APPLICATIONS OF CHERNOFF BOUNDS

SAMIR RAJANI

Abstract. We motivate and derive the theory of Chernoff bounds and examine four of their most significant applications. Chernoff bounds are tight bounds on the tail distributions of the sums of independent random variables. We discuss the sampling problem, which appears most frequently when polling a representative subset of a large population. We also provide a treatment of the set balancing problem, in which we guarantee with high probability balance in binary characteristics between groups using a trivial randomized algorithm. We then bound the discrepancy of a hypergraph, the maximum difference between the number of red and blue vertices in a single edge. Finally, we discuss the permutation routing problem, which guarantees linear-time communication on an n-dimensional network with high probability.

Contents

1. Moment Generating Functions and Markov's Inequality

We seek to derive a probabilistic tool known as the Chernoff Bound, a useful bound on deviation from the expected value of the sum of independent random variables. First, we introduce moments, which are typically used to calculate values like the mean, variance, and standard deviation of a random variable.

Definition 1.1. The nth moment of a random variable X is given by $\mathbb{E}[X^n]$.

Definition 1.2. The moment generating function of a random variable X is given by $M_X(t) = \mathbb{E}[e^{tX}]$.

To justify this definition and nomenclature, we will demonstrate the relationship between the moment and the moment generating function. In particular, the moment generating function allows us to derive moments by differentiating. This, however, requires the commutativity of differentiation and expectation; it turns out this assumption is valid when the moment generating function exists near zero. While a formal proof of this statement is beyond the scope of this paper, more information can be found in [1].

Theorem 1.3. *If the moment generating function exists in a neighborhood around zero, for a random variable X with moment generating function $M_X(t)$, $\mathbb{E}[X^n] = M_X^{(n)}(0)$ for all $n > 1$.*

Proof. Since the moment generating function exists in an neighborhood around zero, we can assume commutativity of differentiation and expectation. We have that:

$$\begin{aligned} M_X^{(n)}(t) &= \mathbb{E}[\frac{d}{dX^n}(e^{tX})|_{x=0}] \\ &= \mathbb{E}[X^n e^0] \\ &= \mathbb{E}[X^n] \end{aligned}$$

□

We next derive an elementary yet ubiquitous property, linearity of expectation. This is an essential tool for manipulating and deriving the expectations of random variables.

Theorem 1.4. *For any discrete random variables $X_1, X_2, \cdots, X_n$, $\mathbb{E}[\sum_{i=1}^n X_i] = \sum_{i=1}^n \mathbb{E}[X_i]$.*

Proof. For $n = 2$, we have that:

$$\begin{aligned} \mathbb{E}[X+Y] &= \sum_x \sum_y (x+y) \Pr(X = x \cap Y = y) \\ &= \sum_x \sum_y x \Pr(X = x \cap Y = y) + \sum_x \sum_y y \Pr(X = x \cap Y = y) \\ &= \sum_x x \sum_y \Pr(X = x \cap Y = y) + \sum_y y \sum_x \Pr(X = x \cap Y = y) \\ &= \sum_x x \Pr(X = x) + \sum_y y \Pr(Y = y) \\ &= \mathbb{E}[X] + \mathbb{E}[Y] \end{aligned}$$

The proof follows from induction on n.

□

We now construct and prove Markov's Inequality, a rather primitive tail bound. We examine this bound not for its occasional usefulness but rather for its role in developing the remarkably strong Chernoff Bound.

Theorem 1.5. *Given a non-negative random variable X, for all $a > 0$,*

$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

.

Proof. Let $a > 0$ be arbitrary. We define an indicator variable I as follows:

$$I = \begin{cases} 0 & X < a \\ 1 & X \geq a \end{cases}$$

In the case that $x < a$, we have that $I = 0$, so $I \leq \frac{X}{a}$. Further, in the case that $x \geq a$, we have that $\frac{X}{a} \geq 1$, so $I \leq \frac{X}{a}$ again. Finally, we have:

$$\begin{aligned} \Pr(X \geq a) &= \mathbb{E}[I] && \text{since } I \text{ is an indicator variable} \\ &\leq \mathbb{E}[\frac{X}{a}] && \text{since } I \leq \frac{X}{a} \\ &= \frac{\mathbb{E}[X]}{a} && \text{by linearity of expectation} \end{aligned}$$

□

2. Chernoff Bounds

As stated previously, Markov's inequality plays a crucial role in the formulation of Chernoff bounds. Instead of applying Markov's inequality to the random variable X itself, we apply it to the moment generating function. In general, Chernoff bounds will be achieved by varying the parameter t of the moment generating function until the tightest bound is attained. This process can be summarized succinctly as such:

$$\begin{aligned} \Pr(X \geq a) &= \Pr(e^{tX} \geq e^{ta}) && \text{equivalent conditions for } t > 0 \\ &\leq \min\{\frac{\mathbb{E}[e^{tX}]}{e^{ta}} \mid t > 0\} && \text{by Markov's Inequality} \end{aligned}$$

$$\begin{aligned} \Pr(X \leq a) &= \Pr(e^{tX} \geq e^{ta}) && \text{equivalent conditions for } t < 0 \\ &\leq \min\{\frac{\mathbb{E}[e^{tX}]}{e^{ta}} \mid t < 0\} && \text{by Markov's Inequality} \end{aligned}$$

Definition 2.1. Bernoulli trials are trials of random variables which are independent, take on values of 0 and 1, and retain their probability distributions across trials. Poisson trials are trials of random variables which are independent and take on values of 0 and 1, but whose probability distribution may vary across trials.

Notation 2.2. For a sequence of Poisson trials with $\Pr(X_i = 1) = p_i$, and $X = \sum_{i=1}^{n} X_i$, we define $\mu = \mathbb{E}[X] = \sum_{i=1}^{n} p_i$, where the second equality follows from linearity.

Lemma 2.3. *Given a sequence of Poisson trials X_i, the following inequality holds: $M_X(t) \leq e^{\mu(e^t-1)}$.*

Proof. We first expand the moment generating function of X to get it in terms of each individual trial. We have:

$$M_X(t) = \mathbb{E}[e^{tX}] = \mathbb{E}[e^{t\sum_{i=1}^{n} X_i}] = \mathbb{E}[\prod_{i=1}^{n} e^{tX_i}] = \prod_{i=1}^{n} \mathbb{E}[e^t X_i] = \prod_{i=1}^{n} M_{X_i}(t)$$

The fourth equality follows from the independence of Poisson trials. Next, we bound the moment generating function of each trial as follows:

$$\begin{aligned} M_{X_i}(t) &= \mathbb{E}[e^{tX_i}] \\ &= p_i e^t + (1-p_i)e^0 \\ &= 1 + p_i(e^t - 1) \\ &\leq e^{p_i(e^t-1)} && \text{for all } y \in \mathbb{R},\ 1 + y \leq e^y \end{aligned}$$

Finally, we plug this bound into the first equation we derived:

$$M_X(t) = \prod_{i=1}^{n} M_{X_i}(t) \leq \prod_{i=1}^{n} e^{p_i(e^t-1)} = e^{\sum_{i=1}^{n} p_i(e^t-1)} = e^{\mu(e^t-1)}$$

□

Theorem 2.4. *For independent Poisson trials X_i, the following inequality holds for all $\delta > 0$: $\Pr(X \geq (1+\delta)\mu) \leq (\frac{e^\delta}{(1+\delta)^{(1+\delta)}})^\mu$.*

Proof. Using the general form for the Chernoff bound discussed earlier, we have that:

$$\begin{aligned} \Pr(X \geq (1+\delta)\mu) &= \Pr(e^{tX} \geq e^{t(1+\delta)\mu}) && \text{equivalent statements for } \delta > 0 \\ &\leq \frac{\mathbb{E}[e^{tX}]}{e^{t(1+\delta)\mu}} && \text{by Markov's inequality} \\ &\leq \frac{e^{\mu(e^t-1)}}{e^{t(1+\delta)\mu}} && \text{by Lemma 2.3} \end{aligned}$$

As mentioned previously, we'd like to choose an optimal value of t to obtain as tight a bound as possible. In other words, the goal is to choose a value of t that minimizes the right side of the inequality, accomplished through differentiation below:

$$\begin{aligned} \frac{d}{dt}[e^{\mu(e^t-1-t-t\delta)}] &= 0 \\ e^{\mu(e^t-1-t-t\delta)}(e^t - 1 - \delta) &= 0 \\ e^t &= 1 + \delta \end{aligned}$$

Hence, the ideal choice of t for our bound is $\ln(1+\delta)$. Substituting this value into our expression, we find that $\Pr(X \geq (1+\delta)\mu) \leq (\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}})^{\mu}$. □

This bound is quite cumbersome to use, so it is useful to provide a slightly less unwieldy bound, albeit one that sacrifices some generality and strength.

Theorem 2.5. *For independent Poisson trials X_i, the following inequality holds for $0 < \delta < 1$:* $\Pr(X \geq (1+\delta)\mu) \leq e^{-\frac{\mu\delta^2}{3}}$.

Proof. We rewrite the right-hand side of our expression to yield

$$\Pr(X \geq (1+\delta)\mu) \leq (e^{\delta-(1+\delta)\ln(1+\delta)})^{\mu}$$

. This new form lends itself to a Taylor expansion:

$$\ln(1+\delta) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}\delta^k}{k}$$

$$(1+\delta)\ln(1+\delta) = \delta + \sum_{k=2}^{\infty} \frac{(-1)^{k+1}\delta^k}{k} + \sum_{k=2}^{\infty} \frac{(-1)^k\delta^k}{k-1}$$

$$(1+\delta)\ln(1+\delta) = \delta + \sum_{k=2}^{\infty}(-1)^k\delta^k(\frac{1}{k-1} - \frac{1}{k})$$

Assuming $0 < \delta < 1$, we have that

$$(1+\delta)\ln(1+\delta) \geq \delta + \frac{\delta^2}{2} - \frac{\delta^3}{6} \geq \delta + \frac{\delta^2}{2} - \frac{\delta^2}{6} = \delta + \frac{\delta^2}{3}$$

. Plugging into our altered expression, we have:

$$\begin{aligned}\Pr(X > (1+\delta)\mu) &\leq (e^{\delta-(1+\delta)\ln(1+\delta)})^{\mu} \\ &\leq e^{\delta-(\delta+\frac{\delta^2}{3})}\mu \\ &= e^{-\frac{\mu\delta^2}{3}}\end{aligned}$$

□

We present without proof a slightly different, but analogous theorem for bounds on the lower end of the tail distribution. A proof of this theorem can be found in [1].

Theorem 2.6. *For independent Poisson trials X_i, the following inequality holds for $0 < \delta < 1$:* $\Pr(X \leq (1-\delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}$.

By simply loosening the upper end Chernoff bound and adding it with the lower end bound, we obtain the following corollary.

Corollary 2.7. *For independent Poisson trials X_i, when $0 < \delta < 1$, the following inequality holds:* $\Pr(|X-\mu| \geq \delta\mu) \leq 2e^{-\frac{\mu\delta^2}{3}}$.

3. A Sampling Theorem

To estimate the percentage of people in a large population who hold some binary opinion or have some binary characteristic, we often sample a smaller subset of the population, insisting that the subset is representative of the larger population. To guarantee a certain accuracy with high probability, we must bound deviations from the expectation; one way this can be done is with a Chernoff bound for $\bar{X}$, the average of the results of Bernoulli trials X_i. Note that to apply a Chernoff bound, people must be chosen with replacement so that indicator variables are independent. We formulate and prove a theorem that models this situation below. Note that, to improve legibility, we use the notation $\exp\{x\}$ to denote the exponential function e^x.

Theorem 3.1. *Given independent 0-1 random variables X_i with $X = \sum_{i=1}^{n} X_i$, $\Pr(X_i = 1) = p$, and $\bar{X} = \frac{X}{n}$, if $n \geq \frac{3}{\theta^2}\ln(\frac{2}{\delta})$, then $\Pr(|\bar{X} - p| \leq \theta) \geq 1 - \delta$.*

Proof. To apply a Chernoff Bound, we first compute $\mathbb{E}[X]$; the second equality follows from linearity.

$$\mathbb{E}[X] = \mathbb{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathbb{E}[X_i] = pn$$

Applying the Chernoff Bound, we have

$$\begin{aligned}\Pr(|\bar{X} - p| \geq \epsilon p) &= \Pr(|X - pn| \geq \epsilon pn) \\ &\leq 2\exp(-\frac{\epsilon^2}{3}pn)\end{aligned}$$

We would like a formula directly in terms of the deviation from the expectation, not as a multiple of the expectation. Hence, we let $\epsilon = \frac{\theta}{p}$ so that $\theta = \epsilon p$.

$$\begin{aligned}\Pr(|\bar{X} - p| \geq \theta) &\leq 2\exp(\frac{-\theta^2}{3p^2} \cdot pn) \\ &= 2\exp(\frac{-n\theta^2}{3p}) \\ &\leq 2\exp(\frac{-n\theta^2}{3}) && \text{since } p \leq 1\end{aligned}$$

To have $\Pr(|\bar{X} - p| \leq \theta) \geq 1 - \delta$, we need $\Pr(|\bar{x} - p| \geq \theta) \leq \delta$. Hence, we require that

$$\begin{aligned}2\exp(\frac{-n\theta^2}{3}) &\leq \delta \\ -\frac{-n\theta^2}{3} &\leq \delta \\ -\frac{n\theta^2}{3} &\leq \ln(\frac{\delta}{2}) \\ n &\geq \frac{3}{\theta^2}\ln(\frac{2}{\delta})\end{aligned}$$

□

4. Set Balancing

Suppose you're in a charge of running a clinical trial, and you'd like to split test subjects into a control group and an experimental group. As such, you'd like to minimize any possible bias from an imperfect dichotomy, such as a discrepancy in the average age between the two groups. Almost immediately, a question arises: if we assign each participant to a group at random, how confident can we be that biases like age and gender aren't significant influences in the conclusions we draw from the experiment. In mathematics, the set balancing problem models such a situation, and the question we have posed can be answered adequately through Chernoff bounds.

We represent the situation with a matrix; each column models a person, and each row models a trait.

In the proof of the following theorem, we use Big-O notation; a definition is provided below. We use this notation even when Big-Theta notation might be more precise because we are generally interested in probabilistic guarantees rather than rigorous mathematical descriptions of the function in question.

Definition 4.1. We say that a function $f(n) = O(g(n))$ if there exist constants $c, n_0 > 0$. such that for all $n \geq n_0$, $f(n) \leq c \cdot g(n)$.

Theorem 4.2. *Let* $\mathbf{A}$ *be an* $n \times m$ *matrix, with each* $a_{ij} \in \{0, 1\}$*, and let* $\vec{b}$ *be an* m*-dimensional vector with each* $b_k \in \{-1, 1\}$*, where each possibility is chosen with probability* $\frac{1}{2}$*. Let* $\vec{c}$ *be the* n*-dimensional vector that denotes the product of* $\mathbf{A}$ *and* $\vec{b}$*. Then, the following inequality holds for* $i \in \{1, \cdots, n\}$:

$$\Pr(\max\{|c_i|\} \geq \sqrt{4m \ln n}) \leq O(n^{-1})$$

Proof. Take row i of matrix $\mathbf{A}$ to be arbitrary. Denote by s the sum of the elements in that row, $\sum_{j=1}^{m} a_{ij}$. In the case that $s \leq \sqrt{4m \ln n}$, we have $|c_i| \leq \sqrt{4m \ln n}$, since each value is getting multiplied only by 1 or -1. This eliminates the possibility of $s = 0$, which would prevent us from evaluating the Chernoff bound. Otherwise, if $s > \sqrt{4m \ln n}$, taking each non-zero term $a_{ij} b_j$ in the sum as an independent random variable taking on values 1 and -1 with equal probability, we have:

$$\begin{aligned} \Pr(|c_i| \geq \sqrt{4m \ln n}) &\leq 2e^{\frac{-4n \ln m}{2k}} && \text{by the Chernoff Bound} \\ &\leq \frac{2}{n^2} && \text{since } k \leq m \end{aligned}$$

Now, we consider the probability that this occurs for any given row:

$$\begin{aligned} \Pr(\bigcup_i |c_i| \geq \sqrt{4m \ln n}) &\leq \sum_{i=1}^{n} \Pr(|c_i| \geq \sqrt{4m \ln n}) && \text{by the union bound} \\ &= O(n^{-1}) \end{aligned}$$

This is a useful result in particular because large values of n, which have the best probability guarantees, are precisely the studies which render manual set balancing unreasonable.

□

5. Discrepancy and Hypergraph Coloring

We next prove a useful bound regarding the edges of a two-colored hypergraph.

Definition 5.1. A hypergraph $H = (V, E)$ is a pair of vertices V and edges E, where V is a finite set and E is a set of subsets of V consisting of two or more elements.

Definition 5.2. A two-coloring of a hypergraph $H = (V, E)$ is an assignment of either red or blue to each vertex $v \in V$, each with probability $\frac{1}{2}$.

Definition 5.3. Given a two-coloring of a hypergraph $H = (V, E)$ with edges $E_1, E_2, \cdots, E_m$, the discrepancy of H, written $\mathrm{Disc}(H)$, is given by the following: $\mathrm{Disc}(H) = \max_{1 \le i \le m}\{|\text{red vertices in } E_i - \text{blue vertices in } E_i|\}$.

Theorem 5.4. *Let $H = (V, E)$ be a hypergraph with n vertices and m edges. If H is randomly two-colored, then with probability $1 - O(m^{-1})$, $\mathrm{Disc}(H) \le \sqrt{12n \ln m}$.*

Proof. We begin by creating an indicator variable I_v corresponding with the color of each vertex:

$$I_v = \begin{cases} 0 & \text{if vertex v is colored blue} \\ 1 & \text{if vertex v is colored red} \end{cases}$$

First, consider a single edge $E_i \in E$. Define $I = \sum_{v \in E_i} I_v$ to be the number of red vertices in edge E_i. Let k denote the total number of vertices in the edge, $|E_i|$. It is apparent that $\mathrm{Disc}(E_i) = 2|\frac{k}{2} - I|$, because any deviation from an equal number of red and blue vertices increases the number of red vertices by 1 and decreases the number of blue vertices by 1. Notably, since $\mathbb{E}[I] = \frac{k}{2}$, this formula is similar to that for deviations from the expected number of red vertices. Hence, using a Chernoff bound on I is appropriate here; we do this first with a placeholder λ, and then choose an appropriate value for λ to guarantee low probability of significant deviations. We have:

$$\begin{aligned} \Pr(|I - \frac{k}{2}| \ge \lambda) &= \Pr(|I - \frac{k}{2}| \ge (\frac{\lambda}{\mu})\mu) \\ &\le 2\exp\{-\frac{\mu(\frac{\lambda}{\mu})^2}{3}\} && \text{by the Chernoff bound} \\ &= 2\exp\{-\frac{2\lambda^2}{3k}\} \end{aligned}$$

Now, substituting $\lambda = \sqrt{3k \ln m}$, we have:

$$\begin{aligned} \Pr(\mathrm{Disc}(E_i) \ge \sqrt{12k \ln m}) &= \Pr(|I - \frac{k}{2}| \ge \sqrt{3k \ln m}) \\ &\le 2\exp\{\frac{-2(3k \ln m)}{3k}\} \\ &= 2\exp\{\ln m^{-2}\} \\ &= \frac{2}{m^2} \end{aligned}$$

Finally, we can combine the edges together. Because each edge contains at most n vertices, we have:

$$\begin{aligned}\Pr(\bigcup_i \mathrm{Disc}(E_i) \geq \sqrt{12n\ln m}) &\leq \Pr(\bigcup_i \mathrm{Disc}(E_i) \geq \sqrt{12k\ln m}) \\ &\leq \sum_i \Pr(\mathrm{Disc}(S_i) \geq \sqrt{12n\ln m}) \quad \text{by the Union Bound} \\ &\leq m(\frac{2}{m^2}) \\ &= O(m^{-1})\end{aligned}$$

□

6. Permutation Routing

A discussion regarding probabilistic analysis of networks is ostensibly one which would require domain-level knowledge of how the network is being used: who is communicating with whom? However, the ability of a randomized algorithm to establish performance guarantees means we can study networks without this data. Given an n-dimensional hypercube, which we will use as a mathematical model of a network, we seek to route packets from a starting node (sender) to a destination node (receiver) within a short amount of time, with high probability.

Without domain knowledge, we are unsure of the distribution of senders and receivers. However, by routing packets to a random intermediate node before sending them to their destination, we can guarantee a reasonably fast running time. Note that we do not necessarily eliminate congestion, but we are able to bound the probability that high congestion will occur. Let us now introduce the hypercube network and the proposed routing algorithm; the proof of our bound using this routing algorithm is based off [1].

The n-dimensional hypercube network has 2^n nodes numbered 0 through $2^n - 1$ using their binary representations. Two nodes are connected by an edge when their binary representations differ in exactly one digit. In the permutation routing problem, each node is the sender and receiver of exactly one packet. The routing algorithm that immediately comes to mind is called the bit-fixing algorithm, in which we traverse an edge to alter each digit of the starting node if necessary, from left to right. Instead, we present the randomized bit-fixing algorithm, which follows the bit-fixing algorithm to send a packet to an intermediate node before using the same algorithm to send it to final destination:

Algorithm 6.1 (Randomized Bit-Fixing Algorithm)**.**
Let $\vec{x}$ and $\vec{y}$ denote the binary representations of nodes x and y.

Given a sender $\vec{x}$ and receiver $\vec{y}$, choose a random node $\vec{z}$ of the hypercube, generated by selecting each bit z_i as either 0 or 1 with equal probability.

[Phase 1] For i = 1 to n, if $x_i \neq z_i$, route the packet from $(z_1, \cdots, z_{i-1}, x_i, \cdots, x_n)$ to $(z_1, \cdots, z_i, x_{i+1}, \cdots, x_n)$.

[Phase 2] For i = 1 to n, if $z_i \neq y_i$, route the packet from $(y_1, \cdots, y_{i-1}, z_i, \cdots, z_n)$ to $(y_1, \cdots, y_i, z_{i+1}, \cdots, z_n)$.

Theorem 6.2. *On a hypercube with $N = 2^n$ nodes, the randomized bit-fixing algorithm takes $O(\log N)$ steps to route all packets from their sender to their receiver with probability $1 - O(N^{-1})$, where a step consists of each packet traversing at most one edge and each directed edge carrying at most one packet.*

Proof. In our discussion of Phase 1, we assume that no packets enter into Phase 2 until all packets have reached their intermediate node. We introduce some notation: denote by $T_1(P)$ the time (number of steps) it takes for packet P to complete Phase 1, and denote by $X_1(e)$ the number of packets which cross edge e in Phase 1.

Claim 1: For an arbitrary packet P that crosses edges $e_1, \cdots, e_m$ in Phase 1, the following inequality holds: $T_1(P) \leq \sum_{i=1}^{m} X_1(e_i)$.

Proof. For any given edge a packet must cross next, the greatest number of steps it takes to traverse that edge is $X_1(e_i)$, a bound realized when the packet waits in a queue $X_1(e_i) - 1$ long and traverses the edge on the next step. In any other case, the packet has higher priority on the queue, and takes less steps to traverse that edge. We sum across each edge in the path to attain the inequality. □

Consider a sequence of edges $E = (e_1, \cdots, e_m)$ following the bit-fixing algorithm. We define $T_1(E) = \sum_{i=1}^{m} X_1(e_i)$. Hence, $T_1(E)$ is an upper bound on the number of steps for a packet to traverse a sequence of edges E. We now seek to bound $T_1(E)$ with high probability. To do so, we consider packets that could possibly interfere with the route of another packet. We say that packet P is impeding edge sequence E at edge (v_i, v_{i+1}) if it is located at some vertex v_i on E and its jth bit has not yet been flipped, where v_i and v_{i+1} have different values in the jth position. Let I denote the total number of packets impeding E. Let E be the edge sequence connecting two arbitrary vertices $a = (a_1, \cdots, a_n)$ and $b = (b_1, \cdots, b_n)$ with the bit-fixing algorithm, and let v_i be an arbitrary vertex connected to some edge $e = (v_i, v_{i+1})$ of E. For some $j \in \{1, \cdots, n\}$, we have that:

$$v_i = (b_1, \cdots, b_{j-1}, a_j, \cdots, a_n)$$
$$v_{i+1} = (b_1, \cdots, b_j, a_{j+1}, \cdots, a_n)$$

To find the expected number of packets that impede E at any point, we first make two observations. In order to impede at $e \in E$, a packet's sender node must end with $a_j, \cdots, a_n$, because if it ends with any other sequence, a bit after j must be flipped to reach v_i, which would imply j has already been flipped. It follows that there are 2^{j-1} possible sending nodes for a packet that impedes at e. Furthermore, in order to impede at e, a packet's destination node must start with $b_1, \cdots, b_{j-1}$, because if it starts with any other sequence, by the time the jth bit is ready to be flipped, the packet will not be at v_i. Hence, for each of the 2^{j-1} possible starting nodes, the probability that the packet starting at that node will impede at e is $\frac{1}{2^{j-1}}$. Let I_e denote the number of packets impeding at $e \in E$. We have that:

$$\begin{aligned}
\mathbb{E}[I] &= \mathbb{E}[\sum_{k=1}^{m} I_e] && \text{packets can impede at } m \text{ edges} \\
&= \sum_{k=1}^{m} \mathbb{E}[I_e] && \text{by linearity} \\
&= \sum_{k=1}^{m} 2^{j-1} \frac{1}{2^{j-1}} \\
&= \sum_{k=1}^{m} 1 \\
&= m
\end{aligned}$$

To apply a Chernoff Bound to I, we must express it as a sum of independent, random 0-1 variables. Although we derived I by considering each edge, multiple packets can impede at a single edge, meaning we cannot apply such a bound. However, we can express I as the sum of indicator variables I_k, where, for $k \in \{0, \cdots, 2^n - 1\}$,

$$I_k = \begin{cases} 0 & \text{if the packet sent by node } k \text{ is impeding } E \\ 1 & \text{if the packet sent by node } k \text{ is not impeding } E \end{cases}$$

Hence, $I = \sum_{k=0}^{N-1} I_k$. Indeed, each I_k is independent, because the starting and ending nodes, which determine the value of I_k, are chosen independently for each packet. We now prove a lemma to simplify our bound.

Lemma 6.3. *For independent Poisson trials X_i, if $\theta \geq 6\mu$, then $Pr(X \geq \theta) \leq 2^{-\theta}$.*

Proof. We set $\theta = (1+\delta)\mu$, and it follows that $\delta \geq 5$. Weakening the bound and substituting θ into our previous Chernoff Bound, we have:

$$\Pr(X \geq (1+\delta)\mu) \leq (\frac{e^{\delta}}{(1+\delta)^{1+\delta}})^{\mu} \leq (\frac{e}{1+\delta})^{(1+\delta)\mu} \leq (\frac{e}{6})^{\theta} \leq 2^{\theta}$$

□

Since $6n \geq 6m$, using our lemma, we can apply a Chernoff Bound to the total number of impeding packets I as follows: $\Pr(I \geq 6n) \leq 2^{-6n}$. Given that $I \leq 6n$, which is true with high probability, we now bound $\Pr(T_1(E) \geq 30n)$.

Claim 2: If a packet leaves the edge sequence of another packet, then the first packet can no longer delay the second.

Proof. Suppose the first packet is located at $(b_1, \cdots, b_{j-1}, a_j, \cdots, a_n)$, where the edge sequence of the first packet follows the bit fixing algorithm from $(a_1, \cdots, a_n)$ to $(b_1, \cdots, b_n)$. If the first packet leaves the edge sequence, some bit aside from a_j will be flipped. To delay the first packet, the second packet would have to be located at the same node as the first. This is not possible, because after the bit in the kth position has been flipped, the bit in $\min\{j, k\}$ cannot be altered for either packet according to the bit-fixing algorithm. □

To condition on the event $I \leq 6n$, we must compute the conditional probability $\Pr(T_1(E) \geq 30n \mid I \leq 6n)$. Suppose $I \leq 6n$, and consider a single packet P impeding at e on the edge sequence E of another packet. The probability that P crosses e is at most $\frac{1}{2}$. Indeed, the packet at least has the choice of whether or not to flip the jth bit. The chance that the packet's own bit-fixing algorithm hasn't yet considered the jth bit further reduces these odds; this bound suffices for our analysis.

To apply a Chernoff Bound, consider the sum Z of 0-1 random variables Z_i. If an impeding packet has a chance of crossing an edge $e \in E$, let Z_i be given as such:

$$Z_i = \begin{cases} 0 & \text{if the packet crosses } e \\ 1 & \text{if the packet does not cross } e \end{cases}$$

Note that by our previous analysis, if $Z_i = 0$, the packet that left the edge cannot impede on E. If $Z = \sum_{i=1}^{36n} Z_i \geq 6n$, then $T_1(E) \leq 30n$, because once $6n$ packets do not cross an edge $e \in E$, no more packets can impede on E, and the total number of edges crossed will be at most $36n - 6n = 30n$. Hence, we have:

$$\begin{aligned} \Pr(T_1(E) \geq 30n \mid I \leq 6n) &\leq \Pr(Z \leq 6n) \\ &\leq e^{\frac{-18n(2/3)^2}{2}} && \text{by the Chernoff Bound} \\ &= e^{-4n} \\ &\leq 2^{-3n-1} && \text{since } n \geq 1 \end{aligned}$$

Given that $\Pr(I \geq 6n) \leq 2^{-6n}$ and $\Pr(T_1(E) \geq 30n \mid I \leq 6n\} \leq 2^{-3n-1}$, we can use conditional probabilities to split $\Pr(T_1(E) \geq 30n)$ into cases. We will also loosen the bound slightly, in the process eliminating the need to compute an additional conditional probability. This is ultimately inconsequential since $\Pr(I \geq 6n)$ is already quite small.

$$\begin{aligned} \Pr(T_1(E) \geq 30n) &= \Pr(T_1(E) \geq 30n \mid I \geq 6n) \Pr(I \geq 6n) \\ &\quad + \Pr(T_1(E) \geq 30n \mid I \leq 6n) \Pr(I \leq 6n) \\ &\leq \Pr(I \geq 6n) + \Pr(T_1(E) \geq 30n \mid I \leq 6n) \\ &\leq 2^{-6n} + 2^{-3n-1} \\ &\leq 2^{-3n} \end{aligned}$$

Since there are 2^n nodes on the hypercube, the number of possible edge sequences is 2^{2n}. Hence, the probability that our bound $T_1(E) \leq 30n$ does not hold for all edge sequences is given by

$$\begin{aligned} \Pr(\bigcup_E T_1(E) \geq 30n) &\leq 2^{2n} 2^{-3n} && \text{by the union bound} \\ &= O(N^{-1}) \end{aligned}$$

It follows that $\Pr(\bigcap_E T_1(E) \leq 30n) = 1 - O(N^{-1})$. Through a symmetry argument (Phase 2 has the edge sequences randomized by randomizing the sender node, rather than the receiver node), this bound also holds for Phase 2. Hence, we have in total $O(\log N)$ steps with probability $1 - O(N^{-1})$.

□

Acknowledgments

I would like to thank my mentor Anthony Santiago Chaves Aguilar for his excellent reading recommendations, thoughtful guidance, and invaluable advice throughout the research process. I would also like to thank Peter May for organizing the REU program and all of the instructors and mentors for introducing us to many fascinating topics.

References

[1] Mitzenmacher, Michael, and Eli Upfal. Probability and Computing: Randomized Algorithms and Probabilistic Analysis. Cambridge UP, 2005.

[2] Chawla, Shuchi. "Chernoff Bounds." 11 Oct. 2004. Carnegie Mellon University School of Computer Science, www.cs.cmu.edu/afs/cs/academic/class/15859-f04/www/scribes/lec9.pdf. Lecture.

[3] Tarjan, Robert. "Advanced Algorithm Design: More Chernoff Bounds." 2009. Princeton University Department of Computer Science, www.cs.princeton.edu/courses/archive/fall09/cos521/Handouts/probabilityandcomputing.pdf. Lecture.