Lecture Notes 2

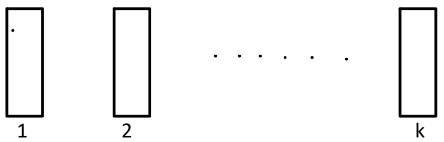
By Rong Zeng

2017年9月28日 14:17

2. Exact Median Finding

Divide and Conquer

(1) Splitting the input of size n into n/5 groups of 5 each, M1,M2,......Mk (k=n/5).



$T = 25 * \frac{n}{5}$

Figure 1

(2) Find n/5 group-medians {m1, m2, ...mk}, as figure 1, using selection sort for each group firstly, and the time needed at this step is $T = 25 * \frac{n}{5}$ (in which 25 is the time for finding medians in each group using selection sort).

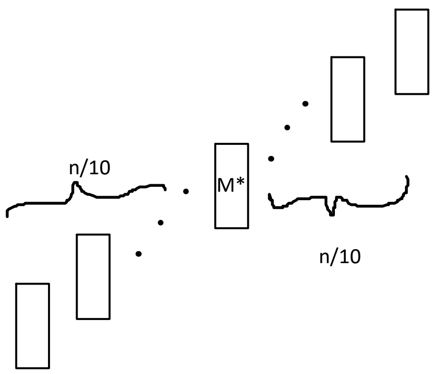


Figure 2

(3) Find M* = median of {m1, m2, ...mk}, k=n/5.

(4) There are n/10 numbers in the group-medians{m1, m2, ...mk} that is bigger than M*, and n/10 numbers smaller; As depicted in figure 2, then in the left n/10 groups, there must be at least 3 numbers smaller than M* in each group, and in the right n/10 groups, at least 3 numbers bigger than M* in each group, that is, we can throw away 3n/10 numbers(smaller or bigger than M*) at least, which means 7n/10 numbers have been left at this step, at most.

Master Theorem

If T(n) = $\sum_{i=1}^{k} T(a_i * n) + n$ and $a_i \le 1$ && $\sum_{i=1}^{k} a_i \le 1$

Then

T(n) = O(n).

Time Complexity: $T(n) = 25 * \frac{n}{5} + T\left(\frac{n}{5}\right) + T\left(\frac{7n}{10}\right) + n$

According to the Master Theorem: T(n) = O(n).

3. Random Median Finding on Small Memory Computer

- F is the memory size

(1) Pick F elements to memory

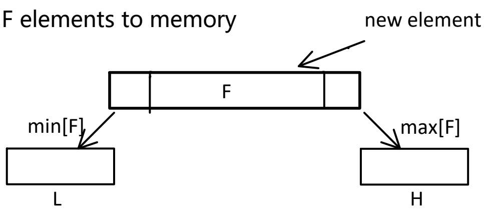


Figure 3

(2) Randomly choose a new element from the remainder, throw away max[F] to set H, or min[F] to set L, according as H<L or H>=L respectively.
(3) Repeat (2) until all elements have entered the memory once.
(4) Find the median in F.

- High probability that the median is in F.
  That is, for all x ∈ L , x <= median and for all y ∈ H , median >= y.

Probability that throwing away median

Firstly we assume B is the set of smaller half of the input numbers (median not included), and T the set of the bigger half.

Figure 4

When selecting the elements into memory, every time if we choose x ∊ B , Z = -1 , if we choose y ∊ T, then Z= 1.

Define W = $Z_1 + Z_2 + \cdots + Z_n$;

$W_k = Z_1 + Z_2 + \cdots + Z_k$;

Then the median in F when | W_k | < $|F|$ for all k<=n.

Example

A walking model where walking left or right has the equal probability(1/2).

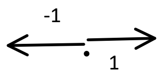


Figure 5

we have L+M+H=n, and W defined as above, then

Probability(W=L)=$\binom{n}{L+\frac{n-L}{2}}/2^n$

Reference:
J. I. Munro and M. S. Paterson, Selection and Sorting with Limited Storage, Theoretical Computer Science, vol. 12, pp. 315-323, 1980.

Random Walk Model

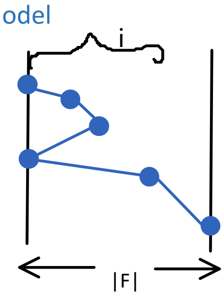


Figure 6

We define $X_n(i)$ as the probability of hitting the line F in the next n steps, in which i refers to the distance from the origin point, as depicted in figure 6.

Then the probability that median is in F would be

$$X_n(0) = \max_{0 \le i \le n} Pro(|W_i| \ge |F|)$$

According to definition

$$X_n(0) = X_{n-1}(1)$$

And

$$X_n(i) = 1 \quad \text{if} \quad i>|F| \qquad \text{(hit the line F)}$$

Then we have

$$X_n(i) = \frac{1}{2}[X_{n-1}(i-1) + X_{n-1}(i+1)]$$