Advanced Algorithms: Lecture 13 Notes

Dana Randall Spring 2024

Sarthak Mohanty

Tail Inequalities Continued

The Markov inequality actually tell us much more than we originally let on. It is easily extneded by realizeing that for any function $\phi(x)$ which is non-negative and strictly monotonically increasing,

$$\mathrm{P}\left(X \geq t\right) = \mathrm{P}\left(\phi(X) \geq \phi(t)\right).$$

We now any have number of ways to modify the bound, as

$$\mathrm{P}\left(X \geq t\right) \geq \frac{\mathrm{E}\left[\phi(x\right]}{\phi(t)},$$

for any such ϕ. Moreover, it holds for any random variable X (not just non-negative!) as we only need $\phi(X) \geq 0$ to apply Markov.

A **Chernoff bound** is simply an application of Markov with $\phi(t) = e^{\lambda}t$ for some $\lambda > 0$;

$$\mathrm{P}\left(X \geq t\right) \leq e^{-\lambda t}\,\mathrm{E}\left[e^{\lambda X}\right].$$

We can once again use this to bound the probability that X deviates significantly from μ:

$$\mathrm{P}\left(X \geq (1+\delta)\mu\right) \leq e^{-(1+\delta)\lambda\mu}\,\mathrm{E}\left[e^{\lambda X}\right].$$

This is particularly useful when X is a sum of independent [0, 1] random variables $X_1, X_2, \ldots, X_N$. Suppose $\mathrm{E}\left[X_i\right] = p_i$ and $\mu = \mathrm{E}\left[X\right]$. Then the Chernoff bound on their sum is

$$\begin{aligned}
\mathrm{P}\left(X_1 + \cdots + X_N \geq (1+\delta)\mu\right) &\leq e^{-(1+\delta)\lambda\mu}\,\mathrm{E}\left[e^{\lambda(X_1+\cdots+X_N)}\right] \\
&= e^{-(1+\delta)\lambda\mu}\,\mathrm{E}\left[e^{\lambda X_1}e^{\lambda X_2}\ldots e^{\lambda X_N}\right] \\
&= e^{-(1+\delta)\lambda\mu}\prod_{i=1}^{N}\mathrm{E}\left[e^{\lambda X_i}\right] \\
&= e^{-(1+\delta)\lambda\mu}\prod_{i=1}^{N}(p_i(e^{\lambda} + (1-p_i)\cdot 1) \\
&= e^{-(1+\delta)\lambda\mu}\prod_{i=1}^{N}(1 + p_i(e^{\lambda} - 1))
\end{aligned}$$

The first order Taylor approximation for e^x tells us that $1 + x \leq e^x$ for all $x \in \mathbb{R}$. Thus

$$\begin{aligned} \mathrm{E}\left[e^{\lambda X}\right] &\leq \prod_i e^{p_i(e^\lambda - 1)} \\ &= e^{\sum_{i=1}^N p_i(e^\lambda - 1)} \\ &= e^{(e^\lambda - 1)\mu}, \end{aligned}$$

and so for all $\lambda \geq 0$,

$$\mathrm{P}\left(X > (1+\delta)\mu\right) \leq e^{-(1+\delta)\lambda\mu} e^{(e^\lambda - 1)\mu} = e^{((e^\lambda - 1) - (1+\delta)\lambda)\mu}.$$

The value of λ that minimizes the right hand side above is $\lambda = \ln(1+\delta)$. Plugging this in and simplifying gives us

$$\begin{aligned} \mathrm{P}\left(X > (1+\delta)\mu\right) &\leq e^{((e^{\ln(1+\delta)} - 1) - (1+\delta)\ln(1+\delta))\mu} \\ &= e^{(\delta - (1+\delta)\ln(1+\delta))\mu} \end{aligned}$$

Now we will 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

$$\mathrm{P}\left(X > (1+\delta)\mu\right) \leq e^{\frac{-\delta^2\mu}{3}} \quad (0 < \delta < 1).$$

A very similar calculation shows that:

$$\mathrm{P}\left(X < (1-\delta)\mu\right) \leq e^{\frac{-\delta^2\mu}{2}} \quad (0 < \delta < 1).$$

Exercise. Let $X \sim \text{Bin}(n, p)$. Using Markov's inequality, find an upper bound on $\mathrm{P}(X \geq \alpha n)$, where $p < \alpha < 1$. Evaluate the bound for $p = \frac{1}{2}$ and $\alpha = \frac{3}{4}$.

Solution. Note that X is a nonnegative random variable and $\mathrm{E}[X] = np$. Applying Markov's inequality, we obtain

$$\mathrm{P}(X \geq \alpha n) \leq \frac{\mathrm{E}[X]}{\alpha n} = \frac{pn}{\alpha n} = \frac{p}{\alpha}.$$

For $p = \frac{1}{2}$ and $\alpha = \frac{3}{4}$, we obtain $\mathrm{P}\left(X \geq \frac{3n}{4}\right) \leq \frac{2}{3}$.

On the other hand, we can use Chebyshev's inequality to obtain

$$\begin{aligned}
\mathrm{P}(X \geq \alpha n) &= \mathrm{P}(X - np \geq \alpha n - np) \\
&\leq \mathrm{P}(|X - np| \geq n\alpha - np) \\
&\leq \frac{\operatorname{var}(X)}{(n\alpha - np)^2} \\
&= \frac{p(1-p)}{n(\alpha - p)^2}
\end{aligned}$$

For $p = \frac{1}{2}$ and $\alpha = \frac{3}{4}$, we obtain

$$\mathrm{P}\left(X \geq \frac{3n}{4}\right) \leq \frac{4}{n}.$$

Note that Markov is better when n is small, but as n increases, Chebyshev gives us a better estimate, inversely linear in x. However, we can do much better than both approaches with Chernoff. Left as an exercise for the reader.

More General Chernoff Inequality

Chernoff bounds may also be applied to general sums of independent, bounded random variables, regardless of their distribution; this is known as **Hoeffding's inequality.** We will not cover it in this course, but it is important to prove some bounds in learning theory, among other things.