CS 6550: Randomized Algorithms **Spring 2019**

Lecture 3: Chernoff Bounds

January 15, 2019

Lecturer: Eric Vigoda *Scribes: Federico Bruvacher and Michael Wigal*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

3.1 Chernoff Bounds

3.1.1 Markov and Chebyshev Inequality

For a Random Variable X, we denote $\mathbf{E}[X] = \mu$ and $\mathbf{Var}(X) = \sigma^2$.

Lemma 3.1 (Markov Inequality) *Let X be a non-negative random variable, and $a > 0$, then*

$$\boldsymbol{Pr}(X > a) \leq \frac{\mu}{a}$$

Lemma 3.2 (Chebyshev Inequality) *Let X be a non-negative random variable for which $\boldsymbol{Var}(X)$ exists, then for all $k > 0$*

$$\boldsymbol{Pr}(|X - \mu| > k\sigma) \leq \frac{1}{k^2}$$

A more general form being,

$$r \geq 0 \boldsymbol{Pr}(|x - \mu| > r) \leq \frac{\boldsymbol{Var}(X)}{r^2}$$

Proof: Note that $Y = (X - \mu)^2$ is a non-negative random variable, we can then apply the Markov Inequality to Y. ■

Note that Chebyshev does not always give a good bound. We give an example. Let

$$X_i = \begin{cases} 1 \text{ with probability } \frac{1}{2} \\ 0 \text{ with probability } \frac{1}{2} \end{cases}$$

Let $X = \sum_{i=1}^{n} X_i$, then from previous lecture we know $\mathbf{E}[X] = \frac{n}{2}$ and $\mathbf{Var}(X) = \frac{n}{4}$, with $\sigma = \frac{\sqrt{n}}{2}$.
Note for $n = 1000$, $X = \text{Bin}(1000, \frac{1}{2})$ by the Chebyshev Inequality we have,

$$\mathbf{Pr}(X \geq 750) = \frac{1}{2}\mathbf{Pr}(|X - 500| \geq 250) \leq \frac{1}{2}\frac{250}{250^2} = 0.002$$

We can calcultate this probability directly,

$$\mathbf{Pr}(X \geq 70) = \sum_{i=750}^{1000} \binom{1000}{i} 2^{-1000} \approx 60 \times 10^{-58}$$

Note that the Chebyshev Inequality is significantly off.

3.1.2 Chernoff "argument" for Bin$(n, \frac{1}{2})$

Note that if $X = \text{Bin}(n, \frac{1}{2})$, using Chernoff Bounds we can obtain bounds,

$$\mathbf{Pr}(X \geq \mu + t\frac{\sqrt{n}}{2}) \leq e^{-t^2/2}(*)$$
$$\mathbf{Pr}(X \leq \mu - t\frac{\sqrt{n}}{2}) \leq e^{-t^2/2}$$

We first argue (*) to show the inuition behind the general Chernoff Bound.
Proof: We first want to transform $X = X_1 + \cdots + X_n$ such that it has mean 0. Let

$$Y_i = -1 + 2X_i = \begin{cases} 1 \text{ with probability } \frac{1}{2} \\ -1 \text{ with probability } \frac{1}{2} \end{cases}$$

Note then that $\mathbf{E}[Y] = 0$. Since $\mathbf{Var}(Y_i) = 1$, this implies $\mathbf{Var}(Y) = n$.
Note that $Y_1 + \cdots Y_k$ can be interpreted as a unbiased random walk on the integers starting at 0.
The original bound we wanted was $X \geq \frac{n}{2} + t\frac{\sqrt{n}}{2}$, which is equivalent to $Y \geq -n + 2(\frac{n}{2} + 2(\frac{n}{2} + t\frac{\sqrt{n}}{2}) = t\sqrt{n}$.
When we are far away from 0, adding 1 can be approximated by instead multiplying by $(1 + \lambda)$ for some tiny λ. If λ is small enough then $(1 + \lambda)(1 + \lambda) = 1 + 2\lambda + \lambda^2 \approx 1 + 2\lambda$.
Let $Z_i = (1 + \lambda)^{Y_i}$ where we will choose λ later. Note

$$Z_i = \begin{cases} 1 + \lambda \text{ with probability } \frac{1}{2} \\ \frac{1}{1+\lambda} \text{ with probability } \frac{1}{2} \end{cases}$$

Note then that $Z = Z_1 \cdot Z_2 \cdots Z_n = (1 + \lambda)^{Y_1} ... (1 + \lambda)^{Y_n} = (1 + \lambda)^Y$.
What we have done is transformed the random walk model, where if the random walk was at u, then in the new model the random walk would be at $(1 + \lambda)^u$. Since Z is now a non-negative random variable, we can now utilize the Markov Inequality. Since Y_i are pairwise independent, so are the Z_i. It follows,

$$\mathbf{Pr}(X \geq \frac{n}{2} + t\frac{\sqrt{n}}{2}) = \mathbf{Pr}(Y \geq t\sqrt{n})$$
$$= \mathbf{Pr}(Z \geq (1 + \lambda)^{t\sqrt{n}}$$

For example, by a smart choice of lambda and Taylor Series approximation, $1 + \lambda \approx e^{\frac{1}{\sqrt{n}}}$.

$$\mathbf{Pr}(Z \geq (1 + \lambda)^{100\sqrt{n}}) = \mathbf{Pr}(Z \geq e^{100})$$

Note that e^{100} is a big number, thus Markov Inequality would give a good bound. To make things rigorous,

$$\begin{aligned} \mathbf{E}[Z_i] &= \frac{1}{2}(1 + \lambda) + \frac{1}{2}(\frac{1}{1 + \lambda} \\ &= \frac{1}{2}(\frac{\lambda^2 + 2\lambda + 2}{1 + \lambda}) \\ &= 1 + \frac{\lambda^2}{2 + 2\lambda} \\ &\leq 1 + \frac{\lambda^2}{2} \end{aligned}$$

It follows that $\mathbf{E}[Z] \leq (1+\frac{\lambda^2}{2})^n$. Note then that for $\lambda = \frac{1}{\sqrt{n}}$,

$$\begin{aligned}
\mathbf{Pr}(Z \geq (1+\lambda)^{t\sqrt{n}}) &\leq \frac{\mathbf{E}[Z]}{(1+\lambda)^{t\sqrt{n}}} \\
&= \frac{(1+\frac{\lambda^2}{2})^n}{(1+\lambda)^{t\sqrt{n}}} \\
&= \frac{(1+\frac{t^2}{2n})^n}{(1+\frac{t}{\sqrt{n}})^{t\sqrt{n}}} \\
&\leq \frac{e^{\frac{t^2}{2}}}{e^{t^2}} = e^{\frac{-t^2}{2}}.
\end{aligned}$$

where we are cheating on the denominator of $\leq$ inequality. ∎

3.1.3 Chernoff Inequality

Now that we have given an intuition on the proof we can move to the rigorous proof.

Let $X_i, \cdots, X_n$ be Independent Bernoulli R.V.s where $0 \leq X_i \leq 1$ Let $X = \sum_{i=1}^n X_i$ and $\mu = \mathbf{E}[X]$ For all $0 < \varepsilon \leq 1$

$$\mathbf{Pr}(X \geq \mu(1+\varepsilon)) \leq e^{-\mu\cdot(\varepsilon^2/3)}$$
$$\mathbf{Pr}(X \leq \mu(1-\varepsilon)) \leq e^{-\mu\cdot(\varepsilon^2/2)}$$

We want to know :

$$\mathbf{Pr}(X \geq \mu(1+\varepsilon))$$

Note as X is non-negative, we can choose an arbitrary t, then we exponentiate both sides and raise both sides to the power t for some arbitrary t,

$$\mathbf{Pr}(e^X \geq e^{\mu(1+\varepsilon)})$$
$$\mathbf{Pr}(e^{tX} \geq e^{t\mu(1+\varepsilon)})$$

We know the applying Markov's inequality:

$$\mathbf{Pr}(e^{tX} \geq e^{t\mu(1+\varepsilon)}) \leq \frac{\mathbf{E}[e^{tX}]}{e^{t\mu(1+\varepsilon)}} \tag{A}$$

Because the $X_i = \text{Bernoulli}(p_i)$ and $1+x \leq e^x$ then

$$\mathbf{E}[e^{tX_i}] = p_i e^t + (1-p_i) = 1 + p_i(e^t - 1) \leq e^{p_i(e^t-1)}$$

Then the moment generating function:

$$\mathbf{E}[e^{tX}] \leq \prod_{i=1}^n e^{p_i(e^t-1)} = e^{\mu(e^t-1)} \tag{B}$$

Let's substitute B in A:

$$\mathbf{Pr}(e^{tX} \geq e^{t\mu(1+\varepsilon)}) \leq (\frac{e^{\varepsilon-1}}{e^{t(1+\varepsilon)}})^{\mu} = (e^{\varepsilon-(1+\varepsilon)\log(1+\varepsilon)})^{\mu} \tag{C}$$

In the last equality we plugged in $t = ln(1+\varepsilon)$to minimize.

Taylor expansion for : $\log(1+\varepsilon) = \varepsilon - \varepsilon^2/2 + \varepsilon^3/3 + \cdots$ Then: $(1+\varepsilon)\log(1+\varepsilon) = \varepsilon - \varepsilon^2/2 + \varepsilon^2 + \varepsilon^3/3 - \varepsilon^3/2 + \cdots \geq \varepsilon + \varepsilon^2/2 - \varepsilon^3/6 = \varepsilon + \varepsilon^2/3$ Using this in C:

$$(\frac{e^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}})^{\mu} \leq e^{\varepsilon^{2/3}\mu}$$