

Quicksort

Andreas Klappenecker

Texas A&M University

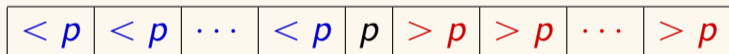
© 2018–2019 by Andreas Klappenecker. All rights reserved.

Quicksort

Suppose that we want to sort an array $A[1..n]$ of length n .

Quicksort picks a **pivot** element p uniformly at random.

Then partitions the array A into three parts: **left**, **pivot**, and **right**.



Partition requires $n - 1$ comparisons with the pivot element p .

Then quicksort recursively sorts left and right parts.

Proposition

The expected number of comparisons made by randomized quicksort on an array of size n is at most $2n \ln n$.

Expected Number of Comparisons

Let P denote the random variable giving the sorting order of the pivot element p . Thus, $P = k$ means that the pivot element is the k -th smallest element of the array.

Let X_n denote the number of comparison done by quicksort on an array of length n . Sorting an array of length n yields the expected number $E[X_n]$ of comparisons

$$E[X_n] = \sum_{k=1}^n E[X_n \mid P = k] \Pr[X = k].$$

Expected Number of Comparisons

Probability that Pivot is k -th Smallest Element

Since the pivot is chosen uniformly at random, we have

$$\Pr[P = k] = \frac{1}{n}.$$

Expected Number of Comparisons

Let $E[X_n]$ denote the expected number of comparisons for an array of length n . Then

$$E[X_n \mid P = k] = (n - 1) + E[X_{k-1}] + E[X_{n-k}],$$

since we need $n - 1$ comparisons with the pivot. If the pivot is the k -th smallest element, then the left partition has $k - 1$ elements, and the right partition has $n - k$ elements.

Expected Number of Comparisons

Let $T(n) = E[X_n]$ denote the expected number of comparisons for arrays of length n .

$$\begin{aligned} T(n) = E[X_n] &= \sum_{k=1}^n (n - 1 + E[X_{k-1}] + E[X_{n-k}]) \Pr[P = k] \\ &= \sum_{k=1}^n (n - 1 + T(k - 1) + T(n - k)) \frac{1}{n} \\ &= n - 1 + \frac{1}{n} \sum_{k=1}^n (T(k - 1) + T(n - k)). \end{aligned}$$

Expected Number of Comparisons

Let $T(n)$ denote the expected number of comparisons in quicksort for arrays of length n .

$$T(n) = \begin{cases} n - 1 + \frac{2}{n} \sum_{k=1}^{n-1} T(k) & \text{if } n > 0, \\ 0 & \text{if } n = 0. \end{cases}$$

Our guess is that $T(n) \leq cn \ln n$, since most pivots lead to splits that are not too imbalanced. It turns out that we can choose $c = 2$.

Proposition

$$T(n) \leq 2n \ln n.$$

Proof.

Basis. The inequality holds for $n = 0$, since $T(0) = 0$ and $\lim_{x \rightarrow 0} x \ln x = 0$, so $n \ln n = 0$ for $n = 0$.

Inductive Step. We assume that $T(k) \leq 2k \ln k$ holds for all k in the range $0 \leq k < n$. We need to show that this implies

$$T(n) \leq 2n \ln n.$$

$$\begin{aligned}T(n) &= n - 1 + \frac{2}{n} \sum_{k=1}^{n-1} T(k) \\&\leq n - 1 + \frac{2}{n} \sum_{k=1}^{n-1} 2k \ln k \\&\leq n - 1 + \frac{2}{n} \int_1^n x \ln x \, dx \\&= n - 1 + \frac{2}{n} \left(n^2 \ln n - \frac{n^2}{2} + \frac{1}{2} \right) \leq 2n \ln n. \quad \square\end{aligned}$$

Since $2x \ln x$ is monotonically increasing on $[1, n]$, we are allowed to bound the sum by the integral

$$\begin{aligned} \sum_{k=1}^{n-1} 2k \ln k &\leq \sum_{k=1}^{n-1} \int_k^{k+1} 2x \ln x \, dx \\ &= \int_1^n 2x \ln x \, dx \\ &= \left(x^2 \ln x - \frac{x^2}{2} \right) \Big|_1^n \end{aligned}$$