

第 章算法分析题

学号: 2209060322

姓名: 梁桐

班级: 计算机 2203

7.2

解: (1) 以概率 my/n 抽取记录, 该方法的标准差是 $\sqrt{m(1-m/n)}$, 有时可能不满足要求

假设在前 t 次考察的记录中, 已抽取 k 个记录, 接下来第 $t+1$ 次考察取得第 $k+1$ 个记录的概率为

$$\frac{\binom{n-t-1}{n-k-1}}{\binom{n-t}{n-k}} = \frac{n-k}{n-t}$$

按此概率抽取记录是无偏的

抽样算法:

```

Void sample(int n, int m, int S[]) {
    RandomNumber rnd;
    int x=0, y=0, k;
    while (y < m) {
        double u = rnd.fRandom();
        k = u * n;
        if (u * (n-x) < (n-y)) {
            S[k]++;
            y++;
        }
        x++;
    }
}
    
```

(2) 如果事先不知道文件中记录的个数, 通常可以先做一次扫描, 确定记录的个数后再抽样。另一个较好的方法是在扫描时预先随机抽取 $p > m$ 个记录, 然后对抽取出的 p 个记录做二次抽样, 从中随机抽取 m 个记录。

7-8.

解: 用主教材中的有序链表类 OrderList, 对有序链表做 n 次插入运算, 第 i 次插入运算平均需要 $O(\sqrt{i})$ 运算时间。含位逆算法 $O(n/k)$ 取 $k=\sqrt{n}$ 则 $O(\sqrt{n})$ 。

插入 n 次总的 $O(n)$ 可表示为 $\sqrt{1} + \sqrt{2} + \sqrt{3} + \dots + \sqrt{n}$

$$\int_0^n \sqrt{x} dx + O(n)$$

$$Max < \sum_{i=1}^n (\sqrt{i+1} - \sqrt{i}) = \sqrt{n+1} - 1$$

$$O(n) < \frac{2}{3} \cdot n^{\frac{3}{2}} + \sqrt{n} - 1$$

$$O(n) = n^{\frac{3}{2}}$$