CSE 694 – Prob. Analysis and Randomized Algo.

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, etc.)

PTCF $=$ *Probability Theory Concepts and Facts*

Example 1: Ramsey Numbers

- Let $R(a, b)$ be the smallest integer n such that in any 2-edge-coloring of K_n with red and blue, there exists either a red K_a or a blue K_b.
- Analogy: $R(a, b)$ is the smallest n so that in any set of n people there must be **either** a mutual acquaintances, **or** b 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.

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).

A Pigeonhole Principle Proof

We'll show that $\binom{n}{k}2^{1-\binom{k}{2}} < 1$ implies, there exists a 2-edge-coloring of K_n with **neither** a red K_k **nor** a blue K_k (i.e. no monochromatic K_k).

- Let $[n]$ be the set of vertices
- Let $\Omega =$ set of all 2-edge-colorings of K_n
- For any $S \in \binom{[n]}{k}$, the number of colorings for which S is monochromatic is $2 \times 2^{\binom{n}{2}-\binom{k}{2}}$
- The number of colorings for which every $S \in \binom{[n]}{k}$ is monochromatic is at most
$$\binom{n}{k} \times 2 \times 2^{\binom{n}{2}-\binom{k}{2}} = 2^{\binom{n}{2}}\binom{n}{k}2^{1-\binom{k}{2}}.$$
- But, the total number of colorings is $2^{\binom{n}{2}}$, and
$$2^{\binom{n}{2}}\binom{n}{k}2^{1-\binom{k}{2}} < 2^{\binom{n}{2}} \Leftrightarrow \binom{n}{k}2^{1-\binom{k}{2}} < 1$$

Probabilistic Method Proof #1

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

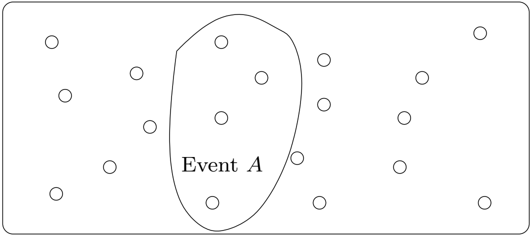
$$\text{Prob}[A_S] = \frac{\#\text{ colorings making } S \text{ mono.}}{\text{total } \#\text{ colorings}} = \frac{2 \times 2^{\binom{n}{2}-\binom{k}{2}}}{2^{\binom{n}{2}}} = 2^{1-\binom{k}{2}}$$

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

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

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

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** a good way to think about probability spaces

PTCF: The Union Bound

Lemma

Let $A_1, A_2, \dots$ be any finite or countably infinite sequence of events. Then,

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

Note:

- this bound hold for **any** probability space (not just simple ones).
- simple but extremely useful!

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 A_S be the event that S is monochromatic, then

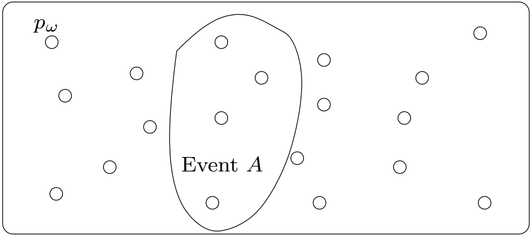
$$\text{Prob}[A_S] = \text{Prob}[S \text{ is blue}] + \text{Prob}[S \text{ is red}] = 2 \times \frac{1}{2^{\binom{k}{2}}} = 2^{1-\binom{k}{2}}$$

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

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

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

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, $p_\omega = \frac{1}{|\Omega|}, \forall \omega$
- This is **not** the most general definition.

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: Sperner Lemma

Lemma (Sperner, 1928)

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

- The collection of $\lfloor n/2 \rfloor$-subsets of $[n]$ satisfies the condition
- Suffices to show that, for any such $\mathcal{F}$, $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$.
- Fix $F \in \mathcal{F}$, choose a permutation $\pi \in S_n$ uniformly at random
- Let A_F be the event that $F = \{\pi_1, \dots, \pi_k\}$ for some k, then

$$\text{Prob}[A_F] = \frac{k!(n-k)!}{n!} = \frac{1}{\binom{n}{k}} \geq \frac{1}{\binom{n}{\lfloor n/2 \rfloor}}$$

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

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

Example 1: Randomized Min-Cut

Min-Cut Problem

Given a multigraph G, find a cut with minimum size.

RANDOMIZED MIN-CUT(G)

```
for i = 1 to n − 2 do
    Pick an edge e_i in G uniformly at random
    Contract two end points of e_i (remove loops)
end for
// At this point, two vertices u, v left
Output all remaining edges between u and v
```

Analysis

- Let C be a minimum cut, $k = |C|$
- If no edge in C is chosen by the algorithm, then C will be returned in the end, and vice versa
- For $i = 1..n-2$, let A_i be the event that $e_i \notin C$ and B_i be the event that $\{e_1, \dots, e_i\} \cap C = \emptyset$

$$
\begin{aligned}
& \text{Prob}[C \text{ is returned}] \\
= \quad & \text{Prob}[B_{n-2}] \\
= \quad & \text{Prob}[A_{n-2} \cap B_{n-3}] \\
= \quad & \text{Prob}[A_{n-2} \mid B_{n-3}]\,\text{Prob}[B_{n-3}] \\
= \quad & \dots \\
= \quad & \text{Prob}[A_{n-2} \mid B_{n-3}]\,\text{Prob}[A_{n-3} \mid B_{n-4}] \cdots \text{Prob}[A_2 \mid B_1]\,\text{Prob}[B_1]
\end{aligned}
$$

Analysis

- At step 1, G has min-degree $\geq k$, hence $\geq kn/2$ edges
- Thus,

$$\text{Prob}[B_1] = \text{Prob}[A_1] \geq 1 - \frac{k}{kn/2} = 1 - \frac{2}{n}$$

- Now we estimate $\text{Prob}[A_2 \mid B_1]$.
  - At step 2, the min cut is still at least k, hence $\geq k(n-1)/2$ edges
  - Thus, similar to step 1 we have

$$\text{Prob}[A_2 \mid B_1] \geq 1 - \frac{2}{n-1}$$

- In general,

$$\text{Prob}[A_j \mid B_{j-1}] \geq 1 - \frac{2}{n-j+1}$$

- Consequently,

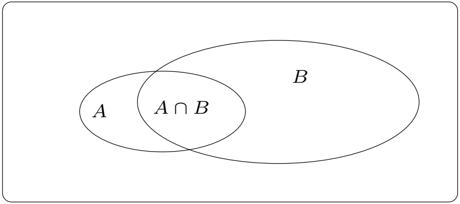
$$\text{Prob}[C \text{ is returned}] \geq \prod_{i=1}^{n-2} \left(1 - \frac{2}{n-i+1}\right) = \frac{2}{n(n-1)}$$

Lower the Failure Probability

- The basic algorithm has failure probability at most $1 - \frac{2}{n(n-1)}$
- How do we lower it?
- Run the algorithm multiple times, say $m \cdot n(n-1)/2$ times, return the smallest cut found
- The failure probability is at most

$$\left(1 - \frac{2}{n(n-1)}\right)^{m \cdot n(n-1)/2} < \frac{1}{e^m}.$$

PTCF: Independence Events and Conditional Probabilities



- The conditional probability of A given B is

$$\text{Prob}[A \mid B] := \frac{\text{Prob}[A \cap B]}{\text{Prob}[B]}$$

- A and B are independent if and only if

$$\text{Prob}[A \mid B] = \text{Prob}[A]$$

- Equivalently, A and B are independent if and only if

$$\text{Prob}[A \cap B] = \text{Prob}[A] \cdot \text{Prob}[B]$$

PTCF: Mutually Independence and Independent Trials

- A set $A_1, \dots, A_n$ of events are said to be independent or mutually independent if and only if, for any $k \leq n$ and $\{i_1, \dots, i_k\} \subseteq [n]$ we have
$$\text{Prob}[A_{i_1} \cap \dots \cap A_{i_k}] = \text{Prob}[A_{i_1}] \dots \text{Prob}[A_{i_k}].$$

- If n independent experiments (or trials) are performed in a row, with the ith being "successful" with probability p_i, then
$$\text{Prob}[\text{all experiments are successful}] = p_1 \cdots p_n.$$
(Question: what is the sample space?)

Example 2: Randomized Quicksort

RANDOMIZED-QUICKSORT(A)

$n \leftarrow$ length(A)
if $n = 1$ **then**
    Return A
else
    Pick $i \in \{1, \dots, n\}$ uniformly at random, $A[i]$ is called the *pivot*
    $L \leftarrow$ elements $\leq A[i]$
    $R \leftarrow$ elements $> A[i]$
    // the above takes one pass through A
    $L \leftarrow$ RANDOMIZED-QUICKSORT(L)
    $R \leftarrow$ RANDOMIZED-QUICKSORT(R)
    Return $L \cdot A[i] \cdot R$
end if

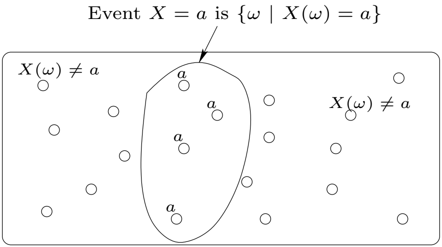
Analysis of Randomized Quicksort

- The running time is proportional to the number of comparisons
- Let $b_1 \leq b_2 \leq \cdots \leq b_n$ be A sorted non-decreasingly
- For each $i < j$, let X_{ij} be the indicator random variable indicating if b_i was ever compared with b_j
- The expected number of comparisons is

$$\mathsf{E}\left[\sum_{i<j} X_{ij}\right] = \sum_{i<j} \mathsf{E}[X_{ij}] = \sum_{i<j} \mathsf{Prob}[b_i \ \& \ b_j \text{ was compared}]$$

- b_i was compared with b_j if and only if either b_i or b_j was chosen as a pivot before any other in the set $\{b_i, b_{i+1}, \ldots, b_j\}$
- Hence, $\mathsf{Prob}[b_i \ \& \ b_j \text{ was compared}] = \frac{2}{j-i+1}$
- Thus, the expected running time is $\Theta(n \lg n)$

PTCF: Discrete Random Variable



- A random variable is a function $X : \Omega \to \mathbb{R}$
- $p_X(a) = \text{Prob}[X = a]$ is called the probability mass function of X
- $P_X(a) = \text{Prob}[X \leq a]$ is called the (cumulative/probability) distribution function of X

PTCF: Expectation and its Linearity

- The expected value of X is defined as

$$\mathsf{E}[X] := \sum_a a\, \mathsf{Prob}[X = a].$$

- For any set $X_1, \ldots, X_n$ of random variables, and any constants $c_1, \ldots, c_n$

$$\mathsf{E}[c_1 X_1 + \cdots + c_n X_n] = c_1 \mathsf{E}[X_1] + \cdots + c_n \mathsf{E}[X_n]$$

This fact is called linearity of expectation

PTCF: Indicator/Bernoulli Random Variable

$$X : \Omega \to \{0, 1\}$$

$$p = \mathsf{Prob}[X = 1]$$

X is called a Bernoulli random variable with parameter p

If $X = 1$ only for outcomes ω belonging to some event A, then X is called an indicator variable for A

$$\begin{aligned} \mathsf{E}[X] &= p \\ \mathsf{Var}[X] &= p(1-p) \end{aligned}$$

Las Vegas and Monte Carlo Algorithms

Las Vegas Algorithm

A randomized algorithm which always gives the correct solution is called a Las Vegas algorithm.
Its running time is a random variable.

Monte Carlo Algorithm

A randomized algorithm which may give incorrect answers (with certain probability) is called a Monte Carlo algorithm.
Its running time may or may not be a random variable.

Example 3: Max-E3SAT

- An E3-CNF formula is a CNF formula φ in which each clause has *exactly* 3 literals. E.g.,

$$\varphi = (x_1 \vee \bar{x}_2 \vee x_4) \wedge (x_1 \vee x_3 \vee \bar{x}_4) \wedge (\bar{x}_2 \vee \bar{x}_3 \vee x_4)$$

- Max-E3SAT Problem: given a E3-CNF formula φ, find a truth assignment satisfying as many clauses as possible

A Randomized Approximation Algorithm for Max-E3SAT

- Assign each variable to $\textsc{true}/\textsc{false}$ with probability $1/2$

Analyzing the Randomized Approximation Algorithm

- Let X_C be the random variable indicating if clause C is satisfied
- Then, $\mathsf{Prob}[X_C = 1] = 7/8$
- Let S_φ be the number of satisfied clauses
- Hence,

$$\mathsf{E}[S_\varphi] = \mathsf{E}\left[\sum_C X_C\right] = \sum_C \mathsf{E}[X_C] = 7m/8 \leq \frac{\text{OPT}}{8/7}$$

  (m is the number of clauses)
- So this is a randomized approximation algorithm with ratio $8/7$

Derandomization with Conditional Expectation Method

- Derandomization is to turn a randomized algorithm into a deterministic algorithm
- By conditional expectation

$$\mathsf{E}[S_\varphi] = \frac{1}{2}\mathsf{E}[S_\varphi \mid x_1 = \text{ TRUE}] + \frac{1}{2}\mathsf{E}[S_\varphi \mid x_1 = \text{ FALSE}]$$

- Both $\mathsf{E}[S_\varphi \mid x_1 = \text{ TRUE}]$ and $\mathsf{E}[S_\varphi \mid x_1 = \text{ FALSE}]$ can be computed in polynomial time
- Suppose $\mathsf{E}[S_\varphi \mid x_1 = \text{ TRUE}] \geq \mathsf{E}[S_\varphi \mid x_1 = \text{ FALSE}]$, then

$$\mathsf{E}[S_\varphi \mid x_1 = \text{ TRUE}] \geq \mathsf{E}[S_\varphi] \geq 7m/8$$

- Set $x_1 =$TRUE, let φ' be φ with c clauses containing x_1 removed, and all instances of $x_1, \bar{x}_1$ removed.
- Recursively find value for x_2

PTCF: Law of Total Probabilities, Conditional Expectation

- Law of total probabilities: let $A_1, A_2, \dots$ be any sequence of mutually exclusive events, then
$$\mathsf{Prob}[A] = \sum_{i \geq 1} \mathsf{Prob}[A \mid A_i] \, \mathsf{Prob}[A_i]$$
- The conditional expectation of X given A is
$$\mathsf{E}[X \mid A] := \sum_a a \, \mathsf{Prob}[X = a \mid A].$$
- Let $A_1, A_2, \dots$ be any sequence of mutually exclusive events, then
$$\mathsf{E}[X] = \sum_{i \geq 1} \mathsf{E}[X \mid A_i] \, \mathsf{Prob}[A_i]$$
- In particular, let Y be any discrete random variable, then
$$\mathsf{E}[X] = \sum_y \mathsf{E}[X \mid Y = y] \, \mathsf{Prob}[Y = y]$$

Example 1: Probabilistic Packet Marking (PPM)

The Setting

- A stream of packets are sent $S = R_0 \to R_1 \to \cdots \to R_{n-1} \to D$
- Each R_i can overwrite the SOURCE IP field
- D wants to know the set of routers on the route

The Assumption

- For each packet D receives and each i, Prob$[F = R_i] = 1/n$ (*)

The Questions

1. How does the routers ensure (*)?
2. How many packets must D receive to know all routers?

Coupon Collector Problem

The setting

- n types of coupons
- Every cereal box has a coupon
- For each box B and each coupon type t,

$$\text{Prob}\left[B \text{ contains coupon type } t\right] = \frac{1}{n}$$

Coupon Collector Problem

How many boxes of cereal must the collector purchase before he has all types of coupons?

The Analysis

- $X =$ number of boxes he buys to have all coupon types.
- For $i \in [n]$, let X_i be the additional number of cereal boxes he buys to get a new coupon type, after he had collected $i-1$ different types

$$X = X_1 + X_2 + \cdots + X_n, \quad \mathsf{E}[X] = \sum_{i=1}^{n} E[X_i]$$

- After $i-1$ types collected, a new box contains a new type with prob

$$p_i = 1 - \frac{i-1}{n}$$

- Hence, X_i is *geometric* with parameter p_i, implying

$$\mathsf{E}[X_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$$

$$\mathsf{E}[X] = n \sum_{i=1}^{n} \frac{1}{n-i+1} = nH_n = n \ln n + \Theta(n)$$

PTCF: Geometric Distribution

- A coin turns head with probability p, tail with $1 - p$
- $X =$ number of flips until a head shows up
- X has geometric distribution with parameter p

$$\begin{aligned} \text{Prob}[X = n] &= (1-p)^{n-1}p \\ \text{E}[X] &= \frac{1}{p} \\ \text{Var}[X] &= \frac{1-p}{p^2} \end{aligned}$$

Additional Questions

- We can't be sure that buying nH_n cereal boxes suffices
- Want Prob$[X \geq C]$, i.e. *what's the probability that he has to buy C boxes to collect all coupon types?*
- Intuitively, X is far from its mean with small probability
- Want something like

$$\text{Prob}[X \geq C] \leq \text{some function of } C, \text{ preferably} \ll 1$$

i.e. (large) deviation inequality or tail inequalities

Central Theme

The more we know about X, the better the deviation inequality we can derive: Markov, Chebyshev, Chernoff, etc.

PTCF: Markov's Inequality

Theorem

If X is a r.v. taking only non-negative values, $\mu = \mathsf{E}[X]$, then $\forall a > 0$

$$\mathsf{Prob}[X \geq a] \leq \frac{\mu}{a}.$$

Equivalently,

$$\mathsf{Prob}[X \geq a\mu] \leq \frac{1}{a}.$$

If we know $\mathsf{Var}[X]$, we can do better!

PTCF: (Co)Variance, Moments, Their Properties

- Variance: $\sigma^2 = \mathsf{Var}[X] := \mathsf{E}[(X - \mathsf{E}[X])^2] = \mathsf{E}[X^2] - (\mathsf{E}[X])^2$
- Standard deviation: $\sigma := \sqrt{\mathsf{Var}[X]}$
- kth moment: $\mathsf{E}[X^k]$
- Covariance: $\mathsf{Cov}[X,Y] := \mathsf{E}[(X - \mathsf{E}[X])(Y - \mathsf{E}[Y])]$
- For any two r.v. X and Y,

$$\mathsf{Var}[X+Y] = \mathsf{Var}[X] + \mathsf{Var}[Y] + 2\,\mathsf{Cov}[X,Y]$$

- If X and Y are independent (define it), then

$$\begin{aligned} \mathsf{E}[X \cdot Y] &= \mathsf{E}[X] \cdot \mathsf{E}[Y] \\ \mathsf{Cov}[X,Y] &= 0 \\ \mathsf{Var}[X+Y] &= \mathsf{Var}[X] + \mathsf{Var}[Y] \end{aligned}$$

- In fact, if $X_1, \dots, X_n$ are mutually independent, then

$$\mathsf{Var}\left[\sum_i X_i\right] = \sum_i \mathsf{Var}[X_i]$$

PTCF: Chebyshev's Inequality

Theorem (Two-sided Chebyshev's Inequality)

If X is a r.v. with mean μ and variance σ^2, then $\forall a > 0$,

$$\text{Prob}\big[|X-\mu| \geq a\big] \leq \frac{\sigma^2}{a^2} \text{ or, equivalently } \text{Prob}\big[|X-\mu| \geq a\sigma\big] \leq \frac{1}{a^2}.$$

Theorem (One-sided Chebyshev's Inequality)

Let X be a r.v. with $\mathsf{E}[X] = \mu$ *and* $\mathsf{Var}[X] = \sigma^2$, *then* $\forall a > 0$,

$$\begin{aligned}
\text{Prob}[X \geq \mu + a] &\leq \frac{\sigma^2}{\sigma^2 + a^2} \\
\text{Prob}[X \leq \mu - a] &\leq \frac{\sigma^2}{\sigma^2 + a^2}.
\end{aligned}$$

Back to the Additional Questions

- Markov's leads to,

$$\text{Prob}[X \geq 2nH_n] \leq \frac{1}{2}$$

- To apply Chebyshev's, we need $\text{Var}[X]$:

$$\text{Prob}[|X - nH_n| \geq nH_n] \leq \frac{\text{Var}[X]}{(nH_n)^2}$$

- Key observation: the X_i are independent (why?)

$$\text{Var}[X] = \sum_i \text{Var}[X_i] = \sum_i \frac{1-p_i}{p_i^2} \leq \sum_i \frac{n^2}{(n-i+1)^2} = \frac{\pi^2 n^2}{6}$$

- Chebyshev's leads to

$$\text{Prob}[|X - nH_n| \geq nH_n] \leq \frac{\pi^2}{6H_n^2} = \Theta\left(\frac{1}{\ln^2 n}\right)$$

Example 2: PPM with One Bit

The Problem

Alice wants to send to Bob a message $b_1 b_2 \cdots b_m$ of m bits. She can send only **one** bit at a time, but always forgets which bits have been sent. Bob knows m, nothing else about the message.

The solution

- Send bits so that the fraction of bits 1 received is within ϵ of $p = B/2^m$, where $B = b_1 b_2 \cdots b_m$ as an integer
- Specifically, send bit 1 with probability p, and 0 with $(1-p)$

The question

How many bits must be sent so B can be decoded with high probability?

The Analysis

- One way to do decoding: round the fraction of bits 1 received to the closest multiple of of $1/2^m$
- Let $X_1, \dots, X_n$ be the bits received (independent Bernoulli trials)
- Let $X = \sum_i X_i$, then $\mu = \mathsf{E}[X] = np$. We want, say

$$\text{Prob}\left[\left|\frac{X}{n} - p\right| \leq \frac{1}{3 \cdot 2^m}\right] \geq 1 - \epsilon$$

which is equivalent to

$$\text{Prob}\left[|X - \mu| \leq \frac{n}{3 \cdot 2^m}\right] \geq 1 - \epsilon$$

This is a kind of concentration inequality.

PTCF: The Binomial Distribution

- n independent trials are performed, each with success probability p.
- $X =$ number of successes after n trials, then

$$\text{Prob}[X = i] = \binom{n}{i} p^i (1-p)^{n-i}, \ \forall i = 0, \dots, n$$

- X is called a binomial random variable with parameters (n, p).

$$\begin{aligned} \mathsf{E}[X] &= np \\ \mathsf{Var}[X] &= np(1-p) \end{aligned}$$

PTCF: Chernoff Bounds

Theorem (Chernoff bounds are just the following idea)

Let X be any r.v., then

1. *For any $t > 0$*

$$\text{Prob}[X \geq a] \leq \frac{\mathsf{E}[e^{tX}]}{e^{ta}}$$

   In particular,

$$\text{Prob}[X \geq a] \leq \min_{t>0} \frac{\mathsf{E}[e^{tX}]}{e^{ta}}$$

2. *For any $t < 0$*

$$\text{Prob}[X \leq a] \leq \frac{\mathsf{E}[e^{tX}]}{e^{ta}}$$

   In particular,

$$\text{Prob}[X \geq a] \leq \min_{t<0} \frac{\mathsf{E}[e^{tX}]}{e^{ta}}$$

(E^{tX} is called the moment generating function of X)

PTCF: A Chernoff Bound for sum of Poisson Trials

Above the mean case.

Let $X_1, \dots, X_n$ be independent Poisson trials, $\text{Prob}[X_i = 1] = p_i$, $X = \sum_i X_i$, $\mu = \mathsf{E}[X]$. Then,

- For any $\delta > 0$,

$$\text{Prob}[X \geq (1+\delta)\mu] < \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu;$$

- For any $0 < \delta \leq 1$,

$$\text{Prob}[X \geq (1+\delta)\mu] \leq e^{-\mu\delta^2/3};$$

- For any $R \geq 6\mu$,

$$\text{Prob}[X \geq R] \leq 2^{-R}.$$

PTCF: A Chernoff Bound for sum of Poisson Trials

Below the mean case.

Let $X_1, \dots, X_n$ be independent Poisson trials, $\text{Prob}[X_i = 1] = p_i$, $X = \sum_i X_i$, $\mu = \text{E}[X]$. Then, for any $0 < \delta < 1$:

1. $$\text{Prob}[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}} \right)^{\mu};$$

2. $$\text{Prob}[X \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}.$$

PTCF: A Chernoff Bound for sum of Poisson Trials

A simple (two-sided) deviation case.
Let $X_1, \dots, X_n$ be independent Poisson trials, $\text{Prob}[X_i = 1] = p_i$, $X = \sum_i X_i$, $\mu = \mathsf{E}[X]$. Then, for any $0 < \delta < 1$:

$$\text{Prob}[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}.$$

Chernoff Bounds Informally

The probability that the sum of independent Poisson trials is far from the sum's mean is exponentially small.

Back to the 1-bit PPM Problem

$$\begin{aligned}\text{Prob}\left[|X-\mu| > \frac{n}{3\cdot 2^m}\right] &= \text{Prob}\left[|X-\mu| > \frac{1}{3\cdot 2^m p}\mu\right] \\ &\leq \frac{2}{\exp\{\frac{n}{18\cdot 4^m p}\}}\end{aligned}$$

Now,

$$\frac{2}{\exp\{\frac{n}{18\cdot 4^m p}\}} \leq \epsilon$$

is equivalent to

$$n \geq 18p\ln(2/\epsilon)4^m.$$

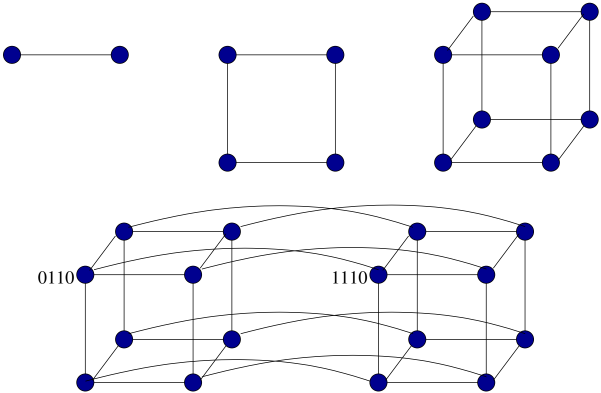
Example 3: Oblivious Routing on the Hypercube

- Directed graph $G = (V, E)$: network of parallel processors
- Permutation Routing Problem
  - Each node v contains one packet P_v, $1 \leq v \leq N = |V|$
  - Destination for packet from v is π_v, $\pi \in S_n$
  - Time is discretized into unit steps
  - Each packet can be sent on an edge in one step
  - Queueing discipline: FIFO
- Oblivious algorithm: route R_v for P_v depends on v and π_v only
- Question: in the worst-case (over π), how many steps must an oblivious algorithm take to route all packets?

Theorem (Kaklamanis et al, 1990)

Suppose G has N vertices and out-degree d. For any deterministic oblivious algorithm for the permutation routing problem, there is an instance π which requires $\Omega(\sqrt{N/d})$ steps.

The (Directed) Hypercube



- The n-cube: $|V| = N = 2^n$, vertices $\mathbf{v} \in \{0,1\}^n$, $\mathbf{v} = v_1 \cdots v_n$
- $(\mathbf{u}, \mathbf{v}) \in E$ iff their Hamming distance is 1

The Bit-Fixing Algorithm

- Source $\mathbf{u} = u_1 \cdots u_n$, target $\pi_u = v_1 \cdots v_n$
- Suppose the packet is currently at $\mathbf{w} = w_1 \cdots w_n$, scan $\mathbf{w}$ from left to right, find the first place where $w_i \neq v_i$
- Forward packet to $w_1 \cdots w_{i-1} v_i w_{i+1} \cdots w_n$

Source	010011
	110010
	100010
	100110
Destination	100111

- There is a π requiring $\Omega(\sqrt{N/n})$ steps

Valiant Load Balancing Idea

Les Valiant, *A scheme for fast parallel communication*, SIAM J. Computing, 11: 2 (1982), 350-361.

Two phase algorithm (input: π)

- **Phase 1:** choose $\sigma \in S_N$ uniformly at random, route P_v to σ_v with bit-fixing
- **Phase 2:** route P_v from σ_v to π_v with bit-fixing

This scheme is now used in designing Internet routers with high throughput!

Phase 1 Analysis

- P_u takes route $R_u = (e_1, \ldots, e_k)$ to σ_u
- Time taken is k ($\leq n$) plus queueing delay

Lemma

If R_u and R_v share an edge, once R_v leaves R_u it will not come back to R_u

Theorem

Let S be the set of packets other than packet P_u whose routes share an edge with R_u, then the queueing delay incurred by packet P_u is at most $|S|$

Phase 1 Analysis

- Let H_{uv} indicate if R_u and R_v share an edge
- Queueing delay incurred by P_u is $\sum_{v \neq u} H_{uv}$.
- We want to bound

$$\text{Prob}\left[\sum_{v \neq u} H_{uv} > \alpha n\right] \geq \;\; ??$$

- Need an upper bound for $\mathsf{E}\left[\sum_{v \neq u} H_{uv}\right]$
- For each edge e, let T_e denote the number of routes containing e

$$\sum_{v \neq u} H_{uv} \leq \sum_{i=1}^{k} T_{e_i}$$

$$\mathsf{E}\left[\sum_{v \neq u} H_{uv}\right] \leq \sum_{i=1}^{k} \mathsf{E}[T_{e_i}] = k/2 \leq n/2$$

Conclusion

- By Chernoff bound,

$$\text{Prob}\left[\sum_{v \neq u} H_{uv} > 6n\right] \leq 2^{-6n}$$

- Hence,

Theorem

With probability at least $1 - 2^{-5n}$, every packet reaches its intermediate target (σ) in Phase 1 in $7n$ steps

Theorem (Conclusion)

With probability at least $1 - 1/N$, every packet reaches its target (π) in $14n$ steps

Example 1: 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
- $\mathbf{f}(\mathbf{x})$ 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

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

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

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

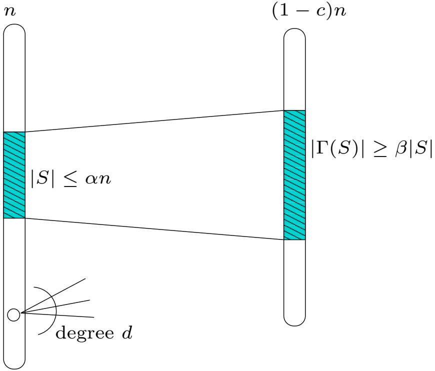
(Yes, using "magical graphs.")

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 $\beta > d/2$
- 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 A_s be the 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,

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

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

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

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

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

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

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

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

Example 2: Non-Adaptive Group Testing

- A $t \times n$ 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

Existence of Good d-disjunct Matrix

- Set a_{ij} to 1 with probability p
- The probability that $\mathbf{A}$ is **not** d-disjunct is at most

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

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

- This is < 1 as long as

$$t \geq 3(d+1)\ln\left[(d+1)\binom{n}{d+1}\right]$$

- In particular, for large n, there exist d-disjunct matrices with rate

$$\frac{\log n}{t} \approx \frac{1}{3(d+1)^2}$$