Binary Interpolation Search Complexity Analysis
Design and Analysis of Algorithms

Niraj Kumar (CS2011)

March 2021

0.1 Binary Interpolation Search Analysis

Theorem 0.1. *Suppose the keys in the interpolation search are drawn from a uniform distribution independently then the expected number of probes C used in Binary Interpolation to determine the next subspace(to be searched) is bounded by a constant.*

Proof. Let the initial size of the search space be n. Now the keys x_i are independently drawn from a uniform distribution, the indicator that a particular key is less than or equal to y is Bernoulli distributed with probability $p = (y - A[0])/(A[n-1] - A[0])$.

Hence, the number of keys X, which are less or equal to y is binomially distributed as well. Notice that X acts as a random variable here.

Now as per binomial distribution, the probability of exactly k keys less than or equal to y is $\binom{n}{k} p^k (1-p)^{n-k}$. Because the distribution of X is binomial its expected value is $\mu = np$ and the variance will be $np(1-p)$.

Now, C can be calculated as follows

$C = \sum_{k=1}^{\infty} kP[\text{exactly } k \text{ probes are used}] = \sum_{k=1}^{\infty} P[\text{atleast } k \text{ probes are used}]$

Now, at least two probes are always used, first one for mid and second with $mid + \lceil \sqrt{n} \rceil$ or $mid - \lceil \sqrt{n} \rceil$ depending on mid. And for $k \geq 3$,

$P[\text{atleast } k \text{ probes are used}] \leq P[(y's \text{ index}) - np] \geq (k-2)\lceil \sqrt{n} \rceil$

Now we can apply Chebyshev's inequality to upper bound this probability by $\frac{\sigma^2}{((k-2)\lceil \sqrt{n} \rceil)^2}$

$$P[|X - \mu| \geq t] \leq \frac{\sigma^2}{t^2}$$

for the random variable X with μ being the mean and σ^2 being the variance.

Also, $\sigma^2 = p(1-p)n \leq \frac{1}{4}n$, and hence

$$\begin{aligned} C &\leq 2 + \sum_{k=3}^{\infty} \frac{\sigma^2}{((k-2)\lceil \sqrt{n} \rceil)^2} \\ &\leq 2 + \sum_{k=3}^{\infty} \frac{1}{4} \frac{n}{((k-2)\lceil \sqrt{n} \rceil)^2} \\ &\leq 2 + \sum_{k=3}^{\infty} \frac{1}{4} \frac{1}{(k-2)^2} \\ &= 2 + \frac{1}{4} \sum_{k=3}^{\infty} \frac{1}{k^2} \\ &\leq 2.42 \end{aligned}$$

□

0.1.1 Average Case Analysis

Let $T(n)$ be the average number of probes required for finding a particular key in the array of size n. To find this we must first calculate the expected number of probes C, which are used by the interpolation search to reduce the search space of size n to $\sqrt{n}$.
Also, C is bounded by a constant C'. Hence,

$$T(n) \leq C' + T(\sqrt{n})$$

Now assuming $n = z^{2^k}$ for some $k, z \in N$ and that $T(z)$ is small. Then,

$$T(n) = T(z^{2^k}) \leq C' + T(z^{2^{k-1}}) \leq C' + C' + T(z^{2^{k-2}}) \leq C' + C' + C'(ktimes) + T(z) = kC' + T(z)$$

Also, as $n = z^{2^k}$ and $z \geq 2$ if $n > 1$ we have $\log_2(z) \geq 1$ and hence,

$$k = \log_2(\log_2(n)) = \frac{\log_2(\log_2(n))}{\log_2(z)} \leq \log_2(\log_2(n))$$

$$\implies T(n) \leq C' \times \log(\log(n)) + T(z)$$

$$T(n) \leq 2.42 \log(\log n)$$

Hence, the average case complexity of searching using binary Interpolation search is $\Theta(\log \log n)$ assuming the distribution of keys to be uniform.

0.1.2 Worst Case Analysis

Consider the case when the list is exponentially distributed instead of a uniform distribution, then the algorithm will access each element in linear fashion as the index calculation formula will be incremented by almost 1 at each step.

$$index = left + \frac{(right - left)(x - A[left])}{(A[right] - A[left])}$$

$$left + \frac{(right - left)(x - c^a)}{c^b - c^a}$$

for some a, b, c.

Now,

$$left + \frac{(right - left)(x - c^a)}{(c^b - c^a)} \sim left + \frac{(x - c^a)}{(c^b - c^a)} \sim left + 1$$

Hence, number of comparisons is $O(n)$.