# 习题二

2.1 某应用需要在 10 人中以加密方式共享一个 100bit 的信息 $s$ 使得其中任意两人根据自己收到的信息能够恢复原始信息但任意一人无法根据自身收到的信息了解 $s$ 的任何情况。为此，10 位相关人员依次编号为 0,1,2,…,9。一种共享信息的方法如下。选择一个长度为 101 比特的素数 $q$,并将其剩余域记为 GF($q$)。在 GF($q$)中均匀一致地选定元素 $f$，并利用拉格朗日插值法获得一个系数取自 GF($q$)的一次多项式 $p(x)=(x-10)*s - (x-11)*f$ 使得 $p(10)=f$,$p(11)=s$。第 $i$ 个人收到的信息定义为 $p(i)$。

(a)请你说明如何根据计算从任意两人收到的信息中恢复 $s$。

(b)请你利用概率知识说明任何人仅凭自己收到的信息无法获知 $s$ 的任意有价值信息。

2.2 投掷一枚均匀硬币 $n$ 次，如果第 $i$ 次投掷和第 $j$ 次投掷出现同一面，则令 $X_{ij}=1$,否则令 $X_{ij}=0$。证明：$X_{ij}$（$i<j$）两两独立但不相互独立。

2.3 假设 $x_1,x_2,\ldots,x_n$ 是从$[0,2^k)$中两两独立地均匀抽取的 $n$ 个数。证明：用桶排序 $x_1,x_2,\ldots,x_n$ 时，时间复杂度的数学期望仍然是线性的。

2.4 身份证号码的前 6 位表示地区编码，中间 8 位是生日，最后 $k$ 位是 $k$ 个[0,9]之间的随机数字。为了确保身份证号码以 99%的概率具有唯一性，试建立模型确定 $k$ 的取值。

2.5 在开放寻址散列法中，散列表是一个数组 $A$ 且每个桶均无拉链。数组中每个位置要么包含一个散列项要么是空的。对每个待散列对象 $x$，散列函数 $h$ 定义了数组中的探测位置序列 $h(x,0), h(x,1),\ldots$。Insert($x$)如下操作：按照 $h$ 定义的探测位置序列 $h(x,0), h(x,1),\ldots$在数组中寻找空位置 $k$，并将 $x$ 存入 $A[k]$。Find($x$)如下操作：依次探查 $h$ 定义的探测位置序列 $h(x,0), h(x,1),\ldots$中的每个位置；如果 $A[h(x,i)]=x$,则返回 $h(x,i)$，否则，返回-1 表明 $x$ 未出现在散列表中。

假设用具有 $2n$ 个存储位置的数组作为开放寻址散列表存储 n 个数据项，并且 $h(x,i)$服从 0,1,…,$2n$-1 上的均匀分布，$h(x,1), h(x,2)\ldots$相互独立。用 $X_i$（$1\le i\le n$）表示第 i 次执行 Insert 操作时探查的位置个数，$X=\max_i X_i$ 表示 $n$ 次插入操作中各次操作的最大探查次数。

(1) 证明：Insert($x$)需要探查 $a$ 个存储位置的概率至多为 $2^{-a}$；

(2) 证明：$\Pr[X_i>2\log n] \le 1/n^2$;

(3) 证明：$\Pr[X>2\log n] \le 1/n$;

(4) 证明：$E[X] = O(\log n)$。

2.6 假设我们有 $m$ 首歌曲和 $n$ 名听众，每名听众有一个自己喜欢的歌曲列表。如果两名听众共同喜欢的歌曲越多，则说明听众的音乐口味越趋于相同。试利用本章所学内容，设计一个高效的算法找出口味相同的听众对。允许大家在完成作业时对题目中未明确的部分进行自由发挥。