College of Computer & Information Science
Northeastern University
CS 7880: Algorithmic Power Tools

Fall 2009
18 September 2009
Scribe: Eric Miles

Lecture Outline:

- Chernoff & Chernoff-Hoeffding Bounds
- Symmetric Lovasz Local Lemma

In this lecture, we study Chernoff bounds, and compare them with the other tail bounds we have seen previously. The statement of the related Chernoff-Hoeffding bound is also included for completeness. We then begin studying the Lovasz Local Lemma, which is used to bound the probability that the union of several partially dependent events does not occur, and see the proof of the symmetric version.

1 Chernoff Bounds

As with the tail bounds we saw in the last lecture, Chernoff's formula is used to bound the probability that a random variable deviates from its mean. For this bound, however, the random variable under consideration will have a very specific structure. Let $X_1, \ldots, X_n$ be independent $\{0,1\}$-valued random variables, and define $p_i = \Pr[X_i = 1] = \mathbb{E}[X_i]$. Then, $X = \sum_i X_i$ will be the random variable we study. Note that $\mathbb{E}[X] = \mathbb{E}\left[\sum_i X_i\right] = \sum_i p_i$ by linearity of expectations; we will denote this value by μ.

Given this setup, the quantity we wish to upper bound is $\Pr[X \geq (1+\delta)\mu]$. Chernoff's formula will give a bound which is exponentially better than either Markov or Chebyshev by considering the (weighted) sum of all moments of X. To see how this can be done, recall the Taylor expansion of e^X:

$$e^X = \sum_{k=0}^{\infty} \frac{X^k}{k!} = 1 + X + \frac{X^2}{2!} + \frac{X^3}{3!} + \cdots$$

So, if we define the function $f_{\text{Cher}}(X) = e^{tX}$, for some t to be chosen later, and apply Markov's inequality to $\Pr[f_{\text{Cher}}(X) \geq f_{\text{Cher}}((1+\delta)\mu)]$, we can hope for an improvement.

Theorem 1 (Chernoff Bound)**.** *Let X be the random variable defined above, and let $\mu = \mathbb{E}[X]$. Then for any $\delta > 0$:*

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

Proof. For any fixed $t > 0$, e^{tX} is an increasing function in X. Therefore, $X \geq (1+\delta)\mu$ if and only if $e^{tX} \geq e^{t(1+\delta)\mu}$, and so $\Pr[X \geq (1+\delta)\mu] = \Pr[e^{tX} \geq e^{t(1+\delta)\mu}]$. We will bound the latter probability.

$$
\begin{array}{rcll}
\Pr[e^{tX} \geq e^{t(1+\delta)\mu}] & \leq & \frac{1}{e^{t(1+\delta)\mu}} \cdot \mathbb{E}[e^{t\sum_i X_i}] & \text{Markov's inequality} \\
& = & \frac{1}{e^{t(1+\delta)\mu}} \cdot \mathbb{E}[\prod_i e^{tX_i}] & \\
& = & \frac{1}{e^{t(1+\delta)\mu}} \cdot \prod_i \mathbb{E}[e^{tX_i}] & \text{independence of the } X_i \\
& = & \frac{1}{e^{t(1+\delta)\mu}} \cdot \prod_i [1 + p_i(e^t - 1)] & \\
& \leq & \frac{1}{e^{t(1+\delta)\mu}} \cdot \prod_i e^{p_i(e^t-1)} & 1 + x \leq e^x \ \forall x \in \mathbb{R} \\
& = & \frac{1}{e^{t(1+\delta)\mu}} \cdot e^{\mu(e^t-1)} & \sum_i p_i = \mu \\
& = & e^{-\mu(t(1+\delta)+1-e^t)} &
\end{array}
$$

We now want to choose t to maximize $t(1+\delta)+1-e^t$. Setting the derivative equal to 0 and solving for t gives $t = \ln(1+\delta)$. Plugging this value into the last line gives the stated bound. $\square$

Before comparing this with the other tail bounds, it will be helpful to simplify the exponent.

Lemma 1. *For $0 \leq \delta \leq 1$, $(1+\delta)\ln(1+\delta) - \delta \geq \delta^2/2$.*

Proof. The trick here is to use the Taylor expansion of $\ln(1+\delta)$:

$$\ln(1+\delta) \ = \ \sum_{k=1}^{\infty} (-1)^{k-1} \frac{\delta^k}{k} \ = \ \delta - \frac{\delta^2}{2} + \frac{\delta^3}{3} - \frac{\delta^4}{4} + \cdots$$

(Note that this only holds for $\delta \in (-1, 1]$, which is enough for our purposes.) Plugging this in and combining like terms gives

$$(1+\delta)\ln(1+\delta) = \delta + \sum_{k=2}^{\infty} (-1)^k \frac{\delta^k}{k(k-1)} = \delta + \frac{\delta^2}{2} - \frac{\delta^3}{6} + \cdots$$

Finally, dropping all but the first two terms gives the lemma. $\square$

So, the Chernoff bound can be restated as $\Pr[X \geq (1+\delta)\mu] \leq e^{-\mu\delta^2/2}$ for $\delta \in [0,1]$. Returning to our example from last lecture ($\delta = p_1 = \cdots = p_n = \frac{1}{2}$), we see that the Chernoff bound does indeed give a probability which vanishes exponentially with n: $\Pr[X \geq \frac{3}{4}n] \leq e^{-n/16}$. In contrast, Markov's and Chebyshev's inequalities gave bounds of $\frac{2}{3}$ and $\frac{4}{n}$, respectively.

A related inequality is known as the Chernoff-Hoeffding bound. This can be used to bound the absolute, rather than the relative, error in the case when each X_i is an independent *real-valued* random variable distributed (not necessarily uniformly) in the range $[0,1]$.

Theorem 2 (Chernoff-Hoeffding Bound)**.** *Let $X_1, \ldots, X_n \in [0,1]$ be independent random variables, and let $X = \sum_i X_i$. Then, for any δ such that $0 \leq \delta < \mathbb{E}[X]/n$:*

$$\Pr\left[\frac{X - \mathbb{E}[X]}{n} \geq \delta\right] \leq e^{-2n\delta^2}$$

Chernoff and Chernoff-Hoeffding bounds were first given in [2] and [4].

2 Lovasz Local Lemma

In designing algorithms, one often wants to bound the probability that the union of several events does not happen. Formally, let $E_1, \ldots, E_n$ be a collection of "bad" events defined over some sample space, and assume that $\forall i: \Pr[E_i] \leq p < 1$ for some fixed p. One question we can ask is, "Does there exist a point in the sample space which is outside all of these events?", i.e. $\Pr\left[\bigwedge_i \overline{E_i}\right] \overset{?}{>} 0$. Our ability to answer this question depends largely on what we know about the dependence between the events.

If the events are completely independent, then $\Pr\left[\bigwedge_i \overline{E_i}\right] = \prod_i \Pr[\overline{E_i}] \geq (1-p)^n$, which is greater than 0 for all $p < 1$. If we know nothing about the dependencies between events, the best we can do is the so-called union bound: $\Pr\left[\bigwedge_i \overline{E_i}\right] = 1 - \Pr[\bigvee_i E_i] \geq 1 - \sum_i \Pr[E_i] \geq 1 - pn$. This is greater than 0 only for $p < \frac{1}{n}$. The Lovasz Local Lemma (LLL) gives a non-trivial answer to our question even with only limited information on the dependencies between events.

Theorem 3 (LLL (Symmetric))**.** *Let $E_1, \ldots, E_n$ be a collection of events such that $\forall i : Pr[E_i] \leq p$. Suppose further that each event is dependent on at most d other events, and that $e \cdot p \cdot (d+1) \leq 1$. Then,* $\Pr\left[\bigwedge_i \overline{E_i}\right] > 0$.

This is referred to as the *symmetric* LLL because the probability of each event is bounded by the same value. The only requirement is that the number of dependencies for each event is bounded; notably, no assumptions about the underlying structure of the dependencies are made. If we instead made the much stronger assumption that the events can be grouped into $n/(d+1)$ blocks such that there are no dependencies between blocks, then we could apply the union bound on each block. (Here, assume w.l.o.g. that the blocks are $\{E_1, \ldots, E_{d+1}\}, \{E_{d+2}, \ldots, E_{2d+2}\}, \ldots, \{E_{n-d}, \ldots, E_n\}$.)

$$\Pr\left[\bigwedge_i \overline{E_i}\right] = \prod_{i=1}^{\frac{n}{d+1}} \Pr\left[\bigwedge_{j=i(d+1)-d}^{i(d+1)} \overline{E_j}\right] = \prod_{i=1}^{\frac{n}{d+1}} \left(1 - \Pr\left[\bigvee_{j=i(d+1)-d}^{i(d+1)} \overline{E_j}\right]\right) \geq (1 - p(d+1))^{\frac{n}{d+1}}$$

This quantity is greater than 0 for $p < 1/(d+1)$. By comparison, Theorem 3 needs $p \leq 1/(e(d+1))$, only a constant factor loss. We will prove Theorem 3 via the following lemma.

Lemma 2. *Assume the conditions of Theorem 3. Then, $\forall\, T \subseteq \{E_1, \ldots, E_n\}$ and $\forall\, E_i$:*

$$\Pr\left[E_i \,\middle|\, \bigwedge_{E_j \in T} \overline{E_j}\right] \leq ep$$

Proof of Theorem 3 assuming Lemma 2.

$$\begin{aligned}
\Pr[\bigwedge_i \overline{E_i}] &= \Pr[\overline{E_1}] \cdot \Pr[\overline{E_2} \mid \overline{E_1}] \cdot \Pr[\overline{E_3} \mid \overline{E_1} \bigwedge \overline{E_2}] \cdot \ldots \cdot \Pr[\overline{E_n} \mid \bigwedge_{i<n} \overline{E_i}] \\
&\geq (1 - ep)^n \\
&\geq \left(1 - \tfrac{1}{d+1}\right)^n \\
&> 0
\end{aligned}$$

$\square$

Proof of Lemma 2. The proof is by induction on the size of T. For the base case $|T| = 0$, the lemma holds because $\Pr[E_i] \leq p < ep$. Assume it holds on all sets up to size $k-1$. Fix T to be a set of size k and fix an event E_i. Let $S \subseteq T$ be defined such that E_i depends on all events in S and no events in $T \setminus S$. Assume that $|S| \geq 1$ (i.e. that E_i depends on some event in T); the lemma is trivially true otherwise. The chain rule for probabilities, and the fact that $S \cup (T \setminus S) = T$, gives

$$\Pr\left[E_i \wedge \bigwedge_{E_j \in S} \overline{E_j} \;\middle|\; \bigwedge_{E_\ell \in T \setminus S} \overline{E_\ell}\right] = \Pr\left[\bigwedge_{E_j \in S} \overline{E_j} \;\middle|\; \bigwedge_{E_\ell \in T \setminus S} \overline{E_\ell}\right] \cdot \Pr\left[E_i \;\middle|\; \bigwedge_{E_j \in T} \overline{E_j}\right]$$

This can be rewritten as

$$\Pr\left[E_i \;\middle|\; \bigwedge_{E_j \in T} \overline{E_j}\right] = \frac{\Pr\left[E_i \wedge \bigwedge_{E_j \in S} \overline{E_j} \;\middle|\; \bigwedge_{E_\ell \in T \setminus S} \overline{E_\ell}\right]}{\Pr\left[\bigwedge_{E_j \in S} \overline{E_j} \;\middle|\; \bigwedge_{E_\ell \in T \setminus S} \overline{E_\ell}\right]}$$

Note that the left hand side is the probability we are trying to bound. We will bound each half of the fraction on the right hand side separately. For the numerator, we have

$$\Pr\left[E_i \wedge \bigwedge_{E_j \in S} \overline{E_j} \;\middle|\; \bigwedge_{E_\ell \in T \setminus S} \overline{E_\ell}\right] \leq \Pr\left[E_i \;\middle|\; \bigwedge_{E_\ell \in T \setminus S} \overline{E_\ell}\right] \leq p$$

The second inequality follows from the fact that E_i is independent of the events in $T \setminus S$. To bound the denominator, denote $S = \{E_{j_1}, \ldots, E_{j_{|S|}}\}$. Then,

$$\begin{aligned}
\Pr\left[\bigwedge_{E_j \in S} \overline{E_j} \;\middle|\; \bigwedge_{E_\ell \in T \setminus S} \overline{E_\ell}\right] &= \Pr\left[\overline{E_{j_1}} \mid \bigwedge_{E_\ell \in T \setminus S} \overline{E_\ell}\right] \cdot \Pr\left[\overline{E_{j_2}} \mid \overline{E_{j_1}} \wedge \bigwedge_{E_\ell \in T \setminus S} \overline{E_\ell}\right] \cdot \ldots \\
&\geq (1 - ep)^d \\
&\geq \left(1 - \tfrac{1}{d+1}\right)^d \\
&\geq 1/e
\end{aligned}$$

The second line comes from the fact that $|S| \leq d$, and by the inductive hypothesis (there are at most $k-1$ events on the r.h.s. of each conditional probability by our assumption that $|S| \geq 1$ and $|T| = k$). The last line comes from the fact that the sequence $\left\{\left(1 - \frac{1}{n+1}\right)^n\right\}_{n \geq 1}$ approaches $1/e$ from above. Combining the bounds on the numerator and the denominator completes the proof. $\square$

The Lovasz Local Lemma was first proved in [3]. The textbook by Alon & Spencer [1] is a good reference for this and related material.

References

[1] N. Alon, J.H. Spencer. *The probabilistic method* (2nd edition). Wiley-Interscience. Chapter 5 (2000)

[2] H. Chernoff. *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations.* Annals of Mathematical Statistics. 23:4, pp. 493-507 (1952)

[3] P. Erdős, L. Lovász. *Problems and results on 3-chromatic hypergraphs and some related questions.* in A. Hajnal, R. Rado, and V. T. Ss, eds.. Infinite and Finite Sets (to Paul Erdős on his 60th birthday). Volume II, pp. 609-627 (1975)

[4] W. Hoeffding. *Probability inequalities for sums of bounded random variables.* Journal of the American Statistical Association. 58:301, pp. 13-30 (1963)