quicksort

Canonical name	Quicksort
Date of creation	2013-03-22 13:59:19
Last modified on	2013-03-22 13:59:19
Owner	thouis (3290)
Last modified by	thouis (3290)
Numerical id	13
Author	thouis (3290)
Entry type	Definition
Classification	msc 68P10
Related topic	SortingProblem

Quicksort is a divide-and-conquer algorithm for sorting in the comparison model. Its expected running time is $O(n \lg n)$ for sorting n values.

Algorithm

Quicksort can be implemented recursively, as follows:

Algorithm QUICKSORT(L)

Input: A list L of n elements

Output: The list L in sorted order

if $n > 1$ **then**

  $p \leftarrow$ random element of L

  $A \leftarrow \{x | x \in L, x < p\}$

  $B \leftarrow \{z | z \in L, z = p\}$

  $C \leftarrow \{y | y \in L, y > p\}$

  $S_A \leftarrow Quicksort(A)$

  $S_C \leftarrow Quicksort(C)$

  return $Concatenate(S_A, B, S_C)$

else

  return L

Analysis

The behavior of quicksort can be analyzed by considering the computation as a binary tree. Each node of the tree corresponds to one recursive call to the quicksort procedure.

Consider the initial input to the algorithm, some list L. Call the Sorted list S, with ith and jth elements S_i and S_j. These two elements will be compared with some probability p_{ij}. This probability can be determined by considering two preconditions on S_i and S_j being compared:

- S_i or S_j must be chosen as a pivot p, since comparisons only occur against the pivot.
- No element between S_i and S_j can have already been chosen as a pivot before S_i or S_j is chosen. Otherwise, would be separated into different sublists in the recursion.

The probability of any particular element being chosen as the pivot is uniform. Therefore, the chance that S_i or S_j is chosen as the pivot before any element between them is $2/(j-i+1)$. This is precisely p_{ij}.

The expected number of comparisons is just the summation over all possible comparisons of the probability of that particular comparison occurring. By linearity of expectation, no independence assumptions are necessary. The expected number of comparisons is therefore

$$
\begin{aligned}
\sum_{i=1}^{n} \sum_{j>i}^{n} p_{ij} &= \sum_{i=1}^{n} \sum_{j>i}^{n} \frac{2}{j-i+1} && (1) \\
&= \sum_{i=1}^{n} \sum_{k=2}^{n-i+1} \frac{2}{k} && (2) \\
&\leq 2 \sum_{i=1}^{n} \sum_{k=1}^{n} \frac{1}{k} && (3) \\
&= 2nH_n = O(n \lg n), && (4)
\end{aligned}
$$

where H_n is the nth Harmonic number.

The worst case behavior is $\Theta(n^2)$, but this almost never occurs (with high probability it does not occur) with random pivots.