Quicksort

Andreas Klappenecker

Texas A&M University

© 2018 by Andreas Klappenecker. All rights reserved.
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* $2n \ln n$. 
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

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

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

\[
T(n) = E[X_n] = \sum_{k=1}^{n} \left( n - 1 + E[X_{k-1}] + E[X_{n-k}] \right) \Pr[P = k]
\]

\[
= \sum_{k=1}^{n} \left( n - 1 + T(k-1) + T(n-k) \right) \frac{1}{n}
\]

\[
= n - 1 + \frac{1}{n} \sum_{k=1}^{n} (T(k-1) + T(n-k)).
\]
Let $T(n)$ denote the expected number of comparisons in quicksort for arrays of length $n$.

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

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 $c = 2$. 
Proof by Induction

Proposition

\[ T(n) \leq 2n \ln n. \]

Proof.

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

\[ T(n) \leq 2n \ln n. \]
Proof by Induction

\[ T(n) = n - 1 + \frac{2}{n} \sum_{k=1}^{n-1} T(k) \]

\[ \leq n - 1 + \frac{2}{n} \sum_{k=1}^{n-1} 2k \ln k \]

\[ \leq n - 1 + \frac{2}{n} \int_{1}^{n} x \ln x \, dx \]

\[ = n - 1 + \frac{2}{n} \left( n^2 \ln n - \frac{n^2}{2} + \frac{1}{2} \right) \leq 2n \ln n. \]
Estimation of Sums

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

$$
\sum_{k=1}^{n-1} 2k \ln k \leq \sum_{k=1}^{n-1} \int_{k}^{k+1} 2x \ln x \, dx
$$

$$
= \int_{1}^{n} 2x \ln x \, dx
$$

$$
= \left( x^2 \ln x - \frac{x^2}{2} \right) \bigg|_{1}^{n}
$$