Bounding deviation

We already have . . .

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 $a > 0$,
\[
\Pr[|X - E[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2}.
\]

These are generic. Chernoff/Hoeffding bounds (specific) give tighter bounds for sums of independent variables and related distributions.

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 $\mu = E[X]$. 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 $\mu = pn$ and $\delta = 1$,
\[
\Pr[X \geq 2\mu] \leq \left(\frac{e}{4}\right)^{pn} = \exp(-\Omega(n)).
\]

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 $\mu = pn$. Due to independence, $\text{Var}[X_i] = p - p^2$ and $\text{Var}[X] = (p - p^2)n = \mu(1 - p)$. With Chebyshev’s inequality
\[
\Pr[X \geq (1 + \delta)\mu] \leq \Pr[|X - \mu| \geq \delta\mu] \leq \frac{\mu(1 - p)}{\delta^2\mu^2} = \frac{1 - p}{\delta^2} = O(1/n).
\]

Thus, Chebyshev gives an inverse polynomial tail whereas Chernoff gives us an exponential tail.
Comparing with Chebyshev’s inequality

Theorem (4.4, basic Chernoff)

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

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

\[ \Pr[X \geq (1 + \delta)\mu] \leq \frac{\mu(1 - p)}{\delta^2 \mu^2} = \frac{1 - p}{\delta^2} = O(1/n). \]

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

However, both give us constant concentration bound for a window of width \( O(\sqrt{n}) \), although Chernoff’s constant is much better.

Chernoff bounds — upper tail

Proof of Thm 4.4.

The event of interest is

\[ X \geq (1 + \delta)\mu \iff e^X \geq e^{(1+\delta)\mu} \]

which, in turn, is equivalent to \( e^X \geq e^{(1+\delta)\mu} \) for any \( t > 0 \).

\[ \Pr[X \geq (1 + \delta)\mu] = \Pr[e^X \geq e^{(1+\delta)\mu}] \]

\[ \leq \frac{E[e^X]}{e^{(1+\delta)\mu}} \quad \text{(by Markov’s Inequality)} \]

\[ \leq \frac{e^{\mu(d-1)}}{e^{(1+\delta)\mu}}. \quad \text{(by the last Lemma)} \]

Chernoff bounds — upper tail

Lemma

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

\[ E[e^{tX}] \leq e^{t^2(\mu)^2}. \]

Proof.

Consider

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

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

\[ E[e^{X_i}] = p_i \cdot e^t + (1 - p_i) \cdot 1 = 1 + p_i(e^t - 1) \leq e^{x/(d-1)}, \quad \text{(by } 1 + x \leq e^x \text{ for } x \in \mathbb{R}) \]

\[ \Rightarrow \quad E[e^{X}] \leq \prod_{i=1}^{n} e^{x/(d-1)} = e^{\sum_{i=1}^{n} p_i(x/(d-1)) = e^{\mu(x/(d-1))}. \]

This holds for any \( t > 0 \), and we want to pick \( t \) to minimize the right hand side, which is \( \text{RHS} := e^{\mu(t/(d-1))'. \text{ Differentiate the exponent,} \]

\[ (\ln \text{RHS})' = \mu e^x - (1 + \delta)\mu. \]

Thus, \( \text{RHS} \) decreases if \( t \leq \ln(1 + \delta) \) and increases if \( t \geq \ln(1 + \delta) \). Its minimum is taken at \( t = \ln(1 + \delta) \).
Chernoff bounds — upper tail

Proof of Thm 4.4 (cont.)

Now take $t = \ln(1 + \delta)$ (and note this is $> 0$) to see

$$\Pr[X \geq (1 + \delta)\mu] \leq e^{t(1 - (1 + \delta)\mu)}$$

$$\leq \frac{e^{t(1 + \delta) - t}}{\ln(1 + \delta)}(1 + \delta)\mu$$

$$= \left(\frac{e^\delta}{(1 + \delta)(1 + \delta)}\right)^\mu.$$  \[\square\]

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.)

Already have

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

Comparing the RHS with $\leq e^{-\mu\delta^2/3},$ we want

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

We will show the following $f$ is always negative for $\delta \in (0, 1)$

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

Differentiating,

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

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

Chernoff bounds — upper tail

Proof of Thm 4.4 (2.) cont.

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

Differentiate again

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

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

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

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 \( f'(\delta) < 0 \) on \((0, 1)\). Hence \( \delta = 1 + \delta \ln(1 + \delta) < -\delta^2/3, \) proving (2.). \( \square \)

---

Chernoff bounds — upper tail

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 \( \delta = R/\mu - 1 \geq 5 \). By Thm 4.4 (1)

\[
\Pr[X \geq (1 + \delta)\mu] \leq \left( \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^\mu \\
= \left( \frac{e^{\delta/\mu}}{1 + \delta} \right)^{(1+\delta)\mu} \\
\leq \left( \frac{e}{6} \right)^R \leq 2^{-R} \]

\( \square \)

---

Chernoff bounds — upper tail

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 \( \delta = R/\mu - 1 \geq 5 \). By Thm 4.4 (1)

\[
\Pr[X \geq (1 + \delta)\mu] \leq \left( \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^\mu \\
= \left( \frac{e^{\delta/\mu}}{1 + \delta} \right)^{(1+\delta)\mu} \\
\leq \left( \frac{e}{6} \right)^R \leq 2^{-R} \]

Thm 4.4 (1.) is the strongest. The other two are slightly weaker but easy to use.
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.
- Bound of (2.) is slightly better than the bound for $\geq (1+\delta)\mu$.
- No (3.) Why?

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 the 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 $\mu = E[X] = \frac{n}{2}$, $\Var[X] = \frac{n}{4}$.

Consider the probability of being further than $5\sqrt{n}$ from $\mu$.

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

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

Choose $\delta = 5\sqrt{n}/\mu = 10\sqrt{n}/n = \frac{10}{\sqrt{n}}$. Then by Chernoff

$$\Pr[|X - \mu| \geq 5\sqrt{n}] \leq 2e^{-\mu \delta^2/3} = 2e^{-\frac{10n^2}{100 \cdot 100 \sqrt{n}}} = 2e^{-16.6\ldots}.$$  

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.
Comparison with Chebyshev

For i.i.d. coin flips,
\[ \Pr[|X - \mu| > D] \leq p \]

<table>
<thead>
<tr>
<th>Deviation ( p )</th>
<th>Constant</th>
<th>( O\left(\frac{1}{n^2}\right) )</th>
<th>( \exp(-\Omega(n)) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( D ) for Chebyshev</td>
<td>( \Omega(\sqrt{n}) )</td>
<td>( \Omega(\sqrt{n} \cdot n^{\alpha/2}) )</td>
<td>( \exp(\Omega(n)) )</td>
</tr>
<tr>
<td>( D ) for Chernoff</td>
<td>( \Omega(\sqrt{n}) )</td>
<td>( \Omega(\sqrt{n \log n}) )</td>
<td>( \Omega(n) )</td>
</tr>
</tbody>
</table>

Proof of Thm 4.7 cont.

Use the last estimate
\[ E[e^{tX}] = \prod_{i=1}^{n} E[e^{tX_i}] \leq e^{t^2 n/2}; \]

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 will once again consider the moment generating function \( e^{tX} \):

\[ E[e^{tX}] = \frac{e^t + e^{-t}}{2} \leq e^{t^2/2}, \]

where the last estimate is due to Taylor expansion.

Unbiased +1/−1 variables

Proof of Thm 4.7 cont.

Use the last estimate
\[ E[e^{tX}] = \prod_{i=1}^{n} E[e^{tX_i}] \leq e^{t^2 n/2}; \]

\[ \Pr[X \geq a] = \Pr[e^{tX} \geq e^{ta}] \leq \frac{E[e^{tX}]}{e^{ta}} = e^{t^2 n/2 - ta}. \]

Once again, minimizing the exponent gives us \( t = a/n \) and

\[ \Pr[X \geq a] \leq e^{-a^2/2n}. \]
Once again, minimizing the exponent gives us $t = a/n$ and

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

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

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