Randomized Algorithms

Kousha Etessami

Bounding the probability of deviations from expectation

We have already seen . . .

Theorem (3.1, Markov's Inequality)

Let X be any random variable that takes only non-negative values. Then for any $a > 0$,

$$\Pr[X \geq a] \;\leq\; \frac{E[X]}{a}.$$

Theorem (3.2, Chebyshev's Inequality)

For every $a > 0$,

$$\Pr[|X - E[X]| \geq a] \;\leq\; \frac{Var[X]}{a^2}.$$

These are generic. Chernoff/Hoeffding bounds give specific, much tighter, bounds for sums of independent random variables and related distributions.

Chernoff Bounds – a first look, and applications

To gain intuition, we first state and use a couple of specific Chernoff bounds.

Theorem (Two special cases of Chernoff bounds)

Suppose we conduct a sequence of n *mutually independent Bernoulli trials,* $X_i \in \{0, 1\}$, *with probability* p *of "success" (i.e., getting a 1, i.e., heads) in each trial. Let* $X = \sum_{i=1}^{n} X_i$ *be the binomially distributed r.v. that counts the total number of successes (recall that* $E[X] = pn$*). Then:*

1. *For all* $\epsilon > 0$, $\Pr(|X - E[X]| \geq \epsilon n) \leq 2e^{-2n\epsilon^2}$.
2. *For all* $R \geq 6E[X]$, $\Pr(X \geq R) \leq 2^{-R}$.

First simple application of Chernoff bounds

Question: A biased coin is flipped 200 times consecutively, and comes up heads with probability 1/10 each time it is flipped. Give an upper bound on the probability that it will come up heads at least 120 times.

First simple application of Chernoff bounds

Question: A biased coin is flipped 200 times consecutively, and comes up heads with probability $1/10$ each time it is flipped. Give an upper bound on the probability that it will come up heads at least 120 times.

Solution: Let X be the r.v. that counts the number of heads. Recall: $\mathrm{E}[X] = pn = (1/10) \cdot 200 = 20$. By the given Chernoff bound (2.),

$$\Pr[X \geq 120] = \Pr[X \geq 6\mathrm{E}[X]] \leq 2^{-6\mathrm{E}[X]} = 2^{-(6\cdot 20)} = 2^{-120}. \quad \square$$

Note: By using Markov's inequality, we are only able to determine that $\Pr[X \geq 120] \leq \frac{\mathrm{E}[X]}{120} = \frac{20}{120} = \frac{1}{6}$.

But by using Chernoff bounds, which are specifically geared for large deviation bounds for binomial and other distributions which arise as sums of independent r.v.'s, we get that $\Pr[X \geq 120] \leq 2^{-120}$.

This is a vastly better upper bound‼

A simple but fundamental application of Chernoff bounds

One of the most basic tasks in statistics: learning a distribution

Suppose you are given a biased coin, which lands heads with probability p each time it is flipped. **But** you are not told what p is. You want to learn an estimate of what p is from samples.
This is the parameter estimation problem.

To estimate p, you can of course flip the coin n times, count the number of times, X, that it lands heads, and give the estimate:

$$\text{“}p \text{ is roughly } \frac{X}{n}.\text{”}$$

But how big does n have to be for your estimate $\frac{X}{n}$ to (probably) be “good”?

A simple but fundamental application of Chernoff bounds

One of the most basic tasks in statistics: learning a distribution

Suppose you are given a biased coin, which lands heads with probability p each time it is flipped. **But** you are not told what p is. You want to learn an estimate of what p is from samples.
This is the parameter estimation problem.

To estimate p, you can of course flip the coin n times, count the number of times, X, that it lands heads, and give the estimate:

"p is roughly $\frac{X}{n}$."

But how big does n have to be for your estimate $\frac{X}{n}$ to (probably) be "good"? Concretely, how many independent random samples (coin flips), n, do you need in order to be sure that, say:

$$\Pr[|\frac{X}{n} - p| > \frac{1}{30}] \leq \frac{1}{25} \quad ?$$

Note that $\Pr[|\frac{X}{n} - p| > \frac{1}{30}] = \Pr[|X - pn| > \frac{1}{30}n]$.

So we can use Chernoff bound (1.) to bound n.

Question: How many random samples n do I need to make sure that:

$$\Pr[|X - pn| > \frac{1}{30}n] \leq \frac{1}{25} \quad ?$$

Solution: Let X be the r.v. that counts the number of heads from n coin flips. Recall that $\mathrm{E}[X] = pn$.
By Chernoff bound (1.), $\Pr[|X - pn| \geq \epsilon n] \leq 2e^{-2n\epsilon^2}$.
Let $\epsilon = \frac{1}{30}$, and let $n = 1800$, then

$$\Pr[|X - pn| \geq \frac{1}{30}n] \leq 2e^{-2 \cdot 1800 \cdot (\frac{1}{30})^2} = 2e^{-4} = 0.0366 \leq 0.04 = \frac{1}{25}.$$

Thus, 1800 random samples (i.e., coin flips) suffice to make sure that, with probability at least $\frac{24}{25}$, your estimate $\frac{X}{n}$ for the coin's bias p is correct to within additive error at most $\frac{1}{30}$.

These kinds of bounds are crucial in statistical analysis.
(For things like "confidence intervals" , etc.)

Recall the Central Limit Theorem

One of the most fundamental theorems in all of probability and statistics. Let $X_1, \ldots, X_n$ be independent identically distributed (i.i.d.) r.v.'s, with $\mathrm{E}[X_i] = \mu_i$ and $\mathrm{Var}[X_i] = \sigma_i^2 > 0$, where $\mu_i = \mu_j$ and $\sigma_i = \sigma_j$, for all $i, j \in [n]$. Let $X = \sum_{i=1}^{n} X_i$, $\mu = \mathrm{E}[X] = n\mu_i$ and $\sigma^2 = \mathrm{Var}[X] = n\sigma_i^2$. The Central Limit Theorem tells us what the distribution of X looks like in the limit as $n \to \infty$.

Theorem (Central Limit Theorem)

As $n \to \infty$, the distribution of $\frac{X-\mu}{\sigma}$ approaches the standard normal (Gaussian) distribution $\mathbf{N}(\mathbf{0}, \mathbf{1})$. Thus, for any fixed $\beta > 0$,

$$\lim_{n\to\infty} \Pr[|X - \mu| > \beta\sigma] = \frac{1}{\sqrt{2\pi}} \int_{\beta}^{\infty} e^{-t^2/2} dt \quad \left(\approx\approx \frac{1}{\sqrt{2\pi}\beta} e^{-\beta^2/2} \right)$$

However: this theorem is only a statement about the limit behavior, and says nothing about the rate of convergence, nor behavior for particular n. Also, it describes probability of deviating by a factor β times the standard deviation σ of X, not the probability of arbitrary deviations.

(credit: A. Sinclair's lecture notes)

Humble origins: framework for proving Chernoff bounds

Proving Chernoff bounds starts by simply applying Markov's inequality to an exponential of the sum $X = X_1 + \ldots + X_n$ of independent r.v.'s. For any $t > 0$:

$$\Pr[X \geq a] = \Pr[e^{tX} \geq e^{ta}] \leq \frac{\mathrm{E}[e^{tX}]}{e^{ta}}.$$

($\mathrm{E}[e^{tX}]$ is known as the "moment generating function" of the r.v. X.
It is also used commonly in proofs of the central limit theorem.)

Hence,

$$\Pr[X \geq a] \leq \inf_{t>0} \frac{\mathrm{E}[e^{tX}]}{e^{ta}}$$

Similarly, for $t < 0$, we have:

$$\Pr[X \leq a] = \Pr[e^{tX} \geq e^{ta}] \leq \frac{\mathrm{E}[e^{tX}]}{e^{ta}}.$$

Hence

$$\Pr[X \leq a] \leq \inf_{t<0} \frac{\mathrm{E}[e^{tX}]}{e^{ta}}$$

Chernoff bounds are obtained by suitably choosing t, depending on context.

Chernoff bounds — upper tail

Poisson trials - sequence of Bernoulli variables X_i with varying p_i's.

Theorem (4.4, basic form)

Let $X_1, \ldots, X_n$ be independent Bernoulli random variables with parameter p_i for $i \in [n]$.Let $X = \sum_{i=1}^{n} X_i$, and $\mu = E[X] = \sum_{i=1}^{n} p_i$. Then for any $\delta > 0$,

$$\Pr[X \geq (1+\delta)\mu] \ \leq \ \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right)^{\mu}.$$

For example, if $\mathrm{E}[X] = \mu = pn$, and $\delta = 1$, and $p = 1/4$, then

$$\Pr[X \geq 2\mu] \ \leq \ \left(\frac{e}{4} \right)^{pn} = (0.67958)^{pn} = (0.67958)^{n/4} \leq 2^{-(n/8)}$$

Comparing with Chebyshev's inequality

Theorem (4.4, basic Chernoff)

$$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu .$$

Consider the case where $p_i = p$ and $\mu = pn$. $\mathrm{Var}[X_i] = p - p^2$, and due to independence $\mathrm{Var}[X] = (p-p^2)n = \mu(1-p)$. With Chebyshev's inequality

$$\begin{aligned}\Pr[X \geq (1+\delta)\mu] &\leq \Pr[|X-\mu| \geq \delta\mu] \\ &\leq \frac{\mu(1-p)}{\delta^2\mu^2} = \frac{1-p}{\delta^2\mu} = \frac{1-p}{\delta^2 pn} = O(1/n).\end{aligned}$$

Thus, Chebyshev gives an inverse polynomial tail whereas Chernoff gives us an exponential tail.

Chernoff bounds — upper tail

Lemma

Let $X_1, \ldots, X_n$ and X be the same as before and $\mu = E[X]$. For any $t \in \mathbb{R}$,

$$E[e^{tX}] \leq e^{\mu(e^t - 1)}.$$

Proof.

Consider

$$\mathrm{E}[e^{tX}] = \mathrm{E}\left[e^{t(\sum_{i=1}^n X_i)}\right] = E\left[\prod_{i=1}^n e^{tX_i}\right].$$

The X_i and hence the e^{tX_i} are mutually independent, so by Thm 3.3, $\mathrm{E}[e^{tX}] = \prod_{i=1}^n \mathrm{E}[e^{tX_i}]$. Each e^{tX_i} has expectation

$$\begin{aligned} \mathrm{E}[e^{tX_i}] &= p_i \cdot e^t + (1 - p_i) \cdot 1 \\ &= 1 + p_i(e^t - 1) \\ &\leq e^{p_i(e^t - 1)} \qquad \text{(by } 1 + x \leq e^x \text{ for } x \in \mathbb{R}) \end{aligned}$$

$$\Rightarrow \quad \mathrm{E}[e^{tX}] \leq \prod_{i=1}^n e^{p_i(e^t-1)} = e^{\sum_{i=1}^n p_i(e^t-1)} = e^{\mu(e^t-1)}. \qquad \square$$

Chernoff bounds — upper tail

Proof of Thm 4.4.

The event of interest is

$$X \geq (1+\delta)\mu \iff e^{tX} \geq e^{t(1+\delta)\mu}$$

for any $t > 0$. Thus

$$\begin{aligned}
\Pr[X \geq (1+\delta)\mu] &= \Pr[e^{tX} \geq e^{t(1+\delta)\mu}] \\
&\leq \frac{\mathrm{E}[e^{tX}]}{e^{t(1+\delta)\mu}} && \text{(by Markov's Inequality)} \\
&\leq \frac{e^{\mu(e^t-1)}}{e^{t(1+\delta)\mu}}. && \text{(by the last Lemma)}
\end{aligned}$$

Chernoff bounds — upper tail

Proof of Thm 4.4.

The event of interest is

$$X \geq (1+\delta)\mu \iff e^{tX} \geq e^{t(1+\delta)\mu}$$

for any $t > 0$. Thus

$$\begin{aligned}
\Pr[X \geq (1+\delta)\mu] &= \Pr[e^{tX} \geq e^{t(1+\delta)\mu}] \\
&\leq \frac{\mathrm{E}[e^{tX}]}{e^{t(1+\delta)\mu}} && \text{(by Markov's Inequality)} \\
&\leq \frac{e^{\mu(e^t-1)}}{e^{t(1+\delta)\mu}}. && \text{(by the last Lemma)}
\end{aligned}$$

This holds for any $t > 0$. For any $\delta > 0$, we can let $t = \ln(1+\delta) > 0$. This immediately yields

$$\begin{aligned}
\Pr[X \geq (1+\delta)\mu] &\leq e^{\mu(e^{\ln(1+\delta)}-1)-(\ln(1+\delta))(1+\delta)\mu} \\
&\leq \frac{e^{\mu(1+\delta-1)}}{(1+\delta)^{(1+\delta)\mu}} \\
&= \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}. \quad \square
\end{aligned}$$

Chernoff bounds — upper tail

Theorem (4.4, full)

Let $X_1, \ldots, X_n$ be independent Bernoulli random variables with parameter p_i for $i \in [n]$.Let $X = \sum_{i=1}^{n} X_i$, and $\mu = E[X]$.

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 $0 < \delta \leq 1$,*

$$\Pr[X \geq (1+\delta)\mu] \;\leq\; e^{-\mu\delta^2/3};$$

3. *For $R \geq 6\mu$,*

$$\Pr[X \geq R] \;\leq 2^{-R}.$$

Chernoff bounds — upper tail

Proof of Thm 4.4 (2.)

We already have

$$\Pr[X \geq (1+\delta)\mu] \quad \leq \quad \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}.$$

So we want to show $\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \leq e^{-\delta^2/3}$, for $\delta \in (0,1]$. Taking logarithms on both sides, we want to show

$$\delta - (1+\delta)\ln(1+\delta) \leq -\delta^2/3$$

Equivalently, we want to show

$$f(\delta) := \quad \delta - (1+\delta)\ln(1+\delta) + \delta^2/3 \leq 0$$

for $\delta \in (0,1]$. Differentiating f, we get:

$$f'(\delta) = 1 - \ln(1+\delta) - (1+\delta)\frac{1}{1+\delta} + \frac{2\delta}{3}$$

$$= -\ln(1+\delta) + \frac{2\delta}{3}.$$

Chernoff bounds — upper tail

Proof of Thm 4.4 (2.) cont.

$$f'(\delta) = -\ln(1+\delta) + \frac{2\delta}{3}.$$

Differentiate again

$$f''(\delta) \quad = \quad -\frac{1}{1+\delta} + \frac{2}{3}$$

Note

$$f''(\delta) \begin{cases} < 0 & \text{for } 0 < \delta < 1/2; \\ = 0 & \text{for } \delta = 1/2; \\ > 0 & \text{for } \delta > 1/2. \end{cases}$$

Chernoff bounds — upper tail

Proof of Thm 4.4 (2.) cont.

$$f'(\delta) = -\ln(1+\delta) + \frac{2\delta}{3}.$$

Differentiate again

$$f''(\delta) \quad = \quad -\frac{1}{1+\delta} + \frac{2}{3}$$

Note

$$f''(\delta) \begin{cases} < 0 & \text{for } 0 < \delta < 1/2; \\ = 0 & \text{for } \delta = 1/2; \\ > 0 & \text{for } \delta > 1/2. \end{cases}$$

Also $f'(0) = 0$, $f'(1) \approx -0.026 < 0$ (check $\delta = 1$ in top equation). Since f' decreases from 0 to 1/2 and then increases from 1/2 to 1, we have that $f'(\delta) < 0$, for all $\delta \in [0, 1]$.

Chernoff bounds — upper tail

Proof of Thm 4.4 (2.) cont.

$$f'(\delta) = -\ln(1+\delta) + \frac{2\delta}{3}.$$

Differentiate again

$$f''(\delta) = -\frac{1}{1+\delta} + \frac{2}{3}$$

Note

$$f''(\delta) \begin{cases} < 0 & \text{for } 0 < \delta < 1/2; \\ = 0 & \text{for } \delta = 1/2; \\ > 0 & \text{for } \delta > 1/2. \end{cases}$$

Also $f'(0) = 0$, $f'(1) \approx -0.026 < 0$ (check $\delta = 1$ in top equation). Since f' decreases from 0 to 1/2 and then increases from 1/2 to 1, we have that $f'(\delta) < 0$, for all $\delta \in (0, 1]$.

By $f(0) = 0$, this implies that $f(\delta) \leq 0$ in all of $[0, 1]$.
Hence $\delta - (1+\delta)\ln(1+\delta) \leq -\delta^2/3$, proving (2.). □

Chernoff bounds — upper tail

(3.) For $R \geq 6\mu$,

$$\Pr[X \geq R] \ \leq 2^{-R}.$$

Proof of Thm 4.4 (3.)

Let $R = (1+\delta)\mu$ and thus for $R \geq 6\mu$, we have $\delta = R/\mu - 1 \geq 5$. By Thm 4.4 (1.)

$$\begin{aligned}\Pr[X \geq (1+\delta)\mu] &\leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \\ &= \left(\frac{e^{\frac{\delta}{1+\delta}}}{1+\delta}\right)^{(1+\delta)\mu} \leq \left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu} \\ &\leq \left(\frac{e}{6}\right)^{R} \leq 2^{-R}.\end{aligned}$$

□

Chernoff Bounds (lower tail)

Theorem (4.5)

Let $X_1, \ldots, X_n$ be independent Poisson trials such that $\Pr[X_i = 1] = p_i$ for all $i \in [n]$. Let $X = \sum_{i=1}^{n} X_i$, and $\mu = E[X]$. For any $0 < \delta < 1$, we have the following Chernoff bounds*:*

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

2. $$\Pr[X \leq (1-\delta)\mu] \;\leq\; e^{-\mu\delta^2/2};$$

- Proof is similar to Thm 4.4 (see book).
- Note the bound of (2.) is slightly better than the bound for $\geq (1+\delta)\mu$.

Concentration

Corollary (4.6)

Let $X_1, \ldots, X_n$ be independent Bernoulli rv such that $\Pr[X_i = 1] = p_i$ for all $i \in [n]$. Let $X = \sum_{i=1}^{n} X_i$, and $\mu = E[X]$. Then for any $\delta, 0 < \delta < 1$,

$$\Pr[|X - \mu| \geq \delta\mu] \;\leq\; 2e^{-\mu\delta^2/3}.$$

- For most applications, we will want to work with a *symmetric* version like in this Corollary.
- We "threw away" a bit in moving from the $\left(\frac{e^{\pm\delta}}{(1\pm\delta)^{1\pm\delta}}\right)^{\mu}$ versions, but they are tricky to work with.

Analysing a collection of coin flips

Suppose we have $p_i = 1/2$ for all $i \in [n]$.

We have $\mu = \mathrm{E}[X] = \frac{n}{2}$, $\mathrm{Var}[X] = \frac{n}{4}$.
Consider the probability of being further than $5\sqrt{n}$ from μ.

Chebyshev $\Pr[|X - \mu| \geq 5\sqrt{n}] \leq \frac{\mathrm{Var}[X]}{25n} = \frac{1}{100}$

Analysing a collection of coin flips

Suppose we have $p_i = 1/2$ for all $i \in [n]$.

We have $\mu = \mathrm{E}[X] = \frac{n}{2}$, $\mathrm{Var}[X] = \frac{n}{4}$.
Consider the probability of being further than $5\sqrt{n}$ from μ.

Chebyshev $\Pr[|X - \mu| \geq 5\sqrt{n}] \leq \frac{\mathrm{Var}[X]}{25n} = \frac{1}{100}$

Chernoff Work out the δ – we need $\mu\delta = 5\sqrt{n}$, so need $\delta = 5\sqrt{n}/\mu = 10\sqrt{n}/n = \frac{10}{\sqrt{n}}$. Then by Chernoff

$$\Pr[|X - \mu| \geq 5\sqrt{n}] \leq 2e^{-\mu\delta^2/3} = 2e^{\frac{-10^2 \cdot n}{2 \cdot 3 \cdot \sqrt{n}^2}} = 2e^{-16.6\ldots}.$$

This is much smaller than the Chebyshev bound (though note it doesn't depend on n).

Get much improved bounds because Chernoff uses specialised analysis for sums of independent Bernoulli variables.

Unbiased $+1/-1$ variables

In fact, for the case of unbiased variables, we can do even better than $2e^{-\mu\delta^2/3}$. We first switch to +1/-1 variables.

Theorem (4.7)

Let $X_1, \ldots, X_n$ be independent random variables with $\Pr[X_i = 1] = 1/2 = \Pr[X_i = -1]$ *for all $i \in [n]$. Let $X = \sum_{k=1}^{n} X_k$. Note $\mu = E[X] = 0$. Then for any $a > 0$,*

$$\Pr[X \geq a] \;\leq\; e^{-a^2/2n}.$$

Proof. We again consider the moment generating function $\mathrm{E}[^{tX_i}]$. We have

$$\mathrm{E}\left[e^{tX_i}\right] = \frac{1}{2}e^t + \frac{1}{2}e^{-t} \leq e^{t^2/2}$$

where the inequality follows by Taylor expansions of e^t and e^{-t}:

$$\begin{aligned} e^t &= 1 + t + \frac{t^2}{2} + \frac{t^3}{3!} + \ldots + \frac{t^i}{i!} + \ldots \\ e^{-t} &= 1 - t + \frac{t^2}{2} - \frac{t^3}{3!} + \ldots + (-1)^i \frac{t^i}{i!} + \ldots \end{aligned}$$

So

$$\frac{1}{2}e^t + \frac{1}{2}e^{-t} = \sum_{i=0}^{\infty} \frac{t^{2i}}{(2i)!} \leq \sum_{i=0}^{\infty} \frac{(t^2/2)^i}{i!} = e^{t^2/2}.$$

Unbiased $+1/-1$ variables

Proof of Thm 4.7 cont.

Thus we have

$$\mathrm{E}\left[e^{tX}\right] = \prod_{i=1}^{n} \mathrm{E}\left[e^{tX_i}\right] \leq e^{t^2 n/2};$$

Unbiased $+1/-1$ variables

Proof of Thm 4.7 cont.

Thus we have

$$\mathrm{E}\left[e^{tX}\right] = \prod_{i=1}^{n} \mathrm{E}\left[e^{tX_i}\right] \leq e^{t^2 n/2};$$

Hence

$$\Pr[X \geq a] = \Pr[e^{tX} \geq e^{ta}] \leq \frac{\mathrm{E}\left[e^{tX}\right]}{e^{ta}} = e^{(t^2 n/2) - ta}.$$

This time setting $t = a/n$ we get:

$$\Pr[X \geq a] \leq e^{-a^2/2n}.$$

□

Unbiased $+1/-1$ variables

Proof of Thm 4.7 cont.

Thus we have

$$\mathrm{E}\left[e^{tX}\right] = \prod_{i=1}^{n} \mathrm{E}\left[e^{tX_i}\right] \leq e^{t^2 n/2};$$

Hence

$$\Pr[X \geq a] = \Pr[e^{tX} \geq e^{ta}] \leq \frac{\mathrm{E}\left[e^{tX}\right]}{e^{ta}} = e^{(t^2 n/2) - ta}.$$

This time setting $t = a/n$ we get:

$$\Pr[X \geq a] \leq e^{-a^2/2n}.$$

□

The lower tail is completely symmetric. Think $-X$.

$$\Pr[X \leq -a] = \Pr[-X \geq a] \leq e^{-a^2/2n}.$$

Unbiased $+1/-1$ variables

Corollary (4.8)

Let $X_1, \ldots, X_n$ be independent random variables with $\Pr[X_i = 1] = 1/2 = \Pr[X_i = -1]$ for all $i \in [n]$. Let $X = \sum_{k=1}^{n} X_k$. Note $\mu = E[X] = 0$. Then for any $a > 0$,

$$\Pr[|X| \geq a] \leq 2e^{-a^2/2n}.$$

Unbiased 0/1 variables

Consider $Y_1, \ldots, Y_n$ with $\Pr[Y_i = 1] = \Pr[Y_i = 0] = 1/2$ for all $i \in [n]$.
Define $X_i = 2Y_i - 1$ for every $i \in [n]$. Then

$$X_i = \begin{cases} 1 & \text{if } Y_i = 1 \\ -1 & \text{if } Y_i = 0 \end{cases}$$

Corollary (4.9, 4.10)

For $Y = \sum_{i=1}^n Y_i$ and $X = \sum_{i=1}^n X_i$, we have

$$\Pr[Y \geq \tfrac{n}{2} + a] = \Pr[X \geq 2a] \quad \leq \quad e^{-2a^2/n};$$

$$\Pr[Y \leq \tfrac{n}{2} - a] = \Pr[X \leq -2a] \quad \leq \quad e^{-2a^2/n}.$$

Let $\mu = \mathrm{E}[Y] = n/2$ and $a = \delta\mu$, for $\delta > 0$. Note $\mu + a = (1+\delta)\mu$. Thus

$$\Pr[Y \geq (1+\delta)\mu] = e^{-2a^2/n} = e^{-2\delta^2\mu^2/(2\mu)} = e^{-\delta^2\mu}.$$

Biased 0/1 variables – revisited

We didn't yet quite manage to prove one of the original Chernoff bounds I stated, namely:

Theorem

Let $X_1, \ldots, X_n$, be independent Bernoulli trials, $X_i \in \{0, 1\}$, with $\Pr[X_i = 1] = p$. Let $X = \sum_{i=1}^{n} X_i$. Hence we have $E[X] = pn$. Then

$$\textit{for all } \epsilon > 0, \ \Pr[|X - E[X]| \geq \epsilon n] \leq 2e^{-2n\epsilon^2}.$$

This has slightly better constants in the exponent on the RHS than what we have shown so far. It follows by a slightly more involved analysis.
(This is exercise 4.13 in the book.)

References

- Read Chapter 4 of book [MU], sections 4.1-4.5.
- We will continue with Chernoff/Hoeffding bounds, and their applications, next time.