习题三

3.1 数值随机算法计算数值 a 的精度可以表示为置信区间 Pr[$x\in[a-\delta, a-\delta]$]> 1-γ。试利用切尔诺夫界为第 2 章计算π的数值随机算法之一建立置信区间，使得我们可以根据置信水平和置信区间估计所需随机实验的次数。

3.2 给定两个集合 $A,B(A\subseteq B),|B|=n$。下面的算法通过数据抽样计算$|A|$。

输入： $B,|B|=n, A\subseteq B$的成员资格判定条件 C

输出： $|A|$

1. k=0
2. for i=0 to m do
3. 从 B 中均匀随机抽取样本 x
4. If x 满足条件 C then k = k+1
5. 输出 kn/m

利用 Chernoff 界给出 Pr[算法输出值$\in[(1-\lambda)|A|,(1+\lambda)|A|]]>1-\gamma$时算法中 m 的取值。

3.3 QuickSort 排序过程可以视为算法的递归调用过程，因此整个算法的执行过程可以视为一棵递归调用树，算法的每次调用对应树中的一个结点，结点间的边表示直接嵌套的调用关系。在每次调用 QuickSort 时，首先从当前数据子集(记其大小为 s)中随机选择划分元素将当前子集划分为两个子集合；如果划分得到的两个子集的大小均不超过 $2s/3$,则称递归调用树中相应节点为*好结点*，否则称之为*坏结点*。

(a)证明：在任意从树根到叶子的路径上，好结点的数量不超过 $c_1\log_2 n$,其中 c_1 是一个常数；

(b)证明：任意从树根到叶子的路径上所含结点的数量不超过 $c_2\log_2 n$ 的概率至少为 $1-1/n^2$, 其中 c_2 是一个常数；

(c) 从树根到叶子的最长路径上所含结点的数量不超过 $c_2\log_2 n$ 的概率至少为 $1-1/n$, 其中 c_2 是(b)中的常数；

(d)利用 a,b,c 的结论，证明：QuickSort 在 $O(n\log n)$时间内排序 n 个数据对象的概率至少为 $1-1/n$。

3.4 设 $X_0=0$, 而 $X_{j+1}(j\geq 0)$是从$[X_j, 1]$均匀随机抽取的值，令 $Y_k=2^k(1-X_k)$。证明：序列 $Y_0,Y_1,\ldots$是一个鞅。

3.5 利用本章所学内容，分析如下随机排序算法的时间复杂性。

输入： n 个不同的值 $x_1,x_2,\ldots,x_n$

输出： $x_1,x_2,\ldots,x_n$ 排序后的结果

步骤：1. 从 $x_1,x_2,\ldots,x_n$ 均匀随机抽取 y_1

2. For k=2 To n
3. 从$\{x_1,\ldots,x_n\}\backslash\{y_1,y_2,\ldots,y_{k-1}\}$中均匀随机抽取 y_k;
4. If $y_k<y_{k-1}$ Then goto 1;
5. 输出 $y_1,y_2,\ldots,y_n$;

3.6 给定 n 个数据对象 $o_1,o_2,\ldots,o_n$ 和相应的正权值 $w_1,w_2,\ldots,w_n$。试利用你已经掌握的知识设计一种数据结构用于从 $o_1,o_2,\ldots,o_n$ 中实现随机抽样时 o_i 被抽中的概率为 $w_i/(w_1+w_2+\ldots+w_n)$,且每次抽样的时间均为 $O(\log n)$。