which has a donkey behind it.
- Monte then asks the Guest, "*Would you like to switch your choice of i ?*"

Our Question: *Should the Guest switch?*

Example 1: Monte Hall Problem (continued)

Marilyn vos Savant writes a column "*Ask Marilyn*" in Parade magazine (she reportedly has an extremely high IQ) . In a 1990 column, she gave her opinion on the Monte Hall Problem, saying that Switch is the correct choice!

Many readers doubted and protested about Marilyn's answer:

How can the new info about donkey <u>from a different door</u> affect the location of car?

But Marilyn turned out to be correct! Here's the analysis.

- Let's formulate it in the formal probability language.

P = (U, *p*) , where U ={a, b} ("a" represents the situation when Guest's initial pick is the *correct* door with car behind it; "b" the other case)

$$p(a) = 1/3, \; p(b) = 2/3$$

- Let T be the event that "action *Switch* would lead Guest to the car".

Clearly, T={b}, hence Pr{T} = p(b) = 2/3. (Note non-Switch gives 1/3 success probability.)

Example 1: Monte Hall Problem (continued)

According to Wikipedia, Paul Erdös, a famous mathematician, remains unconvinced until he was shown a computer simulation 🙂

💡 *But imagine 100 doors, with 98 doors revealed!*

A Question about Passwords

- Mr. Zhang is a data center manager in a university with 30,000 faculty and students. He assigns a random m-bit password x_i to each faculty/student $i \in \{1,2,\ldots, 30000\}$. What value of m *should Mr. Zhang choose?*

  Requirement: $x_i \neq x_j$ for all $i \neq j$.

  For example: *Smallest* m such that $\Pr\{ x_i \neq x_j \text{ for all } i \neq j \} > 1 - 10^{-10}$.

- Along this line, there is a famous problem which we will discuss next. (Mr. Zhang's problem will be left as a homework problem.)

Example 2: Birthday Paradox

In a party of *n* random people, how likely is it to have two people with same birthday? Assume there are 365 days in a year, and all days are equally likely to be a birthday. Let q(n) stand for this probability.

- If one does experiments, it turns out empirically q(23) is about ½, meaning that in a group as small as 23 people, there is a fifty-fifty chance to have two people with the same birthday
- A counter-intuitive result (hence called a paradox)!

  As n gets larger, q(50)=0.97, and q(70)=0.999

Example 2: Birthday Paradox (continued)

How do we explain this mathematically?

Consider the *Probability Space*:

- $P = (U, p)$, where $U = \{(x_1, x_2, \ldots, x_n) \mid 1 \leq x_k \leq 365 \text{ for all } k\}$

  $p(u) = 1/|U|$

  $T = \{(x_1, x_2, \ldots, x_n) \mid \text{there exists } x_j = x_k \text{ for some } j \neq k\} \subseteq U$
- Analyze $\Pr\{T\}$.

<u>*Theorem*</u> Let $q(n) = \Pr\{T\}$. Then $q(n)$ is a non-decreasing function of n.

- For all $n>0$, $q(n) = 1 - (1 - 1/365)(1 - 2/365)\ldots(1 - (n-1)/365)$
- $q(22)<0.5$, $q(23)>0.5$

Consider the *complemented* (or, *negated*) event $\bar{T} = U - T$. Then

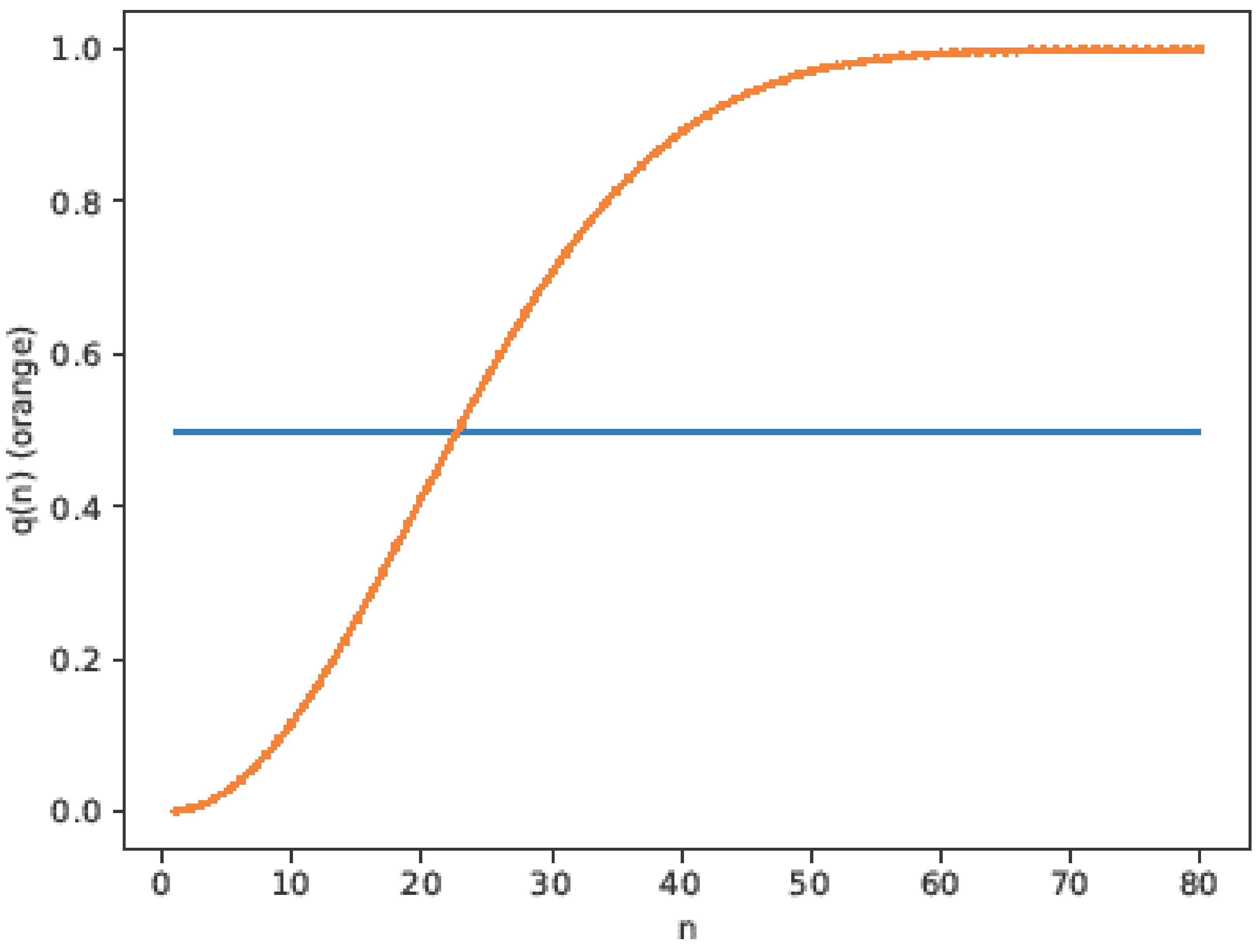
$$|T| = |U| - |\bar{T}|. \qquad (1)$$

Now, each element $(i_1, i_2, \cdots, i_n) \in \bar{T}$ can be uniquely specified by picking i_1 (365 choices), then i_2 (364 choices), ..., i_n (365-n+1 choices). Thus

$$|\bar{T}| = 365 \cdot 364 \cdots (365 - n + 1). \qquad (2)$$

It follows from (1), (2) that

$$\begin{aligned} &\Pr\{T\} \\ = \frac{|T|}{|U|} &= 1 - \frac{|\bar{T}|}{|U|} \\ &= 1 - \frac{365 \cdot 364 \cdots (365 - n + 1)}{365^n} \\ &= 1 - 1 \cdot (1 - \frac{1}{365}) \cdot (1 - \frac{2}{365}) \cdots (1 - \frac{n-1}{365}). \end{aligned}$$



q(n) (orange)
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
60
70
80
n

The series expansion for e^x is $\Sigma_{i\geq 0}\frac{x^i}{i!}$ for all x. When $|x|$ is small, a reasonable approximation is

$$e^{-x} \approx 1 - x.$$

(Also in fact $e^{-x} \geq 1 - x$ for $x \geq 0$.)

Thus, for $n \leq 80$, intuitively

$$\begin{aligned}
\Pr\{T\} &= 1 - (1 - \frac{1}{365}) \cdot (1 - \frac{2}{365}) \cdots (1 - \frac{n-1}{365}) \\
&\approx 1 - e^{\frac{-1}{365}} e^{\frac{-2}{365}} \cdots e^{\frac{-(n-1)}{365}} \\
&= 1 - \exp(-\sum_{1\leq i<n} \frac{i}{365}) \\
&= 1 - \exp(-\frac{n(n-1)}{2 * 365}) \\
&\equiv d(n).
\end{aligned}$$

To see what n makes $q(n)$ rise above 0.5, look at when $d(n)$ rises above 0.5.

Consider the solution of $d(x) = 0.5$.

$$\exp(-\frac{x(x-1)}{2*365}) = 0.5,$$

i.e

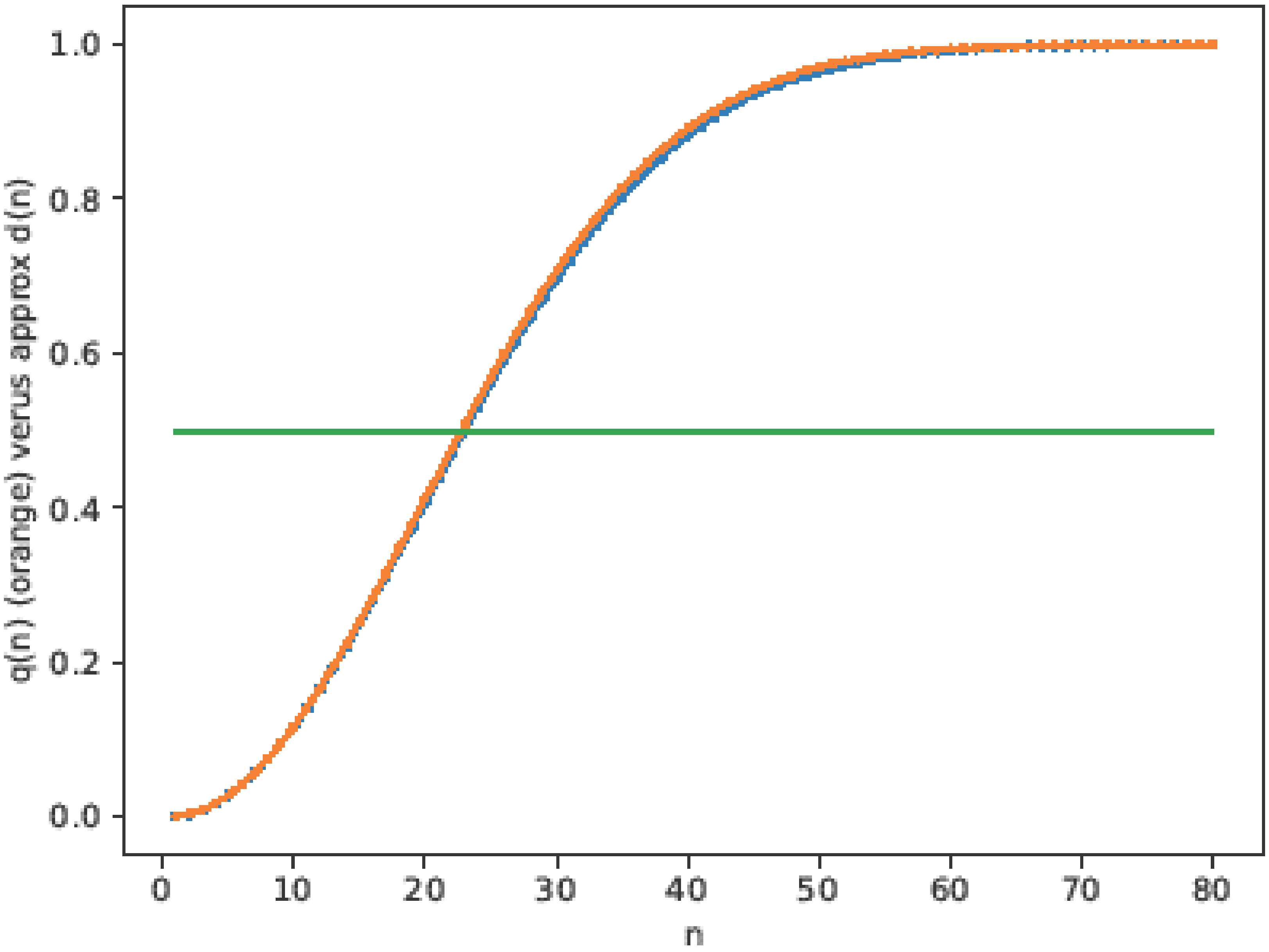
$$\frac{x(x-1)}{2*365} = \ln 2 = 0.69.$$

So, roughly,

$$x = (2*365*0.69)^{1/2} = 22.44.$$

Exactly the correct crossover location $22 < n < 23$!

In fact $d(n)$ approximates $q(n)$ very well (except when n is really small). That's why we get the crossover location exactly. We will show you the numerical values. In the homework we will show how to treat this issue *analytically*.



q(n) (orange) verus approx d(n)
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
60
70
80
n

Error of Approximation seems small

	Pr{T}	Approx	error
n	*q(n)*	*d(n)*	*d(n)-q(n)*
20	0.412	0.406	- 0.006
21	0.444	0.437	-0.007
22	0.476	0.469	-0.007
23	0.507	0.500	-0.007
24	0.538	0.530	-0.008
25	0.569	0.560	-0.009

Note $|d(n)-q(n)| \ll q(n) - q(n-1)$

i.e. d(n) is much closer to q(n) than to q(n-1)

For probability space with uniform probability function *p*, we have *Pr{T}=|T|/|U|*, evaluating probability of an event *T* is the same as *counting the size* of T.

- *Kindergarten* counting rules:
  - -- Addition rule: If a set S is the disjoint union of S_k, then $|S| = \sum_k |S_k|$.
  - -- Multiplication rule: If each item s in S can be uniquely specified as $s = (i_1, i_2, \ldots, i_m)$, where $1 \leq i_k \leq c_k$ then $|S| = c_1 \cdot c_2 \cdots c_m$
- These elementary rules are surprisingly useful in solving many probability problems, e.g. in solving the birthday paradox, we implicitly used the multiplication principle.
- Our next example shows an example where both principles are utilized.

Example 3. Online Auction Problem (aka. *Beauty contest*, *Secretary's problem*)

Suppose you're selling a concert ticket online to $n=10^6$ interested bidders:

Given a stream of n distinct offers $x_1,x_2,,,x_n$, you have to make decision in real-time.

You want to maximize the probability of accepting the highest offer.

Strategy k: $(k<n)$

1) Skip the first k offers

2) Accept x_j if j is the first j satisfying $x_j > \max\{ x_1,x_2,,,x_k \}$

(* If no x_j is selected, clearly the strategy has failed.)

Analysis of Strategy k

Consider the *Probability Space*:

P = (U,p), where U= the set of all permutations of {1,2,…,n}, $p = 1/|U|=1/n!$

Let T be the event of *success* (i.e. when the best offer j gets selected by Strategy k)

Fact. A permutation $x = (x_1,x_2,\ldots,x_n)$ is in T iff the following are true:

(1) $j > k$ (where j is defined by $x_j = n$)

(2) $\max\{x_1,x_2,\ldots,x_{j-1}\} = \max\{x_1,x_2,\ldots,x_k\}$

Thus, $\Pr\{T\} = |T|/|U| = |T|/n!$

For each $k+1 \le j \le n$, let $T_j \subseteq T$ be the subset of those permutations x satisfying $x_j = n$

Lemma 1. $|T| = \sum_{k+1\le j\le n} |T_j|$.

Proof. By addition principle, as T_j's are disjoint.

Lemma 2. $|T_j| = n!\, k/(n(j-1))$.

Proof. Using multiplication principle. (Omitted)

Theorem $\Pr\{T\} = h_{n,k} = \frac{k}{n}\left(\frac{1}{k} + \frac{1}{k+1} + \cdots + \frac{1}{n-1}\right).$

e.g $h_{n,k} \approx 38\%$ for n=8, k=4

U

Coming back to our Online Auction Problem with $n = 10^6$

Consider the *Harmonic Numbers* $H_n = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} \approx \ln n + C + \ldots$

Choose $k = \lceil n/e \rceil = \lceil \mathrm{n}/2.718 \rceil$ ceiling function

Then $h_{n,k} = \frac{k}{n}(H_{n-1} - H_{k-1}) \approx \frac{k}{n}(\ln n - \ln(n/e)) = 1/e \approx 38\%$

Amazingly, by choosing k=n/e, you can land on exactly the <u>highest</u> offer
with probability close to 40% among a million offers!

For more complex problems and questions, we need to develop additional probability concepts and a set of essential tools useful for analysis and calculations.

Essential Probability Tools #1: The Union Bound

1) Let $T, T_1, \ldots, T_m$ be events, and $T \subseteq \cup_i T_i$. Then $\Pr\{T\} \leq \sum_i \Pr\{T_i\}$.
2) If T_i's are disjoint and $T = \cup_i T_i$ then $\Pr\{T\} = \sum_i \Pr\{T_i\}$.

This simple bound often yields surprisingly powerful results, as illustrated by a celebrated result on Ramsey numbers by Paul Erdös.

What are Ramsey numbers?

Ramsey numbers: simplest case

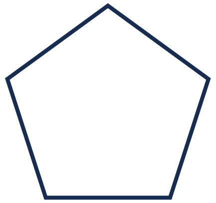
Among 6 people, there must exist <u>either</u> 3 mutual friends, <u>or</u> 3 mutual strangers.

(called the Friendship Theorem)

<u>proof</u>. Construct the *<u>Friendship Graph</u>: an edge between 2 friends*

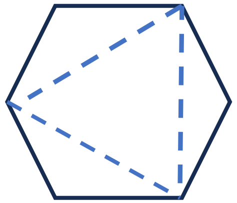
no edge between 2 strangers

5 vertices



G has no triangle
and no anti-triangle

6 vertices



Any 6-vertex graph has
Either a triangle
or an anti-triangle

<u>Ramsey's Theorem</u> *For any integer k≥3, there exists an integer N>0 such that among N people, either ∃ k mutual friends or k mutual strangers.*

- The smallest such N is called the k-th Ramsey number, R(k). For example R(3)=6.

- You are going to prove Ramsey's theorem in Homework #1

  by showing that $R(k) \le \binom{2k-2}{k-1} < 4^k$

- How about <u>lower bound</u> for $R(k)$?

  A famous result by Paul Erdös gives a lower bound:

<u>Theorem</u> (Paul Erdös 1947) For all $k \ge 3$, $R(k) \ge \lfloor 2^{k/2} \rfloor$

Proof. Let $n = \lfloor 2^{k/2} \rfloor$. Let P=(U,p) where U is the set of all graphs on n vertices, and p is the uniform probability function on U. In other words, a random graph G is obtained by setting x_{ij} randomly to 0 or 1 with equal prob for each pair of vertices {i,j}.

Let T be the event that "G contains no clique of size k and no independent set of size k". We prove $\Pr\{T\} > 0$. Equivalently, we show $\Pr\{\overline{T}\} < 1$ where $\overline{T}$=U-T. For any subset V of k vertices, let A_V, B_V be the event that V forms a clique (or an independent set) in the random G, respectively.

By definition,

$$\bar{T} = (\cup_{V,|V|=k} A_V) \cup (\cup_{V,|V|=k} B_V).$$

By the Union Bound, we have for any $k \geq 3$, with $n = \lfloor 2^{k/2} \rfloor$

$$\begin{aligned}
\Pr\{\bar{T}\} &\leq \sum_{V,|V|=k} (\Pr\{A_V\} + \Pr\{B_V\}) \\
&= 2\binom{n}{k}\frac{1}{2^{\binom{k}{2}}} \leq 2\,\frac{n^k}{k!}\,\frac{1}{2^{\binom{k}{2}}} \\
&\leq 2\,\frac{(2^{\frac{k}{2}})^k}{k!}\,\frac{1}{2^{\binom{k}{2}}} = 2\,\frac{2^{\frac{k^2}{2}}}{k!}\,\frac{1}{2^{(k^2-k)/2}} \\
&= 2\,\frac{2^{k/2}}{k!} < 1.
\end{aligned}$$

Significance of Erdös' Theorem:

- It is a novel idea to prove the existence of a mathematical object with certain sophisticated properties without explicitly constructing it.
- Erdös' 1947 paper started an important field "Probabilistic Method" in combinatorics, number theory, theoretical computer science.

Long-standing Open Problem: Give an *explicit* construction, for each k, a graph on $n = \lceil c^k \rceil$ vertices that contains no clique and no independent set of size k, where $c > 1$ is some constant.

* For computer scientists, this means a construction by an algorithm running in polynomial time in n

* Best constructive bounds known (roughly): for some small constant $\epsilon > 0$

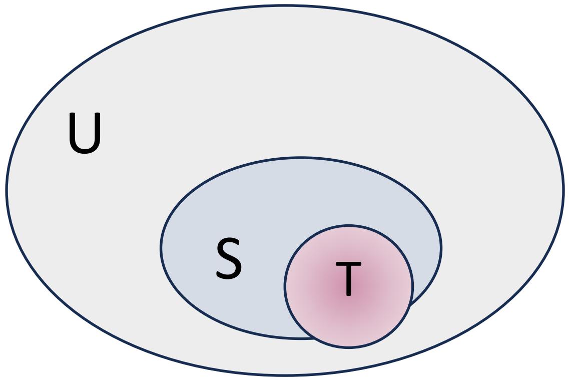
$$R(k) > 2^{2^{(\log k)^\epsilon}}.$$

We next introduce an important concept of conditional probability.

Definition For events S, T, the *conditional probability* of S (given T) is defined as:

$$\Pr\{S \mid T\} = \begin{cases} \Pr\{S \cap T\} / \Pr\{T\}, & \text{if } \Pr\{T\} > 0\,; \\ 0, & \text{if } \Pr\{T\} = 0. \end{cases}$$

* We often write $\Pr\{S \cap T\}$ as $\Pr\{S \wedge T\}$ (S *AND* T in logical sense); write $\Pr\{S \cup T\}$ as $\Pr\{S \vee T\}$ (S *OR* T in logical sense).



Example 2: P=(U,p) where U is the set of all college students, and p is the uniform probability function. Let $S \subseteq U$ be the event consisting of all students who are "smart". Let's say Pr{S} = 40%.

Let T=Tsinghua students. Then Pr{S|T}=100%.

It is easy to verify that $\Pr\{S \cap T\} = \Pr\{T\} \cdot \Pr\{S|T\}$.

This basic equation has the following important generalizations.

- ***Essential Probability Tools #2:***

2A: The Chain Rule (for Conditional Probability)

$$\Pr\{S_1 \cap S_2 \cap \cdots \cap S_m\} = \prod_{1 \le j \le m} \Pr\{S_j \mid S_1 \cap S_2 \cap \cdots \cap S_{j-1}\}.$$

(e.g. $\Pr\{S_1 \cap S_2 \cap S_3\} = \Pr\{S_1\} \cdot \Pr\{S_2 \mid S_1\} \cdot \Pr\{S_3 \mid S_2 \cap S_1\}$.)

2B: Distributive Law (Law of Total Probability)

Let $T \subseteq W_1 \cup W_2 \cup \cdots \cup W_m$.

Then $\Pr\{T\} \le \sum_{1 \le j \le m} \Pr\{W_j\} \cdot \Pr\{T \mid W_j\}$.

Furthermore, if W_j's are disjoint, then the above inequality is equality.

Essential Probability Tools #1 & 2 are generalizations of the Kindergarten Addition and Multiplication rules).

- Let us revisit the Birthday problem and the Auction problem using these new tools.

1. **Birthday Problem**: Analysis of probability of birthday coincidence of n people

$P = (U, p)$ where $U = \{x = (x_1, \cdots, x_n) \mid 1 \le x_i \le 365\}$ and p is uniform over U.

Recall $T = \{(x_1, x_2, \ldots, x_n) \mid \text{there exists } x_j = x_k \text{ for some } j \neq k\} \subseteq U$

For each j,

let S_j = the set of x satisfying $x_j \notin \{x_1, \cdots, x_{j-1}\}$.

$\bar{T}$: the event that all birthdays x_i are distinct

By Chain Rule,

$$\begin{aligned}
\Pr\{\bar{T}\} &= \Pr\{S_1 \cap S_2 \cap \cdots \cap S_n\} \\
&= \prod_{1 \le j \le n} \Pr\{S_j \mid S_1 \cap S_2 \cap \cdots \cap S_{j-1}\} \\
&= \prod_{1 \le j \le n} \left(1 - \frac{j-1}{365}\right)
\end{aligned}$$

QED

2. Auction Problem: Analysis of Strategy k in selecting best bid (*revisited*)

First recall the notations:

$P = (U, p)$ where p is uniform over U, and

U: the set of all $n!$ permutations of $\{1, 2, \cdots, n\}$

Any $x = (x_1, \cdots, x_n) \in U$ is a stream of bid sequence. Strategy k tries to select the highest bid x_j.

T: the set of x for which Strategy k successfully selects the highest bid n

W_j: the set of x with $x_j = n$; , clearly all W_j are *disjoint*

- As we analyzed before:

$T = \cup_{k+1\le j\le n} T_j$ where T_j is the set of x satisfying:

(a) $x \in W_j$, and

(b) max of $\{x_1, \cdots, x_{j-1}\}$ occurs in $\{x_1, \cdots, x_k\}$

- Thus $T \subseteq \cup_{1\le j\le n} W_j$

By Distributive Law,

$$\begin{aligned}
\Pr\{T\} &= \sum_{k+1\le j\le n} \Pr\{W_j\} \cdot \Pr\{T \mid W_j\} \\
&= \sum_{k+1\le j\le n} \frac{1}{n} \cdot \frac{k}{j-1} \\
&= \frac{k}{n}\left(\frac{1}{k} + \frac{1}{k+1} + \cdots + \frac{1}{n-1}\right).
\end{aligned}$$

QED

The End