物联网信息处理实验四

实验要求：

使用自己所擅长的程序语言编写 AMS 算法(估计数据流元素的二阶矩)。假设所有元素为 1-n(整数)，m_i 表示元素 i 的出现次数(在整个文件中的出现次数)，则二阶矩等于$(m_1)^2+(m_2)^2+(m_3)^2+...+(m_n)^2$。

实验步骤：

1：以 stream_for_ams.txt 文件为自己所写程序的输入，读取文件中数据(数值范围是 $1-10^5$)；

2：请编写一个精确算法 A，来计算整个文件 stream_for_ams.txt 中所有数据的二阶矩的真实值(179866)；

3：假设文件中的数据为 e1,e2,e3,…,eN，设采样点个数为 n 个，即随机从数据中选取 1 个位置后，利用 AMS 算法计算估计的二阶矩值(参看书第二版，112 页)；

4：使用书上 110 页（组合估计）的技巧；假设有 m 分组，每个分组包含 n 个位置；对于每组计算二阶矩估计的平均值作为改组的二阶矩估计，然后用 m 个组的二阶矩估计的中位数，作为最终的二阶矩估计值；

5：设真实的二阶矩为 M，令 m=1,n=1，重复实验步骤 4，20 次，可以得到 20 个的估计值(二阶矩)M_1 M_2 ,…,M_{20}，计算平均误差 error_sum $=\{[(M_1-M)^2+(M_2-M)^2+(M_3-M)^2+...++(M_{20}-M)^2]/20\}^{0.5}$；

6：令 m=10,n=10，重复步骤，比较当 m=1,n=1 时的平均误差与 m=10,n=10 的平均误差。

7： 在数据流长度未知的情况下，给定选择位置个数 100（样本集合大小），综合水库采样算法和 ams 算法，设计编写新的算法，该算法在任意时刻都可以给出数据流中当前元素二阶矩的估计；

实验考察要求：

1：分析并讲解自己所编写程序；

2：分析 m=1,n=1 时和 m=10,n=10 时，平均误差有区别的原因；

3：结合 ams 和水库算法，编写出可以处理数据大小未知的情况的二阶矩估计算法；

4：不允许组队，如果在考察中，发现两人的程序完全相同，则两人本次的实验成绩为 0；