Discrete

Mathematics - P173

第一类斯特林数(Stirling number of the first kind)，记为 $c(n,k)$，或 $\left[{n \atop k}\right]$

对于整数 $1 \le k \le n$，将n个人分配围坐于k张圆桌，其中每张圆桌均非空

如果每个人都有相同的左右邻居，则认为是同一种分配方式，考虑不同的安排方式数

对于整数 $1 \le k < n$，有 $c(n+1,k) = c(n,k-1) + n\,c(n,k)$

组合证明过程有：考虑第n+1个人的安排方式

当第n+1个人单独一桌时，则剩余n个人坐在剩余k-1桌

即有 $c(n,k-1)$ 种不同安排方式

当第n+1个人与其他人同坐一桌时，先考虑前n个人的安排

n个人坐在k桌有 $c(n,k)$ 种不同的安排方式

而对于每一种安排，将第n+1个人安排在前n个人中之一的左边

于是有 n 种不同的安排方式

即合计有 $n\,c(n,k)$ 种不同的安排方式

于是合计有 $c(n+1,k) = c(n,k-1) + n\,c(n,k)$ 种不同的安排方式

对于正整数 n，有 $\sum_{j=1}^{n} c(n,j) = n!$

证明过程有：基础步骤：当 $n=1$ 时，$c(1,1) = 1 = 1!$

递归步骤：假设对于任意 $n \in \mathbb{Z}^+$，$P(n)$ 为真，则考虑 $P(n+1)$

$$\sum_{j=1}^{n+1} c(n+1,j) = c(n+1,n+1) + \sum_{j=1}^{n} c(n+1,j)$$
$$= 1 + \sum_{j=1}^{n} [c(n,j-1) + n\,c(n,j)]$$
$$= 1 + \sum_{k=1}^{n-1} c(n,k) + n \sum_{j=1}^{n} c(n,j)$$
$$= \sum_{k=1}^{n} c(n,k) + n \sum_{j=1}^{n} c(n,j)$$
$$= (n+1) \sum_{k=1}^{n} c(n,k)$$
$$\overset{(IH)}{=} (n+1)\cdot n! = (n+1)!$$
即 $P(n+1)$ 为真

根据数学归纳法，对于 $n \in \mathbb{Z}^+$，有 $\sum_{j=1}^{n} c(n,j) = n!$

于是可以递归地定义 $c(n,k)$，$n,k \in \mathbb{Z}^+$

$$c(n,k) = \begin{cases} 0, & n < k \\ 0, & k = 0 \\ 1, & n = k \\ c(n-1,k-1) + n\,c(n-1,k), & \text{otherwise} \end{cases}$$

```
firstStirling :: Int -> Int -> Int
firstStirling n k
  | n < k      = 0
  | k == 0     = 0
  | n == k     = 1
  | otherwise  = firstStirling (n-1) (k-1)
                 + (n-1) * firstStirling (n-1) k
```

Discrete

Mathematics - P174

拉姆齐定理在图论中有[redacted]种不同的描述

对于所有 n 个顶点的图，其中或者包含一个[redacted] r 顶点的团，或者包含一个[redacted] s 顶点的独立集

使这个性质成立的最小自然数 n，记为 $R(r,s)$，其中 $r,s \in \mathbb{Z}^+$

团 (clique)，对于图 $G(V,E)$，存在一个 k 个顶点的点集 $V' \subseteq V$

使得 $\forall x,y \in V'$ $\{x,y\} \in E$，则称 $G_{V'}$ 为图 G 中的一个 k 顶点的团

独立集 (independent set)，对于图 $G(V,E)$，存在一个 k 个顶点的点集 $V' \subseteq V$

使得 $\forall x,y \in V'$ $\{x,y\} \notin E$，则称 V' 为图 G 中的一个 k 顶点的独立集

对于正整数 r,s，有 $R(r,s) \leq R(r-1,s) + R(r,s-1)$

考虑一个有 $R(r-1,s)+R(r,s-1)$ 个顶点的完全图 $G(V,E)$

并且对所有边指派[redacted]蓝色或红色

从图中随机选一个点 $u \in V$，则可以对 $V \setminus \{u\}$ 进行划分，

使得 $M = \{v \in V \mid (u,v)\text{是蓝色}\}$，$N = \{w \in V \mid (u,w)\text{是红色}\}$

于是可知 $M \cup \{u\} \cup N = V$，且两两交集为空

且有 $R(r-1,s) + R(r,s-1) = |M|+|N|+1$

于是有 $|M| \geq R(r-1,s)$ 或 $|N| \geq R(r,s-1)$

考虑 $|M| \geq R(r-1,s)$ 的情形，(同理可得 $|N| \geq R(r,s-1)$ G 中存在一个 s 顶点的红色团)

如果子图 G_M 中包含一个 s 顶点的红色团，则可知[redacted]

否则子图 G_M 中应包含一个 $r-1$ 顶点的蓝色团，

[redacted]又 M 中每个顶点与 u 的边均为蓝色，则存在一个 r 顶点的蓝色团

于是有 $R(r-1,s) + R(r,s-1) \geq R(r,s)$

当 $R(r-1,s)$，$R(r,s-1)$ 均为偶数时，有更强的结论 $R(r,s) \leq R(r-1,s)+R(r,s-1)-1$

令 $p = R(r-1,s)$，$q = R(r,s-1)$ 均为偶数，$t = p+q-1$ 为奇数

考虑 t 个顶点的完全图 $G(V,E)$ 并对所有边指派蓝色或红色

取 $G(V,E)$ 的蓝色子图 $G'(V,B)$，令 d_i 表示点 $v_i \in V$ 的度

则有 $\sum_{i=1}^{t} d_i$ 为偶数，而 t 是奇数，则必定存在一个 i 使得 d_i 为偶数

以这个点进行划分 $M = \{u \in V \mid (u,v_i)\text{是蓝色}\}$，$N = \{w \in V \mid (w,v_i)\text{是红色}\}$

于是有 $|M|+|N|+1 = t = p+q-1$，且 $|M| = d_i$，$|N| = t-1-d_i$ 均为偶数

即[redacted] $|M| \geq p-1$ 或 $|N| \geq q$，又 $p-1$ 为奇数，于是有 $|M| \geq p$

于是可以利用[redacted]相似的方法证明 $G(V,E)$ 满足 $R(r,s)$ 的条件

即有 $R(r-1,s)+R(r,s-1)-1 \geq R(r,s)$

Discrete Mathematics - P175

试验 (experiment)，从一组可能的结果中得出一个结果的过程
a procedure that yields one of a given set of possible outcomes

样本空间 (sample space)，一个试验可能结果的集合
the set of possibe outcomes of the experiment

事件 (event)，一个试验样本空间的子集 (subset of the sample space)
即对于样本空间 S，事件 $E \subseteq S$，

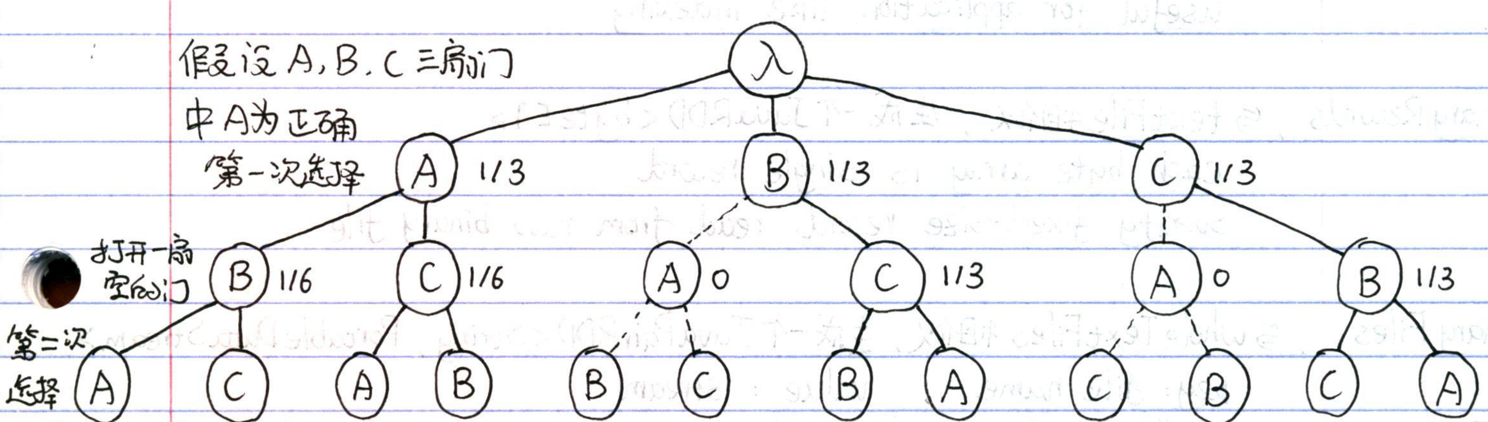
拉普拉斯定义 (Laplace's definition)，对于结果具有相等可能性的有限样本空间 S
S is a finite nonempty sample spase of equally likely outcomes
事件 E 的概率为该事件包含的结果次数除以可能结果的总次数
number of successful outcomes of this event
divided by number of possible outcomes
即有 $p(E) = |E|/|S|$，且有 $p(E) \in [0,1]$

抽样 (sampling)，分为有放回抽样 (sampling with replacement)
和无放回抽样 (sampling without replacement)

蒙地厅大厦的3门问题 (the Monty Hall Three-Door Puzzle)

1. 3扇门中的1扇指向是大奖，另外两扇为空
2. 从3扇门中选择一扇，但不打开
3. 出题方未选择的门中，打开一扇空的门
4. 选择坚持打开第一选择的门，或是打开未打开也未选择的门

假设A,B,C三扇门中A为正确



选择不换，则胜率为 $(\frac{1}{6}+\frac{1}{6})=\frac{1}{3}$，选择换，则胜率为 $(\frac{1}{3}+\frac{1}{3})=\frac{2}{3}$

Discrete Mathematics - P176

概率分布(probability distribution)，对于有限或可数无限个结果的样本空间S

对于每个结果 $s\in S$，指派一个概率 $p(s)$

使得满足条件：$0\le p(s)\le 1$，$s\in S$

且 $\sum_{s\in S} p(s)=1$

即每个结果的概率是一个不超过1的非负实数

所有可能结果的概率之和应该是1

进行试验时结果之一必定出现

则从样本空间的所有事件集合M到非负实数的函数

$p: M\to R$，称为试验的概率分布

均匀分布(uniform distribution)，对于有n个可能结果的样本空间S，有$n\in Z^+$

则对于任意结果 $s\in S$，$p(s)=1/n$

对于样本空间为S的事件$E\subseteq S$，其概率为E中结果的概率之和

即 $p(E)=\sum_{s\in E} p(s)$

特别地有，当事件E为无限集合时，

$\sum_{s\in E} p(s)$ 是一个收敛的无穷级数(convergent infinite series)

对于样本空间S中有限个或可数无穷个两两不相交事件(pairwise disjoint) $E_1, E_2, \cdots$

有 $p(\bigcup_i E_i)=\sum_i p(E_i)$

条件概率(conditional probability of E given F)，对于事件E,F且有$p(F)>0$

则给定条件F下E的条件概率记为 $p(E|F)$

且有 $P(E|F)=P(E\cap F)/P(F)$

独立事件(independent event)，对于事件E,F，

称事件E,F是独立的，当且仅当 $P(E\cap F)=P(E)P(F)$

当F为空事件时，$P(E\cap F)=0=P(E)P(F)$，则E,F独立平凡地为真

当$P(F)>0$时，有$P(E|F)=P(E\cap F)/P(F)=P(E)$

即直觉上独立事件相互间不影响发生概率

两两独立事件(pairwise independent event)，对于事件 $E_1,E_2,\cdots,E_n$

如果对于任意 $1\le i<j\le n$，有 $P(E_i\cap E_j)=P(E_i)P(E_j)$

则称事件 $E_1,E_2,\cdots,E_n$ 是两两独立的

Discrete Mathematics - P177

相互独立事件(mutually independent event)，对于事件 $E_1, E_2, \cdots, E_n$

如果对任意 $2 \le m \le n$，$1 \le i_1 < i_2 < \cdots < i_m \le n$

都有 $P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_m}) = P(E_{i_1})P(E_{i_2})\cdots P(E_{i_m})$

则称事件 $E_1, E_2, \cdots, E_n$ 是相互独立的

特别地有，为了证明 $E_1, E_2, \cdots, E_n$ 为相互独立事件

需要检查的条件有 $\sum_{m=2}^{n} \binom{n}{m}$ 个

于是有 $\sum_{m=2}^{n} \binom{n}{m} = \sum_{m=0}^{n} \binom{n}{m} - \binom{n}{1} - \binom{n}{0}$

$= 2^n - n - 1$

特别注意，$E_1, E_2, \cdots, E_n$ 相互独立 $\longrightarrow$ $E_1, E_2, \cdots, E_n$ 两两独立，但反之不成立

对于数字 2，3，5，30 中随机抽取 1 个数字 k

令事件 A 为 $2 \mid k$，事件 B 为 $3 \mid k$，事件 C 为 $5 \mid k$

则有 $P(A) = \frac{1}{2}$，$P(B) = \frac{1}{2}$，$P(C) = \frac{1}{2}$

且有 $P(AB) = \frac{1}{4}$，$P(BC) = \frac{1}{4}$，$P(AC) = \frac{1}{4}$，但 $P(ABC) = \frac{1}{4} \neq \frac{1}{8}$

于是 A, B, C 是两两独立的，但不是相互独立的

伯努利试验(Bernoulli trial)，指具有两种可能结果的试验

通常记为成功(success)和失败(failure)

二项分布(binomial distribution)，指对于成功率为 p，失败率为 $q = 1-p$ 的 n 次独立伯努利试验中

令函数 $b(k; n, p) = \binom{n}{k} p^k q^{n-k}$ 表示正好成功 k 次的概率

则称这个分布函数为二项分布

通常记作 $X \sim B(n, p)$

另外对于所有 $0 \le k \le n$，有 $\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = (p+1-p)^n = 1$

随机变量(random variable)，为一个从样本空间到实数集的函数 $f: S \to R$

对试验的每个结果指派一个实数值

注意随机变量并不是一个变量，而是一个函数

一个随机变量 X 在样本空间 S 中的分布

是指对于所有的 $r \in X(S)$，有序对 $(r, P(X=r))$

其中 $p(X=r)$ 是 X 取值为 r 的概率

通常是通过对每个 $r \in X(S)$ 指派 $p(X=r)$ 来描述分布

Discrete Mathematics - P178

碰撞(collision)，指对于hash函数，两个不同的关键字映射到相同的hash值

概率算法(probabilistic algorithm)，指在一步或多步中做出随机选择的算法

概率方法(probabilistic method)，指证明与集合中具有给定性质个体的有关结果的证明技巧

通过对个体指派概率，证明具有给定性质的个体的概率为正数

即对于一个给定的集合S，从中随机地选取一个元素

令事件E为该元素具有给定的性质，则$\bar{E}$为不具有给定性质

通过证明$\bar{E}$的概率$P(\bar{E})<1$，即不具有给定性质的概率小于1

则有$P(E)>0$，即集合S中存在具有给定性质的元素

对于整数$k \geq 2$，有拉姆齐数$R(k,k) \geq 2^{k/2}$

证明过程有，当$k=2$时，$R(2,2)=2 \geq 2^{2/2}$

当$k=3$时，$R(3,3)=6 \geq 2^{3/2}$

当$k \geq 4$时，考虑人数$n < 2^{k/2}$

令事件E表示n个人中存在k个人或者两两是朋友，或者两两是敌人

由于n个人中的k个人的组合共有$\binom{n}{k}$种选择，记为$S_1, S_2, \cdots, S_{\binom{n}{k}}$

则令事件E_i为组合S_i的k个人或者两两是朋友，或者两两是敌人

即有$E = \bigcup_{i=1}^{\binom{n}{k}} E_i$

根据假设有，对于n个人中的任意两个人，是朋友或是敌人的概率均为$\frac{1}{2}$

则在组合S_i的k个人中，有$\binom{k}{2}$个不同对的人

于是可知$Pr(E_i) = 2 \cdot (1/2)^{\binom{k}{2}} = 2 \cdot (1/2)^{k(k-1)/2}$

则根据布尔不等式有

$$Pr(E) = Pr(\bigcup_{i=1}^{\binom{n}{k}} E_i) \leq \sum_{i=1}^{\binom{n}{k}} Pr(E_i)$$
$$= \binom{n}{k} \cdot 2 \cdot \left(\frac{1}{2}\right)^{k(k-1)/2}$$

又对于整数$1 \leq k \leq n$，有$\binom{n}{k} \leq n^k / 2^{k-1}$

则$Pr(E) = \binom{n}{k} \cdot 2 \cdot \left(\frac{1}{2}\right)^{k(k-1)/2} \leq \frac{n^k}{2^{k-1}} \cdot 2 \cdot \left(\frac{1}{2}\right)^{k(k-1)/2}$

又根据假设$n < 2^{k/2}$

$$\text{则 } Pr(E) \leq \frac{n^k}{2^{k-1}} \cdot 2 \cdot \left(\frac{1}{2}\right)^{k(k-1)/2} < 2^{(k/2)k} \cdot 2^{-(k-1)} \cdot 2 \cdot 2^{-k(k-1)/2}$$
$$= 2^{k^2/2 - k + 1 + 1 - k^2/2 + k/2} = 2^{2-k/2} \leq 2^{2-4/2} = 1$$

即n个人中存在k个人或者两两是朋友，或者两两是敌人的概率小于1

即n个人中既不存在k个人两两是朋友，也不存在k个人两两是敌人的概率大于0

于是可知当$n < 2^{k/2}$时，无法满足拉姆齐数的条件，于是有$R(k,k) \geq 2^{k/2}$成立

Discrete Mathematics - P179

循环赛 (round-robin tournament)，对于具有m个游戏者的循环赛，其中 $m \geq 2$

每两个游戏者进行一次游戏，游戏的结果为其中一人获胜，另一人失败

对于正整数 $k<m$，考虑有m个游戏者的循环赛

这个循环赛的结果可能存在关于k的性质

对于每一个k个玩家的集合，存在1个玩家

这个玩家赢了这个集合中的k个玩家

并且假设任意两个玩家间的游戏，其中一方获胜都是等可能的

且不同游戏间的结果相互独立

则考虑对于给定的k，找到最小的m值使性质成立

令事件E为，对于每个包含k个玩家的集合，存在1个玩家赢了集合中的k个玩家

令事件F_i为对于第i个包含k个玩家的集合，其中 $i=1,2,\cdots,\binom{m}{k}$

不存在1个玩家赢了集合中的k个玩家

于是有 $\bar{E} = \bigcup_{i=1}^{\binom{m}{k}} F_i$

根据布尔不等式，有 $P(\bar{E}) = P(\bigcup_{i=1}^{\binom{m}{k}} F_i) \leq \sum_{i=1}^{\binom{m}{k}} P(F_i)$

对于任一事件F_i对应的k个玩家的集合S_i

其余m-k个人中的任何一个人赢了这k个玩家的概率为 $(\frac{1}{2})^k$

又所有游戏相互独立，则对于剩余的m-k个人

$P(F_i) = [1-(\frac{1}{2})^k]^{m-k} = (1-2^{-k})^{m-k}$

于是有 $P(\bar{E}) \leq \sum_{i=1}^{\binom{m}{k}} P(F_i) = \binom{m}{k}(1-2^{-k})^{m-k}$

对于给定的正整数k，

注意到 $\binom{m}{k} \in \Theta(m^k)$，

于是有当 $m \to +\infty$ 时，$\binom{m}{k}(1-2^{-k})^{m-k} \to 0$

即对于 $\varepsilon=1$，存在一个正整数 M

使得 $\forall m>M \ |\binom{m}{k}(1-2^{-k})^{m-k}|<1$

于是对于给定的正整数k，总能取到足够大的正整数m

使得 $P(\bar{E}) \leq \binom{m}{k}(1-2^{-k})^{m-k}<1$

即 $P(E)>0$，即存在一种循环赛结果具有性质的概率大于1

则通过给定的k值，可以得到m的最小值的一个上界

如对于 $k=1$，有 $m\cdot(\frac{1}{2})^{m-1}<1$

则有 $m \geq 3$

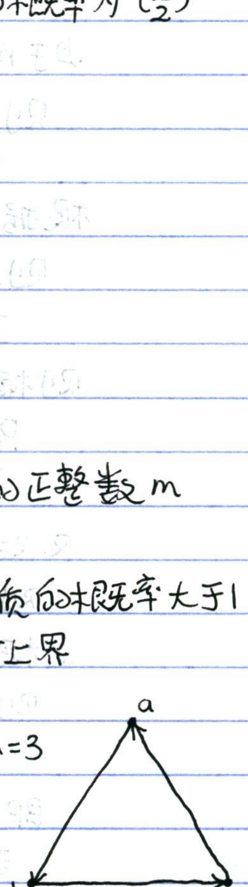
对于 $k=2$，有 $\frac{m(m-1)}{2}\cdot(\frac{3}{4})^{m-2}<1$

则有 $m \geq 21$

对于 $k=3$，有 $\frac{m(m-1)(m-2)}{6}\cdot(\frac{7}{8})^{m-3}<1$

则有 $m \geq 91$

$k=1, m=3$



Discrete Mathematics - P180

Ⅰ型错误 (type Ⅰ error, error of the first kind)

false negative: incorrect rejection of true null hypothesis

Ⅱ型错误 (type Ⅱ error, error of the second kind)

false positive: failure to reject false null hypothesis

判断 \ 真实	TRUE	FALSE
accept	correct (True Positive)	type Ⅱ error (False Positive)
reject	type Ⅰ error (False Negative)	correct (True Negative)

期望线性性质 (linearity), 期望运算符 (expected value operator) $E[\cdot]$ 是线性的 (linear)

对于在样本空间 S 上的随机变量 $X_1, X_2, \cdots, X_n$,

对于实数 a, b, 有 $E(aX+b) = aE(X)+b$

以及 $E(X_1+X_2+\cdots+X_n) = E(X_1)+E(X_2)+\cdots+E(X_n)$

证明过程有, $E(aX+b) = \sum_{s\in S} p(s)(aX(s)+b)$

$= a\sum_{s\in S} p(s)X(s) + b\sum_{s\in S} p(s)$

$= aE(X)+b$

基础步骤: 当 $n=2$ 时, $E(X_1+X_2) = \sum_{s\in S} p(s)(X_1(s)+X_2(s))$

$= \sum_{s\in S} p(s)X_1(s) + \sum_{s\in S} p(s)X_2(s)$

$= E(X_1)+E(X_2)$

递归步骤: 假设对于任意正整数 $n\geq 2$, $P(n)$ 为真, 则考虑 $P(n+1)$

$E(X_1+X_2+\cdots+X_n+X_{n+1}) = \sum_{s\in S} p(s)(X_1(s)+X_2(s)+\cdots+X_n(s)+X_{n+1}(s))$

$= \sum_{s\in S} p(s)(X_1(s)+X_2(s)+\cdots+X_n(s)) + \sum_{s\in S} p(s)X_{n+1}(s)$

$= E(X_1+X_2+\cdots+X_n) + E(X_{n+1})$

$\overset{(IH)}{=} E(X_1)+E(X_2)+\cdots+E(X_n)+E(X_{n+1})$

根据数学归纳法, 对于任意 $n\geq 2$, $E(X_1+X_2+\cdots+X_n) = E(X_1)+E(X_2)+\cdots+E(X_n)$

马尔可夫不等式 (Markov's inequality), 对于非负随机变量 X, 常数 $t>0$

有 $Pr\{X\geq t\} \leq E(X)/t$

证明过程有, 对于 X 的样本空间 S, 定义指示器变量 I

即 $I = \begin{cases} 1, & X\geq t \\ 0, & X<t \end{cases}$

于是有 $I \leq X/t$ 分别对于 $X\geq t$ 和 $X<t$ 的情形成立

则有 $Pr\{X\geq t\} = E(I) \leq E(X/t) = E(X)/t$

Discrete Mathematics - P181

平均情形下的计算复杂度，可以转化为计算一个随机变量的期望值

令样本空间S为可能的输入 $a_1, a_2, \cdots, a_n$

令随机变量X对输入 $a_j\ (j=1,2,\cdots,n)$ 的赋值为

当 a_j 作为输入时，算法所需的操作数

则平均情形的复杂度 $E(X) = \sum_{j=1}^{n} p(a_j)X(a_i)$

其中 $p(a_j)$ 为输入 a_j 出现的概率

逆序对　对于由n个不同的可比较的元素构成的排列中，考虑逆序对的期望

令指示器变量 $I_{i,j}$ 为在排列 $A[1..n]$ 中，第i个元素与第j个元素构成一个逆序对

即对于 $1 \le i < j \le n$，$I_{i,j} = \begin{cases} 1, & A[i] > A[j] \\ 0, & A[i] < A[j] \end{cases}$

则令随机变量X为给定排列 $A[1..n]$ 中的逆序对数量

于是有 $X = \sum_{1\le i<j\le n} I_{i,j}$

则有 $E(X) = E(\sum_{1\le i<j\le n} I_{i,j}) = \sum_{1\le i<j\le n} E(I_{i,j})$

又对于随机排列，元素 $A[i]$ 出现在 $A[j]$ 是等可能的

即有 $E(I_{i,j}) = Pr(I_{i,j}) = \frac{1}{2}$

则有 $E(X) = \sum_{1\le i<j\le n} E(I_{i,j}) = \sum_{1\le i<j\le n} \frac{1}{2}$

$= \binom{n}{2} \cdot \frac{1}{2} = n(n-1)/4$

插入排序　对于n个不同元素的随机排列应用插入排序，考虑比较次数的期望

令 $X_i\ (i=2,3,\cdots,n)$ 随机变量为当前 $i-1$ 个元素 $a_1, a_2, \cdots, a_{i-1}$ 已排序

将元素 a_i 插入正确位置所用的比较次数

则有比较次数的随机变量 $X = X_2 + X_3 + \cdots + X_n$

于是有 $E(X) = E(X_2+X_3+\cdots+X_n) = E(X_2)+E(X_3)+\cdots+E(X_n)$

考虑将 a_i 插入前 $i-1$ 个元素时，从小到大顺序比较

则当 a_i 是前i个元素中第k $(1\le k\le i)$ 大的元素时

$X_i(k) = k$，即使用k次比较

于是有 $X_i = X_i(1) + X_i(2) + \cdots + X_i(i)$

即 $E(X_i) = E(X_i(1)+X_i(2)+\cdots+X_i(i)) = \sum_{k=1}^{i} E(X_i(k))$

$= \sum_{k=1}^{i} \frac{1}{i}\cdot k = i(i+1)/2 \cdot \frac{1}{i} = (i+1)/2$

则有 $E(X) = \sum_{i=2}^{n} (i+1)/2 = \sum_{j=3}^{n} \frac{j}{2} = \sum_{j=1}^{n+1} \frac{j}{2} - \frac{1}{2}(1+2)$

$= \frac{(n+1)(n+2)}{4} - \frac{3}{2} = \frac{n^2+3n-4}{4}$

于是有插入排序的平均比较次数为 $\Theta(n^2)$ 的

Discrete Mathematics - P182

几何分布 (geometric distribution)，对于样本空间 $S = \mathbb{Z}^+$ 的离散概率分布

对于 $k = 1, 2, 3, \cdots$，$P\{X=k\} = (1-p)^{k-1}p$

其中 p 为每一次独立试验的成功概率，且有 $0<p<1$

则随机变量 X 可描述为获得第一次成功所需的试验次数

称随机变量 X 服从参数为 p 的几何分布，记为 $X \sim G(p)$

对于服从参数 $0<p<1$ 的几何分布的随机变量 $X \sim G(p)$，有 $E(X) = \frac{1}{p}$

$$E(X) = \sum_{k=1}^{\infty} k P\{X=k\} = \sum_{k=1}^{\infty} k(1-p)^{k-1}p$$
$$= p\sum_{k=1}^{\infty} k(1-p)^{k-1} = p \cdot \frac{1}{p^2} = \frac{1}{p}$$

比安内梅公式 (Bienaymé's formula)，对于两两独立的随机变量 $X_1, X_2, \cdots, X_n$

有 $V(X_1+X_2+\cdots+X_n) = V(X_1) + V(X_2) + \cdots + V(X_n)$

证明过程有，基础步骤：对于 $n=2$，X_1, X_2 是 S 上独立的随机变量

则 $V(X_1+X_2) = E[(X_1+X_2)^2] - [E(X_1+X_2)]^2$
$= E(X_1^2+2X_1X_2+X_2^2) - [E(X_1)+E(X_2)]^2$
$= E(X_1^2) + E(2X_1X_2) + E(X_2^2)$

由于 X_1, X_2 是独立随机变量 $\quad - E(X_1)^2 - 2E(X_1)E(X_2) - E(X_2)^2$

$E(X_1X_2) = E(X_1)E(X_2) \quad = [E(X_1^2) - E(X_1)^2] + [E(X_2^2) - E(X_2)^2]$
$= V(X_1) + V(X_2)$

递归步骤：对于样本空间 S 上两两独立的随机变量 $X_1, X_2, \cdots, X_n, X_{n+1}$

假设对于任意 $n \geq 2$ 有 $P(n)$ 为真，则考虑 $P(n+1)$

取随机变量 $Y = X_1+X_2+\cdots+X_n$

(注意这一步是不成立的) 则有随机变量 Y 与 X_{n+1} 是独立的

于是 $V(X_1+X_2+\cdots+X_n+X_{n+1}) = V(Y+X_{n+1})$
$= V(Y) + V(X_{n+1})$
$= V(X_1+X_2+\cdots+X_n) + V(X_{n+1})$
$(IH) = V(X_1) + V(X_2) + \cdots + V(X_n) + V(X_{n+1})$

则根据数学归纳法，对于两两独立的随机变量 $X_1, X_2, \cdots, X_n$

有 $V(X_1+X_2+\cdots+X_n) = V(X_1)+V(X_2)+\cdots+V(X_n)$

对于 n 次独立的成功率为 p 的伯努利试验，考虑成功次数的方差

令随机变量 X_i 当第 i 次试验成功时为 1，当第 i 次试验失败时为 0，$i=1,2,\cdots,n$

则有 $V(X) = V(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} V(X_n) = \sum_{i=1}^{n} p(1-p) = np(1-p)$