QuickSort

(read Section 7 on Quicksort in Cormen, Leiserson, Rivest, Stein)

QuickSort has the reputation of being the fastest comparison-based sorting algorithm. Indeed it is very fast on the average but can be slow in bad cases unless precautions are taken.

**Quicksort Algorithm.** Quicksort follows the divide-and-conquer paradigm: it divides the unsorted array into two, it recurses on the two pieces, and it finally combines the two sorted pieces to obtain the sorted array. An interesting feature of quicksort is that splitting separates small from large items and makes combining the sorted pieces a triviality.

```c
void QUICKSORT(int l, r)
    if l < r then
        m = SPLIT(l, r);
        QUICKSORT(l, m - 1);
        QUICKSORT(m + 1, r)
    endif.

Assuming the items are stored in A[0..n - 1], we sort by calling QUICKSORT(0, n - 1).

**Splitting.** The performance of quicksort depends heavily on the performance of the split operation. The effect of splitting from \( l \) to \( r \) is:

1. \( x = A[l] \) is moved to its correct location at \( A[m] \),
2. no item in \( A[l..m - 1] \) is larger than \( x \),
3. no item in \( A[m + 1..r] \) is smaller than \( x \).

Figure 3 illustrates the process with an example. The nine items are split by moving a pointer \( i \) from left to right and another pointer \( j \) from right to left. The process stops when \( i \) and \( j \) cross. To get splitting right is a bit delicate, in particular in special cases. Make sure the algorithm is correct for (i) \( x \) is smallest, (ii) \( x \) is largest, (iii) all items are the same.

```c
int SPLIT(int l, r)
    x = A[l]; i = l; j = r + 1;
    repeat repeat i++ until x <= A[i];
        repeat j-- until x >= A[j];
        if i < j then SWAP(i, j) endif
    until i >= j;
    SWAP(l, j); return j.
```

Special cases (i) and (iii) are ok but case (ii) requires a stopper at \( A[r + 1] \). This stopper must be an item at least as large as \( x \). If \( r < n - 1 \) this is automatically the case. For \( r = n - 1 \) we guarantee this by storing \( A[n] = +\infty \).

**Running Time.** The actions taken by quicksort can be expressed using a binary tree: each (internal) node represents a call and displays the length of the subarray; see Figure 4. The worst case occurs when \( A \) is already sorted. In this case, the tree degenerates to a list without branching. The sum of lengths can be described by the following recurrence relation:

\[
T(n) = n + T(n - 1) = \sum_{i=1}^{n} i = \binom{n + 1}{2}.
\]
The running time in the worst case is therefore in $O(n^2)$. In the best case the tree is completely balanced and the sum of lengths is described by the recurrence relation

$$T(n) = n + 2 \cdot T\left(\frac{n-1}{2}\right).$$

If we assume $n = 2^k - 1$ we can rewrite the relation as

$$U(k) = (2^k - 1) + 2 \cdot U(k-1)$$
$$= (2^k - 1) + 2(2^{k-1} - 1) + \ldots + 2(2 - 1)$$
$$= k \cdot 2^k - \sum_{i=0}^{k-1} 2^i$$
$$= 2^k \cdot k - (2^k - 1)$$
$$= (n + 1) \cdot \log_2 (n + 1) - n.$$

The running time in the best case is therefore in $O(n \log n)$.

**Randomization.** One of the drawbacks of quicksort, as described until now, is that it is slow on rather common almost sorted sequences. The reason are pivots that tend to create unbalanced splittings. Such pivots tend to occur in practice much more often than one would maybe expect. Human and often also machine generated data is frequently biased towards certain distributions, and it has been said that 80% of the time or more, sorting is done on either an already sorted or an almost sorted file.

Such situations can often be helped by adding an element of randomness to the algorithm. In this particular case, we use randomization to make the choice of the pivot independent of the input data. Assume $\text{RANDOM}(\ell, r)$ returns an integer $p \in [\ell, r]$ with uniform probability:

$$\text{Prob} [\text{RANDOM}(\ell, r) = p] = \frac{1}{r - \ell + 1}$$

for each $\ell \leq p \leq r$. In other words, each $p \in [\ell, r]$ is equally likely. The following algorithm splits the array with a random pivot:

```c
int \text{RANDOMSPLIT}(int \ell, r)
  p = \text{RANDOM}(\ell, r); \ \text{Swap}(\ell, p);
  \text{return} \text{SPLIT}(\ell, r).
```

We get a randomized implementation by substituting $\text{RANDOMSPLIT}$ for $\text{SPLIT}$. The behavior of this version of quicksort depends not only on the input, $A[\ell..r]$, but also on $p$, which is produced by a random number generator.

**Average Analysis.** We assume that the items in $A[0..n-1]$ are pairwise different. The pivot splits $A$ into

$$A[0..m-1], \ A[m], \ A[m+1..n-1].$$

By assumption on function $\text{RANDOMSPLIT}$, the probability for each $m \in [0, n-1]$ is $\frac{1}{n}$. Therefore the average sum of array lengths split by $\text{QUICKSORT}$ is

$$T(n) = n + \frac{1}{n} \cdot \sum_{m=0}^{n-1} (T(m) + T(n-m-1)).$$

We multiply by $n$ and obtain a second relation by substituting $n - 1$ for $n$:

$$n \cdot T(n) = n^2 + 2 \cdot \sum_{i=0}^{n-1} T(i), \quad (1)$$

$$(n - 1) \cdot T(n - 1) = (n - 1)^2 + 2 \cdot \sum_{i=0}^{n-2} T(i). \quad (2)$$

Next we subtract (2) from (1), we divide by $n(n+1)$, and we use repeated substitution to express $T(n)$ as a sum:

$$\frac{T(n)}{n + 1} = \frac{T(n - 1)}{n} + \frac{2n - 1}{n(n+1)}$$
$$= \frac{T(n - 2)}{n - 1} + \frac{2n - 3}{(n-1)n} + \frac{2n - 1}{n(n+1)}$$
$$= \sum_{i=1}^{n} \frac{2i - 1}{i(i+1)}.$$

**Bounding the Sum.** We split the sum into two pieces:

$$\sum_{i=1}^{n} \frac{2i - 1}{i(i+1)} = 2 \cdot \sum_{i=1}^{n} \frac{1}{i + 1} - \sum_{i=1}^{n} \frac{1}{i(i+1)}.$$

The second sum is solved directly by transformation to a telescoping series:

$$\sum_{i=1}^{n} \frac{1}{i(i+1)} = \sum_{i=1}^{n} \left( \frac{1}{i} - \frac{1}{i+1} \right)$$
$$= 1 - \frac{1}{n+1}.$$
The first sum is bounded from above by the integral of $\frac{1}{x}$ for $x$ ranging from 1 through $n+1$; see Figure 5. The sum of $\frac{1}{i+1}$ is the sum of areas of the shaded rectangles, and because all rectangles lie below the graph of $\frac{1}{x}$ we can bound the total rectangle area by the integral of $\frac{1}{x}$:

$$\sum_{i=1}^{n} \frac{1}{i+1} < \int_{1}^{n+1} \frac{dx}{x} = \ln(n+1).$$

We plug this bound back into the expression for the average running time:

$$T(n) < (n+1) \cdot \sum_{i=1}^{n} \frac{2}{i+1} < 2 \cdot (n+1) \cdot \ln(n+1) = \frac{2}{\log_2 e} \cdot (n+1) \cdot \log_2(n+1).$$

In words, the running time of quicksort in the average case is only a factor of about $2/\log_2 e = 1.39 \ldots$ slower than in the best case. This also implies that the worst case cannot happen very often, for else the average performance would be closer to the worst case.

**Stack Size.** Another drawback of quicksort is the recursion stack, which can reach a size of $\Omega(n)$ entries. This can be improved by always first sorting the smaller side and simultaneously removing the tail-recursion:

```c
void QUICKSORT(int l, r)
  i = l; j = r;
  while i < j do
    m = RANDOMSPLIT(i, j);
    if m - i < j - m then QUICKSORT(i, m - 1);
      i = m + 1
    else QUICKSORT(m + 1, j);
      j = m - 1
  endif
endwhile.
```

In each recursive call to QUICKSORT, the length of the array is at most half the length of the array in the preceding call. This implies that at any moment of time the stack contains no more than $1 + \log_2 n$ entries. Note that without removal of the tail-recursion the stack can reach $\Omega(n)$ entries even if the smaller side is sorted first.

**Summary.** The average running time of quicksort is in $O(n \log n)$ and the extra amount of memory it requires is in $O(\log n)$. For the deterministic version, the average is over all $n!$ permutations of the input items. For the randomized version the average is the expected running time for every input sequence.