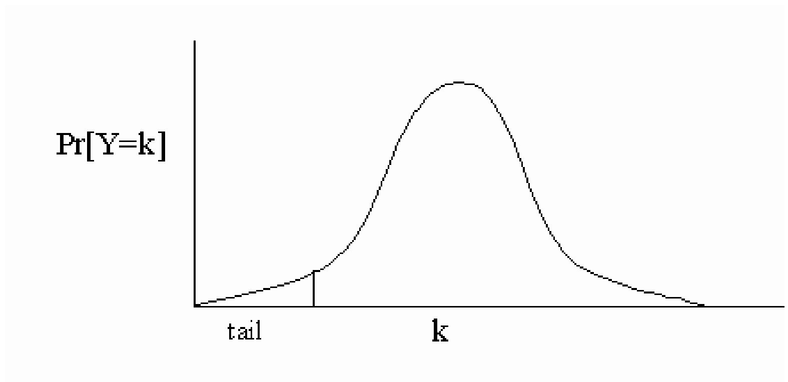
CS174 Lecture 10 John Canny

Chernoff Bounds

Chernoff bounds are another kind of tail bound. Like Markoff and Chebyshev, they bound the total amount of probability of some random variable Y that is in the "tail", i.e. far from the mean.



Recall that Markov bounds apply to any non-negative random variable Y and have the form:

$$\Pr[Y \geq t] \leq \frac{\overline{Y}}{t}$$

where $\overline{Y} = \mathrm{E}[Y]$. Markov bounds don't depend on any knowledge of the distribution of Y. Chebyshev bounds use knowledge of the standard deviation to give a tighter bound. The Chebyshev bound for a random variable X with standard deviation σ is:

$$\Pr[|X - \overline{X}| \geq t\sigma] \leq \frac{1}{t^2}$$

But we already saw that some random variables (e.g. the number of balls in a bin) fall off exponentially with distance from the mean. So Markov and Chebyshev are very poor bounds for those kinds of random variables.

The Chernoff bound applies to a class of random variables and does give exponential fall-off of probability with distance from the mean. The critical condition that's needed for a Chernoff bound is that the random variable be a sum of *independent* indicator random variables. Since that's true for balls in bins, Chernoff bounds apply.

Bernoulli Trials and the Binomial Distribution

The first kind of random variable that Chernoff bounds work for is a random variable that is a sum of indicator variables with the same distribution (Bernoulli trials). That is, if X_i is a random variable with $\Pr[X_i = 1] = p, \Pr[X_i = 0] = (1-p)$, and the X_i are all independent. Tossing a coin is a Bernoulli trial. So is the event that a randomly tossed ball falls into one of n bins $(p = 1/n)$. If

$$X = \sum_{i=1}^{n} X_i$$

is a sum of Bernoulli trials, then X has a Binomial distribution. We derived this already for coins and balls into bins. It is:

$$\Pr[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$$

the Chernoff bounds approximate a generalization of the binomial distribution.

Poisson Trials

There is a slightly more general distribution that we can derive Chernoff bounds for. If instead of a fixed probability we allow every X_i to have a different probability, $\Pr[X_i = 1] = p_i$, and $\Pr[X_i = 0] = (1 - p_i)$, then these event are called Poisson trials. A Poisson trial by itself is really just a Bernoulli trial. But when you have a lot of them together with different probabilities, they are called Poisson trials. But it is very important that the X_i must still be *independent*.

Chernoff Bounds (lower tail)

Let $X_1, X_2, \ldots, X_n$ be independent Poisson trials with $\Pr[X_i = 1] = p_i$. Then if X is the sum of the X_i and if μ is $\mathrm{E}[X]$, for any $\delta \in (0, 1]$:

$$\Pr[X < (1-\delta)\mu] < \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \right)^{\mu}$$

This bound is quite good, but can be clumsy to compute. We can simplify it to a weaker bound which is:

$$\Pr[X < (1-\delta)\mu] < \exp(-\mu\delta^2/2)$$

the simplified bound makes it clear that the probability decreases exponentially with distance δ from the mean.

Example In n tosses of a fair coin, what's the probability of $m < n/2$ heads? Let X be the number of heads, then $\mu = n/2$ and $\delta = (1 - 2m/n)$ is the relative distance of m from μ. The bound gives us a probability of fewer than m heads which is

$$\Pr[X < m] < \exp(-(n/4)(1 - 2m/n)^2)$$

So if we toss the coin 100 times and ask for less than 10 heads, the probability is less than $\exp(-16) = 1.12 \times 10^{-7}$.

Proof of the Chernoff bound First write the inequality as an inequality in exponents, multiplied by $t > 0$:

$$\Pr[X < (1-\delta)\mu] = \Pr[\exp(-tX) > \exp(-t(1-\delta)\mu)]$$

Its not clear yet why we introduced t, but at least you can verify that the equation above is correct for positive t. We will need to fix t later to give us the tightest possible bound. Now we can apply the Markov inequality to the RHS above:

$$\Pr[X < (1-\delta)\mu] < \frac{\mathrm{E}[\exp(-tX)]}{\exp(-t(1-\delta)\mu)]}$$

Notice that $\exp(-tX)$ is a product of independent random variables $\exp(-tX_i)$. This is the heart of the Chernoff bound. The expected value of X is the *product* of the expected values $\mathrm{E}[\exp(-tX_i)]$. So we have that

$$\Pr[X < (1-\delta)\mu] < \frac{\prod_{i=1}^{n} \mathrm{E}[\exp(-tX_i)]}{\exp(-t(1-\delta)\mu)]}$$

Now $\mathrm{E}[\exp(-tX_i)]$ is given by

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

We would like to express this as the exponential of something, so that we can simplify the product expression it appears in. As we did in an earlier lecture, we use the fact that $1 - x < \exp(-x)$ with $x = p_i(1-e^{-t})$, we get

$$\mathrm{E}[\exp(-tX_i)] < \exp(p_i(e^{-t}-1))$$

and from there we can simplify:

$$\prod_{i=1}^{n} \mathrm{E}[\exp(-tX_i)] < \prod_{i=1}^{n} \exp(p_i(e^{-t}-1)) = \exp\left(\sum_{i=1}^{n} p_i(e^{-t}-1)\right) = \exp(\mu(e^{-t}-1))$$

because $\mu = \sum p_i$, and e^{-t} is a constant in the sum. Substituting back into the overall bound gives:

$$\Pr[X < (1-\delta)\mu] < \frac{\exp(\mu(e^{-t}-1))}{\exp(-t(1-\delta)\mu)]} = \exp(\mu(e^{-t} + t - t\delta - 1))$$

Now its time to choose t to make the bound as tight as possible. That means minimizing the RHS wrt t. Taking the derivative of $(e^{-t} + t - t\delta - 1)$ and setting it to zero gives:

$$-e^{-t} + 1 - \delta = 0$$

and solving gives us $t = \ln(1/(1-\delta))$. Making that substitution gives:

$$\Pr[X < (1-\delta)\mu] < \exp(\mu((1-\delta) + (1-\delta)\ln(1/(1-\delta)) - 1))$$

and after cancelling the 1's and applying the exponential, we get:

$$\Pr[X < (1-\delta)\mu] < \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$$

which is the bound we are looking for.

To get the simpler form of the bound, we need to get rid of the clumsy term $(1-\delta)^{(1-\delta)}$. First take the log to give $(1-\delta)\ln(1-\delta)$. Now the Taylor expansion of natural log is

$$\ln(1-\delta) = -\delta - \delta^2/2 - \delta^3/3 - \delta^4/4 \cdots$$

multiplying by $(1-\delta)$ gives:

$$(1-\delta)\ln(1-\delta) = -\delta + \delta^2/2 + \text{all positive terms} > -\delta + \delta^2/2$$

and we can apply exponentiation to give:

$$(1-\delta)^{(1-\delta)} > \exp(-\delta + \delta^2/2)$$

We can substitute this inequality into the earlier bound to get:

$$\Pr[X < (1-\delta)\mu] < \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu} < \left(\frac{\exp(-\delta)}{\exp(-\delta + \delta^2/2)}\right)^{\mu} \quad = \quad \exp(-\mu\delta^2/2)$$

Chernoff Bounds (upper tail)

Let $X_1, X_2, \ldots, X_n$ be independent Poisson trials with $\Pr[X_i = 1] = p_i$. Then if X is the sum of the X_i and if μ is $\mathrm{E}[X]$, for any $\delta > 0$:

$$\Pr[X > (1+\delta)\mu] < \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$$

Proof The proof is almost identical to the proof for the lower tail bound. Start by introducing a t parameter:

$$\Pr[X > (1+\delta)\mu] = \Pr[\exp(tX) > \exp(t(1+\delta)\mu)]$$

compute the Markov bound, convert the product of expected values to a sum, and then solve for t to make the bound as tight as possible. QED

The upper-tail bound can be simplified. Suppose $\delta > 2e - 1$, then

$$\Pr[X > (1+\delta)\mu] < \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu} < \left(\frac{e^{\delta}}{(2e)^{(1+\delta)}}\right)^{\mu} < \left(\frac{e^{\delta}}{(2e)^{\delta}}\right)^{\mu} = 2^{-\delta\mu}$$

which shows once again an exponential drop-off in probability with δ.

By a more complicated argument, which we wont give here, you can show that for $\delta < 2e - 1$, the Chernoff bound simplifies to:

$$\Pr[X > (1+\delta)\mu] < \exp(-\mu\delta^2/4)$$