Chernoff Bounds: what are them and why we are using them

Chernoff Bounds are a set of powerful techniques used to provide tight bounds on the tail probabilities of sums of independent random variables.

- They are particularly useful for assessing the likelihood that the sum deviates significantly from its expected value
- Unlike simpler bounds like Markov's, Chernoff bounds take advantage of the distribution's specific characteristics to offer sharper estimates, especially useful for understanding the decay of tail probabilities exponentially fast

Consider the following footprint exercise:

Let $X_1, X_2, \ldots, X_n$ be independent indicator random variables such that $\Pr(X_i = 1) = 1/(4e)$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = E[X]$. By applying the following Chernoff bound, which holds for every $\delta > 0$,

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

prove that

$$\Pr(X > n/2) < \frac{1}{(\sqrt{2})^n}.$$

What is important in this set of exercises is the set of following steps (always like this):

- Characterize the event X_i (dependent on the type of problem you are dealing with)
  - And find the probability of success
- Characterize the expected value μ
- Use it to find δ
- Apply the bound given by the exercise

Consider we are usually bounding a precise value: apart from some *strange* cases, normally you have to get exactly the number given by the exercise.

So, here, we have to first consider X_i. We know they are independent. Now, we simply need to find the expected value. We already have here the probabilty of success given by $\Pr(X_i = 1) = \frac{1}{4e}$. This applies for all$n$ events since they are independent so:

$$\mu = E[X] = E[\sum_{i=1}^{n} E[X_i] = n * \frac{1}{4e} = \frac{n}{4e}$$

Now, we find δ and this is done according to value we have to bound, given by the exercise or explicitly told here like $\frac{n}{2}$:

$$(1+\delta)\mu = \frac{n}{2}$$

$$(1+\delta)\frac{n}{4e} = \frac{n}{2}$$

$$(1+\delta)\frac{n}{2e} = n$$

$$(1+\delta)n = 2e(n)$$

$$(1+\delta) = 2e$$

$$\delta = 2e - 1$$

Now that we found δ, let's plug it in back in the original bound:

$$\Pr(1+\delta)\,\mu < \left(\frac{(e^{\delta})}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$$

$$= \Pr\left(X > (1+2e-1)\frac{n}{4e}\right) \leq$$

$$\leq \left(\frac{e^{2e-1}}{(1+2e-1)^{(1+2e-1)}}\right)^{\frac{n}{4e}}$$

$$\leq \left(\frac{e^{2e-1}}{(2e)^{(2e)}}\right)^{\frac{n}{4e}}$$

$$\leq \left(\frac{\cancel{e^{2e}} * e^{-1}}{2^{2e} * (\cancel{e^{2e}})}\right)^{\frac{n}{4e}}$$

$$\leq \left(\frac{1}{e}\right)^{\frac{n}{4e}} * \left(\frac{1}{2^{2e}}\right)^{\frac{n}{4e}}$$

$$\leq \left(\frac{1}{e^{-4e}}\right)^{n} * \left(\frac{1}{2^{\frac{2e}{4e}}}\right)^{n}$$

$$\leq \left(\frac{1}{e^{-4e}}\right)^{n} * \left(\frac{1}{2^{\frac{1}{2}}}\right)^{n}$$

$$\leq \left(\frac{1}{e^{-4e}}\right)^{n} * \left(\frac{1}{\sqrt{2}}\right)^{n}$$

To infinity, it dominates the second factor, so we'd have $\left(\frac{1}{\sqrt{2}}\right)^{n}$

Exercise 2 (9 points) Suppose you toss $n \gg 1$ times a coin: applying the following Chernoff bound show that the probability that you obtain more than $n/2 + \sqrt{6n \ln n}/2$ heads is at most $1/n$.[1]

Theorem 1. *Let $X_1, X_2, \ldots, X_n$ be independent indicator random variables such that $E[X_i] = p_i, 0 < p_i < 1$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = E[X]$. Then, for $0 < \delta \leq 1$,*

$$\Pr(X > (1+\delta)\mu) \leq e^{-\mu\delta^2/3}.$$

$\Pr(X_i = 1) = \frac{1}{2}$. In this case, consider $X_i = 1$ if coin is tail, 0 otherwise. These are all independent.

We have $\mu = E[X] = E[\sum_{i=1}^{n} X_i] = n * \frac{1}{2} = \frac{n}{2}$. Now find δ with $(1+\delta)\mu = \frac{n}{2} + \frac{\sqrt{6nln(n)}}{2}$

So, we do the following:

$$(1+\delta)\mu = \frac{n}{2} + \frac{\sqrt{6nln(n)}}{2}$$

$$(1+\delta)\frac{n}{2} = \frac{n}{2} + \frac{\sqrt{6nln(n)}}{2}$$

$$(1+\delta)\frac{n}{2} * 2 = \frac{n + \sqrt{6nln(n)}}{2} * 2$$

$$(1+\delta)n = n + \sqrt{6nln(n)}$$

$$(\cancel{1}+\delta) = \cancel{1} + \frac{\sqrt{6nln(n)}}{n}$$

$$\delta = \frac{\sqrt{6nln(n)}}{n}$$

Now, we apply the bound as follows:

$$\Pr(X > (1+\delta)\mu) \leq e^{\frac{-\mu\delta^2}{3}}$$

$$= \Pr\left(X > \frac{n}{2} + \frac{\sqrt{6nln(n)}}{2}\right) \leq$$

$$\leq e^{\frac{-\frac{n}{2} * \left(\frac{\sqrt{6nln(n)}}{n}\right)^2}{3}}$$

$$\leq e^{\frac{-\frac{\cancel{n}}{\cancel{2}} * \left(\frac{\cancel{6^2 n}ln(n)}{\cancel{n^2}}\right)}{\cancel{3}}}$$

$$\leq e^{-\ln(n)}$$

$$\leq \frac{1}{n}$$

as the exercise wanted.

Problem 2 (10 points) Suppose you throw n balls into $\frac{n}{6\ln n}$ bins[1] independently and uniformly at random. Applying the following Chernoff bound show that, with high probability, the bin with maximum load (load = number of balls in the bin) contains at most $12\ln n$ balls. (Hint: focus first on one arbitrary bin and bound the probability of that bin's load exceeding $12\ln n\ \ldots$)

Theorem 1. *Let $X_1, X_2, \ldots, X_n$ be independent indicator random variables such that $E[X_i] = p_i, 0 < p_i < 1$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = E[X]$. Then, for $0 < \delta \leq 1$,*

$$\Pr(X > (1+\delta)\mu) \leq e^{-\mu\delta^2/3}.$$

Consider $X_i = 1, 2, \ldots n$ since we can assign the i^{th} ball to any possible bin. So, this is uniform at random. The load of the specific bin is given by $X = \sum_{i=1}^{n} X_i$.

First, we have to find μ:

$$\mu = E[X] = \sum_{i=1}^{n} E[X_i] = n * \frac{6\ln(n)}{n} = 6\ln(n)$$

Since each ball is assigned to a bin chosen uniformly at random, we have

$$\Pr(X_i = 1) = \frac{1}{m} * n = \frac{1}{\frac{n}{6\ln(n)}} * n = \frac{6\ln(n)}{n}$$

To apply the Chernoff bound, we set $12\ln(n)$ equal to $(1+\delta)\mu$ so:

$$(1+\delta)\mu = 12\ln(n)$$

$$(1+\delta)6\ln(n) = 12\ln(n)$$

$$(1+\delta) = 2$$

$$\delta = 1$$

Now, we apply the bound:

$$\Pr(X > (1+\delta)\mu) \leq e^{\frac{-\mu\delta^2}{3}}$$

$$\Pr(X > 1+1)\,6\ln(n)) \leq e^{-\frac{6\ln(n)}{3}}$$

$$\leq e^{-2\ln(n)}$$

Recall the property of exponentials and logarithms there, so:

$$\leq e^{\ln(n^{-2})}$$

Recall from the exercise hint that $\ln(n) = \log_e(n)$

So, we have:

$$e^{\ln(n^{-2})} = \frac{1}{n^2}$$

as the exercise wanted. We showed with high probability the bin with maximum load containing at most $12\ln(n)$ balls. We applied this for *one* bin, so we have to use now the union bound; simply use the previous result multiplying by all bins, so $m = \frac{n}{6n(\ln(n))}$: $\frac{n}{6\ln(n)} * \frac{1}{n^2} = \frac{1}{6n\ln(n)}$

Written by Gabriel R.

To characterize the *no bin will exceed*, use the complement event → $1 - \frac{1}{6nln(n)} = 1 - o\left(\frac{1}{n}\right)$

Exercise 2 (11 points) Let S be a set of n distinct positive integers, and let WORK(S) be a procedure which, given input S, returns an integer by performing n^2 operations. Now consider the following randomized algorithm:

```
RAND_REC(S)
   if |S| <= 1 then return 1
   x = WORK(S)
   p = RANDOM(S)
   S1 = {s in S such that s < p}
   S2 = {s in S such that s > p}
   if (|S1| >= |S2|) then
      y = RAND_REC(S1)
   else
      y = RAND_REC(S2)
   return x + y
```

Applying the following Chernoff bound show that the complexity of RAND_REC(S) is $O(n^2 \log n)$ with high probability. (Hint: recall the analysis of randomized QuickSort.)

Theorem 1. *Let $X_1, X_2, \ldots, X_n$ be independent indicator random variables such that $E[X_i] = p_i, 0 < p_i < 1$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = E[X]$. Then, for $0 < \delta \leq 1$,*

$$\Pr(X < (1-\delta)\mu) < e^{-\mu\delta^2/2}.$$

Recalling the analysis of Randomized QuickSort: the event E can be characterized as "in the first $l = \log_{\frac{4}{3}}(n)$ nodes of P there have been $< \log_{\frac{4}{3}}(n^2)$ lucky choices". We are studying this specific event:

- $X_i, 1 \leq i \leq l = \log_{\frac{4}{3}}(n^2)$
- $X_i = 1$ if at the i^{th} vertex of P there is a lucky choice of the pivot
- $\Pr(X_i = 1) = \frac{1}{2}\ \forall i$
- X_i are independent

We want the probability of $P\left(\sum_{i=1}^{l} X_i < \log_{\frac{4}{3}}(n^2)\right)$ to bound $X = \sum_{i=1}^{l} X_i$. Given $X = \sum_{i=1}^{l} X_i$, its expected value is as follows:

$$\mu = E[X] = E[\sum_{i=1}^{l} X_i] = \sum_{i=1}^{l} E[X_i] = \sum_{i=1}^{l} \frac{1}{2} = \frac{1}{2} * l = \frac{1}{2}\log_{\frac{4}{3}}(n^2)$$

Now, let's apply the following Chernoff bound (the first):

$$\Pr(X < (1-\delta)\mu) < e^{\frac{-\mu\delta^2}{2}}, 0 < \delta \leq 1$$

$$\downarrow$$

$$(1-\delta)\mu = \log_{\frac{4}{3}}(n^2)$$

$$(1-\delta)\frac{1}{2}\log_{\frac{4}{3}}(n^2) = \log_{\frac{4}{3}}(n^2)$$

$$(1-\delta)\log_{\frac{4}{3}}(n^2) = 2\log_{\frac{4}{3}}(n^2)$$

$$(1-\delta) = 2$$

$$-\delta = 1$$

$$\delta = -1$$

We then apply the Chernoff lemma as follows:

$$\Pr\left(X < \log_{\frac{4}{3}}(n)\right) < e^{-\log\left(\frac{4}{3}\right)(n^2)}$$

$$= e^{-\frac{ln(n^2)}{ln\left(\frac{4}{3}\right)}}$$

$$= \left(e^{-ln(n^2)}\right)^{\frac{1}{\ln\left(\frac{4}{3}\right)}}$$

$$= \left(\frac{1}{n^2}\right)^{\frac{1}{\ln\left(\frac{4}{3}\right)}\simeq 3{,}47}$$

$$= \left(\frac{1}{n^2}\right)^3 = (n^{-2})^{-3} = n^{-6} = \frac{1}{n^6}$$

Exercise 2 (10 points) Suppose you have a randomized algorithm for a minimization problem A that returns the correct output with probability at least $1/n$, where n is the input size. Show how to obtain an algorithm for A that returns the correct output with high probability. (Hint: for the analysis use this inequality: $(1+x/y)^y \leq e^x$ for $y \geq 1$, $y \geq x$.)

Here, we use Karger since the hint uses that inequality and the only point in the program we saw that is exactly that analysis – so, that's why we use that in place of a "normal" Chernoff Bound.

Characterize the event of getting a probability *at least* $\frac{1}{n}$ using Karger's analysis; run different times the analysis and fix a constant d, $d > 0$:

$$\Pr\left(X > \frac{1}{n}\right) > \frac{1}{n^d}$$

Since the probability is at least $\frac{1}{n}$, so we characterize using Karger. Here, we will characterize the probability of failure (the complement with respect to the previous, so $1-\frac{1}{n}$, running k times to reduce the error probability.

We want to find a value for k such that $Pr\left(1-\frac{1}{n}\right)^k \leq \frac{1}{n^d}$. In this case, it's standard the use of this inequality:

$$\left(1+\frac{x}{y}\right)^y \leq e^x, y \geq 1, y \geq x$$

This inequality is derived from the exponential function and the binomial expansion. It represents an upper bound on the expression $\left(1+\frac{x}{y}\right)^y$, showing that it grows slower than e^x.

Now, we use Karger's analysis, in place of using $k = n^2$ we use $k = n$ and everything comes naturally.

By choosing $k = dn\ln(n)$ it follows that (it holds $d = 1$ coming from Karger):

$$\left(1 - \frac{1}{n}\right)^{k=n} \leq e^{-1} = \frac{1}{e} \rightarrow \text{is } \underline{\text{not}} \text{ in the form } \frac{1}{n^d}$$

Recall the following from the Karger analysis (choice of k and rest of reasoning is that):

$$e^{-\ln(n^d)} = \frac{1}{n^d}$$

Continuing using Karger and the inequality:

$$\left(\left(1 - \frac{1}{n}\right)^n\right)^{\ln(n^d)} = \left(1 - \frac{1}{n}\right)^{n\ln(n)}$$

Let's wrap up:

$$\left(1 - \frac{1}{n}\right)^{k=n\ln(n)} = \left(\left(1 - \frac{1}{n}\right)^n\right)^{\ln(n)}$$

$$\leq (e^{-1})^{\ln(n)} = e^{-\ln(n)} = \frac{1}{n}$$

<u>For reference from theory, Karger's analysis is this.</u>

$\Pr(\textit{the k runs of FULL\_CONTRACTION do not return the min cut}) \leq \left(1 - \frac{2}{n^2}\right)^k \leq \frac{1}{n^d}$ for some constant $d > 0$

The previous one is the probability of an unsuccessful event, so we want it very low, something like $\frac{1}{n^d}$.

We want to find a value for k such that $\left(1 - \frac{2}{n^2}\right)^k \leq \frac{1}{n^d}$. In this case, it's standard the use of this inequality:

$$\left(1 + \frac{x}{y}\right)^y \leq e^x, y \geq 1, y \geq x$$

This inequality is derived from the exponential function and the binomial expansion. It represents an upper bound on the expression $\left(1 + \frac{x}{y}\right)^y$, showing that it grows slower than e^x. The probability of not contracting the minimum cut in each iteration needs to be bounded and manipulated to ensure the overall algorithm's success probability is high.

By choosing $k = \frac{dn^2\ln(n)}{2}$ it follows that:

$$\left(\left(1 - \frac{2}{n^2}\right)^{n^2}\right)^{\ln(n^d)} \leq e^{-\ln(n^d)} = \frac{1}{n^d}$$

Given I am curious, I asked myself: why exactly that value for k?

Consider the probability of success if $\frac{2}{n^2}$ while the failure is, by complement, $1 - \frac{2}{n^2}$ which, amplified by k runs, becomes $\left(1 - \frac{2}{n^2}\right)^k$. The constant d is the desired level of confidence to keep the wanted threshold (in this case $\frac{1}{n^d}$) as low as possible. Then, using some good old GPT-4:

To find k, we need $(1-\frac{2}{n^2})^k \leq \frac{1}{n^d}$. Using the approximation for exponential functions for small x, $(1-x) \approx e^{-x}$, we get:

$$(1-\frac{2}{n^2})^k \approx e^{-\frac{2k}{n^2}}$$

Setting this equal to $\frac{1}{n^d}$, we have:

$$e^{-\frac{2k}{n^2}} \approx \frac{1}{n^d}$$

$$-\frac{2k}{n^2} \approx -d \ln n$$

$$k \approx \frac{dn^2 \ln n}{2}$$

Moving on:

$$\left(1-\frac{2}{n^2}\right)^{k=n^2} \leq e^{-2} = \frac{1}{e^2} \rightarrow \text{is } \underline{\text{not}} \text{ in the form } \frac{1}{n^d}$$

Recall the following:

$$e^{-\ln(n^d)} = \frac{1}{n^d}$$

Let's apply that:

$$\left(\left(1-\frac{2}{n^2}\right)^{n^2}\right)^{\ln(n^d)} = \left(1-\frac{2}{n^2}\right)^{n^2 \ln(n^d)}$$

Let's wrap up (here, in the prof. notes, d magically disappears, but I assume it to be 1 so this works):

$$\left(1-\frac{2}{n^2}\right)^{\boxed{k=\frac{dn^2\ln(n^d)}{2}}} = \left(\left(1-\frac{2}{n^2}\right)^{n^2}\right)^{\frac{\ln(n^d)}{2}}$$

$$\leq (e^{-2})^{\frac{\ln(n^d)}{2}} = e^{-\ln(n^d)} = n^d = \frac{1}{n^d}$$

Then, by choosing that value for k the Karger's algorithm succeeds with high probability:

$$\Rightarrow \Pr(KARGER\ succeeds) > 1 - \frac{1}{n^d}$$

So, in the end, the Karger algorithm accumulates the size of the min-cut with probability at least $\frac{1}{n^d}$.

Exercise 2 (9 points) For $n \gg 1$, let $X_1, X_2, \ldots, X_n$ be independent indicator random variables such that $Pr(X_i = 1) = (6 \ln n)/n$ (recall that $\ln n = \log_e n$). Let $X = \sum_{i=1}^{n} X_i$ and $\mu = E[X]$. By applying the following Chernoff bound

$$Pr(X > (1+\delta)\mu) < e^{-\mu\delta^2/2} \quad \text{for } 0 < \delta \leq 2e - 1$$

prove that

$$Pr(X > 10 \ln n) < \frac{1}{n^c}$$

for some positive constant c to be determined.

Here we have already $\Pr(X_i = 1) = \frac{6\ln(n)}{n}$ and we need to find $\mu = E[X] = E[\sum_{i=1}^{n} X_i] = n * \frac{6\ln(n)}{n} = 6\ln(n)$. Now, we find δ:

$$(1+\delta)\mu = 10\ln(n)$$

$$(1+\delta)6\ln(n) = 10\ln(n)$$

$$(1+\delta) = \frac{5}{3}$$

$$\delta = \frac{2}{3}$$

Now, we use the bound:

$$\Pr(X > 10\ln(n)) = \Pr(X > (1+\delta)\mu) = \Pr\left(X > \left(1 + \frac{2}{3}\right)\mu\right)$$

$$< e^{-\frac{6\ln(n)}{2} * \frac{4}{9}}$$

$$< e^{-\frac{4}{3}\ln(n)}$$

$$< e^{\ln(n)^{-\frac{4}{3}}}$$

$$< n^{-\frac{4}{3}}$$

$$= \frac{1}{n^{\frac{4}{3}}}$$

Which is then verified for the constant $c = \frac{4}{3}, c > 0$ as showed.

There are $m = 6nln(n)$ jobs to be assigned randomly to n processors (Note: remember $\ln(x) = \log_e(x)$. Consider a processor p and show that, with high probability on n, processor p does not receive more than $12\ln(n)$ jobs (Hint: define an appropriate indicator variable and apply the following Chernoff Bound).

Theorem 1: Let $X_1, X_2, \dots X_n$ be independent indicator random variables with $E[X_i] = p_i, 0 < p_i < 1$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = E[X]$. Then, for all $0 < \delta \leq 1$,

$$\Pr(X > (1+\delta)\mu) \leq e^{\frac{-\delta^2 \mu}{2}}$$

Solution: Let $X_1 = 1, 2, \dots m$ (with $m = 6nln(n)$ jobs and $X_i = 1$ when the i^{th} job gets assigned to processor p. Here, $\Pr(X_i = 1) = \frac{1}{n}$ with X_i independent between each other. The number of jobs received by the processor p is then $X = \sum_{i=1}^{m} X_i$ given it holds for each processor.

Now, find $\mu = E[X] = \sum_{i=1}^{m} E[X_i] = m * \frac{1}{n} = \frac{6n(\ln(n))}{n} = 6\ln(n)$

Then, we find δ:

To apply the Chernoff bound, we set $12\ln(n)$ equal to $(1+\delta)$ so:

$$(1+\delta)\mu = 12\ln(n)$$

$$(1+\delta)6\ln(n) = 12\ln(n)$$

$$(1+\delta) = 2$$

$$\delta = 1$$

Now, we apply the bound:

$$\Pr(X > (1+\delta)\mu) \leq e^{\frac{-\mu\delta^2}{3}}$$

$$\Pr(X > 1+1)\,6\ln(n)) \leq e^{-\frac{6\ln(n)}{3}}$$

$$\leq e^{-2\ln(n)}$$

Recall the property of exponentials and logarithms there, so:

$$\leq e^{\ln(n^{-2})}$$

Recall from the exercise hint that $\ln(n) = \log_e(n)$

So, we have:

$$e^{\ln(\mathrm{n}^{-2})} = \frac{1}{n^2}$$

as the exercise wanted. We showed with high probability the bin with maximum load containing at most $12\ln(n)$ jobs. We applied this for *one* bin, so we have to use now the union bound; simply use the previous result multiplying by all jobs, so $m = \frac{n}{6n(\ln(n))}$: $\frac{n}{6\ln(n)} * \frac{1}{n^2} = \frac{1}{6nln(n)}$

To characterize the *no job will exceed*, use the complement event → $1 - \frac{1}{6nln(n)} = 1 - o\left(\frac{1}{n}\right)$

Written by Gabriel R.