Probability and Statistics - P48

选票问题

考虑在美国总统选举中，如果候选人在一个州里获得的选票数最多

则该候选人赢得分配给该州的全部选举团票，

选举团票数正比于该州的人口数，即对于人口为n的州，票数为cn，其中c为常数

实际上约为cn+2，即众议员与参议员各有一张选票

考虑在一次区的选举中，一个选民的平均权力，

称为平均权力，即某个选民的一张选票起关键作用相关的量

假定共有2n+1个选民参与投票，共有2个候选人参选

注意 2n个选民的情形是类似的

如果对于除指定的选民之外的2n个选民中，2个候选人各有n个选民支持

则认为指定选民的选票是关键的

假定2n个选民的投票是相互独立，且支持2个候选人是等可能的

则P{来自人口2n+1的州的公民的选票起关键作用}

$$= \binom{2n}{n}\left(\frac{1}{2}\right)^n\left(\frac{1}{2}\right)^n = \frac{(2n)!}{n!\,n!\,2^{2n}}$$

又根据斯特林公式(Stirling's formula)有

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n = \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n}$$

于是 $P\{2n+1\} = \frac{(2n)!}{n!\,n!\,2^{2n}} \approx \left(\sqrt{2\pi}\,(2n)^{2n+\frac{1}{2}} e^{-2n}\right) / \left(\sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n}\right)^2 \cdot 2^{2n}$

$= 1/\sqrt{\pi n}$

由于选票起关键作用时影响 $c(2n+1)$ 张选举团票

于是平均权力 $P\{2n+1\}\cdot c(2n+1) \sim 2c\sqrt{n/\pi}$

可见平均权力与州的人口数的平方根成正比

即大州中的一票相比小州中的一票有更大的平均权力

泊松随机变量(Poisson random variable)，又称泊松分布(Poisson distribution)

对于参数$\lambda>0$，随机变量X服从参数为λ的泊松分布，记为$X \sim \text{Poisson}(\lambda)$

并有概率分布列 $p(i) = Pr\{X=i\} = e^{-\lambda}\frac{\lambda^i}{i!}$，其中 $i=0,1,2,\cdots$

注意由于 $e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3}x^3 + \cdots = \sum_{k=0}^{\infty}\frac{1}{k!}x^k$

则 $\sum_{i=0}^{\infty} p(i) = \sum_{i=0}^{\infty} e^{-\lambda}\frac{\lambda^i}{i!} = e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = e^{-\lambda}\cdot e^{\lambda} = 1$

泊松分布的期望 $E[X] = \sum_{i=0}^{\infty} i\cdot p(i) = \sum_{i=0}^{\infty} i\cdot e^{-\lambda}\frac{\lambda^i}{i!}$

$$= \sum_{i=0}^{\infty} e^{-\lambda}\frac{\lambda^i}{(i-1)!}$$

$$= \lambda e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^{i-1}}{(i-1)!}$$

$$= \lambda\cdot e^{-\lambda}\cdot e^{\lambda} = \lambda$$

即泊松分布的期望等于其参数λ

Probability and Statistics - P49

泊松随机变量

泊松分布的二阶矩，对于服从参数为λ的泊松分布的随机变量X，其中$\lambda>0$

则 $E[X^2] = \sum_{i=0}^{\infty} i^2 e^{-\lambda}\frac{\lambda^i}{i!} = \lambda\sum_{i=1}^{\infty} i e^{-\lambda}\frac{\lambda^{i-1}}{(i-1)!}$

$= \lambda\left(\sum_{i=1}^{\infty}(i-1)e^{-\lambda}\frac{\lambda^{i-1}}{(i-1)!} + \sum_{i=1}^{\infty} e^{-\lambda}\frac{\lambda^{i-1}}{(i-1)!}\right)$

$= \lambda\left(\sum_{i=0}^{\infty} i e^{-\lambda}\frac{\lambda^i}{i!} + \sum_{i=0}^{\infty} e^{-\lambda}\frac{\lambda^i}{i!}\right)$

$= \lambda(\lambda+1)$

于是服从参数为λ的泊松分布的方差为

$Var(X) = E[X^2] - (E[X])^2 = \lambda(\lambda+1) - \lambda^2 = \lambda$

即泊松分布的方差也等于其参数λ

考虑服从参数为n，p的二项分布的随机变量 $X \sim B(n,p)$

其中n充分大，p充分小，而np趋近于一个常数

即有 $n \to +\infty$，$p \to 0^+$，$np \to \lambda$，其中 $0<\lambda<+\infty$

则考虑随机变量X的概率分布列 $p(i) = Pr\{X=i\}$

有 $Pr\{X=i\} = \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{i!(n-i)!}\cdot p^i (1-p)^{n-i}$

$= n(n-1)\cdots(n-i+1)\cdot\frac{p^i}{i!}(1-p)^{n-i}$

$= \frac{n(n-1)\cdots(n-i+1)}{n^i}\cdot\frac{(np)^i}{i!}\cdot\frac{(1-p)^n}{(1-p)^i}$

注意当i足够大时，$n(n-1)\cdots(n-i+1)/n^i \to 0$

即 $Pr\{X=i\} \to 0$

而当i显著地小于n时，$n(n-1)\cdots(n-i+1)/n^i \to 1$

又 $p \to 0^+$，于是 $(1-p)^i \to 1$

而 $(1-p)^n = (1-p)^{\frac{1}{p}\cdot np} = [(1-p)^{1/p}]^{np}$

当 $p \to 0^+$ 时，有 $\frac{1}{p} \to +\infty$，$(1-p)^{1/p} \to e^{-1}$

于是有 $[(1-p)^{1/p}]^{np} \to e^{-np}$

则 $Pr\{X=i\} \to e^{-np}\cdot\frac{(np)^i}{i!} = e^{-\lambda}\frac{\lambda^i}{i!}$

此时随机变量X可以近似地看作服从参数为λ的泊松分布，即 $X \sim Poisson(\lambda)$

即当独立地进行n次成功率为p的伯努利试验时，

如果n充分大，p充分小，而np保持适中

则其结果的成功次数可以近似地看作服从参数为$\lambda=np$的泊松分布

再考虑其期望与方差，有 $E[X] = np \to \lambda$

又 $(1-p) \to 1$，则有 $Var[X] = np(1-p) \to \lambda$

与随机变量X近似地服从参数为λ的泊松分布的结论相符

于是对于实践中类似的随机变量X，可以通过经验确定成功次数期望λ

然后使用参数为λ的泊松分布代替二项分布

Probability and Statistics - P50

弱相依条件 (weak dependence), 已知对于参数为$\lambda=np$的泊松分布, 其中$0<\lambda<+\infty$

是对n次独立重复的成功率为p的试验中成功次数的较好近似

其中n充分大, 而p充分小, 且$\lambda=np$保持适中

实际中, 如果重复试验并不独立, 而是弱相依 (weakly dependent) 的

参数为$\lambda=np$的泊松分布仍是足够好的近似

对于配对问题, n个人随机地从n顶帽子中选择一顶, 考虑恰好拿到自己帽子的人数

令E_i表示第i次试验, 即第i个人拿到了自己的帽子, 其中$i=1,2,\cdots,n$

则有$Pr\{E_i\}=\frac{1}{n}$, 且有$Pr\{E_i|E_j\}=\frac{1}{n-1}$, 其中$i\neq j$

可知当$n\to+\infty$时, $\lim_{n\to+\infty} Pr\{E_i|E_j\}/P\{E_i\}=1$

即当n足够大时, E_i与E_j的相依性变弱

则可以近似地认为E_i和E_j是独立的

于是可以认为拿到自己帽子的人数可以近似地看作服从$\lambda=n\cdot\frac{1}{n}=1$的泊松分布

则有$Pr\{X=i\}=e^{-1}/i!$, 其中$i=0,1,2,\cdots$

错排的概率$Pr\{X=0\}=e^{-1}$, 与常规方法得到的结果一致

对于生日悖论问题, 考虑使至少存在两人生日相同的概率达到50%所需的人数

令E_{ij}表示对于$1\le i<j\le n$, 第i个人与第j个人的生日相同

假定一年的天数$n=365$, 则考虑$\binom{n}{2}$个试验E_{ij}之间的相关性

则有$Pr\{E_{ij}\}=\frac{1}{365}$, $Pr\{E_{ij}|E_{kl}\}=\frac{1}{365}$

$Pr\{E_{ij}|E_{jk}\}=\frac{1}{365}$, $Pr\{E_{jk}|E_{ij}E_{ik}\}=1$,

其中E_{ij}与E_{kl}有$1\le i<j\le n$, $1\le k<l\le n$, i, j, k, l各不相同

E_{ij}, E_{jk}, E_{ik}有$1\le i<j<k\le n$

于是可知虽然$\binom{n}{2}$个试验E_{ij}并不独立, 但相依性很弱

则可以认为生日相同的对数近似地服从参数为$\lambda=\frac{1}{365}\cdot\binom{n}{2}$的泊松分布

令$Pr\{没有2个人生日相同\}=Pr\{X=0\}\le\frac{1}{2}$

则$\exp\{-\frac{1}{365}\binom{n}{2}\}\le\frac{1}{2}$

$n(n-1)\ge 730\ln 2$

于是有使$Pr\{X=0\}\le\frac{1}{2}$的n为23, 与常规方法得到的结果一致

泊松范例 考虑n个事件, 第i个事件发生的概率为P_i, 其中$i=1,2,\cdots,n$

如果所有的事件或者相互独立, 或者至多弱相依

且n充分大, 而P_i充分小, 使得$\sum_{i=1}^{n}P_i$保持适中

则发生的事件个数可以看作近似地服从参数为$\lambda=\sum_{i=1}^{n}P_i$的泊松分布

Probability and Statistics - P51

游程

考虑n次抛掷硬币试验，假设各次抛掷结果相互独立，且正面向上的概率均为p

则过程中出现连续k次正面朝上的概率，其中 $k<n$，$0<p<1$

考虑利用泊松范例近似地得到这个概率

令事件H_i表示第$i, i+1, \cdots, i+k-1$次抛掷均为正面朝上，其中 $i=1,2,\cdots,n-k+1$

则有 $Pr\{H_i\} = p^k$，

注意当k充分大时，p^k趋近于0，但是H_i的发生次数仍无法近似地看作泊松分布

由于某些事件H_i, H_j之间的相关性很强.

如事件H_i已经发生，即第$i, i+1, \cdots, i+k-1$次均为正面向上

则对于事件H_{i+1}，可知 $Pr\{H_{i+1} | H_i\} = p >> Pr\{H_{i+1}\}$

可见事件H_i与H_{i+1}不满足弱相依条件

令事件E_i表示第$i, i+1, \cdots, i+k-1$次抛掷均为正面向上，且第$i+k$次抛掷为反面向上

则 $Pr\{E_i\} = p^k(1-p)$，$i=1,2,\cdots,n-k$

事件E_{n-k+1}表示 $n-k+1, n-k+2, \cdots, n$ 次抛掷均为正面向上

则 $Pr\{E_{n-k+1}\} = p^k$

可知对于$0<p<1$以及足够大的k，$Pr\{E_i\}$都充分小，即 $Pr\{E_i\} \to 0$

对于 $1 \le i, j \le n-k+1$，考虑事件E_i和E_j的独立性，其中 $i \ne j$

当E_i和E_j所涉及的试验没有重叠时，有 $Pr\{E_i | E_j\} = Pr\{E_i\}$

当E_i和E_j所涉及的试验有重叠时，有 $Pr\{E_i | E_j\} = 0$

当p^k充分小时，$Pr\{E_i | E_j\}$可以近似地看作$Pr\{E_i\}$

则令随机变量N表示事件$E_1, E_2, \cdots, E_{n-k+1}$中的事件发生次数

则N可以近似地看作服从参数为$E[N]$的泊松分布

有 $E[N] =$ [redacted] $= E[\sum_{i=1}^{n-k+1} I\{E_i\}]$

$= \sum_{i=1}^{n-k+1} Pr\{E_i\} = (n-k)p^k(1-p) + p^k$

于是N近似地服从参数为 $\lambda = (n-k)p^k(1-p) + p^k$ 的泊松分布

如果n次抛掷中没有长度至少为k的正面游程，则有$N=0$

于是 $Pr\{N=0\} \approx \exp\{-(n-k)p^k(1-p) - p^k\}$

令[redacted]随机变量L_n表示n次试验中连续正面的最大次数

即n次试验中的最大正面游程长度

可知当且仅当试验序列中没有k次连续正面向上时，有 $L_n<k$

即有 $Pr\{L_n<k\} = Pr\{N=0 | k\} \approx \exp\{-(n-k)p^k(1-p) - p^k\}$

Probability and Statistics - P52

游程

考虑n次抛掷硬币的试验，假设各次抛掷结果相互独立，且正面向上的概率均为p

令随机变量L_n表示n次试验中连续正面的最大次数，即最大正面游程长度

有 $\Pr\{L_n<k|p\} \approx \exp\{-(n-k)p^k(1-p)-p^k\}$

当硬币是均匀的，即 $p=\frac{1}{2}$

有 $\Pr\{L_n<k\} \approx \exp\{-(n-k)(\frac{1}{2})^k\frac{1}{2}-(\frac{1}{2})^k\} = \exp\{-\frac{n-k+2}{2^{k+1}}\}$

$= \exp\{-\frac{n}{2^{k+1}}\}\cdot\exp\{-\frac{k+2}{2^{k+1}}\}$

注意当k足够大时，$-\frac{k+2}{2^{k+1}}\approx 0$，即 $\exp\{-\frac{k+2}{2^{k+1}}\}\approx 1$

则有 $\Pr\{L_n<k\}\approx\exp\{\frac{-n}{2^{k+1}}\}$

假设n是2的幂次，并令 $k=\lg n+i$

则有 $\Pr\{L_n<k\}\approx\exp\{\frac{-n}{2^{\lg n+i+1}}\}=\exp\{\frac{-1}{2^{i+1}}\}$

$\Pr\{L_n=k\}=\Pr\{L_n<k+1\}-\Pr\{L_n<k\}$

$\approx\exp\{\frac{-1}{2^{i+2}}\}-\exp\{\frac{-1}{2^{i+1}}\}$

令事件E_i表示第$i,i+1,\cdots,i+k-1$次抛掷均为正面向上，且第$i+k$次抛掷为反面向上

事件E_{n-k+1}表示第$n-k+1,n-k+2,\cdots,n$次抛掷均为正面向上

则考虑连续k次正面的概率的精确表达式

$\Pr\{L_n\ge k\}=\Pr\{$n次抛掷中出现连续k个正面向上$\}=\Pr\{\bigcup_{i=1}^{n-k+1}E_i\}$

则根据事件与概率的容斥恒等式

则有 $\Pr\{\bigcup_{i=1}^{n-k+1}E_i\}=\sum_{r=1}^{n-k+1}(-1)^{r+1}\sum_{i_1<\cdots<i_r}\Pr\{E_{i_1}E_{i_2}\cdots E_{i_r}\}$

令S_i表示事件E_i所涉及的试验编号的集合，如$S_i=\{i,i+1,\cdots,i+k-1\}$

先考虑$E_1,E_2,\cdots,E_{n-k}$中r个事件交的概率，而暂不考虑E_{n-k+1}

即对于$\Pr\{E_{i_1}E_{i_2}\cdots E_{i_r}\}$，其中$1\le i_1<i_2<\cdots<i_r\le n-k+1$

则有 $\Pr\{E_{i_1}\cdots E_{i_r}\}=\begin{cases}0 & ,S_{i_1},\cdots,S_{i_r}\text{中存在相交的集合}\\ p^{rk}(1-p)^r & ,S_{i_1},\cdots,S_{i_r}\text{互不相交}\end{cases}$

考虑$1\le i_1<i_2<\cdots<i_r\le n-k$中使$S_{i_1},S_{i_2},\cdots,S_{i_r}$互不相交的组合数

注意每个集合中包含k+1次抛掷

而由于r个集合互不相交，则合计包含r(k+1)次抛掷

则其余n-r(k+1)次抛掷则分配在r个集合之外的r+1个位置上

可知使$S_{i_1},S_{i_2},\cdots,S_{i_r}$互不相交的组合数即对n-r(k+1)次抛掷的分配数

于是有 $\binom{n-r(k+1)+(r+1)-1}{(r+1)-1}=\binom{n-rk}{r}$

于是在不考虑E_{n-k+1}的情况下

$\sum_{i_1<i_2<\cdots<i_r\le n-k}\Pr\{E_{i_1}E_{i_2}\cdots E_{i_r}\}=\binom{n-rk}{r}p^{rk}(1-p)^r$

Probability and Statistics - P53

游程

考虑n次抛掷硬币的试验，假设各次抛掷结果相互独立，且正面向上的概率均为p

令随机变量L_n表示n次试验中连续正面的最大次数，即最大正面的游程长度

事件E_i表示$i, i+1, \cdots, i+k-1$次抛掷均为正面向上，且第$i+k$次抛掷为反面向上

事件E_{n-k+1}表示第$n-k+1, n-k+2, \cdots, n$次抛掷均为正面向上

则 $Pr\{L_n \ge k\} = \sum_{r=1}^{n-k+1} (-1)^{r+1} \sum_{i_1<\cdots<i_r} Pr\{E_{i_1}E_{i_2}\cdots E_{i_r}\}$

对于$Pr\{E_{i_1}E_{i_2}\cdots E_{i_r}\}$，其中$1 \le i_1 < i_2 < \cdots < i_r \le n-k$

有 $Pr\{E_{i_1}E_{i_2}\cdots E_{i_r}\} = \begin{cases} 0 & , S_{i_1}, S_{i_2}, \cdots, S_{i_r}\text{中存在相交的集合} \\ p^{rk}(1-p)^r & , S_{i_1}, S_{i_2}, \cdots, S_{i_r}\text{互不相交} \end{cases}$

其中S_i表示事件E_i所涉及的试验编号集合

有对于$1 \le i_1 < i_2 < \cdots < i_r \le n-k$的情况下，即不考虑事件$E_{n-k+1}$

有 $\sum_{1 \le i_1 < i_2 < \cdots < i_r \le n-k} Pr\{E_{i_1}E_{i_2}\cdots E_{i_r}\} = \binom{n-rk}{r} p^{rk}(1-p)^r$

于是再考虑 $1 \le i_1 < i_2 < \cdots < i_{r-1} \le n-k$，$i_r = n-k+1$的情形

类似地有 $Pr\{E_{i_1}E_{i_2}\cdots E_{i_r}\} = \begin{cases} 0 & , S_{i_1}, S_{i_2}, \cdots, S_{i_r}\text{中存在相交的集合} \\ p^{rk}(1-p)^{r-1} & , S_{i_1}, S_{i_2}, \cdots, S_{i_r}\text{互不相交} \end{cases}$

其中$i_r = n-k+1$

则考虑$S_{i_1}, S_{i_2}, \cdots, S_{i_{r-1}}, S_{n-k+1}$不相交的组合数

由于$S_{i_1}, S_{i_2}, \cdots, S_{i_{r-1}}, S_{n-k+1}$互不相交，所以共包含$(r-1)(k+1)+k$次抛掷

则其余$n-(r-1)(k+1)-k$次抛掷分配在前$r-1$个集合之外的r个位置上

可知$S_{i_1}, S_{i_2}, \cdots, S_{i_{r-1}}, S_{n-k+1}$的组合数即对$n-(r-1)(k+1)-k$次抛掷的分配数

于是有 $\binom{n-(r-1)(k+1)-k+r-1}{r-1} = \binom{n-rk}{r-1}$

即 $\sum_{1 \le i_1 < i_2 < \cdots < i_{r-1} \le n-k} Pr\{E_{i_1}E_{i_2}\cdots E_{i_{r-1}}E_{n-k+1}\} = \binom{n-rk}{r-1} p^{rk}(1-p)^{r-1}$

于是有
$$\begin{aligned} Pr\{L_n \ge k\} &= \sum_{r=1}^{n-k+1} (-1)^{r+1} \sum_{1 \le i_1 < i_2 < \cdots < i_r \le n-k+1} Pr\{E_{i_1}E_{i_2}\cdots E_{i_r}\} \\ &= \sum_{r=1}^{n-k+1} (-1)^{r+1} \left[\binom{n-rk}{r} p^{rk}(1-p)^r + \binom{n-rk}{r-1} p^{rk}(1-p)^{r-1}\right] \\ &= \sum_{r=1}^{n-k+1} (-1)^{r+1} \left[(1-p)\binom{n-rk}{r} + \binom{n-rk}{r-1}\right] p^{rk}(1-p)^{r-1} \end{aligned}$$

考虑用于计算$Pr\{L_n \ge k\}$的递归式，令事件A_n表示$L_n \ge k$，其中$k \le n$，$P_n = Pr\{A_n\}$

F_j表示n次抛掷中的第一次反面向上出现在第j次抛掷，其中$j = 1, 2, \cdots, k$

H表示n次抛掷中的前k次抛掷均为正面向上

注意到$F_1, F_2, \cdots, F_k$，H形成了一个互不相容的完备组，即事件组中有且仅有一个发生

则 $P_n = Pr\{A_n\} = \sum_{j=1}^{k} Pr\{A_n | F_j\} Pr\{F_j\} + Pr\{A_n | H\} Pr\{H\}$

由于 $Pr\{A_n | F_j\} = P_{n-j}$，$Pr\{F_j\} = p^{j-1}(1-p)$，$Pr\{A_n | H\} = 1$，$Pr\{H\} = p^k$

于是有 $P_n = \begin{cases} 0 & , n<k \\ p^k & , n=k \\ \sum_{j=1}^{k} P_{n-j}\, p^{j-1}(1-p) + p^k & , n>k \end{cases}$

可以通过动态规划在实际计算中提升计算速度

Probability and Statistics - P54

泊松范例　考虑事件发生在一列随机时间点上，并假设存在正常数λ使得

①在任意长度为h的时间区间内，恰好发生一个事件的概率彼此相同

均等于$\lambda h + o(h)$，其中$o(h)$表示任意满足$\lim_{h\to 0} f(h)/h = 0$的函数$f(x)$

②在任意长度为h的时间区间内，发生至少2个事件的概率非常小，等于$o(h)$

③对于任意确定的正整数n与非负整数$j_1, j_2, \cdots, j_n$以及n个互不相交的时间区间

若以事件E_i表示在第i个时间区间内事件正好发生了j_i次

则事件$E_1, E_2, \cdots, E_n$相互独立

通常而言条件①和②说明当h较小时，在长度为h的时间区间内

正好发生1个事件的概率等于λh加上一个远小于h的量

发生多于1个事件的概率等于一个远小于h的量

而条件③说明在一个时间区间内不论发生了什么

从概率意义上看对其他不相交的区间没有影响

考虑区间$[0,t]$，令随机变量$N(t)$表示这个区间内事件发生的次数

取正整数n，将区间$[0,t]$等分为n个互不相交且时间为$\frac{t}{n}$的子区间，则有$h=\frac{t}{n}$

则$Pr\{N(t)=k\} = Pr\{n$个子区间中的k个恰好发生1个事件且其余$n-k$个子区间发生0个事件$\}$

$+ Pr\{N(t)=k$且至少1个子区间发生多于1个事件$\}$

注意到$Pr\{N(t)=k$且至少1个子区间发生多于1个事件$\} \le Pr\{$至少1个子区间发生多于1个事件$\}$

$= Pr\{\bigcup_{i=1}^n$第i个子区间发生多于1个事件$\}$

$\le \sum_{i=1}^n Pr\{\bigcup_{i=1}^n$第$i$个子区间发生多于1个事件$\}$

$= \sum_{i=1}^n o(\frac{t}{n}) = n\cdot o(\frac{t}{n}) = t[o(\frac{t}{n})/\frac{t}{n}] = t\frac{o(h)}{h}$

于是当$n\to+\infty$时，$h=\frac{t}{n}\to 0$，则$Pr\{N(t)=k$且至少1个子区间发生多于1个事件$\}\to 0$

又$Pr\{n$个子区间中的k个恰好发生1个事件且其余$n-k$个子区间发生0个事件$\}$

$= \binom{n}{k}[\frac{\lambda t}{n}+o(\frac{t}{n})]^k[1-\frac{\lambda t}{n}-o(\frac{t}{n})-o(\frac{t}{n})]^{n-k}$

$= \frac{n(n-1)\cdots(n-k+1)}{n^k}\cdot\frac{n^k[\lambda t/n+o(t/n)]^k}{k!}\cdot\frac{[1-\lambda t/n-o(t/n)]^n}{[1-\lambda t/n-o(t/n)]^k}$

当$n\to+\infty$时，$h=\frac{t}{n}\to 0$，k远小于n，则有$\frac{n(n-1)\cdots(n-k+1)}{n^k}\to 1$

$n[\frac{\lambda t}{n}+o(\frac{t}{n})] = \lambda t + t[o(\frac{t}{n})/\frac{t}{n}] \to \lambda t$，

$[1-\frac{\lambda t}{n}-o(\frac{t}{n})]^n = [1-\frac{\lambda t}{n}-o(\frac{t}{n})]^{\frac{n}{\lambda t}\cdot\lambda t}\to e^{-\lambda t}$，$[1-\frac{\lambda t}{n}-o(\frac{t}{n})]^k\to 1$

则$Pr\{n$个子区间中的k个恰好发生1个事件且其余$n-k$个子区间发生0个事件$\}\to\frac{(\lambda t)^k}{k!}\cdot e^{-\lambda t}$

于是有$\lim_{n\to+\infty} Pr\{N(t)=k\} = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$，其中$k=0,1,2,\cdots$

即如果事件满足条件①②③，则在任意固定的长度为t的时间区间内

事件发生次数$N(t)$近似地服从参数为λt的泊松分布

称事件为按强度为λ的泊松过程发生的

常数λ为由经验确定的单位时间内事件发生的强度

Probability and Statistics - P55

泊松分布

对于服从参数为λ的泊松分布的随机变量X，有概率分布列 $\Pr\{X=k\} = e^{-\lambda}\frac{\lambda^k}{k!}$

则考虑无穷序列 $\{P(k)\}$，其中 $P(k) = \Pr\{X=k\} = e^{-\lambda}\frac{\lambda^k}{k!}$，$k=0,1,2,\cdots$

首先有 $P(0) = e^{-\lambda}$，则对于 $k>0$ 有

$$P(k)/P(k-1) = e^{-\lambda}\frac{\lambda^k}{k!} \Big/ e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!} = \frac{\lambda}{k}$$

于是可知无穷序列 $\{P(k)\}$ 有初始条件 $P(0) = e^{-\lambda}$

且满足递归关系 $P(k) = \frac{\lambda}{k}P(k-1)$，$k>0$

另外可知当 $k\leq\lambda$ 时，$P(k)\geq P(k-1)$，即序列 $\{P(k)\}$ 单调递增

而当 $k>\lambda$ 时，$P(k)<P(k-1)$，即序列 $\{P(k)\}$ 单调递减

当 $k=\lfloor\lambda\rfloor$ 时，$P(k)$ 取得最大值

于是可知泊松分布的 mode 即不大于λ的最大整数

考虑服从参数为λ的泊松分布的随机变量X的n阶矩 $E[X^n]$

有 $E[X^n] = \sum_{k=0}^{\infty} k^n\cdot e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} k^{n-1}e^{-\lambda}\frac{\lambda^k}{(k-1)!}$

$$= \lambda\cdot\sum_{k=0}^{\infty}(k+1)^{n-1}e^{-\lambda}\frac{\lambda^k}{k!}$$

$$= \lambda E[(X+1)^{n-1}]$$

于是无穷序列 $\{E[X^n]\}$ 有初始条件 $E[X] = \lambda$

且满足递归关系 $E[X^n] = \lambda E[(X+1)^{n-1}]$，$n>1$

考虑n次独立重复抛掷硬币，正面向上的概率为 $0<p<1$，则正面向上为偶数的概率

令反面向上的概率 $q = 1-p$，正面向上的次数X服从二项分布 $B(n,p)$

则有 $\Pr\{X\text{为偶数}\} = \sum_{k=0}^{\lfloor n/2\rfloor}\Pr\{X=2k\} = \sum_{k=0}^{\lfloor n/2\rfloor}\binom{n}{2k}p^{2k}q^{n-2k}$

$$= \frac{1}{2}\Big[\sum_{k=0}^{\lfloor n/2\rfloor}\binom{n}{2k}p^{2k}q^{n-2k} + \sum_{k=0}^{\lfloor n/2\rfloor}\binom{n}{2k+1}p^{2k+1}q^{n-(2k+1)}$$

$$+ \sum_{k=0}^{\lfloor n/2\rfloor}\binom{n}{2k}p^{2k}q^{n-2k} - \sum_{k=0}^{\lfloor n/2\rfloor}\binom{n}{2k+1}p^{2k+1}q^{n-(2k+1)}\Big]$$

$$= \frac{1}{2}\Big[\sum_{k=0}^{n}\binom{n}{k}p^kq^{n-k} + \sum_{k=0}^{\lfloor n/2\rfloor}\binom{n}{2k}(-p)^{2k}q^{n-2k} + \sum_{k=0}^{\lfloor n/2\rfloor}\binom{n}{2k+1}(-p)^{2k+1}q^{n-(2k+1)}\Big]$$

$$= \frac{1}{2}\Big[\sum_{k=0}^{n}\binom{n}{k}p^kq^{n-k} + \sum_{k=0}^{n}\binom{n}{k}(-p)^kq^{n-k}\Big]$$

$$= \frac{1}{2}[(p+q)^n + (q-p)^n]$$

$$= \frac{1}{2}[1+(1-2p)^n]$$

注意到当 $p\to0$，$n\to+\infty$ 且 $\lambda=np$ 保持适中时

随机变量X近似地服从$\lambda=np$为参数的泊松分布

则 $\Pr\{X\text{为偶数}\} = \frac{1}{2}[1+(1-2p)^n] = \frac{1}{2}[1+(1-2p)^{\frac{1}{2p}\cdot2pn}]$

又 $\lim_{p\to0}(1-2p)^{\frac{1}{2p}} = e^{-1}$

于是当随机变量X服从$\lambda=np$为参数的泊松分布时

$\Pr\{X\text{为偶数}\} = \frac{1}{2}[1+e^{-2\lambda}]$

Probability and Statistics - P56

考虑n次抛掷硬币的试验，假设各次抛掷结果相互独立，且正面向上的概率为p

事件 L_k 表示试验序列出现连续k次正面向上

则通过递推关系得到 $\Pr\{L_k|n\}$

令概率 P_n 表示长度为n的试验序列中出现连续k次正面向上的概率

则对于给定的正整数k，有无穷序列 $\{P_n\}$, $n=0,1,2,\cdots$

有初始条件 $P_0=P_1=P_2=\cdots=P_{k-1}=0$, $P_k=p^k$

再考虑无穷序列的递推关系.

根据如果长度为n的试验序列中出现连续k次正面向上（第一次），其中第k次正面向上的位置

如果不在第n次抛掷，则等同于在长度为n-1的序列中出现了连续k次正面向上

于是有概率为 P_{n-1}

如果在第n次抛掷，则其 $n-k+1, n-k+2, \cdots, n$ 次抛掷均为正面向上

而在前n-k次抛掷中没有出现k次正面向上

于是有概率为 $p^k(1-P_{n-k})$

于是可知满足递推关系 $P_n = P_{n-1} + p^k(1-P_{n-k})$

于是可知对于给定的正整数k，无穷序列 $\{P_n\}$

满足初始条件 $P_0=P_1=\cdots=P_{k-1}=0$, $P_k=p^k$

递推关系 $P_n = P_{n-1}+p^k(1-P_{n-k})$, $n>k$

对于 $n\in N$，有 $\sum_{k=0}^{n} e^{-\lambda}\frac{\lambda^k}{k!} = \frac{1}{n!}\int_\lambda^\infty e^{-x}x^n dx$

证明过程有，
$$\begin{aligned}
\frac{1}{n!}\int_\lambda^\infty e^{-x}x^n dx &= \frac{1}{n!}\int_\lambda^\infty -x^n de^{-x} \\
&= \frac{1}{n!}\left[-e^{-x}x^n\Big|_\lambda^\infty - \int_\lambda^\infty e^{-x} d(-x^n)\right] \\
&= \frac{1}{n!}e^{-\lambda}\lambda^n + \frac{1}{(n-1)!}\int_\lambda^\infty e^{-x}x^{n-1}dx \\
&\cdots \\
&= \frac{1}{n!}e^{-\lambda}\lambda^n + \frac{1}{(n-1)!}e^{-\lambda}\lambda^{n-1} + \cdots + \frac{1}{1!}e^{-\lambda}\lambda + \frac{1}{0!}\int_\lambda^\infty e^{-x}dx \\
&= \sum_{k=0}^{n} e^{-\lambda}\frac{\lambda^k}{k!}
\end{aligned}$$

几何随机变量(geometric random variable)，

考虑独立重复的伯努利试验，每次试验成功率为 $0<p<1$，重复试验直到首次成功

令随机变量X表示需要的试验次数，

则有 $\Pr\{X=k\} = (1-p)^{k-1}p$，其中 $k=1,2,3,\cdots$

考虑 $\sum_{k=1}^{\infty}\Pr\{X=k\} = p\sum_{k=1}^{\infty}(1-p)^{k-1} = p\sum_{k=0}^{\infty}(1-p)^k$
$$= p\cdot\frac{1}{1-(1-p)} = 1$$

即试验最终出现一次成功的概率为1

Probability and Statistics - P57

几何分布

考虑在坛子中有 N 个白球和 M 个黑球，重复有放回地取球，直到取出一个黑球

令随机变量 X 表示 [redacted] 取球次数 [redacted]

则 $Pr\{X=n\} = \left(\frac{N}{M+N}\right)^{n-1} \cdot \frac{M}{M+N} = MN^{n-1}/(M+N)^n$ ，$n \in N^+$

取 $p = \frac{M}{M+N}$ ，则有 $(1-p) = \frac{N}{M+N}$

于是可知 $Pr\{X=n\} = (1-p)^{n-1}p$

即随机变量 X 服从概率为 $p=\frac{M}{M+N}$ 的几何分布

再有 $Pr\{$至少取球 k 次$\} = Pr\{X \geq k\}$ ，其中 $k \in N^+$

$$= \frac{M}{M+N} \cdot \sum_{n=k}^{\infty} \left(\frac{N}{M+N}\right)^{n-1}$$

$$= \frac{M}{M+N} \cdot \left(\frac{N}{M+N}\right)^{n-1} / \left[1 - \frac{N}{M+N}\right]$$

$$= \left(\frac{N}{M+N}\right)^{n-1} = (1-p)^{n-1}$$

即至少取球 k 次当且仅当前 $k-1$ 次取球均为白球

对于服从概率为 p 的几何分布的随机变量 X

期望 $E[X] = \sum_{k=1}^{\infty} k(1-p)^{k-1}p$

$$= \sum_{k=1}^{\infty} (k-1+1)(1-p)^{k-1}p$$

$$= \sum_{k=1}^{\infty} (k-1)(1-p)^{k-1}p + \sum_{k=1}^{\infty} (1-p)^{k-1}p$$

$$= (1-p)\sum_{k=0}^{\infty} k(1-p)^{k-1}p + 1$$

$$= (1-p)\sum_{k=1}^{\infty} k(1-p)^{k-1}p + 1$$

$$= (1-p)E[X] + 1$$

于是有 $E[X] = \frac{1}{p}$

二阶矩 $E[X^2] = \sum_{k=1}^{\infty} k^2(1-p)^{k-1}p$

$$= \sum_{k=1}^{\infty} (k-1+1)^2(1-p)^{k-1}p$$

$$= \sum_{k=1}^{\infty} (k-1)^2(1-p)^{k-1}p + 2\sum_{k=1}^{\infty}(k-1)(1-p)^{k-1}p + \sum_{k=1}^{\infty}(1-p)^{k-1}p$$

$$= \sum_{k=0}^{\infty} k^2(1-p)^k p + 2\sum_{k=0}^{\infty} k(1-p)^k p + 1$$

$$= (1-p)E[X^2] + 2(1-p)E[X] + 1$$

于是 $pE[X^2] = \frac{2(1-p)}{p} + 1$

$$E[X^2] = \frac{2-p}{p^2}$$

则方差 $Var[X] = E[X^2] - (E[X])^2$

$$= \frac{2-p}{p^2} - \frac{1}{p^2}$$

$$= \frac{1-p}{p^2}$$

于是有服从概率为 p 的几何分布的随机变量 X

期望 $E[X] = \frac{1}{p}$ ，方差 $Var[X] = \frac{(1-p)}{p^2}$

Probability and Statistics - P58

几何分布

对于服从概率为p的几何随机变量X，有 $Pr\{X=n+k \mid X>n\} = Pr\{X=k\}$，其中 $n\in N$，$k\in N^+$

证明过程有，$Pr\{X=n+k \mid X>n\} = Pr\{X=n+k \wedge X>n\} / Pr\{X>n\}$

$$= Pr\{X=n+k\} / Pr\{X\geq n+1\}$$

$$= (1-p)^{n+k-1}p / (1-p)^{n+1-1}$$

$$= (1-p)^{k-1}p = Pr\{X=k\}$$

于是也可以描述为几何分布具有无记忆性

对于服从参数为 (n,p) 的二项随机变量 X

有 $E[\frac{1}{1+X}] = \frac{1-(1-p)^{n+1}}{(n+1)p}$

证明过程有，$E[\frac{1}{1+X}] = \sum_{k=0}^{n} \frac{1}{1+k}\binom{n}{k}p^k(1-p)^{n-k}$

$$= \sum_{k=0}^{n} \frac{1}{1+k}\cdot\frac{n!}{(n-k)!\,k!}\cdot p^k(1-p)^{n-k}$$

$$= \sum_{k=0}^{n} \frac{1}{1+n}\cdot\frac{(n+1)!}{(n-k)!\,(k+1)!}\cdot p^k(1-p)^{n-k}$$

$$= \frac{1}{(1+n)p}\sum_{k=0}^{n}\binom{n+1}{k+1}p^{k+1}(1-p)^{(n+1)-(k+1)}$$

$$= \frac{1}{(1+n)p}\left[\sum_{k=0}^{n+1}\binom{n+1}{k}p^k(1-p)^{(n+1)-k} - \binom{n+1}{0}p^0(1-p)^{n+1}\right]$$

$$= \frac{1-(1-p)^{n+1}}{(1+n)p}$$

对于服从概率p的几何随机变量X

有 $E[\frac{1}{X}] = \frac{-p\ln(p)}{(1-p)}$

证明过程有，$E[\frac{1}{X}] = \sum_{n=1}^{\infty}\frac{1}{n}(1-p)^{n-1}p$

$$= \frac{p}{1-p}\sum_{n=1}^{\infty}\frac{1}{n}(1-p)^n$$

注意 对于实数 $a>0$，可以将 $\frac{1}{n}a^n$ 转换为积分形式

即 $E[\frac{1}{X}] = \frac{p}{1-p}\sum_{n=1}^{\infty}\frac{1}{n}(1-p)^n$

$$= \frac{p}{1-p}\sum_{n=1}^{\infty}\int_0^{1-p}x^{n-1}dx$$

又由于积分与求和的顺序可以交换

则 $E[\frac{1}{X}] = \frac{p}{1-p}\sum_{n=1}^{\infty}\int_0^{1-p}x^{n-1}dx$

$$= \frac{p}{1-p}\int_0^{1-p}\sum_{n=1}^{\infty}x^{n-1}dx$$

$$= \frac{p}{1-p}\int_0^{1-p}\sum_{n=0}^{\infty}x^{n}dx$$

又在积分区间 $[0,1-p]$ 内有 $|x|<1$

于是 $E[\frac{1}{X}] = \frac{p}{1-p}\int_0^{1-p}\sum_{n=0}^{\infty}x^n dx$

$$= \frac{p}{1-p}\int_0^{1-p}\frac{1}{1-x}dx$$

$$= \frac{p}{1-p}\left[-\ln(1-x)\right]\Big|_0^{1-p}$$

$$= \frac{-p\ln p}{1-p}$$