# Expected Number of Comparisons in quick-sort

Let ***A*** be a random array, in which every element is distinct.

The distance between two elements is defined as:

If ***a*** is the ***i***-th smallest element in ***A***, ***b*** is the ***j***-th smallest element in ***A***, then the distance between ***a*** and ***b*** is ***|i-j|***.

We will show that when running quick-sort on an array with N distinct items, the probability of comparing any two elements whose distance are ***n*** is ***2/n***.

Assume that the two elements are the i-th smallest and j-th smallest in the array. If an element is randomly selected, there are three cases :

1. the i-th or th j-th smallest is selected as the partition element. Probability ***2/N***. The two elements will be compared. After the partition, they will never be compared again.

2. the element larger than the i-th and smaller than the j-th is selected as the partition element. Probability ***(n-1) / N***. the two elements won't be compared, and after the partition is done, they belong to different partitions and they will never be compared.

3. the element smaller than the i-th or larger than the j-th is selected. Probability ***(N – n – 1)/N***. The two will not be compared, but since they belong to the same partitions, they may be compared in the following partitions.

Let ***P(N, i, j)*** denote the possibility of comparing the ***i***-th smallest and ***j***-th smallest of an array of length ***N***, where ***i < j*** and ***N >=2*** . Since the array is random, from the previous reasoning we have:

Let ***T(N)*** be the proposition :

***P(N, i, i + n)***  ***= 2 / (n+1)***, where ***0<* i *,0 < n and i + n <= N***

Basic step : when ***N = 2***, ***P(2, 1, 2) = 2 / (1+1) = 1***. the only possible case is ***i = 1*** and n ***= 1***

Inductive step: assume that ***P(K)*** is true for ***2 <= K <= N***. And assume that ***0 < i, 0 < n, i + n <= N + 1***.

For , since ***2 <= N, 1 <= k < i*** we have ***2 <= N + 1 – k < N , i – k > 0, i + n – k <= N + 1 – k*** , by inductive hypothesis ***P(N -k, i-k,i+n-k)=2/(n+1)***.

For , since ***i + n + 1 <= k <= N + 1***, we have ***2 <= k – 1 <= N***, and ***i > 0, i + n <= k-1***, by inductive hypothesis ***P(k-1, i, i+n) = 2 /(n+1)***.

, that is

We have shown that ***T(N + 1)*** follows from inductive hypothesis.

Thus ***T(N)*** is true for all ***N >= 2****.*

So the Expected number of comparisons when quick-sort a random array with distinct elements of length ***N*** is :