CS 70 Discrete Mathematics and Probability Theory
Summer 2016 Dinh, Psomas, Ye Discussion 6A

Recall unlike Markov and Chebyshev's inequalities, Chernoff Bounds are a family of bounds. A few (of the many bounds are shown below):

1 For any $\delta > 0$:
$$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$$

2 For any $1 > \delta > 0$:
$$\Pr[X \leq (1-\delta)\mu] \leq \left(\frac{e^{\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$$

3 For any $1 > \delta > 0$:
$$\Pr[X \geq (1+\delta)\mu] \leq e^{\frac{-\mu\delta^2}{3}}$$

4 For any $1 > \delta > 0$:
$$\Pr[X \leq (1-\delta)\mu] \leq e^{\frac{-\mu\delta^2}{2}}$$

5 For $R > 6\mu$:
$$\Pr[X \geq R] \leq 2^{-R}$$

1. Comparing Bounds

Compare and contrast all of the bounds that we have gone over in CS 70. In other words discuss the conditions when each bound can be used, as well as the strengths and weaknesses of each bound.

1 Markov
Constraints: Must have a strictly positive random variable
Good: Used to prove the other 2 bounds
Bad: Usually gives a very bad bound

2 Chebyshev
Constraints: No constraints, works on all r.v.'s
Good: Works on everything, and in some sense justifies are use of variance as a measure of spread of a distribution (see previous discussion sheet)
Bad: Not as tight as Chernoff Bounds

3 Chernoff
Constraints: Random variable must be a sum of independent indicator random variables (for the purposes of this class)
Good: Exponential bounds decay extremely quickly
Bad: Not a very flexible bound

2. Showing off Chernoff

Let $X_1,\ldots,X_n$ be independent Bernoulli random variables that each take value 1 with probability p and 0 with probability $1-p$. You have learned how to use Chebyshev's inequality to say things about the probability that the sum $S = X_1 + X_2 + \cdots + X_n$ deviates from its mean (pn). In this question you will derive another bound called Chernoff's inequality that is much stronger in most cases.

(a) As an example to help you understand the setting better, assume that X_i is the outcome of a coin flip (that is $X_i = 1$ if the coin flip results in heads and otherwise $X_i = 0$). Then $p = 1/2$ and S is the number of heads you observe. Assume that $n = 100$ is the number of coin flips. The expected number of heads you see is $pn = 50$. The exact probability that $S \geq 80$ is $5.5795 \cdot 10^{-10}$. Now using Chebyshev's inequality find an upper bound for this probability. Is your upper bound much larger than the value you computed?

From Chebyshev's inequality:

$$Pr(S \geq 80) = Pr(S-50 \geq 30) \leq Pr(|S-50| \geq 30) \leq \frac{100\cdot 0.5 \cdot 0.5}{30^2} = 0.0278$$

This is a very conservative upper bound.

(b) Back to the general setting, prove that if $f : \{0,1\} \to \mathbb{R}$ is any function, then $f(X_1),\ldots,f(X_n)$ are independent. Hint: write down the definition of independence. If f takes the same value at 0 and 1 then everything should be obvious. It remains to prove it in the case where $f(0) \neq f(1)$.

For $f(X_1)\ldots f(X_n)$ to be independent, we need

$$Pr[f(X_1) = a_1, f(X_2) = a_2, \ldots, f(X_n) = a_n] = Pr[f(X_1) = a_1]\cdot Pr[f(X_2) = a_2]\ldots Pr[f(X_n) = a_n]$$

If f is a constant function, i.e. $f(0) = f(1) = c$, then this is trivially true as the probability is either 1 or 0 depending on the values of a_i.

So, let's look at when it's not a constant function. Say we have $f(0) = c_0$ and $f(1) = c_1$. Further, since there are the only two values our function can take, we will restrict the codomain of our function to be $\{c_0, c_1\}$. f is thus a bijective function from $\{0,1\}$ to $\{c_0,c_1\}$.

This allows us to define an inverse of f as follows: $f^{-1}(c_0) = 0$ and $f^{-1}(c_1) = 1$. With this, we can say $Pr[f(X_i) = a] = Pr[X_i = f^{-1}(a)], a \in \{c_0, c_1\}$

Now, we have:

$$\begin{aligned} Pr[f(X_1) = a_1, f(X_2) = a_2, \ldots, f(X_n) = a_n] &= Pr[X_1 = f^{-1}(a_1), X_2 = f^{-1}(a_2), \ldots, X_n = f^{-1}(a_n)] \\ &= Pr[X_1 = f^{-1}(a_1)]\cdot Pr[X_2 = f^{-1}(a_2)]\ldots Pr[X_n = f^{-1}(a_n)] \\ &= Pr[f(X_1) = a_1]\cdot Pr[f(X_2) = a_2]\ldots Pr[f(X_n) = a_n] \end{aligned}$$

Thus, we have that the $f(X_i)$s are independent.

(c) Now if we fix a number t and let $f(x) = e^{tx}$, then $f(X_i) = e^{tX_i}$. Compute the expected value of $f(X_i) = e^{tX_i}$ and write it in terms of p and t.

$E[f(X_i)] == pe^{t\cdot 1} + (1-p)e^{t\cdot 0}$
$= pe^t + 1 - p$
$>= p(e^t - 1) + 1$

(d) The following is a famous inequality about real numbers: $1+x \leq e^x$. Another variant of the inequality (which can be derived by replacing x by $x-1$) is the following: $x \leq e^{x-1}$. Apply the latter inequality with x being the expected value you computed in the previous step in order to get an upper bound on $E[f(X_i)]$. (You don't need to prove either of these inequalities.)

$E[f(X_i)] = p(e^t - 1) + 1$

$> \leq e^{(p(e^t-1)+1)-1}$

$> = e^{p(e^t-1)}$

(e) Remembering that $f(X_1), \ldots, f(X_n)$ are all independent what is $E[f(X_1)f(X_2)\ldots f(X_n)]$ in terms of $E[f(X_1)], \ldots, E[f(X_n)]$? Use the upper bound you got from the previous step to get an upper bound on $E[f(X_1)f(X_2)\ldots f(X_n)]$. You should be able to express your answer in terms of p, n, and t. Now let $\mu = pn$ be the expected value of S. Re-express your upper bound in terms of μ and t (i.e. remove the occurrences of p and n and rewrite them in terms of μ).

$E[f(X_1)f(X_2)\ldots f(X_n)] = E[f(X_1)] \cdot \ldots \cdot E[f(X_n)]$
$> \leq e^{p(e^t-1)} \cdot \ldots \cdot e^{p(e^t-1)}$

$> = (e^{p(e^t-1)})^n$

$> = e^{np(e^t-1)}$

$> = e^{\mu(e^t-1)}$

(f) Observe that $f(X_1)\ldots f(X_n) = e^{t(X_1+\cdots+X_n)} = e^{tS}$. Let us call e^{tS} the random variable Y. Does it always take positive values? Let's say we are interested in bounding the probability that $S \geq (1+\alpha)\mu$ where α is a non-negative number. Prove that $S \geq (1+\alpha)\mu$ is the same event as $Y \geq e^{t\mu(1+\alpha)}$. Use Markov's inequality on the latter event to derive an upper bound for $\Pr[S \geq (1+\alpha)\mu]$ in terms of μ, t, and α.

The exponential function always takes positive values as long as the exponent is a real number.

We know that e^{tS} is a monotonically increasing function of S, if t is positive. This means that $a \geq b$ if and only if $e^{ta} \geq e^{tb}$.

Thus, the event $S \geq (1+\alpha)\mu$ is the same as $e^{tS} \geq e^{t(1+\alpha)\mu}$ which is just $Y \geq e^{t(1+\alpha)\mu}$.

Using Markov's inequality, we have:

$Pr[S \geq (1+\alpha)\mu] = Pr[Y \geq e^{t(1+\alpha)\mu}]$

$> \leq \dfrac{E[Y]}{e^{t(1+\alpha)\mu}}$

$> = \dfrac{E[e^{tS}]}{e^{t(1+\alpha)\mu}}$

$> = \dfrac{E[e^{t(X_1+\ldots+X_n)}]}{e^{t(1+\alpha)\mu}}$

$> \leq \dfrac{e^{\mu(e^t-1)}}{e^{t(1+\alpha)\mu}}$

$> = e^{\mu(e^t-1)-\mu t(1+\alpha)}$

(g) For different values of t you get different upper bounds for the probability that $S \geq (1+\alpha)\mu$. But of course all of them are giving you an upper bound on the same quantity. Therefore it is wiser to pick a t that minimizes the upper bound. This way you get the tightest upper bound you can using this method. Assuming that α is fixed, find the value t that minimizes your upper bound. For this value of t what is the actual upper bound? Your answer should only depend on α and μ. Hint: in order to minimize a positive expression you can instead minimize its ln. Then you can use familiar methods from calculus in order to minimize the expression.

We want to minimize $e^{\mu(e^t-1)-\mu t(1+\alpha)}$ with respect to t.

This is the same as minimizing $\log\left(e^{\mu(e^t-1)-\mu t(1+\alpha)}\right) = \mu(e^t-1)-\mu t(1+\alpha)$ with respect to t.

So, let's take $g(t) = \mu(e^t-1)-\mu t(1+\alpha)$. We want $g'(t) = 0$.

$$g'(t) = \mu e^t - \mu(1+\alpha) = 0$$

$$\therefore e^t = (1+\alpha) \Rightarrow t = \log(1+\alpha)$$

We see that $g''(t) = \mu e^t > 0$, so this is indeed a minimum.

The upper bound we get with this t is:

$$e^{\mu((1+\alpha)-1)-\mu\log(1+\alpha)(1+\alpha)} = \frac{e^{\mu\alpha}}{\left(e^{\log(1+\alpha)}\right)^{\mu(1+\alpha)}}$$

$$= \frac{e^{\mu\alpha}}{(1+\alpha)^{\mu(1+\alpha)}}$$

(h) Here we want to compare Chernoff's bound and the bound you can get from Chebyshev's inequality. Assume for simplicity that $p = 1/2$, so $\mu = n/2$.

First compute Chernoff's bound for the probability of seeing at least 80 heads in 100 coin flips (the quantity you bounded in the first part). Compare your answer to that part and see which one is closer to the actual value.

Now back to the setting with general n and α, write down the Chernoff bound as c^n where c is an expression that only contains α and not n. This shows that for a fixed value of α, Chernoff's bound decays exponentially in n. Now write down Chebyshev's inequality to bound $\Pr[|S-\mu| \geq \alpha\mu]$. Show that this is also a bound on $\Pr[S \geq (1+\alpha)\mu]$. Write down this bound as γn^β where γ and β are some numbers that do not depend on n. This shows that Chebyshev's inequality decays like n^β. In general an exponential decay (which you get from Chernoff's) is much faster than a polynomial decay (the one you get from Chebyshev's).

We want to find $P(S \geq 80) = P(S \geq (1+0.6)50)$. So, we have $\alpha = 0.6$ and $\mu = 50$.

Plugging in these values, we get a bound of $5.0031 \cdot 10^{-4}$. This is a much better bound than what we got from Chebyshev's inequality.

Now, let's try to get our Chernoff bound in the form c^n.

$$\frac{e^{\mu\alpha}}{(1+\alpha)^{\mu(1+\alpha)}} = \frac{e^{np\alpha}}{(1+\alpha)^{np(1+\alpha)}} = \left(\frac{e^{p\alpha}}{(1+\alpha)^{p(1+\alpha)}}\right)^n$$

So, we have $c = \frac{e^{p\alpha}}{(1+\alpha)^{p(1+\alpha)}}$.

Note: Using the derivation of the Chernoff bound without the approximation in part 4 (as given in note 19b), we get a much more accurate bound of $4.258 \cdot 10^{-9}$.

Now, let's look at Chebyshev's bound. First we have:

$$
\begin{aligned}
P(|S-\mu| \geq \alpha\mu) &= P(S-\mu \geq \alpha\mu) + P(\mu - S \geq \alpha\mu) \\
&\geq P(S-\mu \geq \alpha\mu) \\
&= P(S \geq (1+\alpha)\mu)
\end{aligned}
$$

So, from Chebyshev's inequality, we have:

$$P(S \geq (1+\alpha)\mu) \leq P(|S-\mu| \geq \alpha\mu) \leq \frac{np(1-p)}{(\alpha np)^2} = \frac{1-p}{p\alpha^2} \cdot n^{-1}$$

This gives us $\gamma = \frac{(1-p)}{p\alpha^2}$ and $\beta = -1$.