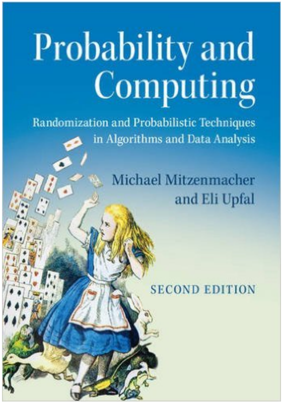
CS155/254: Probabilistic Methods in Computer Science

Chapter 4.1: Large Deviation Bounds



Large Deviation Bounds

A typical probability theory statement:

Theorem (The Central Limit Theorem)

Let $X_1, \ldots, X_n$ be independent identically distributed random variables with common mean μ and variance σ^2. Then

$$\lim_{n\to\infty} \Pr\left(\frac{\frac{1}{n}\sum_{i=1}^n X_i - \mu}{\sigma/\sqrt{n}} \le z\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt.$$

A typical CS probabilistic tool:

Theorem (Chernoff Bound)

Let $X_1, \ldots, X_n$ be independent Bernoulli random variables such that $Pr(X_i = 1) = p_i$. Let $\mu = \frac{1}{n}\sum_{i=1}^n p_i$, then

$$Pr\left(\frac{1}{n}\sum_{i=1}^{n} X_i \ge (1+\delta)\mu\right) \le e^{-\mu n \delta^2/3}.$$

Chernof Bound - Large Deviation Bound

Theorem

Let $X_1, \ldots, X_n$ *be independent,* $0-1$ *random variables with* $Pr(X_i = 1) = E[X_i] = p_i$. *Let* $\mu = \sum_{i=1}^{n} p_i$, *then for any* $\delta \in [0, 1]$ *we have*

$$Prob(\sum_{i=1}^{n} X_i \geq (1+\delta)\mu) \leq e^{-\mu\delta^2/3}$$

and

$$Prob(\sum_{i=1}^{n} X_i \leq (1-\delta)\mu) \leq e^{-\mu\delta^2/2}.$$

Consider n coin flips. Let X be the number of heads.
Markov Inequality gives

$$Pr\left(X \geq \frac{3n}{4}\right) \leq \frac{n/2}{3n/4} \leq \frac{2}{3}.$$

Using the Chebyshev's bound we have:

$$Pr\left(\left|X - \frac{n}{2}\right| \geq \frac{n}{4}\right) \leq \frac{4}{n}.$$

Using the Chernoff bound in this case, we obtain

$$\begin{aligned} Pr\left(\left|X - \frac{n}{2}\right| \geq \frac{n}{4}\right) &= Pr\left(X \geq \frac{n}{2}\left(1 + \frac{1}{2}\right)\right) \\ &+ Pr\left(X \leq \frac{n}{2}\left(1 - \frac{1}{2}\right)\right) \\ &\leq e^{-\frac{1}{3}\frac{n}{2}\frac{1}{4}} + e^{-\frac{1}{2}\frac{n}{2}\frac{1}{4}} \leq 2e^{-\frac{n}{24}}. \end{aligned}$$

The Basic Idea of Large Deviation Bounds:

For any random variable X, by Markov inequality we have:
For any $t > 0$,

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

Similarly, for any $t < 0$

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

Theorem (Markov Inequality)

If a random variable X is non-negative ($X \geq 0$) then

$$Prob(X \geq a) \leq \frac{E[X]}{a}.$$

The General Scheme:

For any random variable X:

1. computing $E[e^{tX}]$
2. optimize

$$Pr(X \geq a) \leq \min_{t>0} \frac{\mathrm{E}[e^{tX}]}{e^{ta}}$$
$$Pr(X \leq a) \leq \min_{t<0} \frac{\mathrm{E}[e^{tX}]}{e^{ta}}.$$

3. symplify

Moment Generating Function

Definition

The moment generating function of a random variable X is defined for any real value t as

$$M_X(t) = \mathrm{E}[e^{tX}].$$

Theorem

Let X be a random variable with moment generating function $M_X(t)$. Assuming that exchanging the expectation and differentiation operands is legitimate, then for all $n \geq 1$

$$\mathrm{E}[X^n] = M_X^{(n)}(0),$$

where $M_X^{(n)}(0)$ is the n-th derivative of $M_X(t)$ evaluated at $t = 0$.

Proof.

$$M_X^{(n)}(t) = \mathrm{E}[X^n e^{tX}].$$

Computed at $t = 0$ we get

$$M_X^{(n)}(0) = \mathrm{E}[X^n].$$

□

Why we can switch the order of the derivative and the expectation?

Assume for simplicity that X has integer values. Let $D(X)$ be the domain of X.

$$M_X(t) = E[e^{tX}] = \sum_{i \in D(X)} e^{ti} Pr(X = i).$$

For finite or uniformly convergent sum:

$$\begin{aligned} M_X^{(1)}(t) &= \frac{d}{dt} E[e^{tX}] = \frac{d}{dt} \left(\sum_{i \in D(X)} e^{ti} Pr(X = i) \right) \\ &= \sum_{i \in D(X)} \frac{d}{dt} e^{ti} Pr(X = i) = E[\frac{d}{dt} e^{ti}] \end{aligned}$$

Theorem

Let X and Y be two random variables. If

$$M_X(t) = M_Y(t)$$

for all $t \in (-\delta, \delta)$ for some $\delta > 0$, then X and Y have the same distribution.

Theorem

If X and Y are independent random variables then

$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

Proof.

$$M_{X+Y}(t) = \mathrm{E}[e^{t(X+Y)}] = \mathrm{E}[e^{tX}]\mathrm{E}[e^{tY}] = M_X(t)M_Y(t).$$

Chernoff Bound for Sum of Bernoulli Trials

Theorem

Let $X_1, \ldots, X_n$ be independent Bernoulli random variables such that $Pr(X_i = 1) = p_i$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \sum_{i=1}^{n} p_i$.

- *For any $\delta > 0$,*

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

- *For $0 < \delta \leq 1$,*

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

- *For $R \geq 6\mu$,*

$$Pr(X \geq R) \leq 2^{-R}. \tag{3}$$

Chernoff Bound for Sum of Bernoulli Trials

Let $X_1, \ldots, X_n$ be a sequence of independent Bernoulli trials with $Pr(X_i = 1) = p_i$. Let $X = \sum_{i=1}^{n} X_i$, and let

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

For each X_i:

$$\begin{aligned} M_{X_i}(t) &= \mathrm{E}[e^{tX_i}] \\ &= p_i e^t + (1 - p_i) \\ &= 1 + p_i(e^t - 1) \\ &\leq e^{p_i(e^t - 1)}. \end{aligned}$$

$$M_{X_i}(t) = \mathrm{E}[e^{tX_i}] \le e^{p_i(e^t-1)}.$$

Taking the product of the n generating functions we get for $X = \sum_{i=1}^{n} X_i$

$$\begin{aligned} M_X(t) &= \prod_{i=1}^{n} M_{X_i}(t) \\ &\le \prod_{i=1}^{n} e^{p_i(e^t-1)} \\ &= e^{\sum_{i=1}^{n} p_i(e^t-1)} \\ &= e^{(e^t-1)\mu} \end{aligned}$$

$$M_X(t) = \mathrm{E}[e^{tX}] = e^{(e^t-1)\mu}$$

Applying Markov's inequality we have for any $t > 0$

$$\begin{aligned} Pr(X \geq (1+\delta)\mu) &= Pr(e^{tX} \geq e^{t(1+\delta)\mu}) \\ &\leq \frac{\mathrm{E}[e^{tX}]}{e^{t(1+\delta)\mu}} \\ &\leq \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} \end{aligned}$$

For any $\delta > 0$, we can set $t = \ln(1+\delta) > 0$ to get:

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

This proves (1).

We show that for $0 < \delta < 1$,

$$\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \leq e^{-\delta^2/3}$$

or that $f(\delta) = \delta - (1+\delta)\ln(1+\delta) + \delta^2/3 \leq 0$
in that interval. Computing the derivatives of $f(\delta)$ we get

$$\begin{aligned} f'(\delta) &= 1 - \frac{1+\delta}{1+\delta} - \ln(1+\delta) + \frac{2}{3}\delta = -\ln(1+\delta) + \frac{2}{3}\delta, \\ f''(\delta) &= -\frac{1}{1+\delta} + \frac{2}{3}. \end{aligned}$$

$f''(\delta) < 0$ for $0 \leq \delta < 1/2$, and $f''(\delta) > 0$ for $\delta > 1/2$.
$f'(\delta)$ first decreases and then increases over the interval $[0, 1]$.
Since $f'(0) = 0$ and $f'(1) < 0$, $f'(\delta) \leq 0$ in the interval $[0, 1]$.
Since $f(0) = 0$, we have that $f(\delta) \leq 0$ in that interval.
This proves (2).

For $R \geq 6\mu$, $\delta \geq 5$.

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

that proves (3).

Theorem

Let $X_1, \ldots, X_n$ *be independent Bernoulli random variables such that* $Pr(X_i = 1) = p_i$. *Let* $X = \sum_{i=1}^{n} X_i$ *and* $\mu = \mathrm{E}[X]$.
For $0 < \delta < 1$:

- $$Pr(X \leq (1-\delta)\mu) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}. \quad (4)$$

- $$Pr(X \leq (1-\delta)\mu) \leq e^{-\mu\delta^2/2}. \quad (5)$$

Using Markov's inequality, for any $t < 0$,

$$\begin{aligned} Pr(X \le (1-\delta)\mu) &= Pr(e^{tX} \ge e^{(1-\delta)t\mu}) \\ &\le \frac{\mathrm{E}[e^{tX}]}{e^{t(1-\delta)\mu}} \\ &\le \frac{e^{(e^t-1)\mu}}{e^{t(1-\delta)\mu}} \end{aligned}$$

For $0 < \delta < 1$, we set $t = \ln(1-\delta) < 0$ to get:

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

This proves (4).
We need to show:

$$f(\delta) = -\delta - (1-\delta)\ln(1-\delta) + \frac{1}{2}\delta^2 \le 0.$$

We need to show:

$$f(\delta) = -\delta - (1-\delta)\ln(1-\delta) + \frac{1}{2}\delta^2 \leq 0.$$

Differentiating $f(\delta)$ we get

$$\begin{aligned} f'(\delta) &= \ln(1-\delta) + \delta, \\ f''(\delta) &= -\frac{1}{1-\delta} + 1. \end{aligned}$$

Since $f''(\delta) < 0$ for $\delta \in (0,1)$, $f'(\delta)$ decreasing in that interval. Since $f'(0) = 0$, $f'(\delta) \leq 0$ for $\delta \in (0,1)$. Therefore $f(\delta)$ is non increasing in that interval.
$f(0) = 0$. Since $f(\delta)$ is non increasing for $\delta \in [0,1)$, $f(\delta) \leq 0$ in that interval, and (5) follows.

Example: Coin flips

Theorem (The Central Limit Theorem)

Let $X_1, \dots, X_n$ be independent identically distributed random variables with common mean μ and variance σ^2. Then

$$\lim_{n\to\infty} \Pr\left(\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma/\sqrt{n}} \le z\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt.$$

$\Phi(2.23) = 0.99$, thus, $\lim_{n\to\infty} Pr\left(\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma/\sqrt{n}} \le 2.23\right) = 0.99$

For coin flips:

$\lim_{n\to\infty} Pr\left(\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - 1/2}{1/(2\sqrt{n})} \le 2.23\right) = 0.99$

$\lim_{n\to\infty} Pr\left(\sum_{i=1}^{n} X_i - \frac{n}{2} \ge 2.23\sqrt{n}/2\right) = 0.01$

$\Phi(3.5) \approx 0.999$, $\lim_{n\to\infty} Pr\left(\sum_{i=1}^{n} X_i - \frac{n}{2} \ge 3.5\sqrt{n}/2\right) = 0.001$

Example: Coin flips

Let X be the number of heads in a sequence of n independent fair coin flips.

$$Pr\left(\left|X - \frac{n}{2}\right| \geq \frac{1}{2}\sqrt{6n \ln n}\right)$$

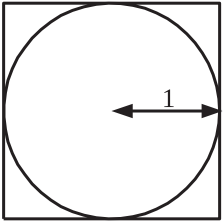
$$= Pr\left(X \geq \frac{n}{2}\left(1 + \sqrt{\frac{6 \ln n}{n}}\right)\right)$$

$$+ Pr\left(X \leq \frac{n}{2}\left(1 - \sqrt{\frac{6 \ln n}{n}}\right)\right)$$

$$\leq e^{-\frac{1}{3}\frac{n}{2}\frac{6 \ln n}{n}} + e^{-\frac{1}{2}\frac{n}{2}\frac{6 \ln n}{n}} \leq \frac{2}{n}.$$

Note that the standard deviation is $\sqrt{n/4}$

Example: estimate the value of π



- Choose X and Y independently and uniformly at random in $[0, 1]$.
- Let
$$Z = \begin{cases} 1 & \text{if } \sqrt{X^2 + Y^2} \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$
- $\frac{1}{2} \leq p = \Pr(Z = 1) = \frac{\pi}{4} \leq 1.$
- $4\mathrm{E}[Z] = \pi$.

- Let $Z_1, \dots, Z_m$ be the values of m independent experiments. $W_m = \sum_{i=1}^{m} Z_i$.
-

$$\mathrm{E}[W_m] = \mathrm{E}\left[\sum_{i=1}^{m} Z_i\right] = \sum_{i=1}^{m} \mathrm{E}[Z_i] = \frac{m\pi}{4},$$

- $\tilde{\pi}_m = \frac{4}{m} W_m$ is an unbiased estimate for π (i.e. $E[\tilde{\pi}_m] = \pi$)
- How many samples do we need to obtain a good estimate?

$$\begin{aligned}
\Pr(|\tilde{\pi}_m - \pi| \geq \epsilon\pi) &= \Pr\left(|W - \frac{m\pi}{4}| \geq \frac{\epsilon m\pi}{4}\right) \\
&= \Pr\left(|W_m - \mathrm{E}[W_m]| \geq \epsilon \mathrm{E}[W_m]\right) \\
= \Pr\left(W_m - \mathrm{E}[W_m] \geq \epsilon \mathrm{E}[W_m]\right) &+ \Pr\left(W_m - \mathrm{E}[W_m] \leq \epsilon \mathrm{E}[W_m]\right) \\
\leq \mathrm{e}^{-\frac{1}{3}\frac{m\pi}{4}\epsilon^2} &+ \mathrm{e}^{-\frac{1}{2}\frac{m\pi}{4}\epsilon^2} \leq 2\mathrm{e}^{-\frac{1}{12}m\pi\epsilon^2}.
\end{aligned}$$

Since it's easy to verify that $\pi > 2$

$$\Pr(|\tilde{\pi}_m - \pi| \geq \epsilon\pi) \leq 2\mathrm{e}^{-\frac{1}{12}m\pi\epsilon^2} \leq \mathrm{e}^{-\frac{1}{6}m\epsilon^2} = \delta$$

For $\epsilon = 0.1$ and $\delta = 0.01$ we need $m \geq 4000$.

Set Balancing

Given an $n \times n$ matrix $\mathcal{A}$ with entries in $\{0, 1\}$, let

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ \dots \\ b_n \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \dots \\ \dots \\ c_n \end{pmatrix}.$$

Find a vector $\bar{b}$ with entries in $\{-1, 1\}$ that minimizes

$$||\mathcal{A}\bar{b}||_\infty = \max_{i=1,\dots,n} |c_i|.$$

Theorem

For a random vector $\bar{b}$, with entries chosen independently and with equal probability from the set $\{-1, 1\}$,

$$Pr(||\mathcal{A}\bar{b}||_\infty \geq \sqrt{4n \ln n}) \leq \frac{2}{n}.$$

The $\sum_{i=1}^{n} a_{j,i} b_i$ (excluding the zero terms) is a sum of independent $-1, 1$ random variable. We need a bound on such sum.

Chernoff Bound for Sum of $\{-1, +1\}$ Random Variables

Theorem

Let $X_1, ..., X_n$ be independent random variables with

$$Pr(X_i = 1) = Pr(X_i = -1) = \frac{1}{2}.$$

Let $X = \sum_1^n X_i$. For any $a > 0$,

$$Pr(X \geq a) \leq e^{-\frac{a^2}{2n}}.$$

de Moivre – Laplace approximation: For any k, such that $|k - np| \leq a$

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{a^2}{2np(1-p)}}$$

For any $t > 0$,

$$E[e^{tX_i}] = \frac{1}{2}e^t + \frac{1}{2}e^{-t}.$$

$$e^t = 1 + t + \frac{t^2}{2!} + \cdots + \frac{t^i}{i!} + \ldots$$

and

$$e^{-t} = 1 - t + \frac{t^2}{2!} + \cdots + (-1)^i \frac{t^i}{i!} + \ldots$$

Thus,

$$\begin{aligned} E[e^{tX_i}] &= \frac{1}{2}e^t + \frac{1}{2}e^{-t} = \sum_{i \geq 0} \frac{t^{2i}}{(2i)!} \\ &\leq \sum_{i \geq 0} \frac{(\frac{t^2}{2})^i}{i!} = e^{t^2/2} \end{aligned}$$

$$\mathrm{E}[e^{tX}] = \prod_{i=1}^{n} \mathrm{E}[e^{tX_i}] \leq e^{nt^2/2},$$

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

Setting $t = a/n$ yields

$$Pr(X \geq a) \leq e^{-\frac{a^2}{2n}}.$$

By symmetry we also have

Corollary

Let $X_1, \ldots, X_n$ be independent random variables with

$$Pr(X_i = 1) = Pr(X_i = -1) = \frac{1}{2}.$$

Let $X = \sum_{i=1}^{n} X_i$. Then for any $a > 0$,

$$Pr(|X| > a) \leq 2e^{-\frac{a^2}{2n}}.$$

Application: Set Balancing

Theorem

For a random vector $\bar{b}$, with entries chosen independently and with equal probability from the set $\{-1, 1\}$,

$$Pr(||\mathcal{A}\bar{b}||_\infty \geq \sqrt{4n \ln n}) \leq \frac{2}{n} \qquad (6)$$

- Consider the i-th row $\bar{a}_i = a_{i,1}, ..., a_{i,n}$.
- Let k be the number of 1's in that row.
- $Z_i = \sum_{j=1}^{k} a_{i,i_j} b_{i_j}$.
- If $k \leq \sqrt{4n \ln n}$ then clearly $Z_i \leq \sqrt{4n \ln n}$.

If $k > \sqrt{4n \log n}$, the k non-zero terms in the sum Z_i are independent random variables, each with probability $1/2$ of being either $+1$ or -1.
Using the Chernoff bound:

$$Pr\left\{|Z_i| > \sqrt{4n \log n}\right\} \le 2e^{-4n \log n/(2k)} \le 2e^{-4n \log n/(2n)} \le \frac{2}{n^2},$$

where we use the fact that $n \ge k$.
The result follows by union bound on the n rows.

Hoeffding's Inequality

Large deviation bound for more general random variables:

Theorem (Hoeffding's Inequality)

Let $X_1, \ldots, X_n$ *be independent random variables such that for all* $1 \leq i \leq n$, $E[X_i] = \mu$ *and* $Pr(a \leq X_i \leq b) = 1$. *Then*

$$Pr(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| \geq \epsilon) \leq 2e^{-2n\epsilon^2/(b-a)^2}$$

Lemma

(Hoeffding's Lemma) Let X *be a random variable such that* $Pr(X \in [a, b]) = 1$ *and* $E[X] = 0$. *Then for every* $\lambda > 0$,

$$\mathrm{E}[E^{\lambda X}] \leq e^{\lambda^2 (a-b)^2}/8.$$

Proof of the Lemma

Since $f(x) = e^{\lambda x}$ is a convex function, for any $\alpha \in (0,1)$ and $x \in [a,b]$,

$$f(X) \leq \alpha f(a) + (1-\alpha) f(b).$$

Thus, for $\alpha = \frac{b-x}{b-a} \in (0,1)$,

$$e^{\lambda x} \leq \frac{b-x}{b-a} e^{\lambda a} + \frac{x-a}{b-a} e^{\lambda b}.$$

Taking expectation, and using $\mathrm{E}[X] = 0$, we have

$$E[e^{\lambda X}] \leq \frac{b}{b-a} e^{\lambda a} + \frac{a}{b-a} e^{\lambda b} \leq e^{\lambda^2 (b-a)^2/8}.$$

Proof of the Bound

Let $Z_i = X_i - \mathrm{E}[X_i]$ and $Z = \frac{1}{n}\sum_{i=1}^{n} X_i$.

$$Pr(Z \geq \epsilon) \leq e^{-\lambda\epsilon}\mathrm{E}[e^{\lambda Z}] \leq e^{-\lambda\epsilon}\prod_{i=1}^{n}\mathrm{E}[e^{\lambda X_i/n}] \leq e^{-\lambda\epsilon + \frac{\lambda^2(b-a)^2}{8n}}$$

Set $\lambda = \frac{4n\epsilon}{(b-a)^2}$ gives

$$Pr(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| \geq \epsilon) = Pr(Z \geq \epsilon) \leq 2e^{-2n\epsilon^2/(b-a)^2}$$

A More General Version

Theorem

Let $X_1, \ldots, X_n$ be independent random variables with $\mathrm{E}[X_i] = \mu_i$ and $Pr(B_i \leq X_i \leq B_i + c_i) = 1$, then

$$Pr(|\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i| \geq \epsilon) \leq 2e^{-\frac{2\epsilon^2}{\sum_{i=1}^{n} c_i^2}}$$

Application: Job Completion

We have n jobs, job i has expected run-time μ_i. We terminate job i if it runs $\beta\mu_i$ time. When will the machine will be free of jobs?
$X_i =$ execution time of job i. $0 \leq X_i \leq \beta\mu_i$.

$$Pr(|\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i| \geq \epsilon \sum_{i=1}^{n} \mu_i) \leq 2e^{-\frac{2\epsilon^2(\sum_{i=1}^{n} \mu_i)^2}{\sum_{i=1}^{n} \beta^2\mu_i^2}}$$

Assume all $\mu_i = \mu$

$$Pr(|\sum_{i=1}^{n} X_i - n\mu| \geq \epsilon n\mu) \leq 2e^{-\frac{2\epsilon^2 n^2 \mu^2}{n\beta^2\mu^2}} = 2e^{-2\epsilon^2 n/\beta^2}$$

Let $\epsilon = \beta\sqrt{\frac{\log n}{n}}$, then

$$Pr(|\sum_{i=1}^{n} X_i - n\mu| \geq \beta\mu\sqrt{n \log n}) \leq 2e^{-\frac{2\beta^2\mu^2 n \log n}{n\beta^2\mu^2}} = \frac{2}{n^2}$$