

Probabilistic Algorithms

Another approach to deal with NP-hard problems: Under a "reasonable probabilistic model" design algorithms which give either exact solutions or good approximate solutions in polynomial time.

The performance of an algorithm is with respect to the probability distribution of the input.

Measures of interest : Expected cost/time, high probability guarantee, asymptotic convergence.

Chernoff Bound

Theorem 1. Let X_1, X_2, \dots, X_n be independent indicator random variables such that, for $1 \leq i \leq n$, $\Pr[X_i = 1] = p_i$, where $0 < p_i < 1$. Then, for $X = \sum_{i=1}^n X_i$, $\mu = E[X] = \sum_{i=1}^n p_i$, and any $\delta > 0$,

$$\Pr(X > (1 + \delta)\mu) < \left(\frac{e^\delta}{(1 + \delta)^{(1+\delta)}} \right)^\mu$$

For $0 < \delta < 1$,

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

For $0 < \delta < 1$,

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

Proof. Upper tail: For any positive real t ,

$$\Pr(X > (1 + \delta)\mu) = \Pr(e^{tX} > e^{t(1+\delta)\mu})$$

By Markov's inequality,

$$\begin{aligned} \Pr(X > (1 + \delta)\mu) &< \frac{E[e^{tX}]}{e^{t(1+\delta)\mu}} \\ &= \frac{E[e^{t \sum_{i=1}^n X_i}]}{e^{t(1+\delta)\mu}} = \frac{E[\prod_{i=1}^n e^{tX_i}]}{e^{t(1+\delta)\mu}} = \frac{\prod_{i=1}^n E[e^{tX_i}]}{e^{t(1+\delta)\mu}} \\ &= \frac{\prod_{i=1}^n (p_i e^t + 1 - p_i)}{e^{t(1+\delta)\mu}} = \frac{\prod_{i=1}^n (1 + p_i(e^t - 1))}{e^{t(1+\delta)\mu}} \\ &< \frac{\prod_{i=1}^n e^{p_i(e^t - 1)}}{e^{t(1+\delta)\mu}} = \frac{e^{\sum_{i=1}^n p_i(e^t - 1)}}{e^{t(1+\delta)\mu}} = \frac{e^{(e^t - 1)\mu}}{e^{t(1+\delta)\mu}} \\ &\leq \left(\frac{e^\delta}{(1 + \delta)(1 + \delta)} \right)^\mu \end{aligned}$$

for $t = \ln(1 + \delta)$

Using $\delta - (1 + \delta) \ln(1 + \delta) \leq -\delta^2/3$ for $0 < \delta < 1$ we get

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

Lower tail:

$$\Pr(X < (1 - \delta)\mu) = \Pr(e^{-tX} > e^{-t(1-\delta)\mu})$$

By Markov's inequality,

$$\Pr(X < (1 - \delta)\mu) < \frac{E[e^{-tX}]}{e^{-t(1-\delta)\mu}}$$

Similar calculations yield

$$< \frac{e^{(e^{-t}-1)\mu}}{e^{-t(1-\delta)\mu}}$$

For $t = \ln(1/(1 - \delta))$

$$\leq \left(\frac{e^{-\delta}}{(1 - \delta)(1-\delta)} \right)^\mu$$

Since $(1 - \delta)^{(1-\delta)} > e^{-\delta+\delta^2/2}$ we have

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

□

Example

Theorem 2. 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 $\delta = \sqrt{\frac{6 \log n}{n}}$

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

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

□

Randomized Quicksort Revisited

Theorem 3. *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 $b > 1$.*

Proof. Recall the Midterm problem on Quicksort.

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**.

We showed:

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

$$\begin{aligned}\Pr(X < c \log n) &= \Pr(X < (1 - (1 - \frac{3}{a}))\mu) \\ &\leq e^{-\mu(1-3/a)^2(1/2)} \leq 1/n^2\end{aligned}$$

for a suitably large constant a .

Thus with probability at least $1 - 1/n^2$ the longest path is at most $ac \log n$. 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.

□

Bin packing

We are given n objects x_1, x_2, \dots, x_n : the size of each object is uniformly distributed over $[0, 1]$. We want to pack in unit sized bins, such that the total number of bins is minimized.

The expected number of bins required is at least $E(\sum_{j=1}^n x_j) = n/2$.

Consider the following algorithm:

$$\text{Let } \alpha = 1 - \frac{6 \log n}{\sqrt{n}}$$

1. Place each element $x_i \geq \alpha$ into a bin on its own. Suppose there are B_1 such elements.
2. Let $N = n - B_1$ be the number of elements yet to be packed.
3. Order the items so that $x_1 \leq x_2 \leq \dots \leq x_N \leq \alpha$.
4. **For** $i = 1, 2, \dots, \lfloor N/2 \rfloor$

- 4.1 **if** $x_i + x_{N-i+1} \leq 1$ put x_i, x_{N-i+1} into one bin
4.2 **else** put into separate bins.

Put item $\lceil N/2 \rceil$ into a separate bin if N is odd.

Theorem 4. For n sufficiently large, the expected number of bins packed by the algorithm is at most $\frac{n}{2} + 7 \log n \sqrt{n}$.

Proof. Each item has size greater than α with probability $6 \log n / \sqrt{n}$, so $E[B_1] = 6 \log n \sqrt{n}$.

We show that for $i = 1, 2, \dots, \lfloor N/2 \rfloor$:

$$\Pr(x_i + x_{N+i-1} > 1) \leq 1/n$$

Then the expected number of bins used in the for loop is less than $\frac{N}{2} + 2$ and the theorem follows.

To show the above bound use Chernoff bound to show

$$\Pr(x_i > \frac{i + 3 \log n \sqrt{n}}{n}) \leq \frac{1}{2n}$$

$$\Pr(x_{N-i+1} > \frac{n - i - 3 \log n \sqrt{n}}{n}) \leq \frac{1}{2n}$$

Use the fact that $x_i > p = \frac{i + 3 \log n \sqrt{n}}{n}$ if and only if there are at most i items of size less than p . \square