15-859(M): Randomized Algorithms **Lecturer:** Shuchi Chawla
Topic: Chernoff Bounds **Date:** October 11, 2004
Scribe: Mugizi Rwebangira

9.1 Introduction

In this lecture we are going to derive **Chernoff bounds**. We will then look at applications of Chernoff bounds to coin flipping, hypergraph coloring and randomized rounding.

9.2 Markov's Inequality

Recall the following Markov's inequality:

Theorem 9.2.1 *For any r.v $X \geq 0$,*

$$\mathbf{Pr}[X > \lambda] < \frac{\mathbf{E}[X]}{\lambda}$$

Note that we can substitute any positive function $f : X \to \mathbb{R}^+$ for X:

$$\mathbf{Pr}[f(X) > f(\lambda)] < \frac{\mathbf{E}[f(X)]}{f(\lambda)}$$

When f is a non-decreasing function, we get that

$$\mathbf{Pr}[X > \lambda] = \mathbf{Pr}[f(X) > f(\lambda)] < \frac{\mathbf{E}[f(X)]}{f(\lambda)}$$

If we pick $f(X)$ judiciously we can obtain better bounds.

9.3 Chebyshev Bounds

As a first application of the above technique, we derive Chebyshev bounds. To do so we pick $f(X) = X^2$

$$\mathbf{Pr}[|X - \mathbf{E}[X]| \geq \lambda] = \mathbf{Pr}\left[(X - \mathbf{E}[X])^2 \geq \lambda^2\right] \leq \frac{\mathbf{E}\left[(X - \mathbf{E}[X])^2\right]}{\lambda^2} = \frac{\mathtt{var(X)}}{\lambda^2}$$

9.4 Chernoff Bounds

For the remainder of this lecture we will focus on Chernoff bounds. Chernoff bounds are typically tighter than Markov's inequality and Chebyshev bounds but they require stronger assumptions.

First we will state our assumptions and definitions. Let X be a sum of n *independent* random variables $\{X_i\}$, with $\mathbf{E}[X_i] = p_i$. We assume for simplicity that $X_i \in \{0,1\}$ for all $i \leq n$. Similar bounds hold for the case when X_is are arbitrary bounded random variables (see Homework 3).

Let μ denote the expected value of X. Then we have

$$\mu = \mathbf{E}\Big[\sum X_i\Big] = \sum \mathbf{E}[X_i] = \sum p_i$$

To establish our bound we pick $f(X) = e^{tX}$, and compute the probability that X deviates significantly from μ:

$$\mathbf{Pr}[X > (1+\delta)\mu] = \mathbf{Pr}\Big[e^{tX} > e^{(1+\delta)t\mu}\Big] \leq \frac{\mathbf{E}\big[e^{tX}\big]}{e^{(1+\delta)t\mu}} \tag{9.4.1}$$

We will know establish a bound on $\mathbf{E}\big[e^{tX}\big]$.

$$\begin{aligned}
\mathbf{E}\big[e^{tX}\big] &= \mathbf{E}\big[e^{t\sum X_i}\big] = \mathbf{E}\big[\textstyle\prod_i e^{tX_i}\big] \\
&= \textstyle\prod_i \mathbf{E}\big[e^{tX_i}\big] \qquad\qquad \text{(by independence)} \\
&= \textstyle\prod_i (p_i e^t + (1-p_i)\cdot 1) \\
&= \textstyle\prod_i (1 + p_i(e^t - 1))
\end{aligned}$$

We now use the following approximation — $\forall x \in \Re,\ 1 + x \leq e^x$. Hence:

$$\begin{aligned}
\mathbf{E}\big[e^{tX}\big] &\leq \textstyle\prod_i e^{p_i(e^t-1)} \\
&= e^{\sum_i p_i(e^t-1)} \\
&= e^{(e^t-1)\mu}
\end{aligned}$$

Substituting $\mathbf{E}\big[e^{tX}\big] \leq e^{(e^t-1)\mu}$ into Equation 9.4.1, we get that for all $t \geq 0$,

$$\mathbf{Pr}[X > (1+\delta)\mu] \leq \frac{e^{(e^t-1)\mu}}{e^{(1+\delta)t\mu}}$$

In order to make the bound as tight as possible, we find the value of t that minimizes the above expression — $t = \ln(1+\delta)$. Substituting this into the above expression, we obtain, for all $\delta \geq 0$:

$$\mathbf{Pr}[X > (1+\delta)\mu] \leq e^{((e^{\ln(1+\delta)}-1)-(1+\delta)\ln(1+\delta))\mu} = [e^{\delta-(1+\delta)\ln(1+\delta)}]^{\mu}$$

We will now try to obtain a simpler form of the above bound.
In particular, we use the Taylor series expansion of $\ln(1+\delta)$ given by $\ln(1+\delta) = \sum_{i\geq 1}(-1)^{i+1}\frac{\delta^i}{i}$.
Therefore,

$$(1+\delta)\ln(1+\delta) = \delta + \sum_{i\geq 2}(-1)^i\delta^i\left(\frac{1}{i-1}-\frac{1}{i}\right)$$

Assuming that $0 \leq \delta < 1$, and thereby ignoring the higher order terms, we get

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

Plugging this into our original expression we obtain:

$$\mathbf{Pr}[X > (1+\delta)\mu] \leq e^{\frac{-\delta^2\mu}{3}} \quad (0 < \delta < 1)$$

A very similar calculation shows that:

$$\mathbf{Pr}[X < (1-\delta)\mu] \leq e^{\frac{-\delta^2\mu}{2}} \quad (0 < \delta < 1)$$

9.4.1 A More General Chernoff Bound

We observe that $\ln(1+\delta) > \frac{2\delta}{2+\delta}$ $\forall\delta > 0$. This implies that

$$\delta - (1+\delta)\ln(1+\delta) \leq \frac{-\delta^2}{2+\delta}$$

Hence we obtain the following bound, which works for all positive values of δ.

$$\mathbf{Pr}[X > (1+\delta)\mu] \leq e^{\frac{-\delta^2\mu}{2+\delta}} \quad (\delta \geq 0)$$

Similarly it can be shown that:

$$\mathbf{Pr}[X < (1-\delta)\mu] \leq e^{\frac{-\delta^2\mu}{2+\delta}} \quad (\delta \geq 0)$$

9.5 Examples

9.5.1 Coin-flipping

We will now use Chernoff bounds to analyze n coin flips of an unbiased coin.

Let $X_i = 1$ if the i^{th} flip is heads and 0 otherwise. Let $X = \sum X_i$ be the number of heads in n flips. It is immediate that we expect to see $\mu = n/2$ heads. We now compute the deviation using Chernoff bounds.

$$\mathbf{Pr}[X \geq \mu + \lambda] = \mathbf{Pr}\left[X \geq \mu\left(1 + \frac{\lambda}{\mu}\right)\right] \leq e^{-\left(\frac{\lambda}{\mu}\right)^2 \frac{\mu}{3}} = e^{-\frac{\lambda^2}{3\mu}}$$

If $\lambda^2 = O(n)$, we get that $\mathbf{Pr}\left[X \geq \mu(1 + \frac{\lambda}{\mu})\right] \leq e^{-O(1)}$. To obtain a lower probability of error, we can use $\lambda = O(\sqrt{n \log n})$ getting $\mathbf{Pr}\left[X \geq \mu(1 + \frac{\lambda}{\mu})\right] \leq \frac{1}{n^C}$.

Hence, w.h.p., $X \in \mu \pm O(\sqrt{n \log n})$.

Let us now compare this with the bound given by Chebyshev's inequality. Note that $\sigma^2(X_i) = p(1-p) = \frac{1}{4}$. Therefore, $\sigma^2(X) = \frac{n}{4}$. So we get

$$Pr[X \geq \mu + \lambda] \leq \frac{\sigma^2}{\lambda^2} = \frac{1}{\log n} \quad \text{with } \lambda = O(\sqrt{n \log n})$$

Chernoff gives a much stronger bound on the probability of deviation than Chebyshev. This is because Chebyshev only uses pairwise independence between the r.v.s whereas Chernoff uses full independence. Full independence can some times imply exponentially better bounds.

9.5.2 Coloring a hypergraph

Consider the following problem. Let U be a universe of n elements, and $S_1, S_2, \ldots S_m$ be subsets of U. Let $E = \bigcup\{S_i\}$. Consider the hypergraph $G = (U, E)$, that is, where the set of vertices is the universe U and each set S_i is an edge in the graph.

We want to find a 2-coloring of G that *balances* each edge, or colors roughly half the elements in each set red, and half the elements blue. In particular, our goal is to minimize the following quantity:

Discrepancy $= \max_i \underbrace{|\#reds\ in\ S_i - \#blue\ in\ S_i|}_{Discrepancy\ of\ set\ S_i}$

Claim 9.5.1 *If each element is colored red w.p. $\frac{1}{2}$ independently then w.h.p. disc(S_i) $\leq \sqrt{12n \log m}$.*

Proof: We use the random variables Y_v to denote the color of the vertex v. Let

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

Let $Y = \sum_{v \in S_i} Y_v$. Then $Disc(S_i) = 2|k/2 - Y|$, where $k = |S_i|$. Note that $\mathbf{E}[Y] = \frac{k}{2}$. Applying Chernoff bounds:

$$\mathbf{Pr}\left[|Y - \frac{k}{2}| \geq \lambda\right] \leq e^{-\frac{2\lambda^2}{3k}}$$

Substituting $\lambda = \sqrt{\frac{3}{k}\log m}$,

$$\mathbf{Pr}\left[Disc(S_i) \geq \sqrt{12k\log m}\right] \leq \frac{1}{m^2}$$

Applying union bound:

$$\mathbf{Pr}\left[\exists i \mid Disc(S_i) \geq \sqrt{12n\log m}\right] \leq \sum_i \mathbf{Pr}\left[Disc(S_i) \geq \sqrt{12n\log m}\right] \leq m\frac{1}{m^2} = \frac{1}{m}$$

■

9.5.3 Randomized Rounding

Let $G = (V, E)$ be an undirected graph. We are given a set of vertex pairs $D = \{(s_i, t_i)\}_{i=1\cdots n}$. We are required to connect each pair of vertices in D with a path, such that no single edge in G is overloaded.

In particular, the solution consists of paths $\{P_i\}$, such that P_i is a path from s_i to t_i in G. Let the *congestion on edge* e be the number of paths containing e.

Problem: Pick P_i such that $max_e(cong(e))$ is minimized.

This problem is NP-hard, but we will give an approximation algorithm for it based on randomized rounding.

We begin by expressing this problem as an integer program. Let $\mathcal{P}_i$ be the set of all paths from s_i to t_i. Let f_P^i denote whether we pick path $P \in \mathcal{P}_i$ or not. That is, $f_P^i = 1$ if path P is picked, else $f_P^i = 0$. Let C denote the congestion in the graph.

We minimize C subject to the following constraints:

1. Exactly one path from s_i to t_i is picked.

$$\sum_{P \in \mathcal{P}_i} f_P^i = 1 \qquad \forall i$$

2. The congestion on every edge is less than C.

$$\sum_i \sum_{e \in P} f_P^i \leq C \quad \forall e$$

3. $f_P^i \in \{0, 1\} \quad \forall i, P$.

Solving this program is NP-Complete, so a natural way to relax the problem is to allow $f_P^i \in [0,1] \;\; \forall i, P$, transforming the above program into a linear program.

Note that the congestion of the optimal solution to the LPless than or equal to the best possible congestion. Therefore an approximation to the congestion of this solution implies an approximation to the problem.

We solve the LP to obtain a fraction solution and then round it to obtain an integral solution. The solution to the LP is a unit flow between every pair of vertices in D. Suppose that this solution has congestion C^*. In order to round this, for every i, we pick a path $P_i \in \mathcal{P}_i$ with probability $f_{P_i}^i$.

Now let us analyze the congestion on an edge e. Note that flow i picks e w.p. $\sum_{\{P \in P_i \;,\; e \in P\}} f_P^i$. Let $X_i^e = 1$ if flow i uses e. Then the congestion on e is given by $X^e = \sum_i X_i^e$.

$$\mathbf{E}[X_i^e] = \sum_{\{P \in P_i \;,\; e \in P\}} f_P^i$$

$$\mathbf{E}[X^e] = \mathbf{E}\left[\sum_i X_i^e\right] = \sum_i \mathbf{E}[X_i^e] = \sum_i \sum_{\{P \in P_i \;,\; e \in P\}} f_P^i \leq C$$

We want to pick a value of k such that $Pr[X_e > kC^*] < \frac{1}{n^3}$. Then, taking a union bound over all edges, we get a k-approximation. (Think about this!)

Claim 9.5.2 *If $C >> \log n$ then the above rounding gives a $(1+\epsilon)$-approximation.*

Proof: If $k = 1+\epsilon$ then using the simpler form of Chernoff, we get $\mathbf{Pr}[X_e > kC] \leq e^{-\frac{\epsilon^2 C}{3}} \leq \frac{1}{n^{\text{const}}}$. ■

When C is small, we do not get a $1+\epsilon$ approximation with a high probability. Instead, we know that for any $k = 1+\delta$, $\delta > 0$, $\mathbf{Pr}[X_e > kC] \leq e^{-\frac{\delta^2 C}{2+\delta}} \simeq e^{-\delta C}$ for $\delta >> 2$. $\delta = O(\log n)$ suffices and we get an $O(\log n)$ approximation.

In order to get a better approximation, we use the stronger form of Chernoff. For any $k = 1+\delta$:

$$\mathbf{Pr}[X_e > kC] \leq \left[\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right]^C \leq \frac{e^\delta}{(1+\delta)^{(1+\delta)}} \quad \text{as } C \geq 1$$

Now if we pick $\delta = O(\frac{\log n}{\log\log n})$, then $(1+\delta)\ln(1+\delta) = O(\ln n)$, and therefore, $(1+\delta)^{(1+\delta)} > n^c$ for some constant c. Therefore we get that with a high probability, the congestion is at most $1+\delta = O(\frac{\log n}{\log\log n})$.

This is essentially the best possible bound achievable using randomized rounding.