Example

Theorem 1. *Consider* n *coin flips, let* X *be the number of heads,*

$$\Pr(|X - \frac{n}{2}| > \frac{1}{2}\sqrt{6n\log n}) \leq \frac{2}{n}$$

Proof.

$$E[X] = n/2$$

We need

$$\frac{n}{2} - \frac{1}{2}\sqrt{6n\log n} \leq X \leq \frac{n}{2} + \frac{1}{2}\sqrt{6n\log n}$$

or

$$X = \frac{n}{2}(1 \pm \sqrt{\frac{6\log n}{n}})$$

Fixing $\delta = \sqrt{\frac{6\log n}{n}}$

$$\Pr(X < (1-\delta)n/2) \leq e^{-\frac{n}{2}\frac{\delta^2}{2}} \leq 1/n$$

$$\Pr(X > (1+\delta)n/2) \le e^{-\frac{n}{2}\frac{\delta^2}{3}} \le 1/n$$

□

What is the probability of more than $\frac{3N}{4}$ heads in N coin flips?

1. Using Markov Inequality:

$$Pr(X \geq 3N/4) \leq 2/3.$$

2. Using Chebyshev's Inequality:

$$Pr(X \geq 3N/4) \leq 4/N.$$

3. Using the Chernoff bound:

$$Pr(X \geq 3N/4) \leq e^{-\frac{N}{2}\frac{1}{4}\frac{1}{3}}.$$

Randomized Quicksort Revisited

View an execution of the randomized quicksort algorithm as the following binary tree of (sub-)problems.

- The root of this tree is the (initial) problem of sorting a given set of n distinct numbers; and a leaf is a subproblem of sorting a singleton set.
- An internal node of this tree is a subproblem of sorting a set S (of size greater than 1).
- Its left child (if any) is the subproblem of sorting a set $S_1 \subset S$ consisting of elements smaller than the pivot.
- Its right child (if any) is the subproblem of sorting a set $S_2 \subset S$ consisting of elements larger than the pivot.

Thus, a run of a quicksort algorithm is described by the above *execution tree*.

High Probability Analysis

Theorem 2. *Randomized Quicksort runs in* $O(n \log n)$ *time with high probability, i.e., with probability at least* $1 - 1/n^b$*, for some constant* $b > 0$.

Proof. Suppose the size of the set to be sorted at a particular node is S. A node in the execution tree is labeled **good** if the pivot element divides the set into two parts, each of size not exceeding $2S/3$. Otherwise the node is called **bad**.

Then we can show that:

1. The probability of a node being labeled good is $1/3$.
2. The number of good nodes in any root to leaf path is bounded by $\log_{3/2} n < c \log n$ for some constant c.

What is the probability that a path of length $ac\log n$ (for some constant $a > 1$) will have at most $c\log n$ good nodes?

The mean $\mu = 1/3(ac\log n)$. Using the Chernoff bound

$$\Pr(X < c\log n) = \Pr(X < (1 - (1 - \frac{3}{a}))\mu)$$

$$\leq e^{-\mu(1-3/a)^2(1/2)} \leq 1/n^2$$

for a suitably large constant a.

Thus with probability at most $1/n$ no root to leaf path has length more than $ac\log n$. Since the total work done at each level of the tree is $O(n)$, the running time is bounded by $O(n\log n)$ with high probability. □

Application: 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 3. *For a random vector $\bar{b}$, with entries chosen independently and with equal probability from the set $\{-1, 1\}$,*

$$Pr(||\mathcal{A}\bar{b}||_\infty \geq \sqrt{12n \ln n}) \leq \frac{2}{n}.$$

Consider the i-th row $\bar{a_i} = a_{i,1}, \ldots, a_{i,n}$. Let k be the number of 1's in that row.

If $k \le \sqrt{12n \ln n}$ clearly $|\bar{a_i} \cdot \bar{b}| \le \sqrt{12n \ln n}$.

If $k > \sqrt{12n \ln n}$, let

$$X_i = |\{j \mid a_{i,j} = 1 \text{ and } b_j = 1\}|$$

and

$$Y_i = |\{j \mid a_{i,j} = 1 \text{ and } b_j = -1\}|.$$

Thus, X_i counts the number of $+1$'s in the sum $\sum_{j=1}^{n} a_{i,j} b_j$, and Y_j counts the number of -1's. Using Chernoff bounds,

$$Pr\left(X_i \geq \frac{k}{2}\left(1+\sqrt{\frac{12\ln n}{k}}\right)\right) \leq e^{-(\frac{k}{2})(\frac{1}{3})(\frac{12\ln n}{k})} \leq e^{-2\ln n}$$

Hence, for a given row,

$$Pr(|X_i - Y_i| \geq \sqrt{12n\ln n}) \leq \frac{2}{n^2}$$

Since there are n rows, the probability that any row exceeds that bound is bounded by $\frac{2}{n}$.