Homework 2

和泳毅 PB19010450

1.下面的排序算法中哪些是稳定的：插入排序、归并排序、堆排序、快速排序和计数排序？给出一个能使任何排序算法都稳定的方法。你所给出的方法带来的额外时间和空间开销是多少？

答：稳定：插入排序、归并排序、计数排序。

在排序之前，先为长度为n的序列的每一个数赋予一个1~n的标签。排序时遇到相等的数，再比较其标签，以保证稳定性。额外的时间复杂度为$O(n)$，空间复杂度为$O(n)$。

2.假设所有元素都是互异的，说明在最坏情况下。如何使快速排序的运行时间为$O(n\log n)$。

答：使用最坏情况为线性时间的选择算法来找到序列的中位数，将中位数作为主元进行快速排序。此时又递归式$T(n)=2T(n/2)+O(n)$，由主方法情况2，得到$T(n)=O(n\log n)$。

3.给定一个整数数组，其中不同的整数所包含的数字的位数可能不同。但该数组中，所有整数中包含的总数字位数为n。设计算法使其可以在$O(n)$时间内对该数组进行排序。

答：假设所有数都是正数且最高位不为0，令m为数的个数，则$m\le n$。

首先用计数排序方法按照位数进行排列。对所有数赋予位数标签，耗时$O(n)$。接着按照位数标签，对数组进行计数排序，耗时$O(m+n)=O(n)$。此时数组已按照位数排列。

然后再用基数排序对位数相同的数进行排序。令m_i为位数为$i\in\{1,2,\dots,n\}$的个数。对i位的数基数排序所花时间为$O(m_i)$，基数排序总耗时$\sum_{i=1}^{n}O(m_i)=O(n)$。
所以计数排序加基数排序总需时间为$O(n)$.

4.SELECT 算法最坏情况下的比较次数 $T(n)=\Theta(n)$，但是其中的常数项是非常大的。请对其进行优化，使其满足：

- 在最坏情况下的比较次数为 $\Theta(n)$。
- 当 i 是小于 $n/2$的常数时，最坏情况下只需要进行 $n+O(\log n)$次比较。

答：当$i \geq n/2$时，仍然使用该SELECT算法。当 $i < n/2$时，使用如下算法：

1. 取$m=\lfloor n/2 \rfloor$，将输入的数组$A[p...p+n-1]$进行划分，第一组（左侧组）为$A[p...p+m-1]$，第二组（右侧组）为$A[p+m...p+n-1]$。如果n为奇数，则将多出的$A[2m+1]$单暂时独分出。对左侧组中的每个数，在右侧组中建立一一对应的映射关系$A[p+j] \leftrightarrow A[p+j+m], j=0,\ldots,m-1$，保持两组中一对数在组内的位置同步变化。
2. 分别对两组的第j个数进行比较，即$A[p]$和$A[p+m]$，$A[p+1]$和$A[p+m+1]$，...，如果$A[p+j] < A[p+m+j]$，则调换这两个元素，把较小元素放到$A[p+m+j]$，较大元素放到$A[p+j]$。完成后右侧组的元素依次比左侧组对应元素小。
3. 对右侧组$A[p+m]...A[p+n-1]$递归执行1、2。（注意，如果递归中对两个分组的元素进行调换，那么需要保持1中的映射关系，他们上一层分组左侧组相应位置也需要调换，依次类推，上上层以及更多层的相应位置都得调换，保证每个左侧组依次大于对应的右侧组。）
4. 每递归一次，组内元素就减少一半。当i大于等于组内元素个数的一半时即$i \geq m/2$，停止递归，对最后分组B_l使用**SELECT**算法。SELECT后，B_l中前i个数是该分组最小的i个数，由于映射关系，倒数第二个分组B_r中的元素依次大于B_l中对应元素。那么B_r和B_l的第i小数就在B_r的前i个数和B_l的前i个数之中，对这$2i$个数使用**SELECT**算法(如果n为奇数，就将最后一个数也加入），得到这两个分组最小的i个数，返回上一层调用当中。
5. 在上一层调用中，对左右分组前i个数共$2i$个数使用**SELECT**算法，然后再返回，直到顶层。最后得到$A[p...p+n-1]$中最小的i个数，其中第i个数即为目标结果。

记在n个元素中找出第i小元素的比较次数为$Comp_i(n)$，$i \geq n/2$时显然有$Comp_i(n)=T(n)$。而当$i<n/2$时，由以上算法，每一层递归分组需要比较$\lfloor n/2 \rfloor$次，并对$2i$个元素使用**SELECT**算法，对$\lceil n/2 \rceil$个元素产生新的分组进行下一次递归，则有递归式$Comp_i(n)=\lfloor n/2 \rfloor + Comp_i(\lceil n/2 \rceil)+T(2i)$。

即

$$Comp_i(n)=\begin{cases} T(n) & i \geq n/2 \\ \lfloor n/2 \rfloor + Comp_i(\lceil n/2 \rceil)+T(2i) & i < n/2 \end{cases}$$

则在最坏情况下的比较次数为 $\Theta(n)$。进一步，当$i<n/2$，证明$Comp_i(n)=n+O(T(2i)\log(n/i))$，进行数学归纳，假设对所有$m<n$该式都成立，有：

$$\begin{aligned} Comp_i(n) &= \lfloor n/2 \rfloor + Comp_i(\lceil n/2 \rceil)+T(2i) \\ &= \lfloor n/2 \rfloor + \lceil n/2 \rceil + f(T(2i)lg(\lceil n/2 \rceil/i))+T(2i) \\ &= n + f(T(2i)lg(\lceil n/2 \rceil/i))+T(2i) \\ &= n+O(T(2i)\log(n/i)) \end{aligned}$$

并且，如果 i 是小于 $n/2$的常数，有：

$$
\begin{aligned}
Comp_i(n) &= n + O(T(2i)\log(n/i)) \\
&= n + O(O(1)\log(n/i)) \\
&= n + O(\log(n) - \log(i)) \\
&= n + O(\log(n) - O(1)) \\
&= n + O(\log(n)).
\end{aligned}
$$