CSE 613: Parallel Programming

Lecture 5.5
(High Probability Bounds)

Rezaul A. Chowdhury

Department of Computer Science
SUNY Stony Brook
Spring 2019

Markov's Inequality

Theorem 1: Let X be a random variable that assumes only nonnegative values. Then for all $\delta > 0$,

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

Proof: For $\delta > 0$, let

$$I = \begin{cases} 1 & \text{if } X \geq \delta; \\ 0 & \text{otherwise.} \end{cases}$$

Since $X \geq 0$, $I \leq \frac{X}{\delta}$.

We also have, $E[I] = \Pr[I = 1] = \Pr[X \geq \delta]$.

Then $\Pr[X \geq \delta] = E[I] \leq E\left[\frac{X}{\delta}\right] \leq \frac{E[X]}{\delta}$.

Example: Coin Flipping

Let us bound the probability of obtaining more than $\frac{3n}{4}$ heads in a sequence of n fair coin flips.

Let

$$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is heads;} \\ 0 & \text{otherwise.} \end{cases}$$

Then the number of heads in n flips, $X = \sum_{i=1}^{n} X_i$.

We know, $E[X_i] = \Pr[X_i = 1] = \frac{1}{2}$.

Hence, $E[X] = \sum_{i=1}^{n} E[X_i] = \frac{n}{2}$.

Then applying Markov's inequality,

$$\Pr\left[X \geq \frac{3n}{4}\right] \leq \frac{E[X]}{3n/4} = \frac{n/2}{3n/4} = \frac{2}{3}.$$

Chebyshev's Inequality

Theorem 2: For any $\delta > 0$,

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

Proof: Observe that $\Pr[|X - E[X]| \geq \delta] = \Pr\left[(X - E[X])^2 \geq \delta^2\right].$

Since $(X - E[X])^2$ is a nonnegative random variable, we can use Markov's inequality,

$$\Pr\left[(X - E[X])^2 \geq \delta^2\right] \leq \frac{E\left[(X-E[X])^2\right]}{\delta^2} = \frac{Var[X]}{\delta^2}.$$

Example: n Fair Coin Flips

$$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is heads;} \\ 0 & \text{otherwise.} \end{cases}$$

Then the number of heads in n flips, $X = \sum_{i=1}^{n} X_i$.

We know, $E[X_i] = \Pr[X_i = 1] = \frac{1}{2}$ and $E[(X_i)^2] = E[X_i] = \frac{1}{2}$.

Then $Var[X_i] = E[(X_i)^2] - (E[X_i])^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$.

Hence, $E[X] = \sum_{i=1}^{n} E[X_i] = \frac{n}{2}$ and $Var[X] = \sum_{i=1}^{n} Var[X_i] = \frac{n}{4}$.

Then applying Chebyshev's inequality,

$$\Pr\left[X \geq \frac{3n}{4}\right] \leq \Pr\left[|X - E[X]| \geq \frac{n}{4}\right] \leq \frac{Var[X]}{(n/4)^2} = \frac{n/4}{(n/4)^2} = \frac{4}{n}.$$

Coin Flipping and Randomized Algorithms

Suppose we have an algorithm that is

- correct (heads) only with probability $p \in (0,1)$, and

- incorrect (tails) with probability $1 - p$.

Question: How many times should we run the algorithm to be reasonably confident that it returns at least one correct solution?

- Las Vegas Algorithm: You keep running the algorithm until you get a correct solution. What is the bound on running time?

- Monte Carlo Algorithm: You stop after a certain number of iterations no matter whether you found a correct solution or not. What is the probability that your solution is correct (or you found a solution)?

Coin Flipping and Randomized Algorithms

Suppose we have an algorithm that is

- correct (heads) only with probability $p \in (0,1)$, and

- incorrect (tails) with probability $1 - p$.

Suppose we run the algorithm k times.

Then probability that no run produces a correct solution is $(1 - p)^k$.

$\therefore$ probability of getting at least one correct solution is $1 - (1 - p)^k$.

Set $k = \ln_{\frac{1}{1-p}} \left(\frac{n^\alpha}{c} \right)$, where $\alpha \geq 1$ and $c > 0$ is a constant.

Then the probability that at least one run produces a correct solution

is $1 - (1 - p)^k = 1 - \frac{c}{n^\alpha}$.

An event Π is said to occur with high probability if $\Pr[\Pi] \geq 1 - \frac{c}{n^\alpha}$.

w.h.p.

Example: A Coloring Problem

Let S be a set of n items.

For $1 \leq l \leq k$, let $S_l \subseteq S$ such that for every pair of $i, j \in [1, k]$ with $i \neq j$, $S_i \neq S_j$ but not necessarily $S_i \cap S_j = \emptyset$.

For each $l \in [1, k]$, let $|S_l| = r$, where $k \leq 2^{r-2}$.

Problem: Color each item of S with one of two colors, red and blue, such that each S_l contains at least one red and one blue item.

Algorithm: Take each item of S and color it either red or blue independently at random (with probability $\frac{1}{2}$).

Clearly, the algorithm does not always lead to a valid coloring (i.e., satisfy the constraints given in our problem statement).

What is the probability that it produces a valid coloring?

Example: A Coloring Problem

For $1 \leq l \leq k$, let R_l and B_l be the events that all items of S_l are colored red and blue, respectively.

Then $\Pr[R_l] = \Pr[B_l] = \left(\frac{1}{2}\right)^r = 2^{-r}$ for every $l \in [1, k]$.

$\therefore \Pr\left[\cup_{l=1}^{k} R_l\right] = \Pr\left[\cup_{l=1}^{k} B_l\right] = k 2^{-r} \leq 2^{r-2} 2^{-r} = \frac{1}{4}$.

Thus $\Pr\left[\cup_{l=1}^{k} (R_l \cup B_l)\right] \leq 2 \times \frac{1}{4} = \frac{1}{2}$.

$\therefore \Pr\left[\cap_{l=1}^{k} (\bar{R}_l \cap \bar{B}_l)\right] = 1 - \Pr\left[\cup_{l=1}^{k} (R_l \cup B_l)\right] \geq 1 - \frac{1}{2} = \frac{1}{2}$.

Hence, the algorithm is correct with probability at least $\frac{1}{2}$.

To check if the algorithm has produced a correct result we simply check the items in each S_l to verify that neither R_l nor B_l holds.

Hence, we can use this simple algorithm to design a Las Vegas algorithm for solving the coloring problem!
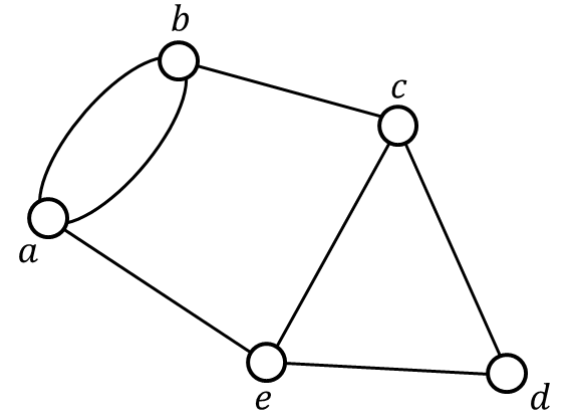
Example: The Min-Cut Problem

Let $G = (V, E)$ be a connected, undirected multigraph with $|V| = n$.

A *cut* in G is a set $C \subseteq E$, such that $G' = (V, E \setminus C)$ is not connected.

A *min-cut* is a cut of minimum cardinality.

The multigraph on the right has a min-cut of size 2: $\{(a, e), (b, c)\}$ and $\{(c, d), (d, e)\}$.

Most deterministic algorithms for finding min-cuts are based on network flows and hence are quite complicated.

Instead in this lecture we will look at a very simple probabilistic algorithm that finds min-cuts with some probability $p > 0$.
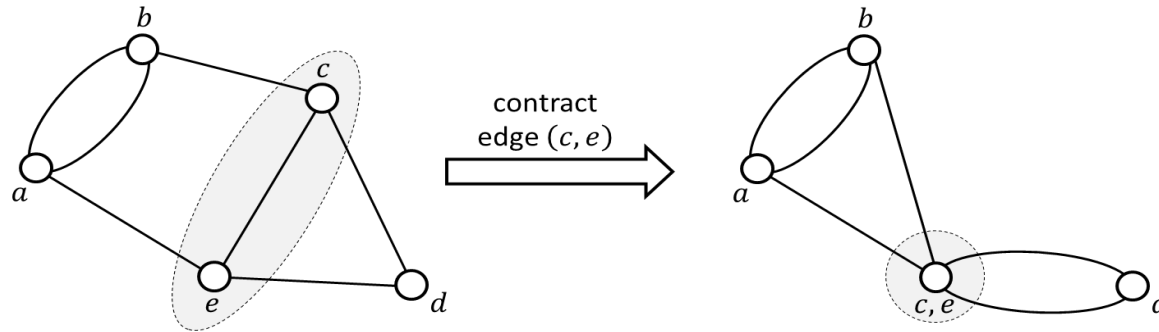
Example: The Min-Cut Problem

We apply the following contraction step $n - 2$ times on $G = (V, E)$:

Select an edge (say, (u, v)) from E uniformly at random.

Merge u and v into a single super vertex.

Remove all edges between u and v from E.

If as a result of the contraction there are more than one edges between some pairs of super vertices retain them all.

Let the initial graph be $G_0 = (V_0, E_0)$, where $V_0 = V$ and $E_0 = E$.

Let $G_i = (V_i, E_i)$ be the multigraph after step $i \in [1, n - 2]$.

Then clearly, $|V_i| = n - i$ and thus $|V_{n-2}| = 2$.

We return E_{n-2} as our solution.

Example: The Min-Cut Problem

Let us fix our attention on a particular min-cut C of G.

What is the probability that $E_{n-2} = C$?

Suppose $|C| = k$.

Then each vertex of $G_0 = G$ must have degree at least k as otherwise G_0 can be disconnected by removing fewer than k edges.

Hence, $|E| = |E_0| \geq k|V_0|/2 = kn/2$.

Let Π_i be the event of not picking an edge of C for contraction in step $i \in [1, n-2]$.

Then clearly, $\Pr[\Pi_1] = 1 - \dfrac{k}{|E_0|} \geq 1 - \dfrac{k}{kn/2} = 1 - \dfrac{2}{n}$

Also $\Pr[\Pi_2 | \Pi_1] = 1 - \dfrac{k}{|E_1|} \geq 1 - \dfrac{k}{k(n-1)/2} = 1 - \dfrac{2}{n-1}$

In general, $\Pr\left[\Pi_i \mid \cap_{j=1}^{i-1} \Pi_j\right] = 1 - \dfrac{k}{|E_{i-1}|} \geq 1 - \dfrac{k}{k(n-i+1)/2} = 1 - \dfrac{2}{n-i+1}$

Example: The Min-Cut Problem

The probability that no edge of C was ever picked by the algorithm is:

$$\Pr\left[\cap_{i=1}^{n-2} \Pi_i\right] \geq \prod_{i=1}^{n-2}\left(1 - \frac{2}{n-i+1}\right) = \frac{2}{n(n-1)} > \frac{2}{n^2}.$$

So $\Pr[E_{n-2} = C] > \frac{2}{n^2}$, and $\Pr[E_{n-2} \neq C] < 1 - \frac{2}{n^2}$.

Suppose we run the algorithm $n^2/2$ times, and return the smallest cut, say C', obtained from those $n^2/2$ attempts.

Then $\Pr[C' \neq C] < \left(1 - \frac{2}{n^2}\right)^{n^2/2} < \frac{1}{e} \Rightarrow \Pr[C' = C] > 1 - \frac{1}{e}.$

Hence, the algorithm will return a min-cut with probability $> 1 - \frac{1}{e}$.

But we do not know how to detect if the cut returned by the algorithm is, indeed, a min-cut.

Still we can design a Monte-Carlo algorithm based on this simple idea to produce a min-cut with high probability!

When Only One Success is Not Enough

In both examples we have looked at so far, we were happy with only one success. The analysis was easy.

But sometimes we need the algorithm to be successful for at least or at most a certain number of times (we will see a very familiar such example shortly).

The number of successful runs required often depends on the size of the input.

How do we analyze those algorithms?
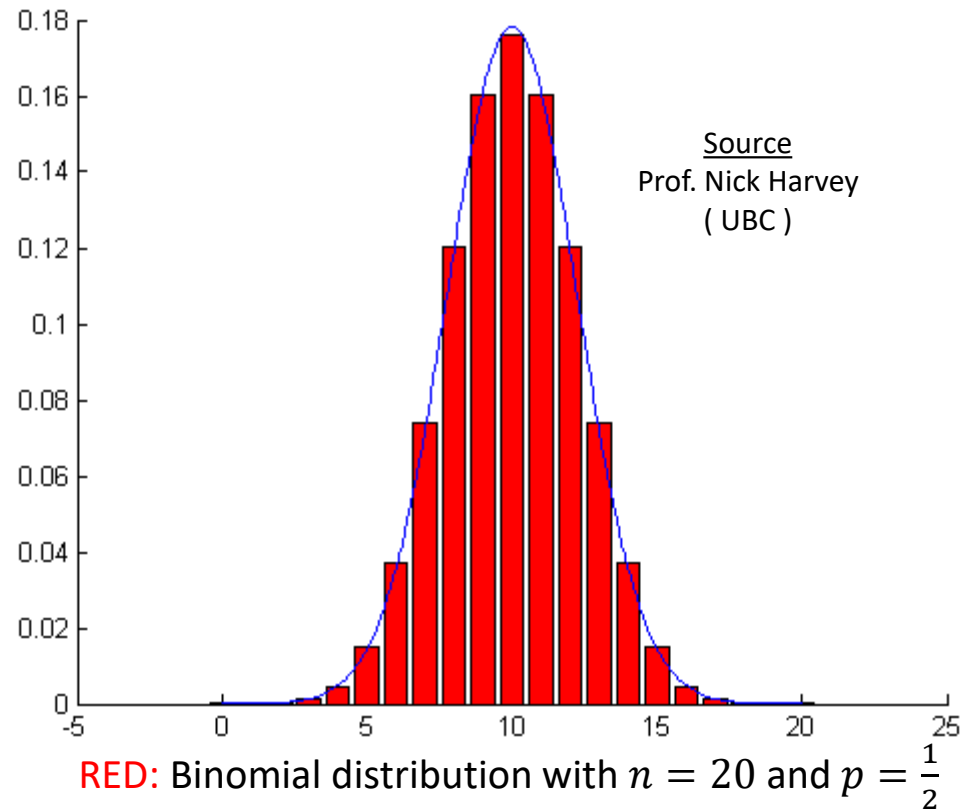
Binomial Distribution

The binomial distribution is the discrete probability distribution of the #successes in a sequence of n independent yes/no experiments (i.e., Bernoulli trials), each of which succeeds with probability p.

Source
Prof. Nick Harvey
(UBC)

RED: Binomial distribution with $n = 20$ and $p = \frac{1}{2}$

Probability mass function:

$$f(k; n, p) = \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad 0 \le k \le n$$

Cumulative distribution function:

$$F(k; n, p) = \Pr(X \le k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}, \qquad 0 \le k \le n$$

Approximating with Normal Distribution

Normal distribution with mean μ and variance σ^2 is given by:

$$f\left(x; \mu, \sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad x \in \Re$$

For fixed p as n increases the binomial distribution with parameters n and p is well approximated by a normal distribution with $\mu = np$ and $\sigma^2 = np(1-p)$.
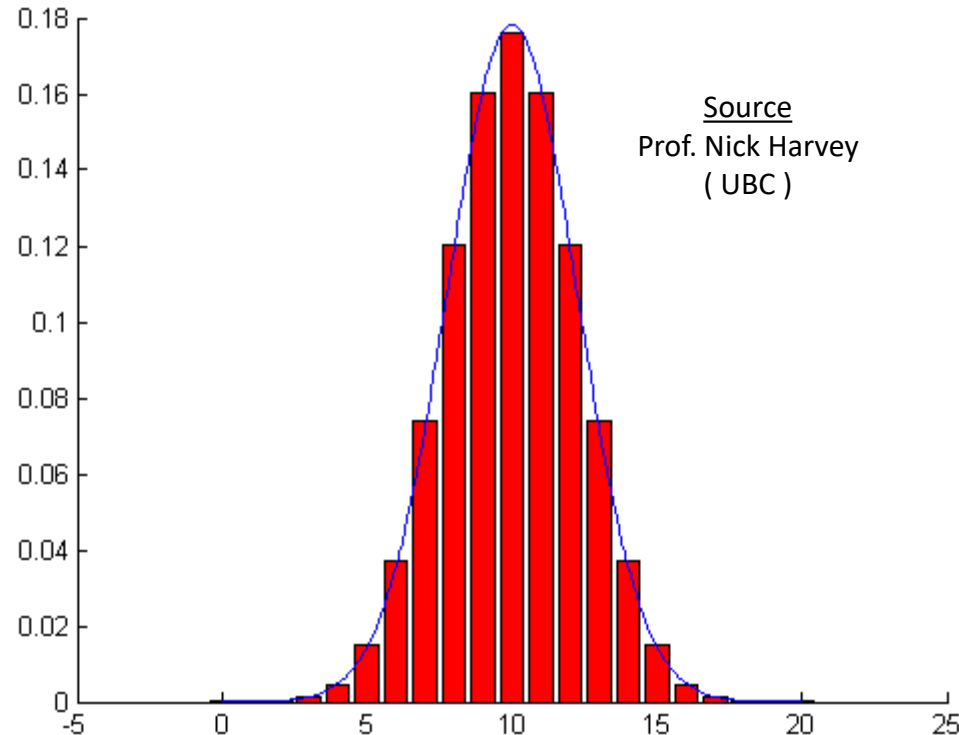
RED: Binomial distribution with

$\qquad n = 20$ and $p = \frac{1}{2}$

BLUE: Normal distribution with

$\qquad \mu = np = 10$

$\qquad$ and $\sigma^2 = np(1-p) = 5$

Approximating with Normal Distribution

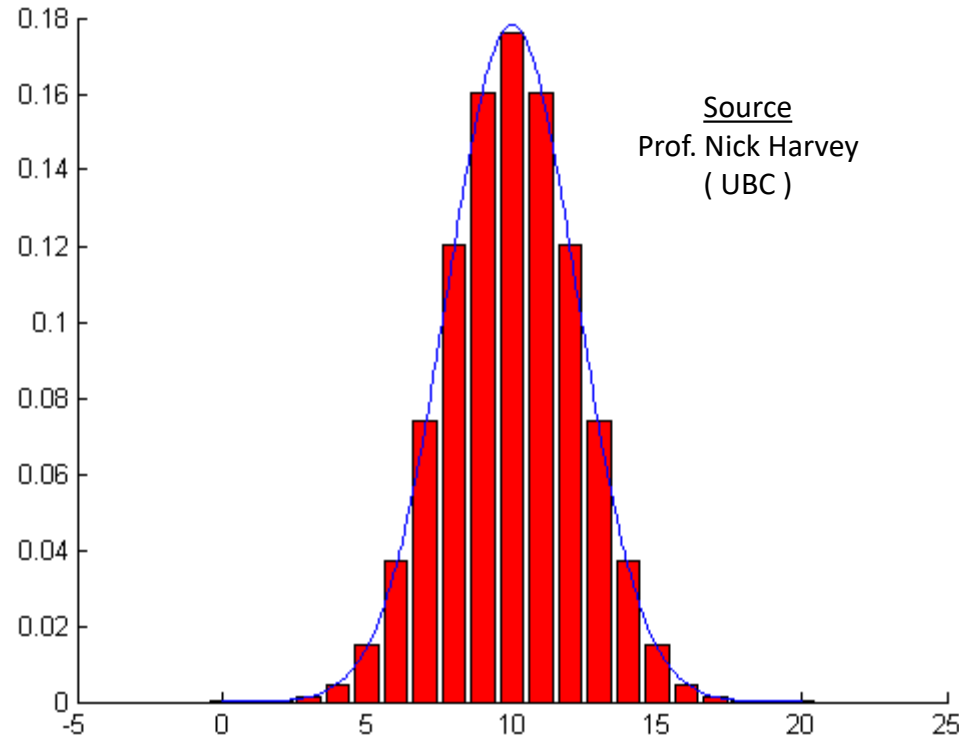Normal distribution with mean μ and variance σ^2 is given by:

$$f\left(x; \mu, \sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad x \in \Re$$

The probability that a normally distributed random variable lies in the interval $(\infty, x]$ is given by:

$$F\left(x; \mu, \sigma^2\right) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right],$$

where, $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt.$

But $\mathrm{erf}(z)$ cannot be expressed in closed form in terms of elementary functions, and hence difficult to evaluate.

Source
Prof. Nick Harvey
(UBC)

Approximating with Poisson Distribution

Poisson distribution with mean $\mu > 0$ is given by:

$$f(k; \mu) = \frac{\mu^k e^{-\mu}}{k!}, \qquad k = 0, 1, 2, \ldots$$

If np is fixed and n increases the binomial distribution with parameters n and p is well approximated by a Poisson distribution with $\mu = np$.
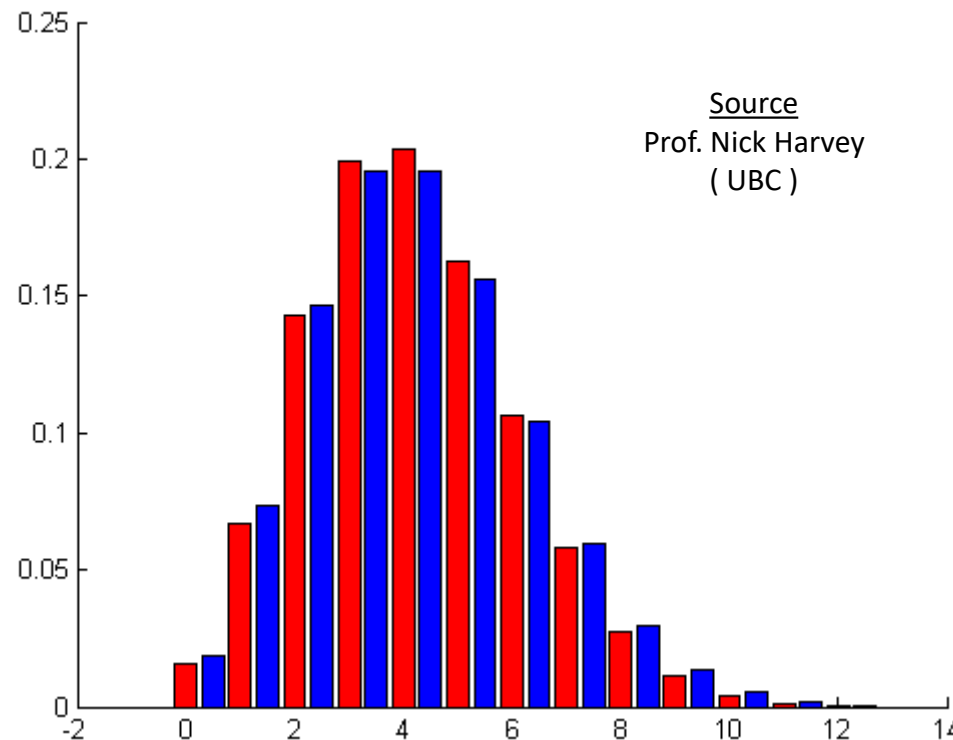
RED: Binomial distribution with

$$n = 50 \text{ and } p = \frac{4}{n}$$

BLUE: Poisson distribution with

$$\mu = np = 200$$

Observe that the asymmetry in the plot cannot be well approximated by a symmetric normal distribution.

Preparing for Chernoff Bounds

Lemma 1: Let $X_1, \ldots, X_n$ be independent Poisson trials, that is, each X_i is a 0-1 random variable with $\Pr[X_i = 1] = p_i$ for some p_i. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = E[X]$. Then for any $t > 0$,

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

Proof: $E\left[e^{tX_i}\right] = p_i e^{t \times 1} + (1 - p_i) e^{t \times 0} = p_i e^t + (1 - p_i)$

$$= 1 + p_i\left(e^t - 1\right)$$

But for any y, $1 + y \leq e^y$. Hence, $E\left[e^{tX_i}\right] \leq e^{p_i(e^t - 1)}$.

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

$$\leq \prod_{i=1}^{n} e^{p_i(e^t - 1)} = e^{(e^t - 1) \sum_{i=1}^{n} p_i}$$

But, $\mu = E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p_i$.

Hence, $E\left[e^{tX}\right] \leq e^{(e^t - 1)\mu}$

Chernoff Bound 1

Theorem 3: Let $X_1, \ldots, X_n$ be independent Poisson trials, that is, each X_i is a 0-1 random variable with $\Pr[X_i = 1] = p_i$ for some p_i. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = E[X]$. Then for any $\delta > 0$,

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

Proof: Applying Markov's inequality for any $t > 0$,

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

$$\leq \frac{e^{\left(e^t - 1\right)\mu}}{e^{t(1+\delta)\mu}} \quad [\text{ Lemma 1 }]$$

Setting $t = \ln(1 + \delta) > 0$, i.e., $e^t = 1 + \delta$, we get,

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

Chernoff Bound 2

Theorem 4: For $0 < \delta < 1$, $\Pr[X \geq (1+\delta)\mu] \leq e^{-\frac{\mu\delta^2}{3}}$.

Proof: From Theorem 3, for $\delta > 0$, $\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$.

We will show that for $0 < \delta < 1$, $\frac{e^\delta}{(1+\delta)^{(1+\delta)}} \leq e^{-\frac{\delta^2}{3}}$

$$\Rightarrow \delta - (1+\delta)\ln(1+\delta) \leq -\frac{\delta^2}{3}$$

That is, $f(\delta) = \delta - (1+\delta)\ln(1+\delta) + \frac{\delta^2}{3} \leq 0$

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

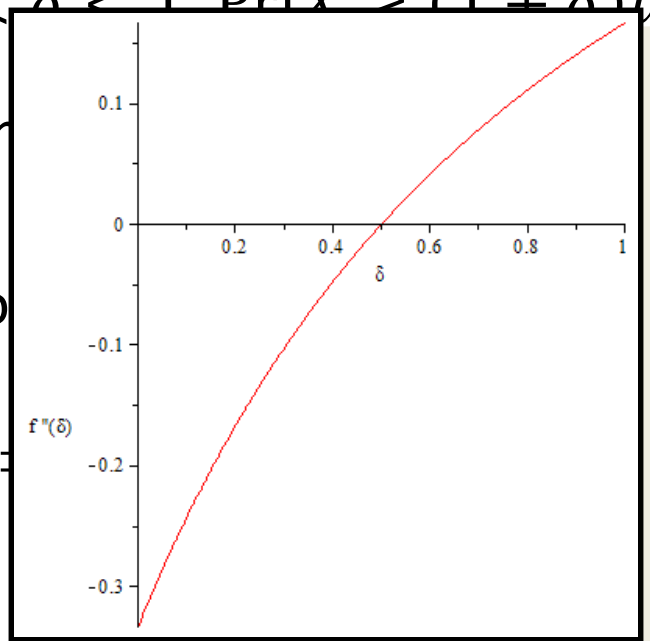Observe that $f''(\delta) < 0$ for $0 \leq \delta \leq \frac{1}{2}$, and $f''(\delta) > 0$ for $\delta > \frac{1}{2}$.

Chernoff Bound 2

Theorem 4: For $0 < \delta < 1$, $\Pr[X > (1 + \delta)\mu] \le e^{-\frac{\mu\delta^2}{3}}$.

Proof: From Theorem ⬚ $+ \delta)\mu] \le \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$.

We will show that fo⬚ $e^{-\frac{\delta^2}{3}}$

$\le -\frac{\delta^2}{3}$

That is, $f(\delta) = \delta - $ ⬚ 0

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

Observe that $f''(\delta) < 0$ for $0 \le \delta \le \frac{1}{2}$, and $f''(\delta) > 0$ for $\delta > \frac{1}{2}$.

Chernoff Bound 2

Theorem 4: For $0 < \delta < 1$, $\Pr[X \geq (1+\delta)\mu] \leq e^{-\frac{\mu\delta^2}{3}}$.

Proof: From Theorem 3, for $\delta > 0$, $\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$.

We will show that for $0 < \delta < 1$, $\frac{e^\delta}{(1+\delta)^{(1+\delta)}} \leq e^{-\frac{\delta^2}{3}}$

$$\Rightarrow \delta - (1+\delta)\ln(1+\delta) \leq -\frac{\delta^2}{3}$$

That is, $f(\delta) = \delta - (1+\delta)\ln(1+\delta) + \frac{\delta^2}{3} \leq 0$

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

Observe that $f''(\delta) < 0$ for $0 \leq \delta \leq \frac{1}{2}$, and $f''(\delta) > 0$ for $\delta > \frac{1}{2}$.

Hence, $f'(\delta)$ first decreases and then increases over $[0,1]$.

Since $f'(0) = 0$ and $f'(1) < 0$, we have $f'(\delta) \leq 0$ over $[0,1]$.
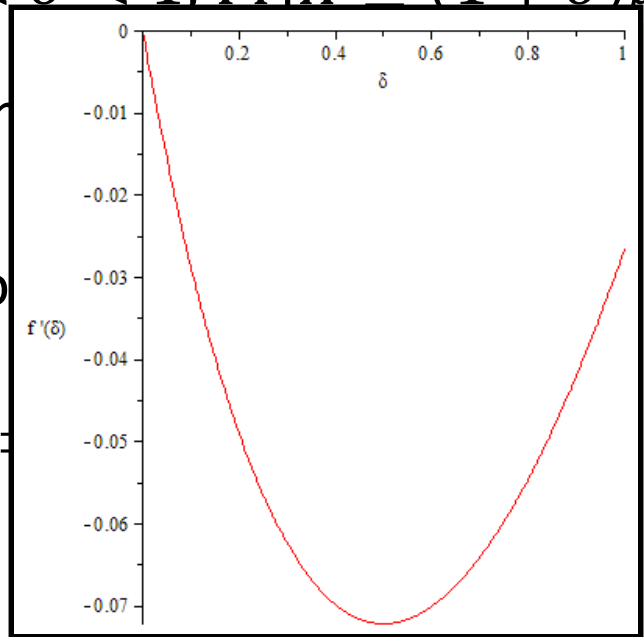
Chernoff Bound 2

Theorem 4: For $0 < \delta < 1$, $\Pr[X \geq (1 + \delta)\mu] \leq e^{-\frac{\mu\delta^2}{3}}$.

Proof: From Theorem $+ \delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$.

We will show that fo $e^{-\frac{\delta^2}{3}}$

$\leq -\frac{\delta^2}{3}$

That is, $f(\delta) = \delta -$ 0

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

Observe that $f''(\delta) < 0$ for $0 \leq \delta \leq \frac{1}{2}$, and $f''(\delta) > 0$ for $\delta > \frac{1}{2}$.

Hence, $f'(\delta)$ first decreases and then increases over $[0,1]$.

Since $f'(0) = 0$ and $f'(1) < 0$, we have $f'(\delta) \leq 0$ over $[0,1]$.

Chernoff Bound 2

Theorem 4: For $0 < \delta < 1$, $\Pr[X \geq (1+\delta)\mu] \leq e^{-\frac{\mu\delta^2}{3}}$.

Proof: From Theorem 3, for $\delta > 0$, $\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$.

We will show that for $0 < \delta < 1$, $\frac{e^\delta}{(1+\delta)^{(1+\delta)}} \leq e^{-\frac{\delta^2}{3}}$

$$\Rightarrow \delta - (1+\delta)\ln(1+\delta) \leq -\frac{\delta^2}{3}$$

That is, $f(\delta) = \delta - (1+\delta)\ln(1+\delta) + \frac{\delta^2}{3} \leq 0$

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

Observe that $f''(\delta) < 0$ for $0 \leq \delta \leq \frac{1}{2}$, and $f''(\delta) > 0$ for $\delta > \frac{1}{2}$.

Hence, $f'(\delta)$ first decreases and then increases over $[0,1]$.

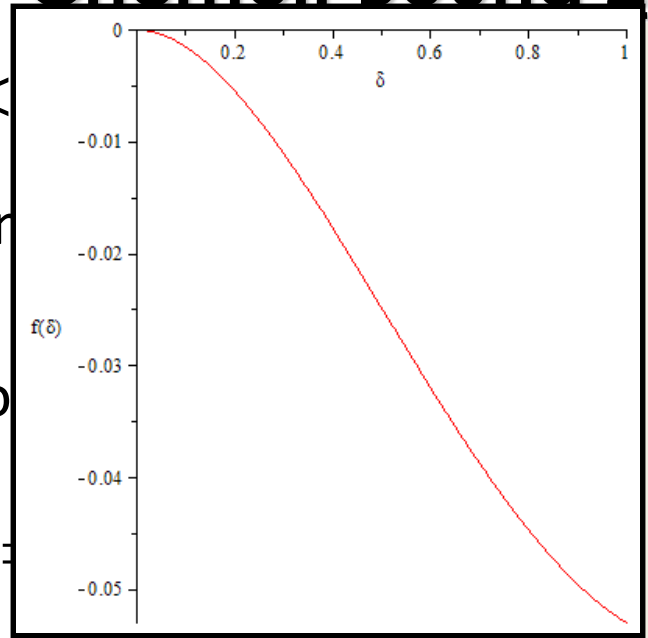Since $f'(0) = 0$ and $f'(1) < 0$, we have $f'(\delta) \leq 0$ over $[0,1]$.

Since $f(0) = 0$, it follows that $f(\delta) \leq 0$ in that interval.

Chernoff Bound 2

Theorem 4: For $0 <$ ... $] \leq e^{-\frac{\mu \delta^2}{3}}$.

Proof: From Theorem ... $+ \delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$.

We will show that fo... $e^{-\frac{\delta^2}{3}}$

... $\leq -\frac{\delta^2}{3}$

That is, $f(\delta) = \delta - (1 + \delta)\ln(1 + \delta) + \frac{\delta^2}{3} \leq 0$

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

Observe that $f''(\delta) < 0$ for $0 \leq \delta \leq \frac{1}{2}$, and $f''(\delta) > 0$ for $\delta > \frac{1}{2}$.

Hence, $f'(\delta)$ first decreases and then increases over $[0,1]$.

Since $f'(0) = 0$ and $f'(1) < 0$, we have $f'(\delta) \leq 0$ over $[0,1]$.

Since $f(0) = 0$, it follows that $f(\delta) \leq 0$ in that interval.

Chernoff Bound 2

Theorem 4: For $0 < \delta < 1$, $\Pr[X \geq (1 + \delta)\mu] \leq e^{-\frac{\mu\delta^2}{3}}$.

Proof: From Theorem 3, for $\delta > 0$, $\Pr[X \geq (1 + \delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$.

We will show that for $0 < \delta < 1$, $\frac{e^\delta}{(1+\delta)^{(1+\delta)}} \leq e^{-\frac{\delta^2}{3}}$

$$\Rightarrow \delta - (1 + \delta)\ln(1 + \delta) \leq -\frac{\delta^2}{3}$$

That is, $f(\delta) = \delta - (1 + \delta)\ln(1 + \delta) + \frac{\delta^2}{3} \leq 0$

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

Observe that $f''(\delta) < 0$ for $0 \leq \delta \leq \frac{1}{2}$, and $f''(\delta) > 0$ for $\delta > \frac{1}{2}$.

Hence, $f'(\delta)$ first decreases and then increases over $[0,1]$.

Since $f'(0) = 0$ and $f'(1) < 0$, we have $f'(\delta) \leq 0$ over $[0,1]$.

Since $f(0) = 0$, it follows that $f(\delta) \leq 0$ in that interval.

Chernoff Bound 3

Corollary 1: For $0 < \gamma < \mu, \Pr[X \geq \mu + \gamma] \leq e^{-\frac{\gamma^2}{3\mu}}$.

Proof: From Theorem 2, for $0 < \delta < 1$, $\Pr[X \geq (1 + \delta)\mu] < e^{-\frac{\mu\delta^2}{3}}$.

Setting $\gamma = \mu\delta$, we get, $\Pr[X \geq \mu + \gamma] \leq e^{-\frac{\gamma^2}{3\mu}}$ for $0 < \gamma < \mu$.

Example: n Fair Coin Flips

$$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is heads;} \\ 0 & \text{otherwise.} \end{cases}$$

Then the number of heads in n flips, $X = \sum_{i=1}^{n} X_i$.

We know, $E[X_i] = \Pr[X_i = 1] = \frac{1}{2}$.

Hence, $\mu = E[X] = \sum_{i=1}^{n} E[X_i] = \frac{n}{2}$.

Now putting $\delta = \frac{1}{2}$ in Chernoff bound 2, we have,

$$\Pr\left[X \geq \frac{3n}{4}\right] \leq e^{-\frac{n}{24}} = \frac{1}{e^{\frac{n}{24}}}.$$

Chernoff Bounds 4, 5 and 6

Theorem 5: For $0 < \delta < 1$, $\Pr[X \leq (1 - \delta)\mu] \leq \left(\dfrac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$.

Theorem 6: For $0 < \delta < 1$, $\Pr[X \leq (1 - \delta)\mu] \leq e^{-\frac{\mu\delta^2}{2}}$.

Corollary 2: For $0 < \gamma < \mu$, $\Pr[X \leq \mu - \gamma] \leq e^{-\frac{\gamma^2}{2\mu}}$.

Chernoff Bounds

Lower Tail	Upper Tail
$\mathbf{0 < \delta < 1}$: $\Pr[X \leq (1 - \delta)\mu] \leq \left(\dfrac{e^{-\delta}}{(1 - \delta)^{(1-\delta)}} \right)^{\mu}$	$\boldsymbol{\delta > 0}$: $\Pr[X \geq (1 + \delta)\mu] \leq \left(\dfrac{e^{\delta}}{(1 + \delta)^{(1+\delta)}} \right)^{\mu}$
$\mathbf{0 < \delta < 1}$: $\Pr[X \leq (1 - \delta)\mu] \leq e^{-\frac{\mu\delta^2}{2}}$	$\mathbf{0 < \delta < 1}$: $\Pr[X \geq (1 + \delta)\mu] \leq e^{-\frac{\mu\delta^2}{3}}$
$\boldsymbol{0 < \gamma < \mu}$: $\Pr[X \leq \mu - \gamma] \leq e^{-\frac{\gamma^2}{2\mu}}$	$\boldsymbol{0 < \gamma < \mu}$: $\Pr[X \geq \mu + \gamma] \leq e^{-\frac{\gamma^2}{3\mu}}$