Probability and Statistics - P42

假设有两个坛子，坛子中即有白球也有黑球

从两个坛子取出白球的概率分别为 p 和 p'，其中 $0 < p, p' < 1$

以如下方式连续地有放回地取球

起始以概率 α 从概率 p 的坛子取球，以概率 $1-\alpha$ 从概率 p' 的坛子取球

如果某一步取出的是白球，则在放回后从同一个坛子取球

如果某一步取出的是黑球，则在放回后从另一个坛子取球

则令 α_n 为第 n 次从概率 p 的坛子（第一个坛子）取球的概率，其中 $n = 1, 2, \cdots$

于是有初始条件 $\alpha_1 = \alpha$

令事件 E_n 表示第 n 次从第一个坛子取球，事件 H 表示取出白球

于是有 $P(E_1) = \alpha$，$P(E_1^c) = 1-\alpha$，$P(H|E_n) = p$，$P(H|E_n^c) = p'$

由于第 $n+1$ 次在第一个坛子取球取决于第 n 次取球的结果

$$P(E_{n+1}) = P(H|E_n)P(E_n) + P(H^c|E_n^c)P(E_n^c)$$

即有

$$\begin{aligned}\alpha_{n+1} &= p\alpha_n + (1-p')(1-\alpha_n) \\ &= \alpha_n(p+p'-1) + (1-p')\end{aligned}$$

于是可知 α_n 满足初始条件 $\alpha_1 = \alpha$

和递推关系 $\alpha_n = \alpha_{n-1}(p+p'-1) + (1-p')$，$n > 1$

注意到 α_n 有显示公式 $\alpha_n = \frac{1-p'}{2-p-p'} + (\alpha - \frac{1-p'}{2-p-p'})(p+p'-1)^{n-1}$

基础步骤：当 $n=1$ 时，$\alpha_n = \frac{1-p'}{2-p-p'} + (\alpha - \frac{1-p'}{2-p-p'})(p+p'-1)^{1-1} = \alpha$

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

$$\begin{aligned}\alpha_{n+1} &= \alpha_n(p+p'-1) + (1-p') \\ &\overset{(IH)}{=} \left[\frac{1-p'}{2-p-p'} + (\alpha - \frac{1-p'}{2-p-p'})(p+p'-1)^{n-1}\right](p+p'-1) + (1-p') \\ &= (\alpha - \frac{1-p'}{2-p-p'})(p+p'-1)^n + \frac{1-p'}{2-p-p'}(p+p'-1) + (1-p') \\ &= (\alpha - \frac{1-p'}{2-p-p'})(p+p'-1)^n + \frac{(1-p')(p+p'-1) + (1-p')(2-p-p')}{2-p-p'} \\ &= (\alpha - \frac{1-p'}{2-p-p'})(p+p'-1)^n + \frac{1-p'}{2-p-p'}\end{aligned}$$

即 $P(n+1)$ 为真

于是根据数学归纳法，对于任意 $n \in N^+$，有 $\alpha_n = (\alpha - \frac{1-p'}{2-p-p'})(p+p'-1)^{n-1} + \frac{1-p'}{2-p-p'}$

令 P_n 表示第 n 次取出白球的概率

则有 $P_n = \alpha_n \cdot p + (1-\alpha_n) \cdot p'$

考虑当 n 足够多大时，α_n 和 P_n 的取值

由于 $0 < p, p' < 1$，则有 $|p+p'-1| < 1$

于是

$$\begin{aligned}\lim_{n\to\infty} \alpha_n &= (\alpha - \frac{1-p'}{2-p-p'}) \cdot \lim_{n\to\infty}(p+p'-1)^{n-1} + \frac{1-p'}{2-p-p'} \\ &= (\alpha - \frac{1-p'}{2-p-p'}) \cdot 0 + \frac{1-p'}{2-p-p'} = \frac{1-p'}{2-p-p'}\end{aligned}$$

$$\begin{aligned}\lim_{n\to\infty} P_n &= \lim_{n\to\infty}\alpha_n \cdot p + p' \cdot \lim_{n\to\infty}(1-\alpha_n) = p \cdot \frac{1-p'}{2-p-p'} + p' \cdot (1 - \frac{1-p'}{2-p-p'}) \\ &= \frac{p+p'-2pp'}{2-p-p'}\end{aligned}$$

注意在次数足够多大时，α_n 与 P_n 都独立于初始条件 $\alpha_1 = \alpha$

Probability and Statistics - P43

随机变量 (random variable)，指定义在样本空间的实值函数

优惠券收集 假设有N种不同的优惠券，每种优惠券都以相同的概率被收集到

每次收集得到一张优惠券，且假设每次收集是相互独立的

如果想收集到一套全套N种优惠券，考虑收集全时收集到的优惠券的总张数

记总张数为随机变量T，则考虑概率$P\{T=n\}$

对于给定的正整数n，首先考虑概率$P\{T>n\}$

令事件A_j表示前n张优惠券中没有第j种优惠券，其中$j=1,2,\cdots,N$

则根据容斥原理 $P\{T>n\}=P(\bigcup_{j=1}^{N}A_j)$ $\quad + \cdots + (-1)^{N+1}P(A_1\cdots A_N)$

$$=\sum_j P(A_j)-\sum_{j_1<j_2}P(A_{j_1}A_{j_2})+\cdots+(-1)^{k+1}\sum_{j_1<\cdots<j_k}P(A_{j_1}\cdots A_{j_k})$$

又对于给定的$k=1,2,\cdots,N$，$j_1,j_2,\cdots,j_k$有$\binom{N}{k}$种选择方式

且对于给定的$j_1,j_2,\cdots,j_k$，对应于事件$A_{j_1},A_{j_2},\cdots,A_{j_k}$

$$P(A_{j_1}A_{j_2}\cdots A_{j_k})=\left(\frac{N-k}{N}\right)^n$$

于是 $P\{T>n\}=\binom{N}{1}\left(\frac{N-1}{N}\right)^n-\binom{N}{2}\left(\frac{N-2}{N}\right)^n+\cdots+(-1)^{N+1}\binom{N}{N}\left(\frac{N-N}{N}\right)^n$

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

另外对于$1\le n<N$，则有$P\{T>n\}=1$

又 $P\{T>n-1\}=P\{T\ge n\}=P\{T=n\}+P\{T>n\}$

于是 $P\{T=n\}=P\{T>n-1\}-P\{T>n\}$

考虑$n=N$时，有$P\{T=N\}=N!\cdot\left(\frac{1}{N}\right)^N$，$P\{T>N-1\}=1$

则有 $N!\cdot\left(\frac{1}{N}\right)^N=1-\sum_{k=1}^{N}(-1)^{k+1}\binom{N}{k}\left(\frac{N-k}{N}\right)^N$

$$=(-1)^0\binom{N}{0}\left(\frac{N-0}{N}\right)^N+\sum_{k=1}^{N}(-1)^k\binom{N}{k}\left(\frac{N-k}{N}\right)^N$$

$$=\left(\frac{1}{N}\right)^N\cdot\sum_{k=0}^{N}(-1)^k\binom{N}{k}(N-k)^N$$

于是有 $N!=\sum_{k=0}^{N}(-1)^k\binom{N}{k}(N-k)^N$

考虑$1\le n<N$时，有$P\{T>n\}=1$

于是 $\sum_{k=1}^{N}(-1)^{k+1}\binom{N}{k}\left(\frac{N-k}{N}\right)^n=1$

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

$$\sum_{k=0}^{N}(-1)^{N-k-1}\binom{N}{k}(N-k)^n=0$$

取$j=N-k$，则有$\sum_{j=0}^{N}(-1)^{j-1}\binom{N}{j}j^n=0$

令随机变量D_n表示前n张收集的优惠券里的优惠券种类数，考虑$P\{D_n=k\}$

事件A表示：n张优惠券都是给定的k种优惠券之一

事件B表示：n张优惠券包含k种优惠券的每一种

则有 $P(A)=\left(\frac{k}{N}\right)^n$，$P(B|A)=1-P\{T_k>n\}=1-\sum_{i=1}^{k-1}\binom{k}{i}\left(\frac{k-i}{k}\right)^n(-1)^{i+1}$

于是 $P\{D_n=k\}=\binom{N}{k}P(AB)=\binom{N}{k}\left(\frac{k}{N}\right)^n\left[1-\sum_{i=1}^{k-1}\binom{k}{i}\left(\frac{k-i}{k}\right)^n(-1)^{i+1}\right]$

$$=\binom{N}{k}\sum_{i=0}^{k-1}\binom{k}{i}\left(\frac{k-i}{N}\right)^n(-1)^i$$

Probability and Statistics - P44

累积分布函数(cumulative distribution function)，又称分布函数

对于随机变量 X，有函数 $F: R \to [0,1]$

即 $F(x) = P\{X \leq x\}$，$x \in (-\infty, +\infty)$

于是有 $\forall_{x \in (-\infty,+\infty)}\ 0 \leq F(x) \leq 1$

对于任意实数 a, b，如果有 $a \leq b$

则事件 $\{X \leq a\}$ 包含于事件 $\{X \leq b\}$

于是有 $F(a) = P\{X \leq a\} \leq P\{X \leq b\} = F(b)$

即累积分布函数 $F(x)$ 为单调非降函数

概率分布列(probability mass function)，对于离散型随机变量 X

即随机变量 X 或者有有限个可能取值，或者有可数无限多个可能取值

有函数 $p(a) = P\{X = a\}$

如果随机变量的可能取值为 $X \in \{x_1, x_2, \cdots\}$

则有 $p(x) \begin{cases} \geq 0, & x \in \{x_1, x_2, \cdots\} \\ = 0, & \text{otherwise} \end{cases}$

且有 $\sum_{x \in \{x_1, x_2, \cdots\}} p(x) = 1$

期望 (expectation)，又称期望值(expected value)

对于离散型随机变量 X，有概率分布列 $p(x)$

则定义 X 的期望 $E[X] = \sum_{x: p(x) > 0} x\, p(x)$

对于指示器变量 $I(A)$，又称示性变量

有 $E[I] = 1 \cdot P\{A\} + 0 \cdot (1 - P\{A\}) = P\{A\}$

期望类似于物理中的重心(center of gravity)的概念

对于离散型随机变量 X，有概率分布列 $p(x)$，可能取值为 x_i，$i = 1, 2, 3, \cdots$

对于任意实值函数 $g(x_i)$，有 $E[g(X)] = \sum_i g(x_i) p(x_i)$

证明过程有：由于离散型随机变量 X 至多有可数无限多个可能取值

则 $g(x_i)$ 也至多有可数多个可能取值 y_j，$j = 1, 2, 3, \cdots$

$$
\begin{aligned}
\text{则有 } \sum_i g(x_i) p(x_i) &= \sum_j \sum_{i: g(x_i) = y_j} g(x_i) p(x_i) \\
&= \sum_j \sum_{i: g(x_i) = y_j} y_j\, p(x_i) \\
&= \sum_j y_j \sum_{i: g(x_i) = y_j} p(x_i) \\
&= \sum_j y_j\, P\{g(x_i) = y_j\} \\
&= E[g(X)]
\end{aligned}
$$

Probability and Statistics - P45

效用 (utility)，假设对于两种行动方案，会产生n种结果之一，$C_1, C_2, \cdots, C_n$

对于方案一，结果C_i的发生概率为p_i，其中 $i=1,2,\cdots,n$

对于方案二，结果C_i的发生概率为q_i，且$\sum_{i=1}^n p_i=1$，$\sum_{i=1}^n q_i=1$

如果令最好的结果为C，最坏的结果为c，

并对每种结果赋值，如$C=1$，$c=0$，$C_i\in[0,1]$

假设两种方式，直接获取结果C_i

或进行随机试验，以概率u得到C，以1-u得到c

则存在概率$u\in[0,1]$，使得$C_i = uC + (1-u)c$

记为$u(C_i)=u$

称无差别的概率值u为结果C_i的效用

则为了决定哪种行动方案是更优的

计算效用的期望 $E_1[u(C_i)] = \sum_i p_i u(C_i)$

$E_2[u(C_i)] = \sum_i q_i u(C_i)$

则当 $E_1[u(C_i)] > E_2[u(C_i)]$，即 $\sum_i p_i u(C_i) > \sum_i q_i u(C_i)$ 时

选择方案一，否则选择方案二

对于离散型随机变量X，有概率分布列$p(x)$，可能取值为x_i，$i=1,2,3,\cdots$

对于任意实数a, b，有$E[aX+b] = aE[X]+b$

证明过程有：$E[aX+b] =$ [redacted] $\sum_i (ax_i+b)p(x_i)$

$= \sum_{i:p(x_i)>0} ax_i p(x_i) + \sum_{i:p(x_i)>0} b\,p(x_i)$

$= a\sum_{i:p(x_i)>0} x_i p(x_i) + b\sum_{i:p(x_i)>0} p(x_i)$

$= aE[X]+b$

矩 (moment)，[redacted]指数学与统计学中对变量分布和形态特点的一组度量

对于离散型随机变量X，有概率分布列$p(x)$

有正整数n，记$\sum_x x^n p(x)$为随机变量X的n阶矩

可知n阶矩实际上是随机变量X的n次幂的期望$E[X^n]$

于是平凡地有，随机变量X的一阶矩 [redacted] (first moment) 即X的期望

如果将X^n替换为$(X-E[X])^n$，则称为X的中心矩 (central moment)

于是有 方差 (variance)：二阶中心矩 $\sum_x (x-E[X])^2 p(x)$

偏度 (skewness)：三阶中心矩 $\sum_x (x-E[X])^3 p(x)$

峰度 (kurtosis)：四阶中心矩 $\sum_x (x-E[X])^4 p(x)$

Probability and Statistics - P46

方差 (variance)，对于随机变量 X，期望 $E[X]=\mu$，记 $Var(X)$ 为随机变量 X 的方差

定义为 $Var(X) = E[(X-\mu)^2] = \sum_x (x-\mu)^2 p(x)$

$$= \sum_x (x^2 - 2x\mu + \mu^2) p(x)$$

$$= \sum_x x^2 p(x) - 2\mu \sum_x x p(x) + \mu^2 \sum_x p(x)$$

$$= E[X^2] - (E[X])^2$$

另外 [redacted] 对于任意实数 a, b，有随机变量 $aX+b$

则 $Var(aX+b) = E[(aX+b-a\mu-b)^2]$

$$= E[a^2(X-\mu)^2] = a^2 Var(X)$$

注意期望与物理学中质量分布的重心类似

而方差与物理中的力学惯性矩 (moment of inertia) 类似

标准差 (standard deviation)，对于随机变量 X，期望 $E(X)=\mu$，方差为 $Var(X)$

则方差的平方根称为标准差，记为 $SD(X) = \sqrt{Var(X)}$

伯努利试验 (Bernoulli trial)，又称伯努利随机变量 (Bernoulli random variable)

表示只有两种可能结果 (1/0，T/F，成功/失败) 的单次随机试验

对于随机变量 X，令 $X = \begin{cases} 1 & \text{当试验结果为成功时} \\ 0 & \text{当试验结果为失败时} \end{cases}$

则有 X 的概率分布列 $\begin{cases} p(0) = P\{X=0\} = 1-p \\ p(1) = P\{X=1\} = p \end{cases}$

其中实数 $0 \le p \le 1$ 表示每次试验成功的概率

二项分布 ([redacted] binomial distribution)，又称二[redacted]项随机变量 (binomial random variable)

对于成功率为 p 的伯努利试验，进行 n 次独立重复试验，令随机变量 X 为其中成功的次数

则称 X 服从参数为 (n, p) 的二项分布，记为 $X \sim B(n, p)$

并有概率分布列 $p(i) = \binom{n}{i} p^i (1-p)^{n-i}$，$i = 0, 1, 2, \cdots, n$

注意到随机变量 X 的 k 阶矩，$k = 1, 2, \cdots$

$$E[X^k] = \sum_{i=0}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=1}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i}$$

$$= \sum_{i=1}^{n} i^{k-1} \cdot i \binom{n}{i} p^i (1-p)^{n-i}$$

$$= \sum_{i=1}^{n} i^{k-1} \cdot n \binom{n-1}{i-1} p^i (1-p)^{n-i}$$

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

$$= np E[(Y+1)^{k-1}]$$

即有随机变量 Y 服从参数 $(n-1, p)$ 的二项分布，$E[(Y+1)^{k-1}]$ 为 Y 的 $k-1$ 阶矩

Probability and Statistics - P47

二项分布

对于离散随机变量X，服从参数为(n, p)的二项分布，$n\in N^+$，$0\le p\le 1$

则有概率分布列 $p(i)=\binom{n}{i}p^i(1-p)^{n-i}$，$i=0,1,2,\cdots,n$

有k阶矩 $E[X^k]=npE[(Y+1)^{k-1}]$，$k\in N^+$

其中Y为服从参数(n-1, p)的二项离散随机变量

于是二项随机变量的期望为

$$E[X]=npE[(Y+1)^0]=npE[1]=np$$

二项随机变量的方差为

$$\begin{aligned}Var(X)&=E[X^2]-(E[X])^2=npE[(Y+1)^1]-(E[X])^2\\&=npE[Y]+npE[1]-(E[X])^2\\&=np\cdot(n-1)p+np-(np)^2\\&=np(1-p)\end{aligned}$$

对于参数为(n, p)的二项随机变量X，其中$0<p<1$，$n\in N^+$

则当k的取值从0到n时，$Pr\{X=k\}$首先单调递增，然后单调递增

当$k=\lfloor(n+1)p\rfloor$时取得最大值，

证明过程有，对于$k=1,2,\cdots,n$，考虑$Pr\{X=k\}/Pr\{X=k-1\}$

根据比值与1的关系判断递增/递减

$$\begin{aligned}\text{则 } Pr\{X=k\}/Pr\{X=k-1\}&=\binom{n}{k}p^k(1-p)^{n-k}\Big/\binom{n}{k-1}p^{k-1}(1-p)^{n-k+1}\\&=\left(\frac{n!}{k!(n-k)!}\cdot p\right)\Big/\left[\frac{n!}{(k-1)!(n-k+1)!}\cdot(1-p)\right]\\&=(n-k+1)p\,/\,k(1-p)\end{aligned}$$

于是 $Pr\{X=k\}\ge Pr\{X=k-1\}$ 当且仅当 $(n-k+1)p/k(1-p)\ge 1$

即 $(n-k+1)p\ge k(1-p)$

$k\le(n+1)p$

又 $k\in N^+$，$n\in N^+$，实数 $0<p<1$，

于是当 $k\le\lfloor(n+1)p\rfloor$ 时，$Pr\{X=k\}\ge Pr\{X=k-1\}$

则 $Pr\{X=k\}$ 在 $k=\lfloor(n+1)p\rfloor$ 时取得最大值

对于参数为(n, p)的二项随机变量X，其中$0<p<1$，$n\in N^+$

计算分布函数 $Pr\{X\le k\}=\sum_{i=0}^{k}\binom{n}{i}p^i(1-p)^{n-i}$，$k=0,1,\cdots,n$

由于 $Pr\{X=k\}/Pr\{X=k-1\}=(n-k+1)p/k(1-p)$

则可以递归地定义概率分布列

$$p(i)=\begin{cases}(1-p)^n, & i=0\\ \frac{p}{1-p}\cdot\frac{n-i+1}{i}\cdot p(i-1), & i=1,2,\cdots,n\end{cases}$$