中国科学技术大学 计算机科学与技术系
University of Science and Technology of China
DEPARTMENT OF COMPUTER SCIENCE AND TECHNOLOGY

算法基础
Foundation of Algorithms

主讲人　徐 云

Fall 2018，USTC

第9章 中值和顺序统计

9.1 最小和最大值

9.2 期望时间为线性的选择

9.3 最坏时间为线性的选择

9.1 最小和最大值

- 最小/最大值：最坏情形W(n)=n-1次比较，时间为Θ(n)
- 同时求最大、最小值
  - 一种方法：独立分别求，比较次数为n-1+n-2=2n-3
  - 另一种方法：

    成对输入x, y，每对比较3次

    ① 比较x, y；

    ② 将min(x, y)与当前最小值比较；

    ③ 将max(x, y)与当前最大值比较；

    总比较次数约为$3\lfloor n/2 \rfloor$。//第一对元素比较一次，最后一组元素若为一个，至多比较二次

第9章 中值和顺序统计

9.1 最小和最大值

9.2 期望时间为线性的选择

9.3 最坏时间为线性的选择

9.2 期望时间为线性的选择

- 基本思想
- RandomizedSelect算法
- 时间分析

基本思想

- 基于分治法的思想
  - 利用快排序的随机划分法，进行问题的划分
  - 具体步骤：

① 划分$A[p..r] \Rightarrow A[p..q-1] \le A[q] < A[q+1..r]$；

// A[q]为划分元

② $k \leftarrow q-p+1$；　　//即A[q]是第k个最小元

③ if (i=k) then　　// k = 左区间长度+1

return A[q]；

if (i<k) then 在左区间中继续找第i个元素；

if (i>k) then 在右区间中继续找第i-k个元素；

临界条件：当区间长度为1时，直接返回该元素

RandomizedSelect算法

```
RandomizedSelect(A, p, r, i)
{ //选择ith元素
    if p=r then  return A[p];                //临界问题处理
    q←RandomizedPartition(A, p, r);
                                  //进行划分，并返回划分元的下标
    k←q-p+1;                          //A[q]是第k个小的元素
    if i=k then                       //A[q]是ith元素
        return A[q];
    else if i < k then               //ith元素落在左区间
        return RandomizedSelect(A, p, q-1, i);
    else                             //ith元素落在右区间
        return RandomizedSelect(A, q+1, r, i-k);
}
```

时间分析

- 最好：每次划分为相等的左右区间

  $T(n)=T(n/2)+n \Rightarrow T(n)=\Theta(n)$

- 最坏：每次划分为不均等的左右区间

  $T(n)=T(n-1)+n \Rightarrow T(n)=\Theta(n^2)$

- 平均(期望)：分析略。

  $T(n)=\Theta(n)$

第9章 中值和顺序统计

9.1 最小和最大值

9.2 期望时间为线性的选择

9.3 最坏时间为线性的选择

9.3 最坏时间为线性的选择

- 算法步骤
- 时间分析

算法步骤

While n > 1 do

step 1. 将n个元素分成5个1组，共$\lceil n/5 \rceil$组。其中最后1组有n mod 5个元素。

step 2. 用插入排序对每组排序，取其中值。若最后1组有偶数个元素，取较小的中值。

step 3. 递归地使用本算法找找$\lceil n/5 \rceil$个中值的中值x。

step 4. 用x作为划分元对A数组进行划分，并设x是第k个最小元。

step 5. if i=k then return x;

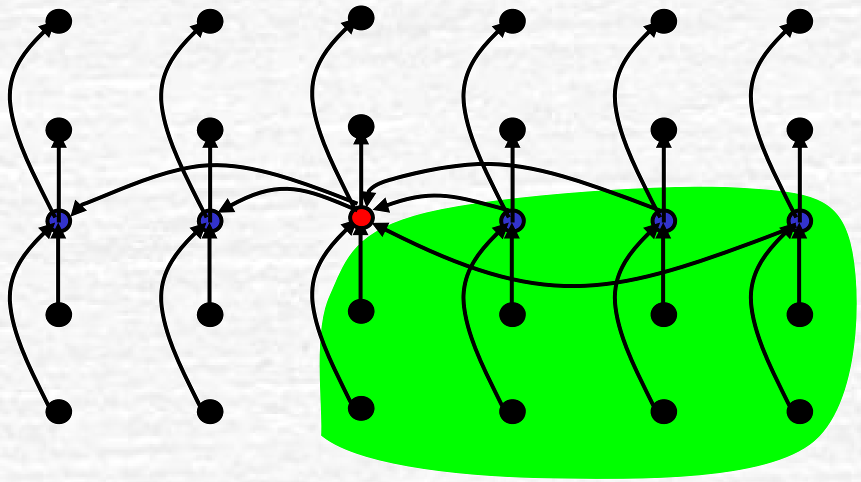
else if i<k then 找左区间的第i个最小元;

else 找右区间的第i-k个最小元;

时间分析（1）

- n个元素中至少有多少个元素>x？

每组按列形式，以每组中值升序次序从左到右排列如下：



图中：箭头指向的元素小于箭尾元素

可以知道，大于x的元素至少有$3(\lceil \lceil n/5 \rceil /2 \rceil -2) \geq 3n/10-6$

同理，小于x的元素至少有$3n/10-6$

- 由上⇒左区间和右区间的最大长度$\leq 7n/10+6$

时间分析（2）

- 运行时间递归式的建立

  step 1, 2: O(n);

  step 3: T(⌈n/5⌉);

  step 4: O(n);

  step 5: 至多T(7n/10+6)

$$\Rightarrow T(n) \le \begin{cases} \theta(1) & if \;\; n \le 140 \\ T(\lceil n/5 \rceil) + T(\frac{7}{10}n + 6) + \theta(n) & if \;\; n > 140 \end{cases}$$

时间分析（3）

- 运行时间递归式的求解

用替代法证：$T(n) \le cn$

$T(n) \le c\lceil n/5 \rceil + c(7n/10+6) + an$ //a为常数

$\le c(n/5+1) + c(7n/10+6) + an$

$= cn/5 + c + 7cn/10 + 6c + an$

$= 9cn/10 + 7c + an$

$= cn + (-cn/10 + 7c + an)$

$\le cn$ //if $-cn/10+7c+an \le 0$

要使$-cn/10+7c+an \le 0$，只要$c \ge 10an/(n-70)$

$\because$ 假定$n>140$，$\therefore$ 有$n/(n-70) < 2$

$\therefore$ 取$c \ge 20a \Rightarrow -cn/10+7c+an \le 0$

故 $T(n)=O(n)$ □

End of Ch9