What is it about?

- Probabilistic thinking!

Administrative Stuff

- 5 assignments (to be done individually)
- 1 final presentation and report (I will assign papers and topic)

First few weeks

- Gentle introduction to concepts and techniques from probability theory
- Done via sample problems from many areas (networking, algorithms, combinatorics, coding, learning theory, etc.)

PTCF $=$ *Probability Theory Concepts and Facts*

Lecture 1: The Probabilistic Method

- Discrete Probability Space
- Events
- The Probabilistic Method
- The Union Bound

Example 1: Tracing A Single Traitor

- Application: "broadcast" to a group of legitimate users
  - DVD or CD-ROM distribution of movies or softwares
  - Pay-per-view subscriptions
  - Online databases
- Some user might be traitor, giving his key(s) to a pirate
- Pirate sells decryption device on black market
- Problem: obtain device, identify the (single) traitor
- Two extremes, both do not work well
  - Single shared key: can't trace the traitor
  - Each person a key: cipher-text too large!

Traitor Tracing and Sperner Family

- Set of keys T, $|T| = t$
- n users, user i given a subset $F_i \subseteq T$ of keys

Claim

To be able to trace a traitor, $F_i \not\subseteq F_j$, for all $i \neq j$.

- A family $\mathcal{F}$ of sets where no member is contained in another is called a Sperner family

Main Questions

- Given n, find the smallest t for which a Sperner family of n sets on $[t] = \{1, \cdots, t\}$ exists
- Dually, given t find the maximum n for which a Sperner family of n sets on $[t]$ exists

A Classic Theorem by Sperner

Theorem (Sperner, 1928)

The maximum size of a family $\mathcal{F}$ of subsets of $[t]$ whose members do not contain one another is $\binom{t}{\lfloor t/2 \rfloor}$.

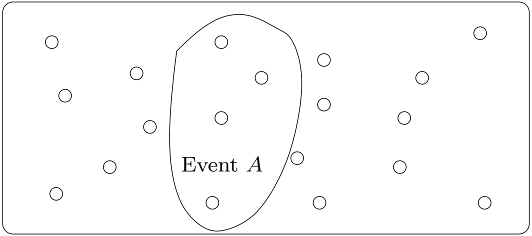
- The collection of $\lfloor t/2 \rfloor$-subsets of $[t]$ is a Sperner family
- Thus, suffices to show that $|\mathcal{F}| \leq \binom{t}{\lfloor t/2 \rfloor}$ for any Sperner family $\mathcal{F}$
- Pick a permutation π of $[t]$ uniformly at random
- For $F \in \mathcal{F}$, let A_F be the event that F is a prefix of π

$$\text{Prob}[A_F] = \frac{k!(t-k)!}{t!} = \frac{1}{\binom{t}{k}} \geq \frac{1}{\binom{t}{\lfloor t/2 \rfloor}}, \quad \text{where } k = |F|$$

- The A_F are mutually exclusive (why?), hence

$$1 \geq \text{Prob}\left[\bigcup_{F \in \mathcal{F}} A_F\right] = \sum_{F \in \mathcal{F}} \text{Prob}[A_F] \geq \frac{|\mathcal{F}|}{\binom{t}{\lfloor t/2 \rfloor}}$$

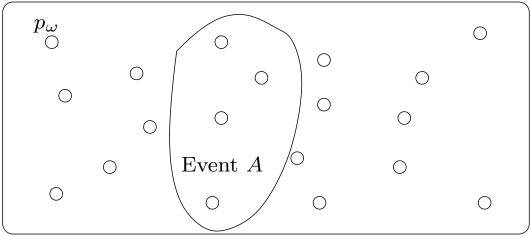
PTCF: Simple Probability Space



Sample Space Ω

- Ω is a finite set of all possible outcomes of some experiment
- Each outcome occurs equally likely
- A subset A of outcomes is an event
  - Think of it as a set of outcomes satisfying a certain property
- $\text{Prob}[A] = \frac{|A|}{|\Omega|}$: the fraction of outcomes in A
- In most cases, **not the way** to think about probability spaces

PTCF: Discrete Probability Space



Sample Space Ω

- Each $\omega \in \Omega$ is assigned a number $p_\omega \in [0, 1]$, such that $\sum_{\omega \in \Omega} p_\omega = 1$.
- For any event A, Prob$[A] = \sum_{\omega \in A} p_\omega$.
- In the simple space, $p_\omega = \frac{1}{|\Omega|}, \forall \omega$
- Note: this is **not** the most general definition, but suffices for now.

PTCF: How do we "assign" the p_ω?

- Could think of it as a mathematical function, like saying "give each outcome ω a number p_ω equal to $1/|\Omega|$"
- That's **not** the probabilistic way of thinking!
- Probabilistic way of thinking:
  - An experiment is an *algorithm* whose outcome is not deterministic
  - For example, algorithms making use of a random source (like a bunch of "fair" coins)
  - Ω is the set of all possible outputs of the algorithm
  - p_ω is the "likelihood" that ω is output

Example 2: Tracing a Group of Traitors

- Suppose we know there are $\leq d < n$ traitors out of n users
- User j gets key set F_j, set system $\mathcal{F} = \{F_j\}_{j=1}^n$

Claim (Property $\mathcal{F}$ must satisfy)

For arbitrary $j_0, j_1, \cdots, j_d \in [n]$,

$$F_{j_0} \not\subseteq F_{j_1} \cup \cdots \cup F_{j_d}.$$

- Such a family $\mathcal{F}$ is called a d-cover-free family
- d-cover-free family of size n on $[t]$ is equivalent to d-disjunct matrix

Non-Adaptive Group Testing

- A $t \times n$ binary matrix $\mathbf{A}$ is called d-disjunct iff the union of any d columns does not contain another column
- Columns are codewords of superimposed codes
- Rate of the code is $R(\mathbf{A}) = \frac{\log n}{t}$
- Want codes with high rates. But, as $n \to \infty$ and $d \to \infty$

$$\frac{1}{d^2 \log e}(1 + o(1)) \leq \limsup_{\mathbf{A}} R(\mathbf{A}) \leq \frac{2 \log d}{d^2}(1 + o(1))$$

  (From Dyachkov, Rykov (1982), and Dyachkov, Rykov and Rashad (1989))
- We'll prove the lower bound

The Probabilistic Method

Want to prove that $t \times n$ d-disjunct matrix exists with small t

Strategy:

- Fix t (which we'll choose later)
- Choose a $t \times n$ matrix $\mathbf{A}$ at random, somehow
- Prove that, with $t = t(d, n)$,

$$\text{Prob}[\mathbf{A} \text{ is } d\text{-disjunct}] > 0.$$

- Or, equivalently

$$\text{Prob}[\mathbf{A} \text{ is not } d\text{-disjunct}] < 1.$$

Existence of Good d-disjunct Matrix

- Set a_{ij} to 1 with probability p
- Fix j_0 and a set $C = \{j_1, \cdots, j_d\} \subseteq [n]$, $j_0 \notin C$
- (j_0, C) is bad for $\mathbf{A}$ if column j_0 is contained in the union of columns in C
- Let $B_{j_0,C}$ be the event that (j_0, C) is bad
- $\mathbf{A}$ is not d-disjunct implies $\bigcup_{(j_0,C)} B_{j_0,C}$, thus

$$\text{Prob}[A \text{ is not } d\text{-disjunct}] \leq \text{Prob}\left[\bigcup_{j_0,C} B_{j_0,C}\right] \leq \underbrace{\cdots}_{how?} < 1$$

PTCF: The Union Bound

Lemma

Let $B_1, B_2, \dots$ be any finite or countably infinite sequence of events. Then,

$$\text{Prob}\left[\bigcup_{i \geq 1} B_i\right] \leq \sum_{i \geq 1} \text{Prob}[B_i]$$

Note:

- this bound holds for **any** probability space (not just simple spaces).
- the bound is simple but extremely useful!

Existence of Good d-disjunct Matrix

$$\text{Prob}\left[\bigcup_{j_0,C} B_{j_0,C}\right] \leq \sum_{j_0,C} \text{Prob}[B_{j_0,C}] = \sum_{j_0,C} \left[1 - p(1-p)^d\right]^t$$

- Set $p = 1/(d+1)$, $\mathbf{A}$ is **not** d-disjunct with probability at most

$$(d+1)\binom{n}{d+1}\left[1 - \frac{1}{d+1}\left(1 - \frac{1}{d+1}\right)^d\right]^t$$

- $f(x) = (1 - 1/(x+1))^x$ is decreasing when $x \geq 1$, and $\lim_{x\to\infty} f(x) = 1/e$, hence $f(x) \geq 1/e$
- RHS is upper-bounded by

$$(d+1)\binom{n}{d+1}\left[1 - \frac{1}{e(d+1)}\right]^t \leq (d+1)\left(\frac{ne}{d+1}\right)^{d+1} e^{-1/e(d+1)}$$

- This is < 1 as long as $t \geq 2e(d+1)^2 \ln\left(en/(d+1)\right)$.

PTCF: Two Very Useful Inequalities

$$1 + x \le e^x, \qquad \forall x \in \mathbb{R} \tag{1}$$

$$\sum_{i=0}^{d} \binom{n}{i} \le \left(\frac{ne}{d}\right)^d, \qquad \forall d \le n \tag{2}$$

The Union Bound Technique

- An extremely simpmle *and* useful technique
- Should be the "first thing to try"

More on the Union Bound and the Probabilistic Method

- Alon and Spencer, "The Probabilistic Method"
- Bolobas, "Random Graphs"

The Union Bound Technique: Main Idea

- A: event our structure exists, want Prob$[A] > 0$ or Prob$[\bar{A}] < 1$
- Suppose $\bar{A}$ implies one of $B_1, \cdots, B_n$ must hold
- (Think of the B_i as the "bad events")
- Then, by the union bound

$$\text{Prob}\left[\bar{A}\right] \leq \text{Prob}\left[\bigcup_i B_i\right] \leq \sum_i \text{Prob}[B_i]$$

- Thus, as long as

$$\sum_i \text{Prob}[B_i] < 1$$

  our structure exists!

We have seen this used in d-disjunct matrix examples.

Example 3: Nice Tournaments

- A tournament is an orientation G of K_n
- Think of $u \to v$ as *"player u beats player v"*
- Fix integer k, G is *nice* if for every k-subset S of players there is another v who beats all of S
- Intuitively, nice tournaments may exist for large n
(Remember the theme: "Sufficiently large space contains locally nice structures")

Existence of Nice Tournaments (Erdős, 1963)

- For every $\{u, v\}$, let $u \to v$ with probability $1/2$
- A: event that a random G is nice
- $\bar{A}$ implies $\bigcup_{|S|=k} B_S$ where $B_S =$ "S is not beaten by **any** $v \notin S$"

$$\text{Prob}[B_S] = \left(1 - \frac{1}{2^k}\right)^{n-k}$$

- Hence, nice tournaments exist as long as $\binom{n}{k}\left(1 - \frac{1}{2^k}\right)^{n-k} < 1$
- What's the order of n for which this holds?

$$\text{use } \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k \text{ and } \left(1 - \frac{1}{2^k}\right)^{n-k} < e^{-\frac{n-k}{2^k}}$$

- Nice tournaments exist as long as $\left(\frac{ne}{k}\right)^k e^{-\frac{n-k}{2^k}} < 1.$
- So, $n = \Omega\left(k^2 \cdot 2^k\right)$ is large enough!

Example 4: 2-coloring of uniform hypergraphs

- Given a k-uniform hypergraph $G = (V, E)$, i.e.
  - E is a collection of k-subsets of V
- G is 2-colorable iff each vertex in V can be assigned with red or blue such that there's no monochromatic edge
- Intuitively, if $|E|$ is small then G is 2-colorable!
- Question is: "how small?"
- An answer may be obtained along the line: "for n *small enough*, a random 2-coloring is good with positive probability"

Theorem (Erdős, 1963)

Every k-uniform hypergraph with $< 2^{k-1}$ edges is 2-colorable!

Example 5: Ramsey Numbers

- The Ramsey number $R(k, k)$ is the smallest integer n such that no matter how we assign red or blue to each edge of K_n, there must exist a monochromatic K_k.
- Analogy: $R(k, k)$ is the smallest n so that in any set of n people there must be **either** k mutual acquaintances, **or** k mutual strangers

Erdős' Quote

Imagine an alien force, vastly more powerful than us landing on Earth and demanding the value of $R(5, 5)$ or they will destroy our planet. In that case, we should marshal all our computers and all our mathematicians and attempt to find the value. But suppose, instead, that they asked for $R(6, 6)$, we should attempt to destroy the aliens.

Notes

- There are (much) more general Ramsey numbers. E.g., $R(a, b)$ is the smallest integer n such that no matter how we 2-color edges of K_n with red and blue, there exists either a red K_a or a blue K_b.
- Or multi-dimensional Ramsey numbers (the above is 2-dim)
- The problem is a generalization of the pigeonhole principle
- Intuition/interpretation:
  - when n is sufficiently large, there must be a monochromatic sub-clique of a given size
  - i.e., in a sufficiency large "space," local "patterns" must emerge. (this theme is manifested in different ways in this course)
  - problem is to find/estimate the threshold

Erdős' Theorem (1947)

Theorem

(i) *If* $\binom{n}{k}2^{1-\binom{k}{2}} < 1$, *then* $R(k,k) > n$.

(ii) *Consequently,* $R(k,k) > \lfloor 2^{k/2} \rfloor$ *for all* $k \geq 3$.

To see (ii), let $n = \lfloor 2^{k/2} \rfloor$.
Then,

$$\binom{n}{k}2^{1-\binom{k}{2}} < \frac{n^k}{k!} \cdot \frac{2^{1+k/2}}{2^{k^2/2}} < \frac{2^{1+k/2}}{k!} \cdot \frac{n^k}{2^{k^2/2}} < 1.$$

We will give two proofs of (i).

Probabilistic Method Proof #1

- Pick a coloring $c \in \Omega$ uniformly at random.
- For any $S \in \binom{[n]}{k}$, let B_S be the "bad" event that S is monochromatic, then

$$\text{Prob}[B_S] = \frac{\#\text{ colorings making } S \text{ mono.}}{\text{total } \#\text{ colorings}} = \frac{2 \times 2^{\binom{n}{2} - \binom{k}{2}}}{2^{\binom{n}{2}}} = 2^{1-\binom{k}{2}}$$

- The probability that some $S \in \binom{[n]}{k}$ is monochromatic is

$$\text{Prob}\left[\bigcup_S B_S\right] \leq \sum_S \text{Prob}[B_S] = \binom{n}{k} 2^{1-\binom{k}{2}} < 1$$

- Thus, there must be some coloring for which no S is monochromatic!

Probabilistic Method Proof #2 (much better than #1!)

- Color each edge of K_n with either red or blue with probability $1/2$
- For any $S \in \binom{[n]}{k}$, let B_S be the "bad" event that S is monochromatic, then

$$\text{Prob}[B_S] = \text{Prob}[S \text{ is blue}] + \text{Prob}[S \text{ is red}] = 2 \times \frac{1}{2^{\binom{k}{2}}} = 2^{1-\binom{k}{2}}$$

- The probability that some $S \in \binom{[n]}{k}$ is monochromatic is

$$\text{Prob}\left[\bigcup_S B_S\right] \leq \sum_S \text{Prob}[B_S] = \binom{n}{k} 2^{1-\binom{k}{2}} < 1$$

- Thus, there must be some coloring for which no S is monochromatic!

Example 6: Error-Correcting Codes

- Message $\mathbf{x} \in \{0,1\}^k$
- Encoding $f(\mathbf{x}) \in \{0,1\}^n$, $n > k$, f an injection
- $C = \{f(\mathbf{x}) \mid \mathbf{x} \in \{0,1\}^k\}$: codewords
- $f(\mathbf{x})$ is sent over noisy channel, few bits altered
- $\mathbf{y}$ is received instead of $f(\mathbf{x})$
- Find codeword $\mathbf{z}$ "closest" to $\mathbf{y}$ in Hamming distance
- Decoding $\mathbf{x}' = f^{-1}(\mathbf{z})$
- Measure of utilization: relative rate of C

$$R(C) = \frac{\log |C|}{n}$$

- Measure of noise tolerance: relative distance of C

$$\delta(C) = \frac{\min_{\mathbf{c}_1 \neq \mathbf{c}_2 \in C} \text{Dist}(\mathbf{c}_1, \mathbf{c}_2)}{n}$$

Linear Codes

- For any $\mathbf{x} \in \mathbb{F}_2^n$, define

$$\text{WEIGHT}(\mathbf{x}) = \text{ number of 1-coordinates of } \mathbf{x}$$

- E.g., $\text{WEIGHT}(1001110) = 4$
- If C is a k-dimensional subspace of $\mathbb{F}_2^n$, then

$$\begin{aligned} |C| &= 2^k \\ \delta(C) &= \min\{\text{WEIGHT}(\mathbf{x}) \mid \mathbf{x} \in C\} \end{aligned}$$

- Every such C can be defined by a parity check matrix $\mathbf{A}$ of dimension $(n-k) \times n$:

$$C = \{\mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{0}\}$$

- Conversely, every $(n-k) \times n$ matrix $\mathbf{A}$ defines a code C of dimension $\geq k$

A Communication Problem

Large rate and large distance are conflicting goals

Problem

Does there exist a family of codes C_k, $|C_k| = 2^k$, for infinitely many k, such that

$$R(C_k) \geq R_0 > 0$$

and

$$\delta(C_k) \geq \delta_0 > 0$$

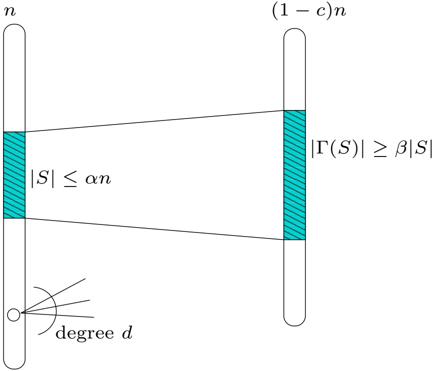
(Yes, using "magical graphs.")

Practicality

Design such a family explicitly, such that the codes are efficiently encodable and decodable.

Magical Graph

(n, c, d, α, β)-graph



c, d, α, β are constants, n varies.

From Magical Graphs to Code Family

- Suppose (n, c, d, α, β)-graphs exist for infinitely many n, and constants c, d, α, β such that $\beta > d/2$
- Consider such a $G = (L \cup R, E)$, $|L| = n, |R| = (1-c)n = m$
- Let $\mathbf{A} = (a_{ij})$ be the $m \times n$ 01-matrix, column indexed by L, and row-indexed by R, $a_{ij} = 1$ iff $(i, j) \in E$
- Define a linear code with $\mathbf{A}$ as parity check:

$$C = \{\mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{0}\}$$

- Then, $\dim(C) = n - \text{rank}(A) \geq cn$, and

$$|C| = 2^{\dim(C)} \geq 2^{cn} \;\Rightarrow R(C) \geq c$$

- For every $\mathbf{x} \in C$, $\text{WEIGHT}(\mathbf{x}) \geq \alpha n$, hence

$$\delta(C) = \frac{\min\{\text{WEIGHT}(\mathbf{x}) \mid \mathbf{x} \in C\}}{n} \geq \alpha$$

Existence of Magical Graph with $\beta > d/2$

- Determine n, c, d, α, β later
- Let $L = [n], R = [(1-c)n]$.
- Choose each of the d neighbors for $u \in L$ uniformly at random
- For $1 \leq s \leq \alpha n$, let B_s be the "bad" event that some subset S of size s has $|\Gamma(S)| < \beta|S|$
- For each $S \subset L$, $T \subset R$, $|S| = s, |T| = \beta s$, define

$$X_{S,T} = \begin{cases} 1 & \Gamma(S) \subseteq T \\ 0 & \Gamma(S) \not\subseteq T \end{cases}$$

- Then,

$$\text{Prob}[B_s] \leq \text{Prob}\left[\sum_{S,T} X_{S,T} > 0\right] \leq \sum_{S,T} \text{Prob}[X_{S,T} = 1]$$

Existence of Magical Graph with $\beta > d/2$

$$
\begin{aligned}
\text{Prob}[B_s] &\leq \binom{n}{s}\binom{(1-c)n}{\beta s}\left(\frac{\beta s}{(1-c)n}\right)^{sd} \\
&\leq \left(\frac{ne}{s}\right)^{s}\left(\frac{(1-c)ne}{\beta s}\right)^{\beta s}\left(\frac{\beta s}{(1-c)n}\right)^{sd} \\
&= \left[\left(\frac{s}{n}\right)^{d-\beta-1}\left(\frac{\beta}{1-c}\right)^{d-\beta} e^{\beta+1}\right]^{s} \\
&\leq \left[\left(\frac{\alpha\beta}{1-c}\right)^{d-\beta}\cdot\frac{e^{\beta+1}}{\alpha}\right]^{s}
\end{aligned}
$$

Choose $\alpha = 1/100$, $c = 1/10$, $d = 32$, $\beta = 17 > d/2$,

$$\text{Prob}[B_s] \leq 0.092^s$$

Existence of Magical Graph with $\beta > d/2$

The probability that such a randomly chosen graph is **not** an (n, c, d, α, β)-graph is at most

$$\sum_{s=1}^{\alpha n} \text{Prob}[B_s] \leq \sum_{s=1}^{\infty} 0.092^s = \frac{0.092}{1 - 0.092} < 0.11$$

Not only such graphs exist, there are **a lot** of them!!!

Some Key Ideas We've Learned

- To show the existence of some combinatorial object, set up some probability space and show that it exists with probability > 0
- The above is essentially a pigeonhole principle kind of proof, casted in probabilistic language
- We will see throughout the course that the probabilistic language is crucial!
- Thinking about probabilities "locally" is better than "globally"
- In a sufficiently large "space," locally nice "patterns" often emerge